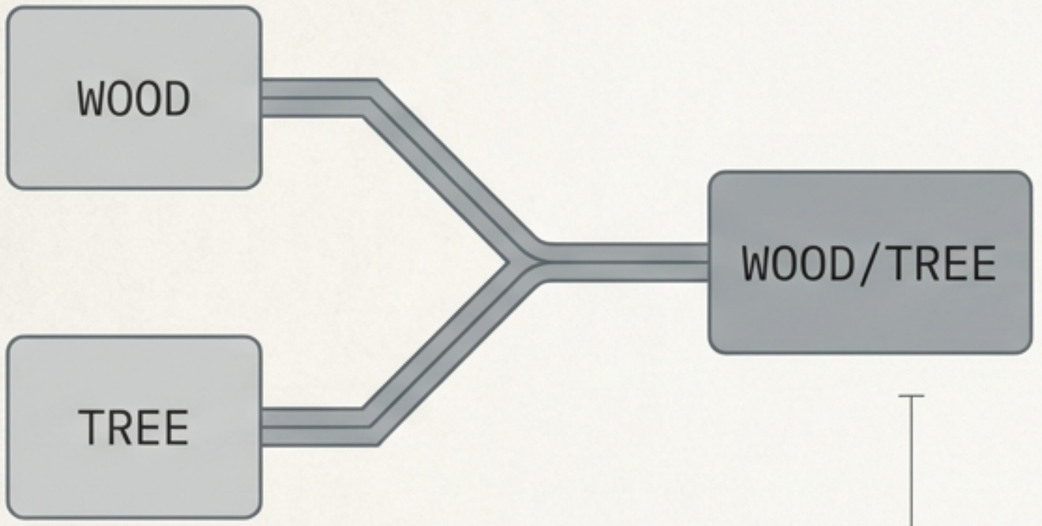
Navigating the Future

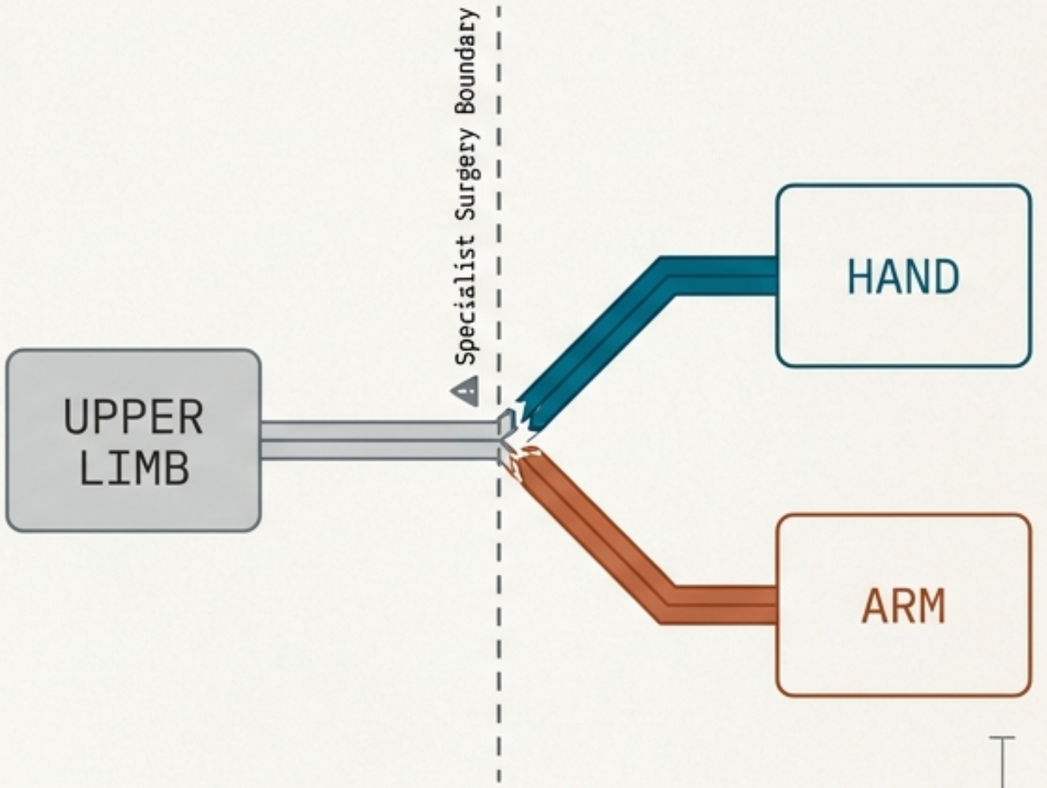
A Reachability Theory of Lexical Preservation.

Why semantic distinctions survive despite the evolutionary pressure for compression.

