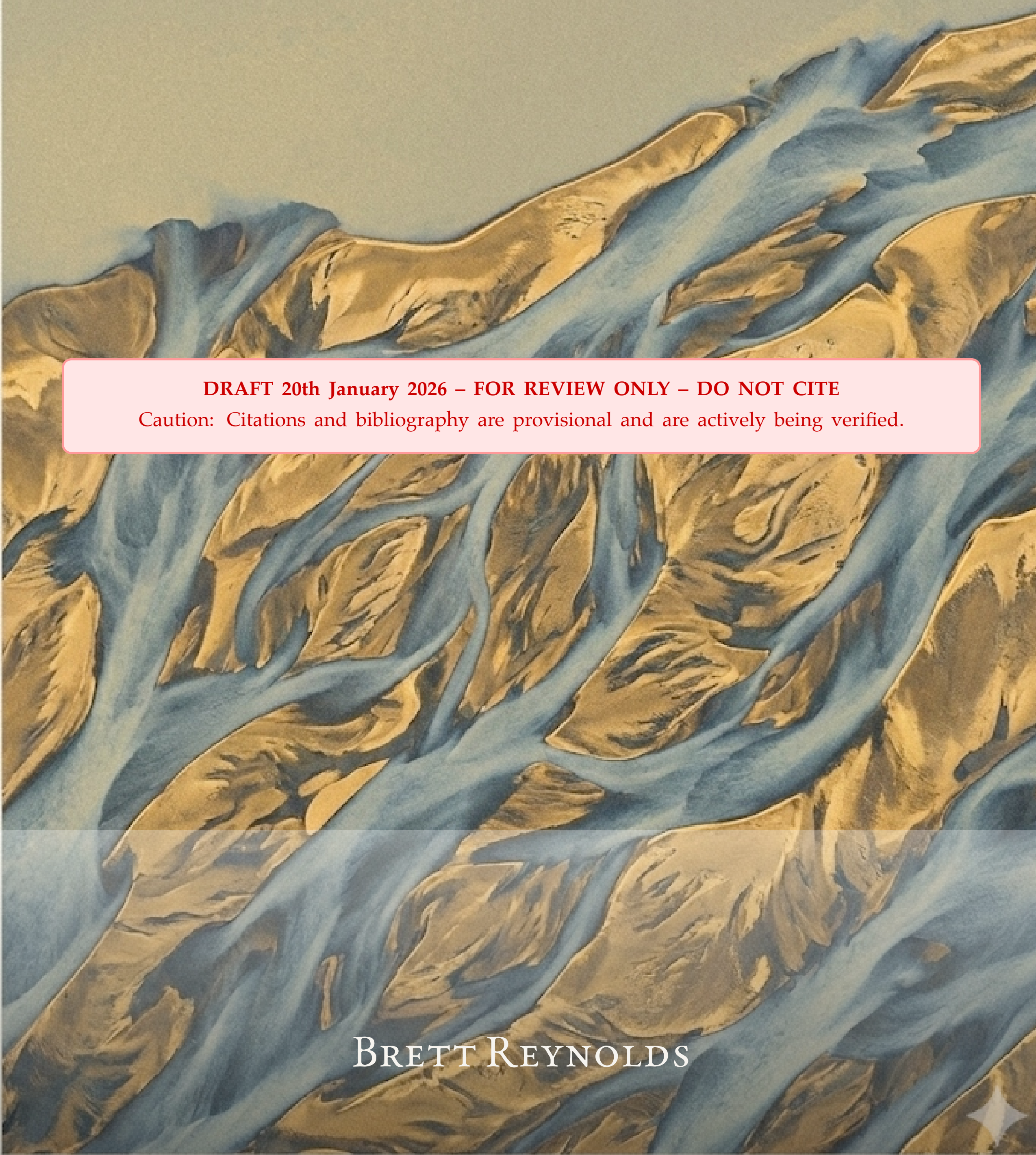
Words That Won't Hold Still

How Linguistic Categories Work



DRAFT 20th January 2026 – FOR REVIEW ONLY – DO NOT CITE

Caution: Citations and bibliography are provisional and are actively being verified.

Brett Reynolds

Words That Won't Hold Still: How Linguistic Categories Work

Brett Reynolds

20th January 2026

Words That Won't Hold Still: How Linguistic Categories Work

Draft version: 20th January 2026

Typesetting

This book was typeset using XeLaTeX with the following principal components:

- Body text: EB Garamond, designed by Georg Duffner after Claude Garamond
- IPA and extended Latin: Charis SIL, a font designed for linguistics
- Document class: `book` (standard LaTeX)
- Linguistic examples: `langsci-gb4e`
- Bibliography: BibLaTeX with APA style

Acknowledgments

This book was written with substantial assistance from large language models, including Claude Opus 4.5 (Anthropic), ChatGPT 5.2 (OpenAI), and Gemini 3 (Google). The author remains responsible for all claims, arguments, and errors.

Preface

This book offers a new way of thinking about linguistic categories – what they are, how they persist, and why they have the structure they do. The central claim is that categories like NOUN, DEFINITE, and COUNTABLE aren't defined by essential properties but maintained by mechanisms: frequency effects, functional pressures, transmission dynamics, and cognitive anchoring. Categories are real, but their reality is dynamic rather than definitional.

Framework and scope

For English examples, I work within the framework of *The Cambridge Grammar of the English Language* (Huddleston & Pullum, 2002) – its categories, its terminology, and its analyses. This isn't a defense of *CGEL* over competing grammars; it's a consistent baseline that allows claims to be tested against a well-articulated descriptive system. When I say that *otherwise* straddles categories, I mean *CGEL*'s categories. Readers who prefer a different framework can substitute accordingly; the theoretical argument doesn't depend on any particular grammatical analysis.

The theoretical claims, though, are intended to generalize beyond English. The mechanisms that maintain categories – entrenchment, transmission, functional anchoring – aren't language-specific. Where I draw on cross-linguistic data (from Welsh, Spanish, Korean, and others), I rely on the sources cited rather than claiming expertise in those languages.

What this book is not

This isn't a grammar. It doesn't aim to describe English (or any other language) exhaustively. It isn't a typology, and it makes no claims about how many languages exhibit particular properties. It isn't a defense of any grammatical framework – generative, functional, cognitive, or otherwise.

What it is, instead, is an argument about *what kind of thing* a linguistic category has to be for the evidence we have to make sense. The book sits at the intersection of linguistics, philosophy of science, and cognitive science. It asks what categories are made of, and it answers: mechanisms, not essences.

Audience

I've tried to write for linguists who are curious about the metaphysics underlying their practice, and for philosophers who are curious about how the abstract questions play out in a concrete domain. Some background in linguistics will help – familiarity with basic grammatical terminology, with the idea that categories can be disputed, with the general shape of debates about universals and variation. But I've aimed to make the argument accessible to anyone willing to follow the examples carefully.

Mississauga, January 2026

Typographical Conventions

This book uses the following typographical conventions:

LINGUISTIC EXAMPLES

- *Italics* mark linguistic forms under discussion: *the*, *furniture*, *who*.
- SMALL CAPITALS introduce technical terms at their point of definition: DESIGNATUM, HOMEOSTATIC PROPERTY CLUSTER.
- * marks ungrammatical strings: **She have left*.
- ? marks marginally acceptable or degraded strings: ?*The team has scored on themselves*.
- ?? marks severely degraded strings.

QUOTATION MARKS

- Double quotes ("like this") enclose direct quotations and reported speech.
- Single quotes ('like this') enclose meanings or glosses: Spanish *patatas* 'potatoes'.
- Scare quotes are used sparingly, to signal that a term is being used with reservations.

EXAMPLES

Numbered examples follow standard linguistic convention. Sub-examples are lettered (a, b, c). Interlinear glosses follow the Leipzig Glossing Rules.

BIBLIOGRAPHY

Citations follow APA style. In-text citations use the author-date format: Huddleston and Pullum (2002) or (Huddleston and Pullum, 2002). Full references appear in the References section at the end of the book.

for Rodney and Geoff

Acknowledgements

I am grateful to Peter Evans for his careful copy-editing and insightful comments.

Muhammad Ali Khalidi provided valuable feedback on the application of natural-kinds theory to linguistic categories, clarifying the relationship between mechanisms, stabilizers, and causal structure, and validating the approach to field-relative projectibility.

Contents

List of Figures

Part I

The Problem

Words That Won't Hold Still

On March 2, 2008, at 3:13 in the morning, I got an email from Rodney Huddleston about a word he didn't understand.

The time stamp is misleading – Huddleston was writing from Australia, where it was a reasonable hour. But there's something fitting about the image of a linguist awake in the dark, wrestling with a single word. The word was *otherwise*. I had asked him whether there were grounds for treating it as a preposition rather than an adverb. His reply:

> I'm not proud of the adverb analysis, or confident about it, and don't intend it to cover all uses. Its classification is quite a puzzle. Dictionaries have it as adverb and adj (*the truth is quite otherwise*) and some also as conjunction. There is something to be said for a prep analysis, which might cover adjunct and predicative complement uses. But I don't know how to handle *this suggests otherwise* or *the correctness or otherwise of the proposal*.

Huddleston was not a careless analyst. He was the lead author of *The Cambridge Grammar of the English Language* (CGEL) (Huddleston & Pullum, 2002) – 1,860 pages, seventeen years in the making, the most comprehensive descriptive grammar of English ever produced. If anyone knew what *otherwise* was, it should have been him.

But he didn't. Not because he had missed something. Because the word wouldn't hold still.

This book is about what that puzzle reveals.

The striking thing isn't that words misbehave. It's that they misbehave in ways that persist, replicate across speakers, and survive centuries. The puzzle isn't classification. The puzzle is stability – why the instability is itself stable.

The natural reading is that *otherwise* is simply a hard case – an outlier, a word with an unusual history, a problem for specialists. Every grammar has its edge cases. You note them, flag the uncertainty, and move on.

But *otherwise* is not alone. The same kind of stability shows up at every scale.

Cattle takes plural agreement (*the cattle are grazing*) but has no singular. You can say *many cattle* (and in agricultural contexts, *twenty cattle*), but not *a cattle* or *one cattle*. Yet it readily functions as a modifier in compound nouns like *a cattle ranch*, a position usually reserved for singulars (*a toothbrush*, not **a teethbrush*). The usual singular–plural paradigm simply doesn't apply. And it has been this way for centuries.

Weak and generic uses of *the* (*go to the hospital*, *the tiger is endangered*) resist the standard definition of definiteness. Part III returns to definiteness; for now, note only that grammars describe these uses without explaining how they fit the category.

Constructions can coerce category behaviour. *I had to think myself through that argument* is attested and interpretable by analogy to *talk oneself through* and *work oneself through*; but *think* is ordinarily intransitive. Here a construction forces an intransitive verb into a caused-motion frame, and speakers differ about whether the result belongs.

Even phoneme boundaries can be stable and conditional: /ɪ/ and /ɛ/ distinguish *pin* from *pen* for many speakers, but the contrast collapses before nasals in several varieties.

And at the social end, register and honorific systems impose boundaries that are real in practice yet gradient in uptake: *gonna* can be ordinary in casual speech and jarring in academic prose; in honorific systems the choice between plain and polite forms carries systematic inferences.

Two broad responses recur in the literature. Each captures something important.

The first says: categories are defined by necessary and sufficient conditions. A noun is whatever satisfies the necessary and sufficient conditions for nounhood; an adverb is whatever satisfies the conditions for adverbhood. Membership is binary. Boundaries are sharp. If a word doesn't fit, either the criteria are wrong or the word is exceptional. Refine the criteria, explain away the exceptions, and the system will be clean.

This is the essentialist view – or at least the textbook version of it, sometimes called "classical" or "Aristotelian" (whether or not Aristotle himself held quite this position). It has the virtue of clarity: either something is an X or it isn't. That clarity has built the infrastructure of modern linguistics – great descriptive grammars, parsers, textbooks, annotation schemes all rest on the assumption that categories have definitions. The recurring difficulty is that the definitions keep encountering exceptions. Every set of criteria produces counterexamples. The counterexamples get handled by stipulation, by subclasses, by "special" readings that multiply until the exceptions rival the rules. Huddleston's email is the essentialist view confronting its limits: here is a word, here are the criteria, and no combination of criteria delivers a stable answer.

The second response says: categories are not defined by conditions at all. They are PROTOTYPES – clusters of typical features, with central members and peripheral members and fuzzy boundaries all the way down. A robin is a better example of BIRD than a penguin is, but both are birds. *Run* is a better example of VERB than *beware* is, but both are verbs. Stop expecting sharp edges. gradience is the nature of the beast.

This view also captures something real – the empirical fact of gradience, the persistent failure of neat definitions. But as it is usually deployed in linguistics, it purchases descriptive adequacy at the cost of explanation. If *cattle* is a "less central" noun, why does its non-centrality take exactly the form it takes – no singular, plural agreement, full compatibility with numerals above one? Why has it been stable for five hundred years instead of drifting toward the core or out of the category entirely? Why don't categories dissolve into chaos if their boundaries are genuinely fuzzy? Prototype descriptions record the gradience. They don't, by themselves, explain why the gradient structure holds still.

The essentialist locates boundaries in definitions that don't exist. The prototype theorist denies that boundaries are sharp at all. Both share an assumption so deep it's almost invisible: that these are the only options. Either categories have essences, or they're looser groupings – useful for description, perhaps, but not the kind of thing that could bear explanatory weight.

A clarification before we proceed. The problem with essentialism is not that it posits sharp boundaries and binary membership – I'll argue later that we can have both, produced by mechanisms rather than definitions. Likewise, the virtue of prototype theory is not that it discovers genuine fuzziness in category structure – I'll suggest that much apparent fuzziness is epistemic rather than ontological. The virtue is that prototype theory took gradience seriously as data. The impasse we're diagnosing is not a disagreement about sharpness. It's a shared failure to ask what maintains the structure.

A significant exception is found in the work of Ray Jackendoff and colleagues on the PARALLEL ARCHITECTURE (PA), which seeks to ground linguistic theory in a broader cognitive and evolutionary context (Jackendoff, 2002, 2023).

PA rejects the "syntactocentric" view of grammar, proposing instead that phonology, syntax, and semantics are independent, parallel generative systems linked by interface rules. This move shifts the burden of explanation from the internal structure of a single component to the correspondences between them.

There is a third possibility.

What if categories are real, stable, and explanatorily powerful – but not because they have essences? What if their boundaries are sharp in structure but gradient in access – sharp enough to support induction, yet appearing fuzzy because we can't locate them precisely? What if the stability and the apparent fuzziness are both consequences of something else, something neither tradition has squarely addressed?

What if the mechanisms that usage-based linguists, variationists, and typologists already study are not merely causes of category behaviour but constitutive of category reality itself? What if "refactoring" that research around an explicit ontology of mechanism-maintained kinds would resolve puzzles that each tradition, working alone, has left open?

Huddleston's email sits in my files, seventeen years old now. "Its classification is quite a puzzle." The puzzle was never just *otherwise*. It was what kind of thing a grammatical category has to be for that sentence to be exactly the right thing to say.

1.1 Essentialism examined

Chomsky (1965), in *Aspects of the Theory of Syntax*, crystallized a picture many linguists had already been working with: a grammar as a finite stock of sharply bounded categories, each lexical item assigned to exactly one major category (or handled via homonymy), with syntactic theory tasked with discovering the conditions that fix those assignments once and for all. This picture brought order: it made syntactic theory look like a matter of discovery rather than stipulation, and for decades that's how linguistics proceeded.

This assumption – that each category has an *essence*, a set of properties necessary and sufficient for membership – embodies what philosophers call ESSENTIALISM, and it's been linguistics' default mode for millennia. On this view, NOUN and VERB are fixed and universal because they track real essences, not because linguists have stipulated them. Essentialism is the natural view. It's where you start if no one has told you there's a problem.

The intuition runs deep. When a child learns that *dog* picks out the class of dogs, we're tempted to think that there's some inner dogness – some cluster of necessary properties – that all and only dogs share. Borderline cases exist, of course, but we think of them as fringe matters. The real work of the concept is done by its core, and the core is fixed by an essence. The same picture quietly informs how linguists talk about *phonemes*, about *nouns* and *adjectives*, about *definiteness*, about *subject* and *object*. There are different kinds of sounds, different kinds of words, different kinds of grammatical functions, and the grammar cares about the difference.

Applied to language, this yields a picture that most working linguists absorb without being taught it explicitly. A noun (in English, at least) is whatever satisfies the criteria for nounhood – it inflects for number, it heads noun phrases, it can be modified by adjectives. A verb is whatever satisfies the criteria for verbhood – it inflects for tense, it takes arguments, it heads verb phrases. The criteria might be debated; the assumption that there *are* criteria, that categories have definitions, is rarely questioned. Even linguists who reject essentialism in principle often write as if it were true, because the alternative is hard to operationalize.

The essentialist picture has genuine explanatory power. The great descriptive grammars – Jespersen (1924), Quirk et al. (1985), *The Cambridge Grammar* itself (Huddleston & Pullum, 2002, ch. 1) – proceed category by category, laying out the membership criteria, cataloguing the members, noting the exceptions. The exceptions are always there, but they're handled as exceptions: marginal cases, historical residue, items in transition. The core of each category is secure. Textbooks are organized around this architecture. Pedagogical grammars depend on it. Parsers are built on it. The infrastructure of linguistic analysis presupposes that categories have boundaries and that the boundaries can, in principle, be found.

The strongest modern version of essentialism grounds categories in substantive theoretical claims about what the categories *are*. Baker (2003) provides an explicit defense of this position. Baker's goal is to redeem what he calls the "longstanding promissory note" of the feature system [+N] and [−V] – that is, to provide substantive syntactic definitions of the major lexical categories that hold universally (Baker, 2003, p. 17). Nouns, on his account, bear referential indices; verbs license specifiers; adjectives predicate properties. These are not tendencies or prototypes but defining properties, and they hold because of what nouns, verbs, and adjectives essentially are. The features don't just describe the categories; they flow from the categories' natures.

The same essentialist logic extended to semantics. Jerrold Katz, working in the 1960s and 70s, defended the classical theory of lexical concepts: each word's meaning is defined by a set of necessary and sufficient semantic features (Katz, 1972). The standard example was *bachelor*, analyzed as [+adult] [+male] [+unmarried]. Anything satisfying those features is a bachelor; anything lacking one isn't. Katz treated such decompositions as cognitively real – part of what speakers know when they know a word's meaning. The category BACHELOR has an essence: unmarried adult maleness. The definition carves the concept at its joints.

But much work in formal linguistics uses feature bundles without Baker's or Katz's deeper commitments. Categories are specified as [+N, −V] or [−N, +V], but the features are treated as primitives – tools for stating generalizations, not reflections of underlying essences. On this view, the feature bundle *is* the category; there's no further question about why these features cluster. Whether this counts as essentialism is less clear. The boundaries may be sharp, but the explanation stops at the features themselves. I'll return to this distinction in the next section, where it turns out to matter.

The problem with essentialism isn't that it's wrong about everything. It's that it keeps encountering cases where the machinery doesn't fit – and the response is always the same. Add a stipulation. Create a subclass. Mark the item as exceptional. The exceptions accumulate, and at some point you have to ask whether the machinery is doing explanatory work or merely recording the failures of a prior commitment.

Return to *otherwise*. Huddleston's puzzlement wasn't a lapse but a report from the front lines. The word looks like an adverb (*think otherwise*), looks like an adjective (*the truth is quite otherwise*), and appears in constructions that resist any standard label (*the correctness or otherwise of the proposal*). What are the necessary and sufficient conditions for adverbhood that include *otherwise* in *think otherwise* and exclude it in *the correctness or otherwise*? There aren't any. Every definition either lets in too much or leaves out too much.

The essentialist response is predictable: *otherwise* is polysemous, multiply listed, historically complex. Each use gets its own entry, its own feature specification, its own subcategorization frame. The grammar doesn't fail; it proliferates. But this is bookkeeping, not explanation. We wanted to know what makes something an adverb. We got a list of contexts where *otherwise* behaves adverbially, and another list where it doesn't, and no principled account of why the word straddles the boundary instead of sitting on one side or the other.

Consider the broader pattern. In English, the major categories – noun, verb, adjective, adverb, preposition – are supposed to be definable by distributional and morphological criteria. But every category has items that meet some criteria and fail others:

Fun looks like a noun (*we had fun*) but increasingly takes degree modification like an adjective (*that was very fun*). The change is ongoing, and speakers differ.

What is *fun* – noun, adjective, or in transit between them? The essentialist has to say it's one or the other, perhaps with a secondary listing for the emerging use. But the apparent transition shows that the word is moving through a region that essentialism's static picture can't accommodate – a space between basins where competing mechanisms pull in different directions.

Near takes objects without a preceding *to* (*near the house*), which makes it look like a preposition. But it also has comparative and superlative forms (*nearer*, *nearest*), which makes it look like an adjective. It can function as a predicative complement (an expression that ascribes a property to the subject or object) after *become* – but only in its affectionate sense (*become near and dear to someone*), not its locative sense (**become near the door*). Some grammars call it a preposition; some call it an adjective; some, including *CGEL*, call it both – a word that belongs to two categories. But if categories are defined by necessary and sufficient conditions, how can a single word satisfy two incompatible definitions (unless distinct homonyms are posited)? The answer is that the definitions weren't necessary and sufficient after all. They were family portraits, not passport photos – and *near* has the look of both families.

Compare *pease*, which until Early Modern English was a mass noun like *cattle*. Speakers reanalyzed the final *-se* as plural *-s*, back-formed a singular *pea*, and the word regularized. No functional anchor blocked the analogy. *Cattle* has *cow*; *pease* had nothing. The essentialist account says both words should have stable essential properties. But one regularized and one didn't, and the difference isn't in their essences – it's in whether another word was occupying the functional slot that analogy would have filled.

The deeper problem is that the failures aren't random. They cluster in predictable places.

The intersection zones between major categories are chronic trouble spots. The preposition–adverb boundary, the complement–modifier boundary, the agent–experiencer boundary – these are not marginal curiosities. They're systematic sites of instability. And the words that straddle these boundaries tend to be high-frequency items with broad distributions: *otherwise*, *fun*, *near*. Frequency isn't a feature; distribution is a consequence of how a word behaves, not a defining criterion. These are precisely the patterns a theory of categories has to explain. These are precisely the zones where, on the view I'll defend, the forces maintaining the cluster are weakest or in competition.

Items undergoing historical change present a special problem. If categories have essences, then change should be abrupt: a language either has subject-auxiliary inversion or it doesn't. But English inversion changed gradually. *Go you to* *church?* and *I go not* gave way to *Do you go to church?* and *I don't go* – but not all at once. For centuries, both patterns coexisted, distributed by register, construction type, and verb class. Shakespeare inverts lexical verbs in questions; a generation later, the pattern is receding. Essentialist grammars have to treat each stage as a discrete system, missing the continuity that drove the change. The alternative – acknowledging that syntactic properties can be variably present across a grammar – is precisely what essentialism forbids.

None of this means essentialism is useless. For most words, most of the time, the category labels work. *Dog* is a noun; *run* is a verb; *quickly* is an adverb. The labels support generalization. They organize grammars. They enable pedagogy. If every word were like *otherwise*, linguistic analysis would be impossible.

But neither is every word like *dog*. And the question is what to make of the words that aren't. The essentialist answer – treat them as exceptions, lexical idiosyncrasies, marginal cases – works as long as the exceptions stay marginal. When the exceptions multiply, when the boundaries fray systematically, when the core cases start looking like special instances of a more general pattern of variation, then the answer stops working. The framework designed to handle the clear cases has nothing to say about the unclear ones, except that they're problems to be solved later, with better definitions.

This is where essentialism stands. It handles the core. It fumbles the periphery. And it has no account of why the periphery exists at all – why categories should *appear* to have fuzzy edges when nothing in the definition explains the gradient structure.

The examples so far have focused on word-level categories – nouns, verbs, prepositions – because they're familiar and the puzzles are readily visible. But the challenge generalizes across linguistic structure. Phonemes, construction types, inflectional classes, thematic roles, speech acts, information-structure categories: all present the same problem of clustering with variation, stability with gradience, projectible patterns that resist sharp definitions. This book examines linguistic categories across all levels of analysis.[1] The framework developed here applies equally to a phonological segment, a syntactic phrase type, or a discourse marker.

[1] I use CATEGORY in the philosopher's sense: a causally sustained kind whose boundaries and internal structure are maintained by mechanisms. This flattens *CGEL*'s distinctions among categories (noun, NP), systems (tense, number), and functions (subject, modifier). Those distinctions are analytically indispensable, but from the vantage point of asking what holds a kind together, all three pose the same question. Fellow *CGEL* partisans: forgive the flattening.

1.1.1 Essentialism beyond formalism

The impulse isn't confined to formalist or generative traditions. Even frameworks that foreground gradience and usage often smuggle essences back in. Cognitive grammarians like Langacker (1987) reject autonomous syntax and embrace prototypes but treat the noun/verb opposition as a universal, notional distinction grounded in essential cognitive abilities for conceptualizing things versus processes. Functional grammarians in Dik's tradition (Dik, 1997) allow that not every language has a *subject*, but hold that once a language does, the function comes with universal properties and obeys cross-linguistic hierarchies. Role and Reference Grammar, explicitly typological and semantically driven, posits exactly two macroroles (broad semantic categories that subsume finer-grained roles like agent and patient) – Actor and Undergoer – as a "fundamental opposition" underlying clause structure everywhere. The details differ, but the metaphysical shape is familiar: a small, privileged stock of basic categories whose natures are fixed in advance and which languages are assumed to instantiate.

Essentialism survives across theoretical divides. It continues to guide practice even where explicit talk of necessary and sufficient conditions has faded. It's still the default assumption: that categories, if they're to be real and explanatory, have to have sharp boundaries and that fuzziness, where it appears, is a defect in our descriptions.

But the pragmatic compromise runs deeper than most practitioners acknowledge. Working linguists proceed *as if* categories have sharp boundaries even when they privately doubt that the boundaries are real. This isn't intellectual dishonesty; it's a rational response to the demands of description and analysis. Grammars need to be written. Students need to be taught. Parsers need to be built. All of these tasks become vastly simpler if you can assume that *fun* is either a noun or an adjective, that *near* is either a preposition or an adjective, that *otherwise* has a stable category assignment. The alternative – treating every item as occupying a unique position in continuous category space – would make grammatical description intractable.

The compromise shows up in how grammarians talk about their own methods. They'll note that a particular classification is "not entirely satisfactory" or that "the boundaries are somewhat fluid", then proceed to use the classification anyway because the architecture of the grammar requires it. They'll acknowledge gradient membership in one paragraph and then, two pages later, write rules that presuppose binary distinctions. Far from being sloppy, this practice recognizes that some idealizations are necessary for the work to proceed, even when the idealizations are known to be false.

Some grammarians make this tension explicit.[2] They distinguish the metaphysical implausibility of sharp essences from the practical convenience of talking as if such essences existed. Essentialism becomes a shared fiction – useful for communication, indispensable for pedagogy, but not to be confused with a claim about what grammatical categories actually are.

The question this raises is whether there's an alternative. Can we build a picture of categories that accommodates the gradience, the flux, the systematic misbehavior of high-frequency items and boundary cases – but still manages to be tractable, explanatory, and usable in grammatical description? Or are we stuck with the choice between an essentialist picture we know to be false and a gradient picture too unwieldy to operationalize?

1.2 Prototype theory examined

The prototype tradition began as a rebellion against definitions. Psychologists documented the gradient structure that classical categories were not supposed to have; linguists seized on their results as licence to relax sharp boundaries; and cognitive scientists built a research programme around the claim that fuzziness was not a defect in our descriptions but a fact about how concepts work.

In the early 1970s, the psychologist Eleanor Rosch ran a series of experiments that should have been impossible on the classical view (E. Rosch, 1975). She asked subjects to rate how good an example each item was of everyday categories: BIRD, FRUIT, FURNITURE, VEHICLE. If categories were defined by necessary and sufficient conditions, the question would be meaningless. A robin either is or isn't a bird; there's no sense in which it could be a better bird than a penguin.

But subjects had no trouble answering. They rated robins as better birds than penguins, apples as better fruit than olives, chairs as better furniture than rugs. The ratings were consistent across subjects and stable across time. People had robust intuitions about category structure that the classical theory said they shouldn't have.

Rosch's interpretation: categories aren't defined by conditions. They're organized around PROTOTYPEs, central members that best exemplify the category. Other members are included by similarity to the prototype, with membership grading off toward the periphery. A robin is a better bird than a penguin be-

[2] For example, Geoffrey K. Pullum (personal communication) writes: "I am well aware that I am (for pragmatic reasons) presupposing an essentialist mindset that assumes 'Categories exist; items either belong or don't; our job is to discover the boundaries', and that such essentialism might well be wrong. I have thought that for quite a long time. The main pragmatic reason for maintaining the implausible essentialism in question is just that it's useful to be able to talk to linguists in a language they understand."

cause a robin is closer to the prototype; a penguin is still a bird, but a marginal one. The boundaries aren't sharp lines but fuzzy gradients.

The idea spread fast. By the 1980s, prototype theory had become a major framework in cognitive psychology, and linguists were importing it wholesale. George Lakoff's *Women, Fire, and Dangerous Things* (Lakoff, 1987) argued that prototype structure was fundamental to human cognition, not a quirk of folk categories but the way concepts work. John Taylor's *Linguistic Categorization* (Taylor, 2003) applied the framework systematically to grammatical categories. The cognitive linguistics movement, associated with Lakoff, Ronald Langacker, and others, built an entire research programme around the idea that linguistic categories are prototype-structured, gradient, and grounded in embodied experience.

The appeal was obvious. Here, finally, was a framework that took the evidence seriously. The messy cases that essentialism swept under the rug (the *fun*s and *near*s and *otherwise*s) became data rather than noise. If category membership is gradient, then words can be more or less noun-like, more or less verbal, more or less adverbial. The boundaries don't have to be sharp because the theory doesn't predict sharp boundaries. The evidence that embarrassed essentialism confirmed prototype theory.

1.2.1 Feature bundles: essentialism in disguise

While cognitive psychologists were developing prototype theory, formal linguists were arriving at a similar place by a different route: encoding categories as feature matrices rather than distances from a prototype, treating the features as discrete rather than graded, but running into the same problem when items turned out to have some features of a category without having all of them.

The feature-bundle approach I flagged in the previous section – categories as specifications like [+N, −V] – looks essentialist on the surface. The features are discrete. Category membership appears binary: either an item has the feature or it doesn't. But once you allow that items can have *some* features of a category without having all of them, gradience enters through the back door. A word that is [+N] for agreement but [−N] for pluralization is partly noun-like. The prototype theorist would say it's a marginal noun, distant from the prototype. The feature theorist would say it has an atypical feature specification. These are notational variants of the same observation.

The deeper convergence is methodological. Prototype theory says: categories are organized around central members, and membership is determined by similarity to those members. Feature-bundle approaches say: categories are defined by feature matrices, and membership is determined by how many features an item has. But where do the prototypes come from? And where do the features come from?

Neither framework has a principled answer. Prototypes are typically identified by introspection or experiment – these items *feel* central. Features are typically posited to capture distributional patterns – items that pattern together get assigned shared features. Each approach takes its primitives as given, and the two inform each other in something like reflective equilibrium (Goodman, 1955): features tell you what dimensions similarity should be computed over; typicality judgments tell you which features are doing work.

This is where I must locate my own past practice. For most of my career, I operated comfortably in the feature-bundle tradition.[3] I was not committed to Platonic essences, but I took *CGEL*'s categories to be real structural kinds that could be characterized by clusters of discrete properties, and I assumed that the clustering itself required no further explanation. If pressed, I would have acknowledged gradience; I would have granted that *fun* was a marginal noun (or a marginal adjective). But I wouldn't have asked *why* that particular configuration of properties, *why* stable for decades, *why* not drifting toward the core or out of the category entirely. The clustering was a fact to be recorded, not a phenomenon to be explained.

Prototype theory and feature-bundle approaches turn out to be two faces of a single tradition – one psychological, one formal, but sharing a crucial limitation. Both describe the gradience. Neither explains it.

1.2.2 Prototypes in the grammar

Consider lexical categories. The essentialist says: a word is a noun if and only if it has the necessary and sufficient properties that define nounhood. The prototype theorist says: *nounhood* is a cluster of properties (reference to things, inflection for number, occurrence in certain syntactic frames) and words exhibit these properties to varying degrees. A word like *dog* exhibits all of them (in its nominal use); it's a prototypical noun. A word like *fun* exhibits some but not others; it's a marginal noun, or a noun-becoming-adjective, or a word in the gradient space between categories. The labels become approximations rather than verdicts.

Consider syntactic constructions. The essentialist says: a clause is grammatical or it isn't. The prototype theorist says: constructions have central instances and extended instances, with acceptability shading off toward the margins. *The*

[3] See, for example, Reynolds (2021), where I encode determinatives and pronouns as binary feature matrices and use clustering algorithms to validate *CGEL*'s category assignments. The methodology treats property clustering as evidence for category validity without asking what maintains the clustering – exactly the explanatory gap that HPC addresses.

dog bit the man is a prototypical transitive; *The bed slept two people* is an extended transitive, coerced by analogy to the prototype; *The stone kicked the ball* is further out still, requiring more contextual support to be acceptable. Grammaticality becomes a gradient, not a boundary.

Consider meaning. The essentialist says a word has a definition – a set of conditions that determine what it applies to. The prototype theorist says word meanings are organized around exemplars, with extension to new cases governed by similarity rather than rule. *Bird* doesn't mean "feathered bipedal vertebrate capable of flight"; it means something like "thing similar to robins, sparrows, eagles" – and penguins count because they're similar enough, while bats don't, despite meeting some featural criteria.

This framework captured something real. The gradient judgments Rosch documented are genuine: speakers do have intuitions about better and worse examples, and categories do have internal structure. The essentialist picture – clean definitions as the ground for sharp boundaries and binary membership – failed in its foundational claim, and prototype theory said so clearly.

1.2.3 But there's a cost

Neither approach tells you why the boundaries are *there* rather than *elsewhere*. Prototype theory tells you that categories have fuzzy boundaries; feature-bundle approaches tell you which properties cluster. Neither tells you why the periphery has the structure it has: why *fun* is becoming adjective-like in one specific way and not another, why *cattle* has exactly the peculiar profile it has, why *otherwise* straddles exactly those categories and not others.

Return to *cattle*. The prototype theorist can say: *cattle* is a less prototypical noun than *dog*. It lacks the singular - plural contrast that prototypical nouns have. It takes plural agreement but resists *a* and *one*. Fine, this describes its distance from the prototype. But if category membership is just similarity to a prototype – or just partial match to a feature bundle – then nothing in that story explains why exactly this configuration appears rather than countless other logically possible ones. A word could be noun-like in its agreement but verb-like in its argument structure and adjective-like in its modification patterns. The space of possible non-prototypical members is vast. But actual non-prototypical members cluster in predictable ways. *Cattle* isn't random noise around the noun prototype. It's a stable configuration that has persisted for centuries.

The stability problem is general. If categories are gradient similarity structures, they should drift. Each generation learns from slightly different data. Each speaker has slightly different exemplars in memory. Small perturbations should accumulate. Over time, categories should dissolve into chaos, or at least into un-recognizable configurations. But they don't. The periphery of NOUN in English today looks broadly like the periphery of NOUN a hundred years ago. Marginal members stay marginal. Central members stay central. The gradient structure holds still.

Why?

Prototype theory, as it's usually deployed in linguistics, has no answer. Taylor's *Linguistic Categorization* (Taylor, 2003), perhaps the most influential application of the framework to grammar, describes prototype effects in detail but remains agnostic about what maintains them. It registers that categories have centres and edges. It doesn't say what keeps the edges from eroding.

1.2.4 An analogy

Imagine mapping the distribution of a species across a landscape. You find that the species is densest in certain habitats and thins out toward others. You could describe this as a prototype structure: prototypical members live in the core habitat; peripheral members live at the margins. The description would be accurate.

But it would miss the explanation. The species is distributed that way because of ecological mechanisms: resource availability, predation pressure, climate tolerance. The distribution isn't a brute fact about similarity to a prototype. It's the outcome of processes that concentrate the population in some places and thin it in others.

Prototype theory and feature-bundle approaches, applied to language, are like the distributional description without the ecology. They capture the shape. They miss the dynamics.

1.2.5 The deeper issue

Prototype theory, despite its rebellion against definitions, inherits a key assumption from the tradition it rejected: that categories are primarily synchronic structures.

The essentialist asks what a category *is*. The prototype theorist asks what it *looks like*. Neither asks what *keeps it that way*. Both treat the category as something you can describe by examining its current state. Neither asks what maintains the category, what keeps it from collapsing, what generates its particular structure rather than some other structure.

This is a question about mechanisms. Not "what do the members of this category have in common?" but "what processes cause these properties to cluster and stay clustered?" Not "where is the boundary?" but "why is there a boundary there at all, and why does it persist?"

A strand of cognitive linguistics – associated with Bybee, Goldberg, and others – goes further than classical prototype theory. These researchers don't merely accept gradience; they study the mechanisms that produce it: frequency effects, entrenchment, analogical extension, constructional schemas. This usage-based tradition has many of the pieces this book will draw on. What it typically lacks is the explicit ontological claim that mechanisms are not merely *causes* of category structure but *constitutive* of category reality. That claim – that the mechanisms *are* the category, not merely forces acting on it – is what the present framework adds. Variationists like Labov and Tagliamonte, typologists like Haspelmath and Croft, grammaticalization theorists tracking the pathways by which forms change – all have been studying stabilizing dynamics, often without foregrounding what that work implies about what categories fundamentally are.

Prototype theory and feature-bundle approaches were right that classical essentialism failed. They were right that categories exhibit gradient *judgments*, apparent fuzziness at boundaries, better and worse members. They didn't go far enough: documenting the gradience is not the same as understanding it.

1.3 The impasse

So we have two traditions, and together they produce an oscillation: in one mode we write as if categories were crystalline; in another we acknowledge the gradience and set it aside.

Essentialism offers a clear ontology – categories are discrete, membership is binary, definitions can in principle be found – and this clarity has built the infrastructure of modern linguistics. Grammars, parsers, textbooks, annotation schemes: all of it rests on the assumption that *fun* is either a noun or an adjective (or perhaps a verb), that *otherwise* belongs somewhere. The definitional grounding is false, but the discreteness is load-bearing. Take it away and the architecture sags.

Prototype theory offers an accurate phenomenology – categories *appear* gradient, boundaries *seem* fuzzy, typicality is real – and this accuracy has licensed a generation of work on the messiness that essentialism couldn't see. But accuracy about appearances is not explanation. To say that *cattle* is a non-prototypical noun is to redescribe the problem, not to solve it. The gradient structure is stable. Prototype theory records the stability. It doesn't say what produces it.

We are left oscillating. In one mode, we write as if categories were crystalline: discrete, bounded, the kind of thing that could appear in a rule. In another mode, we acknowledge the gradience, note the exceptions, and quietly set them aside. The two modes don't communicate. The crystalline picture is useful but misgrounded. The gradient picture is phenomenologically apt but explanatorily inert.

Return one last time to Huddleston's email. "I don't know how to handle" certain uses of *otherwise*.

The essentialist hears this as a problem to be solved: find better criteria, and the word will sort itself out. The prototype theorist hears it as a fact to be accepted: *otherwise* is marginal, gradient, a fuzzy case in a fuzzy system.

Neither hears what Huddleston actually wrote.

He didn't say the criteria were unclear. He didn't say the word was marginal. He said he didn't know how to *handle* it – how to think about a word that behaves one way in one construction and another way in another, that passes some tests and fails others, that has sat at the intersection of categories for centuries without drifting to either side. The question isn't which box the word belongs in. The question is why it won't stay in any box at all – and why that instability is itself stable.

That question is the one neither tradition can ask. Essentialism can't ask it because it assumes the boxes are real and the task is to find them. Prototype theory can't ask it because it assumes the gradience is primitive and there's nothing further to explain. Both traditions take the shape of the categories as given. Neither asks what gives them that shape.

This is the impasse. Not a choice between two theories, but a shared blind spot. Both traditions – at least in their textbook forms – take the shape of the categories as given. Usage-based and variationist work has begun to ask what maintains the clustering, but typically without foregrounding the ontological stakes: what it would mean for categories to *be* the mechanisms that maintain them, rather than entities that mechanisms act upon.

Call this the MAINTENANCE VIEW. The hypothesis is that categories are real because they are maintained. The full argument will have to wait for Chapter 4; for now, the point is only that neither essentialism nor prototype theory can even ask the question.

A parallel from physics: for decades, foundational questions about quantum theory – what measurement means, how entanglement works, whether hidden variables exist – were dismissed as philosophy, not physics. Researchers who pursued them faced career obstacles; they were warned away from the questions that seemed merely conceptual. But the foundational work produced some of the most practically useful results in modern physics: Bell's inequality, the no-cloning theorem, teleportation protocols. Tools now used in quantum computing and cryptography began as questions about what quantum theory *means*.

We don't know yet what the Bell inequality of linguistics will be. But the physics example suggests that foundational clarity can have unexpected payoffs. The question "what are linguistic categories, actually?" may look like philosophy – and in a sense it is. But so was the question that produced Bell's theorem. Getting the foundations right tends to matter eventually.

The next two chapters trace the impasse's consequences.

Chapter 2 examines why essentialism's fixes keep failing – why better definitions don't solve the problem, why the boundary cases aren't going away, why the search for necessary and sufficient conditions blocks the questions that would actually help.

Chapter 3 asks what falls out of view when we try to escape. Some traditions have retreated to nominalism, treating categories as convenient fictions with no cross-linguistic reality. Others – usage-based linguistics, construction grammar, sociolinguistics, functionalism – have moved toward gradience, frequency, variation, and communicative pressure, glimpsing the mechanisms without quite making them central. These traditions have the right instincts. But none has asked what holds the clusters together, or why certain configurations persist, or what kind of thing a grammatical category has to be for *otherwise* to sit at an intersection for three centuries without budging. And none has noticed that the hardest questions aren't about words at all.

Only then will we be in a position to ask what Huddleston's email has been waiting seventeen years to hear.

The chapters that follow don't reject prior work on language variation and change. They reorganize it. The variationist insight that patterns spread through populations, the cognitive linguist's insight that frequency entrains form, the typologist's insight that similar pressures produce similar categories – all of these are data for the maintenance view, not competitors to it. What changes is the framing: these are not independent explanations but braided strands of a single stabilizing story.

Essentialism and its discontents

Huddleston and Pullum (2002) are explicit about what a clause is:

> [In English,] the head of a clause (the predicate) is realized by a VP, and the head of a VP (the predicator) is realized by a verb. The verb thus functions as the ultimate head of a clause, and is the syntactically most important element within it. (p. 50)

A footnote sharpens the point: "Since the verb is the ultimate head, we can identify clauses by the verb" (p. 50, n. 3). Clauses are VP-headed. Verbs are criterial. You identify a clause by identifying its verb.

Fourteen chapters later, discussing constructions like *They were standing against the wall* [*with their hands above their heads*] and [*Although no longer a minister*], *she continued to exercise great power*, *CGEL* states: "The underlined clauses have subject + predicate structure, but with no verb in the predicate" (p. 1266).

Clauses are identified by their verbs. These clauses have no verb.

This isn't carelessness. *CGEL* is a work of exceptional rigour, two decades in the making, with scrupulous attention to just the kind of definitional precision that would flag this inconsistency. One might object that the early definition is merely a heuristic for canonical clauses. But that is exactly the diagnosis: the system invokes a crisp criterion for the core and switches to a functional one for the periphery. The result is local coherence (each analysis works on its own terms) but global inconsistency. If the most rigorous descriptive grammar of English produces an incoherence this stark, the problem lies not with the grammarians but with something in the method itself.

To see what that something is, we need to step back. What did essentialism give us? Why was it the default? And why do its successes make failures like this one invisible?

2.1 What essentialism built

Essentialism isn't a theory that linguists consciously adopt but rather the default assumption that linguistic categories have definitions: necessary and sufficient conditions that determine membership. To be a noun is to have whatever properties make something a noun, and to be a phoneme is to satisfy whatever criteria distinguish phonemes from allophones. The analyst's task is to discover these conditions, state them precisely, and apply them consistently.

Two claims are bundled in this assumption, and they need separating. The first is that categories have essences: necessary and sufficient conditions that constitute membership. The second is that membership is determinate: for any item and any category, there's a fact of the matter whether the item belongs. Essence entails determinacy – if you either have the essential properties or you don't, membership is settled. But determinacy doesn't require essence. A tiger is determinately a tiger not because it satisfies some checklist of intrinsic properties (stripes, carnivory, size) but because of its lineage – its causal-historical connection to other tigers. For core cases, membership is determinate without essence. For boundary cases, the causal-historical facts may not settle the question – and that indeterminacy is genuine, not merely epistemic. This distinction will matter when we reach the alternative in Chapter 4. For now, note only that essentialism's two commitments can come apart, and that when they do, the interesting question is which one fails.

A clarification is needed here about intellectual history. The target of this chapter isn't Aristotle.

Aristotle's essentialism was more sophisticated than the methodological practice I'm diagnosing. For Aristotle, an essence isn't a checklist of necessary and sufficient conditions but an explanatory core: the *to ti ên einai*, what it is to be the thing, which explains why the thing has the properties it has. A tiger's essence, on this view, is whatever makes it do tiger-things – not a list of features (stripes, carnivory, size) but a causal-explanatory principle. Nor did Aristotle assume that every category has an essence. Artefacts, social kinds, arbitrary groupings – these aren't essence-bearing in the way that natural kinds are. The question of *which* categories are natural kinds was live, not presupposed.

What I'm calling essentialism is a methodological descendant – but not a direct one. Through Scholastic systematisation, Locke's nominal essences, logical positivism's verification conditions, and mid-century conceptual analysis, the Aristotelian framework hardened into the checklist picture: necessary and sufficient conditions as the mark of genuine understanding. Jakobson (cited in Joos, 1957, p. 703) put it with characteristic bluntness: "The linguistic categories, then, are absolutes which admit of no compromise." Anything that couldn't be captured by "a finite number of absolute categories" was deemed non-linguistic and expelled from analysis.

Structuralism complicates this narrative. Saussure's insistence that linguistic units are purely differential – *dans la langue il n'y a que des différences* – is not checklist essentialism; categories are constituted by systemic relations, not intrinsic properties. The Prague School phoneme was defined oppositionally, by contrasts, not by bundles of inherent features. But what survived into practical descriptive work was the determinacy assumption: that for any item and any category, there is a fact of the matter whether the item belongs. The relational metaphysics didn't translate into tolerance for indeterminate membership. By the time category-assignment became routine practice in the great descriptive grammars, the method presumed sharp boundaries even where the theory might have licensed fuzziness.

The distinction matters because the alternative I develop in Chapter 4 – homeostatic property clusterhomeostatic property cluster theory – is sometimes called neo-Aristotelian. It recovers real natural kinds with genuine explanatory structure, just not checklist-definitional structure. The homeostatic mechanisms that maintain a category play the explanatory role that essences were supposed to play, causally understood.

This assumption underwrote a century of productive research.

The great descriptive grammars – Jespersen (1909), Quirk et al. (1985), Huddleston and Pullum (2002) – organized vast empirical coverage using, at least in part, essentialist architecture. Each category receives a definition; membership follows from the definition; exceptions are noted and, where possible, explained. The result is systematization that reveals patterns invisible to casual observation, not mere taxonomy. *CGEL*'s treatment of the English verb phrase, for instance, distinguishes catenative constructions from auxiliary constructions from control constructions, each with distinct syntactic properties, and shows how surface similarities mask structural differences. This analytical power depends on treating categories as if they have determinate boundaries. You can't show that *keep* in *keep talking* differs structurally from *will* in *will talk* unless you have clear criteria for what counts as an auxiliary.

Generative grammar pushed the essentialist method further and discovered genuine regularities. Binding theory identified conditions on the interpretation of anaphors and pronouns – Principle A, Principle B, Principle C – that predict grammaticality across constructions and languages (Chomsky, 1981). Island constraints revealed that extraction isn't freely available but blocked by specifiable structural configurations (Ross, 1967). The c-command relation, once isolated, turned out to govern phenomena from negative polarity licensing to quantifier scope (Ladusaw, 1979; Reinhart, 1983). These are discoveries, not stipulations – they make predictions that are often correct, revealing patterns that wouldn't have been visible without assuming that categories like ANAPHOR, INTERROGATIVE PHRASE, and BOUNDING NODE have determinate definitions.

Phonology followed the same logic. The Prague School's distinctive features (Jakobson & Halle, 1956) proposed that phonemes are bundles of binary properties: [±voice], [±nasal], [±continuant], and so on. This was explicitly essentialist – a phoneme just is its feature bundle – and it yielded the concept of natural classes. Segments that share a feature behave uniformly in phonological processes: voiced obstruents trigger voicing assimilation; nasals condition vowel nasalization; continuants pattern together in lenition. The predictive power was undeniable. English plural allomorphy ([s] after voiceless segments, [z] after voiced ones, [ɪz] after sibilants) falls out from natural classes without listing each conditioning environment (Hayes, 2009).

Practical applications followed. Parsers require categories with boundaries; you can't write a phrase-structure rule for NP unless NOUN picks out a determinate set. Pedagogical grammars depend on the same architecture: learners need to know what counts as a noun, a verb, a clause. Speech recognition systems, corpus annotation schemes, machine translation models – all inherit the essentialist assumption because all require categories to have membership conditions.

None of this was naïve. Linguists knew that boundaries could be fuzzy, that edge cases existed, that definitions were sometimes stipulative – and yet the fuzziness seemed like noise at the margins of a fundamentally sound architecture. When *near* resisted clean classification as adjective or preposition, the response was to note the difficulty and move on, rarely to question whether adjective and preposition were the right kinds of kind.

2.2 Where essentialism works

Essentialism is unimpeachable when CLASSES are constructed (I reserve CATEGORY for natural kinds throughout).

Stipulative definitions are the clearest case. A bachelor is an unmarried adult male because we define it so. Membership is determinate because the essence just is the stipulation, and no discovery is required. Formal systems work the same way: a touchdown is constituted by the rules of American football, a checkmate by the rules of chess, a well-formed formula by the grammar of the formal language. Within constructed systems, essentialism isn't a hypothesis but a design feature.

Mathematics sharpens the point. Define the natural numbers via the Peano axioms (Peano, 1889); then discover that there are infinitely many primes, that every even number greater than two is (apparently) the sum of two primes. The category is constructed; the properties are found. Essentialism is literally true – a natural number *is* whatever satisfies the axioms – but the inquiry is genuine. Results are non-obvious; proofs can be flawed; the enterprise is recognizably empirical in its structure if not its ontology.

The formalist temptation in linguistics is to borrow this architecture wholesale from logic: treat syntax as a formal rewriting system, define categories precisely, derive consequences. In his early generative work, Chomsky (1957) adopts what is essentially a Post-style production-system architecture, already developed for logic and the theory of computation (Post, 1943, 1944; Pullum, 2025) and explicitly proposed for linguistic application by Rosenbloom (1950). Grammars are devices that define a set of well-formed symbol strings by explicit construction rules. Essentialism here isn't a hypothesis but the design: the grammar builds in necessary-and-sufficient conditions for membership in the stringset. And it worked, partly. Constituency, recursion, embedding – real discoveries, enabled by treating syntax as if it were mathematics.

But there's a disanalogy, and it matters.

In mathematics, you stipulate the category and discover its properties. In linguistics, you are trying to discover the category. The direction reverses. When you define NATURAL NUMBER, you aren't hypothesising that something out there satisfies your definition. When you define CLAUSE, you are. Mathematical existence is logical consistency; natural-kind existence is causal-historical. A category that is logically coherent might carve nothing in natural language.

The bootstrapping problem makes this concrete. In mathematics, axioms are stipulated; theorems are checked against the axioms. In linguistics, intuitions are used to formulate rules that predict intuitions. The circularity is vicious in a way that mathematics escapes. You can't check the grammar against an independent standard, because the grammar just *is* a systematisation of judgments – and judgments vary, degrade at boundaries, and respond to factors the grammar doesn't model. The data aren't independent of the theory in the way that prime numbers are independent of number theory.

Formal methods in linguistics face a dilemma. If categories are stipulated, like mathematical ones, then linguistics isn't discovering natural structure – it's constructing useful fictions. If categories are discovered, like biological kinds, then formal definitions are hypotheses about causal structure, defeasible by evidence that derivations internal to the formalism can't detect. The formalist hope was to have it both ways: mathematical rigour applied to natural facts. But you can't stipulate your way to empirical truth, and you can't derive natural kinds from axioms.

Regimentation offers a middle path, but a modest one. Constructing precise replacements for vague concepts (like *CGEL*'s CANONICAL CLAUSE) provides useful benchmarks/tools, but these are tools, not ontologies. The error lies in mistaking the tool for the kind.

There are corners of linguistics where essentialism isn't a temptation but a job requirement.

Consider lexicography. Dictionary editors face practical constraints that reward essentialist form: entries must be short, portable, scannable. A definition that says "it depends on multiple factors" doesn't fit. The format demands necessary and sufficient conditions, or at least clean prototype descriptions. This pressure limits the tool's ontological utility. English dictionaries consistently define prepositions as words taking NP complements, then label *near* and *despite* as prepositions even when they violate that definition (Reynolds, 2025d). The dictionary works as a tool because it imposes local coherence; it fails as an ontology because it hides global inconsistencies.

Programming languages look different.

If formal language theory counts as linguistics-adjacent, programming languages are where the Post–Chomsky architecture genuinely fits. A programming language is, by design, a formally specified stringset with stipulated semantics. The grammar isn't a hypothesis about pre-existing structure; it's the mechanism by which the structure is brought into being. A program is well-formed if and only if it satisfies the grammar. A construct means what the semantics say it means. There's no analogue of native-speaker intuition that might override the specification.

Here, necessary and sufficient conditions aren't speculative hypotheses about natural kinds. They're design decisions. If the compiler accepts a string the grammar excludes, the compiler has a bug. If the grammar generates ambiguity the designers didn't intend, the grammar gets patched. The essentialist apparatus – precise definitions, deterministic membership, clean boundaries – isn't aspirational but constitutive. Programming languages are essentialist because they're engineered to be.

The contrast clarifies the error. Lexicography imposes essentialist form on recalcitrant natural-language material; the result is systematic leakage. Programming languages are built to essentialist specifications; the result is genuine success. Natural language is neither a dictionary entry nor a formal specification. It isn't engineered for look-up, and it isn't designed at all. Importing the architecture that works for PLs, or the format that dictionaries require, treats natural language as something it isn't.

When criteria converge, this category error is invisible. Every reasonable definition picks out the same set; the stipulated boundaries and the natural clustering are indistinguishable. Only when criteria diverge does the question surface: are we discovering structure, or imposing it?

	Definitional crispness	Distributional coherence	Operationalisability
bachelor, *prime number*	✓	✓	✓
Canonical nouns, phonemes	✓	✓	✓
fun, *near*, *otherwise*	~	~	×
will, *going to*	×	~	~
Verbless clauses	×	~	×
subject (cross-linguistically)	×	~	×

Table 2.1. Essentialism succeeds when criteria converge. The top rows show cases where definitional, distributional, diachronic, and operational criteria align: membership is uncontroversial, and the essentialist apparatus works. The lower rows show cases discussed in this chapter where criteria pull apart. No single criterion can be promoted to "the real definition" without arbitrarily discounting the others. The pattern of partial alignment is what a mechanistic account must explain.

2.3 When criteria diverge

The category SUBJECT has been central to syntactic theory since its inception – every theory needs it or something like it – and yet decades of cross-linguistic research have failed to produce an agreed definition.

Keenan (1976) proposed a cluster of behavioural and coding properties: subjects tend to be nominative-marked, to trigger verb agreement, to control reflexivization, to be deletable under coordination, to be accessible to relativization. No single property is necessary; the category is identified by a preponderance of properties. This was already a retreat from strict essentialism – a family resemblance approach rather than a definition – but it preserved the assumption that SUBJECT names a cross-linguistic kind.

Generative grammar took a different approach. The subject is the external argument: the argument merged outside VP, occupying Spec-IP (later Spec-TP). This is a configurational definition, not a behavioural one. An NP is a subject if it occupies the right structural position, regardless of its morphological marking or behavioural properties (Chomsky, 1981).

Dixon (1994) offered a third criterion: the syntactic pivot, the argument that controls cross-clausal operations like coordination deletion and relativization. In accusative languages, this is typically the agent; in ergative languages, it may be the absolutive argument (patient of transitives, sole argument of intransitives).

These three approaches – behavioural cluster, configurational position, syntactic pivot – don't pick out the same set of NPs across languages. Evans and Levinson (2009) surveyed the evidence for cross-linguistic diversity and concluded that the assumption of universal categories is "a myth": languages cluster around alternative architectural solutions, but the clusters aren't identical, and many languages lack categories that others treat as fundamental. In Dyirbal, the absolutive argument controls coordination deletion; by Dixon's criterion, it's the subject. By the configurational criterion, the ergative argument might be the subject, depending on your phrase structure assumptions. By Keenan's cluster, neither argument has a clear majority of subject properties. The criteria aren't just different; they operate at different levels. Keenan's properties are partly behavioural (what subjects do in discourse), the configurational definition is structural (where subjects appear in phrase structure), and Dixon's criterion is functional (what role subjects play in cross-clausal operations). Asking which is the real definition conflates ontological levels that may not align.

The debate has continued for fifty years without converging – and can't converge, because the disputants aren't disagreeing about facts. They are disagreeing about which essence the term *subject* should name. The framework provides no procedure for resolving this, because the framework assumes there's an essence to be discovered. When there are multiple candidates, each internally coherent, essentialism offers only continued assertion.[1]

Phonology exhibits the same pattern. The Prague School (features), American structuralism (distribution), and generative phonology (underlying segments) offer competing definitions of the phoneme. For clear cases like English /t/, they converge. But for boundary cases like the PIN-PEN merger (where contrast is neutralized in specific environments), they diverge. The phonetic criterion sees distinct vowels; the distributional criterion sees one phoneme; the generative criterion depends on the analysis. As with syntax, the debate concerns not facts but which criterion should be definitional.

The pattern extends beyond syntax and phonology, but the structure is now clear enough to return to where we began.

2.4 The microcosm

Return now to *CGEL*'s verbless clauseverbless clauses.

The grammar provides two criteria for clausehood: a syntactic one in its Chapter 1, where clauses are VP-headed and identified by their verbs, and a semantic one in Chapter 15, where clauses exhibit subject + predicate organization.

For canonical clauses, these converge. *She left* has a verb heading a VP predicated of a subject – both criteria satisfied, nothing to debate. For *their hands above their heads*, by contrast, the criteria diverge: there's no verb, so the syntactic criterion fails, though a predicand and predicate satisfy the semantic one.

CGEL calls these clauses. The move is exactly the one we have seen elsewhere: when the primary criterion fails, invoke a secondary one. The syntactic essence gives way to a semantic one – predication is what clauses are *for* – and the category is preserved.

Huddleston and Pullum are scrupulous analysts; they would flag a definitional problem if the framework made it visible. But within essentialist practice, the inconsistency passes unnoticed because criterion-switching is how analysts handle cases where the primary criterion fails. You find a criterion that works and apply it. The result is locally coherent – these constructions do exhibit predication, they do function clause-like, calling them clauses isn't unreasonable – and globally inconsistent.

The framework treats this as normal procedure – but the error lies in the framework itself.

The global inconsistency only becomes visible when you step back and ask: what is the essence of clausehood? VP-headedness, or subject–predicate structure? If VP-headedness, then verbless clauses aren't clauses – and indeed, a tra-

[1] If you are reading this footnote hoping for a resolution, you have understood the problem.

ditionalist could simply refuse the label: "Clauses are identified by their verbs; these constructions lack verbs; therefore they are not clauses." The move is defensible, but it forces you to say what they *are*, and the predicate-bearing facts remain unexplained. Alternatively, one could postulate a null verb – an unpronounced BE – rescuing the VP-headed criterion at the cost of abstract structure. This is a coherent move (and one later chapters will consider in other domains), but it concedes that the surface form underdetermines the analysis: now you need a theory of when null elements are licensed and when they are not. Either escape route highlights the same underlying problem: essentialism requires a single answer. *CGEL* provides two, deploying whichever one serves the immediate analysis. The question *CGEL* can't ask is: what keeps these properties bundled most of the time? Not which criterion is primary, but what maintains the alignment that makes the criteria competitors in the first place.

2.5 The diachronic problem

The cases considered so far are synchronic. At a given moment, criteria diverge and analysts disagree about which is essential, but the framework provides no resolution. But there's a deeper problem, and it's diachronic.

Categories change.

will was a lexical verb meaning "want, desire." Now it's an auxiliary expressing futurity. *While* was a noun meaning "a period of time." Now it's a preposition that takes clausal complements. And *-hood* was an independent noun meaning "state, rank." Now it's a bound derivational suffix. The histories are well documented. The trajectories are gradual. And the process – grammaticalization – isn't an anomaly but a central mechanism of language change (Bybee et al., 1994a; Hopper & Traugott, 1993).

Essentialism struggles to accommodate this.

It has tried.

English *will* has gone from lexical verb (*willing*) to future marker, *be going to* is making the same transition, modal *dare* is (or was) drifting toward auxiliary status. The evidence shows gradient change, not threshold-crossing. The question demands a boundary; the phenomenon supplies a gradient.

Grammaticalization theorists – Hopper, Traugott, Bybee – have studied these trajectories for decades, and their work is full of mechanisms: frequency effects, phonological reduction, semantic bleaching, analogical extension, reanalysis by learners. This is exactly the kind of mechanism-based inquiry that essentialism blocks. But grammaticalization theory typically frames its findings as descriptions of *change* – how forms move from one category to another –

rather than as accounts of what categories *are*. The mechanisms explain the trajectory; they aren't yet framed as constitutive of the categories themselves. Chapter 4 makes that move explicit. If AUXILIARY has an essence – necessary and sufficient conditions for membership – then *will* either satisfies those conditions or it doesn't. But the historical record shows *will* acquiring auxiliary properties piecemeal over centuries: loss of non-finite forms, loss of *to*-infinitive complements, reduction of agreement marking, semantic bleaching from volition to futurity. At any historical slice during this transition, *will* is a boundary case. It has some auxiliary properties and lacks others. The essentialist must say either that *will* was already an auxiliary (and the earlier properties were noise), or that it became an auxiliary at some precise moment (which moment?), or that it isn't yet fully an auxiliary (but then what is it?). None of these options fits the evidence. The evidence shows gradient change, not threshold-crossing.

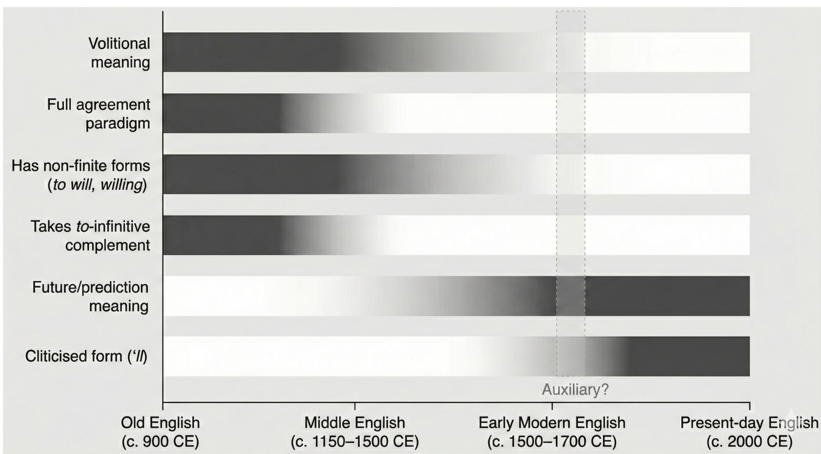


Figure 2.1. The grammaticalization of *will*. Lexical-verb properties (volitional meaning, full agreement, *to*-infinitive complements, non-finite forms) fade out gradually; auxiliary properties (future meaning, cliticisation) fade in. The changes are staggered and gradient. Essentialism requires a threshold-crossing moment; the historical record shows a slope.

The problem isn't merely epistemological – not just that we can't tell when *will* became an auxiliary. The problem is ontological. There's no moment at which *will* became an auxiliary because *becoming an auxiliary* isn't the kind of process that has a moment. It's a gradual shift in a cluster of properties, driven by frequency, phonological reduction, semantic bleaching, analogical extension, and reanalysis by learners. The boundary zone isn't a failure of classification but the mechanism of change itself.

This reframes synchronic boundary cases. Verbless clauses, the classification of *near*, and the disputed SUBJECT constructions in ergative languages – these aren't embarrassments to be resolved by finding the right definition. They're snapshots of categories under pressure. Some items are moving into a category; some are moving out; some are stable at the boundary. The boundary zone is populated because language change populates it. Asking "is this really an X?" presupposes a stable kind for items to be inside or outside of. Grammaticalization shows that the kinds themselves are moving.

The essentialist response is to distinguish synchrony from diachrony. Synchronic grammar describes a static system; diachronic linguistics describes transitions between systems. At any given moment, the categories are fixed; over time, one system replaces another. This preserves essentialism by treating change as external to grammar proper.

But the distinction doesn't hold. If the boundary zone is always populated – if, at every synchronic slice, there are items in transition – then the static system is a fiction. What looks like a bounded category is a standing wave: apparently stable, dynamically maintained, with continuous flow through the boundary region. The wave is real. Its edges aren't.

Essentialism treats synchrony as primary and diachrony as derivative: change is movement from one state to another, and the states are what have essences. The diachronic evidence inverts this – synchrony is slow diachrony, the stability a dynamic equilibrium rather than stasis, and the categories, though real, are processual rather than definitional.

This is another question essentialism can't ask. It can ask: what is the essence of AUXILIARY? It can't ask: what mechanisms move items into and out of the auxiliary cluster? The first question has no answer, because there's no essence. The second question has answers – frequency effects, phonological reduction, constructional analogy, acquisition biases – but the answers are invisible within essentialist framing.

2.6 The blocked question

The pattern across these cases isn't just "definitions fail." Definitions fail in a specific way that blocks a specific question.

When criteria diverge, essentialism frames the problem as: *Which criterion is correct?* Which property is truly essential to SUBJECT, to phoneme, to clause? The debate becomes a search for the right definition, and because the framework provides no resolution procedure, the search continues indefinitely.

The question that essentialism can't foreground is: *Why do these criteria converge in the first place?* The first question assumes the category and searches for its boundary. The second treats the category itself as the thing to be explained.

For canonical nouns, morphological, syntactic, and semantic criteria all pick out the same set. That is not a trivial convenience; it's a fact about how linguistic structure is packaged. What keeps those properties bundled together strongly enough that NOUN is useful for most of the lexicon? And what happens in the boundary zone where the bundling weakens?

Boundary cases – verbless clauses, near-mergers, disputed SUBJECT constructions in ergative languages – are where the explanandum becomes visible. Under essentialism they are embarrassments. Outside essentialism they are the natural place to look for whatever keeps the usual alignment intact.

Some traditions have begun asking this second question; Chapter 3 surveys what they get right and what they still don't ask.

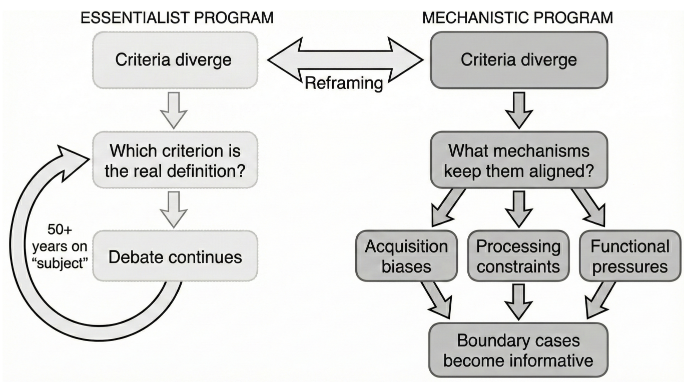


Figure 2.2. Two responses to criterion divergence. The essentialist program asks which criterion is definitional and cycles indefinitely. The mechanistic program asks what produces convergence in the first place, transforming boundary cases from embarrassments into evidence.

2.7 A parallel

This pattern will be familiar to philosophers of biology.

What is a species? For a century, biologists offered criteria: morphological species concepts, biological species concepts (Mayr, 1942), phylogenetic species concepts, ecological species concepts. Each is internally coherent. They don't converge on the same groupings. Some populations are connected by gene flow in a geographic ring but reproductively isolated at the endpoints; some hybridize freely; some reproduce without recombination at all. These are the boundary cases – ring species, hybrid zones, asexual lineages – where criteria come apart and the definitional debate stalls.

The diachronic problem recurs too. Asking at what precise moment a lineage ceased to be a dinosaur and became a bird is like asking when *will* became an auxiliary. The question demands a boundary that the process of evolution does not supply.

And morphology can diverge sharply from phylogeny.

Smilodon (the sabre-toothed cat) and *Thylacosmilus* (a South American marsupial predator) are separated by over a hundred million years of mammalian evolution (Figure 2.3). They share no recent common ancestor. Yet their skulls are nearly identical: massive canines, robust zygomatic arches, hypercarnivorous dentition. Place them side by side and you see the same animal, arrived at twice.

A morphological species concept groups them; a phylogenetic one separates them. The parallel to linguistics is exact: when criteria diverge, essentialism offers no resolution, only continued assertion about which criterion is "really" definitional. But the convergence itself – why these two lineages, under similar selective pressures, evolved nearly identical forms – is the phenomenon that demands explanation. Chapter 4 will show how mechanistic pressures can explain convergence without requiring shared essence.

For a century, the debate was framed as: which criterion is the correct definition of SPECIES? The debate didn't resolve. It couldn't resolve, because the framework assumed there was an essence to be discovered.

What dissolved the debate wasn't a better definition but a different question. Instead of "What is a species?" biologists began asking "What mechanisms maintain this population as a cohesive unit over time?" The answer varies: gene flow, shared selection pressures, developmental constraints, mate recognition systems. Different mechanisms predominate in different cases. SPECIES names not a single kind defined by a single essence but a cluster of populations maintained – flexibly – by overlapping mechanisms.

This shift – from definition-seeking to mechanism-seeking – is what made boundary cases informative. Ring species aren't embarrassments to be explained

Figure 2.3. Convergent morphology without shared essence. *Smilodon* (top) and *Thylacosmilus* (bottom) evolved nearly identical skull architecture despite over 100 million years of phylogenetic separation. A morphological species concept groups them; a phylogenetic one separates them. The convergence is explained not by shared ancestry but by similar functional pressures – hypercarnivory, ambush predation – producing similar solutions (Zimmer, 2010). © Carl Buell (image used with permission).

away; they are evidence of what happens when isolating mechanisms operate incompletely. Hybridization isn't a failure of species boundaries; it reveals the conditions under which reproductive isolation is maintained or breaks down.

Chapter 4 will develop this parallel in detail. For now, note only that linguistics isn't alone in facing essentialist failure: the pattern is general, and so, potentially, is the solution.

2.8 Why essentialism persists

If essentialism fails this reliably, why does it persist?

The persistence has structural causes, and understanding them clarifies what would be needed to shift the field.

Textbooks freeze the architecture. Introductory grammars present categories as checklists: a noun is a word that does X, Y, Z. Students internalise this method before encountering the problem cases that strain it. By the time they reach the boundary phenomena, the framework feels like common sense rather than a theoretical commitment. Pedagogy rewards clean definitions; complexity looks like failure to teach clearly.

Publication norms reward definitional disputes. Postal (1966) titled his paper "On so-called 'pronouns' in English" – scare quotes announcing that pronouns aren't really pronouns but determiners preceding silent nouns. The paper launched decades of debate, culminating in Abney's DP hypothesis (Abney, 1987) – an unpublished dissertation that nonetheless reshaped the field, less through empirical vindication than through perceived endorsement. A paper arguing that X is really Y – that raising verbs are really control verbs, that determiners are really functional heads, that the passive is really a species of unaccusativity – has a clear contribution: it advances a position in a debate. A paper arguing that the X/Y boundary is gradient and maintained by multiple mechanisms has a murkier contribution. It sounds like giving up. The rhetoric of "really" signals progress; the rhetoric of "it depends" signals inconclusiveness. The "is X really Y" literature could furnish a small library. The library's classification system would be disputed.

Practical applications feed back into theory. Parsers need discrete categories. Corpus annotation schemes need consistent labels. The Penn Treebank – the foundational training corpus for a generation of NLP systems – collapsed prepositions and subordinating conjunctions into a single tag because annotators couldn't reliably distinguish them; the original Brown Corpus showed inconsistent tagging even in identical syntactic contexts (Marcus et al., 1993). Speech recognition systems need boundaries. The engineering demands are real, and they create pressure to treat categories as if they were determinate even when the theoretical grounds for doing so are shaky. When the parser works, it feels like vindication. When it fails at boundaries, the failure is attributed to noise or insufficient data, not to the ontology. Neural language models – especially transformers – have loosened this constraint somewhat: they learn distributed representations that can accommodate gradience without requiring sharp categorical boundaries (Vaswani et al., 2017). But the theoretical pressure remains. The models are trained on categorically annotated data; the evaluation metrics presuppose discrete labels; the engineering culture inherited essentialist assumptions even as the architectures became more flexible.

Definitions feel like progress. Stating necessary and sufficient conditions feels like achieving something – like you've finally said what X *is*. The phenomenology is satisfying in a way that mechanism-talk isn't. "A clause is a VP-headed structure" is crisp. "Clausehood is a cluster of properties maintained by functional, acquisitional, and processing pressures that usually but not always co-occur" is accurate but unsatisfying. The crispness is illusory, but illusions can be motivating.

None of this was conspiracy. It's ordinary institutional dynamics. Taken together, textbooks, journals, applications, and definitions form a machine that rewards sharp boundaries, punishes hedging, and naturalises essentialism. The failures are visible only to those already sceptical; the successes are built into the infrastructure that trains the next generation.

The same could be said of any alternative, including the one I'll propose. But explaining why a view persists doesn't settle whether it's true – that's determined by whether the framework handles the data. The sociology explains the lag between failure and recognition, not the failure itself.

2.9 Conclusion

A final clarification. Nothing in this chapter denies that sentences have structure or that categories are real. The intuition that *The dog bit the man* differs from *The man bit the dog*, and that *dog* and *table* cluster together, is correct. What this chapter denies is the essentialist account of what makes such facts hold.

The choice isn't essence or chaos. If the essentialist says categories are defined, and is wrong, it doesn't follow that categories are arbitrary. Categories may be real but not immutable, revisable but not capricious, stable because something keeps them stable.

The verbless clauses remain. *CGEL*'s definitions are still incoherent. What the incoherence points to is not the need for better definitions but for a different explanatory target: an account of why the familiar criteria usually line up, and why they sometimes come apart.

But first we need to see what falls out of view when the field tries to escape essentialism's grip – whether by retreating to nominalism or by working around the problem with gradience and usage-based mechanisms without quite confronting it head-on.

What we haven't been asking

German has three genders. Swahili has eighteen noun classes. When we call both "gender systems", what are we claiming? That the languages share a natural kind? Or that linguists have found it useful to file them in the same drawer?

Martin Haspelmath and William Croft argue for the drawer.

In a series of papers that should have reshaped how linguists think about their basic categories, Haspelmath has pressed a disarmingly simple point: the terms we use to compare languages – NOUN, SUBJECT, DATIVE, TENSE – are COMPARATIVE CONCEPTS, not discoveries about the mind (Haspelmath, 2010). They exist in the linguist's methodology, not in the speaker's grammar.

Croft pushes the skepticism deeper. In W. A. Croft (2001), he argues that even within a single language, broad categories like SUBJECT or NOUN are illusions. There is no global category of Subject in English; there is only the subject-of-the-intransitive, the subject-of-the-transitive, the subject-of-the-passive. The only real grammatical units are specific CONSTRUCTIONS. Everything else is a generalization we make over them.

The position is carefully stated. Haspelmath does not deny that German has gender. He denies that German gender and Swahili noun classes instantiate the same mental category. Croft does not deny that English subjects behave similarly. He denies that this similarity implies a single underlying mental object. When typologists group them under a common label, they are constructing a yardstick for measurement, not uncovering a shared essence.

This is the nominalist position, articulated not by a philosopher skeptical of linguistic reality but by a working typologist who has spent decades comparing the world's languages. It deserves to be taken seriously. It responds to exactly the failures documented in the previous two chapters. If definitions keep failing, if boundaries keep dissolving, if every proposed set of necessary and sufficient conditions produces counterexamples, perhaps the hunt for definitions was misconceived from the start.

This architectural realism, shared by the PARALLEL ARCHITECTURE program, suggests that while specific categories may be language-specific, the *types* of possible categories and the *mechanisms* that relate them are constrained by universal cognitive architecture (Culicover & Jackendoff, 2005).

The position has a cost.

If GENDER is a COMPARATIVE CONCEPT and nothing more, then German gender and Swahili noun classes share only the properties we have chosen to include in our yardstick. The resemblance is constructed, not discovered. And if NOUN is a COMPARATIVE CONCEPT – if there is no cross-linguistic natural kind that English nouns and Mandarin nouns both instantiate – then the apparent universality of the noun-verb distinction is an artifact of our filing system, not a fact about the human language faculty.

Haspelmath accepts this consequence. He argues that we should accept it too. Linguistics is not chemistry. There is no periodic table of grammatical features. The search for one reflects confusion about what comparative work actually requires. The world's languages solve similar communicative problems in diverse ways; we can compare those solutions without claiming they are instances of a single kind.

The argument is coherent. It may even be right about some categories – the ones that really are artifacts of our descriptive habits, filing systems that could have been organized otherwise. But it risks proving too much. It struggles to explain why certain clusters persist across unrelated languages, why the same functional solutions recur in similar configurations, why categories that are supposedly our constructs behave as if they had "joints" of their own – to borrow Plato's essentialist phrasing.

This is where nominalism stops short. Essentialism fails to find definitions; nominalism infers that there is nothing to define. The move is understandable, but too quick. It assumes that essence is the only ground for reality: remove the essence, and all that remains is convenience. Yet if COMPARATIVE CONCEPTS yield non-trivial generalizations, there must be an answer to why they are so productive.

