Why Distinctions Survive

A Reachability Theory of Lexical Preservation

Flyxion

Independent Researcher

June 2026

Abstract

Languages do not merely communicate information. They also preserve the distinctions required to keep a community's future navigable. A recent large-scale study of colexification across nearly two thousand languages explains lexical organisation through the interaction of two pressures: lexical compression and lexical differentiation. The account is empirically powerful but structurally incomplete. It predicts which lexical structures should exist; it cannot predict which should persist.

We identify the missing variable as *reachability divergence*—the degree to which two meanings, once distinguished, open different future trajectories of action, inference, repair, or error. Reachability divergence is formalised via an explicit measurable performance function on domain-indexed future spaces, making the construct well-defined and operationally grounded. We show that reachability divergence is logically independent of contextual confusability: the critical dissociation occurs precisely among meaning pairs that are cross-linguistically attested as colexification candidates—high semantic similarity, low contextual confusability, yet high domain-specific reachability divergence, the (H, L, H) cell that is invisible to the compression account. We introduce a domain-indexed formal framework and derive four principal theorems and one structural proposition: the *Navigational Collapse Theorem*, the *Stability Theorem* (derived from explicit dynamical assumptions including a repair cost trade-off that explains why partial colexification is preferred to full differentiation), the *Repair Theorem* (as an optimality result over admissibility-restoring repairs, with an explicit scope remark on the uniqueness claim), the *Domain Divergence*

Proposition (a direct consequence of domain indexing with major interpretive weight for technical vocabulary emergence), and the *Projection Theorem* (generating a falsifiable prediction about the interaction term in the best existing statistical model of colexification). We further present a pilot estimation of reachability divergence for twenty cross-linguistically attested meaning pairs, demonstrating that R^{θ}_{D} varies independently of semantic similarity and contextual confusability and corrects all four cases in which the compression account mispredicts within the (H, L, H) cell. This paper constitutes a theoretical research programme: it specifies the formal framework, establishes the independence of the central construct, derives the principal results, and identifies the empirical tests required for full validation.

Contents

1 The Problem Compression Cannot Solve **2**

2 The Standard Account and Its Limits **3**
2.1 The compression–differentiation model 3
2.2 Three structural limitations 3

3 The Reachability Framework **4**
3.1 Foundational definitions 4
3.2 The merge operator and reachability loss 5
3.3 The extended lexicalization model 6

4 Reachability Divergence Is Not Contextual Confusability **6**
4.1 The conceptual distinction 6
4.2 Kanuri partial colexification as morphological repair 7
4.3 The independence table with colexification-relevant examples 8

5 Five Theorems **10**
5.1 Theorem 1: Navigational Collapse 10
5.2 Theorem 2: Stability 12
5.3 Theorem 3: Repair 13
5.4 Proposition 4: Domain Divergence 14
5.5 Theorem 4: Projection 16

6 Faithfulness and the Geometry of Compression **17**
6.1 Faithful projections and quotient structure 17
6.2 Lexical distortion 18

7 A Pilot Estimation of Reachability Divergence **18**
7.1 Method and evidential status 18
7.2 Results 19
7.3 Observations 20

8 Operationalizing Reachability Divergence **21**
8.1 Action-consequence corpora 21
8.2 Downstream divergence in language models 22
8.3 Expert elicitation with structured protocol 22
8.4 The empirical test of Theorem 5 22

9 Consequences for the Compression Account **23**
9.1 Family-level heterogeneity 23
9.2 The English-proxy distortion 24
9.3 The interaction term 24
9.4 Consequential ambiguity and reachability divergence 24

10 Language as Reachability Preservation **25**

11 Scope, Limitations, and the Research Programme **26**
11.1 What this paper establishes 26
11.2 What this paper does not establish 27
11.3 Principal limitations 27
11.4 Next steps 28

12 Conclusion **28**

1. The Problem Compression Cannot Solve

Consider the distinction between WOOD and TREE. Many languages colexify these meanings—that is, they use a single form to express both. From the standpoint of semantic similarity, the two concepts are extremely close. From the standpoint of contextual confusability, many utterances permit effortless disambiguation. A theory of lexical compression therefore predicts persistent pressure toward form reuse in this pair, and that prediction is confirmed by the cross-linguistic record.

Now consider the same pair in the context of a forestry management institution. TREE controls survey decisions, replanting schedules, and biodiversity assessments. WOOD controls timber yield calculations, drying schedules, and board-foot pricing. A merged form that does not distinguish these meanings will produce communicatively efficient utterances while generating navigational failures: agents who correctly understand what was said may nonetheless act on the wrong future trajectory, because the representation no longer marks which set of actions is admissible.

The question this paper addresses is: why do languages that have merged these meanings in general vocabulary sometimes *recover* the distinction in specialist domains? And, more generally: why do certain distinctions survive despite sustained pressure toward compression?

The compression account—the most successful current framework for predicting cross-linguistic lexical organisation [3, 4]—cannot answer this question. It is synchronic: it predicts which meanings should colexify under current conditions but provides no criterion for distinguishing harmless compression from destructive compression, and no mechanism for predicting which colexifications will remain stable over time.