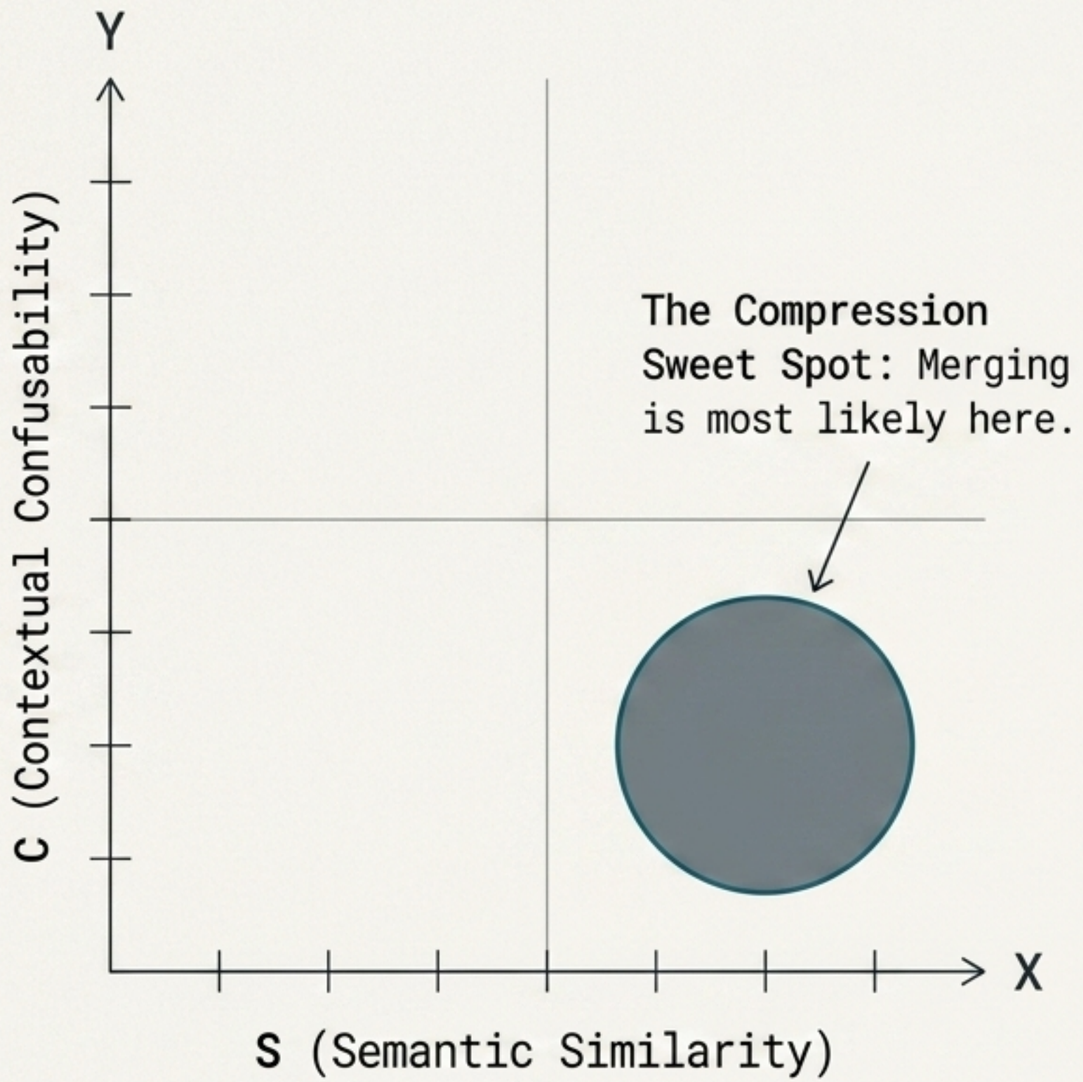
The Paradox of Compression



In many languages, WOOD and TREE seamlessly merge into a single word (e.g., Arabic *shajara*).



Languages face constant pressure to compress vocabulary. Yet some conceptually similar words strictly separate in specialized discourse. Why?