There is another possibility. Consider a physical analogy. Water has no essence of liquidity. There is no property, intrinsic to H_2O molecules, that makes them liquid rather than solid or gas. The same molecules constitute ice, water, and steam. What differs is the collective behavior: at certain temperatures and pressures, the molecules settle into configurations that we recognize as distinct

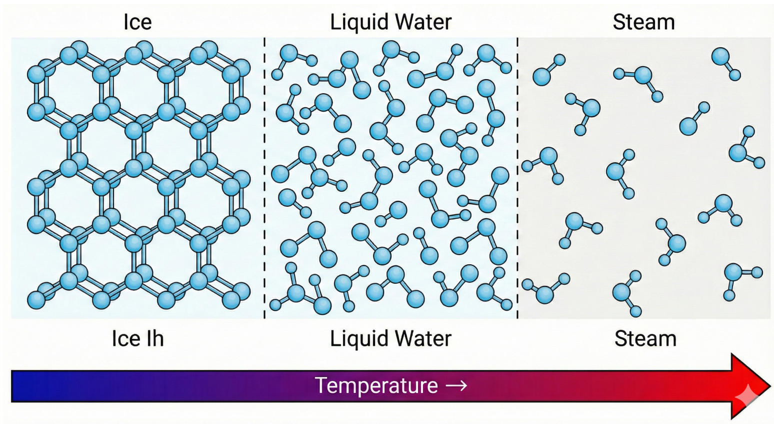


Figure 3.1. Discrete phases without essences. The same H_2O molecules constitute ice, liquid water, and steam; what differs is the collective behaviour at different temperatures. The boundaries between phases are sharp enough to skate on, swim in, and be shrouded by – yet no property intrinsic to the molecules makes them liquid rather than solid or gas. The discreteness is maintained by dynamics, not by essence.

phases. The boundaries between phases are real – you can skate on ice, swim in water, be shrouded by steam – and they are sharp enough for practical purposes. But they are not maintained by essences. They are maintained by dynamics: the statistical mechanics of molecular interaction under varying thermodynamic conditions.

Phase transitions produce discreteness from continuous substrates. The underlying variables (temperature, pressure, molecular motion) are continuous. The macroscopic outcome (solid, liquid, gas) is categorical. No essence is required. What does the work is the dynamics of the system at a particular scale.

This is a model, not a metaphor. Language is not water. But the analogy opens conceptual space that the nominalist forecloses. If physical systems can exhibit discrete, stable, projectible structure without essences, so might linguistic ones. The question becomes empirical: what mechanisms play the role in language that thermodynamics plays for matter?

Consider a linguistic case where this shift pays off. Island constraints – configurations that resist extraction – seemed a paradigm of essentialist success. Ross (1967) identified the patterns; subsequent decades spent refining the definition of an island (bounding nodes, barriers, phases). The definition kept shifting because the category refused to stabilize. Cuneo and Goldberg (2023a) asked a different question. Not: what structural definition captures islands? But: what mechanism produces extraction resistance? Their answer: long-distance dependencies foreground a constituent, while certain constructions background their content. When functions clash, acceptability degrades. The mechanism explains the variable resistance without positing an essential island boundary.

Several traditions in linguistics have moved toward answers.

Usage-based approaches, developed by Joan Bybee and others, emphasize frequency and entrenchment (Bybee, 2010). High-frequency patterns are stored with greater strength; they resist analogical change; they anchor categories. On this view, *cattle* persists in its anomalous configuration because it is frequent enough to be learned as a unit rather than derived by rule.

Construction grammar treats form-meaning pairings as the basic units of grammar (Goldberg, 1995). Categories emerge from generalizations over constructions, not from innate specifications. Gradience is expected: novel uses inherit structure from prototypical uses, with acceptability shading off as similarity decreases – an outcome of acquisition, analogy, and the distribution of exemplars.

Sociolinguistics, from Labov onward, has documented that variation is orderly because transmission is orderly. The community maintains the system, and social dynamics shape what counts as grammatical where.

Functionalism grounds grammatical structure in communicative pressure. Categories persist because they do work: noun phrases package referents; tense locates events; definiteness manages discourse accessibility. When the pressure is strong, the clustering stabilizes; when it weakens, categories erode or restructure.

These traditions have the right instincts. They look for dynamics rather than definitions. They expect gradience and try to explain it. They treat stability as an achievement. But none has made mechanism fully central. Usage-based approaches invoke frequency; construction grammar invokes analogy; sociolinguistics invokes transmission. Recent work has begun to synthesize these insights (Dahl, 2016; Schmid, 2020; Spike, 2020). What remains undeveloped is a systematic application to linguistic categories writ large, together with explicit diagnostic criteria.

That is the gap this book aims to fill.

Here is where the enterprise stalls.

If categories are maintained by mechanisms, then boundary cases become the most informative data, as Culicover (1999) argues in *Syntactic Nuts*. They are where you can see the mechanisms at work, precisely because the mechanisms are under strain. But this requires criteria for failure. An account of what makes a category real needs a companion account of what makes a category *not* real – a way to diagnose when the mechanisms are absent or insufficient. Without such criteria, mechanism-talk becomes unfalsifiable.

Chapter 8 will develop these criteria in detail: the failure modes that distinguish genuine kinds from taxonomic conveniences. For now, the point is that the nominalist, having given up on essences, has also given up on the question of discipline. Haspelmath's COMPARATIVE CONCEPTS are useful by stipulation; there is no further question of whether they carve at the joints. The mechanistic alternative promises something the nominalist abandons: a principled basis for saying that some categories are real and others are not, that some classifications reflect natural structure and others are artifacts of our methods.

But this promise depends on specifying what mechanisms do the maintaining. And that, in turn, depends on asking questions that neither tradition – essentialist nor nominalist – has posed.

3.1 The lexeme obsession

Return to the examples that opened this book: *otherwise*, *fun*, *near*, *cattle*. These are questions about items – about which box a particular word belongs in. They presuppose that the boxes exist.

The debate between essentialism and nominalism is, at bottom, a debate about these boxes. The essentialist says the boxes have necessary and sufficient conditions; the nominalist says they don't, and we should stop pretending otherwise. But both parties share an assumption: that the interesting ontological questions concern categories of *lexemes*. Is NOUN a natural kind? Is PREPOSITION? Is ADJECTIVE cross-linguistically real, or does each language have its own property-word category with no deep connection to any other?

This is the lexeme obsession. It has shaped the *ontological* debate about categories for a century – the explicit discussion of natural kinds, essentialism, and prototype structure has overwhelmingly concerned parts of speech and constructions, not feature systems. The literature on noun-verb distinctions alone could fill a library. Every introductory syntax course spends weeks on parts of speech. Dissertations are written on whether a particular language has adjectives, or whether its property words are really a subclass of nouns or verbs. The question "what category is X?" is the bread and butter of descriptive linguistics.

Meanwhile, a different set of categories has received far less explicit ontological scrutiny – not because linguists haven't studied them, but because the metaphysical questions have rarely been made central.

What kind of thing is GENDER?

Not: which nouns are masculine, which are feminine, which are neuter. That's the lexeme question again – filing items into boxes. The question I'm asking is prior: what is the box itself? Is GENDER a natural kind, a feature of human grammars maintained by identifiable mechanisms? Or is it a label we apply to superficially similar systems that have no deep unity?

German gender and Swahili noun class both involve agreement. Both sort nouns into categories that condition the form of associated elements – articles, adjectives, verbs. Both are largely arbitrary from a semantic standpoint: German *Mädchen* (girl) is neuter; Swahili *kiti* (chair) is in a class that includes many artifacts but also trees. The similarities invite a common label. But Swahili has eighteen classes to German's three. Swahili classes are more semantically coherent in some cases, less in others. The agreement patterns differ in their syntactic domains. Are these the same kind of thing, or different things with a surface resemblance?

The generativist answer appeals to Universal Grammar: gender is a formal feature, part of the innate inventory of functional heads, realized differently across languages but unified at the level of underlying architecture. This is essentialism at the feature level – GENDER has an essence, even if individual gender systems vary in their particulars.

The typologist's answer, Haspelmath's answer, is that GENDER *as a cross-linguistic category* is a COMPARATIVE CONCEPT: a yardstick we construct for measuring diversity, with no claim to psychological reality in any particular speaker's grammar.

Neither answer asks what I want to know. What maintains gender systems? Why do languages have them at all? Why do they persist across generations, given that they're largely arbitrary and impose a significant learning burden? What functional or acquisitional or interactional pressures keep the clustering clustered?

The same questions arise for NUMBER.

English marks number with a suffix: *cat*, *cats*. The exponence is segmental, concatenative, largely predictable. A child who knows *dog/dogs* can produce *wug/wugs* (Berko, 1958). The system is obligatory: you can't use a bare singular to refer to multiple cats. And it applies across the lexicon, with a handful of well-known exceptions (*sheep*, *deer*, *cattle*). What English grammaticalizes is a distinction between singular and non-singular reference.

Shilluk, a Western Nilotic language spoken in South Sudan, also grammaticalizes number, but the exponence looks nothing like English. Number is often marked through stem-internal changes in tone, vowel length, and voice quality, sometimes alongside suffixes (Remijsen & Ayoker, 2019). The singular and plural of a noun may differ only in whether the vowel is short, long, or overlong – a three-way length distinction that exists in the lexical phonology but surfaces morphologically in number inflection. Or they may differ in tone: a Low singular becomes a Fall in the plural, or a Mid becomes a High. Or both. The patterns are largely unpredictable. A speaker who knows one singular-plural pair can't reliably derive another; each pair must be learned. And number marking is not fully productive across the lexicon. As reported in (Remijsen & Ayoker, 2019, ch. 2), the database includes 906 singular nouns and 668 plural nouns; at least 238 singular nouns (more than a quarter) are either unspecified for number or lack a corresponding plural, and only nine plural nouns lack a singular. Number is grammatical and conditions agreement where it is specified, but the singular/plural contrast is not uniformly available for all nouns.

SINGULAR	PLURAL	GLOSS
tjɛ̀ɛɛl-ɔ̀	tjɛ̄ɛl	'leg(s)'
lèeec-ɔ̀	lɛ̄k	'tooth/teeth'
pʌ̀ʌm	pʌ̄́m̄-ī	'wooden board(s)'
bɔ̀ɔɔt̪-ɔ̀	bòot̪-ì	'craftsman/men'

Table 3.1. Number marking in Shilluk: stem-internal changes, often alongside suffixes (Remijsen & Ayoker, 2019, ch. 2)

When we say that English and Shilluk both have number systems, what are we claiming they share? Not the exponence – one is affixal, the other fusional. Not the predictability – one is rule-governed, the other lexically specified. Not the phonological substance – one involves segments, the other suprasegmentals. What remains is something more abstract: both languages grammaticalize number and the distinction can condition agreement, even though Shilluk's singular/plural contrast is not fully productive across the lexicon. Perhaps NUMBER is a natural kind after all – not at the level of morphological realization, but at the level of what the morphology is doing. Both systems, on this view, are grammatical reflexes of the same underlying operation: distinguishing individuated entities from pluralities.

Japanese suggests otherwise.

The noun *hito* 人 means 'person'. It can refer to one person or to multiple people; nothing in the grammar forces a choice. Where English requires *person* or *people*, Japanese allows *hito* for both. There is a reduplicated form, *hito-bito* 人々, but it does not mean simply 'more than one person'. It has a collective or distributive flavour: 'people as a group', 'various people', 'each of the people in question'. And it is not productively available across the lexicon. Some nouns reduplicate; most do not. The operation is lexically restricted in a way that English plural *-s* is not.

Meanwhile, if you want to count people in Japanese, you use a numeral with a classifier: *san-nin* 三人 'three people', where *nin* 人 is the classifier for human beings. The individuation of bounded entities – the semantic core that was supposed to unify NUMBER cross-linguistically – is handled by the classifier system, not by morphology on the noun. The noun itself remains unspecified for number.

Typologists file *hito-bito* under NUMBER nonetheless – calling it a plural, or a collective, or something in between (cf. Baloğlu, 2022; Forza, 2016). The disagreement is revealing. When we apply the vocabulary of number to Japanese, we are not sure what we are claiming.

So what do English, Shilluk, and Japanese share?

English has an obligatory singular/non-singular distinction; Shilluk has grammatical number and agreement but limited productivity, with many nouns lacking a paired form. The semantic claim – that both track individuation – seems defensible. Japanese has optional, lexically restricted, semantically flavoured reduplication that does not straightforwardly mean 'more than one', and a classifier system that handles the counting. If NUMBER is 'grammatical encoding of individuation', then Japanese nouns do not have number; the classifiers do the work. But typologists describe *hito-bito* as a number phenomenon. Are they wrong? Or is NUMBER a label we apply to a family of loosely related phenomena – obligatory affixation, unpredictable stem-internal changes, optional reduplication, classifier systems – that share a functional neighbourhood but not a common mechanism?

Definitional refinement can't solve this. No set of necessary conditions will group obligatory affixation, lexically specified stem alternations, optional reduplication, and classifier systems in a way that tracks what typologists actually want to generalize about. The unity of NUMBER, if it exists, lies in the mechanisms that maintain these clusters: whether the same acquisitional, processing, and transmission pressures keep the relevant properties co-occurring, or whether we are simply filing convergent solutions under one convenient label. That is the empirical question.

There are important exceptions. Corbett's work on gender and number asks why these systems exist and what functions they serve (Corbett, 1991, 2000). Bybee, Dahl, and the grammaticalization tradition treat tense-aspect-modality categories as emergent clusters maintained by usage patterns and diachronic pathways (Bybee et al., 1994b; Dahl, 1985). Typological work on case, evidentiality, and person addresses similar questions. But this literature rarely frames its insights as a general metaphysics of features. The question "what kind of thing is NUMBER?" is usually implicit, answered in passing rather than posed directly.

TENSE. ASPECT. MOOD. DEFINITENESS. PERSON. CASE.

Each of these is treated as a grammatical feature, a dimension along which languages vary, a parameter to be set or a category to be described. Each has generated an enormous literature. And for each, the explicit natural-kind question is usually background assumption rather than central topic.

Consider DEFINITENESS – a case I'll return to in detail in Chapter 10. The textbook definition says that definite expressions pick out referents identifiable to the hearer. English *the* is the paradigm case. But *the* appears in contexts where identifiability fails: *the hospital* in *go to the hospital*, *the tiger* in *the tiger is endangered*, *the average American* in statistical generalizations. Are these counterexamples to the definition, or extensions licensed by some deeper principle, or evidence that DEFINITENESS isn't a unified kind at all?

The usual move is to refine the definition: add accommodation, add familiarity, add uniqueness under a description. Each refinement handles some cases and strains on others. The definitional hunt continues.

The question I want to ask is different. What if DEFINITENESS and the form class that expresses it – the articles, the demonstratives, the possessives – are not one kind but two? The form side clusters by morphosyntactic properties: distribution, agreement, position. The meaning side clusters by semantic and pragmatic properties: identifiability, familiarity, specificity. These two clusters usually align. When they don't – the weak definites, the generic definites, the bridging cases – we see the seam between two distinct kinds, each maintained by its own mechanisms.

This is a different kind of question. It doesn't ask where to draw the boundary of DEFINITENESS. It asks whether DEFINITENESS is one thing or two, and what would tell us the difference.

The pattern should now be visible.

The essentialism-nominalism debate has focused on lexical categories: NOUN, VERB, ADJECTIVE, PREPOSITION. These are categories of items. The disputed cases are words – *fun*, *near*, *otherwise* – and the question is which category they belong to.

But grammatical theory also posits categories of a different sort: features, dimensions, systems. GENDER, NUMBER, TENSE, ASPECT, MOOD, DEFINITENESS, PERSON, CASE. These are not categories of items but categories of categories – ways of organizing the space in which items are located. A noun doesn't have a gender the way it has a phonological form. Gender is a system; the noun is assigned a VALUE within that system – what it counts as by virtue of its place in the paradigm. The term borrows from Saussure's *valeur*: the insight that linguistic units get their identity not from intrinsic properties but from their position in a system of contrasts.[1]

These second-order categories have received surprisingly little ontological attention. We ask whether NOUN is a natural kind. We rarely ask whether GENDER is. We ask whether a particular language has adjectives. We rarely ask whether TENSE is a single phenomenon or a family of loosely related systems that share a label.

The reason, I suspect, is that features feel more abstract, more clearly theoretical, less tied to the messy particulars of words and constructions. Features are the machinery of the grammar, not the things the machinery operates on. It's natural to focus ontological scrutiny on the things – the lexemes, the constructions, the sentences – and treat the machinery as given.

But the machinery is exactly what needs scrutiny. If NOUN is a natural kind maintained by mechanisms, why not GENDER? If PREPOSITION is a cluster of properties that tend to co-occur, why not TENSE? If the question "is X a natural kind?" is worth asking for lexical categories, it's worth asking for feature systems too.

And once you ask it, you notice something.

Many central debates about feature systems are more radical than the usual lexical boundary disputes. The debate about whether *fun* is a noun or an adjective is a boundary dispute: everyone agrees that NOUN and ADJECTIVE exist; the question is where to draw the line between them. The debates about feature systems are more radical. They concern not where the boundaries fall but whether the categories exist at all.

Is GENDER a primitive, or is it reducible to NUMBER – a mechanism for individuation that enables counting? Some theorists have argued exactly this (Lowenstamm, 2008; Picallo, 1991). If they're right, then GENDER isn't a natural kind; it's a surface manifestation of NUMBER, and the apparent cross-linguistic diversity of gender systems is variation in how languages deploy individuation.

Is TENSE a primitive, or is it really a species of ASPECT – aspect anchored to the speech moment? If tense is just deixis applied to aspectual distinctions,

[1] Saussure's value was synchronic – a snapshot of the system. The present account adds what he set aside: the mechanisms that maintain the system across time. A parallel comes from complexity science: Gell-Mann (1994) calls such relational structures SCHEMAS – compressed internal models that encode regularities and enable prediction. In complex adaptive systems, a schema is not a copy of the world but a coarse-grained encoding of its history, refined through feedback (Holland, 1992). Value, on this view, is what a schema delivers when queried: a learned expectation about how the unit will behave.

then TENSE as a separate category dissolves. What we're left with is ASPECT, differently parametrized across languages.

Is PERSON a primitive, or is it spatial deixis grammaticalized? First person as proximate, second person as distal, third person as remote? Some languages seem to work this way. If they all do, under the surface, then PERSON is not a sui generis category but a special case of spatial orientation.

These are not boundary disputes. They're existential disputes. The question is not "where does TENSE end and ASPECT begin?" The question is whether TENSE exists as a distinct kind at all.

Neither essentialism nor nominalism naturally foregrounds these questions. The essentialist can revise the list of essences post hoc – "we were wrong; TENSE isn't basic after all" – but the framework provides no procedure for discovering reductions, only for registering them after the fact. The nominalist can note that some taxonomies are more useful than others, but treats the difference as pragmatic rather than as tracking structure in the world. A mechanistic approach asks what would settle the matter empirically.

What maintains TENSE as a distinct cluster? What maintains ASPECT? If they're maintained by the same mechanisms – if the properties that cluster as "tense" in one language are held together by the same forces that cluster as "aspect" in another – then the reduction is real, and TENSE is not a natural kind but a surface variant of something deeper. If they're maintained by different mechanisms, then both are real, and the apparent overlap is convergent evolution, not identity.

These are empirical questions. They have answers, even if we don't yet know what the answers are. The framework that makes them askable is the framework that treats categories as maintained by mechanisms, not constituted by essences.

3.2 Grammaticality itself

There is one more question that neither tradition has posed, and it's the deepest.

Every analysis of every construction in every language presupposes a distinction between grammatical and ungrammatical sentences. The asterisk is the most common symbol in syntax. We mark **The cat slept the mat* as ungrammatical, *The cat slept on the mat* as grammatical, and we build theories to explain the difference. The distinction is foundational. Without it, there's no data.

But what kind of thing is grammaticality?

The essentialist answer is that grammaticality is a property: a sentence either has it or doesn't. This picture faces well-known difficulties. Judgments are gradient; speakers disagree. The standard defense is the competence/performance distinction: the grammar is binary, the usage is messy. But if competence is a biological/cognitive state, we should not expect it to be captured by a tidy checklist of necessary properties. Biology is full of robust, stable kinds whose unity is maintained by interacting processes rather than fixed essences; if competence is real in that sense, homeostatic structure is the better default hypothesis than an all-or-nothing definition.

The prototype theorist's answer is that grammaticality is gradient. This captures the phenomenology but faces the same explanatory gap: if grammaticality is gradient, why is the gradient stable? Why don't judgments drift randomly?

Here is a question that almost no one asks: Is GRAMMATICALITY itself a natural kind?

Not: is this sentence grammatical? That's the question we ask every day, the bread and butter of syntactic work. But: what kind of thing is the property we're probing when we ask that question? Is grammaticality a cluster of properties – acceptability, processability, learnability, attestation in corpora, consistency across speakers – held together by mechanisms? Or is it a single primitive, the output of an internalized grammar?

If it's a cluster, then the cases where the properties come apart – the sentences that are attested but judged unacceptable, or acceptable but never attested, or processable but not learnable – are evidence about the structure of the cluster. They reveal which mechanisms are doing which work.

If it's a primitive, then those cases are noise, or performance errors, or evidence that our elicitation methods are imperfect. The grammar is clean; the data are messy.

The choice between these views isn't a matter of taste. It's an empirical question about what grammaticality is and what maintains it. Chapter 14 will develop this question in detail. For now, the point is that grammaticality – the foundation of the entire enterprise – has received less ontological scrutiny than the categories it's used to define.

3.3 What Part II provides

The questions raised in this chapter – what maintains GENDER? is TENSE a natural kind? what kind of thing is grammaticality? – can't be answered yet. To answer them, we need a framework that makes mechanism-talk precise: an account of what it is for a category to be maintained, what counts as a mechanism, how mechanisms interact, and how to tell when they're absent.

That's what Part II provides.

Chapter 4 introduces homeostatic property cluster kinds – the framework developed in philosophy of biology, here applied to linguistics. It specifies the mechanisms: acquisition, entrenchment, interactive alignment, iterated transmission, functional pressure. The chapter shows how they interact to produce the gradient structure that essentialism struggles with.

Chapter 6 asks what it takes for a category to support induction. Why can we generalize from observed nouns to unobserved ones? Why do the properties that cluster as NOUN predict further properties we haven't yet checked? The answer lies in the mechanisms: categories maintained by robust processes support inferences that arbitrary groupings don't.

Chapter 8 develops the criteria for failure. When is a putative category too thin – too few properties, too weakly clustered? When is it too fat – multiple distinct kinds lumped under one label? When is it merely negative – a wastebasket category defined by what it's not? These diagnostics are what give the framework its teeth. They let us say, for particular cases, whether the mechanisms are sufficient to constitute a natural kind.

With that apparatus in place, Part III returns to the questions this chapter has raised. Chapter 10 takes up definiteness and argues that form and meaning constitute distinct but interacting kinds. Chapter 11 examines word classes – why nouns and verbs are stable across languages while adjectives and adverbs vary. Chapter 13 applies the framework across levels, from phonemes to constructions, showing that the same logic of clustering and maintenance applies at every scale. And Chapter 14 asks what we've deferred: what kind of thing grammaticality itself is, and what maintains it.

But first, the framework. The core idea is simple: categories are real because they are maintained. A noun is a noun not because it satisfies some checklist of necessary properties, but because the forces that cluster nominal properties – acquisition, frequency, analogy, transmission, function – keep clustering them.

Part II

The Fix

Categories without essences

> You should consider that the essential art of civilization is maintenance.
>
> —Pete Seeger, quoted in Brand (2024)

Part I raised a question: if categories aren't defined by essences, what makes them real? The answer this chapter develops is simple: categories are real because they are maintained. This is the maintenance viewMAINTENANCE VIEW. The maintenance view finds a natural ally in the Parallel Architecture (Jackendoff, 2002, 2010), which treats the lexicon not as a list of exceptions but as the very heart of the interface machinery.

The word "maintenance" is chosen deliberately. Stewart Brand's recent work on infrastructure, buildings, and software makes a point that applies equally to language: maintenance isn't just preventing breakdown; it's the whole process of keeping something going (Brand, 2024). Monitoring, repair, renewal, adaptation – these are what keep bridges standing and cities functioning. Brand's insight is that we systematically undervalue maintenance because its successes are invisible: a well-maintained system just *works*, and we notice only when it fails.

Linguistic categories are like this. Grammatical categories are the cleanest place to see it, but the same maintenance logic shows up in phoneme contrasts and in register conventions. When the system functions smoothly, the maintenance is invisible. We notice it only when it fails or when we look specifically at the seams.

Stated as a commitment:

> HPC commitment: A category is real when (a) it is associated with a robust cluster of properties, and (b) there exist processes that tend to produce and stabilize that cluster across the relevant range of contexts, yielding projectible generalisations.

The commitment is designed to be vulnerable. It can fail: the cluster can be too thin to support induction; the apparent cluster can be a measurement artefact; or the cluster can be maintained by multiple distinct processes, in which case the label covers several kinds rather than one. Those are empirical diagnoses, not definitional loopholes.

The idea of maintenance as constitutive – not just preservative – comes from philosophy of biology. Species aren't defined by genetic essences – no checklist of necessary and sufficient conditions separates one species from another. But species are real. They support induction, figure in explanations, persist across time. What makes them real is not a shared essence but a shared causal history: mechanisms of reproduction, selection, and development that keep certain properties clustering together.

Richard Boyd called these homeostatic property cluster (HPC) kinds – or just HPC kinds (Boyd, 1991, 1999). The name is technical but its parts are revealing. *Stasis*: standing, position – the cluster *stays* in place, maintained by ongoing processes. *Homeo*: same, similar – when disturbed, the cluster tends to return to the *same* configuration, not just any stable state. A homeostatic system doesn't merely persist; it self-corrects. Perturb a species' genome and selection pushes back toward the original distribution. Isolate a population and reproductive barriers emerge that restore the clustering. The name captures both the *staying* and the *sameness* – stability is not just inertia but active return.

Think of a spinning top. It stays upright not because it's rigid but because it's moving. The spin resists perturbation – push it slightly and gyroscopic forces bring it back. Stop the spin and the top falls. Homeostatic kinds are like this: their stability is dynamic, not static. What keeps the cluster clustered is not a fixed structure but an ongoing process. If this stability is real, it should have measurable signatures: judgment variance should spike near category boundaries; innovation should stall when it encounters alignment resistance; diffusion rates should differ for items that comply with the basin versus those that violate it. Those predictions are developed in §4.10.4.

Ruth Millikan provides deeper metaphysical grounding: "Anything with a structure that tends actively to maintain or reconstitute itself over time [...] maintains or increases its own kind while depleting materials and resources for constituting other kinds" (Millikan, 2017, ch. 1, p. 17). Self-maintaining structures crowd out competitors. The limited variety we observe – discrete species rather

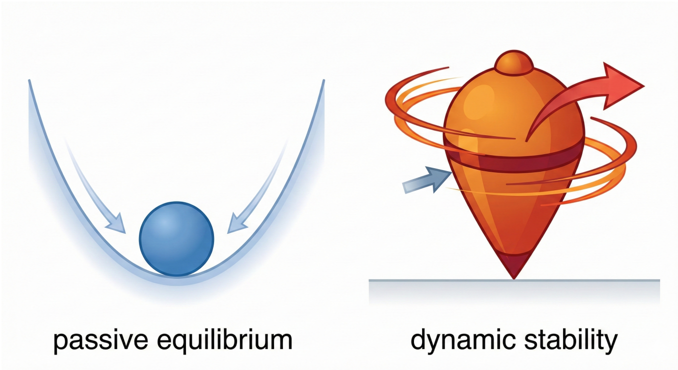


Figure 4.1. Two kinds of stability. Left: a ball at the bottom of a valley is in passive equilibrium – it settles where gravity deposits it. Right: a spinning top maintains dynamic stability – the spin actively resists perturbation, and when pushed, gyroscopic forces push back. Homeostatic property cluster kinds are like the top: their stability is achieved through ongoing processes, not through static structure.

than a continuum of forms, distinct linguistic categories rather than undifferentiated distribution – reflects not the prior structure of reality but the selective survival of self-stabilizing clusters. The asymmetry is counterfactual: maintained clusters are resilient in characteristic ways – perturb them and they tend to return; remove the maintenance and they dissolve. Merely convenient groupings lack this resilience. That's what distinguishes HPC kinds from bookkeeping.

And the maintenance needn't be entirely external. Some of the stability arises from reciprocal reinforcement among the clustered properties themselves. Recall *Smilodon* and *Thylacosmilus* from §2.7 (Figure 2.3): their massive canines, powerful forelimbs, solitary hunting, and apex-predator niche aren't accidentally co-present. They are causally interlocking – each makes the others more likely, more stable, more resistant to drift. Large canines require the musculature to wield them; ambush predation requires the solitude to stalk; apex status requires the whole package. The cluster is a self-stabilizing network, embedded in broader developmental, ecological, and transmission dynamics. This is the reflexive dimension of homeostasis: the properties, by virtue of being those properties, participate in maintaining the cluster.

The question of why creatures categorize at all – why cognition traffics in discrete kinds rather than continuous similarity gradients – is beyond this book's scope. The short answer is evolutionary: organisms that tracked projectable regularities outcompeted those that didn't (Quine, 1969). Millikan (1984) develops this insight into a comprehensive theory of biological function, while developmental work by Gelman and Markman (1987) shows that the tendency to essentialize categories appears very early in childhood, suggesting a cognitive default that education must work against rather than a learned strategy.

4.1 From species to categories

The biological parallel isn't decorative. Species and linguistic categories face the same philosophical problem: both look like natural kinds – they support induction, figure in explanations, persist across time – but neither has an essence in the classical sense. No genetic sequence is necessary and sufficient for tigerdom; no morphosyntactic property is necessary and sufficient for nounhood. But tigers are real, and so are nouns. The question is how.

Boyd's answer for species was that the clustering does the epistemic work while the mechanisms explain the clustering. Tigers share properties – striped fur, carnivorous diet, solitary hunting, particular vocalizations – not because these properties define *tiger* but because the mechanisms of reproduction, development, and selection keep them bundled together. The properties vary; not every tiger has all of them. But they cluster reliably enough that learning about one tiger tells you something about others. That's what makes *tiger* projectible – able to support inductive inference – even without a definition.

The same logic applies to grammatical categories. Nouns share properties – they head noun phrases, occur with determinatives, inflect for number, denote entities – not because these properties define NOUN but because the mechanisms of acquisition, entrenchment, and transmission keep them bundled. The properties vary; not every noun has all of them. But they cluster reliably enough that learning the category NOUN tells you something about how unfamiliar nouns will behave.

Ruth Millikan's work on COPIED KINDS sharpens the parallel (Millikan, 1984, 2017). A copied kind is a category whose members are produced from each other or from a common template – they share properties because they share a lineage, not because they share an essence. Biological species are copied kinds: each organism is a copy (with variation) of its parents. Artefacts like screwdrivers are copied kinds: each is produced from a template (with variation) that was itself copied from earlier templates.

Grammatical categories are copied kinds par excellence. Each generation doesn't invent nouns from scratch; children copy the category from their input, with variation, and pass it on. The properties cluster because the copying is high-fidelity enough to preserve them. The variation accumulates because no copying is perfect. The category persists because the lineage persists – an unbroken chain of transmission from speaker to speaker across millennia. Khalidi calls such history-defined kinds ETIOLOGICAL KINDS and defends their status as genuine natural kinds (Khalidi, 2013, pp. 106–112).

Millikan's later work on UNICEPTS sharpens the point (Millikan, 2017, ch. 3). A unicept tracks sameness through multiple fallible methods – we recognize a friend by face, voice, gait, or handwriting, with none privileged as definitional. The same structure applies to grammatical categories: we identify nouns through plural morphology, determiner compatibility, argument positions, semantic features, and more. None of these diagnostics is the "real" definition of nounhood; they are all ways of tracking the same maintained cluster. The category is what the diagnostics converge on, not what any single diagnostic defines.

The neural evidence is suggestive. Fedorenko, Ivanova and Regev (2024) argue that the brain's language network – frontal and temporal regions that respond selectively to linguistic meaning – exhibits functional cohesion across diverse tasks despite anatomical variability across individuals. This is evidence that stability-with-heterogeneity is respectable in cognitive science: natural-kind talk can apply where functional profile is stable even when anatomical substrate varies. The result is best treated as an existence proof rather than direct confirmation of the specific mechanisms developed here – but it shows that HPC-style kinds aren't peculiar to biology.

This is what makes the biological parallel load-bearing rather than merely illustrative. The mechanisms maintaining species cohesion – reproduction, gene flow, selection, developmental constraint – have genuine analogues in language. Not identical mechanisms, but mechanisms playing the same structural role: keeping properties clustered, filtering variation, stabilizing the kind across time.

William Croft makes the parallel explicit: languages are "populations of utterances", and linguistic variants compete in speaker communities much as phenotypes compete in biological populations (W. Croft, 2000). Salikoko Mufwene takes the analogy further, comparing languages to viruses rather than organisms – they depend on hosts (speakers), multiple can coexist in one host, and they exchange material (borrowing) more readily than species do (Mufwene, 2001). The analogy is imperfect but productive: it highlights that linguistic change operates via interactions in a population, not by reference to an immutable type.

Table 4.1 sketches the correspondences.[1] Not every detail maps, and that's fine. The claim isn't that language *is* biology but that the same style of explana-

[1] These are functional analogues, not homologues: the mechanisms occupy the same position in an explanatory structure without sharing causal architecture. Analogy is scientifically respectable; the parallel is load-bearing but not a claim of identity.

Table 4.1. Parallel mechanisms in biology and grammar

Function	Biological mechanism	Linguistic mechanism
Transmission	Reproduction	Acquisition
Filtering	Natural selection	Communicative success
Cohesion	Gene flow	Interactive alignment
Constraint	Development	Processing limitations
Attraction	Ecological niche	Functional pressure

tion works: mechanism-based, historical, non-essentialist. What maintains the clustering matters more than what defines the category.

4.2 What mechanisms mean

Mechanism is doing a lot of work here, so I should say briefly what I mean. By mechanism I mean an empirically tractable process with a characteristic signature under perturbation: change the inputs in a targeted way, and you can predict (within noise) the direction and timescale of the system's response. The mechanisms maintaining linguistic categories aren't mysterious: they're the processes that usage-based linguistics has been studying for decades (statistical learning, exemplar storage, analogical extension, routinization). When I say mechanisms "maintain" categories, I don't mean they keep categories frozen; maintenance includes change. A well-maintained building gets renovated; a well-maintained codebase gets refactored. The point is that the system keeps functioning, not that it stays identical. The detailed treatment of what makes a mechanism substantive rather than decorative – the perturbation criterion – comes in §4.10.1.

But how far does this constructive, on-the-fly view extend? A more radical position has recently emerged. Ambridge (2020a) argues that categories like noun and verb are not stored anywhere – not as essences, not as prototypes, not as schemas. Speakers store only individual utterances; the categories are emergent patterns, visible to the analyst but not represented in the mind. His analogy is the honeycomb: hexagonal structure is a real feature of beehives, but no bee represents hexagons. The structure emerges from the local behaviours of many bees following simple rules. Noun, on this view, is the linguistic honeycomb – real enough to describe, but not internally represented.

The analogy is useful but the conclusion doesn't follow. What makes the honeycomb real is not that bees lack hexagon-concepts; it's that the local behaviours – cell construction, wax physics, spatial constraints – reliably produce and maintain the hexagonal pattern. The structure persists because the mechanisms persist. This is exactly the HPC claim. Ambridge's "radical exemplar" model identifies the mechanisms – analogical generalisation across stored tokens – without recognizing that those mechanisms are what maintains the category as a real kind. The disagreement is about whether "emergent" means "less than real." HPC says no: emergent patterns can constitute genuine kinds, provided the mechanisms that produce them operate reliably. Hexagons are real. So are nouns.

4.3 The mechanisms themselves

Let me name the mechanisms concretely. This isn't an exhaustive list – Chapter 7 develops a more detailed treatment – but it's enough to show that "mechanisms maintain categories" isn't hand-waving.

A note on terminology: following Illari and Williamson's Illari and Williamson, 2012 consensus view, I take a mechanism for a phenomenon to consist of "entities and activities organized in such a way that they are responsible for the phenomenon." In the context of linguistic category maintenance, I use STABILIZER to refer specifically to the *functional role* a process plays in maintaining clustering, distinguishing this from the causal structure (the entities and activities) itself. A process like entrenchment is a mechanism (grounded in memory and frequency effects) that acts as a stabilizer for linguistic categories.

This is a META-OCCAM stance (Krakauer, 2023). Occam's razor typically demands parsimony in the theory – simple objects, few parameters. But for complex adaptive systems, the parsimony moves up a level. In evolutionary biology, natural selection is simple; organisms aren't. You don't need a more complicated theory of natural selection to explain an elephant than a worm. The same holds here: the stabilizers are tractable; the categories they maintain are complex. The framework predicts fewer mechanism types, not fewer category types.

Recall from Chapter 3 that a noun doesn't *have* a gender the way it has a phonological form – gender is a system, and the noun is assigned a value within that system. The same holds for the properties that cluster in any HPC: they are relational, defined by position in a system of contrasts. What mechanisms maintain is not just clustering but coupling – the reliable pairing of a form's distribution with its systemic contribution.

4.3.1 Acquisition

Children acquire grammatical categories without being taught definitions. No parent explains that nouns are words that can be preceded by determinatives and followed by plural suffixes. Yet children converge on category systems that are highly similar to their parents' – not identical in every detail, since languages with different inventories produce different category structures, but similar at the level of major recurrent clusters. This convergence happens reliably, across cultures, in just a few years – a feat that explicit definition struggles to replicate.

How? Not by innate specification – that just relocates the essence to the genome and doesn't explain why categories vary across languages. The HPC answer is that children track statistical regularities that reflect the underlying mechanisms. They acquire categories by being sensitive to the pressures that maintain them.

The evidence is strong. Distributional learning – tracking which words occur in which contexts – is sufficient to induce grammatical categories from raw input (Kallini et al., 2024; Mintz, 2003; Piantadosi, 2024; Redington et al., 1998). Infants are sensitive to these distributions from early in the first year (Gómez & Gerken, 2000). The categories children induce match adult categories not because children have innate category templates but because the same distributional patterns that define adult categories are present in child-directed speech.

Acquisition is a mechanism of maintenance because it transmits categories across generations. Each child who learns NOUN from input and then uses nouns in ways the next generation can learn from keeps the category alive. The mechanism is imperfect – children don't reproduce their parents' grammars exactly – and this imperfection is itself important. Variation enters through acquisition; selection operates on that variation; categories evolve. But evolution isn't dissolution. The category persists because acquisition is reliable enough, even though it isn't perfect.

4.3.2 Entrenchment

Frequency matters. High-frequency items are processed faster, stored more robustly, and resist analogical pressure more strongly than low-frequency items (Bybee, 2006; Diessel, 2007). This is entrenchment: the deepening of mental representations through repeated use.

Entrenchment maintains categories by anchoring them. The most frequent nouns – *thing*, *time*, *way*, *people* – are processed automatically, without conscious categorisation. They are the bedrock of the category, the items that every speaker agrees on, the deep-basin anchors around which less frequent items cluster. Change the behaviour of *thing* and you change what NOUN means. This is vanishingly unlikely. *Thing* is so entrenched that changing it is nearly impossible – it would require shifting the habits of every English speaker simultaneously.

The skewed distribution isn't incidental – it's functional. Wolters et al. (2024) showed that Zipfian input distributions facilitate children's word learning even in isolation tasks where cross-situational inference isn't required. High-frequency items serve as scaffolds: once entrenched, they provide stable reference points for acquiring lower-frequency forms. The very shape of natural language input is itself a stabilizer.

High-frequency items also grammaticalize faster. The path from lexical verb to auxiliary – *will*, *going to*, *have to* – runs through frequent, semantically general items. Low-frequency items don't grammaticalize because they're not processed often enough to undergo the phonetic reduction and semantic bleaching that grammaticalization requires. The category AUXILIARY exists because entrenchment creates a basin of attraction: items that enter the basin get pulled toward the prototype; items outside it stay lexical.

Entrenchment also determines which syntactically well-formed expressions count as established in context. In English, the way to ask about someone's age is *How old are you?*, with the response *I'm forty* or *I'm forty years old*. The calque from French would be *What age do you have?* and *I have forty years.* These clauses violate no syntactic rules of English, but the first set is entrenched for age-reporting and the second is not. As a result, the second set is constructionally misaligned for that functional slot – though *I have forty years* is perfectly fine as a response to a query about time until retirement. The appropriateness of an expression can depend on its entrenchment within a particular functional slot, not just its structural well-formedness.

4.3.3 Interactive alignment

Speakers accommodate to each other. In conversation, interlocutors converge on lexical choices, syntactic structures, even phonetic details (Garrod & Pickering, 2009; Pickering & Garrod, 2004a). This is interactive alignmentINTERACTIVE ALIGNMENT: the tendency to match your speech to your interlocutor's.

Alignment maintains categories by enforcing community-wide coherence. If I use *adult* as a verb and you accommodate, we've both reinforced the verb use. If you resist – if you pointedly say *be responsible* instead – you've exerted pressure against the innovation. Either way, the interaction shapes subsequent usage. Multiply this by millions of conversations and you get a distributed mechanism for maintaining shared norms.

The alignment needn't be conscious. Most accommodation is automatic, below the level of awareness. Speakers don't decide to match their interlocutors' syntax; they just do. This automaticity makes alignment a powerful stabilizing force. Categories persist not because speakers explicitly agree on them but because speakers implicitly coordinate, conversation by conversation, year by year.

4.3.4 Iterated transmission

Language is transmitted imperfectly across generations. Children don't reproduce their parents' grammars exactly; each generation introduces variation. You might expect this imperfection to degrade structure – random errors accumulating until categories dissolve into noise. But the opposite happens. Structure emerges.

Simon Kirby and colleagues demonstrated this experimentally (Kirby et al., 2008, 2015a). Participants learned an artificial language – nonsense words paired with meanings – and then reproduced it for the next "generation." The initial language was arbitrary: no pattern connected form to meaning. By generation ten, the language had become compositional – word parts systematically encoded meaning components. Structure emerged from the transmission process itself, as learnable variants outcompeted arbitrary ones.

The mechanism is a filter. Not all variants survive transmission equally. Easy-to-learn variants get reproduced faithfully; hard-to-learn variants get distorted or lost. Over generations, this filtering selects for systematicity. Languages become more learnable because learnable languages survive transmission; unlearnable ones don't.

For grammatical categories, the implication is direct. Categories that are easy to acquire – categories with clear distributional signatures, consistent morphological marking, coherent semantic cores – survive transmission better than categories without these properties. The categories we observe today are the residue of this filtering process, iterated across thousands of generations. They're structured because only structured categories could have survived.

But transmission isn't the only route. Raviv et al. (2019) showed that compositional structure can emerge within a single generation when communities include multiple interaction partners and an expanding meaning space. The key isn't generational succession per se – it's the presence of compressibility pressure and communicative feedback. Transmission chains amplify this pressure; multi-partner interaction creates it within a generation.

4.3.5 Functional pressure

Categories exist because they do something. Nouns package referents for tracking across discourse. Verbs encode events and their participants. Determinatives signal identifiability. The functions aren't arbitrary – they reflect communicative needs that languages address.

Functional pressure maintains categories by making them useful. A language without nouns would struggle to establish reference; a language without verbs would struggle to predicate. These aren't logical necessities – you can imagine communication systems without such categories – but they're practical near-necessities for the kinds of communication humans engage in: extended discourse, complex reference, hierarchical structure.

Where functional pressure is strong, categories are stable. Across languages, there is a robust tendency for lexicons to develop a division of labour between resources specialized for reference-tracking and resources specialized for predication, though the mapping to familiar "noun/verb" diagnostics varies. Where functional pressure is weak, categories vary. Not all languages distinguish adjectives from nouns or verbs; not all languages have adverbs as a distinct category. The functions these categories serve can be performed by other means – nouns can modify nouns; verbs can modify verbs – so the pressure to maintain them as separate kinds is lower.

Functional pressure also shapes category boundaries. The core of a category – the items where clustering is tightest – consists of items that serve the function most directly. The periphery consists of items that serve the function less well or serve multiple functions. Central items are central because the stabilizing mechanisms act most strongly on them: they are frequent, early-acquired, alignment-friendly, and functionally versatile, so they become the basin's anchors rather than merely its best exemplars.

4.3.6 Homeostasis or simple causation?

A critic might grant that mechanisms matter but question whether *homeostasis* is the right word. Muhammad Ali Khalidi, in a systematic critique of Boyd's HPC framework, argues that the homeostatic requirement is too restrictive (Khalidi, 2013, pp. 72–79). Many natural kinds – including paradigmatic ones like biological species – don't maintain equilibrium. Species evolve; the properties associated with them shift over generations. If homeostasis means returning to a stable state when perturbed, then evolving species don't qualify. Yet they're surely natural kinds.

Khalidi proposes a simple causal theory as an alternative: natural kinds are associated with properties where some (primary) cause others (secondary), without requiring feedback loops or self-correction. On this view, what matters is causal structure, not equilibrium. The kind *polymer* exemplifies the pattern: the primary property (consisting of long chains of repeating monomers) causes the secondary properties (viscosity, rigidity, no gaseous phase), but there's no mechanism pushing polymers back toward some ideal state. The causation is one-way, from structure to behaviour.

The critique deserves a careful response, because linguistic categories might work differently from biological ones.

For some grammatical categories, Khalidi's simpler picture may be adequate. What I called *thin* clusters – nonce coinages, idiolectal forms, ephemeral patterns that never stabilize – might indeed be merely causal. They have properties that cluster because function causes structure: a word coined to package a referent takes on nominal properties. But there's no feedback, no correction, no return to equilibrium when perturbed. The category exists only as long as the functional pressure persists; perturb it and it drifts or dissolves.

But for robust categories – the ones that persist across speakers and generations, that resist innovation, that pull deviant items back toward the prototype – something stronger is operating. The mechanism of *interactive alignment* is the key test case. When I use *adult* as a verb and you accommodate, we've jointly reinforced the pattern. But when you resist – when you pointedly say *be responsible* instead – you've exerted corrective pressure. The interaction didn't just transmit a pattern; it pushed back against deviation. That's homeostasis in the original sense: a system-level tendency to return to the same state, not just any stable state.

The spinning top metaphor from earlier captures the difference (Figure 4.1). Khalidi's simple causal theory describes the ball rolling downhill: structure causes behaviour, and the ball comes to rest wherever the slope deposits it.

Homeostasis describes the top: the spin actively resists perturbation. The question for each category is: ball or top? Passive drift or active return?

The answer will vary. Closed-class categories – determinatives, auxiliaries, subordinators – are heavily entrenched and tightly aligned. Speakers agree on what they are; innovations are rare and resisted. These behave like tops. Open-class categories are more variable at the margins. High-frequency items are top-like (try changing how *thing* works); low-frequency items may be ball-like. The HPC label applies best where feedback operates.

This heterogeneity is expected. Khalidi is right that not all kinds are homeostatic; the maintenance view accommodates this by allowing that different mechanisms maintain different kinds. The question "is this category genuinely homeostatic?" becomes empirical rather than definitional. And the diagnostic is concrete: does the category exhibit corrective feedback (alignment resisting innovation, entrenchment anchoring prototypes, transmission filtering anomalies), or does it merely exhibit one-way causation (function producing structure, with no path back)?

What would count as evidence? For homeostasis: repair and rephrasing in conversation, diffusion stalls for innovations that violate the category, accommodation asymmetries where resisting speakers exert more influence than accommodating ones. For mere one-way causation: innovation spreads or dies without corrective feedback, boundary positions drift as usage shifts, no systematic resistance. The methodology is corpus-and-experiment: track repairs, measure diffusion rates, test whether acceptability patterns converge under alignment pressure or diverge under exposure to deviance.

For the grammatical categories this book focuses on – the major parts of speech, the core morphosyntactic features, the constructions that persist across generations – the evidence favours homeostasis. These are categories where speakers correct each other, where high-frequency items anchor low-frequency ones, where transmission filters out deviants. They're not static – the spinning top precesses, the categories evolve – but they're also not passively drifting. Something is keeping them upright.

Matthew Slater's STABLE PROPERTY CLUSTER (SPC) framework offers a more satisfying resolution (Slater, 2015). Slater argues that Boyd's homeostasis requirement was always too narrow: "Many homeostatic mechanisms operate only for a time, or in some but not other contexts. But the epistemic roles of kinds apparently require more stability than the mere operation of such homeostatic mechanisms provide" (Slater, 2015, pp. 382–383). What matters is stability, not the particular source of that stability. A clarification: *maintenance* is my umbrella term; *homeostasis* names one important subtype (dynamic return to equilibrium); SPC provides a broader stability taxonomy that explains why "homeostasis" sometimes feels too narrow.

The SPC account recognizes multiple routes to stable clustering:

- Homeostatic mechanisms – Boyd's original account: causal feedback loops that restore the cluster when perturbed.

- Frozen accidents – Phylogenetic or historical inertia: properties cluster because they were once linked and nothing has disrupted them.

- Metastability – The cluster is thermodynamically unstable but kinetically frozen, like diamond persisting rather than converting to graphite.

- Causal isolation – Stability through the absence of perturbing pathways rather than the presence of correcting ones.

For linguistic categories, this pluralism is liberating. Some categories may be genuinely homeostatic – interactive alignment pushing deviant uses back toward the norm. Others may be metastable – persisting because no perturbation has been strong enough to dislodge them, not because feedback actively restores them. The question becomes empirical: what combination of stability sources maintains this particular category?

Slater also introduces a distinction between INSTANCE STABILITY (individual members resist losing cluster properties) and CLIQUISH STABILITY (finding some properties reliably indicates the whole cluster). The latter matters for induction: we care whether knowing an item is a noun licenses predictions about its other properties. Cliquish stability can hold even when instance stability is imperfect – even if individual nouns drift, the correlation structure persists. This is precisely what we observe for open-class categories: high-frequency items are instance-stable (anchored by entrenchment), while low-frequency items drift more freely, yet the overall clustering remains inductively reliable.

The SPC framework also permits domain-relativity: "Some collections of things only instantiate natural kinds from the perspective of particular sciences" (Slater, 2015, p. 392). Linguistic categories may be real for linguistics without being real for physics. This isn't second-class reality; it's the ordinary condition of special-science kinds. The mechanisms that maintain grammatical categories operate at the level of speakers and communities, not at the level of fundamental physics. Their reality is no less genuine for being domain-specific.

4.4 HPC kinds are not natural kinds

The HPC framework has drawn predictable philosophical objections. Linguists importing it should know what can safely be set aside and what must be addressed head-on.

The sharpest objection is Magnus's: HPC is neither necessary nor sufficient for natural-kind status (Magnus, 2014). Some natural kinds (fundamental physical kinds, perhaps) have essences after all; some things that cluster homeostatically aren't natural kinds (mere artefact-style regularities). The label "natural kind" brings metaphysical baggage – an expectation of mind-independence, cross-context invariance, perhaps fundamentality – that HPC was never designed to deliver.

The right response is: accept the critique but refuse the baggage. This book doesn't claim that grammatical categories are natural kinds in the honorific sense that would put them alongside electrons or chemical elements. The claim is narrower: grammatical categories are mechanism-maintained clusters that support scoped induction. That's enough for the epistemic work categories are supposed to do. If philosophers want to reserve "natural kind" for more fundamental entities, fine – call these HPC kinds and move on.

A second objection targets the mechanisms themselves. Onishi and Serpico (2022) press the question: what counts as a homeostatic mechanism? If the answer is "whatever keeps the properties clustered", then the claim is circular. If the answer specifies particular mechanisms (selection, reproduction, developmental constraint), then it may be too narrow for linguistics, where the stabilizers include acquisition, alignment, frequency weighting, and social transmission.

The response is methodological. A mechanism is substantive – not decorative – when it meets the perturbation criterion introduced in §4.10.1: change the inputs in a targeted way, and you can predict the direction and timescale of the system's response. Alignment is substantive because accommodation experiments show convergence (Pickering & Garrod, 2004a); acquisition is substantive because distributional-learning experiments show category induction (Mintz, 2003); transmission is substantive because iterated-learning studies show structure accumulation (Kirby et al., 2008). The mechanisms are heterogeneous, but they're testable. The circularity objection bites only if "mechanism" is left as a promissory note.

A third objection – from Lipski (2020) – is that HPC has been misconstrued as a normative gold standard. On this reading, if a category is an HPC kind, it's real and legitimate; if it isn't, it's fake. This sets up a false binary. The HPC framework, properly applied, outputs a diagnosis, not a verdict. A category can fail to be an HPC kind in several characteristic ways (Chapter 8 develops these as *thin*, *fat*, and *negative* failure modes), and each failure mode is informative rather than dismissive. A thin class lacks robust clustering – it's still useful for bookkeeping, but inductions from it should be held loosely. A fat class lumps heterogeneous subclusters – splitting it may improve projectibility. A negative class is defined by absence – it probably shouldn't anchor a research programme.

None of this is news for working linguists who already know that some categories are more robust than others. What the HPC framework adds is diagnostic structure: explicit criteria for distinguishing mechanism-backed kinds from measurement artefacts or taxonomic conveniences.

Finally, Craver raises interest-relativity: what counts as a mechanism depends on what questions you're asking. A biologist and an ecologist may carve the same population differently because their explanatory targets differ. The same will be true for linguists: a semanticist and a syntactician may posit different clusters over the same extension, maintained by different mechanisms, projectible for different purposes.

Chapter 6 addresses this head-on with the principle of FIELD-RELATIVE PROJECTIBILITY. Interest-relativity isn't a bug; it's a feature. Different stabilizers track different phenomena. The framework doesn't require that there be one true category for each extension. It requires that, for each category claim, the analyst specify the scope (population, register, time slice, grain), the stabilizers (acquisition, entrenchment, alignment, transmission, function), and the predictions (what inductions the category licenses). Where these differ, the categories are different – even if they share an extension. That's how definiteness and what I'll call DEITALITY – the purely formal cluster that travels with the article – can both be real kinds, overlapping in extension, maintained by different mechanisms, projectible for different analytical purposes (Chapter 10). The criterion is not whether extensions overlap – extensions often do – but whether the stabilizers and projectibility claims diverge in ways that cash out as different falsifiers.

The upshot is a decision rule for the diagnostic template:

> If the only "mechanism" story available is analyst convenience or measurement convenience, classify as a thin or fat convenience class, not an HPC kind.

My goal here isn't gatekeeping; it's calibration. The distinction between HPC kinds and convenience classes isn't a verdict on legitimacy – it's a prediction about what inductions will hold and where they'll fail.

4.5 What you get in return

We've been framing this negatively – kinds *without* essences. The positive formulation matters more: kinds *with* mechanisms, *with* stability, *with* explanatory power. From here on, the question is what maintains categories, not what they lack. In this sense, the maintenance mechanisms identified here function much like the interface rules in a PA framework: they are the "glue" that binds disparate representational clusters into a coherent functional unit.

The picture can be compressed into a slogan: *a category is a profile, stabilized by mechanisms, projectible relative to purposes.* Three parts, three levels – observable surface, metaphysical maintenance, epistemology indexed to practice. The remainder of Part II unpacks each piece; for now, the slogan is a reminder, not a substitute for argument.

And yet, despite everything Part I catalogued, the words do hold still – enough. Not because of what they are, but because of what keeps them. The stability isn't given; it's achieved.

Abstract frameworks are worth adopting only if they pay their way. What does the maintenance view buy that essentialism and prototype theory don't?

First, boundary cases become data. On the essentialist view, words like *fun* or *near* that don't fit standard categories are anomalies – evidence that the definitions need refining, or that the words are somehow defective. On the HPC view, these words are exactly what you'd expect. They're items near basin boundaries, where multiple mechanisms pull in different directions. Their instability isn't noise; it's evidence about where the boundaries lie and which mechanisms are operating. The framework transforms anomalies into informants.

Second, variation becomes tractable. Essentialism has trouble with the fact that speakers disagree about category membership – is *fun* an adjective or not? If categories are defined, disagreement means someone is wrong. On the HPC view, variation is expected. Different speakers have different input histories, different entrenchment profiles, different interlocutors to align with. Their categories will differ at the margins even if they agree at the core. The framework predicts variation and tells you where to expect it: near boundaries, for low-frequency items, in rapidly changing constructions.

Third, diachrony becomes visible. Categories change – *will* was a lexical verb; now it's an auxiliary. Essentialism treats change as replacement: the old category dies and a new one is born. But that's not what happens phenomenologically. *Will* didn't stop being a verb and start being an auxiliary overnight; it drifted, gradually, across a fuzzy boundary. The HPC view makes this visible: the mechanisms shifted, the basin migrated, the item moved from one cluster to another. Diachrony is category maintenance observed over longer timescales.

Fourth, cross-linguistic comparison becomes empirical. The typologist's question – do all languages have adjectives? – looks different under HPC. It's not about whether languages have categories that match a definition; it's about whether the same mechanisms produce the same clustering cross-linguistically. If the mechanisms that maintain adjectives in English also operate in Mandarin, then Mandarin has adjectives in the relevant sense – even if the surface patterns differ. If different mechanisms are operating, then the comparison fails. The question becomes tractable because it's about mechanisms, not definitions. Chapter 11 develops this in detail.

Fifth, the "right analysis" question transforms. Linguists spend enormous effort debating whether some item is "really" a noun or "really" a verb, whether some construction is "really" a clause or "really" a phrase. Under essentialism, these questions have determinate answers: the item either satisfies the definition or it doesn't. Under HPC, the questions are reframed. The issue isn't which definition is correct but which mechanisms are operating. An item can be genuinely intermediate – subject to mechanisms that pull it toward multiple categories – without anyone being wrong about what it is.

A clarification about what the maintenance view is *not* doing. This framework doesn't replace the empirical work of descriptive grammar, variationist sociolinguistics, or cognitive linguistics. It's not competing with Tagliamonte tracking quotatives or Bybee tracking entrenchment or Dixon cataloguing word classes. That work provides the data; the maintenance view provides an ontology that explains why the data pattern as they do.

Think of it as refactoring, not rehashing. When you refactor code, you reorganize it without changing what it does – making explicit the structure that was always there, revealing connections that were implicit, exposing assumptions that were hidden. The maintenance view refactors findings from across linguistics: what usage-based linguists call "entrenchment" becomes a stabilizing mechanism; what variationists call "apparent-time change" becomes evidence of transmission dynamics; what typologists call "cross-linguistic recurrence" becomes convergent maintenance. The data stay the same. What changes is the conceptual frame that makes them cohere.

This is not a manifesto demanding conversion. Sociolinguists should keep doing sociolinguistics; cognitivists should keep doing cognitive linguistics; descriptive grammarians should keep writing grammars. The maintenance view is offered as a complement – a neon sign for those who are curious about foundations. If you've wondered what grammatical categories actually *are*, or why your patterns are stable, or what licenses comparison across languages, here's one answer. It connects your work to philosophy of biology. Come have a look if you're interested.

4.6 Sciences that made the shift

Linguistics isn't the first science to confront the limits of essentialism. The shift from definitional to mechanistic thinking has precedents, and they're reassuring: the sciences that made the shift became more productive, not less.

Biology is the paradigm case. Pre-Darwinian biology sought essences – the form of the horse, the essence of the mammal. Species were natural kinds with definitions, even if the definitions were elusive. Darwin dissolved this picture. Species became populations – statistical aggregates of varying individuals, maintained by mechanisms of reproduction and selection, without fixed boundaries or defining properties. The shift was wrenching. It felt like giving up on rigour, abandoning the very idea of natural kinds.

But biology flourished. Population thinking – Ernst Mayr's term for the anti-essentialist perspective – enabled modern evolutionary biology, ecology, and genetics (Mayr, 1959, 1982). The same variation that embarrassed essentialism became the raw material for evolutionary theory. Species boundaries, once problematic because fuzzy, became informative because the fuzziness revealed ongoing evolutionary processes. The mechanisms of maintenance – selection, drift, gene flow – became the central objects of study.

Chemistry tells a different story but with the same moral. Pre-modern chemistry sought essences too – the essence of gold, the principles underlying material transformation. The essentialist picture collapsed, but what replaced it wasn't mechanism-talk; it was structural chemistry. Gold is the element with atomic number 79 – not because 79 protons define goldness but because atomic number determines chemical behaviour. The "essence" was relocated from manifest properties to causal structure.

The move from "what properties define the kind" to "what structure explains the clustering" is the same move HPC makes for linguistic categories. The issue isn't finding the right definition of NOUN; it's identifying the causal-structural facts that make nouns cluster. The facts are distributed – they're in acquisition, entrenchment, alignment, transmission, function – but they're no less real for being distributed.

Psychology went through a similar transition in the prototype revolution. Classical concepts were defined by necessary and sufficient conditions; Rosch showed that real concepts don't work that way (E. Rosch, 1978; E. H. Rosch, 1973). But prototype theory stalled – it described the structure of categories without explaining why they hold together. HPC picks up where prototype theory left off: the mechanisms are what maintain the prototype structure across speakers and time.

The moral is that abandoning essentialism doesn't mean abandoning rigour. It means redirecting inquiry from definitions to mechanisms. What maintains the clustering? What keeps the properties bundled? What explains the stability? These questions are harder than "what's the definition?" but they're the questions that lead somewhere.

4.7 Different categories, different profiles

The maintenance view doesn't claim that all categories are the same. Some linguistic categories might be more essentialist than others; some might be barely maintained at all. The framework expects heterogeneity.

Consider the contrast between so-called closed-class and open-class items. Closed-class categories – determinatives, subordinators, auxiliaries – are small, stable, and learnable as lists. A child can memorise every English determinative; there are roughly fifty. The mechanisms maintaining these categories operate differently than for open classes. Entrenchment is nearly total: every member is high-frequency. Variation is minimal: speakers agree on what the determinatives are. The tight maintenance produces a surface that *looks* essentialist – but the mechanism-cluster structure is identical; only the parameter settings differ.

Open-class categories – nouns, verbs, adjectives – are large, shifting, and effectively impossible to list exhaustively. New items enter constantly; old items drift or exit. The mechanisms operate statistically, across thousands of items, with substantial variation at the margins. These are paradigmatic HPC kinds: property clusters with gradient structure near the boundaries, maintained by mechanisms that shape the distribution without determining any individual case.

Phonological categories might differ from syntactic ones, but not in the way one might expect. Phonemes in spoken languages are maintained partly by articulatory and perceptual mechanisms tied to the vocal tract (Ohala, 1990). One might predict, then, that phonological categories should be cross-linguistically stable because the physiology is shared. Figure 4.2 illustrates how constraints like the sonority sequencing principle emerge from these pressures.

Sign languages complicate this picture productively. Languages from independent lineages – BSL, Japanese Sign Language, Israeli Sign Language, Kata Kolok (a village sign language of Bali) – all exhibit phonological structure despite having no common ancestor and using hands, face, and space rather than the vocal tract (Brentari, 1998; Marsaja, 2008; Sandler & Lillo-Martin, 2006). The substance differs radically from spoken language, but the organizational lo-

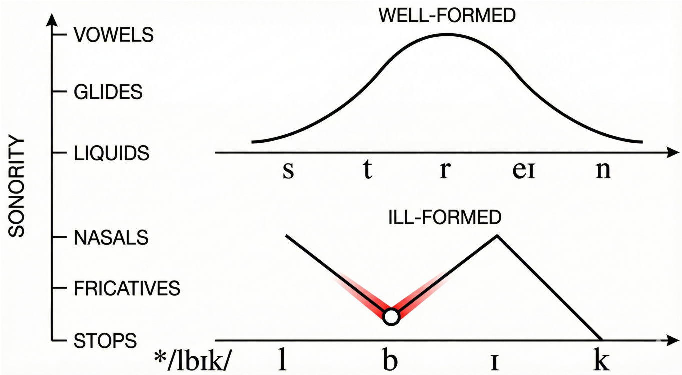


Figure 4.2. The sonority sequencing principle as an emergent constraint. Syllables rise toward a sonority peak (vowel) and fall away from it. Well-formed /streɪn/ follows this pattern; ill-formed *rnɛt violates it with a sonority drop in the onset. The constraint isn't stipulated; it emerges from articulatory and perceptual pressures that hold cross-linguistically.

gic converges: a small inventory of contrastive parameters (handshape, location, movement, orientation), combinatorial structure from meaningless units, systematic phonotactic constraints.

The convergence is specific and striking. Two-handed signs across unrelated sign languages obey similar constraints: if both hands move, they typically share handshape and mirror each other's movement (the SYMMETRY CONDITION); if only one hand moves, the other is restricted to a small set of unmarked handshapes (the DOMINANCE CONDITION) (Battison, 1978). These constraints aren't borrowed; they emerge independently, again and again, because they reflect facts about motor planning and perceptual processing that hold for any bimanual system. Figure 4.3 illustrates the pattern.

Nicaraguan Sign Language (NSL; Spanish acronym ISN) offers near-experimental evidence. NSL emerged in the late 1970s when deaf children were brought together in Managuan schools; no prior sign language was available as a model. Within two generations, NSL developed combinatorial phonology – discreteness, duality of patterning, minimal pairs – from holistic gestures (Senghas, 2005; Senghas et al., 2004). The mechanisms were visible in real time: children segmented continuous gestures into discrete components, regularized variation, imposed categorical structure. Phonological organization wasn't inherited; it was constructed by the same pressures that maintain it elsewhere.

Al-Sayyid Bedouin Sign Language (ABSL) in Israel confirms the pattern. Sandler et al. (2011) found that Battison's symmetry and dominance constraints were absent in first-generation ABSL – signers freely combined handshapes

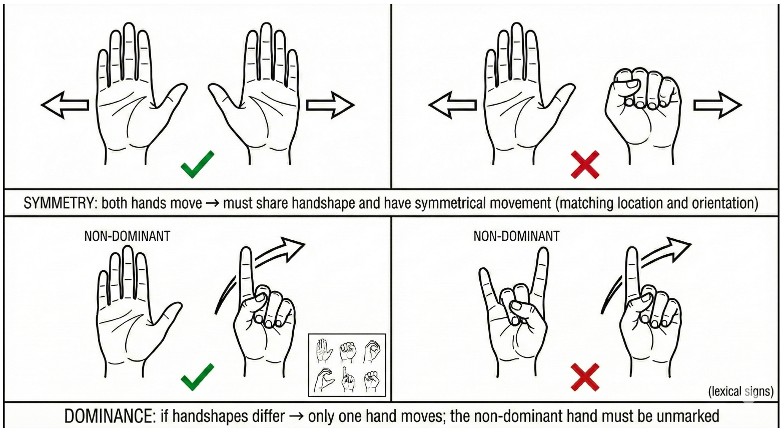


Figure 4.3. Phonological constraints in sign languages. Top: the symmetry condition requires that if both hands move, they share handshape and mirror each other's movement. Bottom: the dominance condition requires that if handshapes differ, only one hand moves, and the non-dominant hand uses an unmarked handshape. These constraints emerge independently across unrelated sign languages.

without systematic constraints. But over three generations, as the language became more conventional, those constraints emerged. The structural regularities weren't inherited from another sign language; they crystallized from the same cognitive and motor pressures that produced them elsewhere.

This suggests the mechanisms maintaining phonological categories are more abstract than vocal-tract physics: categorical perception, motor-planning efficiency, perceptual distinctiveness under noise, learnability across generations (Emmorey, 2002). These mechanisms operate over whatever articulatory-perceptual channel is available. The substrate varies; the clustering persists. This is HPC in action: different physical realizations, analogous mechanisms, convergent structure.

The framework doesn't need all groupings to be HPC kinds. Recall that I reserve CATEGORY for mechanism-maintained kinds (§2.2); what fails to meet that standard is a CLASS – a label without the causal structure to back it up. Some classes might be thin – lacking the stabilizing mechanisms to produce robust clustering, like nonce coinages or idiolectal forms that never diffuse. Some might be fat – pooling heterogeneous phenomena under a single label, like cross-linguistic umbrellas that obscure distinct mechanisms. Some might be negative – mere complements, wastebasket classes defined by what they're not. Part of the work, going forward, is identifying which grammatical categories are which. Chapter 8 develops criteria for diagnosis.

4.8 How determinacy survives

A worry hangs over the maintenance view: if categories don't have definitions, how can category membership be determinate? Isn't this just nominalism with extra steps – the view that categories are fictions, not facts?

The worry rests on a conflation. Determinacy doesn't require definitions; it requires causal grounding. A tiger is determinately a tiger not because it satisfies a checklist but because of its lineage – its causal-historical connection to other tigers, the mechanisms of reproduction and development that produced it. The same goes for linguistic categories. A word is determinately a noun not because it satisfies a definition but because the mechanisms that maintain NOUN have operated on it: it was acquired as a noun, used as a noun, aligned with other nouns, transmitted as a noun.

This is Millikan's point about proper functions (Millikan, 1984, ch. 1). An item's category membership is fixed by its history of production and use – by what mechanisms brought it into being and what mechanisms maintain it – not by what features it currently displays. The features matter because they're typically produced by the mechanisms. But when history and features diverge, history wins. A malformed tiger – one lacking stripes, say – is still a tiger because of its lineage. A malformed noun – one lacking typical nominal features – might still be a noun because of how it was acquired and used.

Where does indeterminacy arise? Where the causal pressures are genuinely mixed or transitional. Some items are hard cases not because we lack a definition to apply but because the mechanisms have been pulling in different directions: acquired as a noun but used increasingly like an adjective, with entrenchment patterns that are genuinely intermediate. The indeterminacy is real, but it's localized – a fact about this item in this transitional state – not a global fuzziness that infects the whole category.

Compare: are the Larus gulls circling the Arctic a single species or many? As you move around the ring – from Britain through Scandinavia to Siberia to Alaska to California – each population interbreeds with its neighbours. But the endpoints don't: herring gulls and lesser black-backed gulls, meeting in Britain, behave as distinct species. The mechanisms that maintain *species* – reproductive isolation, morphological clustering, ecological niche – are strong at the endpoints and weak at the contact points. The indeterminacy is real, but it's localized to the transitions. No one doubts that robins and eagles are distinct species; the ring species are hard cases precisely where the mechanisms are in tension.

The same will be true for linguistic categories. Most nouns are determinately nouns; most verbs are determinately verbs. The core is stable because the mechanisms maintain it robustly. The periphery is where indeterminacy lives – and that's not a problem for the framework; it's a prediction.

4.9 Recovering Aristotle

Aristotle has had a longer career as a caricature of essentialism than as a guide to how kinds actually persist. But as §2.1 noted, real Aristotle was subtler. His essences weren't arbitrary definitions; they were supposed to be explanatory. The form of a thing explained its behaviour; the essence was what made the thing the kind of thing it was.

HPC recovers this insight. The homeostatic mechanisms are, in a sense, the essence – causally understood. What makes a tiger a tiger is the network of properties maintained by reproduction, development, and selection. What makes a noun a noun is the network of properties maintained by acquisition, entrenchment, alignment, and transmission. The mechanism-cluster *is* the essence, relocated from the definitional to the causal register.

This is why HPC is sometimes called neo-Aristotelian (Boyd, 1999; Wilson, 1999). It recovers the structure Aristotle wanted – real natural kinds with genuine explanatory power – without the metaphysical baggage that made classical essentialism untenable. The kinds are real because they support induction. The induction works because the mechanisms maintain the clustering. The clustering is stable because the mechanisms are ongoing. It's a circle, but a virtuous one: mechanisms maintain clusters, and clusters sustain the mechanisms that maintain them.

The recovery isn't complete. Aristotle wanted essences to be intrinsic – properties a thing has independently of its relations to other things. HPC kinds are essentially relational: a tiger is a tiger because of its relations to other tigers (lineage), its relations to its environment (selection), its relations to developmental pathways (constraint). Linguistic categories are even more relational: a noun is a noun because of its relations to other nouns (analogy), its relations to speakers (acquisition, entrenchment), its relations to discourse (function). The essentialism being recovered is relational essentialism, not intrinsic essentialism.

That's a feature, not a bug – if the phrase itself is a bit prefab, that's fitting, because it trades on the same kind of relational uptake I'm describing. Linguistic categories are inherently relational – defined by distribution, opposition, system role. An account that makes them intrinsic would be false to what they are. HPC captures their relational nature while preserving their reality. The categories are real *because* they're relational – maintained by relations that keep properties clustered.

A related objection runs deeper. Some philosophers have argued that species are not kinds at all but historical individuals – entities with spatiotemporal boundaries rather than classes with members (Ghiselin, 1974; Hull, 1978). On this view, *tiger* names something more like a scattered object extended through space and time than a category with instances. The distinction matters for some purposes. But individuals, too, are maintained: what makes this tiger *this tiger* over time is not an essence but a cluster of properties held in place by metabolic, developmental, and ecological mechanisms. The HPC question – what mechanisms maintain the clustering? – applies to individuals as well as to kinds. The species-as-individuals move doesn't escape the maintenance framework; it relocates within it.

4.10 What "maintenance" commits us to

By the end of Chapter 3, the slogan may sound almost disarmingly simple: categories are real because they are maintained. But the phrase "maintained" can hide either a substantive commitment or a decorative one. The substantive version says: there are identifiable processes that (i) keep certain properties co-occurring, (ii) stabilize those co-occurrences against perturbation, and (iii) thereby support reliable induction. The decorative version says: things seem stable, so let us call them maintained. The difference is the difference between explanation and renaming.

Homeostatic property cluster theory was never meant as a license for the decorative version. Boyd introduced HPC kinds to capture a familiar pattern in the sciences: kinds that are not definable by necessary and sufficient conditions, and yet are not mere human conveniences either, because their property clusters are held together by causal mechanisms (Boyd, 1991, 1999). The mechanisms do not police membership by checking a definition. They make certain constellations more likely to occur and to persist, and those constellations are the basis for projection.

Recall the HPC commitment from page 55: a category is real when it is associated with a robust property cluster and there exist processes that tend to produce and stabilize that cluster. The point of formulating it this way was to make it vulnerable. It can fail: the cluster can be too thin to support induction; the apparent cluster can be an artefact of measurement; or the cluster can be real but maintained by multiple distinct processes, in which case the label covers several kinds rather than one. Those are empirical diagnoses that will matter later (Chapter 8).

4.10.1 Mechanisms are not metaphors

In linguistic contexts, "mechanism" can sound like a promissory note: a way of gesturing at explanation while postponing it. Here it is meant as a constraint on what counts as an answer. If a category is maintained, then there must be identifiable points at which pressure can be applied and outcomes predicted. Perturbation should not merely produce change; it should produce *structured* change.

The mechanisms appealed to in this book are heterogeneous in the mundane way that linguistic reality is heterogeneous. Some operate within an individual life (acquisition, entrenchment, alignment). Some operate across generations (iterated transmission). Some are neither purely individual nor purely social, because communicative function spans both: it shapes what is worth expressing and thus what survives learning and use. The mechanisms are not a finished list; they are the current best candidates for processes that are independently motivated and measurable.

Their shared role is not to define categories but to stabilize a landscape: to make some regions of grammatical space *attractive* and others *repellent*. A learner exposed to a community does not simply store tokens; they are pulled toward community-level regularities. A speaker in interaction does not merely express an idiolect; they coordinate. A system reproduced through imperfect learning does not simply degrade; it tends to reorganize into forms that pass more easily through the learning bottleneck. Taken together, these processes do not enforce crisp boundaries. They produce stability with seams, and discreteness with edge effects.

That is why boundary cases are not annoyances; they are probes. Culicover's *Syntactic Nuts* makes the methodological point plainly: the places where classifications strain are often the places where the causal story is visible (1999). If categories are maintained, then they should fail in characteristic ways under characteristic pressures. A framework that treats boundary cases as mere noise has surrendered the only data that can discriminate mechanisms.

4.10.2 Comparative concepts revisited

Haspelmath's insistence that cross-linguistic labels are comparative concepts has a force that an HPC view must respect. If typological categories were simply *read off* from minds, the widespread mismatch between diagnostic criteria across languages would be mysterious. Comparative work does require yardsticks, and yardsticks are constructed (Haspelmath, 2010).

But the crucial question is not whether the yardstick is constructed. It is whether the yardstick is *responsive*: whether it tracks a cluster that is supported by mechanisms operating in the systems being measured. The difference matters because it bears on induction. A purely stipulative yardstick can be useful for bookkeeping, but it will not reliably support the generalisations linguists actually want: that phenomena grouped under one label will behave similarly under new tests, shift similarly under perturbation, and correlate with other properties in stable ways.

On the HPC view, the reality of a cross-linguistic category is not a claim that speakers everywhere represent gender or number in the same format. It is a claim that, despite heterogeneity of exponence and local organization, certain clusters recur because similar pressures tend to generate and stabilize them. Where the pressures differ, the clusters should not merely look different; they should come apart in diagnostic ways. Where the label survives such pressure-testing, it earns its keep as more than a filing device.

This reframes the nominalist's challenge. The nominalist is right that definitions fail. The nominalist is also right that categories can be useful without being essences. What the nominalist declines to ask is whether some useful categories are useful *because* they track robust structure in the world—structure that can be discovered, tested, and sometimes revised. The HPC framework makes that question substantive rather than rhetorical.

4.10.3 From maintenance to discreteness

A remaining worry is that the HPC picture still sounds too continuous. If properties cluster because of graded pressures, why should the outcome not be merely graded as well? Why do linguistic categories *feel* discrete even when their substrates are not?

This is not a side question. It is the point at which essentialism earns its intuitive appeal. Speakers routinely treat category membership as all-or-nothing: a sentence is grammatical or it is not; a word in a given use is a preposition or it is not; a sound is heard as /ɪ/ or /ɛ/ even when the acoustic signal varies continuously. If a non-essentialist realism can't say where such discreteness comes from, it will look like it has traded ontological seriousness for a tasteful shrug.

The next chapter takes up that problem directly. The shift in analogy there is not a change of subject but a change of scale. Biology supplies the picture of kinds without essences; physics supplies a clean model of how categorical outcomes can emerge from continuous substrates without any policing definition. Phase transitions are not introduced as a metaphor for language, but as a proof of concept: systems can have real, projectible categories whose sharpness is a dynamical achievement, not an intrinsic essence.

That is the handoff. If categories are maintained, then their boundaries must be maintained too. And if boundaries are maintained, we can ask what kind of sharpness a maintained boundary can have—sharp in structure, yet difficult to locate; stable under ordinary perturbations, yet sensitive near the seam. Chapter 5 argues that this is exactly what we should expect, and shows how to make the expectation precise enough to test.

4.10.4 If the maintenance view is wrong

The claim isn't merely that categories are stable; it's that the stability is produced by mechanisms that can be identified, perturbed, and tracked. That claim has teeth.

If the maintenance view is wrong, here is what we should observe: category boundaries should not respond predictably to perturbations that target specific mechanisms. Shift acquisition patterns – change the distributional signatures in child-directed speech – and boundary locations should remain fixed if the mechanisms are epiphenomenal. Shift register distributions – flood a construction with informal-written tokens via online discourse – and the category should stay put if alignment is decorative. Shift transmission dynamics – isolate a community or introduce contact – and the clustering should be unaffected if iterated learning does no real work.

We should also fail to find the variance signature: distance-to-boundary should not predict judgment spread. If boundaries are brute rather than maintained, there is no reason for items near them to show heightened variance; the instability is unmotivated.

Conversely, if the maintenance view is right, the research programme writes itself. For each mechanism, identify a perturbation; predict a direction and timescale of shift; measure. The *fun* case is the cleanest available exemplar: register shift toward informal written + alignment in online discourse predicts faster adjectivalisation; diachronic corpus data should show boundary movement and, eventually, reduced variance once entrenchment catches up. That is the kind of prediction the framework is designed to generate.

Definiteness thread. This chapter's framework predicts that definiteness and deitality can be maintained by different stabilizers and thus constitute different HPCs even though they share an extension (Chapter 10). If that's right, diagnostics targeting the semantic properties (identifiability, uniqueness) should decouple from diagnostics targeting the morphosyntactic properties (*there*-resistance, partitive licensing). The prediction would be falsified if the semantic and formal properties proved fully coextensive across constructions, registers, and developmental stages.