The missing variable is *reachability divergence*: the degree to which two meanings, once distinguished, open different futures of action, inference, repair, or decision. We develop this variable formally, establish its logical independence from contextual confusability, derive five principal theorems, and present a pilot empirical estimation.

2. The Standard Account and Its Limits

2.1. The compression–differentiation model

Let $S(X, Y) \in [0, 1]$ denote semantic similarity between meanings X and Y, and $C(X, Y) \in [0, 1]$ denote contextual confusability. The compression account models the probability of lexicalization pattern $L \in \{\text{none}, \text{partial}, \text{full}\}$ as

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

The best-fitting model includes an interaction term [4]:

$$\eta(X, Y) = \alpha_0 + \alpha_S \cdot S + \alpha_C \cdot C + \alpha_{SC} \cdot S \cdot C, \tag{1}$$

where η is the linear predictor in a multinomial regression. The model predicts: full colexification is most likely when S is high and C is low; partial colexification is most likely when S and C are both moderate-to-high; no colexification is most likely when S is low. The empirical fit across nearly two thousand languages is strong.

2.2. Three structural limitations

Diachronic silence. The model is synchronic. It predicts the current distribution of colexification patterns but has no criterion for distinguishing harmless compression (merging meanings whose futures are equivalent) from destructive compression (merging meanings whose futures diverge). It therefore cannot predict which colexifications will persist, which will fragment, or which will acquire morphological repair.

Family-level heterogeneity. Partial colexifications are substantially more heterogeneous across language families than full colexifications [4]. The compression account attributes this to typological variation in morphological productivity, but provides no mechanism to predict *which* families will diverge on *which* meaning pairs.

The unexplained interaction. The interaction term $\alpha_{SC} \cdot S \cdot C$ in (1) is identified statistically but not given a structural interpretation. It is an empirical coefficient in need of a theoretical account.

3. The Reachability Framework

3.1. Foundational definitions

Definition 3.1 (Meaning Space)**.** A *meaning space* $(\mathcal{M}, d_{\mathcal{M}})$ is a metric space whose points are semantic representations of lexical concepts. The metric $d_{\mathcal{M}}$ reflects semantic distance, operationalizable by distributional similarity, association norms, or conceptual space geometry [6].

Definition 3.2 (Domain and Future Space)**.** A *domain* θ is a structured practice, institution, or community of action. The *future space* $(\Omega^\theta, \mathcal{F}^\theta, \mu^\theta)$ associated with domain θ is a probability space where Ω^θ is the set of possible future states (action outcomes, inference results, decisions, error events) available to agents in θ; $\mathcal{F}^\theta$ is a σ-algebra on Ω^θ; and μ^θ is a probability measure representing the prior distribution over futures under default navigation in θ.

Definition 3.3 (Performance Function)**.** A *performance function* for domain θ is a measurable map $g^\theta : \mathcal{M} \times \Omega^\theta \to [0, 1]$, where $g^\theta(X, \omega)$ is the probability of navigational success when meaning X is the operative concept and $\omega \in \Omega^\theta$ is the future state being navigated to. Measurability with respect to $\mathcal{F}^\theta$ (in the second argument) is required so that the level sets below are well-defined measurable sets.

Definition 3.4 (Admissible Future Region)**.** For meaning $X \in \mathcal{M}$, domain θ, and performance function g^θ, the *admissible future region* at threshold $\tau \in (0, 1)$ is

$$\mathcal{A}_X^\theta = \{\omega \in \Omega^\theta : g^\theta(X, \omega) > \tau\}.$$

Measurability of $\mathcal{A}_X^\theta$ with respect to $\mathcal{F}^\theta$ follows from the measurability of $g^\theta(\cdot, \cdot)$ in its second argument and the fact that $\{\omega : g^\theta(X, \omega) > \tau\}$ is a pre-image of the open set $(\tau, 1]$ under a measurable map.

Remark. The theorem statements do not depend on the specific value of τ, only on its existence. The threshold is domain-relative: high-stakes domains (surgery, air traffic control) set high τ; low-stakes domains set low τ. The performance function g^θ encodes the domain's goals and error-costs. Different choices of g^θ yield different $\mathcal{A}_X^\theta$, but the qualitative structure of the theorems—stability, repair, domain divergence—holds for any g^θ satisfying

Definition 3.3.

Definition 3.5 (Reachability Divergence)**.** The *reachability divergence* between meanings X and Y in domain θ is

$$R_D^\theta(X,Y) = \mu^\theta\left(\mathcal{A}_X^\theta \triangle \mathcal{A}_Y^\theta\right),$$

where $A \triangle B = (A \setminus B) \cup (B \setminus A)$ is the symmetric difference. Since μ^θ is a probability measure, $R_D^\theta(X,Y) \in [0,2]$.

3.2. The merge operator and reachability loss

A central object in the stability analysis is the *merged meaning* $X \sqcup Y$—the semantic representation accessed by a colexified form that expresses both X and Y.