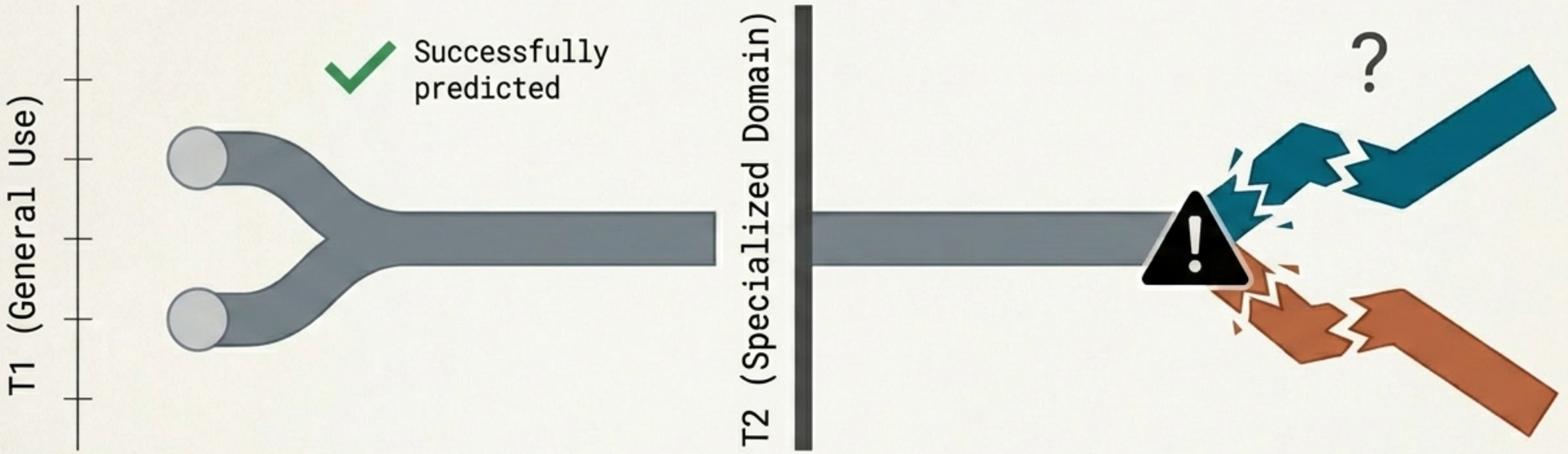


The Standard Paradigm: The Compression Account

$$Pr(L \mid X, Y) = f(S(X, Y), C(X, Y))$$

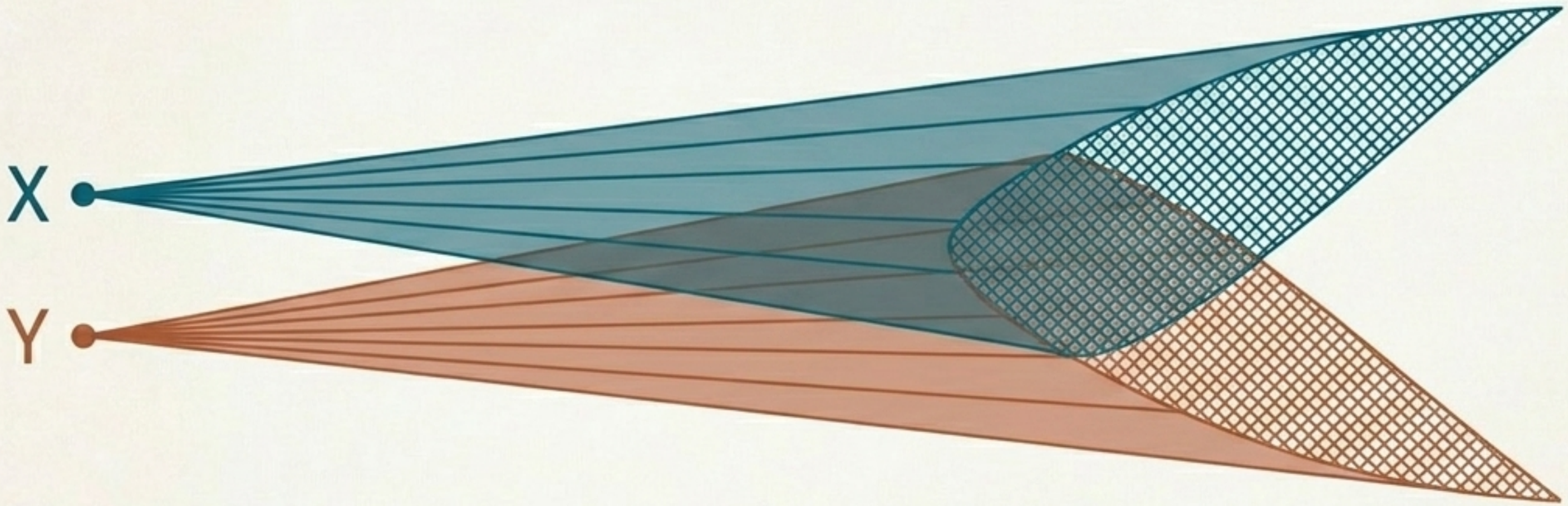
- **The Logic:** Languages merge words when they mean similar things (S) and are easy to tell apart in context (C).
- **The Success:** This model predicts lexical organization across 2,000+ languages, balancing efficient communication against ambiguity.

The Synchronic Blind Spot



The Limitation	The Anomaly
The standard model predicts which structures exist today, but provides no mechanism to distinguish harmless compression from destructive compression over time.	It cannot explain why stable, successfully merged words suddenly fragment in specialized domains despite high similarity and low confusability.

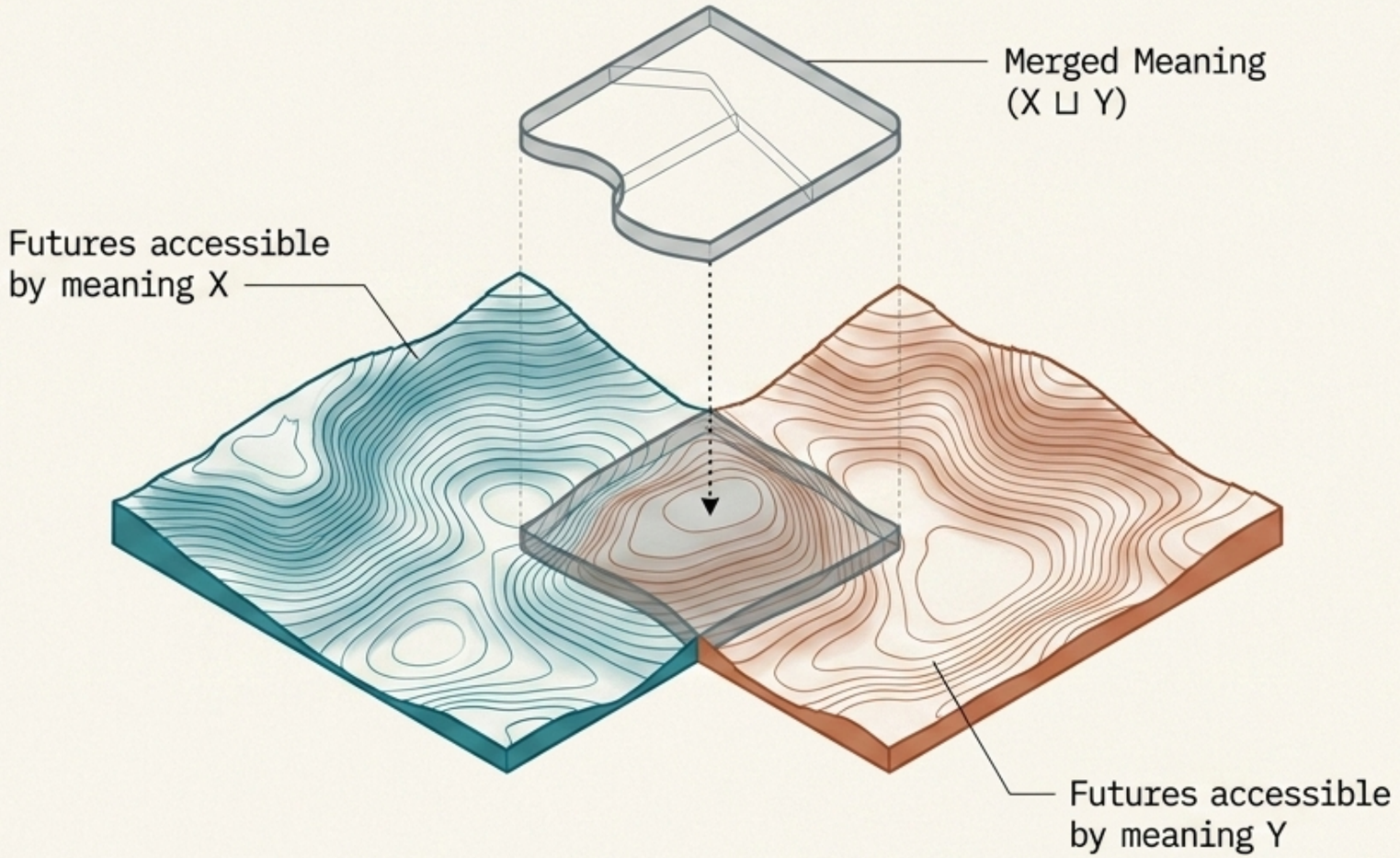
The Missing Variable: Reachability Divergence (R_θ^D)