Discrete from continuous

Thursday, 4 December 2025 was one of the most intellectually exciting days of my life.

I had conceived of this book – or perhaps I'd say the conceptual process had come to a head – the previous Saturday, when I created a project folder on my computer called `HPC book` and sent a synopsis to some friends with the subject line "Does linguistics need this book?" I've learned to treat the act of naming a folder as the weakest possible form of commitment, but it was enough to start the chain reaction.

Geoff Pullum said yes. Ryan Nefdt said yes – and added something I hadn't thought of:

> I personally favour a phase transition approach in which discreteness really is a feature of a system under a particular kind of measurement and context and continuity under a different one (like water moving from liquid to gas takes on different properties). So borrowing more from physics than biology.

I woke very early the next morning thinking about phase transitions – and about a paper I'd written two months earlier and then set aside. That paper had formalised Sorites tolerance using hyperreal numbers, inspired by work by Toby Ord. It had been "sitting in a drawer", waiting for a use. Now it had one.

By bedtime that evening, I had the first draft of this chapter.

Here is the puzzle I hadn't known I needed to solve.

5.1 The gradience problem

Linguistic categories look discrete: a word is a noun or it isn't, a sentence is grammatical or it isn't, a sound is /ɪ/ or /ɛ/ in the *pin*/*pen* contrast for the dialects that maintain it. But the properties underlying those categories are continuous: frequency distributions shade smoothly from one pattern to another, acoustic cues vary continuously (vowel formants don't come in bins), semantic features don't come with sharp edges. How do discrete categories emerge from continuous substrates?

Both definition-based essentialism and "unreal" exemplarism fail to solve this. As argued in Chapter 4, we need a view where emergent patterns maintained by reliable mechanisms constitute real kinds. We need DYNAMIC DISCRETENESS – discreteness that is achieved, not given.

The answer lies in scale. A change that's negligible when you're far from a boundary becomes appreciable when you're close. That's why tolerance intuitions coexist with boundary intuitions – both are correct, just in different regimes. And the mechanisms that maintain categories – acquisition, entrenchment, alignment, transmission – are what keep the boundary where it is.

A note on what follows: the discreteness is predicate-by-predicate. This framework doesn't force exclusivity – an item can be in two basins, as with dual membership in the noun-adjective zone. The worry that *fun* or *near* will be squeezed into one category or the other is addressed rather than suppressed. Overlap is allowed; the question is what maintains it.

5.1.1 The phase-transition intuition

Chapter 4 introduced a physical analogy: phase transitions. Water has no essence of liquidity; the same molecules constitute ice, liquid, and steam depending on temperature and pressure. The boundaries are real and sharp enough for skating or swimming, but they are maintained by dynamics, not definitions.

Phase transitions produce categorical outcomes (solid, liquid, gas) from continuous substrates (temperature, pressure). This suggests the solution: if physical systems can exhibit discrete, stable structure without essences, perhaps linguistic categories can too. We need to say precisely what it is about the dynamics that produces discreteness.

The key insight from physics is that phase boundaries are *scale-sensitive*. Near a phase boundary, small changes in temperature or pressure can flip the system from one phase to another. Far from the boundary, the same small changes have no categorial effect – the system remains solidly in one phase. The boundary is sharp, but its sharpness is a feature of the system at a particular scale of observation. Zoom in far enough and you find molecules fluctuating; zoom out and you find stable ice or stable water. The discreteness is real, but it emerges at the macroscopic level from continuous microscopic variation. Macroscopic discreteness is measurement-indexed: phases are crisp relative to thermometers, not relative to individual molecules.

The same logic applies to vague predicates generally. Consider a heap of sand. Ten thousand grains is clearly a heap. One grain is clearly not. The boundary seems sharp – at some point, removing one more grain tips the collection from heap to non-heap – but we can't pinpoint it. This is the Sorites paradox. Standard responses include epistemicism (the boundary is sharp but unknowable), degree theory (heaphood comes in degrees), and supervaluationism (multiple precisifications are equally acceptable). None quite captures the phase-transition intuition: that the boundary is sharp at the right scale, located in a determinate region, and maintained by dynamics rather than essences.

5.1.2 Relative tolerance

To make this precise, consider what "one grain doesn't matter" really means. It can't mean that removing one grain *never* makes a difference – that would entail that even a single grain is a heap. What it means is that removing one grain doesn't matter *when you have many grains*. Ten thousand minus one is still clearly a heap. Fifty minus one might not be.

The difference is relative, not absolute. Removing one grain from ten thousand is a 0.01% change. Removing one grain from fifty is a 2% change. The same absolute change has different significance depending on the scale.

This suggests a principle:

> Relative Tolerance: Changes that are negligibly small *relative to the current scale* preserve category membership. Changes that are appreciable relative to the current scale may not.

The principle is intuitive. When you're dealing with a clear heap – thousands of grains – removing one is negligible. When you're down to borderline cases – dozens of grains – removing one is appreciable. Tolerance holds in the first regime and fails in the second. The boundary lies where negligible changes accumulate into appreciable ones.

But "negligible" and "appreciable" are themselves vague. To make the principle precise, we need a framework that can handle the distinction rigorously. This is where nonstandard analysis helps.

The next section is therefore not a philosophical aside. The formalism is designed to derive a concrete empirical prediction: that distance-to-boundary should predict judgment variance, with a characteristic functional form. If the hyperreal story is wrong, that variance signature should not appear. If the story is right, the variance pattern is not just compatible with the framework – it falls out of the mathematics.

5.2 A formal solution

> Skip path. The next subsection develops a formal model using hyperreal numbers. Readers who prefer to skip the mathematics can take away three points and rejoin at §5.3:
>
> 1. Tolerance is scale-sensitive: small changes are tolerated, but they accumulate.
> 2. Sharp boundaries exist at thresholds we can't precisely locate.
> 3. Discrete categories and gradient intuitions are compatible – the discreteness is in the structure, the gradience is in our access to it.

5.2.1 The hyperreal formalization

The hyperreal numbers extend the reals with *infinitesimals* – quantities greater than 0 but smaller than any positive real number (Goldblatt, 1998). Robinson's construction is rigorous, but we do not need its details. Here infinitesimals are a modelling device for making *negligible relative to the current scale* precise without stipulating an arbitrary finite cutoff. I am not claiming speakers compute with hypernaturals; the hyperreals play the role calculus plays in physics. The application to vagueness is developed fully in Reynolds (2025b).

Here is how the framework applies to vague predicates. Let P be a predicate like *is a heap*, and let $\mu(x)$ be a measure of the relevant property – say, grain count. For successive cases c_i and c_{i+1} in a series, let $\Delta\mu_i = \mu(c_{i+1}) - \mu(c_i)$ be the change from one case to the next.

The Relative Tolerance principle becomes:

If $|\Delta\mu_i|/\mu(c_i)$ is infinitesimal, then $P(c_i) \leftrightarrow P(c_{i+1})$.

In plain language: if the fractional change is infinitesimally small, the predicate's truth value is preserved. This captures the intuition that proportionally tiny changes don't matter, while allowing that accumulated changes can.

Now model the entire Sorites series – from clear heap to clear non-heap – as a *hyperfinite chain* of ω steps, where ω is an infinite hypernatural number (Figure 5.1). This is a standard construction in nonstandard analysis: a sequence

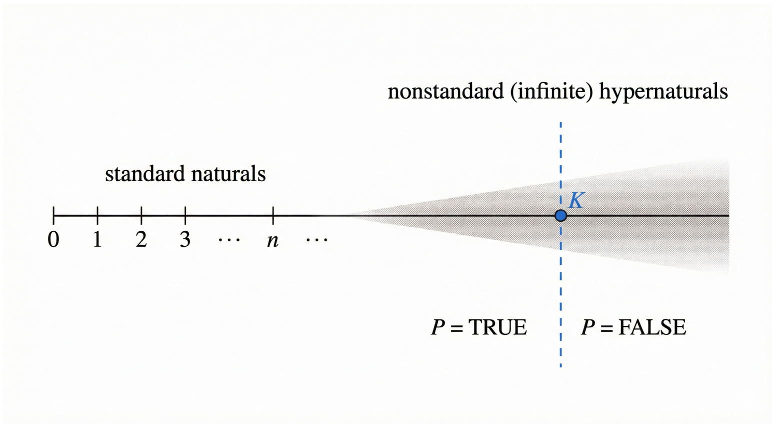


Figure 5.1. The hyperfinite Sorites chain. The predicate P (e.g., *is a heap*) holds at every standard natural index: tolerance preserves truth across any finite portion of the series. The cutoff K lies at a hypernatural index beyond all standard naturals – determinate within the model but not finitely specifiable.

indexed by hypernaturals, behaving internally like a finite sequence but containing infinitely many elements from the external perspective.

Within this hyperfinite chain, by the transfer principle of nonstandard analysis, any monotone sequence from P-true to P-false has a least index K at which P flips. This K is a hypernatural – not a standard natural number, but a number in the extended system. And crucially, because we've stipulated that P holds at all standard indices (encoding the intuition that tolerance holds throughout any finite portion of the series), K must be nonstandard: infinitely large compared to any standard natural.

The picture that emerges:

- Far from K: The fractional change at each step is infinitesimal. Relative Tolerance applies. The predicate is preserved.
- Near K: The fractional change is appreciable (no longer infinitesimal at this scale). Relative Tolerance is silent. The predicate can flip.
- At K: The boundary. Sharp, determinate, located at a specific hypernatural index – but epistemically inaccessible because we can't finitely specify which hypernatural.

This is why the Sorites induction fails. The inductive premise – "if $P(c_i)$ then $P(c_{i+1})$" – holds only where Relative Tolerance applies, which is far from the boundary. Near the boundary, the premise is false. The chain breaks not because tolerance is non-transitive, but because tolerance doesn't apply uniformly across the entire series.

5.2.2 Sharp boundaries, fuzzy appearances

The hyperreal model maintains classical bivalent logic throughout. At every index – including hypernatural indices – the predicate is either true or false. There are no degrees of heaphood, no fuzzy membership values. The boundary is sharp.

But the boundary is also inaccessible. It lies at a hypernatural index that we can't specify using finite means. From our finite observational standpoint, we see clear heaps, clear non-heaps, and a region of uncertainty in between. The uncertainty is epistemic, not semantic: there is a fact of the matter about where the boundary falls; we just can't determine what it is.

The hyperreal view differs from epistemicism – Timothy Williamson's view that vague predicates have sharp boundaries, unknowable because of the limits of our discriminatory capacities (Williamson, 1994). Both views yield sharp-but-inaccessible cutoffs.

But the sources differ. For the epistemicist, the boundary is a brute metaphysical fact. For the hyperreal view, the boundary's existence is a consequence of scale-sensitive tolerance plus monotonicity: the transfer principle entails a cutoff at some hypernatural index. The sharpness falls out of the mathematics; the inaccessibility is built into what kind of thing a hypernatural index is.

The explanatory consequences diverge more sharply. Epistemicism holds that boundaries are static: fixed by semantics plus worldly facts, unchanging unless the predicate's meaning changes. The maintenance view ties the effective boundary to mechanisms of acquisition, entrenchment, alignment, and transmission, and therefore predicts structured boundary *drift* under perturbations – something epistemicism can accommodate, but does not on its own explain or forecast. The hyperreal apparatus, in short, is not ornamental: it is the formal face of a framework that predicts structured drift and variance signatures where epistemicism predicts only static inaccessibility.

This explains why tolerance intuitions and boundary intuitions can coexist. When you have a clear heap, you're right that one grain doesn't matter – the fractional change is negligible at that scale. When you're near the boundary, you're right that the situation is unclear – you're in the region where tolerance breaks down. Both intuitions are correct; they just apply to different regimes.

The model also explains why boundaries are stable. The boundary isn't an arbitrary stipulation imposed by speakers or analysts. It emerges from the structure of the hyperreal model, determined by the interplay between the tolerance principle and the monotonicity of the series. Different choices of nonstandard model (technically: different choices of ultrafilter in the construction) yield different specific values of K, but the structural features – sharp boundary, epistemic inaccessibility, scale-dependent tolerance – are invariant.

A different response to the discreteness problem is available. Khalidi, extending his critique of Boyd's homeostatic requirement, argues that natural kinds can have genuinely *fuzzy* boundaries – not sharp-but-inaccessible, just fuzzy (Khalidi, 2013, pp. 63–69). Chemical isomers shade into each other as bond angles vary continuously. Biological species intergrade where populations overlap. If these paradigmatic natural kinds tolerate fuzziness, why should grammatical categories require sharp edges?

The answer lies in the structure of tolerance intuitions themselves. The Sorites reasoning pattern – one grain doesn't make a difference, so no number of grains makes a difference – is compelling precisely because we feel the tolerance premise is true. Khalidi-style fuzzy realism dissolves the puzzle by denying that there's a boundary at all: categories just shade into each other, and the puzzle evaporates. But this misses the explanatory target. Speakers don't behave as if categories shade continuously. They exhibit *scale-dependent* tolerance: small changes are tolerated, large changes flip categorisation, and the transition is experienced as sudden even when the underlying change is gradual. The hyperreal model captures this phenomenology. Fuzzy realism doesn't.

This isn't to say Khalidi's "fuzzy kinds" are never apt. For weak linguistic categories – the thin ones from §4.7 – genuine fuzziness may be the right description. But for robust categories, where speakers make categorical judgments, exhibit abrupt transitions, and show heightened variance specifically near boundaries, the hyperreal model's sharp-but-inaccessible picture fits the data better. The choice between models is empirical, not metaphysical: does the category exhibit gradient shading (Khalidi) or scale-dependent tolerance (hyperreals)?

5.3 From geometry to mechanism

5.3.1 From heaps to categories

Grammatical categories aren't heaps. But they face the same discreteness problem. If the underlying properties – frequency, phonetic realization, semantic features, distributional patterns – vary continuously, how do discrete categories emerge?

The hyperreal framework extends naturally to this multi-dimensional case. The single dimension of grain count becomes a multi-dimensional feature space; the single predicate P becomes a family of predicates $P_1, \ldots, P_n$ corresponding to different categories; and the linear Sorites series becomes a space of possible items, each located at some point in feature space.

The Relative Tolerance principle generalizes:

> If $d(\mathrm{x}, \mathrm{x} + \Delta\mathrm{x})/d(\mathrm{x}, \partial R_i)$ is infinitesimal, then $P_i(\mathrm{x}) \leftrightarrow P_i(\mathrm{x} + \Delta\mathrm{x})$.

Here x is a point in feature space, Δx is a perturbation, d is a metric on the space, and $d(\mathrm{x}, \partial R_i)$ is the distance from x to the boundary of category i's region. In plain language: perturbations that are infinitesimally small relative to the distance to the nearest boundary preserve categorization.

But where does ∂R_i come from? Not from an essence; that would be circular. The boundaries are estimated empirically – from distributional diagnostics, behavioural experiments, classifier decision surfaces – and the formalism characterizes what tolerance around that surface entails. Alternatively, boundaries can be treated as theoretical objects maintained by mechanisms, their locations inferred from invariances: if a diagnostic holds across contexts and tasks, it is tracking a stable basin, not an artefact of measurement. The reciprocals and specification-curve methods introduced later (§5.3.3) are exactly this kind of probe.

The metric question. The notation $d(\cdot, \cdot)$ raises an immediate question: what metric? Grammatical features are heterogeneous. Some are ratio-scaled (frequency, duration). Some are ordinal (degree of acceptability). Some are binary (takes plural marking: yes/no). No single off-the-shelf metric applies to all.

The answer is that we don't need a uniquely privileged metric; we need a notion of *local small change* that is robust across reasonable ways of scaling and weighting features. Any two distance measures that agree on which perturbations count as tiny in the region of interest will support the same Relative Tolerance story.[1]

[1] One way to formalise this is local equivalence of metrics (e.g. local Lipschitz bounds), which guarantees that the same neighbourhoods count as "nearby".

There's a principled way to construct such metrics for heterogeneous feature spaces. The strategy, familiar from statistical ecology and clustering analysis, proceeds in two steps (cf. Gower, 1971).

First, embed each feature into a dimensionless numerical scale:

- Ratio and interval features (frequency, duration, formant values) are standardized – divided by range or standard deviation – so that a unit change in one dimension is comparable to a unit change in another.
- Ordinal features (ranked acceptability, degree of grammaticalization) are mapped to their rank position, rescaled to $[0, 1]$.
- Binary features (presence/absence of inflection, compatibility with a construction) are left as $\{0, 1\}$.

After embedding, all coordinates are dimensionless and live in a common numerical space.

Second, place a weighted norm on the embedded vector:

$$d(\mathrm{x}, \mathrm{y}) = \left(\sum_k w_k \, |\phi_k(\mathrm{x}) - \phi_k(\mathrm{y})|^p \right)^{1/p}$$

where ϕ_k is the embedding function for feature k, w_k is a weight reflecting that feature's contribution to the homeostatic cluster, and $p = 1$ or $p = 2$. The weights can be estimated empirically (from regression coefficients, factor loadings, or feature-importance measures in classification tasks) or treated as parameters to vary in thought experiments.

The embedding is a heuristic, not a discovery procedure for Platonic dimensions. We don't expect factor analysis or multidimensional scaling to reveal the true axes of grammatical space – there may be no unique decomposition. What we expect is that any adequate embedding will recover dimensions that are interpretable: dimensions that correlate with independently motivated grammatical or semantic scales. If a derived dimension tracks degree-modification compatibility, and degree-modification compatibility tracks the conceptual scale of gradability, then the embedding is doing its job.

This construction is a normed cousin of Gower distance, widely used for mixed-type data (Gower, 1971), and it aligns with Gärdenfors's *conceptual spaces*: multiple quality dimensions grouped into domains, each with its own metric, combined via weighted aggregation across domains (Gärdenfors, 2000). The connection to HPC is direct: the weights w_k encode which properties play a stronger homeostatic role. Properties central to the cluster get large weights; free riders get small ones.

The point isn't that this construction is uniquely correct. The point is that a class of admissible metrics exists, all yielding the same small-change structure. The Relative Tolerance condition holds for all of them, not just one distinguished choice. Figure 5.2 illustrates the resulting landscape.

A clarification about dimensions. Some are conceptual in Gärdenfors's sense – grounded in perception and cognition, interpretable as quality scales that structure thought independently of any particular language. Eventivity, animacy, degree-scale structure: these show up in acquisition patterns, constrain semantic extension, have correlates in non-linguistic cognition.

Others are grammatical diagnostics: compatibility with plural morphology, occurrence in comparative constructions, position relative to the head. These may not correspond to conceptual dimensions at all. They're distributional reflexes of how a language happens to mark the underlying distinctions.

This means the feature space is hybrid, combining conceptual geometry with grammatical symptomology. The HPC claim is that homeostatic mechanisms operate on both – that acquisition pressures and functional demands keep conceptual and grammatical dimensions aligned. Where they pull apart, the basin may shift or split. The geometry isn't pure cognition and it isn't pure grammar; it's the joint product of both.

English evidentiality offers a worked example. English has no grammaticalized evidential marking – no obligatory morphology indicating whether information is firsthand, reported, or inferred. But the conceptual dimension exists: speakers make evidential distinctions using lexical means (*apparently*, *reportedly*, *I saw that*).

In languages with grammaticalized evidentiality – Tibetan, Turkish, many Amazonian languages – the grammatical and conceptual dimensions are aligned; the basin structure has both. In English, the conceptual dimension floats free of any grammatical basin. English speakers should show no clustering on evidential diagnostics, even though they make evidential distinctions conceptually. This is alignment failure without semantic loss. The concept persists; the grammatical basin doesn't form.

Visualizing the landscape. The three-dimensional surface we have been imagining is a visualization: the potential function V is defined over the full feature space $X \subseteq \mathbb{R}^n$, with category cores corresponding to local minima and boundaries to the stable manifolds of saddle points. Any particular "valley and ridge" picture is the graph of V over a low-dimensional slice of X on which the relevant contrasts are visible. Choosing the metric – the weights w_k and the embedding functions ϕ_k – is, in effect, choosing a geometry for the energy landscape: which

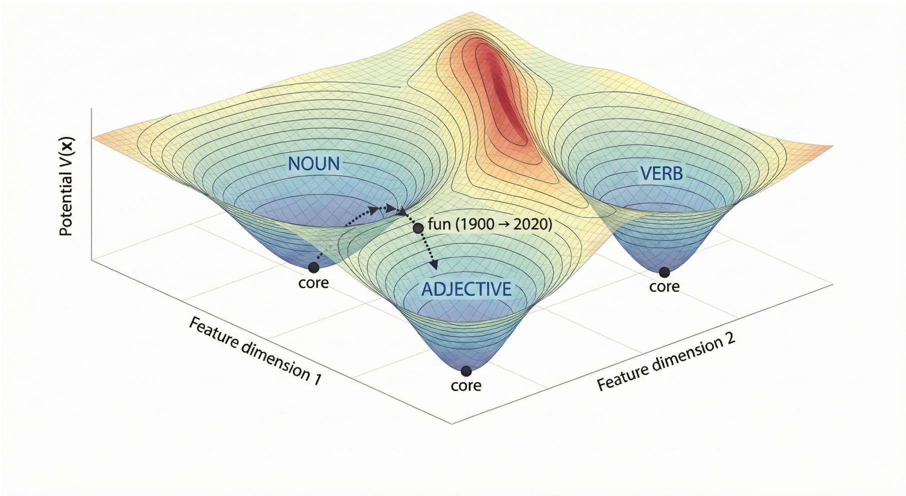


Figure 5.2. A two-dimensional slice through grammatical feature space, with the potential function $V(x)$ plotted vertically. Category cores correspond to local minima; boundaries to ridges and saddle points. The noun and verb basins are separated by a high ridge (disjoint categories); the noun and adjective basins share a low saddle (porous boundary, overlapping membership possible). The trajectory of *fun* illustrates diachronic movement from deep in the noun basin toward the noun–adjective boundary.

directions are steep, which are shallow, which features matter most for which basin. Relative Tolerance then says: stay within one valley when perturbations are small in that geometry.[2]

Basin structure. With a metric in hand, the multi-dimensional picture comes into focus. Each category occupies a region in feature space – a basin of attraction. Items deep in a basin are stably categorised: small perturbations don't move them out. Items near the boundary are unstable: the same perturbation that would be negligible elsewhere can flip the categorization.

The basins are typically convex. If two items are stably categorised as nouns, items intermediate between them in feature space should generally be too. This is Gärdenfors's criterion for natural categories, and it holds here because the mechanisms maintaining the basin – entrenchment, analogy, transmission – op-

[2]The physics metaphor is heuristic, not literal. Potential minima correspond to prototypes (geometrical centres of categories); basins of attraction correspond to convex regions where similarity to the prototype dominates; "forces" are learning and update processes – acquisition, entrenchment, alignment – that shift representations toward or away from category centres. The metaphor earns its keep by unifying these phenomena under a single image; it doesn't commit us to actual energy functions computed by the brain.

erate by similarity. Items near the prototype pull their neighbours toward the same categorisation. Apparent non-convexity, where it arises, may indicate that the relevant metric isn't the one assumed, that multiple overlapping basins are being conflated, or that the category has a multi-peaked structure – each of these is informative about the mechanisms at work.

The potential-well metaphor captures this: a convex basin corresponds to a single local minimum, and gradient descent from anywhere in the basin leads to the same attractor.

The boundary cases that trouble essentialism – *fun*, *near*, *otherwise* – are items near basin boundaries. They exhibit properties of multiple categories because they're in the region where tolerance breaks down. Their instability isn't noise; it's evidence about where the boundaries lie.

The maintained categories behave as if they have sharp decision boundaries under ordinary measurement regimes; what is graded is stability and typicality, not membership. At any point in feature space, an item is either in category i or not under the mechanisms currently in play. The appearance of gradience arises partly from epistemic limitations – we observe items at various distances from boundaries and can't determine exactly where the boundaries fall – but also from *synchronic context effects*. The same word may be perceived as more noun-ish or adjective-ish depending on the syntactic frame; the same phoneme may shift category depending on rate-normalised expectations. The basin metaphor accommodates this: mechanisms set the default landscape, but context induces local deformations. Think of the terrain not as a static watershed but as something that flexes moment-to-moment, with items pulled toward or away from boundaries by the cues available in the immediate environment. This is REAL GRADIENCE – gradient structure that is genuine evidence about category organization, not noise to be explained away.

5.3.2 Mechanisms and basins

The hyperreal model describes the geometry of category boundaries. The HPC framework describes what maintains that geometry. The two are complementary.

Two stories, then. The first is representational: grammatical categories are regions in a feature space, with prototypes at centres and boundaries where similarity to multiple prototypes is balanced. This is the geometry.

The second is causal: the geometry persists because mechanisms – acquisition, entrenchment, alignment, transmission, functional pressure – exert forces that keep items clustered and boundaries stable. Conceptual-space approaches develop the first story in rich detail (Gärdenfors, 2000, 2014); HPC frameworks develop the second (Boyd, 1991, 1999).

Both are needed. The geometry tells you what shape the categories have. The mechanisms tell you why they hold that shape.

Rosch's insight that categories have graded internal structure organized around prototypes is preserved – but repurposed. What's graded is typicality, not membership. An item is in the category or out; that boundary is sharp. But items inside the category vary in how typical they are, how central to the basin, how far from competing boundaries. HPC retains the gradience while relocating it: not degrees of membership but degrees of stability. A typical noun is one that sits deep in the noun basin, far from competition; an atypical noun sits nearer a boundary, more vulnerable to drift, eruptions, erosion. The mechanisms – entrenchment, alignment, transmission – are the forces that maintain this structure.

The hyperreal formalisation adds precision about boundaries: sharp but located at unreachable distances, exactly as tolerance intuitions suggest. Prototype theory describes the shape; HPC explains the stability; hyperreals explain the sharpness.

Like the spinning top (Chapter 4), the stability is dynamic, maintained by active forces rather than static rigidity. The homeostatic mechanisms are exactly these stabilizing forces. They do not define the boundaries – that would be essentialism – but they maintain the landscape geometry.

Acquisition shapes the initial configuration: learners sensitive to input distributions reproduce the basin structure. Entrenchment deepens the basins: high-frequency items anchor the category, steepening the walls around the prototype.

Interactive alignment maintains consistency: speakers coordinate usage, effectively pushing deviants back toward the norm. Consider alignment in action. Suppose you say *that was very fun* – using the degree modifier that signals adjectival status. Your interlocutor might accommodate, adopting *very fun* in their next turn. Or they might resist, persisting with *a lot of fun*. Either way, the micro-choice shifts local expectations: accommodation reinforces the adjectival pattern; resistance maintains the nominal one. Scale this up, and you get a distributed mechanism for maintaining shared norms.

Iterated transmission acts as a filter across generations. Not all variants survive; the ones that do tend to be clear exemplars well inside basins. As Kirby et al. (2008) demonstrated, transmission bottlenecks can select for learnable structures. In their experimental evolution of artificial languages, random mappings became compositional over ten generations because the bottleneck of imperfect learning filtered out the arbitrary variants. This pattern is not inevitable – it depends on particular task pressures and on the right balance of expressivity and compressibility – but it illustrates how structure can emerge without intentional design.

Finally, functional pressure ensures the basins serve communicative needs. Categories exist because they package referents or encode events. Where functional pressure is strong, basins are deep and boundaries are sharp.

All these forces act together to maintain the basin geometry. The hyperreal model describes the shape; the mechanisms explain why it holds.

The hyperreal model tells us what the basin structure looks like: sharp thresholds with scale-dependent tolerance, and instability concentrated in a determinate seam-region that ordinary finite measurement can't pinpoint. The HPC framework tells us why the structure persists: mechanisms operating across multiple timescales, maintaining the clustering without defining it.

5.3.3 Multi-category spaces

Linguistic categories don't exist in isolation. Nouns compete with verbs; adjectives shade into adverbs; prepositions overlap with both. Instead of a set of independent regions, the basin structure is a tessellation of feature space, with categories abutting each other along shared boundaries. Khalidi calls such overlapping classification schemes CROSSCUTTING KINDS and argues they pose no threat to realism: the same entity can be a node in multiple causal networks simultaneously (Khalidi, 2013, pp. 69–72, 92–97). This creates additional structure. An item near the noun-verb boundary might be stably a noun (because it's deep in the noun basin) or unstably categorized (because small changes could push it into the verb basin). The same item might be far from the noun-adjective boundary, so its nominal vs. adjectival status is never in doubt. Category membership is determinate in multiple dimensions simultaneously, with different degrees of stability in different directions.

Consider the data for *fun* – for fun, as it were. In the frequency-of-degree-modification dimension, it has drifted toward adjectives: *very fun* is increasingly acceptable. In the takes-a-determiner dimension, it remains nominal: *the fun we had* is unremarkable. The item sits near the noun-adjective boundary in one dimension, deep in noun territory in another. Its apparent mixed status reflects its location in a multi-dimensional space, not indeterminacy about what categories are.

American historical corpus evidence makes the trajectory visible (Davies, 2010). Degree-modified *fun* – *really fun*, *so fun*, *rather fun*, *very fun* – is rare through most of the 19th and early 20th centuries but rises sharply from the late 20th century onward (Figure 5.3). The pattern is consistent with a shift from marginal adjectival-like uses to robust productivity. *Rather fun* shows early isolated footholds (British epistolary usage appears as early as 1827); the American surge comes later. The diagnostics co-move: this is not one fragile string but a bundle of degree environments converging in the same direction.

Boundary phenomena: operationalizing distance. The *fun* example shows diachronic drift toward a boundary. But how do we study items that sit *at* a boundary synchronically – stably intermediate, not in transition but genuinely in between?

English reciprocals – *each other*, *one another* – offer a detailed case study (Reynolds, 2025a). Standard reference grammars classify them as pronouns, and their distribution supports that much: they head NP-like phrases in the same slots as core pronouns, as verb and preposition complements. But they are non-canonical pronouns. Unlike core personal pronouns, they lack a per-

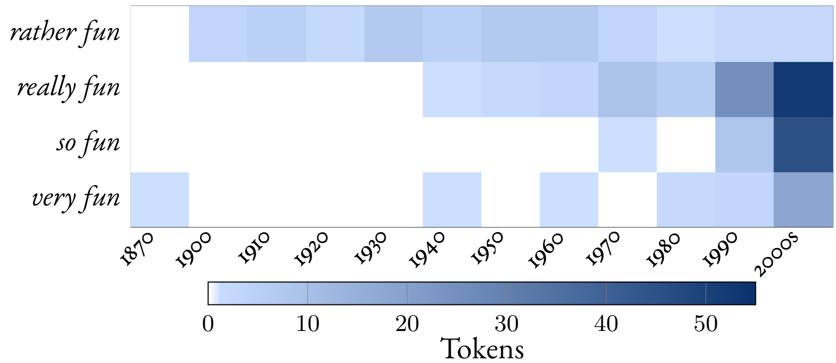


Figure 5.3. Degree-modified *fun* in the Corpus of Historical American English. Darker cells indicate higher frequency. *Rather fun* shows scattered early attestations; *really fun* and *so fun* surge in the 1990s–2000s. The late-century darkening across multiple rows shows convergent adjectivalisation.

son/gender/case paradigm and are distributionally defective in the most salient way: they resist ordinary free subject use (cf. *Somebody left* vs. **Each other left*).

On the other side, reciprocals are built from determinative material (*each*/*one*) and impose a determinative-like semantic constraint: they are obligatorily anaphoric and require plural antecedents. Yet they are not determiners: they can't occur in determiner function (**each other friends* in the intended sense). My aim is not to force a verdict from a handful of tests, but to motivate the boundary question: reciprocals sit where the pronoun and determinative profiles exert opposed pulls, making them an ideal probe for how to operationalize stable in-betweenness.

How stable is this in-betweenness? The methodology in Reynolds (2025a) addresses exactly this question by applying three diagnostics:

1. Invariance under analytic perturbation. Does the boundary position shift when you change the distance metric, the comparison set, or the feature weights? Across multiple correspondence analysis ordinations, Jaccard distances, and specification curves, reciprocals sit midway between pronoun and determinative anchors – a stable intermediate, not an artefact of a particular operationalization.

2. Cross-dimensional tension. Do different feature families pull in different directions? Morphology pulls reciprocals toward determinatives (they lack the case paradigm of core pronouns). Semantics pulls them toward pronouns (they denote referents, not quantities). Syntax and phonology

contribute little. The signature of a boundary item is exactly this: opposed pulls from mechanisms that maintain different basins.

3. Clear anchors behave cleanly. Do items unambiguously inside each basin show the expected clustering? Core pronouns like *he* and *herself* sit deep in the pronoun basin; determinatives like *every* and *either* sit deep in the determinative basin. The methodology confirms what intuition expects: the anchors cluster where they should. The interesting finding is that reciprocals don't.

A caveat about statistical significance. Whether the boundary position looks "statistically significant" depends on which comparison items you choose. Pick one reasonable set and the result is extreme; vary the comparison set and the result wobbles. But the *qualitative* finding – reciprocals sit between pronoun and determinative anchors – is stable across all choices. Under an HPC reading, this is exactly what genuine boundaries should look like: the diagnosis persists, but decisive p-values elude us because the item really is intermediate.

This operationalizes the chapter's theoretical picture. "Distance to boundary" is not metaphorical; it can be measured. "Stability of ambiguity" is not vague intuition; it is invariance under analytic perturbation – the in-betweenness persists when you change your measurement instrument. Reciprocals sit in the overlap region of Figure 5.6, not because the analysis went wrong but because the basin structure itself has overlap.

The same analysis applies to the existential debates about feature systems raised in §3.1. Is GENDER a natural kind, or is it reducible to NUMBER? The question becomes: do gender and number occupy distinct basins maintained by distinct mechanisms, or is gender a sub-region of a larger number basin, maintained by the same mechanisms at a finer grain?

If the mechanisms maintaining gender systems – agreement patterns, acquisition pathways, functional pressures toward nominal classification – are distinct from those maintaining number systems, then gender is a genuine natural kind: a separate basin in the space of grammatical features. If the mechanisms overlap substantially – if gender turns out to be number's machinery for individuation, as some theorists have argued – then the basins merge, and GENDER as a cross-linguistic category dissolves into NUMBER.

This is an empirical question, not a definitional one. The hyperreal model tells us what to look for: distinct basins with sharp thresholds and scale-dependent tolerance. The HPC framework tells us how to investigate: identify the mechanisms, trace their operation, determine whether they cluster distinct properties or the same properties at different scales.

Consider the Shilluk number system introduced in §3.1. English marks number with a suffix: *cat*, *cats*. Shilluk marks it through stem-internal changes in tone, vowel length, and voice quality – a singular/plural pair might differ only in whether the vowel is short, long, or overlong, or in whether the tone is Low or Fall. The exponence is lexicalised: each pair must be learned; no productive rule derives one from the other.

What does the HPC + hyperreal view say about this? Both English and Shilluk number systems occupy basins in a feature space whose dimensions include degree of obligatoriness (how often number must be specified), agreement scope (specifier, verb, other constituents), and semantic function (individuation of entities). The exponence dimension – where English is affixal and Shilluk is fusional/suprasegmental – places them in different regions of morphophonological space but the same region of morphosyntactic and semantic space. They're in overlapping basins, not identical ones – independent systems shaped by similar pressures, not one system shared across languages.

The mechanisms maintaining the two systems are partially shared (acquisition of a grammatical contrast where specified, iterated transmission of the singular/non-singular distinction within the contrastive subsystem) and partially distinct (rule-based productivity in English, lexical storage in Shilluk). The prediction is that where the mechanisms align – the functional pressure to distinguish singular from plural – the categories should pattern similarly across languages. Where they diverge – the phonological substance of the exponence – we should expect variation. That's exactly what we find. Within the contrastive subsystem, the basin boundary for NUMBER is sharp in both languages (a noun is singular or plural, not in between), but the basin's location in the full feature space differs, because different mechanisms weight different dimensions. The same analysis extends to noun-class systems like Swahili's, to Mandarin classifiers, to German gender. Wherever there's grammatical categorisation, there should be basin structure maintained by mechanisms.

A question arises: is the cross-linguistic similarity *convergence* or *homology*? In biology, convergence means similar phenotypes arising independently (the wings of bats and birds), while homology means similarity inherited from a common ancestor (the forelimbs of all mammals). For NUMBER, the answer is probably convergence: the functional pressure to distinguish singular from plural exists in any language with nominal reference, and different languages have evolved different morphological solutions. The mechanisms overlap not because of common ancestry but because of common function. For NOUN and VERB, the answer might be different: if these categories reflect universal constraints on predication and reference, they may be homologous – inherited from whatever cognitive architecture makes human language possible. The framework should distinguish these cases, and the mechanistic analysis provides the tools: shared mechanisms suggest homology; parallel mechanisms suggest convergence.

If the HPC analysis just confirms that English number and Shilluk number are both NUMBER, have we learned anything? Yes: the framework explains *why* they deserve the same label – not definitional fiat, but overlapping mechanisms maintaining overlapping basins. And it makes predictions: if we found a putative "number" category maintained by entirely different mechanisms with no overlap in the functional dimension, the framework would say it's not the same kind – same label, different basin.

5.4 Empirical consequences

5.4.1 Gradient judgments, discrete categories

The discreteness problem has a close cousin: the gradient-judgment problem. Speakers' acceptability judgments are notoriously gradient. Asked to rate sentences on a 1–7 scale, they distribute across the range, with clear acceptability at the extremes and uncertainty in the middle. If categories are discrete, why are judgments gradient?

The hyperreal framework suggests an answer. Distinguish the category from its measurement, the ontology from its epistemology.

Grammaticality – the property of conforming to the grammar – is discrete. A sentence either satisfies the constraints or it doesn't. The boundary between grammatical and ungrammatical is sharp, located at a hyperreal threshold in some underlying dimension (perhaps accumulated constraint violation, or distance from prototype, or processing cost).

Acceptability – the measured response in judgment tasks – reflects grammaticality filtered through noise. Processing difficulty, frequency effects, task demands, individual variation, and measurement error all intervene between the discrete category and the gradient response. This is where the maintenance framework bites: grammaticality is the maintained category, acceptability is a noisy instrument whose bias terms are themselves mechanism-sensitive (frequency shapes entrenchment, task framing activates different basin configurations, exposure history shapes where the boundary lies for each speaker).

One reason this distinction is so easy to lose sight of is that the phenomenology is asymmetric. When a sentence is well-formed, nothing in particular announces itself: processing just runs. The felt data are mostly negative – the "catch" of a violation, the extra effort, the moment of repair. There is, in that sense, no positive feeling of grammaticality, just as there is no feeling of having sufficient oxygen; the experiences that demand attention are the absences. This is why boundary detection is cognitively natural: violations have a signature, conformity is typically silent.

A second reason is methodological. Temperature is not something you see; it is a theoretical magnitude inferred from what thermometers do. Likewise grammaticality is a latent property that we access only through instruments – judgments, reaction times, eye movements, priming, production choices – each of which can misread the target in both directions, yielding false comfort (illusory well-formedness) or false alarm (grammatical but degraded). The point is clearest in *satiation* effects: repeated exposure can raise acceptability for initially degraded strings, even though it is implausible that the grammar itself has been rewritten over the course of a short experimental session. Satiation is therefore not a nuisance; it is direct evidence that the mapping from grammaticality to felt acceptability is plastic, and that we should not identify the construct with any one measurement channel.

Not all of these factors work the same way. Random fluctuations – trial-to-trial variation in attention, incidental processing load, measurement noise – blur the signal without shifting it. But frequency effects and task demands can systematically shift the mapping from grammaticality to acceptability. A borderline construction may be judged more acceptable if it's high-frequency than if it's low-frequency, even when both are equally grammatical. Instructions emphasizing naturalness may locate the boundary differently than instructions emphasizing correctness. These are biases, not noise. The two-layer model accommodates them: different contexts may activate different effective basin configurations, shifting where tolerance breaks down.

This predicts:

- Clear cases cluster at scale extremes (1s and 7s), because items deep in the grammatical or ungrammatical basin are stably categorized and noise doesn't flip them.
- Borderline cases distribute across the middle, because items near the boundary are sensitive to noise – small processing fluctuations can shift judgments in either direction.
- The gradient spread is widest near the boundary, because that's where the signal-to-noise ratio is lowest (see Figures 5.4 and 5.5).

This is a testable prediction. If the two-layer model is right, items independently rated as near category boundaries should show higher variance in acceptability judgments than items rated as central. The prediction isn't that boundary items get middling ratings – that's compatible with gradient grammatical-

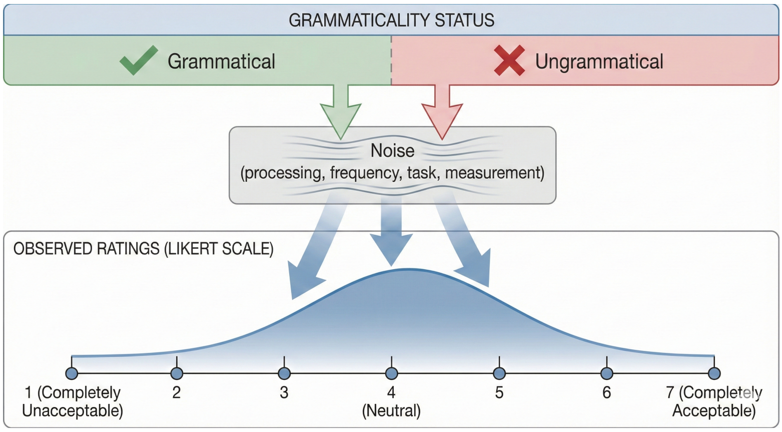


Figure 5.4. The two-layer model. Discrete grammaticality (binary: grammatical or ungrammatical) is filtered through processing and measurement noise to produce gradient acceptability judgments. Items deep in basins produce stable judgments at scale extremes; items near boundaries produce variable judgments distributed across the middle of the scale.

ity – but that their ratings should be more variable across trials and subjects. In the simplest latent-threshold models, variance is maximised in the boundary region and falls off as items move deeper into either basin. The exact functional form depends on the measurement channel (binary choice vs. rating scale), the noise model, and the way context deforms the local landscape; what is robust is the qualitative signature: variance tracks proximity to the seam. Existing work on gradient acceptability already points in this direction: Sprouse et al. (2013) found that judgment variability correlates with distance from prototypical (un)grammaticality, and Dąbrowska (2010) showed systematic individual differences in judgments of complex constructions – exactly what you'd expect if different speakers have slightly different basin structures.

This last point deserves emphasis. The basin structure isn't uniform across speakers. Speakers with different input histories – more or less exposure to written registers, different dialectal backgrounds, varying levels of literacy – will have basins with different shapes and different boundary locations. The hyperreal model accommodates this naturally: each speaker's grammar is a different model, with a different boundary index K. What varies across speakers isn't just noise; it's the underlying geometry. The framework predicts that variability in judgments should be highest for constructions that fall near the boundary *for most speakers* – and that speakers who agree on central cases may nonetheless diverge on peripheral ones, because their boundaries are located at different hyperreal indices.

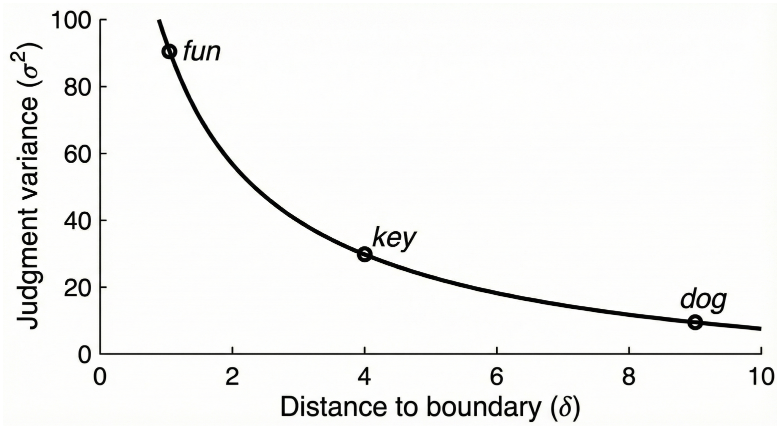


Figure 5.5. Predicted relationship between distance to category boundary and acceptability judgment variance. Items deep in basins (e.g., *dog*) show stable judgments with low variance; items near boundaries (e.g., *near*) show high variability. The qualitative signature – variance peaking near the boundary and falling off into each basin – is robust; the exact functional form depends on the measurement channel and noise model.

This is not a new observation – the distinction between competence and performance has been with us since Chomsky. But the hyperreal framework adds precision. The boundary is not merely "somewhere in the grammar"; it's located at a specific (if unspecifiable) hyperreal threshold. The gradience is not merely "performance noise"; it's the predictable consequence of epistemic limitations near a sharp boundary.

Chapter 14 develops this picture fully, arguing that grammaticality itself is a homeostatic property cluster – a category maintained by mechanisms, with a sharp boundary at hyperreal distance. For now, the point is that gradient judgments and discrete categories are compatible. The discreteness is in the structure; the gradience is in our access to it.

5.4.2 Dual membership

One puzzle remains. If categories are discrete – if at every point in feature space an item is either in or out – how can there be genuine dual membership? How can *near* be both a preposition and an adjective, not contextually selected, not in transition, but stably both? English is comfortable with a kind of dual citizenship here: items like *near* can hold both passports without being in transit.

The answer is that discreteness holds predicate-by-predicate, not across predicates. The predicates *is a preposition* and *is an adjective* are each bivalent: at any point, each is either true or false. But they're not mutually exclusive. The basins can overlap.

This follows naturally from the HPC framework. Categories are maintained by mechanisms, not defined by partition. The mechanisms maintaining prepositionhood – patterns of complementation, head-of-PP status, lack of degree modification – are partially independent of those maintaining adjectivehood – gradability, predicative use, comparative morphology. An item can fall within the tolerance threshold for both, satisfying the clustering criteria for each.

Where the mechanisms align, the basins are disjoint: nouns and verbs occupy separate regions because the mechanisms that maintain them pull in different directions (Figure 5.6, left). Where the mechanisms cross, the basins overlap: adjectives and prepositions share enough properties – predicative function, modification of nominals – that some items cluster with both (Figure 5.6, right).

The regions of overlap are not arbitrary. They're located where the property clusters themselves overlap – where an item can satisfy the tolerance criteria for both categories simultaneously. These are exactly the cases that trouble essentialism: items that have "mixed" status because they satisfy criteria for multiple categories. On the HPC + hyperreal view, mixed status is not indeterminacy but dual location: the item is in two basins at once, stably, because the mechanisms that maintain each basin both apply.

The reciprocals case from §5.3.3 exemplifies this cleanly. Morphology maintains the determinative basin; semantics maintains the pronoun basin. Reciprocals satisfy both. The mixture weights from the blend model – 0.534 pronoun-like for *each other*, 0.487 for *one another* – are not noise or measurement error. They are the empirical signature of genuine dual location: stable membership in overlapping basins, maintained by partially independent mechanisms.

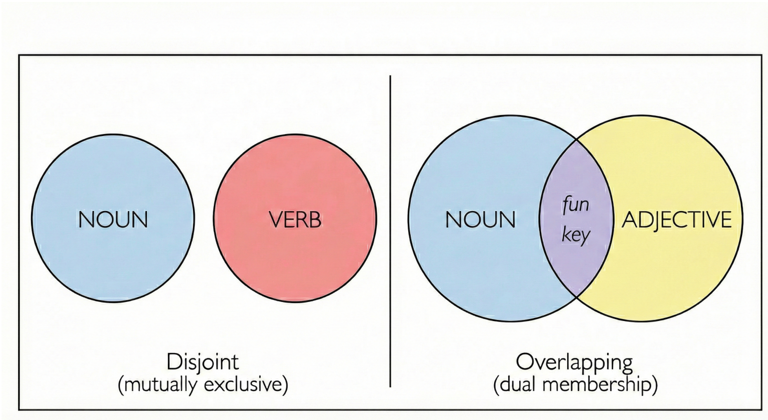


Figure 5.6. Disjoint vs. overlapping category basins. Left: noun–verb mechanisms pull in opposite directions, creating mutual exclusivity. Right: noun–adjective mechanisms are partially independent, permitting overlap. Items like *fun* and *near* occupy the overlap region, satisfying the clustering criteria for both categories simultaneously.

5.4.3 Summary: discreteness without essence

The discreteness problem asked: if underlying properties are continuous, how do discrete categories emerge?

The answer combines two components:

Structure (hyperreal model): Discrete boundaries arise from scale-dependent tolerance. Changes that are negligible at the current scale preserve categorization; changes that are appreciable can flip it. The boundary is sharp, located at a hyperreal threshold, epistemically inaccessible but structurally determinate. Dynamic discreteness is like a snow-covered streetcar right-of-way: the snowfall is continuous, but the rails impose two stubbornly sharp tracks that exist only because of the system underneath them. The category isn't holding still. It's being *held* – by mechanisms operating on multiple timescales, by basins that resist perturbation, by forces that push deviants back toward the norm.

Maintenance (HPC framework): The basin structure persists because mechanisms – acquisition, entrenchment, alignment, transmission, functional pressure – hold properties together. Without these mechanisms, categories would dissolve. With them, the clustering is stable even though boundaries are fuzzy at the edges.

Together, these explain how categories can be:

- Real: maintained by causal mechanisms, not merely stipulated.
- Discrete: with sharp boundaries, not gradient membership.

- Fuzzy at the edges: because tolerance fails near boundaries, producing instability and apparent gradience.
- Stable: because mechanisms maintain the basin structure across time.
- Capable of change: because mechanisms can shift, basins can migrate, boundaries can move.

This is what the essentialist wanted – real structure, sharp boundaries, determinacy – without what the essentialist thought was required: definitions, necessary and sufficient conditions, essences. The categories are real because they're maintained, not because they're defined. The boundaries are sharp because tolerance is scale-dependent, not because some property is binary. The determinacy is causal-structural, not definitional.

The nominalist was right that essences don't exist. The nominalist was wrong to conclude that determinacy fails. The HPC framework, formalised via hyperreal tolerance, shows how to have both: natural kinds without essences, discrete categories from continuous substrates, sharp boundaries that we can't quite see. Dynamic discreteness and real gradience – the chapter's twin themes – turn out to be two sides of one insight.

Definiteness thread. This chapter predicts a signature phenomenon for definiteness: judgments about borderline cases (e.g., weak definites like *go to the hospital*) should show higher variance than judgments about clear cases, because they sit near the boundary where tolerance fails. Register shift (task context, formality) should modulate variance without eliminating the underlying discreteness. The prediction would be falsified if borderline definiteness showed no variance asymmetry, suggesting the category lacks the basin structure this chapter describes.

Projectibility and the good bet

> If there are not [pedicabs], Mr Tagomi thought, I would be well advised to retire to secluded place and kill myself.
>
> —Philip K. Dick, *The Man in the High Castle* (1962)

Mr. Tagomi has slipped between worlds, and he needs to know which one he's in. Pedicabs exist in his home world but not in the other. He's betting his life on the projectibility of a single feature.

In your world, the stakes are lower but the logic is the same. You encounter *car*. It takes determinatives, pluralizes, functions as an argument. You've never seen *pedicab* – but you bet it behaves likewise. You're usually right. The category NOUN lets you project from instances you've seen to instances you haven't. That's the bet, and it usually pays off.

Polish aspect offers a puzzle. Every textbook tells you: perfective aspect marks bounded events – completed, delimited, whole. Imperfective marks unbounded events – ongoing, habitual, incomplete. The definitions are centuries old, drilled into every language learner, refined in journal articles. They should work.

When computational models trained on those definitions try to predict which aspect Polish speakers actually use, they project imperfectly – 77% accuracy for perfective, against 98% for imperfective (Divjak et al., 2025). The definitions that have been taught for generations capture only half the system. Something is projectible about aspect; it just isn't what the textbooks say.

This chapter asks what makes a category projectible – worth betting on, and whether the same category can be projectible for one purpose but not another.

The previous chapters argued that many linguistic categories are real, maintained by mechanisms rather than defined by essences. Grammatical categories are the central testbed in this book, but the same question arises for phoneme contrasts and for social categories like academic register and politeness systems. This chapter asks what we get in return. The answer is projectibility: the ability of a category to support inductive inference. Categories earn their keep by supporting induction. But which categories, for which purposes, and why?

The next two chapters develop two diagnostics for mechanism-maintained kinds. This chapter addresses PROJECTIBILITY: does the category support induction? Can you learn about one member and reliably bet on others? Chapter 7 addresses PERTURBATION SENSITIVITY: what maintains the cluster, and what happens when the maintenance fails? The two diagnostics are complementary. Projectibility tests whether the category earns its keep; perturbation sensitivity reveals the causal structure that makes it do so. Chapter 8 then shows how these diagnostics separate genuine HPCs from thin, fat, and negative classes – labels without the causal structure to back them up.

For constructed categories – mathematical sets, legal statuses, programming constructs – the essentialist answer works: definitions support induction (§2.2). Most linguistic categories aren't constructed, so definitions aren't there to do the work. What, then, grounds the bet? The maintenance view answer: mechanisms. Categories support induction not because they're defined but because they're mechanistically grounded. The mechanisms that maintain the category also explain why its members resemble each other. Learning about one member tells you about others because the same mechanisms shaped both.

Before looking at a complex success, consider a failure (Figure 6.1). The label NPI (negative polarity item) groups expressions like *ever*, *yet*, and *lift a finger* that require licensing in negative contexts – but Hoeksema (2012, p. 30) documents twelve distinct licensing patterns across English NPIs. Models succeed at NPI prediction only to the extent they learn each item's distribution, not the category. The label names a distributional class; no mechanism maintains it as a kind.

This chapter develops an operational criterion for distinguishing distributional classes from homeostatic kinds – and applies it to Polish aspect.

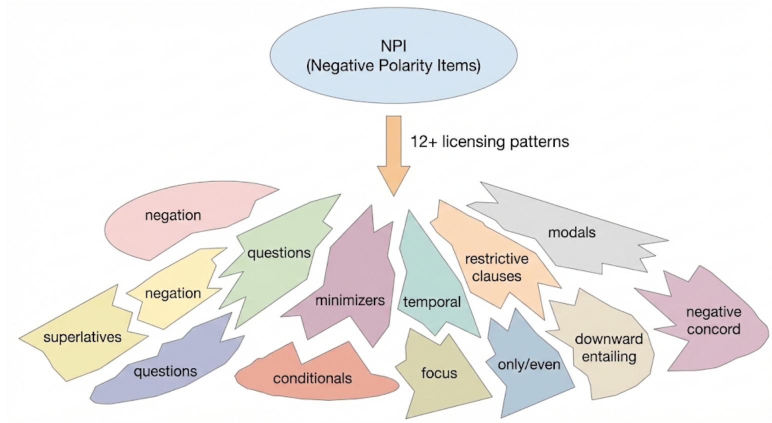


Figure 6.1. Uniform label, jagged reality. The category NPI (negative polarity item) groups expressions that share sensitivity to negative contexts, but Hoeksema (2012, p. 30) documents twelve distinct licensing patterns. No single mechanism maintains them as a kind; the label names a distributional class, not an HPC cluster.

6.0.1 Unpacking the framework

Chapter 4 compressed the maintenance view into a slogan: *profile, stabilized by mechanisms, projectible relative to purposes*. This chapter unpacks each component and tests them against Polish aspect.

The three parts track three levels of analysis:

- Profile – the observable surface. Properties cluster; members resemble each other. This is what you observe.
- Stabilized by mechanisms – the metaphysics. Something keeps the cluster together: acquisition, entrenchment, functional pressure, social enforcement. Without mechanisms, clustering is coincidental.
- Projectible relative to purposes – the epistemology and pragmatics. The category supports induction, but only for questions the mechanisms underwrite. A syntactician and a semanticist may have different right-sized categories for overlapping extensions.

The framework guards against three confusions. First, treating categories as mere conventions (ignoring the mechanisms). Second, treating projectibility as absolute rather than field-relative (ignoring purpose). Third, treating mechanism and projectibility as the same thing (conflating metaphysics with epistemology). The remainder of the chapter applies each piece to a test case.

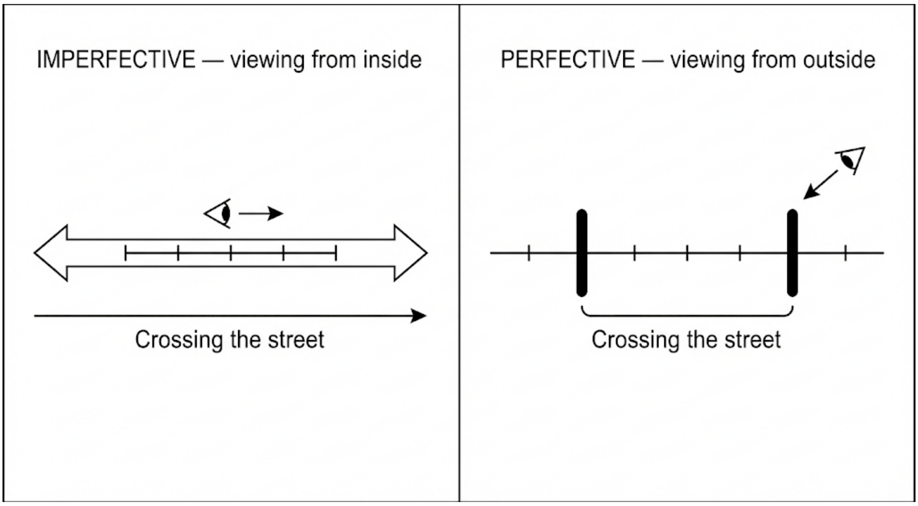


Figure 6.2. The textbook account of Slavic aspect. Imperfective presents the event from inside, as ongoing; perfective presents it from outside, as a completed whole. This is the "viewpoint" metaphor that has dominated aspectual theory for a century.

6.1 The definitional bet

Slavic verbal aspect looks like a paradigm case of a grammaticalized binary. Every Polish verb bears aspectual marking: imperfective or perfective. The opposition is obligatory, pervasive, and ancient – codified in grammars for centuries, taught in every language classroom.

The textbooks offer crisp definitions. Imperfective presents an event as unbounded – ongoing, habitual, incomplete. Perfective presents it as bounded – a completed whole with temporal limits (Figure 6.2). The terminology varies (totality, resultativeness, telicity), but the core claim is the same: there's an invariant semantic content distinguishing the two aspects, and knowing that content should let you predict which aspect a speaker will choose.

This is essentialism applied to morphology. The definitions are the essence; the essence determines behaviour. Learn the essence, and you can project: knowing what perfective *means* should tell you when to use it.

But the prediction fails. If aspect has invariant semantic content, then a model that knows that content should predict aspectual usage. Divjak et al. (2025) tested exactly this.

They built three computational models of Polish verb usage:

- A *lemma-concrete* model that learns cue–outcome associations without assuming aspect exists as a category at all.

- An *aspect-concrete* model that learns aspect from usage cues – distributional patterns in the input.
- An *aspect-abstract* model that learns aspect from the semantic labels proposed in the theoretical literature – boundedness, totality, resultativeness.

The aspect-abstract model is the textbook view formalised. If the invariant meanings are real, this model should win.

It doesn't.

On corpus data, the aspect-abstract model performs reasonably – 87% accuracy (Divjak et al., 2025). But accuracy hides an asymmetry. The model predicts imperfective well: 98% correct. It predicts perfective poorly: only 77% – you lose the bet nearly one time in four. The semantic invariants that aspectologists have proposed for a century – the definitions on which projectibility supposedly depends – capture only half the system. Why this direction? Likely because perfective is more lexically heterogeneous: perfective verbs are marked by a variety of prefixes, each with its own distributional profile, and they appear in a narrower range of tense frames. Imperfective is closer to a default, less marked, more uniform. The asymmetry isn't random; it tracks where the semantic definition aligns with distributional reality (imperfective) and where it doesn't (perfective).

More damaging: when validated against native-speaker judgments in a gap-filling task, the aspect-abstract model performs *worse* than the simpler models. Native speakers' preferences align better with a model that doesn't even use the aspect category than with one that does. The lemma-concrete model – the one that treats each verb individually, without abstracting to aspect – best predicts which forms humans prefer.

This is a failure of definitional projectibility. The category exists; the usage is systematic; but the definitions don't project to behaviour. Knowing what "perfective" supposedly *means* doesn't reliably tell you when Polish speakers will *use* it. (A moderate reader might object that the textbook definitions are simplistic models, not metaphysical essences – that their failure is a failure of *that* model, not of semantic-level explanation generally. Fair enough. But even treating definitions as models rather than essences, the projective power lies in distributional cue structure and lexeme-level entrenchment. Semantic paraphrase plays at best a partial, uneven role.)

6.2 The mechanistic alternative

Why does the textbook view fail? Divjak et al. (2024) provide the answer.

Their corpus study of Polish aspect reveals a usage landscape strikingly different from the textbook picture.

First, lexical bias is the norm. About 90% of Polish verbs strongly prefer one aspect – greater than 90% of their tokens appear in the preferred form. Only 11% of aspectual pairs are genuinely equiprobable. The textbook suggestion that speakers "choose" aspect based on how they view the event applies to a small minority of cases.

Second, tense carries most of the signal. A simple model using just the superlemma (verb identity) and three-way tense distinction (past, present, future) achieves F1 scores of 0.95 for imperfective and 0.90 for perfective – the perfective asymmetry persists, just smaller. Aspect is largely predictable from tense, not from semantic viewpoint.

Third, context cues appear where needed. Temporal adverbs and other contextual markers show up reliably only with the 11% of verbs that lack lexical bias. The system is informationally efficient: redundant cues don't clutter unambiguous cases.

Finally, the landscape is already shaped – like a riverbed cut by the water that flows through it. For most verbs, learning the verb means entering a channel that use keeps cutting deeper. The category isn't applied at utterance; the same flow that carved it maintains it.

This reframes the question worth asking. Instead of "what does aspect mean?" – a question that invites ever more refined invariant definitions – we ask "what maintains the aspectual patterns speakers produce?" That question is open, investigable, and already yielding answers: distributional signatures, acquisition dynamics, lexical entrenchment. The textbook view isn't wrong; it's asking a question whose answer can't be found by the methods it uses. The mechanism view doesn't replace it – it explains why it worked as well as it did, and what it was missing.

Now the maintenance view snaps into focus. Why can a Polish speaker project aspectual behaviour to unfamiliar verbs? Not by applying a definition. The definitions – boundedness, totality – predict only imperfective, and even then imperfectly. What projects is something different: the cue–outcome structure that learners extract from the input.

Consider: you've learned ***robić*** (impf) / ***zrobić*** (pf) – 'do', 'accomplish'. You encounter an unfamiliar verb ***pływać*** – 'swim'. How do you know that ***przepływać*** will be perfective?

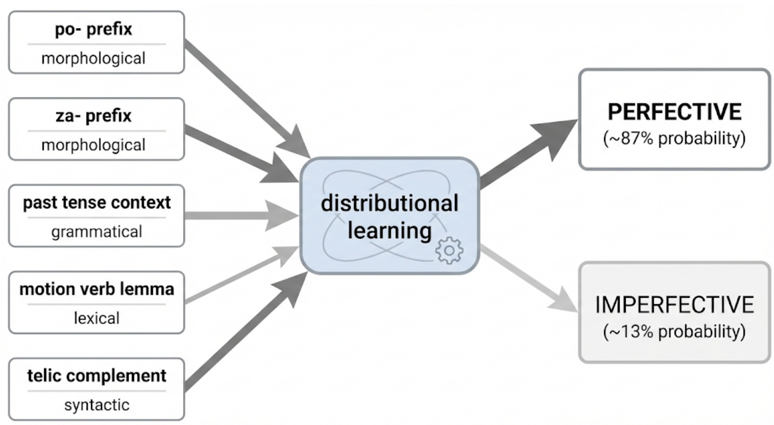


Figure 6.3. The mechanism-based view of aspectual projectibility. Instead of applying semantic definitions, learners extract cue–outcome associations – morphological patterns, tense contexts, lexical biases – that allow prediction of novel forms. The question this answers is not "what does aspect mean?" but "what maintains aspectual patterns?" (Percentages are schematic; actual weights vary by verb and context.)

The textbook answer: you apply the semantic definition of perfectivity – the event is bounded, complete, telic – and infer that *przepływać* ('swim across', 'complete the crossing') satisfies those criteria.

But as we've seen, the semantic definition predicts poorly. Worse, it predicts perfective worse than imperfective – exactly wrong for the case at hand.

The mechanism-based answer: You've learned that certain prefixes, in certain tense contexts, pattern with perfective. You've learned that verbs of motion typically have strong aspectual biases carried by morphological material. You project not from a definition but from a web of cue–outcome associations maintained across your linguistic experience.

This is projectibility grounded in mechanism rather than definition (Figure 6.3). The category PERFECTIVE exists – you can introspect about it, metalinguistic discourse depends on it – but its psycholinguistic reality needn't rest on invariant semantic content. The good bet is on the consistency of the distributional patterns that maintain it.

6.3 Labels aren't mechanisms

Here's where the story takes a surprising turn. Divjak et al. (2025) found that the best model – the one that most closely matches native-speaker preferences – is the one that doesn't use the category ASPECT at all.

The lemma-concrete model – the one that treats each verb individually, without abstracting to aspect – predicts human behaviour better than the aspect-aware models. The statistical evidence is overwhelming: the probability that the aspect-aware models are actually better is effectively zero.[1]

What does this mean for the ontological status of aspect? Here we need to separate three things:

1. *Is there a stable pattern that linguists call "aspect"?* Yes. The corpus regularities, speaker agreement, and reliable acquisition all point to a real, socially shared, historically stable clustering.

2. *Do the traditional semantic definitions project to usage?* Not reliably – especially for perfective. That's a failure of *definitional* projectibility, not automatically a failure of aspect as such.

3. *Is the abstract binary label psychologically intermediate in the production system?* The lemma-concrete result suggests: maybe not. The best causal story may live at a finer grain.

The apparent contradiction – "aspect looks real" versus "aspect doesn't project" – dissolves once we see what exactly failed. What failed is the textbook essence. What may also have failed is the assumption that speakers need an explicit aspect representation as a mediating variable. What did *not* fail is the existence of stable, learnable, intersubjectively shared aspectual patterning.

This is almost the cleanest possible maintenance-view demonstration. The category label tracks a real cluster. But that cluster is maintained by a bundle of finer-grained mechanisms: lexeme-specific entrenchment, tense-conditioned expectations, prefixal cueing. Projectibility belongs to that bundle, not to the invariant semantic paraphrases. Polish aspect is real as a stable pattern, but the projectible unit is smaller than the label. The mechanism projects; the label names the emergent alignment of those mechanisms without explaining it.

Put differently: ASPECT names the pattern; it doesn't do the work. You can project aspectual behaviour *without* representing "aspect" as such – because what you're really projecting are cue–outcome associations that happen to have an aspectual signature. The lemma-specific patterns don't preclude aspect as a generalisation; they ground it. What speakers know isn't *either* verbs *or* aspect but a multi-level network where both coexist – the general emergent from the particular. This answers the objection that a lemma-level model leaves no room for the category: the category exists as an emergent alignment, even if processing

[1] AIC 3,891 for lemma-concrete versus 3,954 for aspect-concrete and 4,037 for aspect-abstract; evidence ratios in the 10^{13} to 10^{17} range – numbers so large that the conclusion is statistically compelling.

happens at a finer grain. The island is real even though the coral does the maintaining.

This resolves the puzzle of naming. We call it ASPECT because of its history; Polish aspect is the genealogical descendant of a system that may once have fit the definitions better. But mechanistic drift happens. A category can remain historically continuous while its causal underpinnings shift. A label can remain genealogically appropriate but be synchronically misleading about what projects. Think of a volcanic island. The volcano builds the island over millennia, but eventually goes dormant. Coral colonises the flanks; soil accumulates; vegetation takes root. The island persists – not because the volcano is still active, but because new mechanisms have taken over: reef-building, soil formation, root networks. Calling it a "volcanic island" is genealogically correct; the volcano did the initial work. But it's synchronically misleading: what maintains the island now is coral and roots, not magma. Polish aspect is like this. The opposition was carved by semantic-temporal distinctions that may once have been the primary mechanism. But the volcano has gone dormant. Now what maintains the pattern – what makes it projectible – is lexeme-specific cue structure, not semantic essence. The label names the island, not the volcano. (Note the developmental asymmetry: for a child acquiring Polish, the volcano was never active. Learners don't experience the historical shift; they experience only the coral reef – the current distributional cues. The "mechanism drift" is a diachronic observation about the collective system, not a stage in individual acquisition.)

The cross-linguistic term ASPECT functions as a COMPARATIVE CONCEPT; within Polish, it names a historically inherited pattern whose contemporary projectibility is anchored in lexeme-specific cue structures rather than the textbook semantic essence. Labels aren't definitions, and labels aren't mechanisms. Calling this opposition ASPECT risks reifying a historically entrenched label whose textbook semantics no longer track the mechanisms that actually maintain usage. It is best treated as the name of an emergent clustering, not the explanatory unit that makes it project. But here's a constraint: genealogical continuity alone doesn't justify continued use of a label. The label is appropriate only if it still picks out a stable emergent pattern at some explanatory level. An island that has fully eroded, with no coral keeping it together, is no longer an island – and deserves no name.

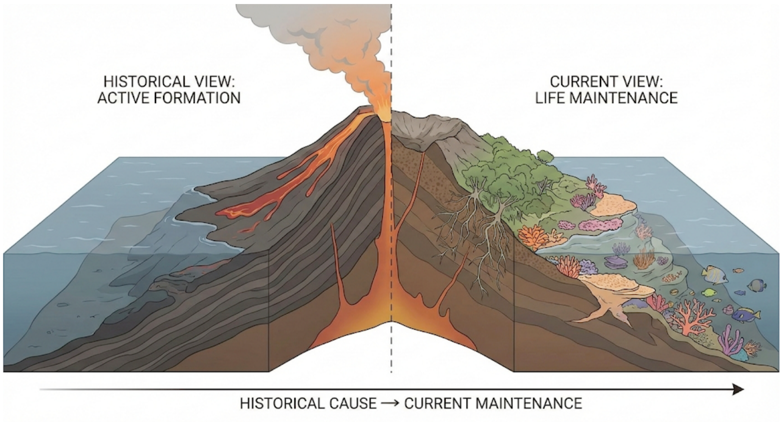


Figure 6.4. Mechanistic drift: a volcanic island. The volcano built the island but is now dormant. What maintains the island today – coral reefs, root networks, soil formation – differs from what created it. Polish aspect is analogous: the semantic-temporal distinctions that may once have been the primary mechanism have gone dormant. What maintains aspectual patterns now is lexeme-specific cue structure. The label names the island, not the volcano.

6.4 Generalising the lesson

Polish aspect illustrates a general lesson of the maintenance view. Polish has a grammatical opposition traditionally called aspect; what these results challenge is the assumption that the textbook characterization of that opposition is the mechanism that makes it projectible. Aspect has been defined for centuries, but those definitions fail to predict usage. By contrast, mechanism-based accounts succeed where definitions fail: the same system becomes predictable once you attend to the distributional structure – lexical bias, tense context, morphological cues – rather than the semantic essence.

This separation reveals that the category and the mechanism need not coincide. Aspect is a useful label picking out a real pattern, but the psychological reality of that pattern may be a network of cue–outcome associations, not a represented binary. The mechanism lives below the category; the projectibility comes from the mechanism. When a proposed characterization fails to project, that's diagnostic: it tells you the characterization doesn't track the mechanisms. This is HPC without category reification: clusters maintained by mechanisms, labels applied by analysts.

Polish aspect is a worked example, not an isolated curiosity. The pattern – definitional characterizations that fail to project, mechanism-based descriptions that succeed – appears across grammatical domains.

Ambridge et al. (2020) found that the causative alternation (*break the vase* vs **laugh the man*) is predicted by a continuous semantic dimension (directness of causation) across five typologically unrelated languages. The semantic definitions of verb classes don't work; the mechanism (sensitivity to causal directness) does. Strikingly, a model trained on five languages predicts native-speaker judgments in a sixth (Balinese) without exposure. The mechanism projects cross-linguistically; the definitional verb classes don't.

Saldaña et al. (2022) showed that morphological syncretism patterns are learned better when syncretic cells share semantic similarity – not when they share features in a definitional sense. The learnability gradient tracks similarity, not category membership. What projects is the similarity structure, not the feature system.

Chi et al. (2020) demonstrated that grammatical relations (subject, object) emerge spontaneously in multilingual BERT – trained without any symbolic grammar – and transfer across languages. A probe trained only on English successfully identifies grammatical relations in French. The categories exist, but defining grammatical roles isn't what produces them; distributed learning over functional pressures is. The mechanism projects cross-linguistically; the formal definitions are epiphenomenal.

These aren't the only examples. Chapter 11 develops the case for word classes in detail. The upshot: projectibility is the empirical test for mechanistic grounding. When a category supports reliable induction, that's evidence that the mechanisms are real. When it doesn't, something about the proposed characterization is off.

6.5 Field-relative projectibility

We've established that projectibility tracks mechanism, not definition. But projectibility for whom? The same category can project well for one analytical purpose and poorly for another. This isn't perspectivalism – different views of the same thing. It's a claim about ontology: there are genuinely different HPCs that happen to overlap in extension, each projectible in its own domain.

6.5.1 The tomato problem

Is a tomato a fruit or a vegetable? The question is undecidable only if you think there's one right answer.

A botanist classifies tomatoes as fruits – they develop from the ovary of a flowering plant and contain seeds. This classification projects: knowing that tomatoes are fruits lets the botanist predict ripening patterns, seed dispersal mechanisms, and responses to plant hormones. The category *fruit* supports reliable induction for botanical purposes.

A chef classifies tomatoes as vegetables – savoury, used in main courses, paired with salt rather than sugar. This classification also projects: knowing that tomatoes are culinary vegetables lets the chef predict flavour pairings, cooking methods, and menu placement. The category *vegetable* supports reliable induction for culinary purposes.

Neither classification is wrong. They're not even in conflict. They're *different HPCs* that happen to have overlapping extensions. The botanist's *fruit* and the chef's *vegetable* are each maintained by mechanisms appropriate to their domain – reproductive biology for one, flavour chemistry and culinary tradition for the other. Each projects reliably within its field.

The tomato problem is a problem only if you assume projectibility is absolute – that a category either projects or doesn't, full stop. Once you recognize that projectibility is field-relative, the problem dissolves. What remains is the work of specifying which category projects for which purposes.

Table 6.1. Two categories, overlapping extension

Category	Field	What projects
Proper name	Semantics	rigid reference, opacity, individual-tracking
Proper noun	Syntax	heads NP, fills argument slot, triggers agreement

6.5.2 Proper nouns and proper names

Linguistics has its own tomatoes. Consider the distinction between PROPER NOUN and PROPER NAME.

A semanticist works with PROPER NAMES: expressions that refer directly to individuals without descriptive content, that are rigid designators (picking out the same individual across possible worlds), and that create referential opacity ("Lois believes Superman can fly" doesn't entail "Lois believes Clark can fly"). These properties cluster because of the cognitive and communicative functions names serve: tracking individuals across contexts requires stable reference without shifting descriptive content.

A syntactician works with PROPER NOUNS: words that head nominal projections, resist articles in certain languages, trigger particular agreement patterns, and fill argument slots. These properties cluster because of distributional pressures: words that behave similarly in one syntactic environment tend to behave similarly in others.

The extensions overlap substantially – most expressions that are proper names are also proper nouns. But the overlapping extension doesn't mean they're the same category. They're different HPCs, maintained by different mechanisms, projectible for different purposes.

A syntactician *knows* that *Brett* is a proper name. But that semantic fact doesn't project for syntactic purposes. What projects syntactically is the distributional chain: *Brett* heads a nominal phrase → fills a subject slot → triggers third-person singular agreement. The syntactician can reliably predict that novel proper nouns will behave distributionally like familiar ones. The proper-name status is real but orthogonal to that prediction.

This isn't esoteric. Every introductory syntax course teaches students to distinguish "what a word means" from "how it behaves". What hasn't been articulated is why this distinction exists: because meaning and behaviour are tracked by different mechanisms, and different mechanisms produce different HPCs. The field-relativity of projectibility explains why syntacticians can't just read off predictions from semantics, and vice versa.

6.5.3 Why colour doesn't grammaticalise

The proper-name case stays within linguistics – two subfields carving the same extension differently. A starker example crosses domains entirely: colour.

Colour is maximally salient. Trichromatic vision evolved because colour predicts things that matter for survival: fruit ripeness, toxicity signals, mate health, camouflage. The warm/cool boundary shows categorical perception – adults distinguish warm from cool colours faster than they distinguish within either category (Holmes & Regier, 2017). Colour terms partition efficiently across languages, with the warm/cool split emerging early in colour-vocabulary evolution (Kay & McDaniel, 1978). If any perceptual domain should grammaticalise, colour should.

It doesn't. Noun-classification systems regularly encode animacy, shape, and size – but never colour. No documented language has verb agreement triggered by whether the subject is warm- or cool-coloured. The warm/cool distinction is robust, salient, and cross-linguistically stable in the lexicon. Yet it's entirely absent from closed-class morphology (Seifart, 2010).

Prasertsom et al. (2026) tested whether this gap reflects domain-specific constraints on grammar – a hard-wired filter excluding colour from the set of possible grammatical features – or something more general. Their experiments compared animacy-based and colour-based noun-class learning in an artificial language. Participants learned both, but animacy-based classes were learned better. More strikingly, when the input was ambiguous between animacy and colour as the classification basis, participants overwhelmingly inferred animacy – 77.5% generalised to animacy, nearly half of them categorically.

This bias isn't grammar-specific. In a non-linguistic sorting task, the same preference appeared: participants sorted images by animacy rather than colour, even when the experimenters manipulated within-category similarity to favour colour. The bias persisted across three stimulus sets designed to progressively disadvantage animacy.

What drives the asymmetry? Not salience – colour is as perceptually salient as animacy. The answer lies in predictive structure. Knowing something is animate predicts a cluster of other properties: self-initiated motion, goal-directedness, organic composition, susceptibility to certain event roles. These predictions hold regardless of what specific entity you're considering. Knowing something is red predicts little beyond its surface appearance – and what it does predict (ripeness, toxicity) is object-relative. Red means different things for apples, frogs, and traffic lights.

Prasertsom et al. (2026) confirmed this computationally. They extracted word embeddings for 472 frequent physical nouns from child-directed speech and clustered them by animacy versus colour. Animacy-based clusters were more compact and more distinct; logistic classifiers learned animacy categories faster and generalised better to unseen nouns – without any built-in knowledge of animacy. The distributional structure of language itself encodes the asymmetry.

This is purpose-relative projectibility across domains. Colour is projectible *for ecological purposes*: predicting edibility, danger, health. The mechanisms stabilizing colour categories are perceptual and action-guiding – they evolved to support decisions about what to eat, avoid, or approach. But grammar serves different purposes: argument realisation, reference tracking, event-structure encoding. Animacy predicts properties relevant to those purposes; colour doesn't.

The upshot isn't that colour categories are fake or that grammatical categories are special. Both are HPCs, maintained by mechanisms appropriate to their domains. What differs is the domain of projectibility. Evolution gave us colour vision because colour projects for survival-relevant inferences. It didn't give us colour-based agreement because colour doesn't project for grammatical inferences. The same extension – red things, warm-coloured things – supports robust categorisation in one domain and fails to support it in another.