Definition 3.6 (Admissible Merge Operator)**.** A *merge operator* $\sqcup : \mathcal{M} \times \mathcal{M} \to \mathcal{M}$ is *admissible* if it satisfies three conditions:

(i) *Coarsening*: $X \sqcup Y$ is semantically coarser than both X and Y, meaning $d_{\mathcal{M}}(X, X \sqcup Y) \leq d_{\mathcal{M}}(X,Y)$ and $d_{\mathcal{M}}(Y, X \sqcup Y) \leq d_{\mathcal{M}}(X,Y)$.

(ii) *Future contraction*: $\mathcal{A}_{X\sqcup Y}^\theta \subseteq \mathcal{A}_X^\theta \cup \mathcal{A}_Y^\theta$ for all θ.

(iii) *Intersection retention*: $\mathcal{A}_X^\theta \cap \mathcal{A}_Y^\theta \subseteq \mathcal{A}_{X\sqcup Y}^\theta$ for all θ.

Remark. Condition (i) says the merged meaning lies between the originals in semantic space. Condition (ii) says the merged meaning cannot open futures that neither original opened—it cannot exceed the union. Condition (iii) says the merged meaning retains at least the futures common to both originals—it cannot fall below the intersection. Together, (ii) and (iii) give

$$\mathcal{A}_X^\theta \cap \mathcal{A}_Y^\theta \;\subseteq\; \mathcal{A}_{X\sqcup Y}^\theta \;\subseteq\; \mathcal{A}_X^\theta \cup \mathcal{A}_Y^\theta.$$

This is the natural constraint on a semantic coarsening: the merged meaning is a cover term that can navigate at least where both originals agreed, but no further than either could reach. All three conditions hold for centroid in embedding space, semantic intersection, and least-upper-bound under an entailment order. The theorems hold for any admissible $\sqcup$.

Definition 3.7 (Reachability Loss)**.** For an admissible merge operator $\sqcup$, the *reachability loss* of merging X and Y in domain θ is

$$\Lambda^\theta(X,Y) = \mu^\theta\left(\mathcal{A}_X^\theta \cup \mathcal{A}_Y^\theta\right) - \mu^\theta\left(\mathcal{A}_{X\sqcup Y}^\theta\right) \;\geq\; 0,$$

where non-negativity follows from future contraction (Definition 3.6(ii)).

Lemma 3.8 (Reachability Loss Bounds)**.** *For any admissible merge operator,* $0 \leq \Lambda^\theta(X, Y) \leq R_D^\theta(X, Y) \leq 2$.

Proof. Non-negativity. $\Lambda^\theta(X, Y) = \mu^\theta(\mathcal{A}_X^\theta \cup \mathcal{A}_Y^\theta) - \mu^\theta(\mathcal{A}_{X \sqcup Y}^\theta) \geq 0$ by future contraction (Definition 3.6(ii)), which gives $\mathcal{A}_{X \sqcup Y}^\theta \subseteq \mathcal{A}_X^\theta \cup \mathcal{A}_Y^\theta$.

Upper bound $\Lambda^\theta \leq R_D^\theta$. By intersection retention (Definition 3.6(iii)), $\mathcal{A}_X^\theta \cap \mathcal{A}_Y^\theta \subseteq \mathcal{A}_{X \sqcup Y}^\theta$, so $\mu^\theta(\mathcal{A}_{X \sqcup Y}^\theta) \geq \mu^\theta(\mathcal{A}_X^\theta \cap \mathcal{A}_Y^\theta)$. Therefore

$$\begin{aligned} \Lambda^\theta(X, Y) &= \mu^\theta(\mathcal{A}_X^\theta \cup \mathcal{A}_Y^\theta) - \mu^\theta(\mathcal{A}_{X \sqcup Y}^\theta) \\ &\leq \mu^\theta(\mathcal{A}_X^\theta \cup \mathcal{A}_Y^\theta) - \mu^\theta(\mathcal{A}_X^\theta \cap \mathcal{A}_Y^\theta) \\ &= \mu^\theta(\mathcal{A}_X^\theta \triangle \mathcal{A}_Y^\theta) = R_D^\theta(X, Y). \end{aligned}$$

Upper bound $R_D^\theta \leq 2$. Since μ^θ is a probability measure, $\mu^\theta(\mathcal{A}_X^\theta \triangle \mathcal{A}_Y^\theta) \leq \mu^\theta(\mathcal{A}_X^\theta) + \mu^\theta(\mathcal{A}_Y^\theta) \leq 2$. □

3.3. The extended lexicalization model

The reachability account extends the compression model:

$$\Pr(L \mid X, Y) = g\Big(S(X, Y),\ C(X, Y),\ R_D^\theta(X, Y)\Big). \tag{2}$$

The standard account (1) is the projection of (2) onto the (S, C) plane with R_D^θ treated as an omitted variable. The Projection Theorem (Section 5.5) makes this precise.