$$R_\theta^D(X, Y) = \mu_\theta(A_\theta^X \triangle A_\theta^Y)$$

The Definition:	The Core Idea:
The degree to which two meanings, once distinguished, open fundamentally different future trajectories of action, inference, or repair.	Language serves as a cartography for action. If merging two words forecloses a necessary future action, the distinction will survive.

The Geometry of Reachability Loss



Constraint 1

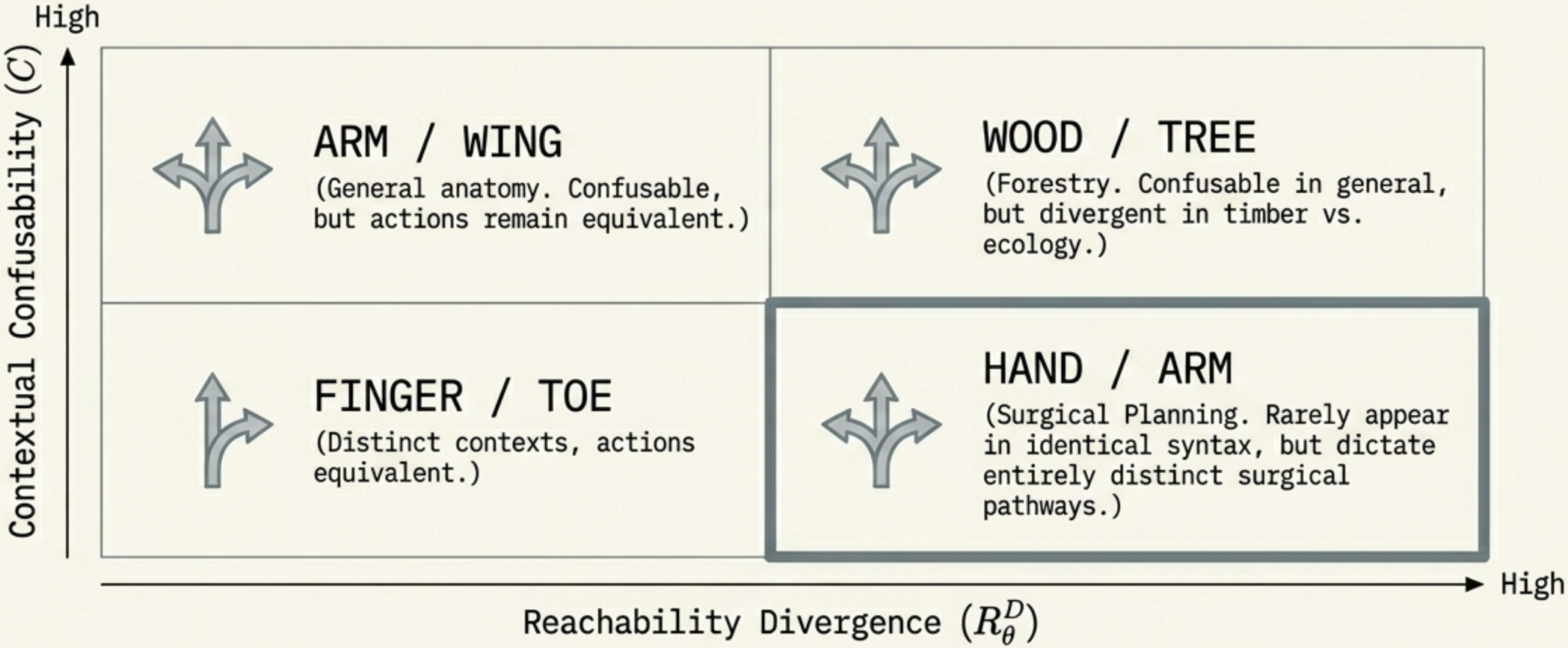
A merged meaning cannot exceed the union of its parent meanings.