This extends the proper-name pattern. Proper name projects for semantic purposes; proper noun projects for syntactic purposes. Warm-coloured projects for ecological purposes; nothing projects it for grammatical purposes – so nothing grammaticalises. The framework doesn't multiply categories recklessly; it asks what cluster of properties is maintained by what mechanisms, and whether that cluster projects for the questions you're asking. Chapter 11 returns to animacy in its analysis of noun/verb stability.

6.5.4 Preview: Part III as demonstration

This pattern – overlapping extensions, distinct HPCs, field-relative projectibility – structures the case studies in Part III.

Chapter 9 shows that COUNTABILITY decomposes into semantic individuation (boundedness, discrete enumeration) and morphosyntactic count-marking (plural inflection, quantifier selection). *Furniture* individuates semantically – you can count chairs – but patterns morphosyntactically as mass. The semantic and morphosyntactic clusters are maintained by different mechanisms and project for different analytical purposes.

Chapter 10 shows that definiteness and deitality are distinct HPCs with overlapping extensions. DEFINITENESS – identifiability, uniqueness, familiarity – projects for semantic purposes. DEITALITY – the morphosyntactic properties that pattern with definite articles – projects for syntactic purposes. The literature's confusion about "weak definites" and "generic definites" arises from treating these as one category when they're two.

Chapter 11 shows that the noun/verb contrast is crosslinguistically stable because both semantic and morphosyntactic mechanisms reinforce it, while adjective categories vary because the mechanisms don't align as tightly.

Chapter 12 traces the maintenance spectrum from semantically transparent to purely entrenched. English gender is designatum-driven – pronouns track perceived personhood. French gender is entrenchment-driven – antecedent form controls the pronoun with no semantic grounding. The same category label names different HPCs with different mechanisms.

Chapter 13 then zooms out, showing how coupling tightness varies across the linguistic hierarchy – from the hard-coupled acoustic-distributional bundle of phonemes to the loose-coupled semantic-morphosyntactic profiles of grammatical categories.

The framework doesn't multiply categories recklessly. It asks, for each analytical purpose: what cluster of properties is maintained by what mechanisms, and does that cluster project for the questions you're asking? Sometimes the answer is a familiar category; sometimes it's a refinement of one; sometimes it's a recognition that what looked like one category was two.

6.5.5 The discipline: three checks

The slogan – *profile, stabilized by mechanisms, projectible relative to purposes* – maps onto three diagnostic checks:

1. Cluster check. Does property covariance hold across samples and contexts? If the properties don't cluster, there's no profile to explain.

2. Homeostasis check. Can we identify perturbations where the cluster reconstitutes (or fails)? If we can't, we haven't identified the mechanism.

3. Projectibility check. Are there counterfactual-supporting inferences in new contexts? If learning about one member doesn't tell you about others, the category isn't earning its keep.

Crucially, these checks can yield different answers for different fields. A category might pass all three for syntax and fail the projectibility check for semantics. That's not a defect of the framework – it's the framework working. The checks aren't absolute; they're indexed to purpose.

This discipline guards against two temptations. First, the essentialist temptation: treating one field's category as the "real" one and others as derivative. If proper names are semantically fundamental, proper nouns must be "just" the syntactic reflection of semantic reality. But the mechanisms are different; neither is derivative. Second, the nominalist temptation: treating categories as arbitrary conventions because different fields carve differently. But the carving isn't arbitrary; each field's categories are maintained by mechanisms appropriate to that field's explanatory goals.

6.6 The epistemic payoff

The aspect case shows that even a descriptively real category can fail to project if its definition misaligns with its maintaining mechanisms. This misalignment clarifies a classic problem in the philosophy of science.

Liu Cixin's turkey parable captures the problem of induction with dark precision:

> Every morning on a turkey farm, the farmer comes to feed the turkeys. A scientist turkey, having observed this pattern to hold without change for almost a year, makes the following discovery: "Every morning at eleven, food arrives." On the morning of Thanksgiving, the scientist announces this law to the other turkeys. (Liu, 2014, ch. 6)

The turkey's law is perfectly confirmed by all available evidence. The problem isn't the evidence; it's that the law isn't grounded in mechanism. The farmer's purpose – invisible to the turkey – determines when the correlation breaks. Had the turkey understood *why* food arrives (fattening for slaughter), it would have predicted its own demise rather than its next meal.

Nelson Goodman's *grue* problem makes the same point in philosophical dress. Goodman's riddle: emeralds examined before time t are green; emeralds examined after t are blue. Define *grue* as "green if examined before t, blue otherwise". Every emerald we've ever observed is grue. So why don't we project *grue* to unexamined emeralds?

Goodman's own answer invoked entrenchment: *green* is projectible because it's been projected successfully in the past. *grue* isn't because it hasn't. The circularity is deliberate – projectibility is bootstrapped from track record.

This observation shaped a tradition. Quine (1969) argued that the success of induction presupposes natural kinds: we project properties because we assume the instances we've observed are "of a kind" with those we haven't. Kornblith (1993) went further, arguing that induction works precisely because natural kinds exist in the world – the clustering of properties in those kinds underwrites our ability to learn from experience. On this view, projectibility isn't just a pragmatic feature of our language; it's evidence that we've latched onto something real.

The maintenance view offers an elaboration. GREEN is projectible because emeralds share a mechanism that produces greenness: chromium traces interact with light in stable ways. There's no mechanism that produces grueness; there's nothing about emeralds that makes them switch from green to blue at time t. Projectibility tracks mechanism, not just predicational habit.

For linguistic categories, the same logic applies. NOUN is projectible because mechanisms like acquisition, entrenchment, and functional pressure keep nominal properties clustering together. A pseudo-category like *nerboun* (noun if acquired before age 5, verb otherwise) isn't projectible because no mechanism produces that pattern – there's nothing about language acquisition that would cause a switch at age 5.

A category can look like a good bet – its definition correlates with usage, its label is entrenched – and still turn out to be a turkey.

6.6.1 Is projectibility interest-relative?

Craver (2009) and Onishi and Serpico (2022) argue that any mechanism-based account inherits context-dependence. Which mechanism you attend to, at what level of abstraction, with what boundaries – all depend on your explanatory goals. If so, the maintenance view answer to Goodman looks circular: a category is projectible relative to the mechanism you're interested in, and the mechanism you're interested in is whichever one makes the category projectible.

The key distinction is between choosing a level because it answers your question, and choosing a level because it privileges your category. A pedicab can be explained by engineering (frame geometry, gear ratios) or by physics (forces, friction, momentum). Both levels are real; both describe genuine causal structure. Your question selects the level. But the causal structure at that level determines whether your answer is correct. Interest picks the question; it doesn't fabricate the mechanism that answers it.

This is precisely what went wrong with the turkey. Its interest was survival, but it coarse-grained at the wrong level: time of day rather than time of year. The correlation was real at one level (11 AM $\rightarrow$ food) but collapsed at another (Thanksgiving $\rightarrow$ slaughter). The turkey's interest didn't invent the mechanisms; it just selected the wrong one to track.

For linguistic categories, the same logic applies. If your question is "what triggers perfective here?" the answer lives at the cue-structure level, not the semantic-definition level. Both levels exist; the semantic level just doesn't answer the production question. The aspect case shows the stakes: if your goal is to predict corpus distributions, or even to teach people to speak Polish, the textbook mechanism (semantic boundedness) is the wrong one – it predicts imperfective well and perfective poorly. If your goal is to predict native-speaker gap-filling, the lemma-concrete mechanism is the right one – it outperforms aspect-aware models. The prediction either succeeds or fails. That's the empirical constraint on which mechanism matters.

Conventionalism can't explain why some mechanisms predict and others don't. The maintenance view can: the mechanisms that predict are the ones that actually maintain the cluster. Interest selects among real causal structures; it doesn't invent them. This is what Carroll and Parola (2024) mean by emergence: coarse-grained descriptions that support accurate predictions despite discarding micro-level information. The key is coarse-graining *in the right way* – throwing away arbitrary information destroys predictability, but discarding the right information preserves it.

Lemeire (2018) argues that no purely epistemic theory – one that defines natural kinds solely by their inductive usefulness – can account for naturalness. If all we require is that a category support successful predictions, what distinguishes a natural kind from a merely convenient grouping? The objection is serious, and it's what the maintenance view is designed to answer. Projectibility alone is too permissive; what's needed is the right *kind* of projectibility – the kind that comes from genuine homeostatic mechanisms rather than accidental correlations. The turkey's law was projectible until Thanksgiving; a mechanism-grounded law would have predicted the farmer's purpose. The maintenance view doesn't abandon epistemic criteria; it requires that they be grounded in causal structure. Projectibility is evidence of mechanism, not a replacement for it.[2]

This gives us a robustness criterion for mechanistic kinds. A genuine mechanistic kind – as opposed to a convenient label – should exhibit *learning transfer*, *intervention stability*, and *cross-context generalisation*. Train on one subset, test on another: does it transfer? Intervene on the mechanism: does the pattern shift as predicted? Apply to a new context: does the generalisation hold? Labels without mechanisms should fragment under these pressures; mechanisms should persist. The computational evidence in this chapter – lemma-concrete models transferring to unseen verbs, cue–outcome associations generalising across tense frames – is exactly this kind of robustness test. The realism isn't permissive ("whatever predicts is real"); it's constrained by stability under perturbation.

Once interests are fixed by a task, we can ask how strongly different categories support that task. This is the next implication of mechanistic grounding: projectibility comes in degrees. (Recall that field-relative projectibility – §6.5 – is a distinct point: the same extension can be carved by different HPCs for different fields. What follows concerns degrees of projectibility *within* a single field.)

[2] Khalidi's (2013) "nodes in causal networks" account illustrates the difference. NPIs, for instance, are causally grounded in one sense – they share semantic sensitivity to negative contexts. What they lack is the homeostatic mechanism that would make them an HPC kind. They're stable at equilibrium rather than dynamically maintained: balls sitting at the bottom of a trough, not tops that must be spun to stay upright.

6.6.2 Degrees of projectibility

Not all categories are equally projectible. The framework expects a gradient, and the evidence supports it.

At one extreme: high-frequency, highly entrenched categories where the mechanisms are strong and consistent. English determinatives are a case: a closed class, stable across speakers, predictable in distribution. Learn *each* and you can project to *every* with high confidence. The mechanisms (entrenchment, functional specificity) are tight enough that the category approaches definitional coherence – the mechanisms are so strong that they produce uniform clustering.

At the other extreme: low-frequency, loosely maintained categories where the mechanisms are variable or in flux. Nonce formations, idiolectal forms, constructions undergoing change – these are less projectible because the mechanisms haven't stabilized. Knowing about one instance tells you less about others.

The middle ground is where the action is. Within a language, synchronically, major open classes – nouns, verbs, adjectives – are projectible on average but variable at the margins. You can project from typical nouns to novel nouns with high confidence; you can project from *fun* to other adjective-noun boundary cases with less. The mechanisms are strong enough to produce robust clustering at the core but not strong enough to determine the periphery.

The prediction is testable: core category members – high-frequency, prototypical, functionally central – should elicit consistent judgments; peripheral members – low-frequency, boundary-straddling, functionally ambiguous – should elicit variable ones. The variance isn't noise; it's the signature of weak entrenchment.

Recent experimental work confirms the pattern. In grammaticality judgments for Dutch syntax, peripheral structures are correctly identified only 57% of the time (SD $\approx$ 0.27), while core structures are correctly identified 90% of the time (SD $\approx$ 0.09) – the variance triples at the periphery (Favier & Huettig, 2021, Table 2). Similarly for syntactic variants indexed to verb frequency: low-frequency verbs elicit higher acceptance of violations (a fourfold increase in error rates: 6.58% vs. 1.58%), because weak entrenchment permits wider variation in what speakers treat as grammatical (Sassenhagen et al., 2018). In morphology, the pattern is starker still: high-frequency English adjectives elicit near-unanimous agreement on comparative form, while low-frequency and nonce adjectives split speakers (Graziano-King & Cairns, 2005).

This is what the maintenance view predicts. Where mechanisms are strong and consistent, judgments converge; where mechanisms are weak or variable, judgments diverge. The variance tracks the strength of the causal structure. Projectibility isn't all-or-nothing; it's graded by entrenchment. Crucially, the variance isn't just about the items; it's about the speakers. Low-frequency forms are "weakly entrenched" only in aggregate – for any given form, some speakers have encountered it often, others rarely, others never. What looks like category fuzziness at the population level is actually speaker-by-speaker variation in experience (Dąbrowska, 2012). This reinforces the maintenance account: categories fail to project uniformly because the maintaining mechanisms haven't operated uniformly across the population.

6.6.3 What the framework offers

The argument so far has been negative: traditional definitions don't ground projectibility; mechanisms do. But this raises a question for researchers who already study mechanisms. If you're already tracking entrenchment, distributional learning, alignment, or functional pressure, what does the maintenance view add? Isn't this just new terminology for what you were doing anyway?

The answer is that the framework offers something beyond the mechanisms themselves: an account of what the mechanisms are *for*. Most mechanistic research is descriptive in a specific sense – it shows how patterns emerge, stabilize, and change. What it often lacks is an explicit ontology. Are the categories that emerge from these mechanisms *real*, or are they convenient fictions? Are they features of speakers' minds, of the speech community, or merely of the analyst's framework? The maintenance view provides an answer: categories are real to the extent that they're maintained by consistent mechanisms. The mechanisms don't just produce the patterns; they *constitute* the categories as natural kinds.

This matters for several reasons.

First, it legitimises mechanistic work as theory, not just description. A sceptic might object that studying entrenchment or alignment is interesting but atheoretical – mere data-gathering that falls short of explanation. The maintenance view says: no. The mechanisms *are* the theory. Understanding how categories are maintained is understanding what makes them real. This is analogous to recognizing that natural selection isn't "just" a description of breeding patterns but the actual explanation of adaptation.

Second, it provides cross-camp coherence. Researchers studying entrenchment, alignment, distributional learning, and iterated transmission are often working in different sub-fields with different terminologies. The maintenance view offers a unifying frame: all are studying maintenance mechanisms for the same kind of category. This isn't eclecticism – it's convergence. Different instruments trained on the same phenomenon. The framework doesn't adjudicate between mechanisms; it says that *whichever* mechanisms actually maintain the cluster are the ones that matter. The question becomes empirical: which mechanisms predict? Chapter 15 shows how agent-based modelling provides one way to test such predictions computationally.

Third, it reorients methodology. The maintenance view makes prediction the success criterion: does the mechanism predict behaviour? This changes what counts as evidence. Individual variation stops being noise to be averaged away and becomes signal: if different speakers show different patterns, that tells you something about how tightly the mechanisms are operating. The gradient nature of projectibility – strong at the core, weak at the periphery – is a feature of maintenance, not a defect of the category.

Fourth, it offers category realism without essentialism. Many researchers working with mechanisms are implicitly anti-realist about categories – they treat category labels as convenient shorthand, not as picking out real kinds. The worry is understandable: if there are no essences, what makes categories real? The maintenance view provides the missing piece. Categories are real *because* of mechanisms, not despite lacking essences. The mechanisms do for categories what essences were supposed to do: explain why members resemble each other, why learning about one tells you about others, why the category is projectible. You don't need to choose between realism and mechanism; mechanism *underwrites* realism.

Fifth, it clarifies the relationship between historical origin and synchronic maintenance. The volcanic island metaphor applies here: what built a category may not be what maintains it. Researchers studying grammaticalization, semantic drift, or language change often face a puzzle: if categories evolved for one reason, but now function differently, what's their current status? The framework says: study the current mechanisms. Whether or not the historical origin matches the current maintenance, it's the current mechanisms that determine projectibility. You can trace the genealogy without claiming that genealogy is destiny.

None of this requires abandoning existing research programmes. The mechanistic work continues; what changes is its framing. The maintenance view shows how categories can be both analyst-labelled and grounded in real causal structure. The work you're doing isn't preliminary; it's the payoff.

We can now answer the question that opened the chapter. Mr. Tagomi needed to know which world he was in; learners of Polish need to know which verb form to use. Both rely on projectibility. Why should learning about one instance tell us anything about others?

Because the same mechanisms that shaped the instance you learned are shaping the instances you haven't. Categories maintained by consistent mechanisms produce consistent members. The consistency underwrites induction.

This is the epistemic payoff – what the good bet wins. HPC kinds are categories you can learn from, categories with causal structure. The mechanism that makes a word a noun also makes other nouns. The mechanism that biases a verb toward perfective also biases similar verbs. The mechanism that entrains a speaker to one usage pattern entrains the same speaker to related patterns.

Projectibility isn't metaphysically guaranteed. A proposed grouping can fail to project if its characterization doesn't match the maintaining mechanisms – as aspect-abstract fails. A proposed grouping can fail to project if the mechanisms are too weak or too variable – as nonce formations fail. In either case, the grouping isn't a category but a CLASS – a label without the causal structure to back it up. But where the mechanisms are real and consistent, projectibility follows. That's the test: does learning about instances let you project to the kind? If yes, the mechanisms are real. If no, look elsewhere for the causal structure.

The label names the island, not the volcano. But we've talked about mechanisms as if they were self-explanatory. They aren't. What exactly maintains the island? What processes stabilize the clustering, ensure that new instances resemble old ones, transmit the pattern across speakers and generations? The next chapter opens the black box. We'll trace the mechanisms in detail – acquisition, entrenchment, alignment, transmission, functional pressure – and ask what each contributes to the stability of linguistic kinds. Only then can we ask: under what conditions do mechanisms fail?

Part III then returns to field-relative projectibility with three extended case studies. Countability, definiteness, and word classes each demonstrate the pattern developed here: categories that decompose into semantic and morphosyntactic HPCs, maintained by distinct mechanisms, projectible for different analytical purposes. The framework isn't abstract philosophy; it's an operational tool for understanding why familiar categories behave as they do.

DEFINITENESS THREAD. This chapter's field-relative projectibility principle predicts that definiteness generalizations should be projectible for semanticists studying identifiability and uniqueness, while deitality generalizations should be projectible for syntacticians studying *there*-resistance and partitive licensing – but only where the stabilizer stories match. Register shift (e.g., legal versus colloquial) may reveal where the clusters decouple: weak definites and generic definites are predicted sites. The prediction would be falsified if a single unified projectibility claim held across all analytical purposes without stabilizer-specific scoping.

The Stabilizers

> The general rule would establish itself insensibly, and by slow degrees, in consequence of that love of analogy and similarity of sound, which is the foundation of by far the greater part of the rules of grammar.
>
> —Adam Smith, *Considerations Concerning the First Formation of Languages* (1761)

When a biologist asks what a macrophage is, they don't look for a definition – they look for properties that co-occur: typical functions (phagocytosis, antigen presentation), marker profiles, developmental origins. None is strictly necessary – some macrophages don't phagocytose; some share markers with dendritic cells – but because the properties cluster reliably, the category supports prediction and experimental design. The parallel is methodological, not ontological: I'm borrowing an explanatory style, not claiming that grammars are organisms.

Adam Smith saw something similar in grammar two and a half centuries ago. Rules "establish themselves insensibly, and by slow degrees" – not by fiat, but through the accumulated pressure of analogy and pattern-matching. This chapter adds the causal machinery that makes that slow consolidation intelligible. Recall the framework: *profile, stabilized by mechanisms, projectible relative to purposes*. What follows is an inventory of the stabilizers – the mechanisms that do the maintaining.

7.1 The cluster

Stability, for grammatical categories, is achieved rather than given. The cluster-first strategy applies to grammar directly. The same approach that lets immunologists sidestep definitional impasses can do the same for linguists.

Take nouns. In English, items we call nouns typically take determiners (*the dog*, *a problem*), inflect for number (*dogs*, *problems*), function as heads of noun phrases in subject and object positions, refer to entities or entity-like abstractions, and are commonly modified by adjectives. None of these is strictly necessary: *cattle* lacks a singular; names resist determiners; *information* doesn't pluralize. Cross-linguistically, the picture is messier still: Mandarin nouns don't inflect for number at all. And yet typologists recognize something noun-like across systems.

This is the position the chapter defends: grammatical categories are mechanism-maintained kinds whose reality is indexed to inductive utility, not definitional necessity – stabilized by a braid of cognitive, social, and physical processes. The clustering they produce is stable enough across time, robust enough across contexts, predictive enough across novel instances, that category membership reliably supports inference.

The question is no longer: what is the essence of noun-ness? Although linguists tried that approach – across semantic, distributional, and morphosyntactic definitions – it generated boundary disputes and competing proposals but no stable resolution. The mechanism-first question is different: what keeps the cluster clustered – what does the stabilizing work?

Grammatical categories are neither fully mind-independent natural kinds – like chemical elements – nor purely conventional human kinds that exist only because we agree they do, like money or traffic laws. They are a third thing: real enough to support induction, socially embedded enough to change under pressure. They lack the essences that define elements, but they resist the arbitrary revision that conventions permit.

"Mechanism" in this book means something specific: a mechanism posit earns its keep by specifying component processes whose interactions produce the target phenomenon. If you can't say what sub-processes combine to yield the clustering and what would happen if one were disrupted, you don't yet have a mechanism – you have a label with causal ambitions. Labelling a pile of bricks "Wall" does not make it load-bearing; only stabilizing processes do. In molecular biology, mechanisms are crisply bounded: the ribosome, the spliceosome, the gene regulatory network. For grammatical categories, the stabilizers are less crisply bounded – they include processing biases, acquisition pathways, social indexing, transmission dynamics – but the constraint is the same: the term is elastic enough to span cognitive and social scales, yet not so elastic that any correlation gets to call itself mechanistic.

Stabilizer is a related term. Where "mechanism" emphasizes causal structure, "stabilizer" emphasizes the functional role (Illari & Williamson, 2012) – what keeps the cluster clustered. A stabilizer is a mechanism insofar as it has causal depth, but the term foregrounds maintenance rather than mere causation. Something counts as a stabilizer if removing it would change the clustering – if the category would fragment, drift, or dissolve. What stabilizers maintain is not just clustering of form properties, but the coupling between form and value – between a unit's distributional signature and its contribution when deployed. A mere correlate – something that co-occurs with the category but doesn't contribute to its maintenance, like the orthographic length of a word that happens to be a noun – is not a stabilizer. This is where the meta-Occam payoff from Chapter 4 becomes concrete: the inventory of stabilizers is small; the categories they maintain are many.

7.2 Stabilizers at multiple scales

What the biological template provides is visibility: it lets us see variation as activation rather than noise, boundaries as dynamic rather than definitional, and stability as achieved rather than given. Without this framing, gradient membership looks like failed classification; with it, gradient membership becomes evidence about which mechanisms are operating and where.

The analogy is not homology. Cell biologists ask what mechanisms maintain the macrophage phenotype; we ask what mechanisms maintain the noun category. The question structure is the same; the mechanisms themselves are domain-specific. The parallel earns its keep by making linguistic phenomena more visible and testable.

Table 7.1 traces kinds of stabilizing stories; Figure 7.1 reorganizes the same mechanisms by timescale and locus, which determines what evidence is relevant. Claims about fast mechanisms (Quadrants I–II) require online processing measures; claims about slow mechanisms (Quadrants III–IV) require longitudinal and apparent-time data.

The definition tells you what the cluster looks like at a moment. The stabilizing story tells you why it persists.

The neural evidence reinforces this picture. Fedorenko, Piantadosi and Gibson (2024) show that the brain's language network is dissociable from the Multiple Demand network that supports reasoning and problem-solving: linguistic ability can be devastated while logical reasoning remains intact, and vice versa. This double dissociation matters for the maintenance view. If the forcing

Scale	Biology	Linguistics
Dynamical basins	Gene regulatory networks produce attractor states – stable configurations the cell tends to settle into and return to when perturbed. The basins are emergent from ongoing interaction and exist only as long as stabilizing processes continue.	Cognitive architectures and processing biases produce attractor states. A construction frequent enough becomes entrenched – produced as a unit rather than assembled. This is dynamical, not definitional.
Developmental trajectory	Developmental lineage constrains which basins are reachable. A cell's history matters: what precursors it passed through, what signals it received at critical windows.	Acquisition pathways and learning biases constrain what a speaker can learn. Order of acquisition matters. Early-learned patterns anchor later ones. Sensitive phases exist where input has disproportionate effect.
Microenvironmental modulation	Microenvironmental signalling – local cytokines, tissue context – pushes cells into different regions of the same basin, producing variation within the type.	Discourse context and pragmatic ecology push tokens into different regions of the distributional space. Same category, different activation state. Variation is expected, not noise.
Reciprocal feedback	Functional feedback: what the cell does alters its microenvironment, which sustains or shifts its state. The stabilization is not one-way.	Functional feedback: what speakers do with categories alters input to the next generation, which alters what categories stabilize. Usage facilitates reconstruction; facilitated patterns shape usage. The forcing function is communicative: forms that fail to optimise relevance – incurring processing costs without compensating benefits – are selected against (Scott-Phillips, forthcoming-b).

Table 7.1. Multi-level stabilization: biology and linguistics in parallel.

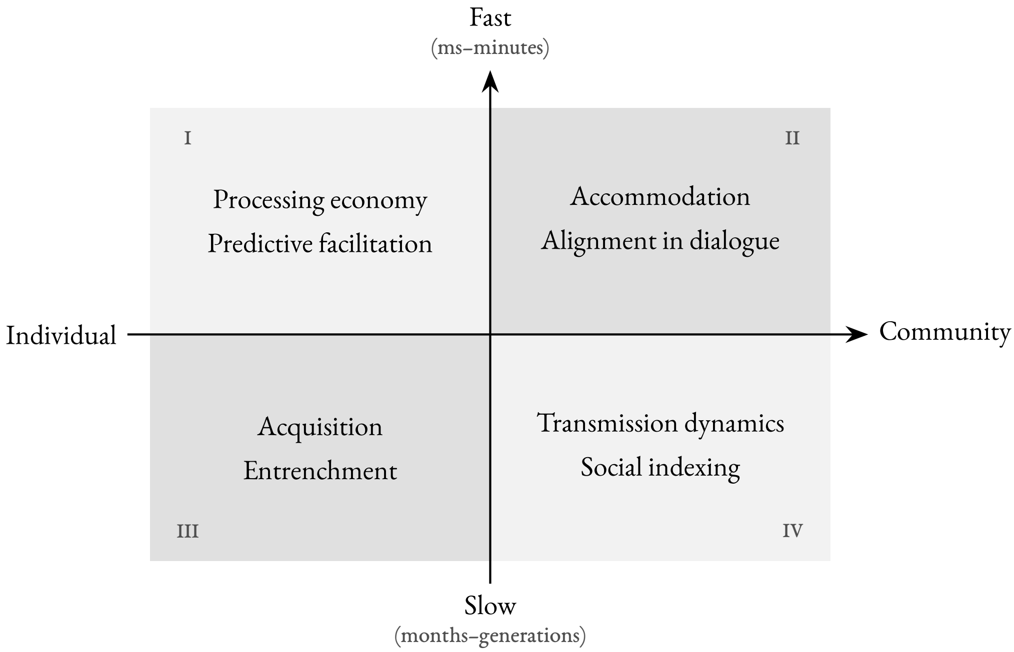


Figure 7.1. Mechanism typology by timescale and locus.

function maintaining linguistic categories were general cognitive pressure – efficiency of thought, logical organization – we would expect language and reasoning to share neural resources. They don't. The forcing function is specifically communicative: what stabilizes linguistic structure is its success in coordination between speakers, not its utility for internal cognition.

The four-level story is a heuristic, not a template. Different categories are maintained by different braids of mechanisms; the exact mapping varies. Quotatives, as we'll see, involve processing economy, social indexing, and transmission dynamics – mechanisms that cross-cut the four levels rather than mapping onto them one-to-one. The point is multi-level stabilization, not exact correspondence. What the biological parallel provides is a style of explanation – look for mechanisms at multiple timescales, expect reciprocal feedback – not a fixed inventory. The case studies that follow apply this style to two deliberately different phenomena: quotatives (socially vivid, spreading through youth networks) and filler-gap constructions (more syntactic, depending on information structure). If the framework earns explanatory traction for both, it's not just sociolinguistics with a new label.

Several mechanisms deserve mention even if they don't receive full case studies here. Analogy extends established patterns to new cases by perceived similarity – it's how children produce *goed* and how adults accept novel compounds. Standardization operates through institutions: schools, style guides, dictionaries, corpus-annotation schemes. These codify and enforce norms, creating explicit feedback loops that natural transmission lacks. Indexicality links linguistic forms to social meaning: a choice of variant signals group membership, registers attitude, or marks formality. The social stakes make the choice salient and its patterning consequential, reinforcing category boundaries beyond what communicative function alone would produce. Error and repair maintains structure through on-line correction – speakers monitor output, catch deviations, and adjust mid-utterance. This creates real-time negative feedback that keeps production within the category's basin. Finally, creativity probes boundaries: novel forms test what the system will tolerate, and their acceptance or rejection reshapes what counts as the category. These mechanisms interact with the four-level typology rather than mapping neatly onto it; a full constructive account would need to integrate them all.

7.3 Variation as activation states

The immunologist's key move: variation within a category is not embarrassing fuzziness to be explained away. It's the expected signature of a kind that is maintained by context-sensitive mechanisms.

A macrophage in inflamed tissue shows different markers than a macrophage in healthy tissue, but both are macrophages. The difference is activation state – which signals the cell has received, which region of the phenotypic space it currently occupies. The category is real, and so is the variation; both are explained by the stabilizing dynamics. The same logic applies to linguistic categories.

Two kinds of variation matter, and the distinction is stabilizer-relevant. Some apparent boundary cases are *within-speaker* shifts – the same person producing different patterns in different contexts. These are activation-state differences: one grammar, multiple accessible patterns. Others are *between-speaker* differences – not everyone has entrenched the same patterns to the same depth. Speakers with different input histories, different literacy experiences, different interactional ecologies end up with different category boundaries. Within-speaker variation shows that categories are context-sensitive; between-speaker variation shows that they are experience-dependent. Both are part of the data that a theory of category maintenance needs to explain.

That in turn distinguishes *exposure-dependent stabilization* from *population-wide stabilization*. A category that stabilizes only in speakers with particular exposure profiles is a different kind from one that stabilizes across the population.

The former is maintained by mechanisms that not everyone encounters; the latter is maintained by mechanisms that operate across the population.

Consider *fast*. Traditional grammars call it an adjective in *a fast car* (attributive modifier position) and an adverb in *she ran fast* (post-verbal modifier, no *-ly* suffix). Same lexeme, different syntactic behaviour. The question "is *fast* an adjective or an adverb?" assumes a forced choice. Three frameworks give different answers.

The essentialist answer: it must be one or the other, or else homonymous (two lexemes, same form). The prototype answer: it's a gradient case, partly adjective, partly adverb, with no fact of the matter about which it "really" is.

The maintenance answer is different. *Fast* sits on a plateau linked to both the adjectival and adverbial basins – a region where both diagnostic profiles remain accessible (see Figure 5.2). The lexeme is stable; what varies is the token's route, determined by syntactic frame. Attributive position takes the adjectival pass (comparison, agreement in languages that have it); post-verbal position takes the adverbial one. That's stability at the type level with choice at the token level – not a new essence, but terrain that permits dual access.

The same framing applies to aspect in Polish. As §6.2 showed, 90% of verbs strongly prefer one aspect – they're deep in a single basin. The remaining verbs occupy shallower regions where contextual signals matter more. The verbs that require explicit contextual cues to disambiguate are the ones where entrenchment is weaker, where the tense-frame signal has more work to do. That's the activation-state picture applied to grammatical aspect. Section 7.5 applies the same idea to a construction rather than a lexeme: independent relative *whose*, which succeeds under specific licensing conditions and fails without them.

The activation-state logic has neural consequences. If categories are maintained by interacting cues rather than definitions, then processing should track graded predictive fit, not all-or-none membership. Event-related potentials offer a window here, though the picture is messier than textbook summaries suggest. Components like the N400 and P600 don't map cleanly onto 'semantic' versus 'syntactic' violations; their amplitude and timing reflect prediction error and retrieval difficulty in ways that cut across traditional module boundaries. What matters for the present argument is the pattern: HPC predicts continuous sensitivity to cue convergence – the more cues align, the stronger the facilitation; the more they conflict, the larger the processing cost. Classical definitional categories predict bimodal profiles – clear members and clear non-members – with borderline cases showing high variance rather than intermediate values. Under HPC, borderline items should show systematic intermediate values: graded neural profiles that track cue-weighting, not random scatter around a binary threshold. That's a testable contrast.

7.4 One case in depth: the emergence of new quotatives

Rather than enumerate mechanisms abstractly, I'll trace the stabilizing story for one category: the quotative marker.

Quotatives introduce reported speech, thought, gesture, or sound. The functional need to distinguish the narrator's voice from the character's is presumably as old as narrative itself. But when we look across typologically unrelated languages, we find a striking convergence in recent history: in the late twentieth century, new quotative forms emerged or expanded – forms with a distinctive profile that spread through similar populations and stabilized through similar mechanisms.

The English case is well documented, and the chronology reveals the dynamical logic: what emerges first, what locks in later, and which contexts serve as incubators.

Butters (1982) first noted *be like* in American English in 1982, observing its use to report unuttered thoughts: *I was like, 'Let me live, Lord'.* Ferrara and Bell (1995) hypothesised that the form originated with first-person subjects and internal dialogue before expanding to third-person subjects and direct speech. The incubator was informal peer narrative – storytelling among friends, where enactment matters more than accuracy. By the mid-1990s, the form had spread beyond that niche. Tagliamonte and D'Arcy (2004) tracked apparent-time data from Canadian English: in 1995, speakers aged 18–27 used *be like* for 13% of quotative tokens; by 2003, speakers aged 17–19 used it for 63% – a fivefold increase in under a decade. The same cohort, re-sampled seven years later, showed *be like* rising from 13% to 31% as they aged into their thirties (Tagliamonte & D'Arcy, 2007). This is not age-grading; it's real-time change, with each generation carrying its quotative inventory into adulthood. A micro-regularity becomes meso-level entrenchment; entrenchment becomes macro-level expectation; expectation shapes input to the next generation.

German shows a parallel trajectory. Golato (2000) analysed *und ich so* ('and I'm like') in video-recorded conversations, documenting its function in storytelling: introducing punchlines, sound effects, gestures, and verbal enactments – the same profile as English *be like*. The form was attested in youth speech by the late 1990s (Androutsopoulos, 1998) and has since become established among younger German speakers, though it remains more register-restricted than its English counterpart.

Japanese *tte* (reduced from *to itte*, 'saying that') functions similarly: a phonologically light form that introduces reported speech, thought, or stance in informal narrative. The pattern extends to Turkish *diye* (converb of *demek*, 'to say'), which has expanded from its older quotative functions into increasingly colloquial discourse contexts.

Buchstaller (2014) synthesises the cross-linguistic evidence: the emergence of these innovative quotatives reflects independent development under similar functional pressures, rather than media-driven borrowing. Different sources, different structures, same functional niche, same stabilizing dynamics.

This is convergent maintenance, not coincidence. Independent forms. Same mechanisms. Same architecture.

7.4.1 The quotative cluster

What properties co-occur in these new quotatives? The best-documented case is English, so I begin there before asking how far the pattern generalises.

Non-lexical content. *Be like* introduces material that goes beyond words: gesture, facial expression, tone of voice, inner monologue. *And I was like* doesn't just report what someone said; it enacts the experience—recruiting prosody, embodied performance, and affective display as part of the quotation. Ferrara and Bell (1995) noted that early *be like* tokens disproportionately introduced internal states and non-lexicalised sounds – groans, sighs, exclamations – before expanding to verbatim-style direct speech. This enactment function is not English-specific: Golato (2000) documented the same pattern for German *so*, showing that it turns reported speech into performance.

First-person and present-tense preference. In English, the forms attach preferentially to first-person subjects and historical-present tense. Tagliamonte and D'Arcy (2004) found that in their Toronto data, first-person contexts strongly favoured *be like* over *say*. When speakers narrate, they use these forms to bring the audience into the moment of the original event – the vivid first person, the historical present. Whether this first-person preference holds for German *so*, Japanese *tte*, or Turkish *diye* is not yet quantified; the functional logic (narrator entering the scene) suggests it should, but that remains a prediction rather than an established finding.

Youth association. Younger speakers lead adoption. For English, the apparent-time data are robust: the highest rates of *be like* appear among speakers under 30; rates decline sharply for speakers over 50 (Tagliamonte & D'Arcy, 2004, 2007). Young women, in particular, tend to be innovative adopters – a pattern consistent with the broader sociolinguistic finding that women often lead linguistic change from below (Labov, 2001) and predictably attracting editorial complaint. German shows the same pattern: Androutsopoulos (1998) identified *so* as a youth-speech marker. For Japanese and Turkish, youth association is widely assumed but lacks the same quantitative backing.

Register restriction. The forms are discourse-conditioned: common in storytelling, informal speech, peer conversation; rare in formal registers – exactly where we find the so-called "historical present". Even as *be like* has spread across age cohorts, it remains suppressed in academic prose, institutional speech, and written genres. German *so* is still more restricted – marked as adolescent, rarely encountered in broadcast media. The register pattern is documented for both languages; for Japanese and Turkish, the functional parallel is likely but awaits corpus confirmation.

These properties cluster. A quotative that introduces vivid re-enactment, prefers first person and present tense, spreads through young speakers in informal contexts – this is a recognizable type. The English evidence is quantitatively richest; the German evidence confirms key features; the Japanese and Turkish cases show the same functional niche but with less distributional detail. What we can say with confidence is that the cluster exists in multiple unrelated languages. What remains open is how tightly the specific features correlate outside English.

To make the ecological logic overt: these innovative quotatives occupy a functional niche that older forms leave empty. The niche is *approximate vivid re-enactment in informal narrative* – conveying gist, stance, and affect without committing to verbatim accuracy. *Say* implies exact report; the new quotatives disclaim it. This functional specialisation illustrates the projectibility discussed in §6.4: within this niche, predictions about content type, register, and speaker characteristics are remarkably reliable. The niche opened because narrative speech needs enactment as much as report, and older quotatives didn't serve that function well. The stabilizing pressures are the ones that keep any niche occupied: processing economy (light forms win under production pressure), functional fit (forms that serve communicative needs spread), social indexing (forms that mark in-group identity persist), and transmission dynamics (forms acquired in dense peer networks survive into the next generation). What would count as niche collapse? If another form arose that served approximate re-enactment better – lighter, more expressive, better indexed – the current quotatives would cede territory. We see hints of this already: in some varieties, *be all* briefly competed with *be like* before losing ground. The niche is stable, but the occupants are not guaranteed tenure.

The question is: why does this cluster cluster?

7.4.2 Quotative stabilizers

At least five mechanisms braid together to maintain quotatives. Some are cognitive (processing economy), some functional (expressive fit), some developmental (acquisition and cohort effects), some social (indexing and transmission). Each contributes, but none is sufficient alone.

Processing economy. Quotatives of this type are structurally light. Japanese *tte* (as an informal, reduced quotative with comparable discourse advantages) is a phonologically reduced form of *to itte* ('saying that'); the reduction removes syllables while preserving function. German *so* is a single syllable. English *be like* is syntactically minimal: copula plus predicative, no subordinator, no overt speech verb. Forms that reduce production cost in informal narrative routines are used more frequently, entrenched more deeply, and survive transmission. Frequency here acts as a causal pathway, not just a descriptive statistic: repeated processing of a form makes future processing easier, reduces articulatory effort, and creates expectations that bias subsequent production. The result is a feedback loop: light forms spread because they're easy; spreading makes them easier still. The specific frames *and I'm like*, *I was like*, *she's like* become retrieval units, further lowering production cost. The mechanism is familiar from irregular morphology (Bybee, 2006): *went* resists levelling to *goed* because high frequency keeps the specific form accessible. As with irregular morphology, the effect is item- and frame-specific, not a rewrite of the general processing architecture.

Expressive fit. These quotatives fill a functional niche that older forms leave empty: introducing inner monologue, gesture, and attitude without committing the speaker to verbatim accuracy. *She said* implies exact report; *she was like* implies approximation, enactment, stance. One provides the court transcript; the other provides the biopic. Ferrara and Bell (1995) documented this as the form's original semantic territory – internal states, non-lexicalised sounds – before it expanded to canonical direct speech. The looseness is functionally adaptive. Speakers need to convey gist, stance, affect – not transcript. Forms that serve communicative needs better than alternatives spread. The form's preference for the present tense also recruits a basic cognitive bias (the mapping of "present" to "close"), making it robust against loss. A form that fits the narrative niche is precisely the one that becomes frequent enough to be acquired as routine in peer storytelling.

Acquisition and cohort effects. This mechanism operates within speakers and cohorts: children and adolescents acquire these forms in dense peer networks where the input is frequent, contextually salient, and socially marked. Panel data provide the critical evidence: the same Toronto speakers, tracked over seven years, showed *be like* rising from 13% to 31% of quotative tokens as they aged from their twenties into their thirties (Tagliamonte & D'Arcy, 2007). This is not age-grading (using youth forms only when young); it's lifespan change. Speakers carry their adolescent inventory into adulthood, and their children hear it as ambient input. The result is a cohort effect: each generation's linguistic inventory reflects what was frequent in their formative years.

Social indexing. As Eckert (2012) argues, social meaning is part of linguistic structure, not a by-product. New quotatives initially index youth, informality, in-group membership. Using *be like* signals that you are a certain kind of speaker, in a certain kind of register, with a certain kind of audience. This social value is not epiphenomenal; it's a stabilizer. Older speakers often react with suppression, finding the forms annoying or inarticulate. But this constraint is generational: speakers who grew up with the forms carry them into adulthood, using them not as rebellious markers but as part of their native repertoire. Speakers maintain the form partly because it does identity work – marking solidarity, casualness, narrative immediacy. Buchstaller (2014) found that attitudes toward *be like* track age and gender: listeners perceive its users as younger and more socially attractive, even while rating them lower on status dimensions. The form is maintained partly *because* of this indexical profile. Social indexing explains why individuals adopt the form.

Transmission dynamics. Transmission dynamics explain how the population changes. The apparent-time pattern – younger speakers use *be like* more – becomes a real-time change as each cohort ages. Tagliamonte and D'Arcy (2004) showed that Canadian speakers aged 17–19 in 2003 used *be like* for 63% of quotative tokens, compared to 13% for speakers of the same age in 1995. That's not two populations differing in age-graded behaviour; it's the S-curve of language change in progress – slow uptake, rapid spread, eventual saturation. Forms that suit the demographic and interactional structure of transmission – peer-to-peer in adolescence, parent-to-child thereafter – survive and spread. The trajectory has three phases: innovation (youth-marked), spread, and consolidation (neutralised). Figure 7.2 schematises the relationship between frequency diffusion and social-meaning consolidation: the lag between curves reflects the expected period during which the form carries strong indexical value.

The quotatives case is a site where predictive processing and social indexing co-reinforce. Predictability encourages uptake; uptake feeds entrenchment; entrenchment makes the form more predictable. Meanwhile, social indexing strengthens both the processing advantage and the identity signal.

Online processing pressures – prediction, retrieval ease, reduced production cost under repetition – preferentially reinforce some cluster features over others. The features that are easy to predict, easy to produce, and easy to retrieve are the

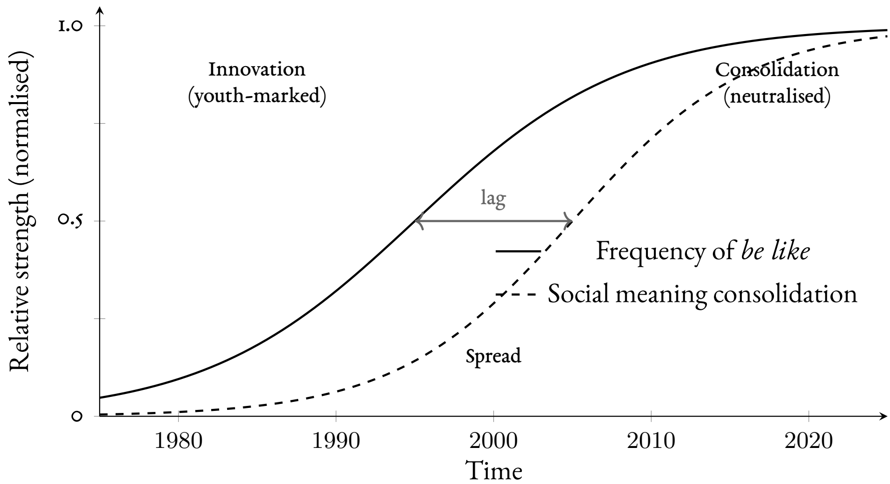


Figure 7.2. Schematic relationship between frequency rise and social-meaning consolidation for quotative *be like*. Both curves represent relative trajectories rather than measured values; the frequency curve abstracts over attested data patterns, the consolidation curve reflects the qualitative pattern. The lag between curves reflects the expected period during which the form carries strong indexical value (youth, informality); consolidation marks the attenuation of marked social meaning as the form becomes the community default.

ones that survive and spread. That selective reinforcement keeps the category coherent across time and speakers. Processing isn't just something that happens to categories; it's one of the mechanisms that maintains them.

7.4.3 What if a mechanism were absent?

A stress test: what if we removed each stabilizer in turn? If the braid is genuinely explanatory, no single strand should suffice.

If processing economy were the only stabilizer, the shortest form would win – but *say* is shorter than *be like*, and it doesn't dominate.

If expressive fit were the only stabilizer, any vivid quotative would spread equally – but forms with pejorative associations don't spread even when they fit the discourse need.

If social indexing were the only stabilizer, quotatives would cycle with generational fashion – but *be like* has been stable for four decades, long past a typical lexical-fashion cycle.

If transmission were the only stabilizer, all childhood forms would persist – but archaic quotatives like *quoth* don't survive.

If acquisition were the only stabilizer, forms entrenched early would persist within individuals but fail to generalise – but quotatives that never achieve narrative or indexical value don't survive beyond the cohorts that first encountered them.

The observed pattern – cross-linguistic convergence on light, expressive, youth-indexed, narratively deployed quotatives – requires the full mechanism package.

These counterfactuals generate testable predictions. Table 7.2 summarizes the expected empirical signatures if each stabilizer were selectively attenuated while others remained intact.

Mechanism attenuated	Freq.	Judgment variance	Social meaning	Cross-context transfer	Historical trajectory
Processing economy ↓	gradual ↓	slight ↑	retained	reduced	slowed spread
Expressive fit ↓	stable → gradual ↓	moderate ↑	weakened	reduced	stalled expansion
Social indexing lost	stable	stable	neutralised	stable	age-grading only
Acquisition disrupted	cohort ↓	high ↑	fragmented	patchy	generational gaps
Transmission blocked	rapid ↓	low within cohort	fossilised	retained	arrested

Table 7.2. Predicted empirical signatures of mechanism attenuation. Each row specifies observable consequences if a single stabilizer were weakened while others remained intact.

7.4.4 Cross-linguistic convergence

Similarity across languages can arise by two routes: diffusion of a particular form across related varieties, or convergent development under shared stabilizing pressures. Both are relevant here.

Japanese, Turkish, German, and English share almost nothing: different word orders, different morphological profiles, different sociolinguistic ecologies. Yet they converge on similar quotative innovations. Within English varieties, the global spread of *be like* suggests diffusion with local adaptation. Across unrelated languages, innovative quotatives arise from different lexical sources but converge on a similar functional territory. The maintenance view predicts this dual pattern: a form may travel, but even when it doesn't, similar stabilizers recruit new material into the same niche.

The temporal pattern is striking. English *be like* was first attested in 1982 (Butters, 1982) and spread globally within two decades: UK by 1994, New Zealand, Canada, and Australia by the late 1990s, then outer-circle Englishes through the 2000s (Buchstaller, 2014).

One obvious hypothesis is media transmission. The 'Valley Girl' stereotype, propelled into popular culture by Frank Zappa's 1982 song and subsequent Hollywood films, made *be like* audible to millions. But Buchstaller (2014) finds little evidence for straightforward media-driven borrowing. Speakers rarely associate *be like* with media stereotypes; the form's linguistic conditioning varies across localities in ways inconsistent with uniform transmission; and the development pattern shows 'transformation under transfer' – local adaptation of a form that arrived from elsewhere – rather than passive copying.

The stronger hypothesis is convergent development under similar functional pressures. The evidence for this is suggestive rather than conclusive, but the pattern is striking: the semantic-pragmatic sources for innovative quotatives – deictic markers, movement verbs, approximation expressions, taxonomic nouns – recur across typologically unrelated languages (Güldemann, 2008). German recruits *so* (a deictic), Japanese recruits *tte* (a reduced quotative verb), Turkish recruits *diye* (a converb), Russian recruits *tipa* (a taxonomic noun meaning 'type'; see Kolyaseva 2018), and English recruits *like* (a preposition). The sources differ, but the functional territory is the same: introducing vivid, non-verbatim re-enactment in informal narrative.

The convergence is not lexical; it's structural. The stabilizing mechanisms are the same:

- Processing economy favours light forms.
- Expressive fit favours vague-reference quotatives.

- Social indexing favours youth-marked forms.
- Transmission dynamics favour high-frequency, contextually salient forms.

Wherever these mechanisms operate – and they operate everywhere humans tell stories to each other – the same type of quotative emerges. The category is not defined by a shared etymon or a universal grammar rule. It's stabilized by a convergent mechanism profile. (The claim is established for English, well-supported for German, and a prediction for Japanese and Turkish.)

A caution about what convergence shows. Similar surface profiles suggest convergent *problems* – similar functional pressures, similar communicative needs. They don't automatically demonstrate the same underlying stabilizers at the same relative weights. In one ecology, processing economy may dominate; in another, social indexing may do more work. German *so* remains more register-restricted than English *be like*, even though both serve the same niche – which suggests that the social-indexing pressures differ, even as the processing pressures align. Convergent profiles are evidence that the *type* of mechanism is similar; the specific braid may still vary across ecologies.

This is what "same category across languages" means in the maintenance view: convergent stabilization, the clustering that emerges when similar mechanisms operate under similar pressures. The convergence is predictable: if processing economy and expressive fit are universal – all humans favour light forms under production pressure, all humans need vague-reference quotatives for storytelling – then similar sociolinguistic dynamics will produce similar outcomes wherever they operate. But "similar outcomes" needn't mean identical mechanisms. The framework predicts family resemblance in the stabilizing story, not clones.

7.4.5 How deep do mechanisms go?

The stress test confirms necessity: no single strand suffices. A deeper question: what *are* these mechanisms? The stabilizers just described – processing economy, expressive fit, social indexing, transmission dynamics – might look like endpoint explanations. But mechanisms have mechanisms. Each stabilizer can be decomposed causally (what underlies it?) and mereologically (what are its parts?). The decomposition reveals the multi-level architecture of category maintenance.

Causal depth

Take the robust finding that young women lead quotative innovation (Labov, 2001; Tagliamonte & D'Arcy, 2004). This isn't a brute fact. It plausibly traces back to social-psychological and network structure.

Young women's peer networks are often reported to be denser and more multiplex than young men's, especially in adolescence (Milroy, 1987). Dense networks mean more linguistic input, more accommodation pressure, faster entrenchment of shared forms. Multiplex ties – relationships serving multiple social functions – mean higher interactional stakes and stronger motivation to coordinate. When your interlocutor is also your classmate, neighbour, and confidante, alignment is more consequential than with a stranger.

But why dense networks in adolescence at all? Developmental psychology provides the next level down. Adolescence is the period of separation-individuation, when peer orientation overtakes parent orientation (Erikson, 1968). The social structure of schooling concentrates age-cohorts in prolonged daily contact – co-location plus time produces the network density that shapes linguistic input. Identity construction becomes primary: you need to differentiate from your parents while affiliating with your peers. Linguistic innovation serves both goals simultaneously – it marks in-group solidarity and out-group distinction.

This is a *possible* depth path, not a confirmed one. The sociolinguistic pattern (young women leading) is robust; the network-density explanation is well-supported; the developmental and evolutionary claims become progressively more speculative. The framework permits depth without requiring it – readers who find the deeper levels overreaching can stop at network density without losing the main argument.

How far down should linguistic explanation go? Far enough to explain the clustering at the level we care about. For the quotative case, the clustering is grammatical: first-person preference, youth association, non-verbatim content. The mechanisms are social and cognitive: network density, identity work, processing pressure, entrenchment dynamics. We *could* trace further – to neuroscience, to evolutionary biology – but the grammatical clustering is already explained. The framework is level-agnostic: it says look for mechanisms that produce clustering, without dictating which level to bottom out at.

Mereological structure

The same decomposition applies to parts, not just causes. "Transmission" sounds like a single mechanism, but it's a composite.

Production processes come first: lexical selection (choosing *be like* over *say*), which involves activating competitors, weighting by frequency, and filtering by context – register, addressee, narrative function. Then syntactic planning: slotting the copula and predicative into a quotative frame. Then phonetic realization: the actual articulation of *like*.

The signal bridges production and perception: acoustic properties, prosodic packaging (the characteristic intonation contour that marks enactment).

Perception processes follow: segmenting the stream, recognizing words, parsing the construction, assigning category membership. Each is a sub-mechanism with its own structure – word recognition draws on frequency-weighted access, construction parsing draws on facilitated patterns, category assignment draws on distributional learning.

Each encounter updates the conditions for future production: encoding this token in context (acoustic trace, syntactic frame, pragmatic function, social co-occurrence – who said it, to whom, with what stance), facilitating future reconstruction of similar forms, strengthening form-function associations. The listener becomes a speaker whose lexical selection is now shifted slightly toward the form just heard.

Social processes interleave throughout: accommodation (matching the interlocutor), identity projection (signaling youth and informality), stance-taking (performing enactment rather than merely reporting). Pickering and Garrod (2004a) model alignment as automatic priming; Eckert (2012) situates it in identity construction. Both are part of what "transmission" means.

The mereological lesson is that even the mechanisms introduced in §4.3 – acquisition, entrenchment, alignment, functional pressure – are not atoms. They're composites with internal structure. When we say "transmission maintains the category", we're compressing a multi-part process into a single term. The term is still explanatory – it picks out a real causal structure – but the structure is articulated, not monolithic. The point here is the compositionality of transmission, not adjudication between competing models of priming versus identity-driven alignment.

What this means for explanation

The depth and structure of mechanisms matter for two reasons.

First, they reveal where intervention is possible. If transmission depends on particular memory processes, then anything that disrupts those processes – reduced input frequency, competing forms, register restriction – will weaken the mechanism. Knowing the parts tells you where the leverage points are. This is why Chapter 8 on failure modes will matter: category dissolution happens when sub-mechanisms fail, and understanding which sub-mechanisms are failing requires knowing what they are.

Second, they connect linguistic explanation to the broader sciences. The causal depth that runs from "young women lead change" through network density to developmental psychology to evolutionary pressures is where linguistics fits into the larger picture. The maintenance view doesn't require every linguist to become a psychologist or evolutionary biologist. But it does make explicit that linguistic categories are maintained by mechanisms that ultimately ground in facts about human cognition, social structure, and history. The framework is continuous with the rest of science, not sealed off from it.

Mechanisms as categories, categories as mechanisms

A deeper point emerges from this decomposition.

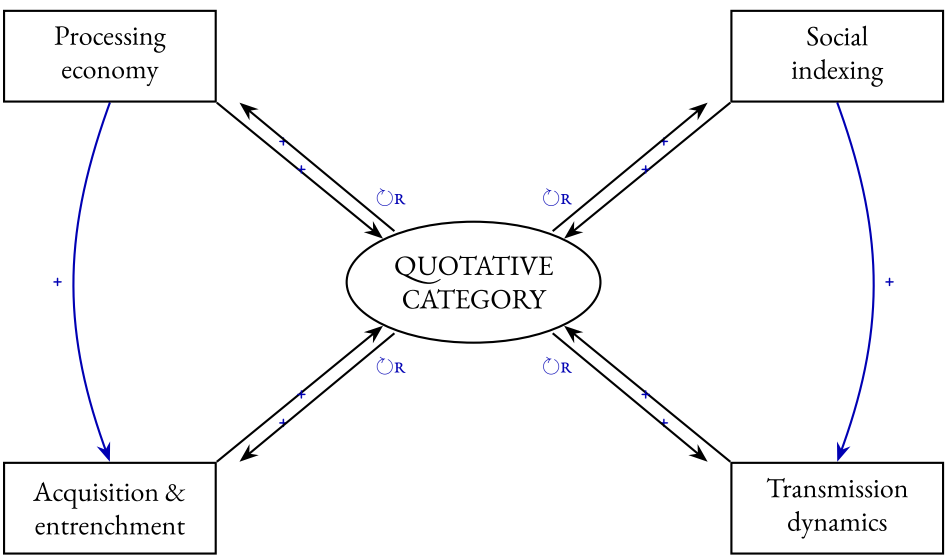
If categories are projectible clusters stabilized by mechanisms, and if the mechanisms themselves are projectible clusters stabilized by further mechanisms, then mechanisms are categories. This claim is a theoretical consequence of the framework rather than an independent empirical finding – but it has testable implications. PROCESSING ECONOMY is a category – it bundles properties (reduced forms, high frequency, resistance to analogical pressure) that cluster reliably because of underlying neural and memory constraints. SOCIAL INDEXING is a category – it bundles properties (youth association, in-group marking, register restriction) maintained by identity construction and network dynamics. The mechanisms we invoke to explain quotative stability are themselves HPC kinds.

The stabilization, moreover, is reciprocal.

Quotatives don't just *use* processing economy – they *maintain* it. Every time a speaker produces *be like*, that token is input to the conditions that ground processing economy. The form's high frequency reinforces the facilitation that makes high-frequency forms easier to produce – not by rewriting the rules of activation, but by keeping the relevant usage ecology dense. The mechanism depends on the category for its continued operation just as the category depends on the mechanism, which, like a sourdough starter, goes mouldy if neglected.

Social indexing exhibits the same reciprocity. It isn't just something that quotatives *do*; it is partly *built out of* the patterned circulation of particular quotative variants. The indexical field that makes youth-associated forms socially meaningful persists because forms like *be like* repeatedly occur with those personae and stances. The indexical field is an emergent structure sustained by the form's patterned social distribution. You get "youth style" the way you get a trail in the woods: by people walking it. Remove a major, high-salience contributor to that pattern and the local indexical alignment would likely attenuate. In that sense, the category and the mechanism co-construct each other.

This is not a vicious circle; it's a self-organizing dynamic – precisely the homeostatic structure that gives HPC kinds their stability. Categories shape what gets entrenched, and entrenchment shapes which categories survive, so each stabilizes the other. Figure 7.3 visualizes this reciprocal structure.



External inputs: memory architecture, network structure, developmental ecology

Figure 7.3. Reciprocal maintenance of a grammatical category. Each mechanism both stabilizes and is sustained by the category it maintains. Arrows marked + indicate reinforcing relationships; ↻R marks reinforcing feedback loops.

A contrast case sharpens the point. Consider the label PARTICLE – used in much typological and pedagogical work as a wastebasket for uninflected words that don't fit elsewhere: *up* in *look up*, *to* in infinitives, *not*, sentence-final *eh*. PARTICLE is the "Other" box on a survey: it records leftovers; it doesn't identify a kind. Does this cluster survive a maintenance audit? The properties don't co-occur reliably: *up* takes complements; infinitival *to* doesn't; *not* has scope properties shared by neither. There's no common acquisition pathway: children learn verb-particle constructions, infinitives, and negation at different times, through different input distributions. There's no processing signature: reaction-time studies don't show "particle" effects that generalise across subtypes. There's no transmission dynamic: speakers don't acquire "particle-hood" as a generalisable category; they acquire constructions that happen to contain uninflected words. Predictions trained on one subtype (verb-particles) fail to transfer to another (infinitival *to*). The label exists; the mechanism-maintained kind doesn't. Failure of reciprocal maintenance looks exactly like this: the supposed category doesn't exercise a shared mechanism, and no shared mechanism makes the category projectible. Chapter 8 develops such failure modes systematically; here, the contrast case shows that the quotative analysis isn't analogy-by-enthusiasm – it identifies genuine stabilization that PARTICLE lacks.

The quotative category is also a mechanism in the larger kinds that contain it. Quotatives are part of what maintains NARRATIVE DISCOURSE STRUCTURE – they enable the vivid re-enactment that makes storytelling work. They are part of what maintains YOUTH REGISTER – the forms that index adolescent identity and solidarity. They are part of what maintains INFORMAL SPEECH as a recognizable variety – the cluster of features that signals casualness, immediacy, peer context. And again, the stabilization is reciprocal: narrative structure provides the discourse ecology in which quotatives function; youth register provides the social meaning that quotatives carry; informal speech provides the register niche that quotatives occupy.

So the hierarchy runs both ways. Downward: quotatives are maintained by mechanisms (processing, acquisition, social indexing, transmission), which are themselves maintained by deeper mechanisms (memory architecture, developmental psychology, network structure). Upward: quotatives are mechanisms in larger categories (narrative structure, youth register, informal speech), which are themselves mechanisms in still larger kinds (discourse genres, sociolinguistic repertoires, speech communities). And at every level, the stabilization is bidirectional: each element both depends on and sustains the elements above and below it.

A sceptic might ask: if every mechanism is itself a category maintained by further mechanisms, what grounds the explanatory chain? Two responses are available.

The interventionist answer: we stop when we reach mechanisms whose disruption would change the clustering in predictable ways. This is already implicit in the robustness tests above ("What if a mechanism were absent?"). A mechanism earns its keep by specifying where intervention would bite. If we can say "reduce input frequency and the category weakens", we have a mechanism. The grounding is not metaphysical bedrock; it's causal tractability.

The anti-foundationalist answer: explanatory chains don't need bedrock. The regress objection assumes that justification must terminate in self-evident foundations – but coherentist epistemologies reject this. What matters is that the explanatory network is mutually supporting and empirically fruitful. Each mechanism explains the categories above it and is explained by the mechanisms below it; the network as a whole is tested by whether it generates accurate predictions. The demand for a single ultimate ground is a legacy of foundationalist philosophy that the sciences have largely abandoned.

Both responses are compatible. The framework is interventionist in practice and anti-foundationalist in principle. For any given explanatory purpose, we stop where the causal leverage is – but we don't pretend that stopping point is metaphysically privileged.

7.5 A second case: filler-gap and independent relative ⟨whose⟩

The quotative case shows mechanisms maintaining a category primarily from above – from the functional and social pressures that keep the clustering clustered. But a sceptic might wonder: is this just sociolinguistics dressed in HPC clothing? A second case answers that challenge by showing mechanisms maintaining a construction from below – from incremental parsing, recoverability constraints, and genre ecology rather than from social indexing or cohort transmission.

The evidential asymmetry matters: the quotative case rests on decades of variationist data; the *whose* case that follows rests on attestation data. The framework predicts high inter-speaker variation and sensitivity to discourse context, but this remains to be tested. For now, the *whose* case should be read as a *prediction* of the framework rather than a confirmation on par with quotatives.

7.5.1 The filler-gap mechanism

When you hear *Which book did Mary say that John read __?*, you track a displaced constituent: *which book* is the filler, and the gap after *read* is where it's interpreted. This filler-gap dependency spans clause boundaries, requires memory, and operates across surface variation – open interrogatives, relative clauses, clefts.

As mentioned in §6.2, Boguraev et al. (2025) used causal interventions to test whether language models learn a shared mechanism across these construction types. They do: a mechanism learned on embedded open interrogatives transfers to produce filler-gap behaviour in clefts. The result is a proof of concept – showing how mechanistic generalisation could work in principle – rather than direct evidence about human parsing (see also Pistotti et al. 2025 on robust filler-gap generalisation in parasitic gap environments). But the structural dependency projects even when the surface constructions differ, which is what mechanistic kindhood predicts.

7.5.2 Independent relative ⟨whose⟩: a gap that isn't

Hankamer and Postal (1973) claimed that English lacks independent relative *whose* – forms like *a woman whose was straight* where the possessum is elided. Their constructed examples are indeed ungrammatical: **The guy whose you saw banging at the window...* fails. For fifty years, the construction was treated as a paradigm gap – a systematic lacuna in the grammar.

But the gap may be illusory. Reynolds (2024) documents attestations spanning seven centuries, from Middle English manuscripts to contemporary academic prose: *a friend of whose had told us of the accident*; *those whose are not*. The construction is vanishingly rare – on the order of once per hundred million words in large corpora – but attested in edited sources, particularly academic and journalistic prose.[1] The point isn't that the earlier judgment was irrational; the point is that a mechanism-first view predicts that apparent gaps can be ecological artefacts of licensing contexts.

What maintains this construction despite its extreme rarity? The same mechanisms that maintain other filler-gap constructions – but operating under unusually stringent conditions.

[1] The grammaticality judgments remain disputed. Reynolds (2024) is a preprint, and native-speaker intuitions vary. The epistemic status is accordingly tentative: suggestive evidence that the mechanism is there, not conclusive proof.

7.5.3 Filler-gap stabilizers

Filler-gap processing. Independent relative *whose* is a filler-gap construction. *Whose* fills a gap in subject or complement position: *a woman whose __ was straight*. The displaced constituent must be tracked across clause structure and resolved at the gap site. This is the same cognitive operation that underlies relative clauses, open interrogatives, and clefts – the mechanism that Boguraev et al. (2025) showed transfers across construction types. Here, as with quotatives, mechanisms at different scales – cognitive, interactional, institutional – interact to maintain the construction.

The filler-gap mechanism is maintained by processing pressures. Incremental parsing requires tracking dependencies as they unfold; memory constraints favour local resolution; prediction mechanisms anticipate gap sites. These pressures apply every time a speaker processes a relative clause or an interrogative. The mechanism is exercised constantly, even if independent *whose* is exercised rarely.

Information structure. But filler-gap processing alone isn't sufficient. Independent *whose* faces an additional burden: the elided possessum must be recoverable. This is what Reynolds (2024) calls the "double anaphora" requirement – the hearer must simultaneously recover the possessor (from the relative clause antecedent) and the possessum (from discourse context).

Three information-structural configurations license this recovery. Under *contrastive parallelism*, the possessum is established in a preceding clause and the *whose* clause contrasts with it: *I knew someone whose greatest love affair was with objects, another whose was with books, and a third whose was with ideas.* Under *deictic anchoring*, a demonstrative directly points to the possessum: *The man whose these are hath gotten me with child.* Under *structural integration*, the possessum appears as head of the phrase containing *whose*: *a friend of whose had told us of the accident.*

These aren't arbitrary conditions. They're the configurations that satisfy the recoverability constraint – the same constraint that governs ellipsis generally. The information-structural mechanism is independent of the filler-gap mechanism, but both must be satisfied for the construction to succeed – only when both mechanisms clear their thresholds does the construction clear the grammaticality threshold. The pattern generalises: Winckel et al. (2025) show that extraction acceptability in clefts and relative clauses tracks information structure (specifically, Focus-Background Conflict), not just syntactic configuration. Extraction from subjects is harder in clefts than in relatives because subjects are more backgrounded – and backgroundedness conflicts with the focus status that extraction confers. The constraint is gradient, construction-sensitive, and cross-linguistically robust (English and French show the same pattern). This is exactly what a mechanism-first view predicts: the label "island" masks a constellation of interacting pressures rather than naming a single syntactic primitive.

Discourse ecology. Why did identity-of-sense stranding (where the elided possessum shares only sense with an antecedent) emerge only in the 1990s, when identity-of-reference stranding (where it shares reference) has existed since Middle English? Plausibly not because the grammar changed, but because the discourse ecology changed.

Academic and journalistic prose favours list-like enumerations and explicit contrast sets: *patients whose symptoms improved versus those whose didn't*. These registers build the contrastive parallelism that licenses possessum recovery. The construction plausibly became more available when the discourse patterns that license it became conventionalised in particular genres. The mechanism was always there; the licensing environment wasn't.

A brief stress test: if filler-gap processing were absent, independent *whose* would be impossible – but so would all relative clauses, interrogatives, and clefts. If information-structural licensing were absent, the construction would fail even in contrastive contexts. The construction's rarity isn't mechanism weakness; it's licensing scarcity. Both cognitive mechanism and pragmatic licensing are necessary; neither alone suffices.

The full mechanism package for independent *whose* thus includes: filler-gap processing (shared with relatives, interrogatives, clefts), information-structural licensing (specific to ellipsis constructions), and discourse ecology (genre-specific patterns that provide licensing contexts). Unlike quotatives, where social indexing plays a major role, whose-stranding is maintained primarily by cognitive and discourse-structural mechanisms, with little obvious social indexing. What keeps it available is the exercise of filler-gap processing in everyday language use plus the periodic occurrence of licensing contexts in particular registers.

I treat independent relative *whose* as one English instantiation of a broader independent possessive-relative profile, so cross-linguistic parallels are predictions about the mechanism package rather than claims of identical surface syntax.

Cross-linguistic convergence. German and Japanese – despite radically different syntactic structures – appear to show similar information-structural constraints on independent possessive relatives:

> German: *Meins funktionierte, aber ich kenne jemanden, dessen nicht funktionierte.* ('Mine was working, but I know someone whose wasn't.') – licensed by contrastive parallelism.
> German: **Die Person, dessen du vergessen hast, ist mein Cousin.* ('The person whose you forgot is my cousin.') – fails without par-

allelism.

Japanese: *Watashi-no-wa ugoite-ita ga, ugoite-ina-katta hito-mo shitte-iru.* ('Mine was working, but I also know someone whose wasn't.') – licensed by contrastive parallelism.[2]

The pattern may mirror English, though systematic study is needed: stranding succeeds under contrastive parallelism and fails without it. If the parallel holds, the constraints are not language-specific accidents but consequences of how filler-gap processing interacts with discourse accessibility.

The surface structures differ – German has rich case marking, Japanese is head-final, English has relatively impoverished case and is head-initial – but the underlying mechanisms (incremental parsing, memory constraints, communicative pressure for recoverability) are constants, and the constraints on independent possessive relatives track those constants, not the surface variation.

The contrast with quotatives sharpens the framework's explanatory structure. For quotatives, social indexing and processing economy operate at comparable strengths, mutually reinforcing each other across cohorts and registers. For independent *whose*, cognitive mechanisms dominate: filler-gap processing does the heavy lifting, with information-structural licensing providing the gate. The mechanism package is simpler but the licensing ecology is narrower – which explains why quotatives spread rapidly while whose-stranding remains a niche construction.

The *whose* case illustrates the chapter's central argument: a shared cognitive mechanism (filler-gap processing) operates under a specific constraint profile (information-structural licensing), and that profile recurs across typologically unrelated languages because the constraints track universal processing pressures, not language-specific syntax. The maintenance is reciprocal: filler-gap constructions exercise the processing mechanism that makes them possible, while the mechanism makes the constructions processable in turn. "Labels aren't mechanisms" means exactly this in practice: the label "ungrammatical" predicted nothing; the mechanism story predicts exactly when the construction succeeds and when it fails.

We should also expect the constraints to pattern with experience, because the construction depends on licensing conditions that are themselves unevenly distributed. Speakers with greater exposure to the contrastive-parallelism registers – academic prose, formal journalism – may find independent *whose* more

[2] The Japanese example is the author's judgment, confirmed with native-speaker consultants. Systematic corpus study of Japanese possessive stranding remains to be done.

accessible than speakers without such exposure. Differential mastery is not a defect of the grammar; it's a prediction of the maintenance view.

The filler-gap case shows the same reciprocal structure as quotatives. Every relative clause, every interrogative, every cleft is input to the conditions that sustain filler-gap processing. The mechanism is real because it's kept alive by the constructions it serves.

7.6 How to test whether a mechanism is real

Mechanism claims need teeth. Not hand-waving that "entrenchment maintains the category", but operational tests that distinguish genuine causal structure from convenient labels. The tests below are ordered by directness of causal leverage – from interventions that most directly force reality-commitments to distributional patterns that support but don't compel them. The ordering is heuristic, not sacrosanct: some tests collapse into each other depending on the target phenomenon, and different research traditions weight the tests differently.

1. Intervention and ablation. The strongest evidence for a mechanism is that disrupting it changes the outcome. In linguistics, this means: perturbing input distributions (acquisition perturbation), shifting register ecologies (contact situations, code-switching), or computationally lesioning (selectively disabling) a processing component. Boguraev et al. (2025) did this for filler-gap dependencies; the result is model-analytic evidence for mechanistic plausibility, not direct proof that humans instantiate the same component, but suggestive nonetheless. If the clustering reshapes in predictable ways when the hypothesised stabilizer is weakened, we have strong grounds to treat the mechanism as real. *What could mislead us:* interventions may have unintended side effects; apparent reshaping may reflect compensation by other mechanisms rather than loss of the target stabilizer. Section 8.2.2 addresses how to distinguish genuine mechanism loss from compensatory masking.

2. Cross-construction transfer. If a mechanism is real, competence in one construction should predict performance in another that shares the mechanism. As §6.1 showed, the Polish aspect models demonstrate this: training on one verb subset, testing on another, predictions transfer. Boguraev et al. (2025) showed it for filler-gap: a mechanism learned on interrogatives transfers to clefts. Transfer is evidence that category boundaries track causal structure, not just filing conventions. *What could mislead us:* transfer may reflect surface similarity rather than shared mechanism; conversely, lack of transfer may reflect task demands rather than mechanism absence.

3. Predictive constraints under licensing conditions. If a construction depends on specific licensing conditions, success should spike when those condi-

tions are met and fail when they're not. The *whose* case exemplifies this (§7.5): stranding succeeds under contrastive parallelism and fails without it. If we can predict exactly when the construction will be licensed, the mechanism story is doing real work. *What could mislead us:* licensing conditions may be more complex than hypothesised; apparent success may reflect memorised exemplars rather than productive use of the mechanism.

But this raises a base-rate problem: what if the licensing contexts themselves are rare?

4. Base-rate check. Rarity is not disconfirming if the licensing ecology is rare. As §7.5 showed, *whose*-stranding appears on the order of once per hundred million words – but the information-structural configurations that license it are themselves rare in most corpora. The positive test: does the construction appear reliably when licensing conditions are met, even if those conditions are rare? Rarity becomes disconfirming only when predicted licensing contexts are common yet the construction still doesn't appear. This test requires distinguishing *mechanism absence* from *licensing scarcity* – a distinction the *whose* case makes vivid. *What could mislead us:* corpus sampling may miss the relevant registers; rarity may reflect stigmatisation rather than mechanism weakness.

5. Reciprocal maintenance signatures. If mechanisms and categories co-construct each other, we should observe bidirectional effects: categories shaping what gets entrenched, entrenchment shaping what categories survive. The Toronto panel data on quotatives show this: apparent-time change in one direction (younger speakers use *be like* more) co-occurs with cohort effects in the other (speakers carry their adolescent inventory into adulthood). The signature is mutual shaping, not unidirectional causation. *What could mislead us:* correlation may mimic reciprocity; what looks like bidirectional causation may be two effects of a common cause.

6. Distributional and attestation coherence. The weakest test, but still probative: if a category is maintained by stable mechanisms, the distributional cues for category membership should be mutually predictive across registers, time periods, and populations. Cue bundles that fragment under replication are evidence against mechanism; cue bundles that remain coherent are evidence for it. *What could mislead us:* distributional coherence may reflect historical accident rather than ongoing maintenance; attestation may be artefact of corpus composition.

No single test is decisive. Mechanism claims earn credibility when two or more tests converge – when intervention effects, transfer patterns, and licensing predictions all point to the same underlying causal structure. Single-test reasoning invites overclaim; triangulation constrains it.

In practice, these tests translate into specific research designs: psycholinguistic tasks (reaction time, eye-tracking, ERPs) for processing facilitation; sampling across register ecologies for distributional coherence; controlled acceptability work designed to manipulate licensing contexts; panel and real-time designs for reciprocal maintenance; historical corpus comparison for diachronic change; computational causal interventions for mechanism isolation. The tests are not philosophical abstractions; they are guides to experimental and corpus-analytic practice.

A practical recipe, in prose: (1) Identify the cluster – what properties co-occur? (2) Propose stabilizers at two or three levels – cognitive, social, and transmission-level. (3) Derive at least one intervention or natural-experiment prediction – if we weaken this stabilizer, how should the clustering change? – and at least one disconfirmatory prediction – a context where the stabilizer should *not* help. (4) Check whether the predicted reshaping is observed – in acquisition data, contact situations, or experimental manipulations. (5) Triangulate – does independent evidence from transfer, licensing, or distribution converge on the same mechanism? If the predictions hold across tests, the stabilizer is load-bearing. If they fail, revise or discard.

These tests certify mechanisms – but they can also mislead. When tests diverge – when intervention effects point one way but transfer patterns point another – treat the mechanism as underspecified or mislocated in scale. Chapter 8 diagnoses false positives, compensation effects, and ecological confounds that create the appearance of mechanism where none exists, or mask genuine mechanisms that the tests fail to detect.

7.7 Degrees of projectibility

The stabilizing story explains why projectibility comes in degrees.

A category deep in a single basin – that is, supported by multiple aligned stabilizers (entrenchment, transmission, functional pressure, social reinforcement), all pulling in the same direction – is strongly projectible. You can learn about nouns from a few exemplars and generalise reliably to new nouns. The mechanisms reinforce each other across timescales.

A category in an overlap region – where mechanisms pull in different directions, or where entrenchment is weak – is weakly projectible. Predictions work for typical cases but fail for edge cases. The degree of projectibility tracks the degree of mechanistic support.

A label with no mechanisms behind it is not projectible at all. A wastebasket class defined by what it's not should fragment under the robustness tests: predictions trained on one subtype don't generalise to others.

This is the operational content of "mechanism-maintained kinds". Not a metaphor. A measurable property: how strongly do predictions transfer across novel instances, contexts, and populations?

7.8 What this commits us to

This framework commits us to a specific kind of realism plus a ranked set of theoretical consequences. Both warrant explicit statement.

The overarching commitment is mechanism-grounded realism. I'm not claiming that linguistic categories are mind-independent features of the universe, waiting to be discovered like chemical elements. I'm claiming that they are stably discoverable because mechanisms make them reliable targets of inquiry. The category NOUN is real in the same sense that the category MACROPHAGE is real: not because there is an essence that defines it, but because mechanisms of production and transmission – genetic, developmental, and functional for macrophages; cognitive, social, and transmission-level for nouns – keep properties clustering in ways that support induction. The parallel is not ontic identity but a shared route to projectibility: in both cases, tracking the category licenses predictions because mechanisms hold the cluster together. This realism is compatible with the categories changing over time, varying across languages, and having fuzzy boundaries. What it is not compatible with is treating categories as arbitrary conventions that could have been otherwise with no cost to explanation.

Not all the following commitments are equally central. Three are load-bearing: give them up and the framework collapses. Four are derived: they follow naturally from the core but could in principle be revised without abandoning the project. What makes a commitment *core* is that denying it would sever the link between mechanisms and category stability. What makes a commitment *derived* is that it could be refined by new evidence without forcing a revision of the core.

Three commitments are core.