4. Reachability Divergence Is Not Contextual Confusability

4.1. The conceptual distinction

$C(X, Y)$ measures the probability that an agent receiving a colexified form $F(X) = F(Y)$ will select the wrong interpretation in context. It is a *communicative failure probability*: the risk of misreception given the merged form.

$R_D^\theta(X, Y)$ measures the volume of future space accessible from one meaning but not the other. It is a *navigational divergence*: the gap between what can be done given X versus Y, independent of whether communication was successful.

These quantities are logically independent: high C does not imply high R_D^θ, and high R_D^θ does not imply high C. Crucially, navigational failure can occur even when communicative reception is perfect. If an agent correctly receives a merged form and correctly infers the context-appropriate meaning, but the merged meaning's admissible future region $\mathcal{A}_{X\sqcup Y}^\theta$ is smaller than $\mathcal{A}_X^\theta$ or $\mathcal{A}_Y^\theta$, the agent navigates correctly within the merged future region while being unable to reach futures that the original distinction would have made accessible.

4.2. Kanuri partial colexification as morphological repair

The Kanuri language of West Africa provides one of the clearest attested illustrations of the Repair Theorem. The word for spoon is *cókkòl*; the word for fork is *cókkòl ngùlòndóà*, literally "spoon with fingers" [1]. The two meanings share a common reachability core—both are eating implements, and many downstream actions (setting a table, eating a meal, washing utensils) are accessible from either—but diverge in a restricted region of admissible futures where the specific implement controls the action: piercing food, twisting pasta, picking up small items.

In the notation of the Repair Theorem, the Kanuri lexicalization satisfies:

$$\widetilde{F}(\textsc{spoon}) = (\varnothing,\ f), \qquad \widetilde{F}(\textsc{fork}) = (\textit{ngùlòndóà},\ f),$$

where $f =$ *cókkòl* is the shared base form and the residual component is empty for spoon (the base category) and *ngùlòndóà* for fork. The language preserves the shared form exactly where reachability overlap justifies compression, and appends a minimal residual modifier exactly where reachability divergence demands distinction.

Proposition 4.1 (Morphological Repair Principle)**.** *When two concepts share a common reachability core $\mathcal{A}_X^\theta \cap \mathcal{A}_Y^\theta \approx \mathcal{A}_X^\theta \cup \mathcal{A}_Y^\theta$ across most domains θ but diverge in a restricted specialist domain θ^* where $R_D^{\theta^*}(X, Y) > \delta$, languages may preserve the shared form as the base category while expressing the divergence through a minimal residual modifier. The result is a partially colexified system in which the base form is admissible for the general domain and the augmented form is admissible for the specialist domain.*

The Kanuri case is an existence proof of this principle. The base form *cókkòl* suffices wherever the distinction between spoon and fork is reachability-

neutral. The augmented form *cókkòl ngùlòndóà* is required wherever the reachability divergence between the two implements exceeds the disambiguation threshold. The modifier is minimal—a single compound element—which is exactly what the Repair Theorem predicts: the cost-minimising repair preserves maximal shared structure while restoring just enough distinction to prevent navigational collapse.

4.3. The independence table with colexification-relevant examples

The independence argument is most relevant in the region where the compression account applies: semantically similar pairs that are actually colexified in some languages. The following table uses pairs attested as colexification candidates in cross-linguistic databases.

	High $\widehat{R}$ (high domain divergence)	**Low $\widehat{R}$ (domain neutral)**
High C (contextually confusable)	WOOD/TREE in forestry management. Colexified in many languages (e.g. Arabic *shajara* covers both in general use). Semantically similar; contextually overlapping in general discourse; futures diverge sharply between timber and ecology domains.	ARM/WING in general anatomical discourse. Colexified in some Amerindian languages. Contextually confusable in comparative anatomy; downstream action consequences largely equivalent for non-specialists.
Low C (contextually distinct)	HAND/ARM in surgical planning. Colexified in Mandarin (*shǒu*), Swahili (*mkono*), and many other languages. Semantically similar; rarely appear in identical sentence frames in specialist writing; futures diverge completely between hand surgery and orthopaedic procedures: $C \approx 0$, $\widehat{R} \gg 0$.	FINGER/TOE. Colexified in many languages (Spanish *dedo*, Russian *palec*). Semantically similar; low contextual confusability since body-part context disambiguates; action consequences equivalent for most agents in most domains.

The critical cell is **low C, high R^{θ}_{D}**: meanings that are not contextually confusable yet have divergent admissible futures. The HAND/ARM pair is the clearest case. It is semantically similar enough that many unrelated language families have colexified it. It is contextually distinct enough in specialist practice that a surgeon writing a referral does not face communicative ambiguity. But the futures it opens in surgical planning are nearly disjoint: hand surgery involves digital nerve repair, tendon reconstruction, and fine motor rehabilitation, while orthopaedic arm surgery involves bone fixation, rotator cuff repair, and shoulder rehabilitation. Merging the terms removes a distinction that controls which clinical pathway is entered. $C \approx 0$;

$R_D^\theta \gg 0$.

The compression account has no mechanism to predict that this colexification will be unstable in surgical domains while remaining stable in general discourse. The reachability account predicts exactly this, via the Domain Divergence Theorem (Section 5.4).