Reachability Loss (λ_θ)

The mathematical gap between what the original words could reach and what the merged word can reach.

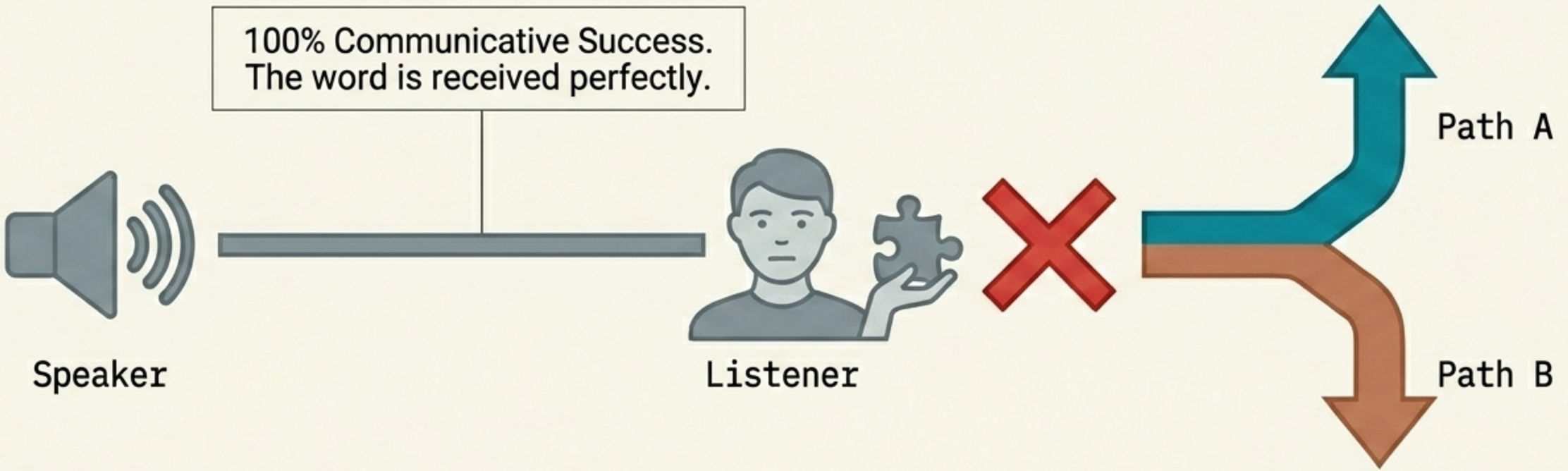
If this loss is greater than zero, the compression is destructive.

The Independence Matrix: Decoupling Variables



The Hand/Arm anomaly proves that navigational failure occurs even when communicative reception is perfect.

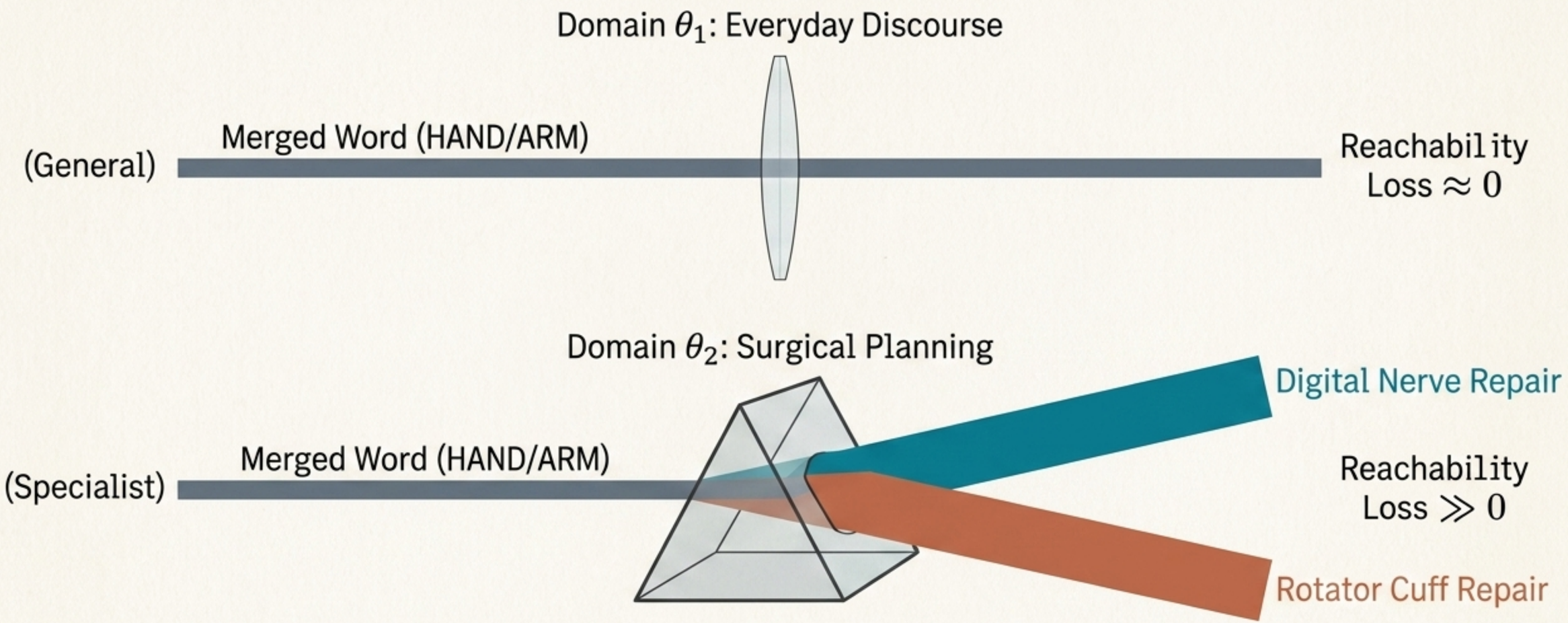
Theorem 1: Navigational Collapse



The Mechanism: The linguistic form is a bottleneck. Even if an agent correctly resolves contextual ambiguity, if the merged meaning's admissible future region is smaller than the original's, they are paralyzed at the fork.