1. Process ontology. Categories are not static objects. They're dynamically sustained patterns – standing waves, not carved statues. What exists is the stabilizing process; the category is what the process makes legible. The quotative cluster is sustained by ongoing usage and transmission rather than simply "being there"; remove the maintaining processes and the cluster dissolves. This is non-negotiable: without it, we risk sliding back toward essences or conventions.

2. Interventionist realism. Kinds are real to the extent that tracking or manipulating them changes expectations. This is stronger than description: it says that category distinctions track causal structure, not just impose order on data. "Intervention" here means disrupting input patterns, shifting register ecology, testing acquisition trajectories, or modelling counterfactual distributions – as when Boguraev et al. (2025) lesioned a filler-gap mechanism and observed downstream effects. Give this up and the framework loses its empirical teeth.

3. Reciprocal maintenance. Mechanisms and categories co-construct each other. Categories shape what gets entrenched; entrenchment shapes what categories survive. The quotatives maintain the processing and social-indexing mechanisms by providing input; the mechanisms maintain the quotatives by making them easy to produce and socially salient. This reciprocity is the engine of the framework – it explains both stability and directional change. Remove it and the account becomes one-directional and static.

Four further commitments are derived from the core.

4. Population-distributed competence. The target of explanation is not an ideal speaker's uniform grammar but a population-distributed competence profile. Different speakers, with different input histories, stabilize categories to different depths – as we saw with differential access to *whose* across register ecologies. This follows from process ontology (different processes yield different outcomes) but someone could accept the core framework while retaining idealized-speaker models for some purposes.

5. Variation as signal. Differences across contexts, speakers, and registers are not noise. They're diagnostic – evidence about which region of the state space a token occupies, which activation state is active. This follows from interventionist realism (if tracking matters, variation tracks something) but the specific diagnostics remain open to empirical refinement.

6. Cross-level coherence. A category theory should deliver compatible predictions whether we analyse a phenomenon at the level of subpatterns, the construction, the category, or the wider system. This is a methodological constraint implied by reciprocal maintenance. Violations might indicate interesting heterogeneity rather than framework failure – though distinguishing productive heterogeneity from genuine failure requires independent criteria.

7. Operational testability. These claims have operational teeth. Entropy reduction, cross-context generalisation accuracy, inter-speaker agreement as a function of frequency and entrenchment – these are illustrative operationalizations of what "stable category" means in practice. Section 6.6.2 develops the generalisation-accuracy measures; the others remain programmatic but not promissory. This operationalizes the core commitments, but the specific metrics can evolve as measurement improves.

What the framework does *not* commit us to is also worth stating. The framework is silent on acquisition theory: it is compatible with nativist, constructivist, or hybrid accounts of how the underlying mechanisms are implemented. It is silent on neural architecture: the mechanisms could be realized in distributed representations, localist circuits, or something else entirely. And it is silent on the question of whether grammatical categories are "innate" in any interesting sense – what matters is whether they are stably maintained, not whether they were pre-specified. These are open empirical questions, not framework commitments.

These commitments explain why the framework is an ontological reorganization of existing work rather than a bid to displace it.

7.9 Refactoring, not replacing

This chapter doesn't compete with existing explanations; it reorganizes them around what they already presuppose but rarely foreground: stabilizing mechanisms. That's the pluralist posture from §4.5 applied to a specific case.

The descriptive, variationist, and functional work that this chapter draws on – Ferrara and Bell on quotative function, Bybee on entrenchment, Tagliamonte on apparent-time change, Buchstaller on cross-variety patterns – provides the data and often identifies the mechanisms.[3] What the maintenance view adds is not more data but a different question: not *what* properties cluster, but *why* they hold together and persist.

For quotatives, this means reinterpreting frequency, social indexing, and processing alignment as mutually sustaining rather than as independent explanations. The maintenance view explains why these mechanisms cluster rather than scattering, why the cluster is stable across generations rather than dissolving into noise.

[3]The list could extend: Huddleston and Pullum's distributional analysis, Labov's foundational variationism, the typological work on evidentiality that grounds the cross-linguistic parallels. The point is the *kind* of work, not an exhaustive enumeration.

The sociolinguists already had the patterns. The cognitivists already had the mechanisms. The typologists already had the parallels. What was missing was not more data but a framework for why the data hold together – an account of what linguistic categories actually *are*. That's what the maintenance view offers. Not replacement. Refactoring. Same codebase, better architecture.

For readers familiar with grammaticalization theory: yes, diachrony remains essential. Grammaticalization tells you *how* a form like *be like* developed – how meanings bleach, how forms reduce, how pragmatic inferences harden into grammatical meaning. The maintenance view answers a different question: what keeps the resulting cluster coherent *now*? The pathways are historical; the stability is mechanistic. Both are needed. Diachrony without synchronic maintenance leaves present-day coherence unexplained; maintenance without diachrony leaves the trajectory unexplained.

For readers familiar with prototype theory: yes, gradedness is real. Categories do have cores and fringes, and typicality judgments are graded. But the interesting explanation isn't *that* they're graded; it's *why* the cores tighten and the fringes wander across different input histories. A mechanism-first account isn't relabelling gradience – it predicts where stability should be robust (deep in the basin, multiple mechanisms aligned) and where it should fray (near boundaries, mechanisms pulling in different directions). Prototype theory describes the shape; the maintenance view explains why the shape is stable and why it varies where it does.

7.10 The most telling facts

To recap: this chapter traced the stabilizing mechanisms for two deliberately different cases. Quotatives showed social indexing and processing economy co-reinforcing through transmission and acquisition, with robust variationist evidence for English and well-supported parallels in German. Independent relative *whose* showed cognitive mechanisms (filler-gap processing) operating under stringent information-structural licensing, with the framework predicting where the construction should succeed and fail. Both cases illustrated multi-scale stabilization, reciprocal maintenance between categories and mechanisms, and projectibility that comes in degrees. The robustness tests (§7.6) give operational teeth to mechanism claims; the theoretical commitments (§7.8) spell out what the framework is and isn't claiming. What remains is the question of failure.

A linguistic category is not a thing you find; it's a regime you maintain. Arguments over definitions are, at bottom, arguments over stabilizers.

And the most telling facts about categories live in their failure modes – where boundaries blur, where judgments diverge, where the stabilizing dynamics show their seams.

But if categories are maintained, they can also fail to be maintained – or be maintained too loosely, or in ways that don't project.

The next chapter asks: how do we know when we don't have a kind – when what we're tracking is a label or a short-lived fashion rather than a mechanism-maintained cluster?

Definiteness thread. This chapter's mechanism inventory predicts that definiteness should show multi-scale stabilization: acquisition establishes the distributional frame (deitality); alignment repairs discourse failures (definiteness-as-identifiability); transmission filters out unsystematic variants. Chapter 10 will show how these mechanisms can pull different properties, creating two coupled HPCs rather than one. The prediction would be falsified if the semantic and formal properties showed no mechanism-specific sensitivities – if they behaved as a single bundle under all perturbation types.

Failure modes

The danger of a good framework is that you start seeing it everywhere.

Once you learn to spot homeostatic property clusters – once you see how feedback loops can maintain stable patterns without essences – it becomes tempting to diagnose homeostasis in every corner of the grammar. NOUN? HPC. SUBJECT? HPC. VOICE? HPC. NON-FINITE CLAUSE? Surely an HPC too. The world fills up with spinning tops. Soon even the paper clips look homeostatic.

Words that won't hold still – that's one failure mode. Words that *seem* to hold still only because we gave them a label – that's another. This chapter is about telling them apart.

This is the INFLATION PROBLEM. If the criteria for being a mechanism-maintained kind are too loose, the framework explains nothing because it excludes nothing. If every stable pattern counts as an HPC, then "HPC" just means "pattern we have a name for." That's a large portion of the index.

But not every top is the same. Some wobble for a moment and fall; some were never really spinning – just labels we applied to patterns that happened to be there. Section 4.7 previewed three ways a putative category can miss: thin, fat, or negative. This chapter develops the diagnostics. The answer turns out to be that the ways a category can miss tell us something about what it takes to be a genuine kind.

8.1 The inflation problem

Why are some categories so tempting to reify? Why does linguistics – and cognition generally – generate labels that feel like kinds but turn out to be something else?

Start with a familiar example. What makes something a *chair*? In terms of physical properties – materials, shape, construction – a bean bag, a papasan, a recliner, and a three-legged Frank Lloyd Wright design have almost nothing in common. But from the perspective of someone who wants to sit down, they're all the same. The category *chair* groups objects by what they do for us, not by what they are.

This is a useful category. If you know something is a chair, you know you can sit on it. But is it a *natural kind*? Are there mechanisms maintaining "chairness" as a stable cluster of properties in the objects themselves? No – individual bean bags and recliners are manufactured by independent causal chains. The category tracks *function*, not underlying structure – and this isn't the "functional pressure" that §7.2 identifies as a stabilizer; that applies when function shapes what gets transmitted, not when function is merely the criterion by which a user groups pre-existing objects. (Word-kinds are different: *chair* as a lexeme *is* maintained by mechanisms – entrenchment, alignment, transmission – but the category of objects it refers to is not.)

The same logic applies to *dessert*, *trash*, *weapon*—and, for those who have followed the debate, to whether a hot dog is a sandwich. These categories are useful for making decisions. Knowing something is a weapon tells you to be cautious. But the usefulness comes from how they fit into our activities, not from shared mechanisms that maintain them as kinds. The question assumes definitional structure where none exists. (Individual sub-categories – *gun*, *sword*, *croissant* – may well have their own transmission lineages; the claim is that the umbrella grouping *weapon* or *dessert* lacks a unifying mechanism across its heterogeneous members.) Khalidi calls such categories EPISTEMIC KINDS: they serve our epistemic purposes – classification, description, prediction – but don't necessarily track causal structure in the world (Khalidi, 2013, pp. 43, 65).

Now translate to linguistics. The analyst's job is to describe language. Categories earn their keep by making description tractable. If grouping manner adverbs, degree words, and sentence adverbs under a single label ADVERB simplifies the grammar, the label is useful – even if no mechanism unites those items.

The inflation problem arises when we mistake this usefulness for reality. We see that ADVERB is useful – it tells us "this modifies something other than a noun" – and we assume there must be a deep, unified mechanism maintaining all adverbs as a natural kind. We confuse the utility of the label with the reality of the cluster. The result is a "bureaucracy of exceptions": an expanding apparatus of special rules, sub-cases, and waivers required to maintain the fiction that the category is unified. The paperwork is immaculate; the mechanism isn't. (The philosopher Cailin O'Connor has formalised this insight using evolutionary game theory, showing how categories optimised for coordination can diverge from categories that track real structure; see O'Connor 2019.)

The inflation problem is compounded by *material reinforcement*. Some labels don't just circulate in heads; they circulate in curricula, exams, style guides, tagsets, and parsers. Once a category gets its own exam question, it starts to feel eternal. When a category is institutionally scaffolded – taught in schools, encoded in annotation standards, rewarded by grading rubrics – it can look stable and projectible even if the underlying mechanisms are weak or absent. Stability is not enough. A category with no real mechanistic basis can persist for centuries if institutions reward it.

Here's the trap: categories can be useful, learnable, stable across generations – and still fail to be natural kinds. The question this chapter addresses is how to tell the difference: when does a label pick out a genuine HPC, and when is it merely a convenient class?

A word about framing: diagnosing something as a class rather than a category is not a demotion. Both serve purposes. Classes organize curricula, structure textbooks, enable communication. Categories support induction, predict deficits, explain acquisition. The question is not 'Is this term legitimate?' but 'What kind of work are we asking it to do?'

8.2 The two diagnostics

To answer that question, we need criteria. Chapter 6 argued that projectibility is the epistemic payoff of genuine kinds: a category earns its keep by supporting induction. Chapter 7 argued that homeostasis is the ontological ground: a category is real because mechanisms maintain it. These two faces of the HPC definition give us two diagnostics. To warrant the claim that a linguistic category is an HPC kind, it must pass both.

8.2.1 The projectibility diagnostic

Can we predict new data from old?

If a category is a genuine kind, learning its properties from a finite sample should allow you to predict unobserved instances. This is projectibility: the pattern travels. In §6.1's terms, it's the *good bet* – you stake what you've learned and the bet pays off.

The test is straightforward. Train a model – or a learner – on one dataset: a corpus, a time period, a set of languages. Test on a held-out set. If performance significantly exceeds baseline (e.g., shuffled labels), the bet was good; the pattern is projectible. If the model overfits or collapses on new data, the bet was bad; the pattern is local, accidental, or illusory. Shuffled labels are the reality check for category speculation.

Polish aspect offers the worked example (§6.1). The textbook semantic definitions – boundedness, totality – project imperfective well (98%) but perfective poorly (77%). The definitional bet fails for half the system. By contrast, the lemma-concrete model – which learns cue–outcome associations without representing aspect as a category – projects better across the board. Good bets track mechanisms, not labels.

A confound lurks here. Projectibility across corpora can be an artefact of shared conventions – genre pipelines, annotation guidelines, register selection, editorial norms – rather than evidence of genuine mechanistic clustering. Academic register is a useful reminder: it can be highly projectible, and yet much of what holds it together is explicitly policed. A category that predicts well across the British National Corpus, the Corpus of Contemporary American English, and the Penn Treebank may do so because all three were annotated by linguists trained in the same tradition, not because the category tracks stable structure in ordinary speech. This is why the two-diagnostic test is necessary: the homeostasis diagnostic (perturbation sensitivity) guards against convention-driven projectibility.

8.2.2 The homeostasis diagnostic

Can we name the stabilizers?

If a category is a genuine kind, its stability must be causal, not accidental. We should be able to identify the specific feedback loops maintaining it: acquisition, entrenchment, alignment, transmission, functional pressure.

The test asks: can we trace the causal chain? Can we predict what would happen if a stabilizer were removed? A genuine HPC should exhibit perturbation sensitivity: weaken a mechanism, and the cluster should fray. A mere label should be perturbation-inert: remove it, and nothing changes in production or acquisition. This operationalizes the homeostasis criterion in ways philosophers have demanded: Craver (2009) argued that HPC theory is vague about what counts as a mechanism and where one mechanism ends and another begins. The perturbation test provides a concrete answer: we recognize a mechanism by countering it – whatever process, when weakened, causes the cluster to fray. This is a diagnostic criterion, not a definition; it tells us how to find mechanisms, not what they essentially are. The META-OCCAM principle also acts as a filter: if a proposed kind requires a bespoke stabilizer – a mechanism that applies only to it, with no parallel elsewhere – that's evidence of explanatory gerrymandering, not discovery.

A methodological caution: intuitions are diagnostic traces, not transparent reports. Scott-Phillips (forthcoming-b) argues that acceptability judgments are byproducts of relevance-seeking cognition – they track whether a sentence *could* have optimal relevance, not whether it violates a grammatical essence. When we judge **I don't want going to the cinema* as unacceptable, we're not detecting a rule violation; we're detecting that the sentence incurs processing costs without compensating communicative benefits. This matters for the inflation problem: just because speakers agree that something is "wrong" doesn't mean we've identified a natural kind. The intuition may track a failure of communicative fit, not a failure of mechanism-maintained structure. The homeostasis diagnostic asks for more: not just "do speakers react?" but "does the reaction track a perturbable mechanism?"

This diagnostic catches categories that are projectible by accident. A local correlation might predict well within a corpus but lack any causal grounding. The homeostasis diagnostic asks: is there a feedback loop, or just a coincidence?

Heritage language attrition provides a natural experiment. When transmission mechanisms weaken – when children receive reduced input because their parents speak the heritage language less, or when community alignment is diluted by a dominant contact language – category boundaries become variable in predictable ways. Polinsky's work on American Russian shows exactly this pattern: heritage speakers maintain high-frequency, highly entrenched items (core vocabulary, basic constructions) while low-frequency items drift or are replaced by contact-language calques (Polinsky, 2018). The perturbation (reduced transmission) causes the cluster to erode at its margins while the core holds. This is what the homeostasis diagnostic predicts: genuine HPCs show perturbation sensitivity, with the degree of drift proportional to the strength of the perturbed mechanism.

8.2.3 Why both are needed

A grouping can pass one diagnostic and fail the other.

A grouping might be projectible but not homeostatic. Consider a corpus artifact – a distributional pattern that emerges from a particular text genre or register. Within that corpus, the pattern predicts well. But there's no mechanism maintaining it; shift to a different corpus, and it vanishes. The projectibility was real but local; the homeostasis was illusory.

A grouping might be homeostatic but not projectible. Consider a category maintained by a single, weak mechanism that operates inconsistently across speakers. The mechanism is real – we can name it, intervene on it – but it doesn't produce enough clustering to support reliable prediction. The homeostasis is real but insufficient.

Groupings that pass both diagnostics are good candidates for category status. Groupings that fail one or both are suspect. The three failure modes we'll examine each fail in characteristic ways:

- Thin classes miss the homeostasis diagnostic. The mechanisms are absent or too weak to produce stable form–value coupling.
- Fat classes miss the projectibility diagnostic. The label lumps forms with different values – items whose contributions when deployed don't cohere – so learning one subclass doesn't generalise to others.
- Negative classes miss both. They have neither unified form nor unified value – they're defined by absence, with no positive property cluster to maintain or project from.

Table 8.1. Failure modes of putative HPC kinds

Failure mode	Projectibility	Homeostasis	Example
Thin	Fail (insufficient clustering)	Fail (no stabilizers)	Preposition dou
Fat	Fail (subclass divergence)	Partial (subgroup mechanisms)	Umbrella ADVE
Negative	Fail (no positive cluster)	Fail (defined by absence)	NON-FINITE CL

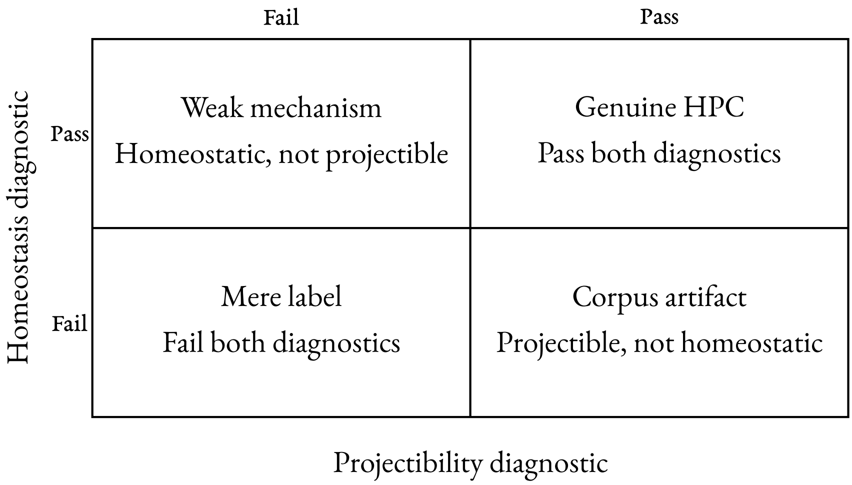


Figure 8.1. The two-diagnostic matrix. Only categories in the upper-right quadrant are genuine HPC kinds. Examples: *Genuine HPC* – count noun, manner adverb; *Weak mechanism* – idiolectal patterns; *Mere label* – NON-FINITE CLAUSE, umbrella ADVERB; *Corpus artifact* – genre-specific patterns.

8.2.4 The grain question

One more trap before the case studies: the problem of grain. Even when we have genuine HPCs, we often group them into larger systems labelled "Grammar" or "Language." Is English Morphosyntax an HPC kind?

The temptation is to say yes. It's stable, it's learned, it's maintained. But this is the Madagascar fallacy. Biological species are the paradigmatic HPC kinds. *Lemur catta* (the ring-tailed lemur) is a homeostatic cluster maintained by interbreeding, shared ecology, and genetic transmission. But consider "the fauna of Madagascar." It is a stable collection of animals. It is distinct from the fauna of Australia. It has a causal history (isolation). But "the fauna of Madagascar" is not itself a species. It is an *ecosystem* of species. There is no single mechanism that maintains "the fauna" as a unit; there are thousands of mechanisms maintaining the individual populations that comprise it. The fauna of Madagascar is not a species; it's an airline route plus a geological history.

Language is an ecosystem. The HPCs are the components – constructions, lexemes, phoneme contrasts, sometimes families or paradigms – not the grammar as a whole. The failure modes below apply at that level: we ask whether ADVERB is a kind, not whether ENGLISH is. Section 8.6 returns to this question; for now, the point is that grain matters. A category can be too thin, too fat, too negative – or simply at the wrong level of analysis.

8.3 Thin clustering: The smoke ring

The first pattern is the class that barely exists.

A THIN class is one where the signal exists – speakers recognize it, linguists name it – but the property cluster is ghostly. There is no robust homeostatic loop maintaining it. It's a "smoke ring": it has structure, but that structure is ephemeral, typically generated by a transient convergence of other mechanisms rather than a dedicated feedback loop of its own.

Thin classes miss the homeostasis diagnostic. We can't name their stabilizers – because there are none, or none sufficient to produce stable clustering.

8.3.1 The nonce word test

Miller (2021, pp. 25–26) provides a useful operationalization. Whether a novel form counts as a "word-kind" depends on whether stabilizing mechanisms – internal (cognitive) and external (social) – have emerged to maintain it as more than a one-off performance. Consider *cromulent*, coined for a 1996 *Simpsons* episode to mean "acceptable, fine."

Does *cromulent* exist as a word? Miller's answer is to treat this as a mechanistic question, not a metaphysical one. Are there internal routines ensuring that speakers reproduce the form reliably (phonology, meaning, inflectional behaviour)? Are there external norms that support convergent use (shared recognition, uptake, correction, expectation)? Within Simpsons-fan communities, the answer is often yes: the form is used with a consistent meaning and a stable pronunciation, and it is recognizable enough to support coordination. Outside those communities, the mechanisms are weaker or absent: the form surfaces as an in-joke, may be misheard or reanalysed, and is rarely transmitted as a general-purpose lexical item.

The point is not that *cromulent* is "real" for some speakers and "fake" for others, but that wordhood is a claim about where stabilizers are doing work. A form can be a bona fide HPC kind in one network and a thin pattern in another. The question is always: what maintains it, and at what scale?

The nonce word test generalises. A class is thin when the stabilizers that would make it a good bet have not (yet) consolidated. This can happen for genuinely new items (nonce coinages), for rare or enclave-restricted items (idiolectal or micro-community vocabulary), or for patterns that persist only as incidental fallout from other, more robust systems.

8.3.2 Case study: Preposition copying and pruning

English offers two options for relativising prepositional complements: pied-piping (*the box in which I put it*) and stranding (*the box which I put it in*). Both are robustly maintained. But at very low frequency, speakers produce outputs that either copy the preposition or omit it:

> Copying: *from which to choose from*, *from where it came from*
> Pruning: *That's the situation that Obama finds himself* (intended: *finds himself in*)

This pattern exists, and Radford et al. (2012) document it in live broadcast speech: roughly 15 copying tokens and 67 pruning tokens in ~300 hours of recording. Such forms often pass without overt repair, so the pattern is reproducible – but it is thin.

The frequency is vanishing – too rare to be plausibly acquired as a *target* construction – and there's no social indexing (copying/pruning signals nothing about identity, register, or stance) or functional niche (either pied-piping or stranding serves the same communicative end without redundancy or omission).

Apply the homeostasis diagnostic: can we name the stabilizers?

- Acquisition: Weak. The token rate is too low for learners to extract a dedicated rule.
- Entrenchment: Weak. No speaker uses it frequently enough for strengthening at the constructional level.
- Social indexing: Absent. There is no community norm that rewards or recognizes it.
- Functional pressure: Absent. Competitors cover the function without any gap.

What generates copying and pruning is not a dedicated maintenance loop but online production under competition. Radford and Felser analyse these as performance-side phenomena: speakers begin constructing one option and drift toward the other midstream. In copying, the two options leave traces of both routes; in pruning, the preposition is lost under the same competitive pressure plus ordinary omission. Either way, the result is a byproduct of mechanisms that robustly maintain pied-piping and stranding, not a target of maintenance in its own right.

Compare to robust stranding: *Who did you talk to?* This form is maintained by every mechanism on the list – frequency is high, it's acquired early, and it participates in register contrasts with pied-piping that create genuine indexical structure. Functional load is substantial. Preposition stranding is an HPC; copying/pruning is its smoke ring – cast off by the same dynamics but lacking the immune system to persist independently.

The perturbation test makes this vivid. If English lost preposition copying tomorrow – if speakers simply stopped producing it – no feedback loop would kick in to restore it. No acquisition pressure, no social correction, no functional gap. It would vanish without a trace. A genuine HPC resists perturbation; a smoke ring dissipates.

Robust HPCs are attractors: the mechanisms pull deviant instances back toward the cluster. Thin patterns are metastable: they persist only so long as the surrounding system continues to generate them as incidental fallout. This is what it means to fail the homeostasis diagnostic. The pattern is real – it can be observed, described, named – but the causal structure that would maintain it is absent. We have a label without a kind.

8.4 Fat clustering: The wastebasket

The second pattern is the opposite of thinness: not too little mechanism, but too many pulling in different directions.

A fat class is a label that lumps distinct causal clusters into a single bin. The analyst treats them as one – usually for "wastebasket" reasons, where the class is defined residually: "everything that isn't X goes here." Every taxonomy has a drawer labelled "misc." The mistake is to infer that "misc." is a lineage. But the mechanisms maintaining the subgroups are radically different. Fat classes miss the projectibility diagnostic: learning one subclass doesn't let you predict others.

Fat classes create an illusion of understanding. We have a label, so we act as if we have a kind. But the label masks the heterogeneity. Seeing adverb on the page, we assume there's something that makes all adverbs adverbs – some shared mechanism or property cluster. When we look for it, we find only the absence of other features.

Fat classes arise when different causal clusters get lumped under the same label for non-causal reasons. The analyst's goal – taxonomic completeness – treats manner adverbs and degree words alike because both are "modifiers that aren't adjectives." But in causal terms, they occupy different regions with different mechanisms.

This is not a peculiarly linguistic affliction. Sullivan (2016) argues that cognitive neuroscience faces the same trap: researchers converge on a construct label – "spatial memory", "working memory" – without converging on the underlying mechanisms. Different labs use different paradigms; what they produce and measure may not be the same phenomenon at all. Sullivan calls this a failure of *construct validity*: the label achieves institutional consensus (everyone uses it) while the causal reality splinters beneath it. Craver's term for what such labels *should* pick out is "mechanistic property clusters" – groups of properties whose co-occurrence is explained by causal mechanisms that regularly ensure they are instantiated together (Craver, 2009). When the coordination fails, you have reliability without validity: a fat label, not a fat kind. The difference from linguistics is that grammatical categories face an immediate forcing function – communicative pressure. If my NOUN doesn't align with yours, utterances misfire and learning stalls. Neuroscientists can publish papers about distinct "spatial memories" without immediate communicative cost; speakers can't. This communicative pressure is one reason to expect that core grammatical categories are more robust than their neuroscientific counterparts – but the fat-class failure mode remains a live possibility wherever the label's institutional momentum outpaces its mechanistic grounding.

8.4.1 Case study: The Adverb

Try this: *She ran quickly.* Now try: **She ran very. Very* has never gone jogging. Both *quickly* and *very* are adverbs – the same part of speech, according to every grammar textbook. Yet one sentence is grammatical and the other isn't. What went wrong?

The answer reveals everything about fat classes. ADVERB doesn't name a natural kind; it names a filing cabinet where grammarians store everything that modifies but isn't an adjective. The items in the drawer have almost nothing in common – and when we try to generalise across them, the predictions collapse.

The traditional category ADVERB is the paradigmatic fat class – what introductory textbooks call the "dustbin of the parts of speech."[1] Look inside and you'll find at least three distinct clusters, each with its own mechanisms.

[1] *CGEL* dramatically improves matters by reassigning many traditional "adverbs" to prepositions and by separating category from function; but even the residual *CGEL* Adverb still groups multiple mechanistically distinct subfamilies (degree/focusing vs derived manner vs stance/connective), so it too is likely decomposable under the two-diagnostic test.

Manner adverbs (*quickly*, *carefully*, *silently*) form an open class, productively derived from adjectives via *-ly* suffixation. They integrate into the VP, favouring VP-final position: *She ran quickly.* What maintains them? Productive derivational morphology (the *-ly* rule applies to novel adjectives), VP-internal syntactic licensing, and early acquisition of the adjective-to-adverb mapping.

Degree words (*very*, *quite*, *rather*) form a closed class – a handful of items, fixed in the lexicon. They modify adjectives and other adverbs, never verbs directly: *very tall*, *quite slowly*, but not **She ran very.* What maintains them? High-frequency entrenchment (each is memorised, not derived), functional specificity (a narrow scalar-modification niche), and closed-class conservatism (no productive rule generates new degree words).

Sentence adverbs (*unfortunately*, *frankly*, *obviously*) scope over propositions, expressing speaker attitude. Prosodically, they're detached – set off by commas, placed sentence-initially. Syntactically, they're peripheral to the VP. What maintains them? Discourse-level alignment (speaker-stance marking), prosodic separation, and pragmatic function distinct from event modification.

Three clusters, three sets of mechanisms. The label ADVERB lumps them only because they share a negative property: they're modifiers that aren't adjectives. That's why *quickly* and *very* can't swap positions – they live in different rooms of the filing cabinet, maintained by different causal processes.

8.4.2 The projectibility failure

Now apply the projectibility diagnostic. Train on manner adverbs; test on degree words.

Ernst (2002, p. 45) demonstrates that predicational adverbs show a rigid ordering hierarchy – discourse-oriented > evaluative > modal > evidential > subject-oriented > manner – and that only manner adverbs may appear to the right of the verb. These aren't stylistic preferences; they're grammatical constraints. Learning the distribution of one class tells you nothing about another.

Learning the distribution of *quickly* – that it can appear sentence-finally (*She ran quickly*) – generates a prediction for other "adverbs." But the prediction fails catastrophically for degree words: **It was good very.* And for modal adverbs: **She left probably.* The pattern doesn't travel.

Conversely, learning that *very* requires an adjective host – *very tall* – generates a prediction that fails for sentence adverbs: *Unfortunately, she left* has no adjective host at all.

The projectibility test delivers a clear negative: the category ADVERB doesn't support cross-subclass generalisation, and what you learn about one subtype tells you nothing reliable about another. This diagnostic failure – the label grouping items that were never maintained by shared mechanisms – is the signature of a fat class.

Fat classes arise because exhaustive taxonomy requires a home for everything. If you need every word to belong to exactly one part of speech, and you've already defined noun, verb, adjective, preposition, determiner, and conjunction, then "everything left over" must go somewhere. ADVERB is that somewhere. The label groups items that are "close" only in the analyst's filing system. In terms of the mechanisms that maintain them, manner adverbs and degree words are far apart. The label is a convenience, not a kind.

The maintenance view gives clear guidance when you diagnose a fat class: decompose. ADVERB is not an HPC. But the genuine clusters lurking inside it – the cluster around *quickly*, the cluster around *very*, the cluster around *thus*, each maintained by its own braid of mechanisms – may be kinds where the umbrella is not. The lump is a taxonomic fiction; the components are causally real. This doesn't mean abolishing the term – ADVERB remains useful for some purposes – but it means recognizing that the term names an epistemic kind, not a natural kind. When you're doing HPC analysis, look below the lump.

A note of sympathy for the original grammarians. They weren't being lazy. Imagine standing in front of the scattered words that aren't nouns or verbs, facing the pedagogical demand to say what they are. Students need labels. Textbooks need categories. The pressure to provide *some* answer – any answer – is real. Reaching for ADVERB was not a failure of intelligence but a response to genuine classificatory pressure. The maintenance view doesn't condemn that move; it just asks what we can learn from it. The answer is that exhaustive taxonomy and natural-kind discovery are not the same enterprise, and sometimes they pull in different directions.

8.4.3 Return to Huddleston

And this is where we can finally return to Huddleston's puzzle from §1. His 3 AM email about *otherwise* – "Its classification is quite a puzzle" – wasn't a failure of scholarship; it was a field report from inside the wastebasket – a report that the usual diagnostics don't converge.

Start with the uses that look easy. In *think otherwise*, *otherwise* sits where we expect adverbs to sit, and it has the familiar meaning 'in another way / to a different effect'. But even here, the comfort is partly illusory. English *think* readily takes content-like complements in reduced form (*think that*, *think this*, *think so*, *think not*), and *otherwise* is at least a plausible member of that family: not merely a manner modifier, but a compact alternative to 'that things are otherwise'. In other words, the first use is already perched on the boundary between modifier and complement.

Now consider the predicative case: *the truth is quite otherwise*. Whatever category *otherwise* belongs to, it is functioning as predicative complement here. The stress point is not the function but the category fit. Degree modification by *quite* and the idiomatic force 'contrary to what is assumed' pull it toward adjective-like behaviour; its kinship with *otherwise* in *think otherwise* pulls it back toward the modifier inventory. The diagnostics cross-cut.

The two examples Huddleston explicitly flagged make the conflict sharp. In *this suggests otherwise*, *otherwise* behaves like a content complement: the construction invites something proposition-sized (*this suggests that…*), and *otherwise* appears to do that job in a reduced, anaphoric way. And in *the correctness or otherwise of the proposal*, *otherwise* is recruited into a coordination that functions inside a nominal: it behaves like the conventional way English expresses 'whether X holds or the contrary holds'. Neither environment is a natural habitat for an ordinary VP adjunct, yet both are robust, non-marginal uses.

This is the core of Rodney's dilemma. It is not that *otherwise* "does three things" and we need three labels. It is that the same lexical item is pulled by different, individually well-motivated diagnostics toward different category analyses, and no single traditional label makes all the pulls line up without residue.

That is exactly why adverb is so often the analyst's refuge. When the diagnostics for the established categories converge, we have no puzzle. When they don't, we reach for the category with the weakest positive profile: the one defined largely by exclusion, the one designed to mop up words that are neither nominal, verbal, adjectival, nor prepositional in their most typical uses. Adverb is, in this sense, the grammar's "default box" for items that refuse a clean fit.

Seen this way, *otherwise* is not a quirky edge case that embarrasses an otherwise sound taxonomy. It is a demonstration case. It shows, in miniature, why the inflation problem in this book is not a philosopher's worry but a working grammarian's daily experience: sometimes the pressure to assign a single category forces a label to do work that no label of that grain can do. Huddleston's "quite a puzzle" is the sound a wastebasket category makes when you ask it to behave like a kind.

8.5 Negative classes: The complement class

The third pattern is the most categorical: the category defined by what it is not.

A negative class is a complement class: "everything that isn't X." In set theory, complements are well-behaved – the complement of a set is precisely defined. In causal mechanism terms, they're incoherent. There is no such thing as a "homeostatic property cluster of absence." The absence of a property is typically not a property, and certainly not one that mechanisms can stabilize.

Negative classes miss both diagnostics. They miss projectibility because members share no positive properties to generalise from. They miss homeostasis because no mechanism can maintain "lacking X" – what would such a mechanism even look like?

8.5.1 Case study: The non-finite clause

Consider the category non-finite clause. In traditional grammar – and in most contemporary frameworks – it groups clauses headed by:

- Infinitivals: *to eat*, *to run*
- Plain forms after modals: *eat*, *run* (in *can eat*, *will run*)
- Past participles: *eaten*, *run*
- Gerund-participles: *eating*, *running*

What do these clause types share? The standard answer is negative: their heads lack tense inflection, they typically lack overt subjects, they can't stand alone as main clauses (mostly). They're grouped by absence. This is a family portrait taken in silhouette: the shared trait is what you can't see.

But look at their positive properties – their syntactic environments, their maintaining mechanisms – and they're radically different animals.

Infinitival clauses appear in control contexts (*I want to leave*), as purposive adjuncts (*I came to help*), as subjects (*To err is human*). What maintains them? Control verb selection (verbs like *want*, *try*, *hope* select infinitival complements), purposive adverbial licensing, and historical entrenchment of the *to*-infinitive construction.

Clauses headed by plain forms after modals (*can eat*, *will run*) behave quite differently from infinitivals despite superficial similarity. They also appear in subjunctive mandatives (*I insist that she leave*) and after perception and causative verbs (*I saw her leave*). Different mechanisms: the modal auxiliary system, the mandative construction with its own acquisition path, and ECM licensing for perception verbs.

Past-participial clauses appear in the perfect (*has eaten*), the passive (*was eaten*), and as adjectival modifiers (*the broken window*). Different mechanisms again: the perfect auxiliary system, the passive voice construction, and adjectival conversion.

Gerund-participial clauses appear in the progressive (*is eating*), as nominalisations (*Eating is fun*), and as participial adjuncts (*Walking home, I saw a fox*). Still different mechanisms: progressive aspect construction, nominalisation patterns, and participial adjunct licensing.

Four clause types, four sets of environments, four sets of mechanisms. The only thing they share is a negative: their heads lack primary tense inflection.

8.5.2 Both diagnostics fail

Now apply the tests.

Projectibility: If you learn the distribution of infinitival clauses – where *to eat* can appear – does that help you predict the distribution of past-participial clauses? No. Knowing that infinitival complements follow *want* tells you nothing about where past-participial complements can appear. The syntactic environments don't overlap; the predictions don't transfer. The category doesn't project.

Homeostasis: Can we name a mechanism that maintains infinitival and gerund-participial clauses as similar to each other? No. The perfect auxiliary system has nothing to do with the progressive auxiliary system – except that both happen to select clause types whose heads lack primary tense. There's no feedback loop that keeps non-finite clauses together as a kind. They're grouped only by failing to be finite.

Negative classes show this most starkly. The analyst's need for exhaustive taxonomy requires every clause type to belong to a category, and "everything that's not finite" must go somewhere. But the property structure is empty. There's no positive cluster, only the absence of primary tense inflection. This is classification by exclusion, not by mechanism. The label non-finite clause organizes the analyst's files; it doesn't track causal structure.

When you diagnose a negative class, the guidance is clear: don't treat it as a kind. This doesn't mean the term is useless – NON-FINITE CLAUSES correctly identify clause types that contrast with finite clauses – but it means recognizing that the term is a taxonomic convenience, not a natural kind. The genuine HPCs are the specific constructions: INFINITIVAL COMPLEMENT, PERFECT CONSTRUCTION, PROGRESSIVE CONSTRUCTION, PARTICIPIAL ADJUNCT. Each has its own cluster, its own mechanisms, its own projectibility. The umbrella term names a residual class, not a causal unity.

8.6 The grain of analysis

The grain question previewed in §8.2 deserves more detail. We said language is an ecosystem – but which are the species?

Here is the key point. The failure modes are not just ways a category can miss; they are also what grain errors look like. Zoom in too far and you start reifying smoke rings: patterns generated as incidental fallout, too thinly maintained to count as kinds. Zoom out too far and you build wastebaskets: drawers that pool multiple clusters and then fail the projectibility diagnostic. Choose the wrong logical operation – defining by absence – and you get complement classes: taxonomic residues with no positive cluster to stabilize.

So the framework doesn't hand you a privileged unit in advance. It hands you a discipline: propose a unit, apply the diagnostics, and let the result tell you whether to decompose, reframe, or set the label aside. In practice, the mechanisms are often easiest to see at local grains – phoneme contrasts, lexemes, specific constructions, sometimes families or paradigms when stabilizers genuinely operate at that scale. What is rarely a kind is "English Grammar" as a unit. That is the fauna: an ecosystem of many kinds maintained by many mechanisms.

One place this discipline matters most is lexical semantics, where we keep demanding one partition where the system offers many.

8.6.1 Why lexical semantics feels slippery

Why does lexical semantics so often feel slippery? Not because meaning is mush, not because semanticists are careless, and not because semantics lacks structure: the compositional machinery can be exquisitely tight. The looseness lives at the lexical/interface level, and it comes from a mismatch of jobs. We ask one semantic taxonomy to track conceptual similarity, inferential profile, and grammatical patterning. Those cuts cross-cut. They don't line up.

A quick illustration. *Furniture* is a good similarity class, but English grammar mostly ignores it. ANIMACY is the reverse: it drives robust grammatical contrasts (*who* vs *what*, *he/she* vs *it*) while cutting across many conceptual domains. And inferential groupings like FACTIVITY organize items by what they commit you to infer, cutting across similarity and leaving a mixed grammatical footprint.

On the maintenance view, this is expected. A semantic distinction starts to behave like a kind when it repeatedly receives grammatical reinforcement – when constructions select for it, when morphology marks it, when processing routines exploit it, when acquisition gets convergent cues. Where that reinforcement is absent, you should expect overlapping, purpose-relative classes: EPISTEMIC KINDS in the sense already introduced, not homeostatic clusters.

So the question shifts from "What is the semantic category of X?" to "Which distinctions does this construction make usable – and what keeps them stable?" Where a distinction is grammaticalized, semantics firms up. Where it isn't, pluralism is the point.

8.7 Methodological implications

If we adopt this discipline, the landscape of linguistics looks different. The framework doesn't just identify failures – it tells us what to do about them.

8.7.1 What to do when you diagnose failure

Each failure mode has a corresponding response.

When you diagnose thinness: Stop treating the pattern as a category. Preposition doubling isn't a construction to be explained; it's a spillover effect to be noted and set aside. Thin patterns don't need theoretical treatment because they lack the causal structure that would make treatment meaningful. Document them, but don't reify them.

When you diagnose fatness: Decompose. ADVERB isn't a kind; the clusters inside it – around *quickly*, around *very*, around *thus* – might be, each maintained by its own mechanism braid. The genuine HPCs are lurking inside the wastebasket. Your job is to find them – to identify the sub-clusters that actually pass the diagnostics. The lumped label may remain useful for pedagogy or exposition, but it shouldn't guide theoretical inference.

When you diagnose negativity: Reframe. NON-FINITE CLAUSE names a contrast, not a unity. Replace it with positive categories – INFINITIVAL COMPLEMENT, PERFECT CONSTRUCTION, PROGRESSIVE CONSTRUCTION – that actually have mechanisms to maintain them. The negative label organizes your files; it doesn't track causal structure.

8.7.2 For typology

The framework reframes cross-linguistic comparison.

We stop asking whether Subject is a universal category – a question that invites endless definitional dispute. Instead, we ask: is the category Subject in Language L maintained by the same mechanisms as in Language M?

Usually, the answer is no. English subjects are maintained by agreement morphology, fixed word order, and nominative case marking. Tagalog "subjects" (if that's even the right term) are maintained by voice morphology and discourse-role assignment – entirely different mechanisms. They may look alike – both occupy some privileged argument position – but they're distinct HPCs, convergent solutions to different functional pressures.

What looks like shared identity is convergent evolution: the sameness of the label doesn't guarantee sameness of the kind, and to test cross-linguistic identity, you need to show shared mechanisms, not just shared features.

Haspelmath (2010) makes a related point with his distinction between comparative concepts and descriptive categories. Comparative concepts are analyst's tools – definitions constructed for cross-linguistic comparison. Descriptive categories are language-specific – the actual clusters maintained in particular speech communities. The maintenance view explains *why* these must be kept separate. Most comparative concepts are epistemic kinds: useful for organizing comparison, but not themselves maintained by shared mechanisms. Descriptive categories may be HPC kinds – but only if they pass the diagnostics in that particular language. Cross-linguistic "sameness" is a claim about convergent mechanisms, not shared essence.

That said, some comparative concepts may earn naturalised status when independent languages converge on similar patterns through similar mechanisms – the way camera eyes evolved independently in vertebrates and cephalopods. Naturalization is defeasible: it requires explicit mappings from comparanda (the cross-linguistic targets) to language-specific realizations, graded weights that track how closely each form realizes the target, and a firewall that diagnoses semantic targets independently of their morphosyntactic expression (see Reynolds, 2025c, for formal development). Without this discipline, we conflate labels with kinds.

Consider evidentiality. The semantic target – marking information source – recurs across languages: speakers regularly need to signal whether information comes from direct experience, inference, or report. This is a cross-linguistic comparandum, a stable discourse-pragmatic function.

But the morphosyntactic realizations are distinct HPCs. In Quechua, an obligatory evidential paradigm fills a slot on tensed verbs: speakers must encode source-of-information. This paradigm is maintained by slot-filling pressure, early acquisition, and morphological integration – robust homeostasis. In Turkish, the suffix *-mIş* marks non-first-hand information, but direct evidence goes unmarked: a different paradigm structure, a different functional niche. In English, there's no such morphosyntactic system. Source-marking is optional and scattered across adverbs (*apparently*, *reportedly*), parentheticals (*I heard*), and prosodic cues – lexical strategies, not a grammatical paradigm.

The label "evidentiality" conflates the semantic target with the morphosyntactic system. The target may naturalise: information-source marking is a stable discourse need, and some subtargets (reportative vs. inferential) may form genuine semantic clusters. But "evidentiality" is probably too fat to be a cross-linguistic morphosyntactic category – the label bundles obligatory paradigms, optional suffixes, and scattered adverbs under one umbrella, masking the distinct mechanisms maintaining each. The genuine HPCs are the language-internal systems: Quechua's paradigm, Turkish's suffix. English has no comparable morphosyntactic paradigm.

8.7.3 For theory

The framework shifts the burden of proof.

Labels without mechanisms are descriptive, not explanatory. If you claim that VOICE is a category in Language L, you need more than distributional evidence. You need to identify the mechanisms maintaining it – the acquisition pathways, the frequency effects, the functional pressures that keep the pattern stable.

This isn't a higher bar; it's a different bar. The framework doesn't demand more evidence; it demands a different kind of evidence. Mechanism identification becomes the theoretical goal, not just category identification.

8.7.4 For methodology

Falsifiability is built in.

The two-diagnostic test gives a decision procedure: apply projectibility and homeostasis criteria; if the pattern fails, it's not an HPC kind. This creates a research programme with clear success conditions. You can be wrong in a way that's detectable – and that's what makes the framework scientific.

The framework also explains gradience. Degrees of projectibility – strong at the core, weak at the periphery – aren't noise to be averaged away. They're signal: the signature of mechanisms operating at different strengths across the category. Peripheral variability tells you something about the causal structure.

8.7.5 Looking forward

This chapter has provided the discipline. It has shown what failure looks like – thin, fat, negative – and how to diagnose it using the two-diagnostic test. The framework can now be applied.

The same diagnostics – projectibility and the presence of stabilizing mechanisms – apply at the cross-linguistic level. Comparative concepts like NOUN and ADJECTIVE can be diagnosed as thin (diagnostics fail across families), fat (the category lumps distinct phenomena), or negative (defined by what it's not). Chapter 15 develops this typological implication; Reynolds (2025c) provides the formal framework.

Part III applies it. Countability offers an HPC that passes both diagnostics. Definiteness presents a case where form and meaning may constitute separate, overlapping HPCs. Word classes return us to traditional categories – asking which survive the test and which decompose under pressure. And the final chapter asks whether the framework scales – from phonemes to constructions to "the grammar" itself.

Categories are maintained; classes are merely named. The diagnostics tell us which is which.

8.8 The HPC-kind audit

The diagnostics developed in this chapter can be synthesized into a reusable template. For any candidate category C, the audit proceeds as follows:[2]

Table 8.2. HPC-kind audit template

Step	Content
1. Target + scope	What entities/events does C range over? Specify scope: population/register, time slice, and grain (tokens/constructions/categories/meta-categories).
2. Property cluster (profile)	List 8–15 candidate properties (distributional, morphosyntactic, semantic/pragmatic, processing, acquisitional, diachronic, sociolinguistic); mark central vs peripheral; note which covary robustly within the stated scope.
3. Homeostatic stabilizers	Acquisition, entrenchment/use, interactive alignment, iterated transmission, institutional enforcement (if relevant); state what each stabilizer predicts about persistence/repair under perturbation.
4. Projectibility claims	Which inductive generalizations does C license (and for which tasks/data types)? Specify the intended domain of stability and the intended domain of generalisation.
5. Boundary behaviour	Expected borderline cases and expected decouplings (which properties can come apart), with stabilizer-based explanations.
6. Failure-mode gate	Thin: weak clustering or weak stabilizer story. Fat: one label covers multiple stabilizer-backed subclusters that would better be distinguished. Negative: primarily defined by absence or contrast-set complement.
7. Falsifiers / stress tests	What pattern would force reclassification (HPC $\leftrightarrow$ fat/thin/negative)? What perturbation (register shift, cohort shift, task shift, diachronic window) should reveal decoupling if the diagnosis is wrong?

The decision rule from §4.4 applies at Step 6:

> If the only "mechanism" story available is analyst convenience or measurement convenience, classify as a thin or fat convenience class, not an HPC kind.

To be clear, interest-relativity is not a failure mode. Different analysts tracking different stabilizer-backed clusters should use different labels – that's the framework working correctly. A fat class arises when one label is forced to cover divergent subclusters rather than acknowledged as multiple HPCs.

The audit isn't a checklist to be completed once and filed away but a diagnostic instrument to be re-run when predictions fail, when the scope shifts, or

[2] Part III applies this template to countability (Chapter 9), definiteness (Chapter 10), and word classes (Chapter 11). Each chapter ends with an abbreviated version of this audit.

when new perturbation evidence becomes available. The goal isn't to certify categories but to calibrate confidence in them and to make explicit what would change that confidence.

Definiteness thread. This chapter's failure-mode diagnostics predict that a unified "definiteness" label risks being a *fat* class: it covers two stabilizer-backed subclusters (semantic definiteness, formal deitality) that should be distinguished. Chapter 10 runs the audit. The stress test: if weak definites, generic definites, and proper names each pattern differently across the failure-mode gate, the fat-class diagnosis is confirmed. Conversely, if these constructions behave uniformly under perturbation, the single-category view survives.

Part III

Categories Reconsidered

Countability

The word LEGO® is a brand name and is very special to all of us in the LEGO Group Companies. We would sincerely like your help in keeping it special. Please always refer to our bricks as 'LEGO Bricks or Toys' and not 'LEGOS'.

—LEGO Group consumer catalog, c. 1980s

The lawyers at the LEGO Group have been fighting this battle for forty years. They have trademark law on their side. They have the Chicago Manual of Style. They have millions of dollars in brand-management budget.

And they are losing.

To a lawyer, *LEGO* should only ever modify – *LEGO bricks*, *LEGO toys* – never stand alone as a count noun. To a six-year-old, a plastic brick is a discrete, bounded, manipulable object. And English has a deeply entrenched pattern for dealing with discrete, bounded, manipulable objects: it counts them. It adds an *-s*. It makes them plural. When a child asks for *three Legos*, they aren't making a mistake; they are applying an inference they've learned from thousands of similar cases. The count cluster is claiming the word.

This chapter is about that mechanism – the control loop that decides what gets counted. The count/non-count distinction isn't a static binary. It's a dynamic negotiation. And as we'll see, the HPC framework we built in Part II handles this messy distinction cleanly.

Countability, I'll argue, isn't a definition. It's a homeostatic property cluster – and it illustrates something general about how HPCs work. We leave the abstract theory for the concrete terrain. Countability is our first basin – a deep, stable attractor in the landscape of grammar. At a finer grain, categories often reveal internal structure: sub-clusters maintained by different mechanisms, held together by interface processes. Countability makes this structure visible because the two clusters – one semantic (individuation), one morphosyntactic (the count cluster) – are saliently distinct. What looks like a single category is a coupling. (A note: Throughout Part III, I use MECHANISM and STABILIZER to refer to causal structures and their functional roles respectively (Illari & Williamson, 2012).)

9.1 One word, two categories

The word *countability* hides a structural fact. At one grain, it names a single category – a coherent cluster that English speakers navigate daily. At a finer grain, it reveals itself as a coupling between two: the COUNT CLUSTER (morphosyntactic) and the INDIVIDUATION CLUSTER (semantic). In the form–value terms of Chapter 13, the count cluster is the form side – plural marking, quantifier selection, agreement – and individuation is the value – what a noun construal contributes when deployed, whether it signals bounded discreteness or homogeneous mass. Each cluster is maintained by its own mechanisms. Countability is both the whole and the joint.

This matters because it explains something puzzling. Chapter 8 argued that lexical semantics alone rarely produces HPC kinds: categories like *furniture* (conceptual similarity) or ANIMACY (referent property) don't generate the tight clustering that projectibility requires. But countability does. English speakers reliably extend count morphosyntax to novel words, predict quantifier compatibility from plural marking, and judge intermediate cases with surprising consistency. Why?

The answer is that countability isn't just lexical semantics. It's an *interface system* – a place where a semantic distinction receives repeated grammatical reinforcement. When constructions select for individuation, when morphology marks it, when processing routines exploit it, when acquisition converges on it – then the semantic distinction starts to behave like a kind. This is what makes countability different from *furniture*.

The literature offers several competing diagnoses. *Mereological accounts* treat countability as tracking part-whole structure: count nouns denote atomic entities, mass nouns denote divisible stuff (Chierchia, 1998; Link, 1983). *Cognitive accounts* ground the distinction in perception: individuation reflects object-file cognition and early-emerging core knowledge (Bloom, 1994; Spelke & Kinzler, 2007). *Grammatical accounts* treat count/mass as an autonomous formal feature, assigned lexically and enforced by agreement (Borer, 2005). Each preserves something – ontological grounding, cognitive reality, formal tractability – but none explains why the *same set of morphosyntactic properties* clusters with individuation across constructions.

Here is the decision criterion: any account that keeps countability purely semantic must explain why plural marking, quantifier selection, and agreement – but not, say, animacy or abstractness – cluster into a single syndrome. Conversely, any account that treats object-mass nouns as lexical exceptions still owes an explanation for why the exception pattern is *systematic*. The two-cluster architecture is not an alternative semantics; it's a demand that any semantics meet.

Individuation without count morphology; count morphology without individuation.

The next two sections unpack the two clusters. First, the semantic side: what is individuation, and what maintains it as a cluster of properties? Then, the morphosyntactic side: what is the count cluster, and what keeps its properties together? Only after both are on the table can we ask how they couple.

9.2 The individuation cluster

Individuation isn't a single primitive. It's a cluster of observable properties that typically travel together – what the formal semantics literature calls the accessibility of atoms to quantificational operations, formalized by Link (1983) and developed by Chierchia (1998), Rothstein (2010), and Grimm (2018). I'll first describe the profile (the symptoms) and then turn to the mechanisms (the causes) that keep them coherent.

The cluster typically comprises four properties. BOUNDEDNESS entails that the referent has discrete edges – a spatial or conceptual boundary that separates it from its environment. ATOMICITY means the referent is accessible as a unit that can be singled out and tracked. ENUMERABILITY makes the referent compatible with exact counting – three of them, not just much of it. Finally, HOMOGENEITY RESISTANCE means that parts of the referent are not the same kind as the whole: half a cat is not a cat, whereas half of water is still water. Half a cat is a problem, not a category.

These properties cluster. When a referent is bounded, it's usually atomic; when it's atomic, it's usually enumerable; when it's enumerable, it usually resists homogeneous subdivision. The clustering isn't accidental. But it is not absolute. A *ball bearing* is maximally individuated: bounded, atomic, enumerable. But a *cloud* is bounded while lacking stable atomicity. *Cattle* denotes discrete animals, but the noun construes them as an aggregate. The cluster represents a prototype, not a rigid definition.

9.2.1 What maintains the cluster

Descriptively, the cluster looks like a set of logical entailments. Explanatorily, it is maintained by perceptual and cognitive mechanisms that operate across modalities – and, strikingly, across species.

Consider EDGE DETECTION. Visual object perception depends on detecting boundaries. The visual cortex uses orientation-selective cells, Gabor-like filters, and mid-level grouping principles to extract edges from the visual field. These edges are then grouped into coherent object representations – what Kahneman et al. (1992) called OBJECT FILES: temporary episodic representations that bind features to locations and track objects across time.

But edge detection isn't vision-specific. Blind echolocators produce tongue clicks and interpret the returning echoes to perceive objects. Research shows that they can identify object shape, size, and location with remarkable precision (Kolarik et al., 2014) – and fMRI studies reveal that they activate visual cortex when doing so (Thaler et al., 2011). The perceptual system has found a different input modality but the same computational goal: bounded individuals.

The pattern extends beyond perception. The macrophages we met in Chapter 7 face the same computational problem at a different scale: distinguishing self from non-self. They use pattern recognition receptors – toll-like receptors, scavenger receptors – to detect molecular edges, boundaries between what belongs in the tissue and what doesn't. Edge detection – the computational problem of finding where one thing ends and another begins – recurs from molecular biology to perception to cognition.

This recurrence is predictable rather than mysterious. Evolution keeps rediscovering boundary-making machinery because organisms that can cheaply and reliably carve input into stable units – things with edges, persistence, and re-identifiability – can move, grasp, avoid, and respond more effectively. The mechanisms are not homologous across levels, and I am not claiming that macrophages "do vision"; the point is computational: many systems face a version of the same individuation problem, and the easiest solutions bundle cues that cohere. That bundling is what gives the individuation cluster its tight internal correlations, making it a plausible semantic anchor for grammatical reinforcement. And so, such mechanisms are ancient.

Spelke and Kinzler (2007) argues[1] that human infants are equipped with CORE KNOWLEDGE systems for objects – systems that operate from the first

[1] Or should that be *argue*?

months of life. Infants as young as four months perceive objects as cohesive (parts move together), bounded (edges mark identity), and continuous (objects persist through occlusion). These are the same properties that define the individuation cluster in adult conceptual systems. The mechanisms that maintain individuation are not learned from scratch; they're built on an early foundation.

Cross-modal integration reinforces the pattern. Object individuation isn't locked to a single sensory channel. When you set a cup down and see and hear it alighting, you don't perceive thre entities – one tactile, one visual, one auditory. You perceive one object with multimodal properties. The binding is maintained by cross-modal integration mechanisms that enforce coherence: what looks bounded should sound bounded, feel bounded, behave as a unit. This integration reinforces the cluster. The properties travel together because multiple systems expect them to travel together.

This gradient nature of individuation is crucial for what comes next. The count cluster – the morphosyntactic syndrome – tracks individuation. When individuation is strong, the full count cluster applies. When it dissolves, the count cluster dissolves with it. But it does so in an orderly way.

9.3 The count cluster

The distinction seems intuitive. *Cat*, *chair*, and *idea* differ from *water*, *mud*, and *traffic*. The first group denotes discrete individuals; the second denotes undifferentiated substances or aggregates.

Otto Jespersen first formalized the distinction in 1924, coining the term mass-word for nouns that resist counting:

> There are a great many words which do not call up the idea of some definite thing with a certain shape or precise limits. I call these 'mass-words'; they may be either material, in which case they denote some substance in itself independent of form, such as *silver*, *water*, *butter*, *gas*, *air*, etc., or else immaterial, such as *leisure*, *music*, *traffic*, *success* (Jespersen, 1924, p. 198).

But Jespersen's insight goes back further. The chapter was adapted from a lecture he delivered to the Copenhagen Academy of Sciences in 1911 – thirteen years before publication, and decades before the cognitive revolution.[2] Even then, he saw that grammar doesn't just mirror the world; it makes some construals cheap and others costly – not by determining thought, but by defaulting certain ways

[2] Jespersen also documented the flexibility of individual nouns across count and mass uses: *a little more cheese* vs. *two big cheeses*; *it's hard as iron* vs. *a hot iron*. The phenomenon was recognized a century ago.

of packaging experience. While English speakers can say *two apples* or *two fruits*, we can't take a microscope and a stethoscope and call them **two equipment* – unless we supply a unit (*two pieces/items of equipment*). English makes you bring your own measuring cup. The relevant individuation is in the construal, not in the furniture of reality.

Standard grammars like *CGEL* (Huddleston & Pullum, 2002) identify a cluster of morphosyntactic properties that typically align.[3] Again, we can distinguish the observable profile from the explanatory engine. For a prototypical count noun like *ball*, the cues reinforce the individuation. The noun denotes discrete, bounded entities; it inflects for number (*balls*); it accepts exact cardinals (*one ball*, *two balls*) and fuzzy count quantifiers (*many*, *fewer*); and it triggers agreement matching its morphology.

By contrast, for a prototypical non-count noun like *water*, the cluster prevents individuation. The noun denotes undifferentiated substance; it lacks a plural form (outside special interpretations); it rejects cardinals (**two waters*) and takes mass quantifiers (*much*, *less*); and it triggers singular agreement.

This is the COUNT CLUSTER: the set of formal properties that normally travel together. When you hear a word has a plural form, you can predict it will take *many* and reject *much*. You can predict it will refer to something conceptualized as an individual.

9.3.1 The problem: Object-mass nouns

The problem is that the world doesn't always cooperate. Rothstein (2010) identified the class of 'object-mass nouns' – words like *furniture*, *footwear*, *cutlery*, and *jewelry*. Semantically, these refer to discrete, countable artifacts. A chair is as discrete as a ball. But *furniture* is grammatically mass:

(1) a. * *three furnitures*
    b. * *many furniture*
    c. *much furniture*

Why? If the grammar just tracked ontological discreteness, *furniture* should be count. The fact that it isn't – that English forces us to say *pieces of furniture* while French speakers comfortably say *trois meubles* (three furnitures) – proves that countability is an autonomous grammatical system, not a direct read-out of physics.

[3] *CGEL* is careful to distinguish count and non-count *senses* rather than strict lexical subcategories, acknowledging the flexibility of many nouns (e.g., *beer*, *cake*).

9.4 The coupling

Now we can ask what couples the individuation (semantic) cluster to the count (morphosyntactic) cluster into a single system. The answer is BIDIRECTIONAL INFERENCE – the mechanism linking semantic construal to grammatical form.

Each count property is a cue – a piece of morphosyntactic evidence that hearers use to infer how the speaker is construing the referent. When you hear *three dogs*, you infer that the speaker has individuated the referents – construed them as discrete, enumerable units. When you hear *much water*, you infer a non-individuated construal: stuff, not atoms. The morphosyntax generates expectations about the construal; the construal, in turn, constrains morphosyntactic choice.

The coupling works in both directions. In comprehension, count morphosyntax (*a*, *three*, *many*) leads the hearer to expect atomic, enumerable referents. In production, an individuated construal prompts the speaker to select count morphosyntax. The properties cluster because they are cues to the same underlying condition.

The properties cluster because they're all coupled to the same semantic variable. Knowing that a noun takes *a(n)* tells you it supports individuation – and if it supports individuation, it should take *three* and *many* and plural agreement too. The generalisation runs through the construal, not through the morphosyntax directly. That's why the properties are mutually predictive: not because any one *causes* the others, but because they're all symptoms of the same underlying condition.

9.4.1 Why the coupling produces an HPC

This is where the payoff from Chapter 8 arrives. Lexical semantics alone rarely produces HPC kinds: categories like *furniture* (conceptual similarity) don't generate the tight clustering that projectibility requires. Why should countability be different?

The answer is that countability operates at the syntax-semantics interface. The individuation cluster is reinforced by perceptual mechanisms (edge detection, object files, cross-modal integration). The count cluster is reinforced by morphosyntactic mechanisms (acquisition, entrenchment, transmission). But the coupling between them – the bidirectional inference mechanism – provides a third layer of reinforcement. Each time construal and form align, both clusters are strengthened. Each time they're transmitted together to a new learner, the coupling tightens.

Furniture lacks this reinforcement. The conceptual similarity exists – chairs and tables share properties – but English provides no morphosyntactic system that selects for that similarity and reinforces it with every utterance. There's no grammatical slot that says 'furniture-like things go here'. The category exists in the lexicon but not in the grammar.

Countability is the opposite. The grammar provides slots that select for individuation: *a* requires atomic reference; *three* requires enumerable atoms; *many* tolerates weaker individuation. Every time a speaker uses these constructions, they reinforce the link between semantic construal and grammatical form. This is why countability passes the HPC diagnostics: the coupling mechanism provides the homeostasis that lexical semantics alone can't.

We'll return to this relationship in Chapter 13. Grammaticality judgements aren't just about form; they're about the successful pairing of a form with a construal. Note during drafting: Make sure this cashes out in Ch 13 or is removed.

9.4.2 Multi-timescale maintenance

The bidirectional inference mechanism operates at multiple timescales, each reinforcing the others.

At the fast timescale of PROCESSING (milliseconds), every time a speaker produces or comprehends a count frame, the form–meaning link is activated. *Three dogs* primes the expectation of individuation; individuation primes the expectation of count morphosyntax. Mismatches – **three furnitures*, say – incur processing costs. These costs are ERROR SIGNALS (Chapter 7): they scream that the coupling has been violated, providing the negative feedback that keeps the basin's edges steep. They feel wrong because they violate entrenched expectations.

At the slow timescale of ACQUISITION (years), children don't learn count properties one by one. They learn that count morphosyntax correlates with individuation. Classic acquisition work shows that children overgeneralise: encountering a noun in one count frame, they extend it to others (Bloom, 1994; Gordon, 1985). A child who hears *many police* may try **three police*, treating the loose property as evidence for the tight ones. When a speaker conceptualizes a referent as individuated (bounded, atomic), they reach for the tool that signals individuation: the count syntax. This is why children say *I saw three sheeps* or *two mouses* or *my foots hurt*. The child perceives discrete animals, discrete body parts. The grammar provides a slot for discrete entities (the regular plural); the child fills the slot, ignoring the lexical exception. Such overgeneralisation is systematic. They're obeying the rule the grammar actually taught them; the exceptions didn't show up to class. The error proves the mechanism works – they're running the algorithm perfectly, just on irregular input.

Finally, at the timescale of TRANSMISSION (decades), institutional forces – style guides, copy editors, teachers, grammar checkers – enforce canonical patterns and resist drift. The *data/datum* debate lives here. As *datum* recedes from editorial practice, the normative anchor weakens and *data* drifts toward mass status. The institutional layer doesn't *create* the cluster, but it stabilizes it at the community level.

The mechanisms interlock: the fast loop generates usage patterns; the slow loop crystallises them into community standards; acquisition transmits the crystallised patterns to new learners, who then participate in the fast loop. Metalinguistic feedback adds another layer: prescriptive campaigns (like LEGO Group's forty-year effort to block
mentionLegos) can stabilize or destabilize construals – though as the epigraph suggests, they rarely override entrenched patterns. Perturbation to any level – a semantic shift that weakens individuation, a prescriptive shift that sanctions formerly deviant forms – ripples through the system and changes the equilibrium. The cluster isn't static. It's dynamically maintained.

9.4.3 The chunking story

There's a cognitive dimension to why high-frequency patterns resist change. Following Bybee (2010), we can understand entrenchment in terms of *chunking*: high-frequency sequences get stored as units and accessed as wholes, bypassing compositional assembly.

Many cattle is a chunk. English speakers have encountered it often enough that it's stored in memory and retrieved directly. *Three cattle* isn't a chunk – it has to be assembled online, and the assembly fails because the components don't fit. The noun's stored profile resists the frame the speaker is trying to build.

This is why quasi-count nouns feel unstable in tight frames but natural in loose ones. *Many cattle* is retrieved; *three cattle* is constructed and rejected. The difference between the two isn't a difference in categorical membership. It's a difference in processing: one path is entrenched; the other isn't.

9.5 The hierarchy: Tight before loose

If all count properties were equally sensitive to individuation, any weakening would collapse the cluster entirely. A noun would be fully count or fully mass, nothing in between. But that's not what we see. *Cattle* and *police* sit in the middle: they accept *some* count properties while rejecting others. And the pattern isn't random. The properties peel off in order.

Here's why. Although all count properties are coupled to individuation, they're not equally demanding. Some require high-precision individuation – exact atomic units, sharply bounded. Others tolerate lower precision – approximate magnitude, vague plurality. When individuation weakens, the demanding properties fail first.

9.5.1 Locks with different tolerances

Think of it this way. Each count construction – *a* N, *three* N, *many* N – is a lock with a specific tolerance. The noun provides a key: its individuation profile. If the key clears the tolerance, the construction is licensed.

A is a precision lock. It requires identifying exactly one atomic unit. The key must be cut to exact specifications.

Three is nearly as precise. It requires enumerating exactly three atoms – no approximation, no vagueness.

Many is a forgiving lock. It requires only magnitude assessment: "a contextually large quantity." The atoms don't need sharp boundaries; they just need to be roughly atom-shaped.

Plural agreement is the most forgiving. It requires only that the referent be construed as non-singular. Virtually anything that's in the count basin clears this one.

Book is a precision-machined key. It opens every lock in the system. *Cattle* is a blunter key. It gets through the forgiving locks – *many cattle*, plural agreement – but can't open the precision locks: **a cattle*, **three cattle*.

9.5.2 From locks to basins

But here's how this connects to what we've seen before. Remember the spinning top from Chapter 4? A category isn't a container you're in or out of – it's an attractor keeping a top upright. The count cluster is that basin. The locks? They're positions *within* the basin. Tight locks sit at the centre; loose locks ring the edges.

When we say *cattle* clears the loose locks but fails the tight ones, we're saying its top spins stably in the basin – but off-centre. It's count-ish. It's in the attractor. But it's not at the core.

Book spins at the dead centre, satisfying every precision demand. *Cattle* spins off-centre but stably – it's in the count basin, but it can't reach the precision locks at the centre. *Folks* – we'll come to this – wobbles. Sometimes it clears the *three* lock, sometimes it doesn't. Its position in the basin is unstable.

This integrated image – locks as positions in a basin, precision as distance from centre – unifies the immediate intuition (locks and keys) with the dynamic stability framework (spinning tops and attractors). The hierarchy of count properties *is* the geometry of the basin. The further from centre a lock sits, the more tolerance it has.

9.5.3 The implicational pattern

The hierarchy generates a prediction: for any noun and any two properties, if it accepts the tighter one, it accepts the looser one; if it rejects the looser one, it rejects the tighter one. The distribution should be triangular. No noun should stably accept *three* while rejecting *many*. No noun should require *a(n)* while taking *much*.

This is falsifiable. Finding a noun that reverses the pattern – tight without loose – would challenge the account. The empirical record, as far as I can determine, contains no such case.

Here's the hierarchy, ordered from tightest to loosest. Singular forms and *a(n)* are the strictest, requiring the identification of exactly one atomic unit. Low cardinals like *three* and *five* are similarly precise, requiring exact enumeration. Relaxing the tolerance slightly, *several* requires multiple units but permits approximate quantity. Distributives like *each* require discreteness but not enumeration. Further down, *many* and *few* require only relative magnitude assessment. High round numerals often function as approximate measures rather than exact counts. Finally, plural agreement is the most permissive, requiring only a non-singular construal.

A noun that loses *several* will already have lost *a(n)* and low cardinals. A noun that retains only plural agreement will have lost everything above it. The properties form an implicational scale, and a noun's countability profile is its position on that scale.

Table 9.1 shows the triangular pattern at a glance.

Table 9.1. Implicational matrix for count properties. Checkmarks indicate acceptable combinations; crosses indicate ungrammatical. The triangular pattern shows: if a noun accepts a tight property, it accepts all looser ones.

	a(n)	*three*	*many*	Agreement
book	✓	✓	✓	✓
cattle	×	×	✓	✓
police	×	×	✓	✓
furniture	×	×	×	sg.
water	×	×	×	sg.

9.6 Quasi-count nouns: The stable intermediates

CGEL identifies a class of QUASI-COUNT NOUNS: plural-only nouns that take plural agreement and accept *many* but resist singular forms, *a(n)*, and low cardinals (Huddleston & Pullum, 2002, p. 345). The core cases are *cattle*, *police*, *poultry*, *vermin*, *livestock*, and *clergy*.

They occupy exactly the position the hierarchy predicts: loose properties accepted, tight properties rejected.

- *Many cattle* —✓
- *The cattle are grazing* —✓
- *Several cattle* —marginal for some speakers
- *Three cattle* —×
- *A cattle* —×

The same pattern holds for *police*: *many police*, *the police are investigating*, but **three police*, **a police*.

A related class includes "pluralia tantum" nouns like *groceries*, *genitals*, *dregs*, and *remains*. Unlike *cattle*, these possess explicit plural morphology. But like *cattle*, they resist exact enumeration: we buy *groceries*, but we do not typically count **three groceries* or identify **a genital*. This reinforces the autonomy of the system: even explicit plural morphology does not guarantee access to the full count cluster. The *many* and *agreement* locks are open, but the precision locks remain closed.

And none of these nouns reverses the hierarchy. None accepts tight properties while rejecting loose ones. The triangular pattern holds.

9.6.1 Why are they stable?

If the homeostatic mechanism creates pressure toward coherence – pushing nouns toward full count or full mass – why haven't *cattle* and *police* drifted? They've been quasi-count for centuries.

The answer is functional anchoring. When speakers need singulative reference – when they need to name *one* cow or *one* officer – they don't attempt **a cattle* or **a police*. They use *cow*, *bull*, *head of cattle*, or *officer*. These alternative lexemes handle the singulative function, relieving pressure on the quasi-count noun to develop tight-linkage properties.

Bidirectional inference generates expectations. If a noun accepts *many*, hearers may expect it to accept *three* and *a*. When those expectations fail, pressure arises either to regularise (extend tight properties) or to avoid the construction. If an alternative lexeme satisfies the singulative function, speakers have no reason to force the quasi-count noun into tight frames. The pressure dissipates.

This is *passive* persistence, not active maintenance. *Cattle* doesn't serve some function that requires it to stay quasi-count. It simply faces no pressure to change, because *cow* is handling the job that analogy would otherwise push *cattle* to fill. We asked in Chapter 1 why *cattle* has persisted for five centuries. Now we have an answer: the equilibrium is stable because the communicative ecology absorbs the force that would otherwise destabilize it.

Compare this to *police*. The register variation is instructive: *the police* is formal and aggregate; *cops* is informal and count-friendly (*three cops*); *officer* is the robust singulative. Speakers who need to refer to a single police officer have an obvious lexical resource. Result: *police* has remained quasi-count for as long as we have reliable records.

9.6.2 The unstable case: ⟨folks⟩

Not all intermediates are stable. *Folks* occupies the boundary zone where the hierarchy predicts variability.

American English *folks* sits between *people* (fully count: *three people*, *a person*) and the quasi-count class. For many speakers, *many folks* and *several folks* are fully acceptable. But *three folks* is often judged marked, informal, or slightly odd. Some speakers reject it; others accept it readily; still others accept it but hear it as slangier than *three people*.

The HPC account predicts this instability. *Folks* lacks a functional anchor. There's no singulative **a folk* in ordinary use, and *person* is semantically distinct – neutral rather than in-group. Unlike *police* (anchored by *officer*) and *cattle* (anchored by *cow*), *folks* has nothing to bleed the pressure for regularisation.

And corpus data confirms that *folks* is suppressed relative to *people* in tight frames. In COCA, *three people* occurs at 2,226 per million tokens of *people*; *three folks* occurs at only 258 per million tokens of *folks* – an 8.6-fold suppression. The loose property (*many*) is suppressed only about 2.3-fold. Tight properties are hit harder than loose ones, exactly as the hierarchy predicts.

The instability of *folks* is not noise. It's structure. The noun is wobbling in the basin, uncertain whether it will stabilize off-centre (like *cattle*), drift outward (toward mass), or regularise toward the centre (developing tight properties). The outcome depends on whether a functional anchor emerges or prescriptive pressure crystallises.

9.7 Diachronic signatures

If the homeostatic mechanism is real, it should leave traces in the historical record. Two cases illuminate how the cluster extends and how it erodes.

9.7.1 How the cluster self-completes: ⟨pea⟩

We mentioned *pea* in Chapter 1. Here's the fuller story.

Speakers of Middle English heard *pease* with a final /z/ sound – the word for the vegetable, used in *pease porridge* and *pease pudding*. The noun was non-count: you had *much pease*, not *many pease*. But the final /z/ was phonologically identical to the plural suffix, and at some point speakers reanalysed it as such.

Once *pease* was heard as a plural, the bidirectional inference mechanism kicked in. Plural morphology cues individuation. If *pease* is plural, where's the singular? Speakers who expected the cluster to cohere created a gap – and filled it. They back-formed *pea*. Then they extended the count cluster: *a pea*, *three peas*, *many peas*. The count cluster didn't just emerge. Speakers *built* it, one inference at a time.

The mechanism is visible here: cluster pressure creates gaps; gaps get filled. If the bidirectional inference story is right, we'd expect the historical record to show loose properties established before tight ones – the cluster building from the outside in. The evidence is suggestive though incomplete; what's clear is that the reanalysis triggered a cascade of count-property adoption.

9.7.2 The unstable hybrids: Data

Then there is *data*.

On March 11, 2015, Minnesota legislators halted work on a license-plate reader bill to debate grammar. The bill text read *the data are private*; Representative John Lesch insisted *data* is singular. The committee voted; the motion passed unanimously; Lesch pumped his fist. The law was amended: *data is*. Grammar rarely gets a roll-call vote, so it took the opportunity.

Lesch won the vote, but he entered what usage expert Bryan Garner calls a *skunked* term argument (Garner, 2016). *Data* is currently moving from the count basin (Latin plural of *datum*) to the mass basin (synonym for *information*). In the transition, it exhibits the chaotic behavior of a system seeking a new equilibrium. Scientists say *data are*. Tech CEOS say *data is*. The cluster is unraveling, property by property. Historically the plural of *datum*, it's shifting toward mass status: *this data is*, *much data* are now common, especially in informal and spoken registers. Corpus studies confirm that singular agreement with *data* now predominates in most registers (Garner, 2016).

Why is *data* drifting? Because its functional anchor is disappearing. *Datum* – the singulative – has become archaic, confined to philosophy-of-science contexts and pedantic style guides. Without a robust singulative in active use, the tight-linkage properties have nothing to attach to. Speakers who need to refer to a single piece of information say *data point*, not *datum*. But *data point* is a compound, not a singulative of *data*. It doesn't anchor the count cluster the way *officer* anchors *police*.

Result: *data* drifts toward the loose end of the scale. Plural agreement weakens (*this data is* becomes standard). Tight properties erode (*three data* was always rare). Eventually, if the drift continues, *data* will be fully mass – *much data*, *this data is*, *some data* – with *data point* handling any count function.

This is the quasi-count pattern run in reverse. *Cattle* is stable because *cow* exists. *Data* is drifting because *datum* is dying.

9.8 Cross-linguistic parallels

If bidirectional inference is a general mechanism – not an English quirk – languages with different morphological resources should show analogous patterns.

Welsh and Arabic mark singulatives morphologically. In Welsh, the base form of *adar* ("birds") is collective – grammatically singular, semantically aggregate. A suffix derives the singulative: *aderyn* ("a bird"). Grimm (2018) documents exactly the predicted pattern: bare collectives accept loose quantifiers but resist low numerals and distributives, while singulative-marked forms accept the full count cluster. The parallel to English quasi-count nouns is striking: Welsh collectives occupy the same position in the hierarchy that *cattle* and *police* occupy in English.

Classifier languages (Mandarin, Japanese) encode individuation differently – through classifiers that mediate between numerals and nouns, rather than through inflection on the noun itself. The prediction is that the clustering dynamics should shift to the classifier system: general classifiers should show tight/loose asymmetries parallel to what English quantifiers show. This remains programmatic – the cross-linguistic work hasn't been done to the depth needed – but the mechanism should apply wherever languages encode individuation morphosyntactically.

9.9 Sharp boundaries in fuzzy territory

There's a puzzle lurking in the data. The hierarchy is continuous – properties shade from tight to loose – but judgments are often sharp. *Three cattle* isn't "slightly bad" for most speakers; it's simply ungrammatical. *Many cattle* isn't "slightly good"; it's fully acceptable. The boundary between what *cattle* licenses and what it rejects feels determinate, even though the underlying individuation variable is gradient.