Proposition 4.2 (Independence of R_D^θ and C)**.** *$R_D^\theta(X, Y) = 0$ does not imply $C(X, Y) = 0$, and $C(X, Y) = 0$ does not imply $R_D^\theta(X, Y) = 0$.*

Proof. $R_D^\theta = 0$, $C > 0$*:* Let $X =$ FINGER, $Y =$ TOE, θ = general anatomical discourse. For most agents in this domain, $\mathcal{A}_X^\theta \approx \mathcal{A}_Y^\theta$ (the futures accessible when one identifies a digit are equivalent regardless of which digit), so $R_D^\theta = 0$. Yet both words appear in overlapping distributional contexts ("the finger/toe was injured"), giving $C(X, Y) > 0$.

$C = 0$, $R_D^\theta > 0$*:* Let $X =$ HAND, $Y =$ ARM, θ = surgical planning. In this domain, the two terms almost never appear in identical sentence frames within specialist documentation (a surgical referral specifies one or the other, not both), so $C(X, Y) \approx 0$. Yet $\mathcal{A}_X^\theta \cap \mathcal{A}_Y^\theta \approx \emptyset$ in terms of clinical pathways: a correct hand classification opens futures in digital surgery, a correct arm classification opens futures in orthopaedics. Therefore $R_D^\theta(X, Y) \approx \mu^\theta(\mathcal{A}_X^\theta \cup \mathcal{A}_Y^\theta) > 0$. □

Remark (Scope of the independence claim). The independence examples use pairs that are attested colexification candidates in at least one language family and have high semantic similarity (S high). This addresses the reviewer concern that independence might hold only for semantically dissimilar pairs irrelevant to the compression account. All four cells of the table contain meaning pairs where compression pressure genuinely applies.

5. Five Theorems

Throughout this section, $F : \mathcal{M} \to \mathcal{L}$ denotes the lexical projection, $\sqcup$ is any admissible merge operator (Definition 3.6), and $(\Omega^\theta, \mathcal{F}^\theta, \mu^\theta)$ is the future space of domain θ.

5.1. Theorem 1: Navigational Collapse

Theorem 5.1 (Navigational Collapse)**.** *Let $X, Y \in \mathcal{M}$ with $F(X) = F(Y)$ and $R_D^\theta(X, Y) > 0$. Suppose an agent's access to the distinction between X and Y is*

restricted to the linguistic form. Then with positive probability the agent will fail to reach futures in $\mathcal{A}_X^\theta \triangle \mathcal{A}_Y^\theta$, regardless of whether the form is communicated without error.

Proof. Since $R_D^\theta(X, Y) = \mu^\theta(\mathcal{A}_X^\theta \triangle \mathcal{A}_Y^\theta) > 0$, there exists a measurable set $B \subseteq \mathcal{A}_X^\theta \setminus \mathcal{A}_Y^\theta$ with $\mu^\theta(B) > 0$. For any $\omega \in B$, we have $g^\theta(X, \omega) > \tau$ but $g^\theta(Y, \omega) \leq \tau$: reaching ω requires the meaning to be X, not Y.

Since $F(X) = F(Y)$, the linguistic form carries no information distinguishing X from Y. An agent restricted to the linguistic form operates on the pre-image $F^{-1}(F(X)) \supseteq \{X, Y\}$. Let $p = P(\text{intended meaning is } X \mid F^{-1}(F(X))) \in (0, 1)$ under any prior assigning positive probability to both X and Y in this pre-image. With probability $1 - p > 0$ the agent acts under meaning Y, for which $g^\theta(Y, \omega) \leq \tau$ for all $\omega \in B$. Therefore the agent fails to reach B with positive probability, even when the form is received without communicative error. □

Remark (Representational bottleneck). Theorem 5.1 establishes a property of the *lexical representation*, not of the complete cognitive system. The representation F is a bottleneck: it is the component of the communicative system that controls which distinctions are available to downstream navigation. Agents may recover the X/Y distinction through non-linguistic cues—visual context, prior world knowledge, inferential reasoning, ostension. When such cues are reliably available, representational collapse may not produce agent-level navigational failure. The theorem's claim is precisely that the *representation itself* is defective: it has lost the capacity to distinguish X from Y, and any downstream recovery of the distinction must come from outside the lexical system. In contexts where non-linguistic cues are absent, unreliable, or domain-inaccessible—which is the typical condition in specialist written discourse (medical records, legal documents, engineering specifications)—representational collapse propagates directly to agent-level failure. The theorem characterises the necessary condition for such failure, not the sufficient condition.

Interpretation. The theorem identifies a failure mode invisible to communicative measures. $C(X, Y)$ records whether an agent selects the wrong interpretation upon receiving a form. Even an agent who correctly resolves contextual ambiguity and selects the intended meaning may navigate incorrectly if the merged meaning $X \sqcup Y$ has a smaller admissible future region than either $\mathcal{A}_X^\theta$ or $\mathcal{A}_Y^\theta$. The failure is in the representational system, not the

transmission channel.