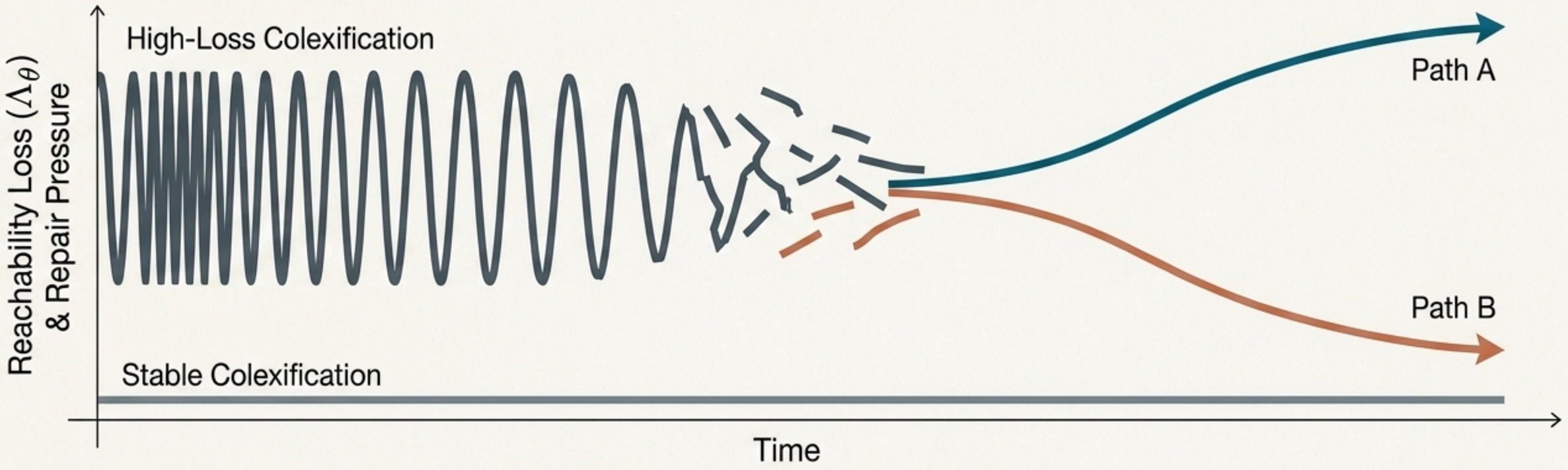
The Insight: The failure lies in the representational system, not the transmission channel. Accurate communication does not guarantee accurate navigation.

Proposition 4: Domain Divergence



The Rule: A colexification stable in one domain may be highly unstable in another. This mechanically predicts the emergence of specialized technical, legal, and medical vocabularies as necessary repairs.

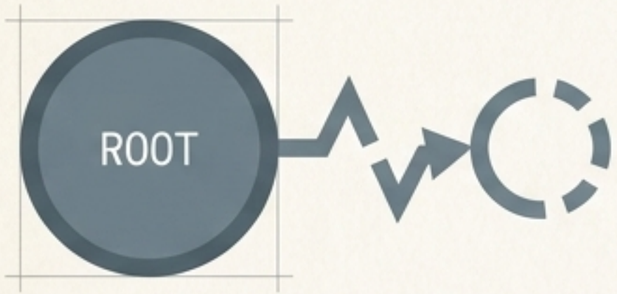
Theorem 2: The Dynamics of Stability



Assumption:	The Diachronic Prediction:
Repair pressure is a non-decreasing function of reachability loss.	High-loss colexifications don't just eventually fragment—they fragment faster. Stable colexifications are only those where the merged meaning can reach everything both original meanings could reach.