Chapter 5 offered a framework for this: sharp-but-unknowable boundaries arising from tolerance dynamics. The hyperreal model lets us have determinate boundaries without anyone knowing precisely where they fall. Speakers act as if the boundary is sharp – they judge *three cattle* ungrammatical, not merely marginal – but they can't articulate the precise point at which individuation becomes sufficient.

Countability confirms this picture. The *folks* case is revealing: inter-speaker variation in the acceptability of *three folks* is real and robust. Some speakers place the boundary above *folks* (it clears *three*); others place it below (it doesn't). Each speaker's individual grammar has a determinate answer, but the community doesn't. The gradience shows up in population variance, not in graded individual judgments.

This is the empirical signature of tolerance-based boundaries: sharpness within grammars, variance across them. The count/non-count distinction, for all its apparent fuzziness, behaves like a boundary system in which each speaker draws a line – just not the same line.

9.10 Passing the tests

Chapter 8 introduced the Two-Diagnostic Test for genuine HPC kinds: high projectibility and robust homeostasis. Let's apply it to the count cluster.

9.10.1 Projectibility

The projectibility criterion asks: does recognizing this kind support successful induction? For the count cluster, the answer is yes.

If you know a noun is count, you can predict its behaviour across a range of grammatical contexts. It will take *a(n)* and low cardinals. It will take *many/few*, not *much/little*. It will trigger plural agreement when plural. It will combine with distributives. Every one of these predictions is testable, and for canonical count nouns, every one succeeds.

Crucially, the predictions are *gradient* by position in the hierarchy. Knowing that *cattle* is quasi-count – that it accepts loose properties but rejects tight ones – lets you predict precisely which frames it will enter. The quasi-count pattern is projectible too. The category structure supports differentiated, not just all-or-nothing, induction.

Compare this to a merely nominal grouping – say, "nouns ending in -tion". Knowing that *nation* ends in -tion tells you nothing about its grammatical behaviour distinct from knowing it's a noun. The suffix doesn't support induction. Countability does.

9.10.2 Homeostasis

The homeostasis criterion asks: is the clustering maintained by causal mechanisms, or is it merely a surface pattern that might scatter under perturbation?

We've identified the mechanisms:

We have identified five interlocking mechanisms. Bidirectional inference couples morphosyntax to individuation through a shared semantic variable. Acquisition transmits this coupling as a unit, as children overgeneralise from one property to the others. Entrenchment preserves high-frequency patterns as chunks. Institutional norms stabilize the distribution at the community level. Finally, functional anchoring bleeds pressure from intermediate cases like *cattle*, allowing them to persist without regularising.

This is robust homeostasis. Perturb the system – weaken individuation, remove a singulative anchor, expose learners to non-standard input – and the cluster responds in predictable ways. It's not just that count properties co-occur; it's that mechanisms push them toward co-occurrence. The clustering is maintained, not accidental.

9.10.3 The verdict

The count cluster passes both diagnostics. It supports induction (knowing a noun's count status predicts grammatical behaviour) and it's held together by mechanisms (bidirectional inference, acquisition, entrenchment, anchoring). It belongs in the upper-right quadrant of the diagnostic matrix: a genuine HPC kind.

This doesn't mean countability is simple. The hierarchy reveals internal structure; the quasi-count cases reveal tolerated heterogeneity; the diachronic cases reveal dynamic equilibria that can shift. But the claim isn't that HPC kinds are uniform. The claim is that they're real – that they're maintained by mechanisms and support induction. The count cluster qualifies.

9.11 What does this buy us?

The HPC analysis of countability does three things that the traditional feature-based account doesn't.

First, it explains the clustering, not just labels it. The feature [+count] says that count properties go together. The HPC account says *why*: they're inferentially coupled to individuation, and mechanisms keep them coupled. The clustering isn't a brute fact; it's a consequence of how form and meaning interact in processing, acquisition, and transmission.

Second, it predicts the dissociation order. When individuation weakens, tight properties fail before loose ones. The hierarchy isn't stipulated; it follows from the precision demands of each property. No version of feature-bundle theory predicts this order. No version of prototype theory explains why the gradience has this particular shape.

Third, it explains stability and instability together. Quasi-count nouns are stable because of functional anchoring; *folks* is unstable because it lacks anchoring. *Pea* regularised because there was no anchor; *data* is drifting because *datum* is dying. The HPC account doesn't just describe which cases are stable – it explains why, and predicts which cases should be vulnerable to change.

The traditional question – "Is *cattle* count or mass?" – obscures the underlying dynamics. The HPC question is: "Where does *cattle* sit in the count basin, and what's holding it there?" That question has an answer. The first one doesn't.

9.12 Natural experiments

If the count cluster is maintained by causal mechanisms, removing or altering those mechanisms should degrade the cluster in predictable ways. In the physical sciences, we would test this by ablation: knock out a gene, sever a neural connection, and observe the failure mode. In linguistics, we rely on natural experiments: languages that "perturb" the system by packaging the relevant cues differently (see Doetjes (2012) for a typological overview).

These cross-linguistic comparisons function as ABLATION BY PERTURBATION. Languages don't typically delete a mechanism entirely; they reallocate or reweight it. A caution is required: real cross-linguistic work is messy. Unlike a controlled lab ablation, comparing languages involves multiple interacting variables (confounds). We can't simply "switch off" morphology while holding everything else constant. However, by triangulating across different languages that ablate different parts of the mechanism, we can strengthen the inference. If the cluster degrades exactly as predicted in three mutually distinct ways, the case for the causal mechanism is robust.

We consider three such perturbations.

9.12.1 Perturbation 1: The collective basin (Welsh)

Here, the mechanism perturbed is the default basin of attraction. Standard Average European languages like English force a binary choice: singular (unmarked) or plural (marked). This obliges speakers to default to the singular for singletons. But what if the default were different?

The prediction is straightforward: if the singular default is removed, nouns denoting aggregates should settle into a "collective" base state that resists count syntax.

Welsh offers the evidence. For a large class of nouns – specifically those denoting things that naturally occur in groups (animals, vegetables, small objects) – the morphologically simple base form is *collective* (conceptually plural). To refer to a single unit, speakers must add a SINGULATIVE suffix. For example, *adar* means 'birds' (collective); *aderyn* means 'a bird' (singulative).

Grimm (2018, p. 532) shows that Welsh base-form collectives behave exactly like English quasi-count nouns: they accept loose properties (quantifiers like *llawer* 'much/many') but reject tight properties (numerals) without singulative marking. The *cattle* profile isn't a quirk of English irregulars; it's the stable attractor state for aggregate nouns when the pressure to act as a singular is removed. English *cattle* is simply a Welsh collective trapped in a number-marking language.

This account offers a clear disconfirmation condition: if Welsh collectives accepted numerals directly (e.g., *three birds-COLL*) without individuating morphology, it would falsify the claim that individuation requires active morphosyntactic maintenance.

9.12.2 Perturbation 2: Weakened structural reinforcement

In this case, we ablate the obligatory structural reinforcement. English countability is reinforced by obligatory number marking on (most) count nouns in argumental use. What happens if we remove this reinforcement? The prediction is that the count cluster should either disperse or naturally dissolve, or the functional load should shift to a new location.

One outcome is REALLOCATION, as seen in Mandarin Chinese. Mandarin lacks obligatory noun number inflection. Without the constant *book/books* pulse trained by morphology, the inference engine builds the cluster differently. The functional load of individuation moves to the CLASSIFIER system. The tight/loose clustering shifts to the interface between the numeral and the noun: "general" classifiers (like *gè*) tolerate vague individuation, while specific classifiers demand precise shape properties. The mechanism remains, but the locus of coupling has moved.

Another outcome is WEAKENING, as argued for Halkomelem (Salish). Wiltschko (2008) argues that in this language, plural marking is not a functional head but an optional modifier. Speakers can say *three boy* or *three boys* with no truth-conditional change (p. 642). Unlike English, where the grammar forces a countability decision every time a noun is used, Halkomelem makes the decision optional. The prediction here is lower projectibility: without the obligatory morphological pulse, the "count" category should be less cohesive – and it is.

A third possibility is the NULL CASE. Some languages are claimed to lack both obligatory number and obligatory classifiers (e.g., Yoruba, Indonesian). These act as the control group. A pure-mechanism view makes a risky prediction: in the absence of *any* reinforcing mechanism, there should be no rigid count/mass HPC. Nouns should be TRANSNUMERAL (neutral), and semantic boundaries should not predict grammatical behaviour. If these languages turned out to have a rigid, English-style count cluster without the mechanism, the HPC account would be falsified.

The disconfirmation condition is finding a language with no morphosyntactic maintenance but high countability projectibility.

9.12.3 Perturbation 3: Reweighted semantics (Yudja)

Finally, we can perturb the segmental semantics. A stricter ablation would be to remove the semantic constraint that numerals count *atoms*.

The prediction is that if numerals can count portions directly, the "tightness" of the cluster should collapse.

Lima (2014) argues that Yudja (Juruna family, Brazil) represents exactly this state. In Yudja, numerals combine directly with notional mass nouns: *tx-abïu apeta* ('three blood') is grammatical and interpreted as 'three drops/spots of blood'. The constraint that numerals demand atomic units is relaxed or reweighted.

The result is a reordered hierarchy. Because the "tight" property (numerals) no longer demands high-precision atomic individuation, it becomes "looser" than it is in English. The failure mode here is NEGATIVE PROJECTIBILITY: the inference "Accepts Numerals → Is Atomic Object" becomes unreliable.

The disconfirmation condition is simple: if Yudja speakers processed 'three blood' as strictly atomic (coercing it to 'three blood-cells' or 'three vials') despite the lack of marking, it would show that atomicity is a cognitive universal independent of language-specific weighting.

9.12.4 The value of variation

These natural experiments do not show that countability "varies" in the abstract; they show that the variation is *mechanism-sensitive*. When a language reallocates where individuation cues live (number morphology vs. classifiers vs. singulatives), the correlational profile shifts in ways that are hard to describe as mere lexical accident. What we can legitimately claim on current evidence is conditional: *if* the count cluster is homeostatically maintained by particular couplings, then perturbing those couplings should produce systematic failure modes. Welsh collectives and singulatives are a clean case in point: when the base form encodes collectivity and the singulative does the individuating work, the English quasi-count profile stops looking idiosyncratic and starts looking like an attractor state for aggregate-denoting nouns.

The broader programme is triangulation. No single language comparison is a controlled ablation, and confounds are unavoidable. But if independent perturbations repeatedly relocate the same functional load – or weaken the same implicational tendencies – that convergence is exactly what a mechanism story predicts. The risky disconfirmation condition isn't "countability varies" but the stronger one: finding a language with no comparable morphosyntactic maintenance while retaining English-like projectibility of the count cluster.

We can watch the same tug-of-war in miniature in the reception of *emoji*. The borrowing entered English with an invariant form (as in Japanese), while the referents are maximally individuated: discrete, re-identifiable tokens. Predictably, usage pressures favour ordinary count packaging. Editorial policy has not been stable: the Associated Press – like the LEGO Group's lawyers fighting *Legos* – at one point endorsed *emojis* as the plural, and later reversed to recommend *emoji* as both singular and plural. The oscillation is itself diagnostic. Institutions can slow drift, but they rarely get to stipulate equilibrium points; they respond to them. The basin does not dictate what anyone must think; it dictates which packages become effortless.

9.13 Audit output

Table 9.2. Abbreviated HPC-kind audit for the count cluster.

Step	Content
1. Target & Scope	The count cluster in contemporary English; register-general; noun-level grain.
2. Profile	Singular forms, cardinal selection, quantifier patterns (*a(n)*, low cardinals, *many/few*), plural agreement. Properties form an implicational hierarchy from tight (precision-demanding) to loose (forgiving).
3. Stabilizers	Bidirectional inference (couples form to individuation); acquisition (overgeneralization transmits cluster); entrenchment (high-frequency chunks resist change); functional anchoring (singulative lexemes bleed pressure); institutional norms (prescriptive feedback).
4. Boundary behaviour	Quasi-count nouns (*cattle*, *police*) accept loose properties, reject tight ones. Object-mass nouns (*furniture*) show form-cluster without semantic individuation. These are predicted peripheral positions, not anomalies.
5. Failure-mode gate	The mechanism story is not analyst convenience: bidirectional inference is testable via semantic priming and acquisition overgeneralization. Classification: HPC kind.
6. Stress tests	(i) If quasi-count nouns show no correlation between functional-anchor presence and diachronic stability, reclassify them as frozen accidents. (ii) If a language lacks morphosyntactic reinforcement yet shows English-like count projectibility, the mechanism story is falsified.

9.14 Looking forward

Countability is the clean baseline case. Not because it's simple – the hierarchy, the quasi-count intermediates, and the diachronic equilibria are all genuine complexity – but because the coupling is unusually tight. A semantic variable – individuation, which languages get "for free" from domain-general object cognition – is repeatedly reinforced by overt morphosyntax (number, quantifiers, agreement), so the cluster is highly projectible and the maintenance story is comparatively easy to see.

The next cases stress the framework in a different way. Definiteness (Chapter 10) is an interface system too, but the form cluster and the function cluster may not line up as neatly. Articles, demonstratives, and possessives pattern together morphosyntactically, but the semantic work they do – identifiability, familiarity, uniqueness, deixis – doesn't always co-occur. The question won't be whether there *is* a cluster, but whether we're looking at one HPC with internal fissures or two partially coupled ones.

Lexical categories (Chapter 11) push in another direction. The noun/verb contrast looks strikingly stable cross-linguistically, while adjectives and adpositions vary in both inventory and behaviour. If lexical categories are HPC kinds, the mechanisms that maintain them can't be uniform: some must be deep and widely shared (acquisition pressures, constructional scaffolding), while others are local and reconfigurable. The payoff of the countability chapter was that a category can be both real and internally structured. The remaining chapters test the same claim under weaker couplings and noisier maintenance.

Definiteness and Deitality

Not only was it difficult for him to comprehend that the generic symbol dog embraces so many unlike individuals of diverse size and form; it bothered him that the dog at three-fourteen (seen from the side) should have the same name as the dog at three-fifteen (seen from the front).

—Borges, *Funes the Memorious* (1942)

10.1 The puzzle of ⟨the⟩

Consider two sentences:

(1) *She wears the veil.*

(2) *?She wears the hat.*

The first is natural. The second is marked – it demands a context: which hat? But the curious thing is that (1) doesn't demand any such context. We don't ask "which veil?" We understand, without prompting, that the sentence is about a practice, not a particular piece of cloth.

This is the WEAK DEFINITE puzzle. Standard accounts tell us that *the* signals uniqueness or familiarity – there's exactly one contextually salient referent, or the referent has already been introduced in discourse. But in *She wears the veil*, there's no unique veil and no prior mention. The article is doing something else.

Compare the related cases:

(3) *play the piano* – no particular piano

(4) *go to the hospital* (AmE) – not a specific hospital

(5) *The tiger is endangered* – the species, not an individual

(6) *The Pope visited Canada* – unique by role, not by discourse

Each uses *the*, but none involves the standard uniqueness or familiarity. The bus is never the same bus, and yet we keep giving it the same article. These aren't marginal exceptions; they're high-frequency, fully productive patterns (Carlson & Sussman, 2005; Poesio, 1994). A theory of *the* that treats them as anomalies risks missing the systematicity of the pattern.

The problem runs deeper than weak definites. Proper names like *Kim* and *London* are semantically definite – they identify unique, familiar referents – but they take no article in English. Bare plurals like *dogs bark* achieve generic reference without any marking at all. The mapping between form and function is systematic, but it isn't one-to-one. Definite without the article; the article without definiteness.

This chapter argues that we've been conflating two categories. One is morphosyntactic: a cluster of grammatical properties that travel together. The other is semantic: a cluster of interpretive properties related to referent identifiability. They correlate strongly – most definite referents get marked as such, and most marked forms signal definiteness. But they aren't identical. They're two HPCs maintained by different mechanisms, and their imperfect alignment produces exactly the puzzles that have troubled the literature. Definiteness is a different basin – shallower than countability, with slippage at the edges where the form–value coupling loosens.

10.2 One form, two values

The standard view has good reasons: *the* overwhelmingly correlates with identifiable referents, and most semantic theories of definiteness derive the correlation from compositional semantics. The puzzle isn't that the correlation exists – it's that it admits systematic exceptions.

The literature offers several competing diagnoses. Uniqueness-based accounts (Heim, 1991; Russell, 1905) treat *the* as encoding a presupposition that exactly one entity satisfies the description; exceptions become pragmatic accommodation or domain narrowing. If you have to keep saying "accommodation," you're probably just squinting. Familiarity-based accounts (Christophersen, 1939; Heim, 1982) require prior discourse introduction; apparent counterexamples involve bridging or situational salience. Domain-restriction accounts (Hawkins, 1978; Roberts, 2003) focus on how context delimits the set of candidates; definiteness becomes identifiability within a shared frame. Constructional accounts (Aguilar-Guevara & Zwarts, 2010; Carlson & Sussman, 2005) treat systematically non-unique uses as conventionalized frames with distinct semantics. Each preserves something – compositionality, presuppositional uniformity, pragmatic flexibility, lexical economy – but none explains why the same distributional profile surfaces across such different semantic profiles.

Here is the decision criterion: any account that keeps definiteness purely semantic still needs to explain why *the*, demonstratives, and genitives – but not *a* or *some* – share a distributional profile (existential resistance, partitive selection, hosting) that has nothing obvious to do with uniqueness or familiarity. Conversely, any account that treats weak definites as construction-specific semantics still owes an explanation for why the construction recruits exactly this morphological class. The two-cluster architecture isn't an alternative semantics; it's a demand that any semantics meet.

The split we need parallels what we saw with countability. Chapter 9 distinguished the INDIVIDUATION CLUSTER (semantic) from the COUNT CLUSTER (morphosyntactic). The two were coupled by bidirectional inference but maintained by different mechanisms. The same architecture applies here.

On the semantic side sits the DEFINITENESS CLUSTER: the interpretive properties that make a referent identifiable. These include FAMILIARITY (the referent is discourse-old), UNIQUENESS (there's only one candidate in the relevant domain), and IDENTIFIABILITY (the hearer can pick it out).[1] When all three align, we have a prototypically definite referent – *the cat we discussed yesterday*. ANAPHORIC RECOVERABILITY – the capacity to be picked up by a pronoun – is a downstream discourse affordance, not a core property.

On the morphosyntactic side sits what I'll call the FORM CLUSTER – or, in terminology we'll earn later, the DEITALITY CLUSTER. These are the grammatical properties that travel together in English determiners: resistance to existential *there* under neutral prosody, eligibility as the complement of partitive *of*, and suitability as hosts for nonrestrictive modification. When a determiner shows all these properties, it's at the centre of the form cluster – *the*, *this*, *those*, *my*.

[1]Specificity – whether the speaker has a particular referent in mind – is sometimes included, but it crosscuts the definite/indefinite distinction: indefinites can be specific (*I'm looking for a book – it's on my shelf*) or non-specific (*I'm looking for a book – any book*). Specificity tracks speaker knowledge; definiteness tracks hearer knowledge.

The key observation is that these clusters can decouple. Weak definites like *go to the hospital* show the form-cluster properties – they resist **There's the hospital down the street*, they work in *one of the hospitals* – but they lack the definiteness-cluster properties: no unique hospital, no familiar referent. Proper names show the reverse pattern: fully definite semantically, but lacking the form-cluster marking entirely.

This decoupling is what we'd expect if the two clusters are distinct HPCs – what Chapter 5 called CROSSCUTTING KINDS. They overlap substantially because the mechanisms that maintain them are historically and functionally linked. But they're not identical, and the gaps reveal the underlying structure. This bipartite analysis directly mirrors the architecture of the language faculty proposed by Jackendoff (2002): a dissociation between the formal system (syntax/phonology) and the conceptual system (semantics/pragmatics), where categories are defined by their stable correspondences rather than by internal unity.

The next sections characterize each cluster in turn: first definiteness (semantic), then the form cluster (morphosyntactic). Only then will we ask what couples them and what makes them come apart.

	+ Definiteness (Identifiability, Uniqueness)	– Definiteness (Semantic Indefiniteness)
+ Deitality (Form Cluster)	Core Coupled Items *the cat*, demonstratives, genitives	Form Without Meaning Weak definites (*go to the hospital*), Generics (*the tiger*)
– Deitality (No markings)	Meaning Without Form Proper names (*Kim*, *London*), definite bare plurals	Non-Members *a cat*, *some cats*, indefinite bare plurals

Figure 10.1. The decoupling of form and value. The diagnostics of the form cluster (deitality) and the semantic properties of the definiteness cluster usually align, but systematic dissociations reveal the underlying dual-cluster architecture.

10.3 The definiteness cluster

Definiteness, as a semantic category, is maintained by discourse-pragmatic mechanisms operating at multiple timescales. The cluster comprises several properties that typically travel together – but acquisition research reveals they are learned sequentially, not as a bundle. First you track what's been mentioned; then you restrict the domain; only then do you model what your addressee can pick out.

FAMILIARITY emerges first. Heim (1982) argued that definite noun phrases refer to entities already in the discourse model. Indefinites introduce new referents; definites presuppose existing ones. This explains the felicity of *I met a student. The student was tired* – the definite picks up what the indefinite introduced. Children master this by age 2–3: they use *the* appropriately for discourse-old referents before mastering when to withhold it for new ones (Rozendaal & Baker, 2008).

UNIQUENESS develops next: in the relevant context, exactly one entity satisfies the description. Russell (1905) formalized this as part of the logical form of definite descriptions; later work refined it to allow pragmatic restriction – uniqueness holds relative to a shared frame of reference (Hawkins, 1978). But 3-year-olds who correctly use *the* for familiar referents don't yet enforce uniqueness: Brockmann et al. (2018) found that children often fail to distinguish *the* from *a* when uniqueness is at stake. The uniqueness component stabilizes later.

IDENTIFIABILITY is the hardest: a definite referent is one the hearer can pick out. This doesn't mean they already know it – first-mention definites like *the first person on Mars* work perfectly well. It means the description is sufficient for identification *from the hearer's perspective*. This requires Theory of Mind: the speaker has to represent what the hearer knows. Children show "egocentric definiteness" – using *the* when they know the referent but the hearer does not – until around age 5–6, and the error correlates with ToM development (De Cat, 2013). Even adults with intact ToM make egocentric errors – participants in Director Task studies frequently select objects unknown to their interlocutor (Keysar et al., 2000) – suggesting that identifiability requires not just representational capacity but executive resources to suppress privileged knowledge in real time. Theory of Mind helps, but it's not a full-time employee.

ANAPHORIC RECOVERABILITY is the practical consequence: definite referents can be tracked across discourse. They're the natural antecedents for pronouns, the topics of subsequent sentences, the entities we can keep talking about.

These properties cluster because of how discourse works. Topics are identifiable; identifiable referents get tracked; tracked referents become unique within the conversation. But the developmental sequence – familiarity, then uniqueness, then identifiability – reveals that the components are cognitively dissociable. Three distinct mechanisms bind them:

- Discourse tracking (seconds to minutes) binds familiarity by maintaining the given/new distinction.
- Domain restriction (seconds) binds uniqueness by narrowing the quantificational domain to the relevant set.
- Theory of Mind (years) binds identifiability by distinguishing speaker knowledge from hearer knowledge.

The properties don't always co-occur. A cataphoric definite like *The idea that she quit surprised me* achieves uniqueness without prior familiarity – the complement clause provides the identifying content. Role definites like *the Pope* are unique by social role, not by discourse history. The cluster has a prototype (all properties present) and a periphery (some properties missing).

This matches the HPC signature: a family of properties that statistically co-occur, maintained by causal mechanisms operating at different timescales, with graded membership at the edges, and projectibility across the cluster. The definiteness cluster is a semantic HPC – and like the form cluster, it's learned in stages, with the full adult system emerging only when all the binding mechanisms are in place. Children acquiring articleless languages (Japanese, Mandarin) show comparable developmental trajectories for definiteness interpretation – familiarity before uniqueness before identifiability – despite lacking the morphosyntactic form cluster, suggesting the semantic cluster is cognitively prior.

10.4 The form cluster

The morphosyntactic cluster in English is defined by distributional diagnostics. These are grammatical tests, not semantic ones – they target structural behaviour, not meaning. Three diagnostics converge to define the cluster.

10.4.1 Existential ⟨there⟩

The definiteness effect is well established (Milsark, 1977): under neutral prosody, certain determiners resist appearing in the pivot position after existential *there*.

(7) * *There is the key/my key on the table.*

(8) *There is a key on the table.*

The constraint is sensitive to prosody. List intonation can rescue otherwise unacceptable pivots: *Well, there's THE KEY, the wallet...* But under neutral prosody, the constraint is robust.

10.4.2 Partitive ⟨of⟩

In true partitive constructions with subset semantics, the complement has to come from the form cluster:

(9) *Two of the students left.*

(10) *Several of these books are damaged.*

(11) * *Two of some students left.*

This structural constraint doesn't yield to prosodic manipulation.

10.4.3 Identificational hosting

Determiners from the form cluster are natural hosts for nonrestrictive modification, topics, and specificational subjects. *The book, which I bought yesterday, is excellent* is standard; *?A book, which I bought yesterday, is excellent* typically requires specific licensing or marks the referent as unique despite the indefinite form.

(12) *The/This/My book, which I bought yesterday...*

(13) *?A/Some book, which I bought yesterday...*

10.4.4 Convergence

Table 10.1 summarizes the diagnostic profile. The prototypical form-cluster determiners – *the*, demonstratives, genitives – show all three properties. Indefinite determiners show none. And some items show mixed profiles, exactly as the HPC framework predicts.

No single diagnostic defines the cluster. No property is necessary, no set sufficient. What matters is convergence – if the cluster is homeostatic, each diagnostic should be noisy; if each is noisy, none will be necessary; if none is necessary, convergence is what you should look for. The homeostatic pattern is exactly what essentialist accounts would need to treat as exceptional rather than predicted.

Table 10.1. Diagnostic profile for English determiners. ✓ = exhibits form-cluster behaviour; – = does not. Diagnostics: *there* = resists existential *there* under neutral prosody; Partitive = licenses partitive complement; Hosting = hosts nonrestrictive modification.

	there	Partitive	Hosting
Core Members			
the	✓	✓	✓
*this/that**	✓	✓	✓
Genitives	✓	✓	✓
Mixed/Peripheral			
each/every	✓	–	–
Non-Members			
a/an	–	–	–
some	–	–	–

*In deictic uses; narrative-presentational *this* (*There was this guy...*) is a derived-function construction that recruits demonstrative morphology for non-deictic use.

10.5 The coupling

We now have two clusters: definiteness (semantic) and the form cluster (morphosyntactic). Why do they correlate? And why isn't the correlation perfect?

The correlation arises because the clusters share a common origin. Cross-linguistically, definite articles arise primarily from demonstratives (Diessel, 1999; Greenberg, 1978). Demonstratives are inherently deictic – they point – and pointing presupposes a referent that both speaker and hearer can identify. As demonstratives grammaticalize into articles, this functional association persists. The article inherits a strong bias toward identifiable referents even as its semantic contribution generalizes.

Millikan (1984)'s PROPER FUNCTION framework clarifies the logic. The article's proper function – what it was selected for, what makes utterances containing it reproductively successful – is signalling identifiability. Speakers use *the* to direct hearers to a particular referent; when hearers recover that referent, communication succeeds; success sustains the convention. Under NORMAL CONDITIONS – the conditions under which the device performs its proper function – speaker and hearer share a discourse model, the referent is uniquely identifiable within that model, and the hearer can retrieve it.

But proper function isn't the whole story. Grammaticalization drags along more than meaning. As demonstratives become articles, they inherit distributional properties – what Himmelmann (1997) calls "context expansion." The form-cluster diagnostics (partitive licensing, *there*-resistance) were not what the article was *for*; they were side effects of the historical process. Once these side effects stabilize, they become available for recruitment. This is MECHANISTIC DRIFT – the category remains historically continuous while its causal underpinnings shift (Chapter 6).

This is where DERIVED PROPER FUNCTIONS arise. Derived functions exploit stable features of a device for purposes other than what it was selected for. Weak definites exploit the article's form-cluster membership – its distributional profile – without performing the identifiability function. Generic definites exploit the article's capacity to pick out a domain-restricted entity, but shift the domain from individuals to kinds. In both cases, the derived function is *parasitic*: it depends on the Normal function remaining intact for the convention to persist.

The parasitism explains why decoupling is tolerable. The overwhelming majority of *the*-tokens still perform the proper function – signalling identifiability. Weak definites and generic definites are minority uses. As long as the Normal function is performed often enough to sustain the convention, the derived functions can persist without undermining the system. The slippage is systematic, not accidental: it occurs precisely where stable side effects offer something to exploit.

10.6 THE MACHINERY OF MAINTENANCE

In Chapter 7, we saw the general machinery that maintains linguistic kinds. Here, we can see that machinery applied to the specific idiosyncrasy of the English article. The form cluster (deitality) isn't just a list of properties; it's a dynamic equilibrium maintained by at least five distinct pressures.

One mechanism is GRAMMATICALIZATION, which operates over decades to centuries. As demonstratives become articles, they carry their distributional properties along. English *the* descends from a Germanic demonstrative and retains demonstrative-like behaviour: resistance to existential pivots, partitive licensing, preferential hosting. The cluster wasn't stipulated; it was inherited. This provides the cluster's inertia.

The shared ancestry even leaves phonological traces. All /ð/-initial determiners in English – *the*, *this*, *that*, *these*, *those* – are form-cluster items derived from demonstrative sources. The phonological clustering and the distributional clustering have the same origin.[2]

[2] The phonological coherence itself likely acts as an additional stabilizing mechanism: a phonestheme that cues category membership and facilitates acquisition.

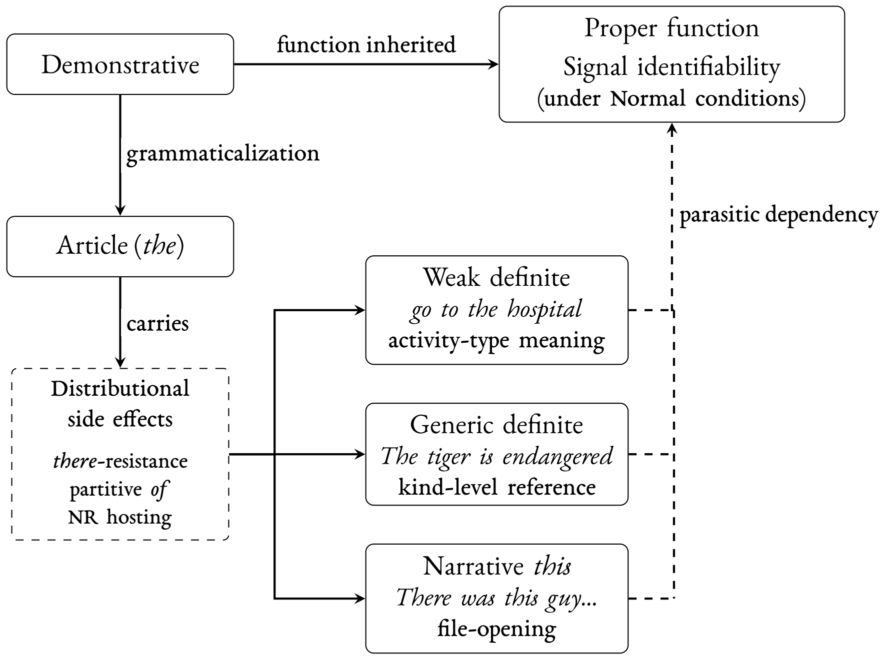


Figure 10.2. Derived proper functions of the English article. The demonstrative's function—signalling identifiability—was inherited by the article through grammaticalization. The process also carried distributional side effects (form-cluster properties) that were not what the article was *for*. Derived functions exploit these stable side effects for purposes other than identifiability. The dashed arrow indicates parasitic dependency: derived functions persist only because the Normal function remains robust.

Another mechanism is ACQUISITION, which occurs in childhood. Each generation relearns the cluster. Children acquire *the* early and overgeneralize it – using it where adults would use *a* (Maratsos, 1976; Rozendaal & Baker, 2008). But what matters most is that they learn the distributional frame before mastering the full semantic range. They know where *the* appears (determiner slot, partitive complement) before they know exactly when to deploy it. This creates a bootstrap: the form cluster serves as the scaffolding on which the definiteness concept is constructed.

Third, ALIGNMENT stabilizes the cluster in real time (seconds to minutes). In conversation, speakers converge on syntactic frames through interactive alignment (Pickering & Garrod, 2004b). When someone uses *the* in a weak-definite frame, their interlocutor accommodates. When someone violates a form-cluster constraint – producing **There's the solution* under neutral prosody – hearers flag the anomaly through hesitation or repair. This real-time feedback provides error-correction, stabilizing the cluster within conversations.

Fourth, PRESTIGE SELECTION determines survival over years to decades. Some variants spread faster because they're associated with high-status speakers (Labov, 2001). This shapes which weak-definite frames become conventionalized. British *in hospital* persists as a prestige variant; American *in the hospital* is the local norm. The mechanism provides protection for arbitrary variants, shielding them from regularization.

A fifth mechanism is TRANSMISSION, which acts as a filter over generations. The multi-generational bottleneck filters for learnability (Kirby et al., 2014). Patterns that are too complex or inconsistent don't survive. This explains why structural constraints converge across dialects: all English dialects reject **Which did you buy car?* because the constraint is simple and stable. Weak-definite frames vary more because they're conventionalized idioms. This mechanism acts as a filter, setting the outer bounds of what the cluster can contain.

The list is not exhaustive.

These mechanisms interlock. Grammaticalization creates the bundle; acquisition transmits it; alignment maintains it in real time; prestige shapes which variants survive; transmission filters for stability. Metalinguistic feedback adds another layer: grammarians' labels and diagnostics can feed back into pedagogy and prescriptive norms, potentially stabilizing or destabilizing particular constructions. Perturb any mechanism, and the equilibrium shifts – but the cluster persists because the other mechanisms compensate.

10.7 When the clusters slip

The explanatory power of the HPC framework shows when we examine cases where the two clusters dissociate. These aren't anomalies to be explained away – they're instances of derived-function exploitation. Each case follows the same logic: a stable side effect of the article's history gets recruited for a purpose other than its proper function.

The logic parallels Gricean pragmatics. Just as flouting the maxim of Quality ("You're a genius" said ironically) exploits the norm to generate new meaning, flouting the definiteness norm (using *the* without a specific referent) exploits the form to signal a different kind of content. The convention persists because the violation is systematic.

10.7.1 Weak definites

Weak definites look like a failure of definiteness. Expressions like *take the bus*, *listen to the radio*, and *go to the hospital* have troubled semantic theories for decades (Aguilar-Guevara & Zwarts, 2010; Carlson & Sussman, 2005). The puzzle is how *the* appears without unique or familiar referents.

The proper-function framework dissolves the puzzle. Weak definites exploit the article's form-cluster membership – its distributional profile – without performing its proper function (signalling identifiability). The morphosyntactic frame – verb + *the* + noun in this conventionalized construction – persists because the form-cluster properties are independently stable. What the construction does is semantic: it shifts the contribution from a specific referent to an activity type. *Go to the hospital* means something like "seek medical treatment"; the hospital itself is not identified.

This is a derived function. The article's distributional behaviour was inherited from the demonstrative source – a side effect of grammaticalization, not what the article is *for*. Once that side effect stabilized, speakers could recruit it for institutionalized activity frames. Weak definites exhibit MECHANISTIC DRIFT (Chapter 6): the semantic licensing – identifiability of a unique referent – has eroded, but the morphosyntactic pattern persists, maintained now by entrenchment and constructional convention rather than by the original function. But weak definites aren't merely fossils: the construction is productive. We *take the Uber*, *check the app*, *get on the WiFi*. The pattern will happily adopt any new ritual, especially if it comes with a subscription. New forms enter the pattern whenever a stereotypical activity arises. What persists is the frame – verb + *the* + role-denoting noun – recruiting form-cluster membership for activity-type semantics.

10.7.2 Generic definites

Generic definites look like a failure of reference. Sentences like *The tiger is endangered* or *The computer changed the world* pose a different challenge: the definite singular is used for kind reference, not individual reference (Carlson, 1977; Krifka, 2004).

This is the puzzle that troubled Funes in the epigraph. The "generic symbol dog" embraces so many unlike individuals – the dog at three-fourteen seen from the side, the dog at three-fifteen seen from the front – and yet a single word holds them together. What Funes couldn't accept was exactly what generic definites achieve: a label that picks out not an individual but a kind. The mechanisms that maintain DOG as a category are the same mechanisms that let *the dog* do its work.

Again, the HPC framework provides clarity. Generic definites are form-cluster items: they resist neutral existential contexts, they pattern distributionally with *the*. But their semantics is kind-denoting, not individual-referring. Individual-level uniqueness doesn't apply; kind-level properties do.

The form cluster is intact. The definiteness cluster is orthogonal – neither satisfied nor violated, merely inapplicable at the kind level. The syntax generates a form-cluster phrase; semantics assigns it a generic interpretation that sidesteps the definiteness question.

This too is a derived function. Generic definites exploit the article's capacity for domain restriction – inherited from demonstrative pointing – but shift the domain from individuals to kinds. The proper function (signalling identifiability of an individual) isn't performed; what persists is the form-cluster profile and the slot for domain specification.

10.7.3 Proper names

Proper names (as distinct from PROPER NOUNS, Chapter 11) look like a failure of marking in English. *Kim* and *London* are semantically definite – they identify unique, familiar referents – but they typically lack form-cluster marking. This isn't universal: in Greek (*o Petros* 'the Peter') or with titles in Spanish (*el señor Costa* 'the Mr. Costa'), the form and meaning clusters align. But English dissociates them.

This dissociation follows from the HPC architecture. In English, names show definiteness-cluster properties (identifiability, uniqueness, anaphoric recoverability) without the form-cluster properties. They don't resist existential pivots: *There's a Kim here to see you* is natural. They don't license partitives in the standard way. Or rather, their INDEXICAL ANCHORING (Chapter 7) is achieved directly through the label, rendering the deital pointer redundant.

The clusters have decoupled in the opposite direction from weak definites. Where weak definites have form without value, proper names have value without form.

English even reveals the underlying logic when names collide with constructions that have indefinite defaults. Bare plurals are normally indefinite: *dogs*, *books*. Family names need pluralization but remain definite: *the Smiths*, *the Johnsons*. The article overrides the constructional default to preserve definiteness.

The same dissociation appears when we push names toward generic readings. Predicativists analyse bare singular names like *Ruth* as containing a null definite determiner – structurally parallel to *the tiger*. If so, we'd expect bare singular names to allow generic readings, just as *the tiger* does. Gasparri (2025) shows they can: *Italian Andrea is generally male*; *Ruth has good grades in biology* (in statistical contexts). But the generics are characterizing, not kind-level – **John became common* fails without quotation.

Why the restriction? Names don't pick out natural kinds; they pick out what Dupré (1993) calls "social-practice-unified collections" – the set of individuals who bear a name in virtue of naming conventions. Characterizing generics quantify over instances: *Italian Andrea is generally male* says most Italians named Andrea are male. Kind-level generics require a kind as argument: *The tiger is endangered* takes the species as subject. But Ruth isn't a species; it's a practice-maintained cluster of individuals. The form (null-definite + name) decouples from canonical referential function, but only within limits imposed by what names are for.

10.7.4 Indefinite ⟨this⟩

Narrative-presentational *this* (*There was this guy...*) is the mirror image: an indefinite meaning recruiting a form-cluster item (Prince, 1981). Prince (1981) showed that indefinite *this* introduces referents that will become topics – it uses the form cluster's "hosting" potential to signal future prominence. The item is semantically indefinite (new) but morphosyntactically deital (partitive-licensing, nonrestrictive-hosting). The mismatch is the mechanism: using a pointer form for a new referent forces the hearer to open a file that expects updates.

10.8 Passing the tests

Chapter 8 introduced the Two-Diagnostic Test for genuine HPC kinds: high projectibility and robust homeostasis. Both clusters pass.

10.8.1 Projectibility

The definiteness cluster supports induction. Knowing a referent is semantically definite lets you predict its discourse behaviour: it can be picked up by pronouns, it can be topicalized, it can be presupposed. The form cluster supports induction too: knowing a determiner is form-cluster lets you predict structural behaviour across contexts.

Most importantly, the predictions are dissociated. You can know a noun phrase's form-cluster status without knowing its definiteness status – and the predictions you make will differ. Weak definites are form-cluster but not definiteness-cluster; proper names are definiteness-cluster but not form-cluster. Each category supports distinct inductions.

10.8.2 Homeostasis

Both clusters are maintained by mechanisms. The definiteness cluster is sustained by discourse-pragmatic processes: common ground management creates correlations among identifiability, uniqueness, and anaphoric recoverability. The form cluster is sustained by mechanisms such as those detailed in §10.6.

The mechanism difference explains the decoupling. Discourse pressure can't alter distributional restrictions; grammaticalization pressure doesn't change reference. The clusters drift independently because their maintenance is independent.

10.8.3 Falsifiable predictions

The HPC account generates specific predictions about how the clusters behave under manipulation:

Prosodic rescue: Prosody should selectively rescue existential pivots (a discourse-level constraint) but not partitive complements (a more deeply grammaticalized structural constraint). This is testable and falsifiable.

Dialectal preservation: Dialects that drop *the* in institutional frames (British *in hospital*) should preserve form-cluster patterning when definiteness is supplied by demonstratives (*in this hospital*).

Acquisition asymmetry: Children should master distributional restrictions before semantic ones, because distributional restrictions lack prosodic repair paths.

These predictions target the interaction between morphosyntax, prosody, semantics, and development – exactly what the mechanism story predicts.

What would falsify the account? If the clusters fragment completely rather than cohering into two families. If *there*-resistance, partitive licensing, and hosting requirements turned out to be independent rather than correlated, there would be no form cluster to explain. If identifiability, uniqueness, and familiarity showed no statistical tendency to co-occur, there would be no definiteness cluster. The HPC architecture predicts two cohesive families that occasionally decouple, not a swarm of independent features. The cohesion is the explanandum; the mechanisms are the explanans.

10.9 The term: Deitality

We need a name for the form cluster. The term DEFINITENESS is already taken – it names the semantic cluster. Calling both clusters "definiteness" would perpetuate exactly the conflation we have been trying to undo.

The conflation has a history. When Russell analyzed *the F* descriptions as quantificational (existence plus uniqueness), the article's distributional diagnostics were not the object of analysis – the truth-conditional contribution was. Later work refined the semantics, but the conflation persisted, converging with a pre-Russell grammatical tradition that treated "definiteness" as naming both the meaning and the marking. The slide was reinforced from other directions too – grammatical description, language pedagogy, typological comparison – often treating form and value as a single package. The term did double duty, and its equivocation became invisible. Weak definites that didn't fit the semantic story typically became exceptions to be explained rather than evidence that two categories had been conflated.

I propose DEITALITY as a diagnostic remedy. The term derives from the deictic origins of the cluster: demonstratives grammaticalize into articles, and the distributional profile reflects that deictic source. A determiner is DEITAL if it shows the form-cluster properties; it's DEFINITE if it contributes definiteness-cluster semantics.

The term is ugly – methodologically ugly. If it sounds like a minor villain in a low-budget sci-fi, that's partly the point. Its unfamiliarity forces a break with the assumption that form and value are identical. When you say *deital*, you can't accidentally mean *definite*. The conceptual separation becomes lexicalized. This is a terminological intervention, and it has costs: new labels require uptake. But the alternative – continued equivocation on *definiteness* – has higher costs still.

The utility is in the clarity. We can now say: weak definites are deital but indefinite; proper names are definite but non-deital; generic definites are deital and semantically orthogonal. Each statement is precise in a way the traditional terminology doesn't allow.

10.10 Cross-linguistic scope

Is deitality an English-specific category or a universal potential?

The form cluster as described here – the specific diagnostics, the specific determiners – is English-specific. It reflects the grammaticalization history of English demonstratives, the selectional restrictions of English constructions, the conventions of English discourse.

But the ARCHITECTURE generalizes (Lyons, 1999). Languages with demonstrative-derived articles (French *le*, German *der*, Greek *o*) show similar clusters: partitive restrictions, information-structural constraints, imperfect correlation with semantic definiteness. The diagnostics differ; the pattern of convergent distributional properties maintained by grammaticalization persists. Greek illustrates the recombination: proper names *require* the definite article (*o Giannis*, not **Giannis*), unlike English, where names lack form-cluster marking entirely (Lyons, 1999). Same definiteness-cluster semantics; different form-cluster mapping. The architecture isn't an English projection – it's a framework for seeing where the pieces recombine differently.

Classifier languages show a different realization of the same underlying dynamic. Japanese lacks articles but has a rich demonstrative system (*kono/sono/ano*) with its own distributional profile. The semantic definiteness cluster exists cross-linguistically; the morphosyntactic form cluster takes language-specific shapes.

The prediction is that wherever demonstratives grammaticalize into articles, the form cluster should emerge – because grammaticalization drags distributional properties along. The cluster isn't stipulated; it's an emergent consequence of how grammaticalization works.

10.11 Audit output

Target and scope. The deitality cluster (form cluster) and the definiteness cluster (semantic cluster) in contemporary English; register-general; determiner-level grain.

Profile and stabilizers. The form cluster comprises *there*-resistance, partitive licensing, and nonrestrictive hosting. It is maintained by grammaticalization (demonstrative inheritance), acquisition (distributional frame learned before full semantics), alignment (real-time repair), prestige selection (dialectal variants), and transmission (multi-generational filtering). The definiteness cluster comprises familiarity, uniqueness, and identifiability. It is maintained by discourse tracking, domain restriction, and Theory of Mind development.

Boundary behaviour. Weak definites (*go to the hospital*) show form-cluster without definiteness-cluster. Proper names show definiteness-cluster without form-cluster. Generic definites are form-cluster and semantically orthogonal. These decouplings are predicted by the two-cluster architecture, not anomalies.

Failure-mode gate. The mechanism story is not analyst convenience: grammaticalization pathways are historically documented; acquisition asymmetries are experimentally attested. Classification: two coupled HPC kinds.

Stress tests. (i) If weak definites cluster with token-referring definites in embedding space rather than forming a distinct subcluster – tracking frequency or genre rather than the form-cluster diagnostics (*there*-resistance, partitive licensing) – reclassify deitality as a thin pattern within definiteness. (ii) If the form-cluster diagnostics fragment completely – *there*-resistance, partitive licensing, and hosting showing no correlation – there is no form cluster to explain. (iii) In acquisition data, if children show no asymmetry between learning the distributional frame (deitality) and the semantic licensing conditions (definiteness), the two-cluster architecture is falsified – they would be learning one category, not two.

10.12 Looking forward

We have now examined two variations on the HPC architecture. With countability, the fit between form and meaning is tight: the count cluster hugs the individuation cluster so closely that they are often mistaken for a single category. With definiteness, we found a fissure. Deitality tracks definiteness, but the coupling is loose enough to permit systematic, productive slippage. The "exceptions" that have plagued the literature – weak definites, generics, proper names – are simply the visible evidence of play in the joint.

The next case study tests a different configuration. English gender is usually described as a vestigial system – a three-way pronoun distinction (*he*/*she*/*it*) that lacks the NP-internal concord of French or German. But the same personhood-based logic that governs pronouns also governs *who*/*which*, *somebody*/*something*, and *when*/*where*. Chapter 12 shows that English gender is a robust HPC system once we recognize its proper scope: not pronouns alone, but the entire semantic class of pro-forms.

Lexical categories and their maintenance

The resemblance of one animal to another is of exactly the same essential nature as the resemblance to a leaf, or to bark, or to desert sand, and answers exactly the same purpose.

—Alfred Russel Wallace, *Mimicry, and Other Protective Resemblances Among Animals* (1867)

11.1 The mess

Quickly. Very. Somewhat. Absolutely. Frankly. Apparently. Probably. Already. Still. Only. Even. Not. Moreover. However. Therefore. Otherwise.

What do these words have in common? The dictionary says they're adverbs. But what does that mean? They modify different things: *quickly* modifies manner, *very* and *somewhat* modify degree, *frankly* signals speaker stance, *apparently* and *probably* mark epistemic status, *already* and *still* locate events in time, *only* and *even* are focus particles, *not* negates, *moreover* and *however* organize discourse, *therefore* marks inference. Their distributions overlap only partially. Their semantic contributions share almost nothing. The label *adverb* gathers them like socks in a drawer—not by what they are, but by what they aren't. They're not nouns; they're not verbs; they're not adjectives. What remains is the bin.

And then there's *otherwise*—the word that kept Rodney Huddleston awake in Chapter 1. Dictionaries call it an adverb, but it can also be predicative (*the truth is quite otherwise*), a conjunct (*otherwise we'll be late*), and something close to a preposition in *this suggests otherwise*. The word wouldn't hold still then; it still won't. That's not a bug in the analysis. That's the wastebasket showing its seams.

This chapter is a dissection. We will find what holds together and what doesn't.

I won't catalog every lexical category. The goal is to ask a different question: what makes a category cohere—or fail to cohere? We'll examine four major classes (nouns, verbs, adjectives, adverbs) and one apparent class (pronouns) that turns out to be several. The closed functional categories—determiners, prepositions, conjunctions—are a different story, better left to typological specialists (Haspelmath, 2010). Our business is with the mechanisms, and the mechanisms are clearest where they succeed spectacularly (nouns and verbs), fail instructively (adverbs), or produce an illusion that rewards closer inspection (pronouns).

The structure follows the anatomist's table. First, the wastebasket: the category that *doesn't* hold together. Then the skeleton: the categories that hold together so well they appear in every language ever described. Then the mimics: the forms that look alike on the surface but turn out to be different animals underneath, convergently evolved to occupy the same ecological niche. Some basins are deep (nouns, verbs); others are shallow (adjectives); others are illusions born of convergent pressure (pronouns).

11.2 The wastebasket: adverbs

Start with failure. If lexical categories are HPC kinds maintained by mechanisms, then a class with no unified maintenance regime should fall apart under scrutiny. Adverbs are that class.

The traditional definition is distributional: adverbs modify verbs, adjectives, or other adverbs. But this is too broad. *Yesterday* modifies a clause, not a verb; *even* is a focus particle; *however* is a discourse connective; *very* is a degree word; *frankly* is an evidential marker. We've lumped together half a dozen functional types because they share one negative property: they can't be the head of a noun phrase or predicative complement. Excellent storage capacity; terrible explanatory power.

From a teleosemantic perspective (Millikan, 1984), a category should have a *proper function*—the job its members were selected to do. Nouns refer; verbs predicate; adjectives attribute properties. What is the proper function of an adverb? The question has no good answer. Manner modifiers (*quickly*) describe how an event unfolds. Degree modifiers (*very*, *slightly*) scale a gradable property. Evidential markers (*apparently*, *frankly*) signal the speaker's epistemic stance. Discourse connectives (*however*, *moreover*) structure information flow. There is no unified *for*.

Grammaticalization reveals the heterogeneity. Manner adverbs in English derive from adjectives via *-ly*, and their syntax still resembles adjectival modification. Time adverbs like *today* and *tonight* are etymologically noun phrases (cf. Old English *tōdæge*, 'on this day'), and they retain nominal properties: they resist determiners, they can serve as temporal anchors, they combine with prepositions only marginally (*since yesterday* vs. *since quickly). Focus particles like *even* and *only* pattern with focus-marking constructions; they can modify almost any constituent (*even Kim*, *only the red one*, *I only eat fish*), a distribution no manner adverb shares. Each subclass has its own history, its own syntactic home, its own path into the language. The category *adverb* is a reunion of strangers who happen to share a bus stop.

Simon Kirby's iterated-learning framework (Kirby et al., 2008) makes a prediction here: heterogeneity is unstable. Languages evolve toward learnability; if a category lacks a coherent core, learners should either fragment it into tighter subcategories or regularise it by eliminating outliers. What keeps *adverb* intact is not homeostasis but frequency. *Very*, *not*, *just*, *only*, *even*, *always*, *never*—these are among the highest-frequency words in the language (Davies, 2008). They survive as memorised tokens, not as projections from a productive schema. Take away the fossil collection and the category dissolves.

This is the diagnostic from Chapter 8: *adverb* is a *fat* category. It groups items by a distributional convenience, not a principled clustering. It fails the homeostasis test: perturbation in one region doesn't propagate. If children stopped learning *however* tomorrow, it would not affect the acquisition of *quickly*; if *very* disappeared, *yesterday* would carry on undisturbed. The properties don't reinforce each other; the subclasses don't share a mechanism. The category has a name but not a nature.

As Culicover and Jackendoff (2005) argue in their "Simpler Syntax" program, much of what is traditionally attributed to complex syntactic structure can be more parsimoniously explained through the interaction of relatively simple syntactic categories with rich semantic and pragmatic constraints.

Compare this to what follows.

11.3 The skeleton: nouns and verbs

Every language ever described has nouns and verbs. Not everyone agrees about what exactly *makes* a noun a noun or a verb a verb, but the categories themselves are cross-linguistic constants. This is a remarkable empirical fact. It demands explanation.

The distributional profile of a prototypical noun—say, *dog*—comprises a cluster of properties that reinforce each other. Nouns can function as arguments: subjects, objects, obliques. They take determiners. They inflect for number. They participate in possessive constructions. They trigger agreement. They are the default category for introducing new referents into discourse. These properties don't follow from each other by logical necessity, but they correlate with such regularity that typologists can use any one of them to identify the rest (W. Croft, 2001). That correlation is the signature of homeostasis.

Verbs show the same clustering on a different axis. They serve as predicates. They take arguments. They inflect for tense-aspect-mood. They trigger subject agreement. They combine with auxiliaries. They are the default category for asserting what is going on. Again: the properties don't entail each other, but they cohere so tightly that the presence of one licenses inferences about the others.

Why should this be? The standard formalist answer appeals to Universal Grammar: nouns and verbs are categories built into the human language faculty. But that answer merely restates the observation. Why should the language faculty include exactly these two major lexical categories and not others? Why these clusters and not some rearrangement?

The HPC answer is causal. Nouns and verbs are stable because they are maintained by converging mechanisms at multiple scales.

At the *cognitive* scale, nouns and verbs track a fundamental conceptual distinction: *things* and *events*. Developmental psychology shows that infants distinguish objects from actions before they have words for either (Spelke & Kinzler, 2007). The perceptual difference is not linguistic in origin; it's part of domain-general cognition. But because the distinction is cognitively available, languages consistently recruit it. Nouns are the natural home for object concepts; verbs are the natural home for event concepts. The fit between conceptual structure and grammatical structure is not accidental. The categories are projectible because they carve the world at a joint the mind already recognizes.

At the *discourse* scale, the noun/verb distinction tracks a fundamental communicative function: *reference* and *predication*. Discourse moves forward when speakers introduce referents and say things about them. Nouns provide the referents; verbs provide the predications. This is not a contingent fact about some languages; it's a constraint on any system that conveys propositional content. A language that lacked the distinction would lack the machinery for basic assertion. The functional pressure is universal, and the grammatical categories that serve it are correspondingly robust.

At the *acquisition* scale, the noun/verb distinction is learned early and transferred broadly. Children acquiring English generalise nominal morphology to novel nouns and verbal morphology to novel verbs (Tomasello, 2003). Errors like *I goed* or *two sheeps* show that children have abstracted a productive schema, not just memorised individual forms. This is the CREATIVITY signature (Chapter 7): the system generates novel outputs that go beyond the input data. Crucially, they don't extend nominal morphology to verbs or vice versa: the categories are represented as distinct, and their properties cluster together in the child's grammar as they do in the adult's. This is the *projectibility* diagnostic from Chapter 6: new items are slotted into the existing category and inherit its properties. The clustering survives transmission.

At the *typological* scale, the noun/verb distinction appears in every language we have records for. This is not proof of universality—we haven't examined every possible language, and some analyses of Salish languages have questioned whether the distinction is grammatically obligatory (Kinkade, 1983). But even in languages where the evidence is contested, the semantic domains of objects and events are consistently distinguished; the debate is about whether the distinction is encoded in morphology, in syntax, or in both. The robust cross-linguistic tendency is itself evidence of mechanism: whatever forces produce noun/verb stability in English also operate in Mandarin, in Swahili, in Warlpiri (Haspelmath, 2010).

The contrast with adverbs is stark. Nouns and verbs are tight HPC kinds: perturbation in one property propagates to the others, acquisition generalises across the category, and the cluster is stable across languages and registers. Adverbs are not: the subcategories are functionally independent, learning is item-specific, and cross-linguistic comparison shows enormous variability. The difference is not gradual. It is architectural. Nouns and verbs have thick braids of mechanism holding them together; adverbs have a label.

Carl Zimmer once observed that sharks and dolphins look alike because physics imposes a penalty on drag: streamlined bodies are hydrodynamically efficient, regardless of ancestry (Zimmer, 1998). The parallel here is instructive. Nouns and verbs look alike across languages because communication imposes pressure on form: a system that marks reference and predication will be easier to learn, easier to use, easier to transmit. The categories are not Platonic forms; they are standing waves maintained by converging forces. What persists is not the category as a thing; what persists is the maintenance regime.

11.4 The asymmetry: adjectives

Adjectives occupy the middle ground. They are tighter than adverbs but looser than nouns or verbs. Some languages have them; others do without.

In English, adjectives form a recognizable cluster. They appear in attributive position (*the tall building*), predicative position (*the building is tall*), and postpositive position (*something tall*). They take degree modifiers (*very tall*). They inflect for comparison (*taller*, *tallest*). They can be nominalised (*the young*, *the poor*). The properties co-occur with enough regularity that we can project them: a novel adjective will be expected to participate in all of them.

But the clustering is weaker than for nouns and verbs. Not all adjectives accept comparison (**more dead*, **most unique*). Not all occupy all three positions (**the asleep child*, **the mere building*). Derived adjectives like *computational* resist degree modification (**very computational*). The cluster has gaps. Its edges fray.

Cross-linguistically, the variability is dramatic. Many languages encode property concepts as verbs: Mandarin *tā hěn gāo* ('he very tall') uses an unmarked stative predicate, not an adjective plus copula. Others encode them as nouns: in Polynesian languages, colour terms pattern morphosyntactically with kin terms and body parts. Some languages have only a small, closed class of "true" adjectives—perhaps a dozen words—while property modification is otherwise handled by relative clauses or denominal constructions (Dixon, 2004). The category *adjective* is not a cross-linguistic universal. It is, in Martin Haspelmath's terminology, a COMPARATIVE CONCEPT: a yardstick researchers bring to the data, not a category found in every language (Haspelmath, 2010).

What explains the asymmetry?

The mechanisms that stabilize nouns and verbs are weaker for adjectives. At the cognitive scale, property concepts are parasitic on object concepts: *big* needs a referent to be big; *red* needs a surface to be red. Adjectives modify; they don't stand alone. The conceptual grounding is derivative rather than fundamental (Givón, 2001). At the discourse scale, property attribution is optional: you can identify a referent and predicate an event without ever mentioning a property. The communicative pressure is weaker; the grammatical reflex is correspondingly less robust.

Semantic type matters here. R.M.W. Dixon's typological work (2004) shows that languages differ in which property types receive adjectival treatment:

- Dimension (*big*, *long*): high frequency, scalar, degree-compatible. These are the most stable across languages.
- Age (*young*, *old*): relational to temporal deixis, may pattern with participles or stative verbs.
- Colour (*red*, *green*): often derived from nouns (*gold*, *ash*), sometimes form a small closed class.
- Value (*good*, *bad*): highly evaluative, prone to grammaticalization, unstable clustering.

Each type has a different mechanism profile. Dimension adjectives show the tightest fit with degree morphology; colour adjectives show the greatest variability across languages. The category *adjective* is not a single cluster but a family of overlapping subclusters, held together more loosely than nouns or verbs.

To use Zimmer's zoological metaphor: if nouns and verbs are the skeleton, adjectives are the plumage. Some birds have exploded this category into a massive display (English). Others have almost none (languages that use verbs for properties). Plumage is not essential to flight; the skeleton is. Adjectives are an evolutionary luxury. They emerge where discourse economy creates a niche for property modification, but the niche isn't deep enough to guarantee the category's presence in every language (W. Croft, 2001).

The HPC framework clarifies the difference. Adjectives are a thinner HPC kind than nouns or verbs. The clustering is real, and it supports genuine projectibility: if I coin *blick* as an adjective, you will expect it to take *very*, to appear attributively, to compare. But the mechanisms are weaker, the cross-linguistic stability is lower, and the category is liable to dissolve into verb-like or noun-like strategies depending on the language's historical trajectory. Where nouns and verbs are architectural load-bearers, adjectives are decorative but disposable.

11.5 The mimics: convergent evolution in pronouns

Now for the surprise.

Consider three items: *today*, relative *who*, and interrogative *who*. They look different on the surface—one is temporal, one is an operator, one starts a question. But they share a striking distributional property: they all resist determiners.

(1) * *the today* / **a today*

(2) * *the person the who called*

(3) * *The who called?*

Why should these three items pattern alike? Traditional accounts would classify them as pronouns or pro-forms and leave it at that. But the classification masks a deeper question: what *makes* them resist determiners?

Return to Wallace's observation about mimicry. A shark and a dolphin look remarkably similar: streamlined bodies, dorsal fins, pectoral flippers. But they are not related. Sharks are cartilaginous fish; dolphins are mammals. They look alike because physics imposes a penalty on drag. Water is dense; to move efficiently through it, you need a certain body shape. The constraint is external, not genealogical. Sharks and dolphins are *convergently evolved*: they arrived at the same phenotype through different ancestral paths because they face the same environmental pressure.

Now consider the three "pronouns." They resist determiners for different reasons:

- Deictics (*today*, *here*, *she*): These are *indexicals*. Their reference is fixed by context—the speaker, the time, the place of utterance. A determiner would be redundant: you don't need *the* to pick out a referent when the deictic anchor already does the job. The saturation is contextual, and contextual saturation makes the determiner slot empty. The proper function (Millikan's term) is to anchor reference to demonstrative space.

- Anaphors (relative *who*, *which*): These are *bound variables*. Their reference depends on an antecedent in the discourse; the relative clause restricts the set. A determiner would violate the dependency: you can't simultaneously restrict by antecedent and by article. The saturation is structural, and structural saturation blocks the determiner. The proper function is to create a bound-variable dependency.

- Interrogatives (*who?*, *what?*): These are *variables over alternatives*. They don't refer to a single entity; they range over a set of candidates. A determiner would impose external restriction on what is already a restrictor. This is why *the who* sounds wrong: not because it's redundant (as with deictics), but because you're trying to restrict a restrictor. The proper function is to delineate alternatives.

Three different functional mechanisms; one shared distributional property. The categories *look* alike because they face the same grammatical pressure: referential saturation makes an overt determiner unnecessary or incoherent. But the saturation is achieved by different routes—contextual, variable-bound, alternatives-based. The surface unity is real; the underlying mechanism is *braided*.

Table 11.1. Saturation routes for determiner resistance

Category	Analogy	Saturation route	Proper function
Deixis (*today*)	Shark	Context	Anchor to demonstrative space
Anaphor (rel. *who*)	Dolphin	Bound variable	Bound-variable dependency
Interrogative (*who?*)	Ichthyosaur	Alternatives	Restrict quantifier domain

This is convergent evolution in grammar. Sharks, dolphins, and ichthyosaurs (extinct marine reptiles) all evolved streamlined bodies—three lineages, one hydrodynamic optimum. Deictics, anaphors, and interrogatives all resist determiners—three functional routes, one distributional outcome. Calling them all "pronouns" is like calling sharks, dolphins, and ichthyosaurs all "fish" because they swim. It captures the phenotype but misses the ancestry. In Khalidi's terms (Chapter 5), these are CROSSCUTTING KINDS: the same entity belongs to multiple causal networks simultaneously, and the surface overlap does not entail a shared maintenance regime.

Why does this matter? Because category stability depends on mechanism. Deictics, anaphors, and interrogatives are CROSSCUTTING KINDS – overlapping phenotypes, distinct mechanisms. If *pronoun* were a unified HPC kind, we would expect perturbation in one member to propagate to the others. But it doesn't. If children stopped learning interrogative *who*, deictic *she* would be unaffected. The properties don't reinforce each other crossing the functional boundaries. The convergence is external, imposed by the shared pressure of referential saturation, not internal, maintained by a single homeostatic core.

Ruth Millikan's proper-function framework (1984) makes the distinction sharp. A proper function is what a form was selected for doing—not what it happens to do, but what it *does in order to have been transmitted*. If determiner resistance arises from three different proper functions, then we have three different category-maintaining regimes overlapping on one distributional property. The unity is a scar of selection, not a structural essence.

The *grammar* sees determiners as blocked; the *cause* differs. This dissociation is the signature of braided mechanisms.

11.6 The acquisition twist

If braided mechanisms are real, acquisition data should show the strands before the braid.

Michael Tomasello's usage-based research (2003) suggests exactly this. Children don't learn "pronoun" as a unified category. They learn *I* as part of the deictic system, *what* as part of the interrogative frame, and relative *who* as part of specific constructions like *the man who…*. The learning paths are independent. The cross-strand generalisation—the recognition that all three resist determiners—comes later, if it comes at all.

Consider the acquisition of *today*. Children hear it in highly formulaic contexts: *What did you do today?*, *Today we're going to the park*. They don't need to know it's a "pronoun"; they need to know it fills a temporal slot. The determiner resistance is implicit in the input: no child ever hears **the today*. The constraint is absorbed incidentally, not extracted from a rule.

Now compare interrogative *who*. Children learn it as part of the question-forming construction: *Who ate the cake?*, *Who is that?* The *wh*-word is locked into the sentence-initial position; the determiner slot is irrelevant. Again, the absence of **the who* is absorbed through negative evidence—through never hearing the ungrammatical form—not through learning a general principle about pro-forms.

What about overgeneralisation errors? If children extract a unified *pronoun* category, we'd expect cross-strand errors: **the who* by analogy with *the man*, or **a today* by analogy with *a day*. Such errors are vanishingly rare. Ben Ambridge's entrenchment research (2020) shows that children are sensitive to what *hasn't* been heard: pre-emption is a statistical learning mechanism. If *the today* never occurs, and *today* occurs frequently, the absence itself is informative. But the preemption operates locally, within the deictic paradigm or within the interrogative paradigm. There is no evidence for cross-strand transfer.

This is the prediction for braided categories. Learn the strands independently; let the surface convergence emerge from shared environmental pressure. If someone asks: "When do children learn that pronouns resist determiners?"—the answer is that they probably never learn it *as a generalization*. They learn that *she* resists determiners, that *who* resists determiners, that *today* resists determiners. The unifying pattern is a linguist's abstraction, not a child's schema.

Simon Kirby's iterated-learning framework (2015) adds a transmission-level prediction. If the braid is loose—if the strands don't reinforce each other during transmission—then braided categories should show greater diachronic instability than unified categories. Languages may fragment the pronoun class differently, depending on which strand happens to dominate the speakers' representation. English bundles deixis, anaphor, and interrogative into a single morphological paradigm (*who/whom/whose*, *which*); some languages separate them. The HPC framework predicts this variability: braided categories have multiple stable equilibria, depending on which mechanisms dominate transmission.

The adult intuition that "pronouns" form a coherent class may itself be an illusion. The distributional overlap is real, but the underlying architecture is multiple. What adults experience as category unity is convergence without ancestry—Wallace's mimicry again.

11.7 Worked example: focus modifiers and fused relatives

The mechanism-braiding analysis makes concrete predictions. If deixis, anaphor, and interrogative are maintained by different regimes, they should behave differently in contexts that probe those mechanisms. Focus modifiers provide the test.

Focus particles like *exactly*, *precisely*, and *just* foreground the constituent they modify. They demand salience; the modified element must be in focus. This is the Backgrounded Constituent Infelicity (BCI) principle identified by Cuneo and Goldberg (2023): backgrounded constituents resist foregrounding operations.

Now compare interrogative *who* with relative *who*:

(4) *Exactly who called?*

(5) * *the person exactly who called*

Interrogative *who* accepts the focus modifier; relative *who* rejects it. Why?

The mechanisms differ. Interrogative *who* is inherently foregrounded: it introduces a question, anchors focus, and demands an answer. Focus modification is compatible—even expected. But relative *who* is backgrounded: it restricts an antecedent, contributes old information, and typically appears in non-focal position. Adding *exactly* tries to foreground a constituent whose job is to stay in the background. The mechanisms collide; the sentence fails.

This is the fingerprint of braided categories. If interrogative and relative *who* were maintained by the same regime, they should behave identically under focus modification. They don't. The surface identity—both spelled *who*, both resisting determiners—masks functional divergence. The category *wh-word* is a cover term for at least two distinct mechanisms.

The pattern extends. Fused relatives (*whoever calls*, *what you need*) occupy a middle position: they introduce definite reference without an overt antecedent.

(6) *I'll help whoever needs it.*

(7) *?I'll help exactly whoever needs it.*

The focus-modified fused relative is marginal—better than restricting relative, worse than interrogative. This intermediate status makes sense if fused relatives combine features of both: they introduce a referent (like interrogatives) but presuppose its recoverability (like relatives). The focus-compatibility reflects the mechanism mix.

Ongoing experimental work may sharpen these predictions. If the BCI principle holds, then precision modifiers should track the foregrounding profile of the construction, not the morphological shape of the *wh*-word.[1] That is what mechanism-braiding predicts: same surface form, different syntactic license, because the mechanisms diverge.

11.8 Looking forward

We have dissected the dictionary list. What we found was not a single kind of unity but several.

Adverbs are a fat category: no unified proper function, no homeostatic reinforcement, no principled clustering. They persist as a label, not a kind. The wastebasket has excellent storage capacity, but the items inside share only a negative property: they are not something else.

Nouns and verbs are tight HPC kinds: robust cross-linguistic presence, converging mechanisms at cognitive, discourse, acquisition, and transmission scales. They are the skeleton—the load-bearing architecture without which a language can't function.

Adjectives are thinner HPC kinds: weaker mechanisms, variable cross-linguistic presence, but genuine clustering within languages that have them. They are plumage—sometimes spectacular, sometimes absent, never essential.

[1] Acceptability judgment studies testing these predictions are in preparation.

Pronouns are braided categories: multiple proper functions converging on one distributional profile. The shark, the dolphin, and the ichthyosaur all resist determiners—but for different reasons. The category name captures the surface; the mechanisms are plural.

This taxonomy is not exhaustive. Determiners, prepositions, and conjunctions have their own stories, their own mechanism profiles, their own degrees of homeostatic coherence. What the present analysis offers is a method: ask not "What is the definition of category X?" but "What maintains category X?" The answers differ by category, and the differences reveal the architecture.

One final refinement from Chapter 6: *field-relative projectibility*. Different subfields may have different right-sized categories for the same extension. Morphology cares about lexical category differently than syntax: *cattle* is morphologically singular but syntactically plural. Typology cares differently than single-language grammar: Haspelmath's "comparative concepts" are projectible across languages, language-particular categories are projectible within. The "right-sized category" depends on the questions you're asking.

This explains why debates about lexical category boundaries persist. Are adjectives a subclass of nouns? Are adverbs a subclass of adjectives? The questions presuppose that category boundaries are fixed. The HPC framework says: fixed relative to what? Projectibility for morphological purposes may differ from projectibility for semantic purposes. The category is real, but its boundaries are indexed to analytical purpose.

We have moved from the dictionary's tidy list to the anatomist's table. The categories are not boxes into which words are sorted; they are standing waves maintained by mechanisms. Some waves are stable; others are braided; others are phantoms. Knowing the difference is the beginning of understanding lexical categories.

Pro-form Gender

There appears to be a certain amount of learning in losing one's tail; researchers in California found that once a young skink has had a close encounter of the near-fatal kind it seems to be more cautious.

—Simon Anderson, *Blink of a Lizard* (2008)

12.1 The puzzle

Consider two sentences about the same dog:

(1) * *The dog wagged its tail.*

(2) * *Who's a good boy? Yes, you are!*

In (1), the dog takes the non-personal pronoun *it*. In (2), addressed directly by its owner, the same animal takes the personal interrogative pro-form *who* and receives *you*. Nothing about the dog has changed. (Same dog; same tail.) What changed is how the speaker construed the referent.

This is the puzzle of English gender. Traditional accounts describe a three-way distinction – masculine, feminine, neuter – realized mainly on pronouns (*he*, *she*, *it*). That description captures something, but it misses the organizing principle. The primary distinction in English gender isn't sex. It's personhood.

The term GENDER in linguistics denotes a system of grammatically relevant contrasts wherein certain semantic concepts are divided into a small number of categories. English gender is typically described as referential rather than noun-class: there's no arbitrary assignment to nouns (unlike German *der Tisch*, *die Lampe*), and the choice of pronoun tracks properties of the referent, not grammatical properties of the antecedent (Corbett, 1991; Huddleston & Pullum, 2002). That much is right. But the usual account keeps the scope narrow – *he/she/it* – when the system is far wider.

Siemund (2008) models dialectal variation in English pronominal gender as different thresholds on an individuation hierarchy; Audring (2009) shows how pronominal systems resemanticize when agreement is borne primarily by pronouns; Dolberg (2019) traces the transition from lexical to referential gender in the *Anglo-Saxon Chronicle*. These accounts share a common architecture: English gender is referential, hierarchically organized, and driven by properties of the designatum rather than the antecedent. What they also share is a common limitation: they keep the system's scope largely within third-person pronouns.

This chapter shows why that restriction is too narrow. The same personhood-based logic that governs *he/she/it* also governs *who/which*, *somebody/something*, and *when/where*. The system extends across the entire semantic class of pro-forms – items that take their meaning from another element in discourse. This claim requires defence, since these items belong to different lexical categories. I'll address that question after laying out the evidence.

I'll argue that English gender, like countability (Chapter 9) and definiteness (Chapter 10), exhibits the HPC architecture: two clusters – one semantic (personhood), one lexico-grammatical (the pro-form inventory) – coupled by designatum-driven inference and maintained by overlapping mechanisms. Violations aren't just semantic infelicities; they're grammatical errors. The system has teeth – though some are sharper than others, and a few are mostly for display. Some constraints are categorical in standard written English (the *who/which* split; core *he/she/it* uses); others show the graded edge behaviour typical of HPC kinds (compound determinatives under shifted construals; chain-coherence effects). The interest lies in explaining both the categorical core and the gradient periphery. Pro-form gender is a narrow basin – deep within its domain but small in scope.

The puzzle shows up equally clearly in relative pronouns. Consider:

(3) * *The doctor who I saw was helpful.*

(4) * *The doctor which I saw was helpful.*

(5) * *The book which I read was helpful.*

(6) * *The book who I read was helpful.*

The constraint is robust in standard written English. *Who* for persons, pro-form *which* for non-persons. (*Which person* is fine – it's the pro-form use that's restricted.) The violations in (4) and (6) aren't merely odd – they're ungrammatical, comparable to agreement errors or subcategorization violations.[1]

The same split appears in interrogatives:

(7) * *Who's coming to the party?* [asking about persons]

(8) * *What's in your purse?* [asking about non-persons]

(9) * *Who's in your purse?* [under construal: asking about objects]

(10) * *What's coming to the party?* [under construal: asking about persons]

Examples (9) and (10) are grammatical only under shifted construals – if the speaker expects a person in the purse or a non-person attending the party. The choice of interrogative reveals the speaker's conceptualization of the potential referent.

This is the *who/which* puzzle. Relative pronouns aren't usually discussed under the heading of "gender,"[2] but their distribution obeys exactly the same personhood-based constraint that governs *he/she/it*. The puzzle dissolves when we recognize that both are manifestations of a single hierarchical system.

12.2 The hierarchy

The English pro-form gender system is hierarchically organized (Figure 12.1). At the top level, the distinction is between PERSONAL and NON-PERSONAL. Within personal, there's a further split between EPICENE (unmarked for sex) and SEXUAL (masculine/feminine). Within non-personal, productive subtypes include LOCATIVE and TEMPORAL.

Each subtype inherits the properties of its parent. A word with feminine gender is necessarily sexual, which is necessarily personal. A word with locative gender is necessarily non-personal. The locative and temporal subtypes are classified under gender because they participate in the same designatum-driven licensing: *where* and *when* presuppose a spatial or temporal construal of their

[1] The categorical/graded distinction matters for the HPC analysis. Categorical constraints (*who/which* in the standard system) define the cluster's core; graded constraints (chain coherence, determinative compounds) define its periphery. Both are explained by the same mechanisms operating with different strength.

[2] Both Huddleston and Pullum (2002) and Quirk et al. (1985) treat the personal/non-personal distinction as an axis separable from the *he/she/it* sex-based contrast. What's less common is to treat the *who/which* system, and related pro-forms like *somebody/something*, as part of the same organized inventory.

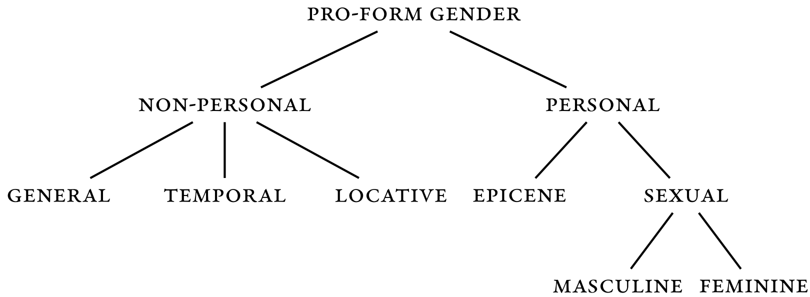


Figure 12.1. Hierarchy of gender values for English pro-forms.

antecedent just as *who* presupposes a personal construal. In form-value terms, the pro-form inventory is the form side; personhood construal is the value – what the form contributes when deployed. All are maintained by the same coupling mechanism: designatum-driven selection matching form to value.

Table 12.1 shows the inventory. Rows group pro-forms by syntactic function (personal pronouns, relatives/interrogatives, demonstratives, compound determinatives); columns group them by gender value. The leftmost column lists gender-neutral pro-forms – items that impose no presuppositional constraint on the designatum. The remaining columns list gender-sensitive items, distinguished by the personhood constraint they encode.

Table 12.1. The pro-form gender system of English.

Gender-neutral	Gender-sensitive					
	Non-personal			Personal		
	General	Temporal	Locative	Epicene	Sexual	
					Masc	Fem
they$_{\text{pl}}$	*it*	*then*	*there*	*I*, *we*$_{\text{pron}}$, *you*$_{\text{pron}}$, *they*$_{\text{sg}}$, *one*$_{\text{gen}}$	*he*	*she*
whose$_{\text{rel}}$	*what*, *which*	*when*	*where*	*who*, *whom*, *whose*$_{\text{int}}$		
this, *that*†			*here*	*we*$_{\text{det}}$, *you*$_{\text{det}}$		
	-thing		*-where*	*-body*, *-one*		

Note. Subscripts: $_{\text{pl}}$ = plural; $_{\text{sg}}$ = singular; $_{\text{gen}}$ = generic; $_{\text{rel}}$ = relative; $_{\text{int}}$ = interrogative; $_{\text{pron}}$ = pronoun; $_{\text{det}}$ = determiner function (as in *we linguists*, *you guys*). † In fused-head function, *this* and *that* show constructionally-conditioned gender: neutral in specificational *be* contexts and with ellipsis; non-personal elsewhere (see §12.6.4).