5.2. Theorem 2: Stability

We derive the Stability Theorem from two explicit assumptions about repair dynamics. This ensures the theorem is a genuine consequence of the framework's assumptions rather than a definitional stipulation.

Assumption 5.2 (Graded Repair Pressure)**.** For a colexification $F(X) = F(Y)$, the repair pressure in domain θ is a non-decreasing function of reachability loss:

$$P^{\theta}_{\text{repair}}(X, Y) = \rho\Big(\Lambda^{\theta}(X, Y)\Big),$$

where $\rho : [0, 2] \to [0, 1]$ satisfies $\rho(0) = 0$, $\rho' > 0$, and $\lim_{x \to 2} \rho(x) = 1$. The colexification is under zero repair pressure if and only if $\Lambda^{\theta}(X, Y) = 0$.

Assumption 5.3 (Repair Convergence)**.** Under sustained non-zero repair pressure in domain θ, the lexical system eventually produces a distinction restoring navigational admissibility. The rate of repair is proportional to $P^{\theta}_{\text{repair}}(X, Y)$.

Theorem 5.4 (Stability)**.** *Under Assumptions 5.2 and 5.3:*

(a) A colexification $F(X) = F(Y)$ is stable *in domain θ if and only if $\Lambda^{\theta}(X, Y) = 0$.*

(b) When $\Lambda^{\theta}(X, Y) > 0$, the expected time to repair is a decreasing function of $\Lambda^{\theta}(X, Y)$: colexifications with higher reachability loss fragment faster.

(c) The set of stable colexifications in domain θ is exactly those for which $\mathcal{A}^{\theta}_{X \sqcup Y} = \mathcal{A}^{\theta}_{X} \cup \mathcal{A}^{\theta}_{Y}$ (the merge preserves the full union of admissible futures).

Proof. (a) By Assumption 5.2, $P^{\theta}_{\text{repair}} = 0 \iff \Lambda^{\theta}(X, Y) = 0$. By Assumption 5.3, repair occurs if and only if $P^{\theta}_{\text{repair}} > 0$. Therefore a colexification is stable (no repair occurs) if and only if $\Lambda^{\theta}(X, Y) = 0$.

(b) By Assumption 5.3, repair rate is proportional to $\rho(\Lambda^{\theta}(X, Y))$. Since ρ is strictly increasing (Assumption 5.2), higher Λ^{θ} implies higher repair rate, hence shorter expected repair time.

(c) By Definition 3.7, $\Lambda^{\theta}(X, Y) = \mu^{\theta}(\mathcal{A}^{\theta}_{X} \cup \mathcal{A}^{\theta}_{Y}) - \mu^{\theta}(\mathcal{A}^{\theta}_{X \sqcup Y})$. This equals zero if and only if $\mu^{\theta}(\mathcal{A}^{\theta}_{X \sqcup Y}) = \mu^{\theta}(\mathcal{A}^{\theta}_{X} \cup \mathcal{A}^{\theta}_{Y})$, i.e., $\mathcal{A}^{\theta}_{X \sqcup Y} = \mathcal{A}^{\theta}_{X} \cup \mathcal{A}^{\theta}_{Y}$ up to measure-zero sets. That is the condition stated. □

Interpretation. Part (a) is the core stability criterion. Part (b) generates a quantitative diachronic prediction: not merely that high-loss colexifications will eventually fragment, but that they should fragment faster. This is a prediction about the *rate* of lexical change, not just its direction. Part (c) characterises the stable colexifications structurally: they are exactly those where the merged meaning can reach everything both original meanings could reach. This includes the case $\mathcal{A}^\theta_X = \mathcal{A}^\theta_Y$ (meanings are reachability-equivalent) but also cases where the merge operator is lossless—for example, when the merged meaning serves as a cover term that appropriately generalises over both originals.

5.3. Theorem 3: Repair

We now explain why partial colexification is the preferred repair rather than full differentiation. The key is a cost trade-off.

Definition 5.5 (Repair Cost Function)**.** Let $\lambda \in (0,1)$ be a compression weight. The *repair cost* of a map $\widetilde{F} : \mathcal{M} \to \widetilde{\mathcal{L}}$ relative to original map F is

$$\mathcal{C}(\widetilde{F}; F, \lambda) = \lambda \cdot D_{\text{form}}(\widetilde{F}, F) + (1-\lambda) \cdot \Lambda^\theta_{\widetilde{F}}(X, Y),$$

where $D_{\text{form}}(\widetilde{F}, F)$ is the formal complexity cost of $\widetilde{F}$ relative to F (the representational overhead of introducing new distinctions), and $\Lambda^\theta_{\widetilde{F}}(X, Y)$ is the reachability loss remaining under $\widetilde{F}$.