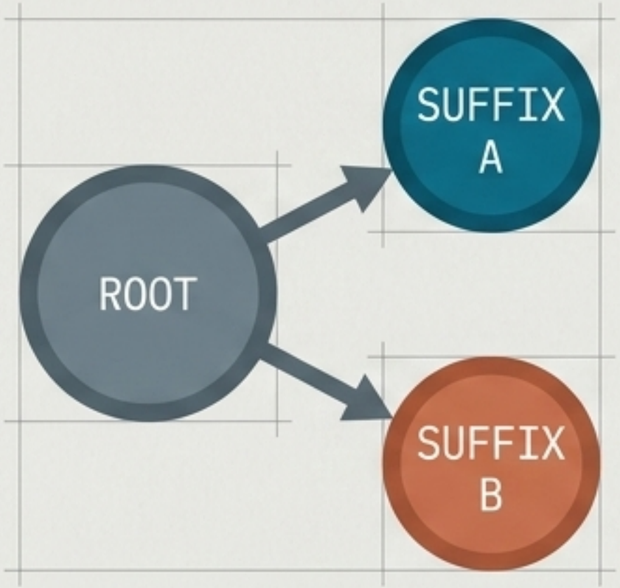
Theorem 3: The Repair Cost Balance

Full Colexification



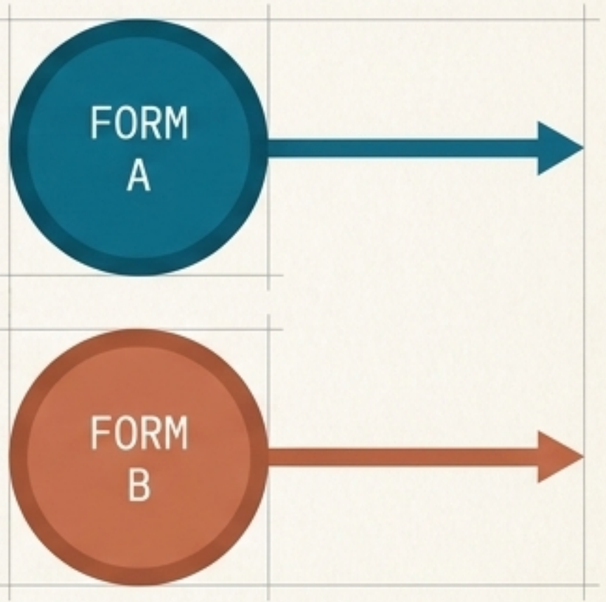
Cost = High Reachability Loss ($\Lambda_\theta > 0$)

Partial Colexification (The Sweet Spot)



Cost = Cost-Minimizer (D_{part})

Full Differentiation

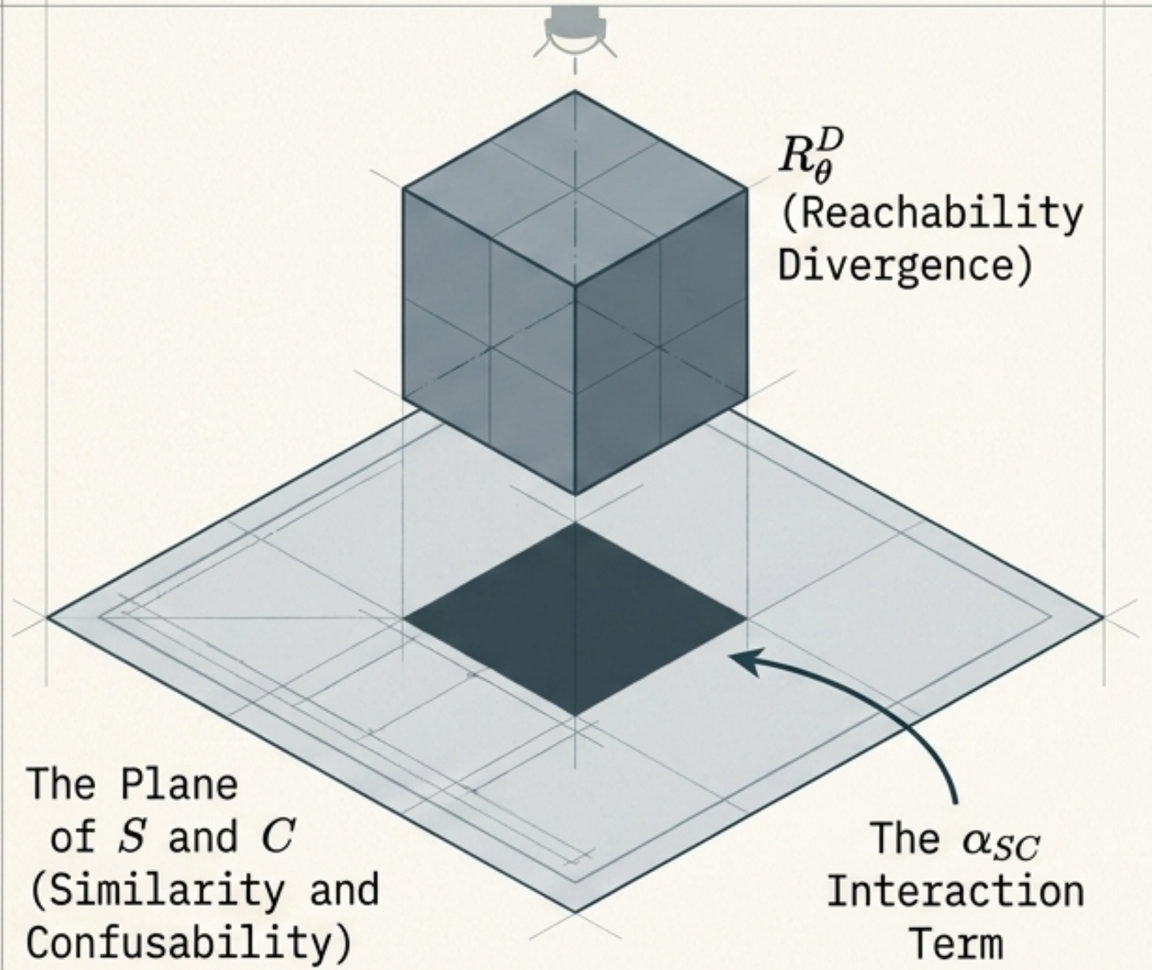


Cost = High Formal Complexity ($D_{\max}$)

$$\text{Cost}(F) = \lambda \cdot D_{form} + (1 - \lambda) \cdot \Lambda_\theta$$

The Optimum: Morphological marking (shared root + new suffix) is the unique mathematical cost-minimizer. It retains the compression value of the root while restoring the necessary navigational distinction.