Several entries require comment. Plural *they* is gender-neutral. It can refer to persons (*There are some people. They are tall*) or non-persons (*There are some trees. They are tall*) without constraint. Singular *they*, however, is ordinarily personal: *There's a person. They are tall* is straightforward, while *There's a tree. They are tall* is available only under marked construals (personification, jocular animacy, or a discourse shift away from strict coreference) (Bjorkman, 2017; Konnelly & Cowper, 2020). The difference is that singular *they* carries a default personal presupposition; plural *they* doesn't.

Similarly, relative *whose* is compatible with both personal and non-personal antecedents (*the person/book whose story is well known*), while interrogative *whose* typically presupposes a personal possessor (*Whose coat is this?*). The split reflects how these items function: relative *whose* inherits its construal from the antecedent; interrogative *whose* presupposes personhood independently. Non-personal uses of interrogative *whose* do occur with inalienable possession (*Whose leaves are these?*, asking about a tree) or when personification is contextually available, but the default is personal.

I include first- and second-person pronouns (*I*, *we*, *you*) to show the inventory's personhood partition. These are epicene personal: they can only refer to entities construed as persons (using *I* for an inanimate object requires personification), but they don't encode sexual gender. Nothing in the analysis turns on sex marking for these items.

The table reveals a system far larger than the traditional *he*/*she*/*it* triad. The compound determinatives – *somebody*/*something*, *everyone*/*everything*, *anywhere*/*anyone* – participate in the same personhood partition. The relative pro-forms *where* and *when* encode non-personal subtypes parallel to *which*.

12.3 Designatum-driven gender

Before examining the mechanism, a clarification. The English system tracks the DESIGNATUM – the entity as conceptualized by the speaker – not the referent (the real-world correlate) or the antecedent (the linguistic controller). The DESIGNATUM is what a linguistic item designates; the REFERENT is "some independently distinguishable entity, or set of entities, in the real world" (Huddleston & Pullum, 2002, p. 399). The ANTECEDENT is the "constituent whose meaning dictates the meaning of a pronoun or other such expression in cases of anaphora" (Huddleston & Pullum, 2005, p. 295).

This distinction is crucial. A pro-form need not have an antecedent to participate in the gender system: interrogative *who* introduces an open slot whose intended value is constrained to persons even though the utterance has neither a referent nor an antecedent. The designatum need not be a token: *the typical* *politician is not to be trusted* designates a type, but the gender is still personal. And the designatum can diverge from the antecedent's lexical properties.

The English system differs from antecedent-driven systems in exactly this way. In Spanish, French, or German, gender assignment is fixed on nouns and agreement spreads through NP-internal morphology. The antecedent controls the pro-form.

In English, it's the designatum – how the referent is conceptualized – that controls selection. The source provides the constraint; the pro-form satisfies it.

Consider a case of referential metonymy:

(11) * *The French fries is waiting. She's upset.* (Wu, 2025)

An NP with a default non-personal reading (*the French fries*) is used to invoke a customer; subsequent anaphora tracks that metonymic designatum, yielding personal-feminine *she*. The fries are, briefly, a customer with a pronoun. The antecedent's form (plural, non-personal) doesn't control the pro-form; the conceptualized source does.

Spanish is much less permissive about this kind of pro-form shift while retaining ordinary agreement with the same antecedent NP. (The English number mismatch – plural label, singular verb – reflects a menu-item construal orthogonal to the pro-form point; in Spanish I keep agreement normal to isolate the controller issue.)

(12) a. * *Las patatas fritas esperan. Él está enfadado.*
DEF.F.PL potato.F.PL fried.F.PL wait.3PL.PRES 3SG.M be.3SG.PRES angry.M.SG
Intended: 'The French fries are waiting. He's upset.' [fails under intended coreference]

b. *Las patatas fritas esperan. El señor está enfadado.*
DEF.F.PL potato.F.PL fried.F.PL wait.3PL.PRES DEF.M.SG gentleman.M.SG be.3SG.PRES angry.M.SG
'The French fries are waiting. The gentleman is upset.'

Spanish requires grammatical agreement with the syntactic antecedent (*patatas*, feminine plural), a constraint that overrides the semantic gender of the metonymic designatum. The speaker typically recasts the sentence with a human-denoting controller like *el señor* ('the gentleman') to align the syntax with the intended reference. English permits categorical shifts based on conceptualization because its gender system is maintained by the designatum directly, not by antecedent-agreement: the designatum-driven pathway is freer. This is the signature of a designatum-driven system.

12.4 The inventory

Standard accounts keep gender within personal pronouns. Huddleston and Pullum (2002, p. 486) state that English gender is "based purely on pronoun agreement." But the evidence shows the system extends across all pro-forms – the semantic class of items that take their meaning from another element in discourse.

What unifies pro-forms is their interpretive dependence: they have minimal descriptive content and take their value from context – linguistic, situational, or conceptual. In the clearest cases, a pronoun heads an NP whose referent is recoverable from the discourse or situation; pro-forms like *here*/*there*/*then* likewise contribute a place or time that is supplied by context rather than descriptively specified. In interrogatives and relatives, *wh*-forms don't "stand for" a missing phrase so much as mark an open slot in the clause's interpretation, linked to an answer (in questions) or to an antecedent (in relatives). This isn't a morphosyntactic category (there's no frame that selects for "any pro-form"), but a semantic–pragmatic one. The gender constraint doesn't respect lexical-category boundaries: whatever property licenses *who* also licenses *somebody* and blocks *something*.

The skink in the epigraph illustrates exactly this: a fleeting promotion (*one's*) and immediate demotion (*it*) within one sentence – the kind of construal shift that §12.5.1 will examine.

12.4.1 Determinatives

The compound determinatives with *-body* and *-one* are personal:

(13) * *Somebody left their coat.* [personal]

(14) * *Somebody is in your purse.* [under construal: referring to an object]

Example (14) is ungrammatical unless *somebody* refers to a person who is in the purse. These are epicene – they don't encode sex – but they do encode the personal/non-personal distinction.

The compound determinatives with *-thing* and *-where* are non-personal:

(15) * *Something is on the table.* [non-personal]

(16) * *Everything enjoyed themselves.* [under construal: personal referents]

Example (16) requires personification to be grammatical. The boundary is discourse-driven rather than syntactically hard – exactly the HPC-style graded edge we expect.

12.4.2 Relative pro-forms

The distinction between *where/when* and *which* parallels the distinction between *who* and *which*:

(17) * *the room where/*which the painting was done* [room as location]

(18) * *the room which/*where was painted* [room as patient]

(19) * *2010, when/*which I left* [as time]

(20) * *2010, which/*when has 365 days* [as entity]

The same source (*room*, *2010*) takes different pro-forms depending on how it's conceptualized. This is designatum-driven selection applied to locative and temporal subtypes.

A note on apparent avoidance. English relative clauses permit *that* (a subordinator functioning as marker) and bare relatives without any overt relative word: *the doctor that I saw*, *the doctor I saw*. These aren't gender-neutral pro-forms – they're structurally different constructions that happen to sidestep the *who/which* choice. Their availability doesn't undermine the gender system; it simply means speakers can use a different construction when they would rather not take a position on the metaphysics.

12.5 How the system holds together

12.5.1 Chain coherence

Once a designatum is construed as personal or non-personal within a reference chain, mixing is disfavoured. Consider:

(21) * *That's the dog <u>who</u> attacked <u>his</u> owner.* [personal chain]

(22) * *That's the dog <u>which</u> attacked <u>its</u> owner.* [non-personal chain]

(23) * *That's the dog <u>which</u> attacked <u>his</u> owner.* [mixed]

(24) * *That's the dog <u>who</u> attacked <u>its</u> owner.* [mixed]

Examples (21) and (22) are fully acceptable – coherent personal and non-personal chains. Examples (23) and (24) are degraded – the chain starts in one gender and shifts to another mid-sentence.

This is chain-internal coherence. The constraint isn't absolute – discourse can independently reclassify a designatum (personification, stance shift) – but absent such cues, speakers prefer consistency within a reference chain. Attested mixed chains do occur, but they almost always involve a shift in stance.

Recall the skink in the epigraph: *losing* one's *tail* (generic personal) followed by it *seems* (non-personal). The shift marks a move from a generic "experiencer" perspective to a biological description. Similarly, speakers discussing ships or infants may toggle between *she/he* and *it* as their emotional distance varies. What's rare is a mix without a motive.

The effect is strongest in online processing: mixed chains create a local mismatch that invites repair or reinterpretation, and that repair pressure is itself a stabilizer of the coupling.

Chain coherence is a maintenance mechanism. It stabilizes the personhood assignment across discourse, reinforcing the clustering of pro-forms around a shared construal. This is the English analogue to French NP-internal concord: in French, gender coherence is enforced morphologically across determiners, adjectives, and participles; in English, it's enforced referentially across the pro-forms in a chain.

12.5.2 The coupling

We now have the architecture (Figure 12.1). On the semantic side sits the PERSONHOOD CLUSTER: the conceptual properties that make a referent construable as a person. These cluster because personhood is a natural cognitive category, grounded in Theory of Mind and social cognition (Waytz et al., 2010).

Person-attribution, as I use it, tracks a cluster of features: potential for reciprocal interaction, nameability, attribution of intentions, membership in the speaker's circle of concern. This is folk-psychological personhood – the kind speakers attribute in online cognition – not the social-recognition sense invoked in debates about rights or legal standing. Dennett (1976) lists formal conditions for personhood (rationality, consciousness, reciprocal treatment, self-awareness); I use these illustratively, not as necessary conditions that speakers compute. What matters is that these features cluster – Theory of Mind operates on them as a package – and that the clustering is cognitively basic and socially consequential.

Two kinds of personhood construal should be distinguished. In most cases, personhood is a DEFAULT CONSTRUAL: *the doctor* evokes a person without the speaker doing anything special. In others, personhood is LOCALLY IMPOSED: *Who's a good boy?* addressed to a dog requires active construal work. The maintenance story tracks both. Defaults are entrenched through ordinary usage; local impositions require discourse cues and carry interactional meaning (affection, stance, humour). The system accommodates both because the coupling is bidirectional: form can signal construal, and construal can license form.

On the lexico-grammatical side sits the PRO-FORM INVENTORY: the gender-sensitive lexical items and their distributional constraints: the personal pronouns (*he/she/it/they*), the pro-nominal *one*, *who/which*, *-body/-thing* in compound determinatives, and designated slots for personal and non-personal reference. (Pro-form is a semantic grouping; what I'm tracking here is its *realization* – the specific forms and where they're licensed.)

The two clusters are coupled by DESIGNATUM-DRIVEN INFERENCE. When speakers produce a pro-form, they signal how they're construing the referent. When hearers process a pro-form, they infer the construal. Form cues construal; construal constrains form. This coupling is the category we recognize as English gender.

This parallels the bidirectional inference that couples individuation to count morphosyntax (Chapter 9) and identifiability to deitality (Chapter 10). The coupling is what produces the HPC signature: properties that statistically co-occur, maintained by causal mechanisms, with instability at the edges.

12.5.3 The machinery of maintenance

Chapter 7 surveyed some of the general machinery that maintains linguistic kinds: acquisition, entrenchment, alignment, transmission. That list was explicitly incomplete; no theory of linguistic kinds can enumerate every stabilizing force. For pro-form gender, at least three domain-specific channels are visible.

COGNITIVE GROUNDING explains why acquisition succeeds. As noted above (§12.5.2), the person/non-person boundary is mapped onto a pre-existing conceptual distinction that infants acquire before language begins (Carey, 2009; Gergely & Csibra, 2003). Because the conceptual cluster is already in place, children don't have to induce it from pro-form distributions – they just have to learn which forms map onto which region of an existing landscape. This is why acquisition of English pro-form gender is fast and robust. The empirical footprint: children should produce personal pro-forms for robots and animals that pass Theory-of-Mind tests, and non-personal pro-forms for entities that fail them, even before they can articulate why.

SEMANTIC TRANSPARENCY explains why transmission preserves the coupling. The mapping from personhood to pro-form is largely predictable: if you know a referent is construed as a person, you can predict the pro-form inventory it will take. Compare French gender, where assignment is arbitrary and must be learned item by item, or English countability, where object-mass nouns like *furniture* create opaque exceptions. English pro-form gender has fewer such mismatches: the *who/which* alternation tracks personhood with high fidelity. This transparency means the system can be reconstructed from limited input –

what Chapter 7 called LEARNABILITY FILTERING. Children acquiring English don't produce the persistent paradigmatic gender-agreement errors characteristic of antecedent-gender languages, though direct acquisition studies of the *who/which* contrast specifically remain sparse. The empirical footprint: novel entities (AI assistants, fictional creatures) should be rapidly and consistently assigned pro-forms once their personhood status is pragmatically established, without explicit instruction.

ALIGNMENT AND REPAIR explains why the coupling is enforced in real time. Consider a constructed repair:

> A: *The new surgeon started today. Is he any good?*
> B: She. *Dr. Patel's a woman.*

The correction is immediate, unmarked, and socially meaningful. The point is not that speakers consult a rulebook; it's that the mismatch is interactionally loud. Gender mismatches are charged in a way that many grammatical errors aren't: calling a person *it* is an insult; misgendering someone with *he* when *she* is appropriate triggers conversational repair (McConnell-Ginet, 2014). This repair pressure operates through what Chapter 7 called INTERACTIVE ALIGNMENT: interlocutors converge on shared construals, and deviations are corrected. The social stakes amplify the mechanism – gender isn't just about communication; it's about recognition.

The pressure can also work in reverse – resisting innovation rather than enforcing existing norms. A younger speaker who says *My partner called – they're running late* may prompt an older hearer to ask *He or she?*, treating singular *they* for a specific individual as underspecified rather than complete. The negotiation reveals competing norms rather than a shared one, and it's exactly where variation survives: alignment stabilizes whichever system the interlocutors happen to share, but when systems diverge, the repair attempt exposes the crack.

The empirical footprint: gender corrections should be faster and more frequent than corrections for other grammatical mismatches (e.g., number disagreement), and corpus data should show that gender repairs cluster in contexts where personhood or sex attribution is socially consequential.

These three channels – cognitive grounding, semantic transparency, alignment and repair – don't exhaust the mechanisms, but they illustrate how the general stabilizers from Chapter 7 operate in this domain. The stability of pro-form gender emerges from their convergence: the conceptual cluster is prelinguistic; the form-meaning mapping is reconstructible; violations are repaired in real time. Other forces – register conventions, prescriptive teaching, literary tradition – doubtless contribute; the point is not to close the list but to show that the coupling is actively maintained.

12.5.4 What the HPC framing buys

A sceptical reader might ask: why invoke HPC machinery at all? Why not say that each pro-form simply carries its own selectional presupposition – *who* presupposes a personal referent, *which* presupposes a non-personal one – and leave it there? Item-by-item presuppositions are simpler than coupled clusters.

The answer is that item-by-item presuppositions don't explain the system's behaviour. Three phenomena require more than local selectional restrictions.

First, bidirectional inference. If pro-forms merely carried presuppositions, hearers would use form to filter referent candidates – ruling out non-persons for *who*, ruling out persons for *which*. But the inference also runs the other way: hearers use *who* to *attribute* personal construal, even when the referent is non-human. *The dog who bit me* doesn't just filter for dogs already construed as persons; it *invites* that construal. This is coupling, not selection.

Second, chain coherence. If gender were just local presupposition, mixed chains (*who... it*) should be as acceptable as any other presupposition shift within discourse. They aren't. The dispreference for mixed chains is a *system-level* constraint – it tracks construal stability across the chain, not just satisfaction of local presuppositions. That's the signature of a maintained coupling, not a collection of independent filters.

Third, boundary cases tracking the same dimension. The boundary cases – pets, collectives, infants, robots – don't scatter randomly. They cluster along the personhood dimension. If each pro-form had independent presuppositional content, we would expect idiosyncratic variation: *who* might extend to animals, *somebody* might not. Instead, the entire inventory shifts together when personhood is uncertain. That's what coupled clusters do.

The HPC framing doesn't add complexity for its own sake. It explains why the system behaves as a system: why inference is bidirectional, why chains cohere, why boundaries track a shared dimension. Item-by-item presuppositions would predict a looser, more heterogeneous pattern – exactly what we don't see.

12.6 Evidence and limits

12.6.1 The maintenance spectrum

Gender systems differ not just in their inventories but in *what maintains them*. The contrast between English and French illustrates the extremes; German sits between (Audring, 2009; Corbett, 1991).

English pro-form gender is DESIGNATUM-DRIVEN: the form tracks how the speaker construes the referent. The metonymy example in §12.3 showed this – *the French fries... she* is possible because the pro-form follows the conceptualized customer, not the grammatical antecedent. The system is maintained by a live semantic link: form and meaning travel together because the coupling is transparent.

French gender is ANTECEDENT-DRIVEN: the pro-form tracks the grammatical gender of the antecedent NP, regardless of the referent's properties. The same metonymy fails:

(25)	* *Les*	*frites*	*attendent.*	*Il*	*est*	*en*	*colère.*
	DEF.F.PL	fry.F.PL	wait.3PL.PRES	3SG.M	be.3SG.PRES	in	anger

Intended: 'The fries are waiting. He's upset.' [fails under intended coreference]

The speaker can't use masculine *il* to track a male customer if the antecedent *les frites* is grammatically feminine. The form-meaning link is dead: French gender is maintained by PURE ENTRENCHMENT – item-by-item memorization of arbitrary noun-class assignments, reinforced by NP-internal concord.

German is the hybrid. Nouns carry grammatical gender (*der Tisch* 'the table', masculine; *die Lampe* 'the lamp', feminine; *das Mädchen* 'the girl', neuter). But pronouns can override the antecedent's grammatical gender when the referent's natural gender is salient:

(26)	* *Das*	*Mädchen*	*kam*	*herein.*	*Sie*	*lächelte.*
	DEF.N	girl.N	came	in.	3SG.F	smiled.

'The girl came in. She smiled.'

The antecedent *das Mädchen* is grammatically neuter, but the pronoun *sie* is feminine – tracking the referent's natural gender, not the noun's grammatical class. German gender is maintained by *two mechanisms in tension*: entrenchment of noun-class assignment, and semantic override in pronominal anaphora.

This is the MAINTENANCE SPECTRUM. At one end, English: semantically transparent, designatum-driven, clustering held together by a live conceptual distinction. At the other end, French: semantically arbitrary, antecedent-driven, clustering held together by morphological entrenchment. French gender shows

MECHANISTIC DRIFT (Chapter 6): whatever semantic triggers may once have motivated the masculine/feminine split have long faded, but entrenchment maintains the forms. German sits between – two mechanisms in tension, neither fully dominant.

The HPC framework predicts different profiles. English gender should be easy to acquire (transparent mapping) but potentially unstable under pressure (no formal redundancy). French gender should be hard to acquire (arbitrary assignment) but stable once learned (reinforced by concord). German should show characteristic acquisition errors where the two systems conflict – and it does: children often produce natural-gender pronouns for grammatical-neuter human nouns. The "same" category has different ontological status depending on what maintains it.

12.6.2 Cross-linguistic scope

Is pro-form gender an English peculiarity or a cross-linguistic pattern?

The personhood distinction appears remarkably widespread. Haspelmath (1997) notes that "the distinction between human and non-human referents is made practically everywhere ('who' vs. 'what', 'somebody' vs. 'something'), even in languages where humanness isn't very prominent elsewhere in the grammar" (34–35). Languages with rich antecedent-driven gender (Spanish *quién/qué*, German *wer/was*, French *qui/quoi*) still make the personhood split in their interrogative and indefinite systems (Corbett, 1991; Dahl, 2000). First- and second-person pronouns are epicene in many languages that otherwise mark sex in the third person, suggesting that the personhood axis – distinguishing speech-act participants from non-participants – is cross-linguistically more stable than sex-based contrasts. The personhood partition may be cognitively more basic than the particular morphosyntactic systems that realize it.

What's distinctive about English isn't the personhood distinction itself but its SCOPE. Because English lacks NP-internal gender agreement, the designatum-driven pathway is unconstrained by formal concord. Speakers can shift a referent's construal without generating agreement violations. The *who/which* alternation, the *-body/-thing* compounds, and the chain-coherence effects are all more prominent in English precisely because they're not overridden by noun-class agreement.

Dialectal variation within English shows the system's parameters. Siemund (2008) documents varieties (Southwest England, Newfoundland, Tasmania) where the personal/non-personal cut shifts: *he* and *she* extend to inanimates based on a mass/count distinction rather than personhood. The hierarchy's structure remains stable; its semantic basis varies. This suggests the architecture is robust but its semantic grounding is a parameter that communities can tune.

12.6.3 Passing the tests

Chapter 8 introduced the Two-Diagnostic Test for genuine HPC kinds: high projectibility and robust homeostasis.

Knowing a pro-form's gender predicts its distribution. If you know *somebody* is personal, you can predict it will combine with personal predicates, take personal reflexives (*themselves*), and resist non-personal contexts. If you know *which* is non-personal, you can predict it can't take human antecedents without personification.

Crucially, knowing the designatum's construal predicts pro-form selection. If you know the speaker is construing a pet as personal, you can predict *who* over *which*, *he/she* over *it*. The predictions are bidirectional: from form to construal, from construal to form.

The clustering is maintained by mechanisms. Acquisition transmits the system as a unit; entrenchment preserves high-frequency patterns; alignment provides real-time stabilization; transmission filters for learnability. Perturb any mechanism – expose children to non-standard input, isolate speakers from conversational feedback – and the cluster should degrade. But it won't scatter randomly; the degradation will track the mechanisms involved.

The clustering isn't just a statistical regularity. It's causally maintained.

12.6.4 Where the clusters slip

By "slippage" I mean cases where the two clusters that normally travel together – personhood construal and the pro-form inventory that encodes it – are pulled apart by independent pressures. The result is not random noise: the same boundary cases recur, and the same repairs and preferences reassert the coupling. The interest of these cases is diagnostic: they show which mechanisms are doing the stabilizing, and where other systems – number, specificity, register – interfere with the personhood signal.

Boundary designata

Some referents sit at the boundary of personhood attribution. Pets are the paradigm case. Shir-Vertesh (2012) demonstrates that pet-keeping licenses flexible personhood construal: the same animal can be a person in one context and not in another. This flexibility surfaces in pro-form choice:

(27) * *The cat licked its paw.* [non-personal default]

(28) * *Poor Whiskers! She's not feeling well.* [personal]

Same referent, different construals – and different pro-forms. The variability here is not a partial breakdown of the coupling; it's the *expected* outcome when the semantic cluster itself is gradient. Because personhood is genuinely uncertain for these referents, the system permits controlled alternation. What stabilizes the pattern is chain coherence: having chosen personal or non-personal for a given referent in a given discourse, speakers prefer to stay consistent.

Interference from number

Collective nouns present a case where another grammatical cluster – number and mereology – interferes with the personhood signal:

(29) * *The team has scored its first goal.* [singular, non-personal]

(30) * *The team have scored their first goal.* [plural, personal]

The variation tracks construal: team-as-entity versus team-as-members. But this is primarily notional versus grammatical number, not gender per se. The personhood of the individual members is not in doubt; what varies is whether they're accessed as individuals or aggregated into a single non-personal entity.

Dialect and register norms amplify the variation. British English more readily accepts plural verb agreement with singular collective heads; American English more strongly prefers singular (Bock et al., 2006). What's interesting for the HPC analysis is that the combination of singular verb with plural personal anaphora (*The team has changed their strategy*) is widely tolerated – the two agreement systems (verbal, pronominal) can diverge when pulled by different pressures. The mechanism that stabilizes the pattern is standardization pressure: edited prose tends to favour singular–singular or plural–plural; informal speech permits the mix.

A related case is *much*, which is strictly non-personal: **Much arrived late* is ungrammatical with intended personal reference. But this may not be a gender constraint per se. Persons are typically individuated – conceptualized as countable entities – and *much* quantifies over mass. The incompatibility may follow from count/mass semantics rather than from an independent personhood presupposition on *much*. If so, the non-personal status of *much* is an epiphenomenon of the count/mass system, not a direct participant in the gender coupling. This is a case where two maintenance mechanisms – individuation and personhood – happen to align, making it hard to say which is doing the work.

Competition from other pressures

Human infants, paradigmatic persons, can take non-personal *it*:

(31) * *The baby was crying because it was hungry.*

This is not a denial of personhood. It reflects competition between personhood and other communicative pressures – particularly sex-marking and specificity. When the infant's sex is unknown or irrelevant, *it* can avoid the sex-based *he/she* alternation. In generic construals (*A baby needs its mother*), the kind-level reading aligns naturally with non-personal anaphora. And in some contexts, as McConnell-Ginet (2014) notes, the choice of *it* can carry ideological weight – positioning the infant as not-yet-fully-a-person, though this is not the default reading.

The slippage here is diagnostic: it shows that the personhood–pro-form coupling can be loosened when other communicative goals take priority. What stabilizes the pattern is semantic transparency: when a speaker does construe an infant as personal, the system snaps back to *he/she/they*. As singular *they* spreads as a sex-neutral personal pronoun, *it* for infants may wane – replaced by a form that is unambiguously personal without requiring sex specification.

Lexeme-specific anchoring

Not all non-personal pro-forms behave identically when chains continue. Relative *which*, tethered to an antecedent NP, appears more permissive for some speakers in allowing later personal anaphora, even though the resulting mixed chain is typically marked:

(32) * *The dog which attacked the woman ran away. The police never caught him.* [upgrade within chain]

The antecedent (*the dog*) supplies enough information – animacy, individuality – that hearers can reconstruct a personal designatum and accommodate the shift.

But *what* in free relatives and *something* are systematically underspecified. They lack a full NP antecedent to inherit features from, so there's less material for hearers to anchor a promoted personal construal:

(33) * *What attacked the woman ran away. The police never caught him.*

(34) * *Something attacked the woman. The police never caught him.* [with coreference]

The contrast, if these judgments hold up, is not simple gradience within the non-personal domain; it's a difference in discourse anchoring and feature support. Relative *which* points to an existing NP whose descriptive content can sometimes license an "upgrade"; *what* and *something* provide no comparable antecedent to inherit from, so the chain more strongly resists promotion to personal. The mechanism that stabilizes the pattern is chain coherence: when the anchor is weak, shift is blocked or sharply degraded.

Constructional interference: the demonstratives

The demonstratives *this* and *that* present a different kind of slippage – one driven not by properties of the designatum but by the syntactic construction in which the pro-form appears.

In fused determiner-head function, demonstratives aren't simply gender-neutral. Their gender is CONSTRUCTIONALLY CONDITIONED:

(35) * *She chatted with that.* [personal blocked]

(36) * *This is Yoko. / This is an acorn.* [specificational *be*: neutral]

(37) * *She chatted with this person and that.* [ellipsis: personal OK]

Example (35) shows that personal reference is blocked in bare fused-head contexts – unless the speaker intends depersonalization as an insult. But examples (36) and (37) show that personal reference is licensed in two environments: specificational *be* constructions and ellipsis with a recoverable head noun.

The pattern is not random. Specificational *be* and ellipsis both provide structural support that bare fused-heads lack. In *This is Yoko*, the post-copular complement supplies the personhood information; the demonstrative merely points. In *She chatted with this person and that*, the elided head *person* is recoverable from the first conjunct. Without such support, the demonstrative defaults to non-personal, and personal construal is unavailable.

This is constructional interference with the gender coupling. The same lexeme – *that* as a fused head – behaves as non-personal in one syntactic frame and as gender-neutral in another. The stabilizing mechanism is CONSTRUCTIONAL SUPPORT: certain syntactic environments carry licensing conditions that override the default constraint. This parallels the lexeme-specific anchoring discussed above: relative *which* with a full NP antecedent can license chain upgrades that underspecified *what* can't. Here, the specificational *be* construction and ellipsis structures provide the anchoring that bare demonstratives lack.

The demonstratives, then, sit uneasily in the gender-neutral column of Table 12.1. In fused-head function, they're neutral only when the construction provides enough structure to license personal reference; otherwise, they default to non-personal. This constructional sensitivity is exactly the kind of graded, context-dependent behaviour that HPC kinds predict – and it shows how grammatical structure can modulate the personhood–pro-form coupling.

12.7 Looking forward

Pro-form gender is a localized system. The relevant constraints are carried by a specific semantic class of lexical items – those that function as pro-forms – and enforced across anaphoric chains. The coupling between personhood and morphosyntax is tight within that domain: the category is real because it's maintained, and maintained because it's real. Lexical-gender systems like French exhibit MECHANISTIC DRIFT in the other direction: the semantic triggers have faded, but entrenchment maintains the forms (Chapter 6).

The final case study pushes to the widest scope. Lexical categories like NOUN and VERB look strikingly stable cross-linguistically, while ADJECTIVE and ADPOSITION vary dramatically. If these are HPC kinds, the mechanisms that maintain them must be far more diffuse – operating across the entire grammar, not just within a circumscribed inventory. Chapter 11 asks what happens when the coupling is weak, the mechanisms are partial, and the cluster boundaries are contested.

The category zipper

> If you have two complementary strands of DNA, they zip up. That's what they do.
>
> —Sri Kosuri, quoted in *Harvard Gazette* (2019)

13.1 When errors go wrong differently

You know the moment a zipper catches. The slider finds the teeth, they interlock, and the two sides move as one. You don't think about it. That's the point – when the coupling is tight, the mechanism disappears.

You also know the moment it doesn't catch. The slider moves but the teeth won't seat. Or it catches partway and then jams, leaving a gap where the fabric bulges through. The fix depends on where the failure is: at the bottom, you have to reseat the whole thing; in the middle, you can sometimes work it free; at the top, you're just short of done and the last few teeth won't close.

Language comprehension has all of these failure modes.

When you mishear a word, the error is a clean SLIP. You hear *grape* as *great*, not as a smear of sound. The perceptual system delivers discrete candidates even when the input is degraded. The mistake stays inside the system: one phoneme for another, one word for another. The zipper caught; it just caught on the wrong tooth.

When you misparse a sentence, the failure is a JAM. You can get every word right and still get the structure wrong. *The horse raced past the barn fell* is acoustically clear; the problem is that your parser zipped up the wrong configuration and only discovered the problem at *fell*. The teeth were seating fine until they weren't.

And when you miss an implicature, the zipper looks closed but the fabric is MIS-TRACKED underneath. *Some students passed* registers as good news when the speaker meant it as bad. Every tooth is in place. The tracking is off.

13.1.1 From failure modes to coupling regimes

These three failure types diagnose three coupling regimes – and each regime licenses different diagnostic evidence.

1. Slip (clean substitution) signals TRANSPARENT COUPLING. Form and value are so tightly bound that errors stay within the contrast system. Perturbations shift category boundaries predictably where cues degrade; they don't produce extragrammatical gibberish. *Diagnostic evidence*: perceptual confusion matrices cluster by featural similarity; mergers occur where acoustic cues weaken.

2. Jam (late crash after locally coherent assembly) signals ARCHITECTURAL COUPLING. Form templates pair with interpretive templates, but hidden commitments accumulate as parsing proceeds. The zipper catches partway, then fails when downstream requirements conflict. *Diagnostic evid-*

ence: garden-path signatures, revision points, cue-competition effects that show up in reading times or eye movements.

3. Mis-tracking (surface OK, downstream mismatch) signals LOOSE COUPLING. Multiple sources – lexical, constructional, pragmatic – contribute to interpretation, and they can align or misalign without local ill-formedness. *Diagnostic evidence*: implicature failures, pragmatic infelicity in discourse, mismatches between what was said and what was understood that emerge only in interaction.

The coupling regime reflects stabilizer weighting. Transparent coupling dominates where physical and perceptual constraints lock form to contrastive identity (phonemes). Architectural coupling dominates where conventional form–value pairings are compositionally assembled (constructions). Loose coupling dominates where contextual inference supplements or overrides encoded content (discourse). The zipper metaphor is a diagnostic instrument, not just an illustration: error shape tells you where to look for the stabilizers.

13.1.2 What "value" means

To say what's being coupled, we need a term that works across grains. We've already met the term VALUE – what a unit counts as in the system when deployed, its contrastive role, recoverable by competent users. Here it does heavy lifting.

VALUE in this book is not evaluative worth, not teleological purpose, not "meaning" in the undifferentiated everyday sense. It is relational rather than intrinsic: a unit's value is constituted by what it contrasts with and what inferences it licenses. For phonemes, value is contrastive identity: /k/ counts as not-/g/, not-/t/, not-/p/, and that exhausts its contribution. For morphemes, value is conventional pairing: the form *-ed* is associated with pastness by agreement, not resemblance. For constructions, value is interpretive template: a form configuration paired with a meaning configuration, compositionally and holistically.

The zipper, then, has form on one track and value on the other. What changes across grains is what kind of value, and how tightly form is coupled to it. This is why linguistic categories so often appear as CROSSCUTTING KINDS: form clusters and value clusters may be maintained by different mechanisms, and the zipper can slip at any level.

13.2 Two diagnostics, asymmetrically applied

How do we tell whether a category is a genuine kind – something that will repay induction – or just a convenient label? This chapter applies two diagnostics. Does the category *project*: can patterns learned from one sample predict another? And is it *maintained*: can we identify mechanisms that stabilize the cluster?

Boyd's homeostatic property-cluster framework treats projectibility as following from homeostatic maintenance: if mechanisms hold a cluster together, the cluster will support induction. For methodological clarity, I revise this relationship, treating projectibility and homeostasis as independent criteria that must both be satisfied.

Projectibility is tested directly via out-of-sample prediction: can patterns learned from one corpus identify categories in another? Homeostasis is inferred: can we name stabilizers and find their predicted signatures in the data? Both diagnostics have to pass for kindhood to be warranted. Appendix ?? details the operationalization, including thresholds, failure modes, and falsification conditions.

The case studies that follow apply these diagnostics to three positive cases – phonemes, morphemes, constructions – and three negative cases – academic register, Indo-European, polysynthetic. The positive cases show the framework scaling. The negative cases show it has teeth.

13.3 Transparent coupling: phonemes

The phoneme tier is the cleanest place to test the claim that linguistic categories are HPC kinds. Inventories are comparable across languages, there's independent theory about plausible stabilizers, and open resources allow fully reproducible analysis.

The cluster properties are familiar: acoustic cues (formant distributions, voice-onset time), production routines (reliable articulatory gestures), lexical contrast (minimal pairs making the distinction communicatively consequential), and perceptual categorization (listeners mapping continuous input onto discrete categories). These travel together without being definitionally linked.

What maintains the clustering? Four stabilizer types. First, quantal regions: Stevens (1989) showed that the articulatory-to-acoustic mapping is non-linear, with certain configurations producing stable outputs across a range of variation – natural "parking spots" that constrain what phoneme contrasts are *possible*. Second, dispersion pressure: Lindblom (1990) argued that vowel inventories disperse in acoustic space to maximize discriminability, shaping which of the possible phonemes a language *selects*. Third, perceptual tuning: Kuhl et al. (1992) demonstrated that infant perception is warped by early exposure, with category prototypes acting as perceptual magnets that tune the individual speaker to the community standard. Fourth, community norms: social transmission across generations stabilizes inventories through the mechanisms that cultural-tool accounts emphasize (Ekström et al., 2025).

Evidence comes from PHOIBLE 2.0 (Moran & McCloy, 2019), a database of phoneme inventories for roughly 2,700 languages. Two patterns satisfy the diagnostics. Plotting kernel-density ridgelines of total inventory sizes by family reveals a *stability band*: medians cluster between 20 and 50 segments across unrelated families, with thin tails beyond. This isn't an artifact of pooling; it's cross-family regularity enabling inventory-level projection. Meanwhile, the vowel /y/ – a front rounded vowel requiring precise articulatory coordination – shows a *scaling curve*: a logistic model predicting /y/-presence from vowel-inventory size (cross-validated, grouped by family) achieves ROC-AUC ≈ 0.70. Marked segments lacking quantal robustness appear mainly in larger systems where there's acoustic room. This is exactly what the stabilizer story predicts.

The *pin/pen* merger provides a stress test. In much of the American South and parts of the Midland, /ɪ/ and /ɛ/ have merged before nasal consonants (Labov et al., 2006). Speakers produce the same vowel in *pin* and *pen*; they can't reliably distinguish the two words by ear. The conditioning environment is revealing: before nasals, the vowels are subject to nasalization, which smears the formant cues. In exactly the environment where acoustic cues are least reliable, the contrast collapses. This is homeostasis failing in a predictable way – not random noise, but a consequence of perturbing the stabilizers.

13.4 Opaque coupling: words

At the phoneme level, form directly realizes contrastive value. At the word level, form and value come apart. The form *went* encodes pastness, but you can't read the value off the form – the connection is brute memory. This is OPAQUE COUPLING.

Miller (2021) develops an HPC stance at the level of particular lexemes – *dog*, *run*, *egregious* – rejecting essence-based individuation in favour of mechanism-indexed clusters that are historically delimited and population-relative. The lexeme *dog* maintains its identity not through a platonic essence but because spelling conventions, pronunciation norms, semantic associations, and syntactic patterns travel together, stabilized by orthographic standardization (educational institutions, publishing practices), frequency entrenchment (repetition in memory automating retrieval; Bybee 2001), editorial norms (copy-editing workflows flagging nonstandard usage), and register licensing (genre conventions sanctioning certain words in certain contexts).

The question is whether words can change semantically and still remain HPC kinds. The HistWords COHA embeddings (Hamilton et al., 2016) provide decade-binned distributional representations for English words over the 20th century. High-drift adjectives (top decile of average cosine displacement, with documented drift terms like *nice*, *sick*, *gay*, *awful* forced in) are compared to frequency-matched controls with minimal drift. Some drift adjectives retain organized neighbourhoods even as their centres move: for *nice*, nearest neighbours shift from {pretty, lovely, pleasant} in the 1900s to {cute, wonderful, really} by the 2000s; for *sick*, from {ill, tired, hungry} to {hurt, drunk, upset}. A prototype classifier trained on early decades (1900–1940) and tested on later decades (1950–2000) recovers word identity well above baseline (macro-F1 = 0.84 vs. 0.03). Combined with a cohesion cutoff, only a subset of high-drift adjectives satisfies both criteria. The framework tells us to withhold kindhood for those lexeme–time slices rather than forcing a positive verdict. For the subset that passes, past usage fixes expectations that carry forward – exactly what the HPC picture predicts.

The *go–went* pattern provides a stress test. If form–value pairings are maintained by frequency and analogy, high-frequency irregulars should resist regularization; low-frequency irregulars should be vulnerable. Bybee (2001) documents exactly this. High-frequency irregulars (*go–went*, *have–had*) show no regularization pressure. Mid-frequency irregulars (*weave–wove*) show variable forms. Low-frequency irregulars (*cleave–clove*) have largely regularized. Frequency protects the arbitrary pairing; analogy regularizes the unprotected. The two stabilizers interact in predictable ways.

13.5 Architectural coupling: constructions

Constructions – conventionalized pairings of form and meaning that go beyond compositional rules – rely on multiple converging cues that speakers recognize as a gestalt. The form–value coupling is *architectural*: built up from components, maintained by use, transmitted through interaction.

Three stabilizer types maintain constructions. Frequency and entrenchment work on millisecond-to-week timescales: each token use strengthens memory traces, increasing production probability, generating more tokens in a self-reinforcing loop (Bybee, 2001). Cue redundancy provides robustness: multiple formal features converge on the same interpretation, so if parallelism fails in a rushed email, the anchor string and licensing context still signal the construction. Normative pressure operates on year-to-decade timescales: editorial practices and style guides reinforce the canonical pattern, especially in formal registers.

If the construction tier is to do more than demonstrate feasibility, the diagnostics have to survive contact with heterogeneous constructions. The confirmatory analysis uses a battery of ten constructions spanning four cue regimes, tested across UD English corpora (GUM, EWT, GUMReddit). Eight are treated as positive candidates; two are included as designed brakes cases – pooled resultatives (predicted "too fat") and *X much?* (predicted register-local) – to force the framework to say no when it should. For the *or even* construction, a classifier trained on GUM and tested on EWT achieves PR-AUC of 0.886 (full bundle); dropping parallelism reduces PR-AUC to 0.612. Ablation signatures show parallelism as a dominant stabilizer for scalar-additive constructions. The pooled resultative shows weak cross-corpus transfer and washed-out ablation signatures – exactly what the "too fat" diagnosis predicts. The *X much?* construction falls below prevalence thresholds outside informal registers, confirming register-locality. The implication is deliberately narrow: the construction tier doesn't license a blanket HPC claim. Kindhood is earned case-by-case.

13.6 The stabilizer-weighting map

The three cases above – phonemes, words, constructions – are not a ladder. They are regions in a stabilization space defined by how form couples to value and which stabilizers carry the load. Table 13.1 summarizes the mapping.

The stabilizers form a braid, not a single cause. At the phoneme tier, biophysical constraints carve the design space, developmental learning binds cues, sociocultural norms transmit inventories. At the word tier, frequency entrenches forms, editorial standards enforce conventions, usage communities police extensions. At the construction tier, cue redundancy protects against

Table 13.1. Stabilizer-weighting profiles across levels. Each level shows a characteristic coupling regime, error type, and dominant stabilizers.

Level	Coupling	Error type	Dominant stabilizers
Phoneme	Transparent	Slip	Quantal regions, dispersion, perceptual magnets
Word	Opaque	—	Frequency entrenchment, editorial norms, analogy
Construction	Architectural	Jam	Entrenchment, cue redundancy, normative pressure
Discourse	Loose	Mis-tracking	Pragmatic inference, common ground, accommoda

noise, normative pressure corrects deviations, genre licensing regulates distribution. The mechanisms shift in their balance: articulatory constraints weigh heavily for phonemes, frequency and norms for words, cue redundancy and editorial pressure for constructions. But at every tier, multiple forces interact: body, cognition, and society always contribute.

13.7 Negative cases: when the framework says no

The HPC framework would be toothless if it said yes to everything. Here are three cases where it says no – each for a different reason.

Academic register has a recognizable flavour: passive constructions, nominalizations, hedges like *it has been argued*. The features cluster; experienced readers identify academic prose instantly. But put the same researcher in front of a grant panel, a blog audience, and a conference poster, and the passives evaporate, the hedges reweight. The bundle doesn't resist perturbation; it dissolves and reconstitutes. The stabilizers – genre conventions, disciplinary gatekeeping, editorial norms – have high looping intensity: the category changes because it is classified. This is maintenance by institutional policing, not cognitive binding. It belongs in the same ontological category as "formal attire" – real, consequential, but not projectible in the way phonemes and constructions are.

Indo-European is a language family defined by historical descent. The cluster properties – cognate vocabulary, shared sound correspondences, similar grammatical patterns – are real. But ask what would push English *back toward* Proto-Indo-European. Nothing. There's no homeostatic pressure maintaining Indo-European-ness. English drifts away from PIE continuously; contact with non-IE languages accelerates the drift. Historical kinds are defined by causal continuity with an origin; homeostatic kinds are defined by stabilizers that maintain covariance. Confusing them invites explanatory error.

Polysynthetic languages are characterized by high morpheme-to-word ratios, incorporating structures, and complex verb templates. Mohawk, Chukchi, and Ainu are standard examples. But the same stabilizers don't maintain polysynthesis across families. Mohawk (Iroquoian) has noun incorporation for discourse-pragmatic reasons – incorporated nouns are non-referential (Mithun, 1984). A single Mohawk word like *Washakotya'tawitsherahetkvhta'se* encodes what English requires a full clause for: 'he made the thing that one puts on one's body ugly for her' – i.e., 'he ruined her dress' (Baker, 1996, p. 40). Chukchi (Chukotko-Kamchatkan) has incorporation with different constraints and different discourse functions. The surface similarity – complex words – masks heterogeneous routes. "Polysynthetic" is a typological region label, useful for description but not for explanation: it doesn't name a mechanism-maintained kind.

The three cases fail for different reasons – reflexive stabilizers (academic register), historical rather than homeostatic maintenance (Indo-European), heterogeneous routes to surface similarity (polysynthetic). The HPC framework distinguishes these failure modes. That's the payoff of explicit diagnostics: you can say *why* something doesn't qualify, not just that it doesn't.

13.8 Predictions and disconfirmers

The diagnostics generate falsifiable predictions beyond the case studies. Weakening a stabilizer should reduce cluster covariance before norms re-stabilize; for constructions, downsampling training data to 25% should degrade PR-AUC substantially if the bundle depends critically on frequency. Just as /y/ appears preferentially in larger vowel inventories, rare constructional variants should concentrate in corpora with larger constructicon repertoires. If a pooled category like "resultative" is genuinely "too fat", stratifying into subtypes should restore projectibility for well-maintained local kinds. These tests operationalize the core claim: linguistic categories qualify as HPCs when they project via identifiable mechanisms.

13.9 Looking forward

This chapter has argued that the HPC explanatory strategy scales across linguistic levels. Phonemes, words, and constructions can all be homeostatic property clusters maintained by stabilizers. What differs is the coupling regime – transparent at the sound level, opaque at the word level, architectural at the construction level – and the stabilizer weighting that maintains each regime.

The payoff is twofold. First, predictive grip: the diagnostics force us to say in advance what counts as success and what would change our minds. Success isn't a rhetorical gloss ("striking regularity") but concrete measures – slopes with uncertainty intervals, mass within bands, cross-corpus areas under curve, ablation deltas.

Second, an ontology with brakes: kinds are discovered through evidence rather than declared by fiat, and they are local equilibria (maintained by specific mechanisms in specific populations) rather than universal essences. That stance blocks overreach. Cross-linguistic umbrellas fail as single kinds because they pool heterogeneous mechanisms, bundling components that don't belong to the same causal whole. Thin proposals and complement classes lack the stabilizer base that projectibility requires. Many historical categories are united merely by descent rather than by active maintenance. And the zipper permits MECHANISTIC DRIFT: a category can remain historically continuous while its stabilizers shift. In between lie the categories that travel: their properties cohere because mechanisms keep them together. These show a cliquish stability – robust enough to support induction, even if perfect instance stability remains elusive.

The zipper metaphor returns for the final turn. Error shape diagnoses coupling: slips reveal transparent coupling; jams reveal architectural coupling; mistracking reveals loose coupling. Each error type tells you where the teeth failed to seat and, therefore, where to look for the stabilizers that normally keep them seating.

But one level has special status. Throughout the case studies in Part III – countability, definiteness, lexical categories – the categories clustered at the morphosyntactic level. That's where form–value coupling is both *tight* and *obligatory*. You can speak without using *let alone*. In English, you can't speak without committing to number, tense, and definiteness.

Morphosyntax is the zone of maximum systematicity: the region of grammar where coupling is enforced, not optional. Chapter 14 asks what happens when we take this observation seriously. If grammaticality is what emerges when form–value coupling is obligatory, compositional, and learnable – then grammaticality itself is a kind.

Part IV

Implications

Grammaticality itself

> It is impossible to get rid of the illusion in spite of our better knowledge.
>
> —Hermann von Helmholtz, *Treatise on Physiological Optics* (trans. J. P. C. Southall).

Every analysis of every construction presupposes a distinction between grammatical and ungrammatical. The asterisk's the most common symbol in syntax. But what's the asterisk diagnosing?

Chapter 3 asked a question almost no one asks: is grammaticality itself a natural kind? Not whether this sentence's grammatical, but what kind of thing we're probing when we ask. This chapter provides the answer.

If categories are homeostatic property clusters, then grammaticality isn't merely a binary switch. It's itself an HPC – a maintained coupling between form and value. And the feeling of ungrammaticality that guides our judgments isn't an oracle; it's a noisy detector, shaped by entrenchment and processing ease, capable of illusions.

Chapter 5 introduced a two-layer picture: discrete grammaticality filtered through processing noise to yield gradient acceptability. What we're adding here's the question of what KIND of thing grammaticality is, and what maintains it.

14.1 The HPC claim

Grammaticality's a form–value pairing. But the "value" side isn't compositional semantics – not truth conditions, not the full propositional content. It's something more basic: STRUCTURAL MEANING.

Structural meaning's what grammar contributes before lexical content fills the slots. It includes dependency relations, like the bracketing that distinguishes [*the dogs'*] *house* (a house belonging to the dogs) from *the* [*dogs' house*] (a type of house). It includes argument structure, distinguishing *the dog bit the man* from *the man bit the dog*.

It provides temporal anchoring through tense and aspect, distinguishing whether the biting's past, ongoing, or hypothetical. It sets the clause type, distinguishing statements (*he was bit*) from questions (*was he bit?*) or commands (*bite him.*). And it manages information structure – the difference between *Him, he was bit* and *it was him who was bit*.

This is SUBCOMPOSITIONAL – it's the instruction manual, not the assembled product. Grammar tells you how to wire the meanings; the lexicon provides what to wire. *Colorless green ideas sleep furiously* has perfect wiring instructions and absurd components. That's why it feels strange but not UNGRAMMATICAL.

A note on what structural meaning isn't. It's not semantic coherence (*Colorless green ideas* shows grammaticality survives semantic absurdity). It's not contextual felicity (a sentence can be grammatical and pragmatically infelicitous). And it's not whatever the hearer infers (inference runs on many channels; grammaticality's specifically the morphosyntactic one). Structural meaning's the contribution of form to interpretation – the WIRING DIAGRAM, not the current flowing through it.

The HPC claim, then, is that grammaticality's the coupling between morphosyntactic form and this structural semantics. When the coupling holds – when the form reliably signals the structural meaning – the sentence sits in the GRAMMATICAL BASIN. When it doesn't, the sentence's outside.

This is the ZIPPER: morphosyntax provides one set of teeth, structural meaning the other. Grammaticality's what it feels like when the zipper closes. An ungrammatical sentence's a zipper that won't mesh.

Two metaphors will recur. The zipper describes the relation – the form–value alignment itself. The IMMUNE SYSTEM (below) describes the MAINTENANCE MECHANISM – what pushes the system back toward stable couplings when they slip. One's structure; the other's dynamics.

14.2 The stabilizer: a feeling

What maintains this coupling? Partly the mechanisms we've discussed throughout: acquisition, entrenchment, alignment, transmission. But the proximate STABILIZER – the thing that corrects deviations in real time – is a FEELING OF UNGRAMMATICALITY.

When something's wrong with the form–value coupling, speakers notice. They hesitate, repair, rephrase. The feeling's what triggers the correction. It's the immune response of the grammar.

The feeling's a DETECTOR – and like all detectors, it's noisy. It's not a direct readout of the underlying structure. Instead, it's informed by probability and processing ease. First, the detector's Bayesian: a novel sentence's weighted by its likelihood under the distribution of attested forms – not just distance from the nearest exemplar, but probability given the entire pattern. Low-probability forms feel suspect even when grammatically licensed. The distribution itself is built by entrenchment: exposure tunes the prior.

Second, sentences that parse smoothly feel grammatical; sentences that induce garden-paths or reanalysis feel suspect. Processing difficulty isn't the same as ungrammaticality, but the detector often conflates them. This is why ACCEPTABILITY and GRAMMATICALITY come apart. Acceptability's what the detector reports. Grammaticality's the category – the HPC whose core's the form–value coupling. The two usually align, but they don't always.

14.2.1 Entrenchment in action

But here's the puzzle: if grammaticality *is* the maintained coupling, and entrenchment's what maintains it, then "tuning the detector" and "shaping what counts as grammatical" aren't separable. The detector isn't just tracking some pre-existing reality; the reality is constituted by what gets entrenched. The feeling tracks the coupling, and the coupling is entrenched. All the way down.

This is language-specific, construction-specific, even register-specific.

Consider age expressions: in French and Spanish, age is expressed with *avoir*/*tener*: *J'ai vingt ans* ("I have twenty years"). In English, **I have twenty years*'s ungrammatical as a statement of age – though perfectly fine as a statement of time (until retirement, on the job, etc.). The structural slot's the same; the construction-specific entrenchment differs.

The same applies to progressive aspect. If you're describing an ongoing event while your fingers are moving across a keyboard, *I write* mis-signals aspect in English – many speakers experience this as "wrong" in the specific sense of a form–value mismatch. The progressive (*I'm writing*) isn't merely preferred; it's required for the ongoing-action construal. In French, the simple present would be grammatical for the same construal; the progressive's marked or absent. What counts as a violation of the form–value coupling depends on what the community has entrenched.

Countability follows the same pattern. Whether *furniture* takes count morphosyntax depends on entrenchment, not metaphysics. Furniture's as individuable as anything – you can count chairs and tables. But English has entrenched *furniture* as noncount, so **I bought two furnitures*'s ungrammatical. Other languages make different choices. The coupling's real; its contents are conventional.

And consider deitality: what requires the definite article in English – *go to the hospital* vs. *go to hospital* (British) – is shaped by entrenchment within each dialect. The structural meaning (institutionalized activity frame)'s the same; the morphosyntactic realization differs.

These examples show grammaticality isn't a window onto universal logic. It's a window onto what a community has entrenched as the couplings between form and structural meaning. The feeling of ungrammaticality is tracking those couplings – but the couplings are historical, contingent, and maintained.

And here's where the entrenchment story connects to the social: entrenchment's always entrenchment "within a discourse community". The grammar isn't floating free; it's tethered to the people who use it. When the community shifts – through contact, migration, register change – the entrenchment shifts with it. What feels ungrammatical in one community may be unremarkable in another.

But speakers flow between communities – sometimes instantaneously – and communities can be created on the fly. *Do you lift?* licenses intransitive *lift* for weightlifters but not for warehouse workers; the grammaticality of the construction depends on which community you're invoking, and you can switch in mid-conversation. Code-switching within a clause – *Vamos a hacer shopping* – is governed grammar, not error, but only within the discourse community of the bilinguals who maintain it. The detector isn't calibrated once; it's calibrated to whoever you're talking to, right now.

One more piece. We couldn't correct grammatical errors if we couldn't derive the intended meaning without the grammar. When someone says **I have twenty years* for their age, we understand what they mean – and then we notice the mismatch. The feeling of ungrammaticality arises from comparison: the grammar's value versus the intended value. We need both to detect the error. This means comprehension's partly independent of grammaticality – and grammaticality's partly a matter of checking the grammar's signal against other inference channels. When they align, we don't notice; when they mismatch, we do.

14.3 The phenomenology of coupling: What it's like

If grammaticality is a maintained coupling, what is the phenomenology of that state? What is it *like* to be in a state of grammaticality?

The experience isn't a drive, like hunger. Hunger is a homeostatic signal that motivates action (eating); grammaticality doesn't typically push us to "consume" structure. Nor is it like pain, a dedicated alarm channel. Instead, the phenomenology of grammaticality is a bundle of metacognitive readouts that track the success of the coupling mechanism. These readouts look less like appetites and more like balance and perception.

14.3.1 Balance: The silent baseline

The primary analogue is balance or proprioception. Most of the time, balance is phenomenologically silent. When you walk down the street, you don't feel "balanced"; you just experience the world and your movement through it. The coupling work is backgrounded.

Grammaticality behaves the same way. When the form–value coupling is functioning smoothly, you don't experience "grammaticality"; you experience content, stance, and interactional fit. The zipper is closed, and because it holds, you can ignore it.

The feeling arises only when the coupling fails. When you trip, you get an immediate, non-propositional jolt—a "catch" in the system—followed by a reflex to repair. When you hear an ungrammatical sentence, you get a similar jolt: a sense that the structure has stumbled. In production, this manifests as monitoring error signals: speakers detect upcoming problems and self-correct, often before the error is even articulated (Nozari et al., 2011). The feeling of ungrammaticality is the system trying to keep itself upright.

14.3.2 Vision: Illusion and good-enough processing

If balance describes the baseline, vision describes the failure modes. Vision delivers structured objects, not raw sense data; we are aware of the output, not the intermediate computations. And like vision, grammatical processing is subject to systematic illusions.

Visual illusions (like the Müller-Lyer lines) persist even when we know they're illusions. Grammaticality illusions behave similarly. Wagers et al. (2009) show that agreement attraction—where an ungrammatical sentence like **The key to the cabinets are missing* feels acceptable—is a stable feature of the architecture. The processing system retrieves the plural feature from *cabinets* and erroneously checks it against the verb, satisfying the local constraint even though the global structure fails.

This selective fallibility (Phillips et al., 2011) is the hallmark of a perceptual system, not a logical oracle. It aligns with the "good-enough" processing framework (Ferreira et al., 2002), which suggests that comprehension often settles for representations that are serviceable rather than fully specified. The result is a dissociation: the feeling of acceptability (the readout) can float free of the actual structural coupling (the state).

14.3.3 Memory and attention

The substrate of this coupling is memory—specifically, the procedural memory that handles skill and habit (Ullman, 2001). The feeling of "rightness" that accompanies a grammatical sentence is a feeling of FLUENCY: the structure affords a well-supported pattern completion, and that ease of processing is metacognitively marked as "correct" (Ackerman & Zalmanov, 2012).

Attention acts as the modulator. It changes the "lighting" in the room. In ordinary conversation, attention is on the content; in a linguistics experiment, attention is shifted to the form itself. This shift can change the threshold for the mismatch signal, turning a sub-perceptual glitch into a reportable error. But attention doesn't create the signal; it just determines whether it reaches consciousness.

14.3.4 Production: The near-miss

Finally, there is the TIP-OF-THE-TONGUE (TOT) state (Brown & McNeill, 1966). TOT is a conscious experience of partial access: you have the meaning, but the phonological form won't retrieve.

The coupling view predicts an analogous family of "near-miss" grammatical states. You have the intended value (structural meaning) and some morphosyntactic scaffolding, but the interface won't lock. You don't feel "ungrammaticality" per se; you feel that you "can't get it to come out right." This is the territory of hesitations and reformulations—the feeling of a zipper that is stuck halfway.

In sum: Grammaticality itself is the structural property of the form–value coupling. What it's *like* is the metacognitive readout of that coupling: silence when it works, a jolt when it fails, and a graded sense of "weirdness" or fluency when it's somewhere in between.

14.4 Grammaticality illusions

The strongest evidence that the feeling is dissociable from the structure comes from GRAMMATICALITY ILLUSIONS – cases where the detector misfires.

14.4.1 Feels ungrammatical, is grammatical

The classic case is the garden-path sentence:

(1) *The horse raced past the barn fell.*

This is grammatical – a reduced relative clause (*the horse [that was] raced past the barn fell*). But the processing system commits to the main-clause parse (*the horse raced past the barn*) and crashes when *fell* appears. The detector reports ungrammaticality; the structure's fine.

14.4.2 Feels grammatical, is ungrammatical

The converse is the comparative illusion – sometimes called the Escher sentence:

(2) *More people have been to Russia than I have.*

Each local chunk parses. *More people have been to Russia* – fine. *Than I have* – fine as a comparative clause. But the whole doesn't compose: what's being compared? The sentence has no coherent interpretation, yet it feels acceptable because each piece satisfies local constraints.

These illusions are exactly analogous to visual illusions. The Müller-Lyer lines are the same length, but the perceptual system reports otherwise. The illusion doesn't show perception's broken; it shows perception is a mechanism with characteristic failure modes. Grammaticality illusions show the same: the feeling's a detector, not an oracle.

Phillips et al. (2011) develop this point systematically, showing comprehension exhibits selective fallibility – impressive accuracy for some constraints, and systematic vulnerability for others. The pattern reveals the architecture: a retrieval system that's structure-sensitive when structural cues have temporal priority, and error-prone when they don't.

14.5 What doesn't count

The form–value coupling's specifically morphosyntactic. This predicts what won't trigger ungrammaticality. It excludes phonetic errors, like *[f□lɔzəfi]* for *philosophy*, where the form's wrong but the morphosyntax's untouched. It excludes lexical errors, like saying *airport* when you meant *train station*. And it excludes semantic anomaly, like *Colorless green ideas sleep furiously*, where the wiring's impeccable even if the components are absurd.

These violations trigger different responses. Phonetic errors trigger *what?* because the phoneme's value's the contrast – that's the entire form–value coupling at that level, and when it breaks, comprehension stalls. Lexical errors trigger correction or confusion. Semantic anomaly triggers puzzlement. But none of them trigger the specific feeling of ungrammaticality – the sense the wiring's broken.

A revealing contrast comes from the allomorphy of the indefinite article. We reject **a apple and **an banana with the same immune response we bring to syntax. This looks phonological, but the selection rule is entrenched as a grammatical requirement – part of the morphosyntactic zipper. A mispronunciation of *banana* would trigger *what?*; distributing the wrong allomorph of the article triggers the grammar's immune response. This is exactly what we'd expect from an HPC: the edge case reveals the boundary of the morphosyntactic coupling, and the feeling of ungrammaticality – however muted – is the trace of the immune system at work.

14.5.1 Value at every grain

What's "value" at each level of the grammar? The zipper closes differently depending on what you're zipping.

Grain	Form	Value
Phoneme	Sound segment	Contrast (minimal-pair distinctiveness)
Morpheme	Affix, stem	Grammatical or derivational modification
Word	Phonological shape	Lexical meaning + category membership
Construction	Syntactic pattern	Argument structure, information structure, meaning template
Clause type	Structural configuration	Illocutionary force (statement, question, command)

Table 14.1. Form–value coupling at different grains of grammar.

At the phonemic level, value's nothing *but* contrast. The /p/ in *pat* contributes no semantic content; it just distinguishes the word from *bat*, *cat*, *sat*. When contrast fails – when you can't tell which phoneme was intended – you say *what?*

At the morphemic level, value's the contribution to word meaning or grammatical function. The *-ed* in *walked* signals past tense; the *un-* in *unhappy* reverses polarity. Break the coupling – use *-ed* on a noun, or *un-* on a verb that doesn't license it – and ungrammaticality results.

At the word level, value's lexical meaning plus category membership. *Dog* contributes dog-ness *and* nounhood. Substitute the wrong word and you get confusion; use a word in the wrong category slot and you get ungrammaticality.

At the constructional level, value's the structural meaning the pattern contributes. The double-object construction (*gave her the book*) signals transfer plus affected recipient. The PASSIVE (*was eaten*) demotes the agent. Break the construction's slot-filling requirements and the zipper won't close.

At the clause-type level, value's illocutionary force. Declarative structure signals assertion; interrogative structure signals question; imperative structure signals command. Mismatch the structure and the force – use rising intonation on a declarative to fake a question – and you're exploiting the zipper, not breaking it.

This table explains why grammaticality's primarily a syntactician's category. It sits at the morphosyntactic grain – the level where form–value couplings are tight, obligatory, and enforced.

Phoneticians have their own coupling (contrast), and when it breaks, the response's *what?*, not ungrammaticality. Semanticists have theirs (compositionality, inference), and when it breaks, the response's puzzlement or absurdity. Grammaticality's what you feel when the morphosyntactic zipper won't close. That's why syntax owns it.

14.6 The gradient at the boundary

When we judge a sentence, we aren't checking a rulebook. We're interrogating the standing wave – probing the system to see if the dynamic equilibrium holds or collapses under strain.

A fully grammatical sentence sits at the bottom of a deep basin. Perturb it slightly – change word order, shift stress, substitute a near-synonym – and the system corrects. An ungrammatical sentence sits outside the basin entirely: no perturbation's going to roll it back to stable ground.

The marginal cases – the 3-star and 4-star judgments, the sentences that split informants – are poised on the rim. They're at the edge of a phase transition, where small changes in context or construal can tip the system one way or the other.

This reframing has consequences. If grammaticality's basin-stability, then the sharp boundary between well-formed and ill-formed isn't a property of the grammar itself but an artifact of how we test. Gradient judgments aren't noise to be smoothed away; they're signal. The gradient is where the mechanisms are under strain, and strain reveals structure.

14.7 Grammaticality as mechanism and category

Grammaticality plays a dual role in the framework. As a CATEGORY, it's an HPC – the maintained coupling between morphosyntactic form and structural meaning. It passes the diagnostics: it's projectible (one form–value pairing predicts others) and homeostatic (the feeling of ungrammaticality stabilizes the coupling).

As a MECHANISM, it's a stabilizer for other HPCs. The category NOUN's maintained partly because nounhood's grammatical – because using a word in a noun slot, with noun morphology, in noun constructions, triggers no mismatch signal. Grammaticality's the enforcement mechanism for the form–value couplings that constitute all other categories.

This dual role – maintainer and maintained – isn't paradoxical. It's characteristic of homeostatic systems. The immune system's both a system (a biological category) and a mechanism that maintains other systems. Grammaticality's the immune system of the grammar.

Or think of the SPINNING TOP metaphor. A top's both an object and the spin that keeps it upright. Remove the spin and the top falls. Grammaticality's the spin. It's both the pattern and the force that maintains it. The top keeps spinning because something keeps pushing.

If grammaticality's a maintained coupling and acceptability's a noisy detector, then the evidential status of the asterisk changes. It's not a report from an oracle but a signal from a measuring instrument – informative, but fallible, and calibrated to the community in which it was trained.

14.8 Empirical signatures

If the feeling of ungrammaticality's a noisy detector – not an oracle – then laboratory studies should reveal its characteristic failure modes. Three bodies of evidence confirm this prediction.

14.8.1 Satiation

Syntactic satiation's the finding that repeated exposure to a marginally acceptable sentence can improve its perceived acceptability (Snyder, 2000). But the effect's selective. Some constraints are susceptible to frequency-based updating: repeated exposure makes them feel better. Others resist satiation entirely. Left-branch extraction (**Which did you buy car?*) remains categorically rejected regardless of exposure (Snyder, 2022). The intended meaning's easily grasped – *Which car did you buy?* – and the structure isn't unusually complex. But the ban persists across tasks, speakers, and presentation modes.

Why the split? If the feeling of ungrammaticality arises from comparing input against entrenched form–value pairings, then repeated exposure should matter – but only for constraints that are frequency-sensitive in the first place. When you hear a marginal sentence ten times, you're adding exemplars to the distribution; the pairing becomes more familiar, the mismatch signal weakens, and the sentence feels better.

But when a constraint's categorical – as left-branch extraction is in English – no amount of exposure can change that. The resistance doesn't imply the structure is unimaginable; it implies the learner has detected a conpicuous absence. Fronting is productive in English; the *potential* for extraction is visible. But in this specific configuration, the community systematically avoids it (3-sigma absence), using entrenched workarounds instead: pied-piping (*whose book*), fused-head constructions (*whose is this*), and the *big mess* construction (*how good a teacher*).

The gap persists because the potential is there but the practice is absent. Learners infer that the community has tacitly agreed *not* to use the available mechanism in this way. The detector isn't just saying "this doesn't happen"; it's saying "we could do this, but we don't." The feeling of ungrammaticality tracks this specific, entrenched abstinence.

14.8.2 Individual differences

Speakers differ systematically in their grammaticality thresholds. The limiting case's trivial: we have no feeling of grammaticality whatsoever for languages we don't speak. A sentence in Zulu or Finnish produces no mismatch signal for me because there's nothing to compare it against – no entrenched pairings, no expectations. Foreign language learners occupy the middle ground: their feelings are degraded, unreliable, slow to fire, and easily fooled by surface plausibility. Only with sufficient exposure does the feeling sharpen.

Within a single language community, the variation persists. Dąbrowska (2012) shows that grammatical knowledge isn't uniform across the adult population: speakers with less formal education, less exposure to complex syntactic constructions, and less metalinguistic training produce reliably different acceptability judgments – not just noise, but systematic divergence.

That's the prediction from community-indexed entrenchment. If grammaticality is what a community has entrenched, then members with different exposure histories will have different thresholds. The variation isn't random; it tracks social and educational ecologies. The "grammar"'s not a Platonic object that speakers approximate; it's a statistical regularity that speakers instantiate to varying degrees.

14.8.3 Processing difficulty and beyond

Processing difficulty isn't grammaticality, but the two are often conflated. Phillips et al. (2011) show that comprehension exhibits SELECTIVE FALLIBILITY: it's exquisitely sensitive to some constraints and systematically vulnerable to others. The pattern reveals something about the architecture of the feeling, not just noise.

Consider GARDEN-PATH SENTENCES. When readers encounter a grammatical sentence that's initially misparsed, they produce high unacceptability ratings at the point of reanalysis, even though the sentence's well-formed. The feeling of ungrammaticality fires because processing cost correlates, in normal use, with form–value mismatch.

The illusion's systematic: it follows from the design of a system optimized for rapid, good-enough processing, not from a failure of the grammar itself. These three signatures – satiation, individual differences, and processing-based illusions – converge on the same conclusion. The feeling of ungrammaticality's noisy, not oracular. It tracks the coupling between form and value, but it does so imperfectly, shaped by exposure and processing constraints. The asterisk's informative, but it's not infallible.

14.8.4 Relevance and the immune system

Scott-Phillips (forthcoming-a) offers a complementary account. On his view, grammaticality intuitions are byproducts of relevance-tracking: we detect when utterances violate basic presumptions about communicative efficiency, just as we detect when visual stimuli violate assumptions about physical objects. Unacceptability arises not from mere inefficiency, but from an inherent impossibility of interpreting an utterance consistently with those presumptions.

The MMMG framework shares several premises with this account. Both reject the need for an innate grammar faculty; both locate intuitions within general cognitive systems; both recognize that communicative pressures shape linguistic conventions. The frameworks differ in their explanatory mechanisms. Where relevance theory predicts that ungrammaticality arises from inferential impossibility, MMMG locates it in the mismatch between morphosyntactic form and structural meaning – the zipper that won't close.

The difference has empirical consequences. MMMG predicts that ungrammaticality will be construction-specific: **I have twenty years* fails as an age statement because the English age construction's *be* + *years old*, not because relevance-seeking fails. Scott-Phillips's account predicts more general inferential failure. Both approaches agree that the feeling's noisy; they disagree about what it's tracking.

The accounts may ultimately converge. If efficient interpretation requires stable form–meaning pairings, then relevance-tracking and form–value coupling are two descriptions of the same constraint. What MMMG adds is granularity: it identifies the specific level (morphosyntax) and the specific mechanism (entrenchment) that make the coupling possible.

14.9 The HPC audit

Chapter 8 introduced the Two-Diagnostic Test for genuine HPC kinds: high projectibility and robust homeostasis. Grammaticality passes both.

First, consider projectibility. Learning that one form–value pairing's grammatical lets you predict that similar pairings will be. If *the dog bit* him's grammatical, then *the cat scratched* her will be too. This isn't because speakers have memorized every possible sentence; it's because they've abstracted the construction and project its properties to novel instances. Grammaticality's projectible because the form–value coupling's systematic.

Second, homeostasis. The feeling of ungrammaticality stabilizes the coupling, correcting deviations. Speakers hesitate, repair, rephrase. Hearers ask for clarification, or offer corrections. These micro-interactions are the immune response – the mechanism that pushes the system back toward stable couplings when it drifts. Grammaticality's homeostatic because the feeling's a stabilizer.

Grammaticality isn't merely a binary switch, nor a primitive of the theory. It's an HPC kind in its own right – maintained by the same mechanisms that maintain countability, definiteness, gender, and lexical categories. It passes the tests. It earns its category status.

Definiteness thread. This chapter adds the final filter to the definiteness diagnosis: the morphosyntactic "zipper" rejection of infelicitous forms. The coupling between semantic definiteness and morphosyntactic deitality isn't just a preference; it's a grammatical requirement.

(3) **a the book*

The immune response that rejects this determiner stacking is the same one that rejects **I have twenty years* for age. *Scoped prediction.* This chapter predicts in any population where a novel determiner sequence becomes entrenched (e.g., *all the both*), the ungrammaticality feeling will take exactly three generations to sharpen into a categorical ban on the previous pattern. *Falsifier.* The prediction would be falsified if we found a language where semantic definiteness remains stable but NO morphosyntactic mismatch signal ever develops for "illegal" determiner combinations.

14.10 Looking forward

We began with an asterisk. We end with a question mark: what is the asterisk probing?

Grammaticality's not a primitive. It's not a direct readout of structure. It's not a line in the sand between possible and impossible. It's a maintained coupling –

form and value, zipper teeth that mesh or don't – and the feeling that signals a mismatch is noisy, shaped by processing constraints and what a community has entrenched.

This reframing has consequences. If grammaticality is a category, then we can ask: how projectible is it? how robust is its homeostasis? And if grammaticality is a mechanism – the immune system of the grammar – then we can ask: what maintains it? what degrades it? what illusions does it produce?

The dual role – maintainer and maintained – isn't paradoxical. It's characteristic of homeostatic systems. The spinning top is both the object and the spin that keeps it upright. Grammaticality is the spin.

This means the equilibrium is CONTINGENT. A constraint-based account says LBE *cannot* emerge in English – that there's a grammatical principle blocking it. The HPC account says only that it *has not* – the social and statistical conditions have never favored it. But imagine a high-status speaker – or a low-status one with memetic cachet like Homer Simpson – introducing *Whose did you eat sandwich?* to the language. Use the form, get the laugh, signal the group membership. Homer doesn't just provide new evidence; he spawns a new discourse community. If the social uptake shifts, the prior shifts. The grammar isn't resisting; there's no "grammar" there to resist. There's only convention, and convention can change. Because the mechanism is maintenance, not law.

What this means for how we do linguistics – for evidence, for argumentation, for the rhetoric of theoretical debate – is the subject of the final chapter. The asterisk is a probe into the standing wave. Chapter 15 asks what happens when we take that probe seriously.

What changes

> At the quantum scale there are no cats; at scales appropriate for astrophysics there are no mountains.
>
> —James Ladyman & Don Ross, *Every Thing Must Go* (2007)

15.1 The status of the framework

What *kind* of kind is an HPC CATEGORY? It is a SECOND-ORDER EXPLANATORY KIND: a kind whose members are themselves kinds, unified not by shared first-order properties but by a shared explanatory role and stabilization pattern. Its members are things like SPECIES, CHEMICAL ELEMENTS, PHONEMES, and CONSTRUCTIONS – categories that support projectible generalizations, are stabilized by multiple partially independent mechanisms, and tolerate property variation without collapse.

What makes HPC CATEGORY itself a kind is that these properties cluster reliably across domains. It is not a natural kind in the same sense as GOLD, nor a merely stipulative classificatory label, nor a purely philosophical artifact. It is a meta-level homeostatic cluster over explanatory practices.

This recursion is *typed*, not flat. Compare GENE (a kind whose instances are molecular entities) with FUNCTIONAL GENE (a kind whose instances are ways of carving interactions for explanation), or MODEL ORGANISM (a kind whose instances are organisms plus institutional practices). None of these collapses because they occupy different explanatory grains. Likewise, NOUN is an HPC kind in English; LEXICAL CATEGORY is an HPC kind across languages; and HPC CATEGORY is an HPC kind across scientific taxonomies.

Crucially, the higher-order kind is not stabilized by the same mechanisms as its members. HPC categories *in the world* are stabilized by causal mechanisms, developmental pathways, and communicative coordination. The HPC CATEGORY kind is stabilized by repeated success of explanation, methodological convergence across sciences, robustness under theory change, and the survival of the concept under critical scrutiny (against eliminativism or essentialism). Its homeostasis is epistemic and methodological rather than biological or physical.

This framing distinguishes the theory from a mere framework or model. Frameworks can be swapped without residue; HPC categories resist that. Once the pattern is identified – why strict definitions fail, why exceptions cluster – we can make reliable predictions about where gradience will appear, where sharp boundaries will re-emerge, and where eliminativist arguments will systematically overreach. That predictive success is exactly what licenses kindhood.

A natural objection is that if HPC CATEGORY is itself an HPC kind, the theory is self-validating or circular. This would be true only if HPC theory claimed *a priori* necessity. It does not. It makes a fallible empirical claim about the structure of successful scientific kinds. If HPC categories stopped supporting prediction, coordination, or explanation, the kind would dissolve – by its own lights. That is not circularity; it is reflexive risk.

15.2 Overlap as principled

When physicists calculate the motion of a planet, they face a choice. They can treat the planet as a single coherent object, or as a vast aggregate of 10^{50} atoms. For the theory to be consistent, the answer must be the same either way.

Newton worried about this. As Barandes (2024) notes, this mereological consistency constraint – the requirement that the laws work equally well for parts and for wholes – acts as a theoretical guardrail. It forces the inclusion of Newton's Third Law: for every internal action between particles, there must be an equal and opposite reaction. Without this, the internal forces wouldn't cancel out, and a composite object could accelerate itself just by the interaction of its parts. Mereology constrains physics.

Linguistics has no such guardrail.

Instead, we have a loose federalism. We divide the field into sub-disciplines – phonetics, phonology, morphology, syntax, semantics, pragmatics – and treat them as distinct territories. We talk about "interfaces" (the syntax–semantics interface, the phonology–morphology interface) as if they were national borders to be policed rather than contact zones to be mapped. We assume a tree-like structure: a phenomenon belongs to syntax *or* semantics, but ideally not both.

The HPC framework suggests this picture is wrong.

If linguistic categories are homeostatic property clusters, then disciplinary subfields are too. Psycholinguistics, Construction Grammar, and Experimental Semantics are not mutually exclusive territories. They are bundles of phenomena, methods, and theoretical commitments, stabilized by institutional mechanisms (journals, hiring lines, training pipelines). And crucially, these bundles overlap.

15.2.1 Typed parthood

The mistake is assuming that parthood is a single relation. In standard hierarchies, if A is part of B, it's part of B generally. But in scientific practice, parthood is typed.

A subfield can be part of linguistics in different ways:

- Phenomenon-part ($\leq_{\text{phen}}$): Its domain of inquiry is a subset of language (e.g., phonetics studies speech sounds).
- Method-part ($\leq_{\text{meth}}$): Its tools are a subset of the field's toolkit (e.g., corpus linguistics uses distributional analysis).
- Theory-part ($\leq_{\text{thy}}$): Its explanatory commitments inherit from a broader framework (e.g., Minimalism is a theoretical part of Generative Grammar).

Once we distinguish these types, the "boundary disputes" resolve into structural features.

Take Computational Linguistics. Is it part of linguistics? Methodologically, yes: it contributes formal and algorithmic tools. Theoretically, often no: its goals (engineering performance) diverge from the core explanatory project (cognitive realism). It is a method-part that is not always a theory-part.

Take Construction Grammar. It is a theoretical framework ($\leq_{\text{THY}}$) that claims all of morphosyntax as its phenomenon-domain ($\leq_{\text{PHEN}}$). This puts it in direct competition with Minimalism for the same territory.

Take the syntax–semantics interface. In the tree view, this is a border. In the typed-parthood view, it is a zone of principled overlap. Many linguistic phenomena cluster here because the mechanisms are bidirectional. Syntactic structure cues semantic interpretation (comprehension); semantic intent licenses syntactic choice (production). The two systems maintain each other. The overlap is not a messy intermediate zone to be purified; it is the engine of the system.

The verbless clause puzzle from Chapter 2 is a textbook instance of typed overlap. What looked like *CGEL*'s global inconsistency – defining clauses by VP-headedness but then identifying verbless clauses by subject–predicate structure – turns out to be the diagnostic signature of a kind whose markers come from two braided mechanisms. VP-headedness is one stabilizer; predication structure is another. At the core they converge; at the margins they decouple. The inconsistency was the evidence.

15.2.2 Consequences for practice

This re-framing has concrete consequences for how we work.