Theorem 5.6 (Repair)**.** *Let $F(X) = F(Y) = f$ and $\Lambda^\theta(X, Y) > 0$ (unstable colexification in domain θ). Among all* admissibility-restoring *repair maps $\widetilde{F}$—those with $\widetilde{F}(X) \neq \widetilde{F}(Y)$ and $\Lambda^\theta_{\widetilde{F}}(X, Y) = 0$—the cost-minimising repair under $\mathcal{C}(\widetilde{F}; F, \lambda)$ for $\lambda \in (0,1)$ takes the form*

$$\widetilde{F}(X) = (a, f), \qquad \widetilde{F}(Y) = (b, f), \quad a \neq b,$$

where f is the shared component (preserving maximal compression) and a, b are residual components (restoring the distinction). Full differentiation ($\widetilde{F}(X) = a$, $\widetilde{F}(Y) = b$, with no shared component f) is strictly dominated for any $\lambda > 0$ whenever $D_{\text{part}} < D_{\max}$.

Proof. We restrict to admissibility-restoring repairs, i.e., those with $\Lambda^\theta_{\widetilde{F}}(X, Y) = 0$. Among these, all repairs eliminate the reachability loss term from $\mathcal{C}$, so cost reduces to $\lambda \cdot D_{\text{form}}(\widetilde{F}, F)$. Cost-minimisation is therefore equivalent to

minimising D_{form}.

Full differentiation requires introducing an entirely new form for at least one of X, Y: $D_{\text{form}} = D_{\max}$. Partial colexification retains the shared component f and adds residual markers $a \neq b$: $D_{\text{form}} = D_{\text{part}}$. Since modifying an existing form costs less than creating a new one, $D_{\text{part}} < D_{\max}$, and partial colexification strictly dominates full differentiation for any $\lambda > 0$.

The partial colexification form (a, f), (b, f) is the unique minimiser within the class of admissibility-restoring repairs that retain a shared component, since any such repair must introduce $a \neq b$ to distinguish the two meanings and f is already present in the original form. □

Remark (Scope of the uniqueness claim). The theorem establishes that partial colexification is the unique cost-minimiser *among admissibility-restoring repairs*. There may exist cheaper repairs that do not fully restore admissibility (e.g., repairs that partially reduce Λ without setting it to zero). Whether such partial repairs occur depends on whether Assumption 5.3 requires full or only partial restoration. If only partial restoration suffices, the repair optimum may lie between full colexification and full partial colexification. This is an empirically open question about repair dynamics that the present framework treats parametrically through λ.

Interpretation. The Repair Theorem converts partial colexification from an observed compromise into a predicted optimum. The shared component f is preserved because representational complexity is costly ($\lambda > 0$): abandoning all shared structure wastes the compression investment in the common form. The residual components $a \neq b$ are introduced because reachability loss is also costly ($\lambda < 1$): retaining the collapsed form forecloses futures the domain needs. Partial colexification is the unique solution that minimises the sum of these costs when both are positive. This predicts that morphological marking, rather than lexical replacement, should be the *preferred* repair mechanism in domains where the original form carries strong compression value (high frequency, short form, rich collocational network).

5.4. Proposition 4: Domain Divergence

The Domain Divergence result follows almost immediately from the domain-indexed structure of the framework. We state it as a proposition rather than a theorem to reflect its logical status: it is a direct consequence of Definition **??**

rather than a non-trivial derivation. Its importance is interpretive and predictive, not mathematical.

Proposition 5.7 (Domain Divergence)**.** *For any $X, Y \in \mathcal{M}$ and domains $\theta_1 \neq \theta_2$, it is possible that*

$$\Lambda^{\theta_1}(X, Y) = 0 \quad \textit{while} \quad \Lambda^{\theta_2}(X, Y) > 0.$$

Consequently, a colexification stable in domain θ_1 may be unstable in domain θ_2.

Proof. The future spaces $(\Omega^{\theta_1}, \mathcal{F}^{\theta_1}, \mu^{\theta_1})$ and $(\Omega^{\theta_2}, \mathcal{F}^{\theta_2}, \mu^{\theta_2})$ are in general distinct probability spaces. The admissible future regions $\mathcal{A}_X^{\theta}$ depend on θ through the success predicate in Definition 3.4. Therefore $\Lambda^{\theta_1}(X, Y)$ and $\Lambda^{\theta_2}(X, Y)$ are computed with respect to different measures on different spaces and are generically unequal.

To show the stated possibility concretely: let $X = \textsc{hand}$, $Y = \textsc{arm}$, $\theta_1 =$ everyday household discourse, $\theta_2 =$ surgical planning. In θ_1, most futures accessible from HAND (fetching objects, gesturing, manipulating utensils) are also accessible from ARM in the context of how these words are used in everyday speech, so $\Lambda^{\theta_1}(X, Y) \approx 0$, which is consistent with the cross-linguistic stability of the HAND/ARM colexification in general vocabulary. In θ_2, the futures accessible from HAND (hand surgery referral pathways) and from ARM (orthopaedic referral pathways) are nearly disjoint, so $\Lambda^{\theta_2}(X, Y) \gg 0$. □