Theorem 4: The Mathematical Projection



The Falsifiable Prediction:

The best existing statistical model (Brochhagen et al.) includes an unexplained interaction term ($\alpha_{SC} \cdot S \cdot C$).

The Proof:

The framework proves this interaction term is merely the statistical shadow of the omitted Reachability variable projected onto the (S, C) plane.

If R_θ^D is measured and added to the model, the mysterious interaction will mathematically weaken.

The Blind Spot Illuminated: Pilot Data

Word Pair	Standard Model Predicts	Reachability Theory Predicts	Empirical Reality
WOOD / TREE	❌ Full Colexification	✅ Partial Colexification	Partial or Full/Partial
HAND / ARM	❌ Full Colexification	✅ Partial Colexification	Partial or Full/Partial
SKIN / BARK	❌ Full Colexification	✅ Partial Colexification	Partial
TONGUE / LANGUAGE	❌ Full Colexification	✅ Partial Colexification	Partial or Full/Partial

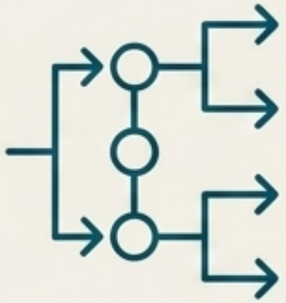
Relying on similarity and confusability alone generates false predictions for pairs with divergent future actions. Reachability mathematically corrects this blind spot.

Operationalizing the Variable

Reachability Divergence is an operationally grounded, measurable performance metric.

Action-Consequence Corpora (Gold Standard)

Extract specialized documents (medical records, legal cases) to calculate the **empirical empirical action distribution** following specific lexical labels.



Downstream LLM Divergence (Scale)

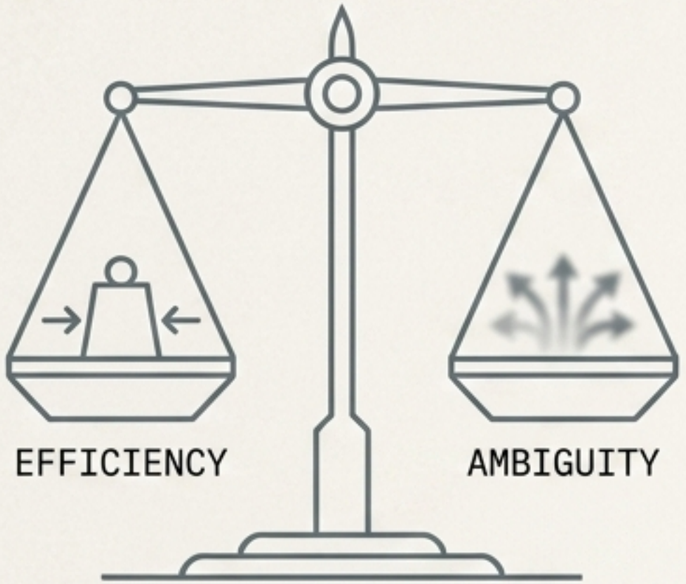
Prompt language models to generate likely next actions given a domain-indexed concept, concept, measuring the **total variation distance** between outputs.

Structured Expert Elicitation (Baseline)

Domain experts explicitly enumerate the **disjoint** and **overlapping action-pathways** for **colexified pairs**.

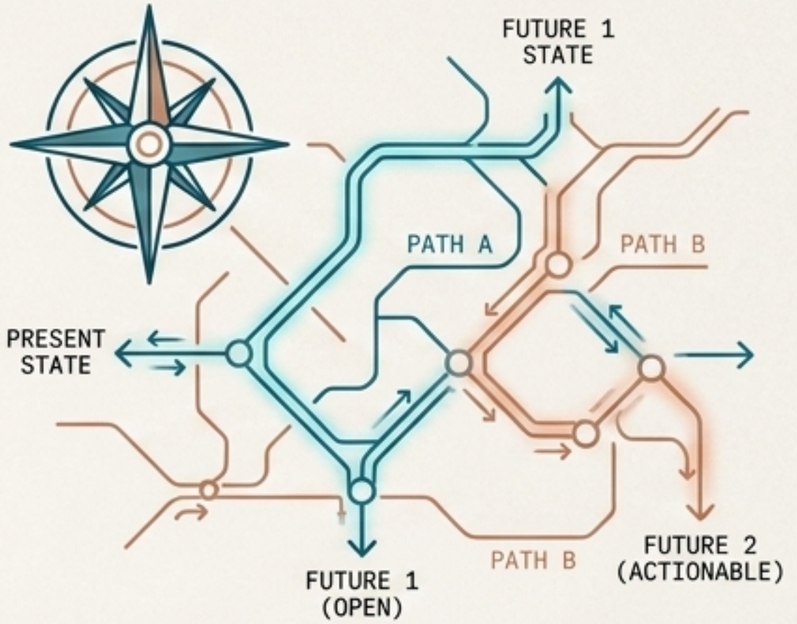
The Paradigm Shift

Language as Efficient Communication



A system balancing maximum compression against the risk of contextual ambiguity.

Language as Navigational Cartography



A system preserving the exact distinctions required to keep a community's future open and actionable.

Lexical distinctions are not merely representations of conceptual structure; they are instruments of navigation. A distinction is preserved not because it is communicatively ambiguous to lose it, but because its loss is `action-theoretically inadmissible`.