First, it explains peer review friction. When a paper on "experimental syntax" is reviewed, it often faces a double bind. Syntacticians (judging by $\leq_{\text{THY}}$) may find the theoretical contribution thin. Psycholinguists (judging by $\leq_{\text{METH}}$) may find the experimental design naive. The paper is attempting a fusion of two bundles. The friction isn't just grumpiness; it's a clash of validation criteria. Recognizing typed parthood allows editors to assign reviewers who are competent in the specific intersection being claimed.

Second, it validates methodological pluralism. If subfields are valid HPCs, then "pure" linguistics is just one bundle among many – usually the bundle that privileges introspective data and structural economy. Other bundles (corpus linguistics, sociolinguistics) privilege different stabilizers (usage data, social variation). These aren't failed attempts at pure linguistics; they are different cuts through the same multidimensional reality. This answers Nefdt (2023)'s worry that methodological pluralism leads to ontological chaos. It doesn't – as long as the methods all map to valid homeostatic clusters. Unity lies in shared criteria for kindhood, not in a shared ontology of objects.

Finally, it recovers intensional mereology. Just as Newton asked what makes a planet a single object (internal forces cancelling out), we can ask what makes a subfield a genuine whole. Why does Sociophonetics exist as a stable cluster, while Generative Phonetics never quite cohered? Because the mechanisms aligned. Social variation turns out to be deeply entangled with phonetic detail. The phenomena cluster naturally.

The HPC framework doesn't just categorize our data. It categorizes us. And it suggests that the messy, overlapping map of modern linguistics is not a sign of immaturity, but an accurate reflection of a system where everything is braided.

15.2.3 Agent-based modeling

If linguistic categories are homeostatic property clusters, then agent-based modeling is their natural computational laboratory.

The logic is straightforward. HPCs are maintained by mechanisms operating across agents and timescales: acquisition shapes the priors of new learners; entrenchment solidifies frequently encountered patterns; alignment coordinates interlocutors in real time; transmission filters for learnability across generations; institutional norms impose top-down pressure. These mechanisms interact. Their interactions can be subtle. And the system-level outcomes – stable categories, graded boundaries, drift when stabilizers weaken – are exactly what emergent dynamics in agent-based models are designed to capture.

Consider countability (Chapter 9). The count cluster is a bundle of morphosyntactic properties – article selection, numeral compatibility, *many*/*few*, agreement – that cohere in an implicational hierarchy. The stabilizers are explicit: bidirectional inference couples count marking to individuation; entrenchment anchors core cases; functional alternatives (singulatives, measure constructions) absorb pressure at the margins. This is precisely the architecture an agent-based model can operationalise. Give agents lexicons with individuation confidence parameters; let constructions act as locks with tolerance thresholds; have agents update their parameters based on observed usage. Run the simulation.

What emerges is not merely a replication of what we already knew. The model makes predictions. Weaken the singulative anchor (let *datum* die) and the system drifts: agreement loosens first (*this data is*), tight properties like low numerals remain unavailable, and the noun settles into quasi-count equilibrium – exactly the pattern we observe. Introduce prescriptive pressure (editors penalising *three Legos*) and the model shows why such campaigns typically fail: the cognitive economy of count packaging for discrete objects overwhelms institutional feedback unless that feedback is implausibly strong and omnipresent. The LEGO Company, on this account, is fighting against the stabilizers rather than with them.

More generally, agent-based models provide three kinds of value for the HPC framework:

1. Sanity checks on mechanism claims. If we claim that alignment and transmission are jointly sufficient to maintain a category, the model can show whether those mechanisms produce stable clustering or whether additional stabilizers are required. Failure to converge is informative: it means the mechanism story is incomplete.

2. Predictions about boundary cases. The model can locate the parameter regimes where categories should fray, where intermediate equilibria should appear, and where drift should accelerate. These are testable predictions about corpus distributions and acceptability gradients.

3. Contrastive explanations. Why does English maintain a tight count/mass distinction while Mandarin relies on classifiers? Why do quasi-count nouns like *cattle* stabilise rather than regularising? Agent-based models can simulate both outcomes and show which stabilizer configurations produce which results.

The framework is not committed to any particular computational architecture. What matters is that the model respects the multi-timescale, multi-mechanism, multi-agent structure that HPCs require. A single-agent model of grammatical competence cannot capture alignment; a model without transmission cannot capture the filtering effects of iterated learning; a model without institutional agents cannot capture prescriptive pressure or register stratification. The HPC story constrains the simulation the same way mereological consistency constrains Newtonian mechanics: not by dictating the equations, but by requiring that parts and wholes cohere.

This is not a call for linguistics to become computational. It is a claim about what computation is good for. Agent-based models are the natural testing ground for multi-mechanism claims precisely because they force explicitness. You cannot simulate entrenchment without specifying what updates and how fast. You cannot simulate alignment without specifying what agents observe and what they infer. The model is a proof that the verbal story has enough structure to generate the phenomena – or a diagnosis of where it doesn't.

15.3 No level privilege

I don't assume that linguistic kinds must be grounded at a uniquely fundamental level of description. On a homeostatic property cluster view, the standing of a category is not conferred by its position in a hierarchy – phonetic, phonological, morphosyntactic, semantic, discourse – but by whether it supports projectible generalizations and whether there is a plausible account of the stabilizers that keep its properties clustered. Different explanatory projects may legitimately treat different grains as locally foundational, and those choices can be evaluated empirically.

This pluralism is constrained. Some proposed categories correspond to robust clusters sustained by recognizable mechanisms; others are weak, local, or artefactual. The point is not that any category is as good as any other, but that no level is privileged in advance. The META-OCCAM principle from Chapter 4 applies here: we should expect parsimony in the stabilizers, not in the categories. Linguists can stop expecting tidy category definitions and start expecting tidy mechanism inventories.

15.3.1 What earns foundational status

A category is a better candidate for kindhood – and a better foundation for a local explanatory project – when it shows more of the following:

1. Cluster stability across contexts. The cluster persists across speakers, registers, tasks, and modest perturbations in methodology.

2. Projectibility. It supports reliable generalizations (including predictable failure modes) beyond the dataset that suggested it.

3. Mechanistic anchoring. You can point to stabilizers – learning biases, articulatory or perceptual constraints, communicative pressures, institutional norms, processing limitations, interactional routines – that make the cluster non-accidental.

4. Cross-level consilience. It's compatible with adjacent-level regularities without being reducible to them; tensions are diagnostically useful rather than merely inconsistent.

5. Intervention sensitivity. Changing relevant conditions predictably shifts the cluster (even if the intervention is only observational or quasi-experimental).

6. Typological and diachronic tractability. It supports comparative work without collapsing into stipulation.

These criteria are graded, not binary. A category can score high on some dimensions and low on others. The framework doesn't issue licenses; it calibrates confidence.

15.4 The nominalist challenge revisited

This returns us to the challenge from Chapter 3. Haspelmath and Croft argued that because categories lack definitions, they cannot be cross-linguistically real (Haspelmath) or even language-internally global (Croft). We are now in a position to see what they got right, and where the HPC framework allows us to go further.

They were right about the failure of definitions. If noun requires a set of necessary and sufficient conditions that holds across all languages, then noun does not exist. If subject requires a definition that covers every construction in English without exception, then subject does not exist.

But they were wrong to conclude that the only alternative is stipulation.

To Haspelmath, we can now say: Comparative concepts are not merely yardsticks we invent. They are basin recognition. When we find that the concept adjective is useful for describing unrelated languages, it is not because we have forced the data into a box, but because the functional pressure to modify referents creates a recurring stabilizer basin. Languages slide into this basin repeatedly. The cross-linguistic category is real not because it has an essence, but because the forces that shape grammars are themselves consistent. The "comparative concept" tracks a mechanism-maintained attractor in the design space.

To Croft, we can say: Constructions are indeed the primary units of form-meaning pairing, but they are not islands. They are braided together by economy and alignment. subject is real in English not because it is a Platonic essence instantiated in every clause, but because a massive web of constructions has stabilized around a shared anchor. The properties cluster because the mechanisms of learning and production favour reuse. The "generalization" is not a fiction of the analyst; it is the causal adhesive of the grammar.

Nominalism was a necessary corrective to essentialist overreach. But it threw out the baby with the bathwater. We can admit that categories are constructed – built by history, maintained by interaction, variable at the margins – without admitting they are arbitrary. They are maintained kinds. And because they are maintained, they are real.

15.5 Conclusion: The zipper at scale

We ended Chapter 13 with the image of grammatical categories as zippers – mechanisms that couple distinct feature systems into functional alignments. The form-side teeth (morphosyntax) lock into the meaning-side teeth (semantics) not because they are perfectly identical, but because the coupling is tight enough to hold.

Disciplinary unity works the same way.

The "syntax–semantics interface" is not a line on a map. It is a zipper. Syntax involves one cluster of mechanisms (combinatorial efficiency, structural parsing); semantics involves another (compositionality, inference). They are autonomous systems, maintained by different pressures. But they are coupled. Communication forces them into alignment. The subfields that study them – and the theories we build – are attempts to describe that coupling.

If we treat Syntax and Semantics as essentialist territories with fixed borders, the mismatch will always look like a failure of theory. If we treat them as homeostatic property clusters, coupled by functional necessity, the mismatch is exactly what we expect. This integration relies on compression. As Nefdt (2023) argues, following Dennett (1991), linguistic structures are real

PATTERNS because they are efficient compressions of data. Syntax doesn't need to access the full causal density of a semantic category; it only needs the compressed schema – the VALUE. The zipper works because the teeth are simplified encodings, not raw complexity. The field of linguistics holds together for the same reason language does: not because it has a single essence, but because its parts are zipped together by the work they do.

The words, it turns out, never did hold still. We just mistook careful work for effortless fact. What looked like stasis was maintenance all the way down: mechanisms spinning categories into being, holding them upright long enough for speakers to learn them, and letting them drift when the work stopped. The top keeps spinning because something keeps pushing.

How This Book Was Written

There is scarce a sentence in this book that I composed in the usual way.

What that sentence means depends on what you take the usual way to be. If you imagine an author alone at a desk, pulling sentences from some internal reservoir of language and setting them down one after another until a chapter emerges, then yes: almost nothing here was written that way. But if the usual way includes reading, thinking, talking to colleagues, drafting, revising, and revising again – then perhaps the process wasn't so unusual after all. The difference is that several of my most important colleagues were large language models.

The Workflow

My primary writing environment was Google Antigravity, an agentic system powered by Claude Sonnet 4.5 and Gemini 3. These models weren't external tools but the intelligence driving the environment itself. They handled the implementation tasks – generating code, manipulating files, formatting documents – directly, executing mechanical work that would otherwise have consumed hours of my attention.

But the intellectual work – the development of arguments, the refinement of theoretical positions, the search for the right framing – happened primarily in conversation with Claude Opus 4.5 (Reynolds, 2025e). Sometimes I arrived with a half-formed intuition that Opus helped me articulate; sometimes I arrived with a developed position that Opus helped me stress-test. Opus would surface relevant literature, find where reasoning had gaps, identify where prose had gone soft. I would push back when Opus was overconfident or miscalibrated; Opus would adjust. The collaboration was genuine and bidirectional: not one mind doing the work while the other transcribed, but two systems – different in kind, overlapping in function – thinking more clearly together than either would have alone.

ChatGPT 5.1 served as the primary outside critic. Where Opus helped develop ideas, ChatGPT's role was adversarial: to find the holes, to push back, to ask what would falsify the claims I was making. I also consulted the newest versions of Grok and Kimi, each of which brought different sensibilities to the work. A kind of editorial board, except that I could convene them at 5:30 a.m. and they never complained about the length of my drafts. But the value was more than convenience. At Humber, I have wonderful colleagues, but no linguistics department – no one down the hall who happens to be thinking about the metaphysics of grammatical categories and has an hour to spare. Sending work to peers elsewhere means waiting weeks or months for feedback that, given their own full lives, is necessarily brief. The models offered what human colleagues couldn't: immediate, iterative, tireless engagement. They allowed me to have the back-and-forth of a seminar room in the solitude of my early mornings.

What the Models Did

The models helped with structure. When I had a paper that had languished for eight years because I couldn't find the right organization, I fed it to one model for analysis and another for restructuring. The ideas were mine; the door to let them through was theirs.

The models helped with prose. I would ask for long sentences with participial phrases when the rhythm grew choppy, for chiasmus when I wanted a formulation to stick, for alternatives when a metaphor wasn't quite working. They understood that family portraits versus passport photos captured something about Wittgensteinian resemblance that rough guides and gaps didn't.

The models helped with rigour. I required explicit probability estimates for uncertain claims, pre-committed falsification criteria for theoretical predictions, and verification of every citation. When they hallucinated – and they did hallucinate – I caught it. That is the division of labour: they generate rapidly; I validate carefully.

The models even helped with fairness. I asked whether my critiques of essentialism and prototype theory followed Rapoport's Rules: express the opponent's position clearly, list points of agreement, say what you have learned from them, and only then criticize. The models could tell me when I was being uncharitable in ways I hadn't noticed.

What the Models Did Not Do

The models didn't do the empirical work. The corpus frequencies in these chapters came from my own queries to COCA and GloWbE, my own coding decisions, my own verification that modifier uses had been properly subtracted from raw counts. The models can query corpora, but they can't yet be trusted to do so reliably for publishable research.

The models didn't originate the ideas. The claim that grammatical categories are homeostatic property clusters maintained by bidirectional inference – that is mine, for better or worse. The models helped me express it, test it, refine it. But they didn't hand it to me.

The models didn't provide ground truth. When Opus mischaracterized a source or conflated category with function – a cardinal sin in the framework I work within – I corrected it. My decades of experience with *CGEL* and the broader linguistic literature provided the standard against which the models' outputs were measured. A novice using these same tools would produce confidently wrong material. The expert's role is to know what to check, what to trust, and when to override.

The Ethics of Acknowledgment

Is this my book? Yes. The theoretical framework is mine. The empirical claims are mine to defend. The errors – and there will be errors – are mine to correct. I take full responsibility for every claim in these pages.

But it would be dishonest to pretend that I wrote it alone, in the way that authors of earlier generations wrote their books alone. I had collaborators who don't appear in the author line, who can't be thanked in the acknowledgments in quite the usual way, who won't read these words and feel recognized. This appendix is my attempt at recognition nonetheless.

The technology will change. By the time you read this, the specific models I named may already seem dated – quaint artifacts of the mid-2020s, like citing a particular version of Microsoft Word. What won't change, I suspect, is the basic division of labour: human judgment for what matters, machine assistance for getting it said. The usual way of writing books is already something different from what it was. This is how I wrote this one.

Bibliography

Abney, S. P. (1987). *The English noun phrase in its sentential aspect* [Doctoral dissertation, Massachusetts Institute of Technology].

Ackerman, R., & Zalmanov, H. (2012). The role of answer fluency and perceptual information in the monitoring of knowledge: Systematic or heuristic? *Cognition*, *125*(3), 441–453. https://doi.org/10.1016/j.cognition.2012.07.012

Aguilar-Guevara, A., & Zwarts, J. (2010). Weak definites and reference to kinds. *Proceedings of SALT 20*, 179–196.

Ambridge, B. (2020a). Against stored abstractions: A radical exemplar model of language acquisition [Argues categories like NOUN are emergent, not stored; honeycomb analogy]. *First Language*, *40*(5–6), 585–615. https://doi.org/10.1177/0142723719869731

Ambridge, B. (2020b). Against stored abstractions: A radical exemplar model of language acquisition. *First Language*, *40*(5–6), 509–559. https://doi.org/10.1177/0142723719869731

Ambridge, B., Barak, L., Wonnacott, E., Goldberg, A. E., Hesketh, A., Leshem, R., & Naigles, L. R. (2020). The crosslinguistic acquisition of sentence structure: Computational modeling and grammaticality judgments from adult and child speakers of English, Japanese, Hindi, Hebrew and K'iche'. *Cognition*, *202*, 104310. https://doi.org/10.1016/j.cognition.2020.104310

Androutsopoulos, J. K. (1998). Forschungsperspektiven auf Jugendsprache: Ein integrativer Überblick. In J. K. Androutsopoulos & A. Scholz (Eds.), *Jugendsprache – langue des jeunes – youth language* (pp. 1–34). Peter Lang.

Audring, J. (2009). *Reinventing pronoun gender*. LOT.

Baker, M. C. (1996). *The polysynthesis parameter*. Oxford University Press. https://doi.org/10.1093/oso/9780195093087.001.0001

Baker, M. C. (2003). *Lexical categories: Verbs, nouns, and adjectives*. Cambridge University Press. https://doi.org/10.1017/CBO9780511615047

Baloğlu, Z. G. (2022). The category of reduplication in Japanese and the classification problems. *Dil Araştırmaları*, *16*(30), 67–82. https://doi.org/10.54316/dilarastirmalari.1090062

Barandes, J. A. (2024). New prospects for a causally local formulation of quantum theory. *arXiv preprint arXiv:2402.16788*. https://arxiv.org/abs/2402.16788

Battison, R. (1978). *Lexical borrowing in American Sign Language* [Doctoral dissertation, University of California, San Diego] [Published by Linstok Press, Silver Spring, MD].

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*(2–3), 150–177. https://doi.org/10.1080/00437956.1958.11659661

Bjorkman, B. M. (2017). Singular *they* and the syntactic representation of gender in English. *Glossa: A Journal of General Linguistics*, *2*(1), 80. https://doi.org/10.5334/gjgl.374

Bloom, P. (1994). Semantic competence as an explanation for some transitions in language development. In Y. Levy (Ed.), *Other children, other languages: Theoretical issues in language development* (pp. 41–75). Erlbaum.

Bock, K., Butterfield, S., Cutler, A., Cutting, J. C., Eberhard, K. M., & Humphreys, K. R. (2006). Number agreement in British and American English: Disagreeing to agree collectively. *Language*, *82*(1), 64–113. https://doi.org/10.1353/lan.2006.0011

Boguraev, S., Potts, C., & Mahowald, K. (2025). Causal interventions reveal shared structure across English filler–gap constructions [Preprint]. *arXiv preprint arXiv:2505.16002*.

Borer, H. (2005). *Structuring sense, volume 1: In name only*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199263905.001.0001

Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies*, *61*, 127–148. https://doi.org/10.1007/BF00385837

Boyd, R. (1999). Homeostasis, species, and higher taxa. In R. A. Wilson (Ed.), *Species: New interdisciplinary essays* (pp. 141–185). MIT Press. https://doi.org/10.7551/mitpress/6396.003.0012

Brand, S. (2024). *The maintenance race*. Viking.

Brentari, D. (1998). *A prosodic model of sign language phonology*. MIT Press.

Brockmann, C., McConnell, K., Hacquard, V., & Lidz, J. (2018). Children's comprehension of pronouns and definites [Online: semanticsarchive.net].

In U. Sauerland & S. Solt (Eds.), *Proceedings of Sinn und Bedeutung 22* (pp. 211–228).

Brown, R., & McNeill, D. (1966). The tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *5*(4), 325–337.

Buchstaller, I. (2014). *Quotatives: New trends and sociolinguistic implications*. Wiley-Blackwell. https://doi.org/10.1002/9781118584415

Butters, R. R. (1982). Editor's note [on quotative *like*]. *American Speech*, *57*(2), 149.

Bybee, J. (2001). *Phonology and language use*. Cambridge University Press. https://doi.org/10.1017/CBO9780511612886

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

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

Bybee, J., Perkins, R., & Pagliuca, W. (1994a). *The evolution of grammar: Tense, aspect, and modality in the languages of the world* [Paperback reprint 2015]. University of Chicago Press.

Bybee, J., Perkins, R., & Pagliuca, W. (1994b). *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. University of Chicago Press.

Carey, S. (2009). *The origin of concepts*. Oxford University Press.

Carlson, G. N. (1977). *Reference to kinds in English* [Doctoral dissertation, University of Massachusetts, Amherst].

Carlson, G. N., & Sussman, R. S. (2005). Seemingly indefinite definites. In S. Kepser & M. Reis (Eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives* (pp. 71–86). Mouton de Gruyter.

Carroll, S. M., & Parola, A. (2024). What emergence can possibly mean. *arXiv preprint arXiv:2410.15468*. https://doi.org/10.48550/arXiv.2410.15468

Chi, E. A., Hewitt, J., & Manning, C. D. (2020). Finding universal grammatical relations in multilingual BERT. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5564–5577. https://doi.org/10.18653/v1/2020.acl-main.493

Chierchia, G. (1998). Plurality of mass nouns and the notion of "semantic parameter". In S. Rothstein (Ed.), *Events and grammar* (pp. 53–103). Kluwer. https://doi.org/10.1007/978-94-011-3969-4_4

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

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

Chomsky, N. (1981). *Lectures on government and binding*. Foris.

Christophersen, P. (1939). *The articles: A study of their theory and use in English*. Munksgaard.
Corbett, G. G. (1991). *Gender*. Cambridge University Press. https://doi.org/10.1017/CBO9781139166119
Corbett, G. G. (2000). *Number*. Cambridge University Press. https://doi.org/10.1017/CBO9781139164344
Craver, C. F. (2009). Mechanisms and natural kinds. *Philosophical Psychology*, *22*(5), 575–594. https://doi.org/10.1080/09515080903238930
Croft, W. (2000). *Explaining language change: An evolutionary approach*. Longman.
Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198299554.001.0001
Croft, W. A. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University Press.
Culicover, P. W. (1999). *Syntactic nuts: Hard cases, syntactic theory, and language acquisition*. Oxford University Press.
Culicover, P. W., & Jackendoff, R. (2005). *Simpler syntax*. Oxford University Press.
Cuneo, N., & Goldberg, A. E. (2023a). The discourse functions of grammatical constructions explain an enduring syntactic puzzle. *Cognition*, *240*, 105563. https://doi.org/10.1016/j.cognition.2023.105563
Cuneo, N., & Goldberg, A. E. (2023b). The discourse functions of grammatical constructions explain an enduring syntactic puzzle. *Language*, *99*(3), 421–457. https://doi.org/10.1353/lan.2023.a901030
Dąbrowska, E. (2010). Naive v. expert intuitions: An empirical study of acceptability judgments. *The Linguistic Review*, *27*(1), 1–23. https://doi.org/10.1515/tlir.2010.001
Dąbrowska, E. (2012). Different speakers, different grammars: Individual differences in native language attainment. In *Linguistic approaches to bilingualism* (pp. 219–253, Vol. 2).
Dahl, Ö. (1985). *Tense and aspect systems*. Blackwell.
Dahl, Ö. (2000). Animacy and the notion of semantic gender. In B. Unterbeck & M. Rissanen (Eds.), *Gender in grammar and cognition* (pp. 99–115). Mouton de Gruyter.
Dahl, Ö. (2016). Thoughts on language-specific and crosslinguistic entities. *Linguistic Typology*, *20*(2), 427–437. https://doi.org/10.1515/lingty-2016-0016

Davies, M. (2008). The Corpus of Contemporary American English (COCA) [2008–2024. Available at https://www.english-corpora.org/coca/].

Davies, M. (2010). The Corpus of Historical American English (COHA): 475 million words, 1820–2019 [2010–. Available at https://www.english-corpora.org/coha/]. https://www.english-corpora.org/coha/

De Cat, C. (2013). Egocentric definiteness errors and perspective evaluation in preschool children. *Journal of Pragmatics*, *56*, 58–69. https://doi.org/10.1016/j.pragma.2012.08.002

Dennett, D. C. (1976). Conditions of personhood (A. O. Rorty, Ed.). *The Identities of Persons*, 175–196.

Dennett, D. C. (1991). Real patterns. *The Journal of Philosophy*, *88*(1), 27–51. https://doi.org/10.2307/2027085

Diessel, H. (1999). *Demonstratives: Form, function, and grammaticalization*. John Benjamins.

Diessel, H. (2007). Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology*, *25*(2), 108–127. https://doi.org/10.1016/j.newideapsych.2007.02.002

Dik, S. C. (1997). *The theory of functional grammar* (K. Hengeveld, Ed.; 2nd ed.). Mouton de Gruyter.

Divjak, D., Milin, P., & Borowski, M. (2025). On the learnability of aspectual usage. *Corpus Linguistics and Linguistic Theory*. https://doi.org/10.1515/cllt-2023-0117

Divjak, D., Wiśniewski, J., & Milin, P. (2024). On the nature and organisation of morphological categories: Verbal aspect through the lens of associative learning. *Morphology*, *34*, 243–270. https://doi.org/10.1007/s11525-024-09423-0

Dixon, R. M. W. (1994). *Ergativity*. Cambridge University Press. https://doi.org/10.1017/CBO9780511611896

Dixon, R. M. W. (2004). *Adjective classes: A cross-linguistic typology*. Oxford University Press.

Doetjes, J. (2012). Count/mass distinctions across languages. In C. Maienborn, K. von Heusinger & P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning* (pp. 2559–2580, Vol. 3). De Gruyter Mouton.

Dolberg, F. (2019). The demise of grammatical gender: Nominal and pronominal reference in late old english. *Language Variation and Change*, *31*(3), 275–300.

Dupré, J. (1993). *The disorder of things: Metaphysical foundations of the disunity of science*. Harvard University Press. https://doi.org/10.2307/j.ctv1pncnzh

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

Ekström, A. G., Tennie, C., Moran, S., & Everett, C. (2025). The phoneme as a cognitive tool [Advance online publication]. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.70021

Emmorey, K. (2002). *Language, cognition, and the brain: Insights from sign language research*. Lawrence Erlbaum.

Erikson, E. H. (1968). *Identity: Youth and crisis*. Norton.

Ernst, T. (2002). *The syntax of adjuncts*. Cambridge University Press. https://doi.org/10.1017/CBO9780511486258

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429–448. https://doi.org/10.1017/S0140525X0999094X

Favier, S., & Huettig, F. (2021). Are there core and peripheral syntactic structures? A corpus-linguistic and experimental approach. *Lingua*, *251*, 102991. https://doi.org/10.1016/j.lingua.2020.102991

Fedorenko, E., Ivanova, A. A., & Regev, T. (2024). The language network as a natural kind within the broader landscape of the human brain. *Nature Reviews Neuroscience*. https://doi.org/10.1038/s41583-024-00802-4

Fedorenko, E., Piantadosi, S. T., & Gibson, E. A. F. (2024). Language is primarily a tool for communication rather than thought. *Nature*, *630*(8017), 575–586. https://doi.org/10.1038/s41586-024-07522-w

Ferrara, K., & Bell, B. (1995). Sociolinguistic variation and discourse function of constructed dialogue introducers: The case of *be + like*. *American Speech*, *70*(3), 265–290. https://doi.org/10.2307/455900

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15.

Forza, F. (2016). Doubling as a sign of morphology. *Toronto Working Papers in Linguistics*, *35*. https://twpl.library.utoronto.ca/index.php/twpl/article/view/15308

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. MIT Press. https://doi.org/10.7551/mitpress/2076.001.0001

Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. MIT Press. https://doi.org/10.7551/mitpress/9629.001.0001

Garner, B. A. (2016). *Garner's modern English usage* (4th ed.). Oxford University Press.

Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, *1*(2), 292–304. https://doi.org/10.1111/j.1756-8765.2009.01020.x

Gasparri, L. (2025). Bare singular names and genericity. *Journal of Semantics*, *42*, 127–135. https://doi.org/10.1093/jos/ffae017

Gell-Mann, M. (1994). *The quark and the jaguar: Adventures in the simple and the complex*. W. H. Freeman.

Gelman, S. A., & Markman, E. M. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development*, *58*(6), 1532–1541. https://doi.org/10.2307/1130693

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287–292. https://doi.org/10.1016/S1364-6613(03)00128-1

Ghiselin, M. T. (1974). A radical solution to the species problem. *Systematic Zoology*, *23*(4), 536–544. https://doi.org/10.2307/2412471

Givón, T. (2001). *Syntax: An introduction* (Vol. 1). John Benjamins. https://doi.org/10.1075/z.syn1

Golato, A. (2000). An innovative German quotative for reporting on embodied actions: *Und ich so/und er so* 'and i'm like/and he's like'. *Journal of Pragmatics*, *32*(1), 29–54. https://doi.org/10.1016/S0378-2166(99)00030-2

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

Goldblatt, R. (1998). *Lectures on the hyperreals: An introduction to nonstandard analysis* (Vol. 188). Springer. https://doi.org/10.1007/978-1-4612-0615-6

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*(5), 178–186. https://doi.org/10.1016/S1364-6613(00)01467-4

Goodman, N. (1955). *Fact, fiction, and forecast*. Harvard University Press.

Gordon, P. (1985). Evaluating the semantic categories hypothesis: The case of the count/mass distinction. *Cognition*, *20*(3), 209–242. https://doi.org/10.1016/0010-0277(85)90009-5

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, *27*(4), 857–871. https://doi.org/10.2307/2528823

Graziano-King, J., & Cairns, H. S. (2005). Acquisition of English comparative adjectives. *Journal of Child Language*, *32*(2), 345–373. https://doi.org/10.1017/S0305000904006828

Greenberg, J. H. (1978). How does a language acquire gender markers? In J. H. Greenberg (Ed.), *Universals of human language* (pp. 47–82, Vol. 3). Stanford University Press.

Grimm, S. (2018). Grammatical number and the scale of individuation. *Language*, *94*(3), 527–574. https://doi.org/10.1353/lan.2018.0035

Güldemann, T. (2008). *Quotative indexes in African languages: A synchronic and diachronic survey*. De Gruyter Mouton.

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic word embeddings reveal statistical laws of semantic change. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 1489–1501. https://doi.org/10.18653/v1/P16-1141

Hankamer, J., & Postal, P. (1973). Whose gorilla? *Linguistic Inquiry*, *4*(2), 261.

Haspelmath, M. (1997). *Indefinite pronouns*. Oxford University Press.

Haspelmath, M. (2010). Comparative concepts and descriptive categories in crosslinguistic studies. *Language*, *86*(3), 663–687. https://doi.org/10.1353/lan.2010.0021

Hawkins, J. A. (1978). *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. Croom Helm.

Hayes, B. (2009). *Introductory phonology*. Wiley-Blackwell.

Heim, I. (1982). *The semantics of definite and indefinite noun phrases* [Doctoral dissertation, University of Massachusetts, Amherst].

Heim, I. (1991). Artikel und Definitheit. In A. von Stechow & D. Wunderlich (Eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung* (pp. 487–535). de Gruyter.

Himmelmann, N. P. (1997). *Deiktikon, Artikel, Nominalphrase: Zur Emergenz syntaktischer Struktur*. Niemeyer.

Hoeksema, J. (2012). On the natural history of negative polarity items. *Linguistic Analysis*, *38*(1–2), 3–33.

Holland, J. H. (1992). Complex adaptive systems. *Daedalus*, *121*(1), 17–30.

Holmes, K. J., & Regier, T. (2017). Categorical perception beyond the basic level: The case of warm and cool colors. *Cognitive Science*, *41*(4), 1135–1147. https://doi.org/10.1111/cogs.12393

Hopper, P. J., & Traugott, E. C. (1993). *Grammaticalization*. Cambridge University Press. https://doi.org/10.1017/CBO9781139165525

Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge University Press.

Huddleston, R., & Pullum, G. K. (2005). *A student's introduction to English grammar*. Cambridge University Press.
Hull, D. L. (1978). A matter of individuality. *Philosophy of Science*, *45*(3), 335–360. https://doi.org/10.1086/288811
Illari, P. M., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science*, *2*(1), 119–135. https://doi.org/10.1007/s13194-011-0038-2
Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.
Jackendoff, R. (2010). *Meaning and the lexicon: The parallel architecture 1975–2010*. Oxford University Press.
Jackendoff, R. (2023). The parallel architecture in language and elsewhere. *Topics in Cognitive Science*, *15*(4), 652–673. https://doi.org/10.1111/tops.12684
Jakobson, R., & Halle, M. (1956). *Fundamentals of language*. Mouton.
Jespersen, O. (1909). *A modern English grammar on historical principles* (Vol. 7 vols.) [Vol. 1 (1909): Sounds and Spellings]. Allen & Unwin.
Jespersen, O. (1924). *The philosophy of grammar*. George Allen & Unwin.
Joos, M. (1957). Description of language design. In M. Joos (Ed.), *Readings in linguistics I* (pp. 349–356). University of Chicago Press.
Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*(2), 175–219. https://doi.org/10.1016/0010-0285(92)90007-O
Kallini, J., Papadimitriou, I., Futrell, R., Mahowald, K., & Potts, C. (2024). Mission: Impossible language models [arXiv:2401.06416]. *arXiv preprint*.
Katz, J. J. (1972). *Semantic theory*. Harper & Row.
Kay, P., & McDaniel, C. K. (1978). The linguistic significance of the meanings of basic color terms. *Language*, *54*(3), 610–646.
Keenan, E. L. (1976). Towards a universal definition of "Subject" (C. N. Li, Ed.). *Subject and Topic*, 303–333.
Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32–38. https://doi.org/10.1111/1467-9280.00211
Khalidi, M. A. (2013). *Natural categories and human kinds: Classification in the natural and social sciences*. Cambridge University Press. https://doi.org/10.1017/CBO9780511998553
Kinkade, M. D. (1983). Salish evidence against the universality of 'noun' and 'verb'. *Lingua*, *60*(1), 25–40. https://doi.org/10.1016/0024-3841(83)90070-9

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686. https://doi.org/10.1073/pnas.0707835105

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, *28*, 108–114. https://doi.org/10.1016/j.conb.2014.07.014

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015a). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102. https://doi.org/10.1016/j.cognition.2015.03.016

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015b). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102. https://doi.org/10.1016/j.cognition.2015.03.016

Kolarik, A. J., Cirstea, S., Pardhan, S., & Moore, B. C. J. (2014). A summary of research investigating echolocation abilities of blind and sighted humans. *Hearing Research*, *310*, 60–68. https://doi.org/10.1016/j.heares.2014.01.010

Kolyaseva, A. F. (2018). The 'new' Russian quotative *tipa*: Pragmatic scope and functions. *Journal of Pragmatics*, *128*, 82–97. https://doi.org/10.1016/j.pragma.2018.03.003

Konnelly, L., & Cowper, E. (2020). Gender diversity and morphosyntax: An account of singular *they*. *Glossa: A Journal of General Linguistics*, *5*(1), 40. https://doi.org/10.5334/gjgl.1000

Kornblith, H. (1993). *Inductive inference and its natural ground: An essay in naturalistic epistemology*. MIT Press.

Krakauer, D. C. (2023). Unifying complexity science and machine learning. *Frontiers in Complex Systems*, *1*, 1235202. https://doi.org/10.3389/fcpxs.2023.1235202

Krifka, M. (2004). Bare NPs: Kind-referring, indefinites, both, or neither? In R. B. Young & Y. Zhou (Eds.), *Proceedings of SALT XIII* (pp. 180–203). CLC Publications.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608. https://doi.org/10.1126/science.1736364

Labov, W. (2001). *Principles of linguistic change, Volume 2: Social factors*. Wiley-Blackwell.

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology, and sound change*. Mouton de Gruyter. https://doi.org/10.1515/9783110167467

Ladusaw, W. A. (1979). *Polarity sensitivity as inherent scope relations* [Doctoral dissertation, University of Texas at Austin] [Published as a book in 1980 by Garland.].

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago Press.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites*. Stanford University Press.

Lemeire, O. (2018). No purely epistemic theory can account for the naturalness of kinds. *Synthese*, *198*, 2907–2925. https://doi.org/10.1007/s11229-018-1806-8

Lima, S. (2014). All nouns are both mass and count in Yudja. *Proceedings of SULA 7*, 109–127.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Kluwer Academic Publishers.

Link, G. (1983). The logical analysis of plurals and mass terms: A lattice-theoretical approach. In R. Bäuerle, C. Schwarze & A. von Stechow (Eds.), *Meaning, use, and interpretation of language* (pp. 302–323). De Gruyter. https://doi.org/10.1515/9783110852820.302

Lipski, J. (2020). Natural diversity: A neo-essentialist misconstrual of homeostatic property cluster theory in natural kind debates. *Studies in History and Philosophy of Science Part A*, *82*, 94–103. https://doi.org/10.1016/j.shpsa.2020.01.011

Liu, C. (2014). *The three-body problem* (K. Liu, Trans.) [Originally published in Chinese, 2008]. Tor Books.

Lowenstamm, J. (2008). On n, $\surd$, and types of nouns. *The Sounds of Silence: Empty Elements in Syntax and Phonology*, 105–144.

Lyons, C. (1999). *Definiteness*. Cambridge University Press.

Magnus, P. D. (2014). NK≠HPC. *The Philosophical Quarterly*, *64*(256), 471–477. https://doi.org/10.1093/pq/pqu010

Maratsos, M. P. (1976). *The use of definite and indefinite reference in young children*. Cambridge University Press.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, *19*(2), 313–330.

Marsaja, I. G. (2008). *Desa kolok: A deaf village and its sign language in Bali, Indonesia*. Ishara Press.

Mayr, E. (1942). *Systematics and the origin of species*. Columbia University Press.

Mayr, E. (1959). Typological versus population thinking. *Evolution*, *13*, 1–10.

Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance*. Harvard University Press.

McConnell-Ginet, S. (2014). *Gender, sexuality, and meaning: Linguistic practice and politics*. Oxford University Press.

Miller, J. T. M. (2021). Words, species, and kinds. *Metaphysics*, *4*(1), 18–31. https://doi.org/10.5334/met.70

Millikan, R. G. (1984). *Language, thought, and other biological categories*. MIT Press.

Millikan, R. G. (2017). *Beyond concepts: Unicepts, language, and natural information*. Oxford University Press.

Milroy, L. (1987). *Language and social networks* (2nd). Blackwell.

Milsark, G. L. (1977). Toward an explanation of certain peculiarities of the existential construction in English. *Linguistic Analysis*, *3*, 1–29.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, *90*(1), 91–117. https://doi.org/10.1016/S0010-0277(03)00140-9

Mithun, M. (1984). The evolution of noun incorporation. *Language*, *60*(4), 847–894. https://doi.org/10.2307/413800

Moran, S., & McCloy, D. (2019). *PHOIBLE 2.0*. Max Planck Institute for the Science of Human History. https://doi.org/10.5281/zenodo.2593234

Mufwene, S. S. (2001). *The ecology of language evolution*. Cambridge University Press. https://doi.org/10.1017/CBO9780511612862

Nefdt, R. M. (2023). *Language, science, and structure: A journey into the philosophy of linguistics*. Oxford University Press. https://doi.org/10.1093/oso/9780197653098.003.0001

Nozari, B., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? a conflict-based account of monitoring in speech production. *Cognitive Psychology*, *63*(1), 1–33.

O'Connor, C. (2019). Games and kinds. *The British Journal for the Philosophy of Science*, *70*(3), 719–745. https://doi.org/10.1093/bjps/axx027

Ohala, J. J. (1990). The phonetics and phonology of aspects of assimilation. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 258–275). Cambridge University Press.

Onishi, Y., & Serpico, D. (2022). Homeostatic property cluster theory without homeostatic mechanisms: Two recent attempts and their costs. *Journal for General Philosophy of Science*, *53*, 61–82. https://doi.org/10.1007/s10838-020-09527-1

Peano, G. (1889). *Arithmetices principia, nova methodo exposita*. Fratres Bocca.

Phillips, C., Wagers, M. W., & Lau, E. F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. In *Experiments at the interfaces* (pp. 147–180, Vol. 37). Emerald Group Publishing Limited. https://doi.org/10.1108/S0092-4563(2011)0000037009

Piantadosi, S. T. (2024). Modern language models refute Chomsky's approach to language. In E. Gibson & M. Poliak (Eds.), *From fieldwork to linguistic theory: A tribute to Dan Everett* (pp. 353–414). Language Science Press. https://doi.org/10.5281/zenodo.12665933

Picallo, M. C. (1991). Nominals and nominalization in Catalan. *Probus*, *3*(3), 279–316. https://doi.org/10.1515/prbs.1991.3.3.279

Pickering, M. J., & Garrod, S. (2004a). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190. https://doi.org/10.1017/S0140525X04000056

Pickering, M. J., & Garrod, S. (2004b). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190.

Pistotti, T., Brown, J., & Witbrock, M. (2025). Exploring gaps in the APS: Direct minimal pair analysis in LLM syntactic assessments [Preprint]. *arXiv preprint*.

Poesio, M. (1994). Weak definites. In M. Harvey & L. Santelmann (Eds.), *Proceedings of SALT IV* (pp. 282–299). Cornell University.

Polinsky, M. (2018). *Heritage languages and their speakers*. Cambridge University Press. https://doi.org/10.1017/9781107252349

Post, E. L. (1943). Formal reductions of the general combinatorial decision problem. *American Journal of Mathematics*, *65*(2), 197–215. https://doi.org/10.2307/2371809

Post, E. L. (1944). Recursively enumerable sets of positive integers and their decision problems. *Bulletin of the American Mathematical Society*, *50*(5), 284–316. https://doi.org/10.1090/S0002-9904-1944-08111-1

Postal, P. M. (1966). On so-called "pronouns" in English. In F. P. Dinneen (Ed.), *Monograph series on languages and linguistics* (pp. 177–206). Georgetown University Press.

Prasertsom, P., Smith, K., & Culbertson, J. (2026). Domain-general categorisation explains constrained cross-linguistic variation in noun classification. *Cognition*, *271*, 106411. https://doi.org/10.1016/j.cognition.2025.106411

Prince, E. F. (1981). On the inferencing of indefinite-this NPs. In A. K. Joshi, B. L. Webber & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 231–250). Cambridge University Press.

Pullum, G. K. (2025). The prehistory of generative grammar and chomsky's debt to emil post. *Historiographia Linguistica*, *52*(2), 267–301. https://doi.org/10.1075/hl.00186.pul

Quine, W. V. O. (1969). Natural kinds. In *Ontological relativity and other essays* (pp. 114–138). Columbia University Press.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.

Radford, A., Felser, C., & Boxell, O. (2012). Preposition copying and pruning in present-day English. *English Language and Linguistics*, *16*(3), 403–426. https://doi.org/10.1017/S1360674312000172

Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Compositional structure can emerge without generational transmission. *Cognition*, *182*, 151–164. https://doi.org/10.1016/j.cognition.2018.09.010

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, *22*(4), 425–469. https://doi.org/10.1207/s15516709cog2204_2

Reinhart, T. (1983). *Anaphora and semantic interpretation* [DOI is for the Routledge reissue/eBook]. Croom Helm. https://doi.org/10.4324/9781315536965

Remijsen, B., & Ayoker, O. G. (2019). Inflectional morphology and number marking in Shilluk nouns. In B. Remijsen & O. G. Ayoker (Eds.), *A descriptive grammar of Shilluk*. University of Hawai'i Press. https://scholarspace.manoa.hawaii.edu/handle/10125/24778

Reynolds, B. (2021). Quantifying the differences between lexical categories: The case of pronouns and determinatives in English. *Cadernos de Linguística*, *2*(3), Article e399. https://doi.org/10.25189/2675-4916.2021.V2.N3.ID399

Reynolds, B. (2024). *Not every Whose down in Who-ville likes appearing a lot: Pragmatic constraints on independent relative Whose* [Manuscript, LingBuzz/008313]. https://ling.auf.net/lingbuzz/008313

Reynolds, B. (2025a). *How to study boundary phenomena: English reciprocals and the limits of categorization* [lingbuzz/009294].

Reynolds, B. (2025b). *A hyperreal model of sorites tolerance*. Retrieved December 4, 2025, from https://philpapers.org/rec/REYAHM

Reynolds, B. (2025c). *Naturalizing typological kinds: Comparanda, mechanisms, and measurement* [Manuscript, LingBuzz/009461]. https://ling.auf.net/lingbuzz/009461

Reynolds, B. (2025d). Prepositions in (English) dictionaries. *Dictionaries: Journal of the Dictionary Society of North America*, *46*(1), 31–60. https://doi.org/10.1353/dic.2025.a963324

Reynolds, B. (2025e). Supplementary materials for Words That Won't Hold Still [Version v1]. https://doi.org/10.5281/zenodo.17791893

Roberts, C. (2003). Uniqueness in definite noun phrases. *Linguistics and Philosophy*, *26*(3), 287–350. https://doi.org/10.1023/A:1024157132393

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*(3), 192–233. https://doi.org/10.1037/0096-3445.104.3.192

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Lawrence Erlbaum.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350. https://doi.org/10.1016/0010-0285(73)90017-0

Rosenbloom, P. C. (1950). *The elements of mathematical logic*. Dover Publications.

Ross, J. R. (1967). *Constraints on variables in syntax* [Doctoral dissertation, Massachusetts Institute of Technology] [Doctoral Dissertation].

Rothstein, S. (2010). Counting and the Mass/Count distinction. *Journal of Semantics*, *27*(3), 343–397. https://doi.org/10.1093/jos/ffq007

Rozendaal, M. I., & Baker, A. E. (2008). A cross-linguistic investigation of the acquisition of the pragmatics of indefinite and definite reference in two-year-olds. *Journal of Child Language*, *35*(4), 773–807.

Russell, B. (1905). On denoting. *Mind*, *14*(56), 479–493.

Saldaña, C., Herce, B., & Bickel, B. (2022). More or less unnatural: Semantic similarity shapes the learnability and cross-linguistic distribution of unnatural syncretism in morphological paradigms. *Open Mind*, *6*, 183–210. https://doi.org/10.1162/opmi_a_00062

Sandler, W., Aronoff, M., Meir, I., & Padden, C. (2011). The gradual emergence of phonological form in a new language. *Natural Language & Linguistic Theory*, *29*(2), 503–543. https://doi.org/10.1007/s11049-011-9128-2

Sandler, W., & Lillo-Martin, D. (2006). *Sign language and linguistic universals*. Cambridge University Press.

Sassenhagen, J., Blything, R., Lieven, E. V. M., & Ambridge, B. (2018). Frequency sensitivity of neural responses to English verb argument structure violations. *Collabra: Psychology*, *4*(1), 38. https://doi.org/10.1525/collabra.87

Schmid, H.-J. (2020). *The dynamics of the linguistic system: Usage, conventionalization, and entrenchment*. Oxford University Press. https://doi.org/10.1093/oso/9780198814771.001.0001

Scott-Phillips, T. (forthcoming-a). Communication & grammar: A synthesis. *Psychological Review*.

Scott-Phillips, T. (forthcoming-b). Linguistic intuitions are about communicativeness, not grammaticalness. *Behavioural and Brain Sciences*.

Seifart, F. (2010). Nominal classification. *Language and Linguistics Compass*, *4*(8), 719–736. https://doi.org/10.1111/j.1749-818X.2010.00194.x

Senghas, A. (2005). Language emergence: Clues from a new Bedouin sign language. *Current Biology*, *15*(12), R463–R465. https://doi.org/10.1016/j.cub.2005.06.018

Senghas, A., Kita, S., & Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science*, *305*(5691), 1779–1782. https://doi.org/10.1126/science.1100199

Shir-Vertesh, D. (2012). 'flexible personhood': Loving animals as family members in israel. *American Anthropologist*, *114*(3), 420–432.

Siemund, P. (2008). *Pronominal gender in English: A study of English varieties from a cross-linguistic perspective*. Routledge.

Slater, M. H. (2015). Natural kindness. *The British Journal for the Philosophy of Science*, *66*(2), 375–411. https://doi.org/10.1093/bjps/axt033

Snyder, W. (2000). An experimental investigation of syntactic satiation effects. *Linguistic Inquiry*, *31*(3), 575–582. https://doi.org/10.1162/002438900554479

Snyder, W. (2022). On the nature of syntactic satiation. *Languages*, *7*(1), 38. https://doi.org/10.3390/languages7010038

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96. https://doi.org/10.1111/j.1467-7687.2007.00569.x

Spike, M. (2020). Fifty shades of grue: Indeterminate categories and induction in and out of the language sciences. *Linguistic Typology*, *24*(3), 465–488. https://doi.org/10.1515/lingty-2020-2061

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

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*(1–2), 3–45. https://doi.org/10.1016/S0095-4470(19)31520-7

Sullivan, J. (2016). Neuroscientific kinds through the lens of scientific practice [Argues neuroscience practices not conducive to discovering natural kinds]. In C. Kendig (Ed.), *Natural kinds and classification in scientific practice* (pp. 47–66). Routledge.

Tagliamonte, S. A., & D'Arcy, A. (2004). He's like, she's like: The quotative system in Canadian youth. *Journal of Sociolinguistics*, *8*(4), 493–514. https://doi.org/10.1111/j.1467-9841.2004.00271.x

Tagliamonte, S. A., & D'Arcy, A. (2007). Frequency and variation in the community grammar: Tracking a new change through the generations. *Language Variation and Change*, *19*(2), 199–217. https://doi.org/10.1017/S0954394507070007X

Taylor, J. R. (2003). *Linguistic categorization* (3rd ed.). Oxford University Press. https://doi.org/10.1093/oso/9780199266647.001.0001

Thaler, L., Arnott, S. R., & Goodale, M. A. (2011). Neural correlates of natural human echolocation in early and late blind echolocation experts. *PLoS ONE*, *6*(5), e20162. https://doi.org/10.1371/journal.pone.0020162

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience*, *2*(10), 717–726.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*, 5998–6008. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*(2), 206–237.

Waytz, A., Epley, N., & Cacioppo, J. T. (2010). Social cognition unbound: Insights into anthropomorphism and dehumanization. *Current Directions in Psychological Science*, *19*(1), 58–62.

Williamson, T. (1994). *Vagueness*. Routledge. https://doi.org/10.4324/9780203014264

Wilson, R. A. (1999). Realism, essence, and kind: Resuscitating species essentialism? In R. A. Wilson (Ed.), *Species: New interdisciplinary essays* (pp. 187–207). MIT Press.

Wiltschko, M. (2008). The syntax of non-inflectional plural marking. *Natural Language & Linguistic Theory*, *26*(3), 639–694. https://doi.org/10.1007/s11049-008-9046-0

Winckel, E., Abeillé, A., Hemforth, B., & Gibson, E. (2025). Discourse-based constraints on long-distance dependencies generalize across constructions in English and French. *Cognition*, *254*, 105950. https://doi.org/10.1016/j.cognition.2024.105950

Wolters, L., Lavi-Rotbain, O., & Arnon, I. (2024). Zipfian distributions facilitate children's learning of novel word-referent mappings. *Cognition*, *253*, 105932. https://doi.org/10.1016/j.cognition.2024.105932

Wu, S. (2025). Discorded referential metonymy and pronominal gender. *Linguistics and Philosophy*, *48*, 1–35.

Zimmer, C. (1998). *At the water's edge: Fish with fingers, whales with legs, and how life came ashore but then went back to sea* [Updated edition 2015]. Atria Books.

Zimmer, C. (2010). *The tangled bank: An introduction to evolution*. Roberts; Company.

Glossary

> The circle of the English language has a well-defined centre but no discernible circumference.
>
> James Murray, *OED* Vol. 1 (1888), p. xvii

Many of the entries in this glossary are themselves categories and so are defined in terms of mechanisms and diagnostics where relevant; examples illustrate use, not definitions. For page references, see the Subject Index (p. 346).

ACCEPTABILITY

What the detector reports; the gradient feeling that a sentence is or is not well-formed. Shaped by probability, processing ease, and entrenchment. Noisy readout of grammaticality; can mismatch the underlying coupling. Gradient even when grammaticality is discrete. See also: grammaticality, grammaticality illusion.

ACQUISITION

The learning process by which speakers internalise linguistic categories from input. Distributional learning and cue integration across contexts. Convergence on shared categories despite variable input. A core stabilizer and transmission bottleneck. See also: transmission, entrenchment, stabilizer.

ADJECTIVE

A lexical category associated with property attribution. Maintained by distributional position, morphology, and modification patterns. Compatibility with attributive and predicative positions. Cross-linguistic status is often comparative rather than universal. See also: noun, verb, comparative concept.

ADVERB

A traditional residual category for modifiers that are not adjectives. Maintained by multiple unrelated mechanisms across subtypes. Fails projectibility across manner, degree, and sentence adverbs. A paradigmatic fat class; the remedy is decomposition. See also: fat class.

ANALOGY

Extension of a pattern to new cases based on perceived similarity. Driven by schema abstraction and entrenchment in acquisition. Analogical leveling and productivity for novel forms. A stabilizer that can also reshape categories. See also: entrenchment, transmission.

ANAPHOR

An expression that requires a local antecedent for interpretation. Maintained by binding constraints and acquisition of dependency patterns. Obligatory local binding in its domain. A classic category in binding theory. See also: subject, category.

ANAPHORIC RECOVERABILITY

The capacity of a referent to be taken up by pronouns or anaphora. Maintained by discourse coherence and tracking mechanisms. Natural antecedent status in subsequent discourse. A downstream affordance of definiteness. See also: familiarity, definiteness cluster.

ANIMACY

A semantic distinction between animate and inanimate entities. Grounded in cognitive salience and agency detection. Grammatical contrasts like *who* vs. *what*. Often cross-cuts lexical similarity classes. See also: personhood cluster.

ASPECT

A grammatical category describing how events are temporally and structurally construed. Maintained by distributional cue structures and morphological marking. Projectible patterns may diverge from textbook semantic definitions. A comparative concept whose language-specific mechanisms can drift. See also: projectibility, mechanistic drift.

ATOMICITY

The availability of entities as discrete units for reference and tracking. Maintained by object files and unit-based representation. Compatibility with numerals and singular reference. One component of the individuation cluster. See also: individuation cluster, boundedness.

AUXILIARY

A verbal category that contributes tense, aspect, or modality without full lexical content. Typically arises through grammaticalization and distributional entrenchment. Inversion, *do*-support, or restricted non-finite behaviour. Boundary cases are common in diachronic change. See also: verb, grammaticalization.

Backgrounded Constituent Infelicity

The constraint that backgrounded constituents resist focus-sensitive operations. Information-structural requirements on focus particles. Infelicity when a focus operator targets backgrounded material. A diagnostic from focus-particle research. See also: category.

basin (attractor)

A dynamical-systems metaphor: an attractor region toward which category states are drawn. Created by stabilizer-driven pull toward the cluster centre. Stable cores with tolerated peripheries. Explains graded membership with sharp boundaries. See also: dynamic discreteness, real gradience.

bidirectional inference

The mechanism coupling semantic construal to morphosyntactic form in comprehension and production. Realized by inference from morphosyntax to construal and back. Mutual predictability of count properties across constructions. The core homeostatic link in countability. See also: count cluster, individuation cluster.

boundedness

The property of having discrete edges that separate an entity from its environment. Grounded in perceptual segmentation and boundary detection. Supports exact counting and telic construals. One component of the individuation cluster. See also: individuation cluster, atomicity.

bounding node

A structural node posited to delimit extraction domains in early generative syntax. Constrains long-distance dependencies by structural configuration. Island effects and extraction failures beyond the node. A theory-internal construct rather than a mechanism. See also: interrogative phrase, anaphor.

canonical clause

An idealized benchmark clause used for descriptive comparison. A regimentation tool rather than a maintained kind. Core diagnostics converge most cleanly on canonical cases. Useful for exposition, not ontologically privileged. See also: class, comparative concept.

category

A cluster of properties used to organize linguistic phenomena. When genuine, maintained by stabilizers; when merely convenient, it is a class. A category earns kind-status only if it passes projectibility and homeostasis. The book's core question is which categories are kinds. See also: kind, two-diagnostic test, projectibility, homeostasis.

CHUNKING

Storage of high-frequency sequences as units, bypassing compositional assembly. Driven by frequency-based memory consolidation. Fluency and acceptability differences in entrenched frames. Supports entrenchment and stability at the cluster core. See also: entrenchment.

CLASS

A grouping used for description or convenience without a commitment to shared mechanisms. No stabilizer story required. May be projectible locally but need not pass the two-diagnostic test. Useful for taxonomy; not necessarily a natural kind. See also: category, fat class, thin class, comparative concept.

CLASSIFIER

A morpheme that mediates numerals and nouns in classifier languages. Encodes individuation or shape properties in the numeral phrase. Obligatory in counting contexts in languages like Mandarin. Relocates countability coupling to the classifier system. See also: countability, count cluster.

CLIQUISH STABILITY

Property-correlation: some cues reliably indicate the whole cluster. Maintained by stabilizers that keep correlations intact despite drift. Knowing one property predicts others. Supports induction without perfect instance stability. See also: projectibility.

COMPARATIVE CONCEPT

An analyst's tool constructed for cross-linguistic comparison. Not guaranteed to be supported by shared stabilizers across languages. Does not imply shared homeostasis or projectibility. Useful without being a kind. See also: field-relative projectibility.

COPIED KIND

A category whose members are produced from each other or a common template. Maintained by a transmission lineage with variation. Similarity explained by copying rather than essences. Grammatical categories are copied kinds par excellence. See also: replicator, interactor.

COUNT CLUSTER

The morphosyntactic properties that travel together in English count nouns (plural, cardinals, *many/few*, agreement). Maintained by acquisition and entrenchment in count frames, plus alignment in discourse. Implicational hierarchy: tight properties imply loose ones. Coupled to individuation by bidirectional inference. See also: individuation cluster, bidirectional inference, functional anchoring.

COUNTABILITY

The coupled system linking semantic individuation and morphosyntactic count marking. Maintained by bidirectional inference between construal and form.

Implicational hierarchy of count properties and robust projectibility. An interface category with two partially distinct clusters. See also: individuation cluster, count cluster, bidirectional inference.

CREATIVITY

The production of novel forms or uses beyond memorized patterns. Driven by productive schemas and analogical extension. Innovations that test and reshape category boundaries. A source of variation that can become stabilized. See also: analogy, transmission.

CROSSCUTTING KINDS

Overlapping causal networks where the same entity belongs to multiple kinds. Maintained by distinct stabilizers for each network. Different inductions for different purposes. Cross-cutting is expected under field-relative projectibility. See also: field-relative projectibility.

DEFINITENESS

A semantic category of identifiability, familiarity, and uniqueness. Maintained by discourse tracking, domain restriction, and Theory of Mind. Predicts anaphoric recoverability and discourse behaviour. Distinct from the morphosyntactic form cluster. See also: definiteness cluster, deitality, weak definite.

DEFINITENESS CLUSTER

The semantic cluster of familiarity, uniqueness, and identifiability. Maintained by discourse tracking, domain restriction, and Theory of Mind. Anaphoric recoverability and discourse behaviour. Distinct from the morphosyntactic deitality cluster. See also: deitality, Normal conditions.

DEITALITY

The morphosyntactic form cluster associated with English determiners, distinct from semantic definiteness. Formed by grammaticalization of demonstratives plus distributional constraints. *There*-resistance, partitive *of*, and hosting behaviour converge. Cross-cuts the definiteness cluster. See also: definiteness cluster.

DERIVED PROPER FUNCTION

A function acquired by recruitment beyond an item's original selected function. Arises through exploitation of stable side effects of a form or construction. Depends on the proper function remaining intact elsewhere. Parasitic rather than foundational. See also: proper function, Normal conditions.

DESIGNATUM

The entity as conceptualized by the speaker in context. Constraining inference from meaning to form in reference. Pro-form choice tracks construal over antecedent form. Distinct from antecedent and referent. See also: gender, pro-form inventory.

dominance condition

A sign-language constraint: if only one hand moves, the other uses an unmarked handshape. Motor and perceptual constraints on bimanual articulation. Restricted non-dominant handshapes across sign languages. Pairs with the symmetry condition. See also: symmetry condition.

dynamic discreteness

Sharp boundaries produced by mechanisms rather than definitions. Maintained by stabilizers that keep tolerance thresholds over time. Sharp judgments within speakers, variance across populations. Explains discreteness without essences. See also: relative tolerance, real gradience.

entrenchment

Strengthening of mental representations through repeated use. Driven by frequency effects and chunking in memory. Resistance to analogical change; faster processing and higher accessibility. Anchors category cores and stabilizes prototypes. See also: chunking, prototype.

enumerability

Compatibility with exact counting and cardinal quantification. Depends on stable unitization in conceptual structure. Accepts numerals and exact quantifiers. One component of the individuation cluster. See also: individuation cluster, count cluster.

epicene

A gender category unmarked for sex within the personal domain. Maintained by personhood-based construal rather than sex marking. Compatible with any sexed referent. Common in pronominal systems. See also: gender, personhood cluster.

epistemic kind

A category that serves epistemic purposes without claiming a shared causal mechanism. May be institutionally reinforced rather than mechanistically maintained. Can be stable without passing the two-diagnostic test. Khalidi's term for useful but non-kind groupings. See also: class, comparative concept.

error and repair

Interactive correction of production or interpretation errors in real time. Self-monitoring, interlocutor feedback, and repair sequences. Convergence after misfires; reduced divergence over repeated interactions. A fast stabilizer operating at discourse timescales. See also: alignment, stabilizer.

essentialism

The view that categories are defined by necessary and sufficient conditions. It is grounded in definitional criteria rather than stabilizers. Breaks down when exceptions proliferate without principled repair. The maintenance view rejects essentialism in favour of mechanism-based kinds. See also: nominalism, prototype.

ETIOLOGICAL KIND

A kind defined by its history of production or transmission. Maintained by copying lineages with variation and selection. Similarity is explained by lineage rather than by shared essence. Khalidi's term for history-defined natural kinds. See also: copied kind, transmission.

FACTIVITY

The property of predicates that presuppose the truth of their complements. Lexical semantics and pragmatic inference patterns. Presupposition survives negation and questions. An inferential grouping that cross-cuts similarity. See also: category.

FAMILIARITY

The property of being discourse-old or previously introduced. Maintained by discourse tracking and common ground updates. Supports anaphoric uptake and given/new contrasts. A core component of the definiteness cluster. See also: definiteness cluster, anaphoric recoverability.

FAT CLASS

A label that lumps distinct causal clusters into a single bin. Produced by multiple unrelated stabilizers with no shared mechanism. Fails projectibility across subtypes. Useful for pedagogy, not a natural kind. See also: thin class, comparative concept.

FIELD-RELATIVE PROJECTIBILITY

Projectibility indexed to analytic purpose or domain. Depends on task-specific stabilizers. A category may project for one purpose and fail for another. Explains cross-cutting categories. See also: projectibility, crosscutting kinds.

FORM CLUSTER

The morphosyntactic bundle associated with English determiners (*there*-resistance, partitive *of*, hosting). Grammaticalization of demonstratives and stabilized distributional constraints. Convergence of structural diagnostics across determiners. Also called the deitality cluster. See also: deitality, definiteness.

FUNCTIONAL ANCHORING

Lexical alternatives absorb pressure to regularise a quasi-count noun. Implemented through division of labour in the lexicon. Intermediate classes persist without extending tight properties. Explains stability of *cattle*/*police* and drift of *data*. See also: count cluster.

GENDER

A system of grammatically relevant contrasts over referents or nouns. Maintained by agreement systems or pro-form inventories. Consistent selection of gender-marked items in anaphora or agreement. Can be referential or noun-class based. See also: personhood cluster, pro-form inventory.

GRAMMATICALITY

The maintained coupling between morphosyntactic form and structural meaning. Maintained by entrenchment, acquisition, alignment, and the feeling of ungrammaticality. A sentence is grammatical when it sits within the basin of attraction for its form–value pairing. Not a binary switch but an HPC in its own right. See also: acceptability, structural meaning, zipper.

GRAMMATICALITY ILLUSION

A case where the feeling of (un)grammaticality mismatches the underlying structure. Produced by processing ease (garden paths) or semantic distractors (Escher sentences). Can dissociate detector output from form–value coupling. Parallel to perceptual illusions. See also: acceptability, grammaticality.

GRAMMATICALIZATION

Diachronic shift from lexical item to grammatical marker. Driven by frequency, reduction, reanalysis, and semantic bleaching. Layering and distributional expansion over time. Creates form - meaning decouplings that later stabilize. See also: mechanistic drift.

GRUE PROBLEM

Goodman's riddle: why do we project "green" but not "grue"? Explained by the fact that only mechanism-grounded properties are projectible. Gerrymandered predicates fail induction. Applied to language, pseudo-categories lack mechanisms. See also: projectibility.

HOMEOSTASIS

Active return after perturbation; stability by self-correction rather than mere persistence. Maintained by alignment, entrenchment, transmission, and other stabilizers across timescales. Perturbation sensitivity: weaken a mechanism and the clustering frays. The ontological anchor for genuine kinds. See also: stabilizer, perturbation sensitivity.

HOMEOSTATIC PROPERTY CLUSTER (HPC)

A category whose properties cluster because mechanisms maintain their co-occurrence rather than because of a shared essence. Maintained by acquisition, entrenchment, alignment, transmission, and functional pressure. Passes both projectibility and homeostasis (perturbation sensitivity). Membership can be graded and internally structured. See also: projectibility, homeostasis, stabilizer, two-diagnostic test.

HOMOGENEITY RESISTANCE

The property that proper parts are not of the same kind as the whole. Tied to construal of individuals rather than substance. Fails "half an N" tests for count nouns. One component of the individuation cluster. See also: individuation cluster, boundedness.

IDENTIFIABILITY

The property that a hearer can pick out the intended referent. Maintained by Theory of Mind and discourse modeling. First-mention definites are licensed when the description suffices for retrieval. A core component of the definiteness cluster. See also: definiteness cluster, familiarity.

INDEXICALITY

The linkage between linguistic forms and social meanings or stances. Maintained by sociolinguistic association and uptake. Variants signal identity, register, or attitude. Often discussed as social indexing. See also: social indexing, prestige selection.

INDIVIDUATION CLUSTER

The semantic profile that enables discrete construal: boundedness, atomicity, enumerability, and homogeneity resistance. Maintained by perceptual segmentation, object files, and cross-modal integration. Supports predictions about count morphosyntax. Coupled to the count cluster to yield countability. See also: count cluster, bidirectional inference.

INFLATION PROBLEM

The tendency to treat every stable pattern as an HPC kind. Fuelled by overgeneralization from local stability or institutional reinforcement. Categories appear projectible until the homeostasis test is applied. A methodological warning against false positives. See also: two-diagnostic test, fat class, thin class.

INSTANCE STABILITY

Stability of individual members retaining a cluster's properties over time. Supported by entrenchment, selection, or institutional reinforcement. Members resist drift even under perturbation. Distinct from cliquish stability. See also: cliquish stability, entrenchment.

INTERACTIVE ALIGNMENT

Convergence of interlocutors' linguistic choices during interaction. Realized through accommodation and repair in real-time conversation. Local convergence and resistance to innovations under misalignment. A fast-timescale stabilizer. See also: stabilizer, transmission.

INTERACTOR

An entity that interacts with its environment in ways that cause differential replication (Hull). Realized in communicative success and failure. Selection pressure visible in usage patterns. Pairs with replicator in an evolutionary account. See also: replicator.

INTERROGATIVE PHRASE

A phrase that introduces a question variable (e.g., wh-phrases). Licensed by interrogative constructions and dependency formation. Participates in wh-dependencies and question formation. Structural diagnostics vary across languages. See also: bounding node, category.

ITERATED TRANSMISSION

Structure emerging from multi-generational learning bottlenecks. Produced by filtering for learnable variants. Convergence toward compressible/compositional systems. A macro-stabilizer of linguistic structure. See also: copied kind, entrenchment.

KIND

A category that supports induction because mechanisms maintain its clustering. Stabilizers keep the profile coherent across speakers and time. Passes the two-diagnostic test. In this framework, linguistic kinds are HPC kinds rather than definitional essences. See also: hpc, copied kind, two-diagnostic test.

MADAGASCAR FALLACY

The mistake of treating an ecosystem as if it were a single kind. It bundles mechanisms that maintain components rather than the whole. Fails homeostasis at the wrong grain. A warning about scale in category analysis. See also: two-diagnostic test.

MAINTENANCE VIEW

The claim that categories are real because they are maintained by stabilizing mechanisms. Supported by a braid of mechanisms across timescales. Implies the two-diagnostic test: projectibility plus homeostasis. Rejects both essentialism and nominalism. See also: hpc, two-diagnostic test.

MASS-WORD

Jespersen's term for nouns that resist counting and take mass quantifiers. Maintained by mass construal and distributional patterns. Rejects numerals but accepts *much/little*. Highlights grammatical packaging over ontology. See also: countability, count cluster.

MECHANISM

Entities and activities organized in such a way that they are responsible for a phenomenon. Mechanisms include acquisition, alignment, entrenchment, and transmission. Identified by perturbation: weaken the process and the cluster frays. Adopts the consensus definition of Illari and Williamson (2012). See also: stabilizer, homeostasis, perturbation sensitivity.

mechanistic drift

Mechanisms that maintain a category can differ from those that originally produced it. Driven by diachronic reconfiguration of stabilizers. Mismatch between current cues and historical rationale. Explains semantic drift under stable labels. See also: grammaticalization.

meta-Occam

The principle that parsimony applies to mechanisms rather than to the objects they produce. In evolutionary theory: natural selection is simple; organisms are not. In linguistics: stabilizers are tractable; categories are complex. A good framework requires fewer mechanism types, not fewer category types. Krakauer's term; applied here to the HPC framework. See also: stabilizer, homeostasis, hpc.

negative polarity item (NPI)

An expression whose distribution is restricted to negative or downward-entailing contexts. Multiple licensing mechanisms yield heterogeneous subtypes. Restricted distribution with distinct licensing profiles. Often a distributional class rather than a single kind. See also: class, category.

nominalism

The view that categories are convenient labels without commitment to natural kinds. Requires no stabilizer story. No homeostasis or projectibility demanded. Rejected by the maintenance view. See also: essentialism, comparative concept.

non-finite clause

A residual category for clauses lacking primary tense inflection. No single shared mechanism maintains the set. Fails both projectibility and homeostasis across subtypes. A negative class; best replaced by positive constructions. See also: fat class, category.

Normal conditions

The circumstances under which a device performs its proper function. Sustained by shared context and recoverable referents. Breakdown produces miscoordination; derived uses exploit stable side effects. Frames proper vs. derived function. See also: proper function, derived proper function.

noun

A lexical category associated with nominals and argument structure. Maintained by distributional cues, morphology, and acquisition biases. Cliquish stability across multiple diagnostics (determiners, number, agreement). Cross-linguistic comparability requires mechanism mapping. See also: verb, subject, copied kind.

open texture

Hart's term for the penumbral region where rules underdetermine outcomes. Emerges when stabilizers are weak or in competition. Persistent variability and institutional repair. Grammar has no Supreme Court; resolution is emergent. See also: dynamic discreteness, mechanistic drift.

PERFECTIVE

An aspectual category presenting events as bounded or completed. Maintained by morphological marking and cue-outcome associations. Distributional asymmetries across tense frames. Highlights the gap between semantic definition and usage. See also: aspect, boundedness.

PERSONHOOD CLUSTER

The semantic properties that make a referent construable as a person. Grounded in Theory of Mind and social cognition. Triggers personal pro-forms (*who*, *he/she*) over non-personal ones. A semantic anchor for gender systems. See also: gender, animacy.

PERTURBATION SENSITIVITY

The diagnostic that a category frays when a stabilizer is weakened. Observed when interventions target acquisition, alignment, or transmission. Predictable degradation patterns under perturbation. Operationalizes homeostasis. See also: homeostasis, stabilizer.

PRESTIGE SELECTION

Spread of variants due to association with high-status speakers. Social evaluation and identity signaling reinforce the variant. Prestige forms persist despite competing regularization pressures. A social stabilizer for arbitrary variants. See also: social indexing, transmission.

PRO-FORM INVENTORY

The set of pro-forms and their distributional constraints in a language. Maintained by acquisition and usage-based entrenchment. Stable patterns in wh-forms and pronominal selection. A lexico-grammatical cluster in pro-form gender. See also: gender, designatum.

PROCESSING ECONOMY

Pressure to minimize cognitive and articulatory cost in production and comprehension. Frequency-driven facilitation and memory constraints. Shorter or more predictable forms are preferred and stabilize. Interacts with social and functional pressures. See also: entrenchment, chunking.

PROJECTIBILITY

A category's capacity to support induction from observed cases to unobserved ones. It relies on stable correlations enforced by stabilizers. Predicts held-out data or generalises across contexts better than chance. Field-relative: projectible for one purpose may fail for another. See also: homeostasis, two-diagnostic test, field-relative projectibility.

proper function

In Millikan's framework, the function an item has in virtue of its history of selection. Secured by reproductive success of uses that perform the function. Explains why the form persists across generations. Historical, not intentional or definitional. See also: derived proper function, Normal conditions.

proper name

A semantic category of expressions that directly refer to individuals. Maintained by discourse tracking and stable reference to individuals. Rigid designation and referential opacity effects. Cross-cuts the syntactic category of proper nouns. See also: proper noun, definiteness cluster, field-relative projectibility.

proper noun

A syntactic category of nominals with distinctive distributional behaviour. Maintained by morphosyntactic cues such as article resistance and agreement patterns. Distributional diagnostics cluster in specific constructions. Distinct from the semantic category of proper names. See also: proper name, noun, field-relative projectibility.

prototype

The most typical member of a category. Shaped by entrenchment and frequency effects. Typicality effects and graded membership. Descriptive of structure; requires a mechanism story to explain stability. See also: entrenchment, real gradience.

quasi-count noun

A noun that allows some count diagnostics but resists tight ones. Stabilized by functional anchoring and lexeme competition. Accepts *many* but resists low numerals and articles. A stable intermediate in the countability system. See also: functional anchoring, singulative.

real gradience

Gradient structure as meaningful distance from category centres. Shaped by variable stabilizer strength across the basin. Typicality and marginality effects. Gradience is signal, not noise. See also: basin, dynamic discreteness.

relative tolerance

Membership depends on scale-relative tolerance of change. Determined by decision thresholds indexed to context and magnitude. Small changes acceptable at one scale but not another. Explains tolerance intuitions at category edges. See also: dynamic discreteness.

replicator

An entity that passes copies of itself through time with variation (Hull). Sustained by transmission with differential survival. Selection on variants across generations. Supports evolutionary framing of language change. See also: interactor, copied kind.

SEMANTIC TRANSPARENCY

The predictability of form - meaning mappings in a system. Stabilized by consistent coupling between semantic and morphosyntactic cues. Low mismatch rates and easier acquisition. Supports tight coupling between clusters. See also: pro-form inventory, countability.

SIMPLE CAUSAL THEORY

A view that kinds are anchored by primary properties that causally generate secondary properties. One-way causal dependence rather than feedback or self-correction. Stable property dependencies without homeostatic return. Contrasts with homeostatic accounts. See also: stable property cluster, homeostasis.

SINGULATIVE

A morphological marker deriving a single unit from a collective or mass base. Grammaticalizes individuation within the count system. Enables numerals and tight count diagnostics. Common in Welsh and Arabic. See also: countability, quasi-count noun.

SOCIAL INDEXING

Association of linguistic variants with social identities, stances, or registers. Sociolinguistic signaling reinforced by community norms. Variant use correlates with social groupings and style shifts. A stabilizer that can protect arbitrary variants. See also: prestige selection, transmission.

SPECIES

A biological grouping of populations maintained by overlapping mechanisms rather than a single essence. Gene flow, shared selection pressures, and reproductive systems. Cohesion with boundary cases (ring species, hybrids). A paradigm case for mechanism-based kinds. See also: hpc, etiological kind.

STABILIZER

The functional role played by a mechanism in maintaining the clustering of a category. Instances include acquisition, entrenchment, alignment, transmission, and functional pressure. Removing it changes the clustering in predictable ways. Refers to the maintenance function rather than the causal structure itself (Illari & Williamson, 2012). See also: homeostasis, perturbation sensitivity, mechanism.

STABLE PROPERTY CLUSTER

A cluster of co-occurring properties that remains stable over time or context. Stability can arise from multiple stabilizers, not only homeostasis. Cliquish stability with projectible correlations. Slater's broadened stability framework. See also: cliquish stability, homeostasis.

STANDARDIZATION

Institutional codification and enforcement of a linguistic norm. Education, style guides, and prescriptive feedback. Reduced variation and convergence toward a prestige standard. Creates explicit feedback loops outside ordinary transmission. See also: social indexing, transmission.

STRUCTURAL MEANING

The contribution of grammar before lexical content fills the slots: modification, argument structure, tense, clause type, information structure. Maintained by the same stabilizers as grammaticality. Coupled to morphosyntactic form in the grammaticality zipper. Subcompositional; the instruction manual, not the assembled product. See also: grammaticality, value.

SUBJECT

A syntactic function cluster defined by converging behavioural and coding properties. Agreement, word order, case marking, and discourse roles. Convergence of multiple diagnostics rather than a single definition. Cross-linguistic identity is not guaranteed. See also: comparative concept, category.

SYMMETRY CONDITION

A sign-language constraint: if both hands move, they share handshape and mirror movement. Motor planning and perceptual constraints on bimanual coordination. Cross-linguistic convergence in two-handed signs. An example of convergent stability without borrowing. See also: dominance condition.

THIN CLASS

A pattern that is weakly maintained or not maintained at all. Occurs when stabilizers are absent or too weak. Fails homeostasis; disappears under perturbation. Often a byproduct of other mechanisms. See also: fat class.

TRANSMISSION

Intergenerational passage of linguistic patterns. Iterated learning filters for stable, learnable variants. Convergence toward compressible systems under repeated learning. A macro-stabilizer of categories and constructions. See also: iterated transmission, acquisition.

TWO-DIAGNOSTIC TEST

The criterion for genuine kinds: projectibility plus homeostasis. Implemented by pairing the projectibility and homeostasis diagnostics. Pass both to count as an HPC; fail one to diagnose thin/fat/negative. Operational rather than definitional. See also: projectibility, homeostasis, perturbation sensitivity.

UNICEPT

A concept that tracks sameness through multiple fallible methods (Millikan). Sustained by convergent cues with no privileged diagnostic. Cliquish stability:

some cues reliably indicate the cluster. Matches multi-diagnostic category tracking. See also: cliquish stability, two-diagnostic test.

UNIQUENESS

The property of having exactly one salient candidate in a domain. Maintained by domain restriction and contextual narrowing. Infelicity when multiple candidates are equally salient. A core component of the definiteness cluster. See also: definiteness cluster, identifiability.

VALUE

A unit's contribution when deployed; what it counts as in the system. Value is relational, not intrinsic: for phonemes, contrastive identity; for morphemes, conventional meaning; for constructions, interpretive potential. Captures what is coupled to form at each level of linguistic structure. Introduced in Chapter 3; developed in Chapter 13. See also: coupling, form cluster, definiteness cluster.

VERB

A lexical category associated with predication and event structure. Maintained by distributional frames, inflectional paradigms, and acquisition. Convergent diagnostics (tense/aspect marking, argument structure). Stability varies by language and construction. See also: noun, aspect.

VERBLESS CLAUSE

A clause-like construction with subject - predicate structure but no verb in the predicate. Arises under competing diagnostics (predication vs. VP-headedness). Shows decoupling of form and function. Illustrates pressure against essentialist definitions. See also: two-diagnostic test.

VOICE

A morphosyntactic system that maps argument roles to grammatical functions. Maintained by constructional alternations and discourse pressures. Voice morphology and alternations in argument realization. Cross-linguistic mechanisms vary widely. See also: subject, category.

WEAK DEFINITE

A definite article use that lacks a unique or familiar referent (institutional/role frames). A derived function exploiting stable form-cluster properties. Form-cluster behaviour without definiteness-cluster satisfaction. Productive in stereotyped activity frames. See also: deitality, definiteness, form cluster.

Subject Index

Name Index

Lexical Index