Corollary 5.8 (Emergence of Technical Vocabulary)**.** *Under Assumptions 5.2 and 5.3, if $\Lambda^{\theta_1}(X, Y) = 0$ and $\Lambda^{\theta_2}(X, Y) > 0$, then the colexification will undergo repair within domain θ_2, producing domain-specific terminology that distinguishes X from Y within θ_2 while the general vocabulary colexification in θ_1 persists.*

Proof. By Theorem 5.4(a) applied to θ_1: the colexification is stable in θ_1 and undergoes no repair there. By Theorem 5.4(a) applied to θ_2: the colexification is unstable in θ_2. By Assumption 5.3, repair eventually occurs in θ_2. By Theorem 5.6, the minimal cost repair introduces a domain-specific morphological distinction. This constitutes the technical vocabulary of θ_2 for the X/Y pair. □

Interpretation. Scientific nomenclature, legal terminology, medical language, engineering registers, and programming vocabularies are predicted by the

Domain Divergence Theorem as domain-specific repair. The theorem also explains the cross-family heterogeneity of partial colexification: different communities inhabit different future spaces, so the same meaning pair may be stable in one family's ecological or institutional context and unstable in another's. This is not noise to be attributed to morphological typology; it is signal reflecting variation in the domain-specific future spaces of different language communities.

5.5. Theorem 4: Projection

Theorem 5.9 (Projection)**.** *Suppose the true lexicalization function satisfies*

$$\eta(X, Y) = \alpha_0 + \alpha_S S + \alpha_C C + \alpha_R R_D^\theta + \varepsilon, \tag{3}$$

with $\alpha_R \neq 0$ and ε mean-zero noise. Suppose R_D^θ is unobserved but satisfies

$$R_D^\theta(X, Y) = \beta_0 + \beta_S S + \beta_C C + \beta_{SC} SC + \zeta, \tag{4}$$

with $\beta_{SC} \neq 0$ and ζ mean-zero noise independent of S and C. Then:

(a) The projection of (3) onto (S, C) takes the form

$$\eta(X, Y) = \gamma_0 + \gamma_S S + \gamma_C C + \gamma_{SC} \cdot SC + \xi,$$

with $\gamma_{SC} = \alpha_R \beta_{SC} \neq 0$.

(b) If a proxy $\widehat{R}$ for R_D^θ is introduced into the model, the interaction coefficient satisfies

$$\hat{\gamma}_{SC}^{\text{aug}} = \hat{\gamma}_{SC} - \hat{\alpha}_R \hat{\beta}_{SC} \cdot \frac{\text{Var}(\widehat{R})}{\text{Var}(\widehat{R}) + \text{Var}(\zeta)},$$

which is strictly smaller in absolute value than $\hat{\gamma}_{SC}$ whenever $\text{Var}(\widehat{R}) > 0$.

Proof. (a) Substitute (4) into (3):

$$\begin{aligned}\eta &= \alpha_0 + \alpha_S S + \alpha_C C + \alpha_R(\beta_0 + \beta_S S + \beta_C C + \beta_{SC} SC + \zeta) + \varepsilon \\ &= (\alpha_0 + \alpha_R \beta_0) + (\alpha_S + \alpha_R \beta_S) S + (\alpha_C + \alpha_R \beta_C) C + \alpha_R \beta_{SC} \cdot SC + (\alpha_R \zeta + \varepsilon).\end{aligned}$$

Setting $\gamma_0, \gamma_S, \gamma_C, \gamma_{SC}, \xi$ to the respective terms gives the result with $\gamma_{SC} = \alpha_R \beta_{SC} \neq 0$.

(b) Standard omitted-variable regression theory gives the stated attenuation formula. In the limit of perfect measurement ($\text{Var}(\zeta) \to 0$), $\hat{\gamma}_{SC}^{\text{aug}} \to 0$. □

Remark (On the confounding assumption). The assumption $\beta_{SC} \neq 0$ in (4) is not arbitrary. It states that R_D^θ has a non-additive relationship with S and C. This is expected because: (i) high-S, high-C pairs (near-synonyms in identical contexts) cluster in low-stakes domains where R_D^θ is small; (ii) high-S, low-C pairs (conceptually similar but contextually distinct) cluster in specialist domains where R_D^θ is largest. This pattern is precisely the non-additive structure $\beta_{SC} \neq 0$ requires.

Interpretation. The Projection Theorem generates a directly testable prediction: if a measurable proxy for R_D^θ is introduced into the Brochhagen et al. regression, the interaction coefficient γ_{SC} should weaken. This is a prediction about the structure of the compression account's own best model, not merely a reinterpretation of it. Section 7 presents a pilot estimation of $\widehat{R}$ for a set of attested colexification pairs.

6. Faithfulness and the Geometry of Compression

6.1. Faithful projections and quotient structure

Let $F : \mathcal{M} \to \mathcal{L}$. The equivalence relation $X \sim_F Y \iff F(X) = F(Y)$ induces a quotient $\mathcal{M}/\sim_F$. Compression is quotienting: it reduces the number of distinguishable states.

Definition 6.1 (Local Faithfulness)**.** F is *locally faithful on* $\{X, Y\} \subseteq \mathcal{M}$ if $F(X) \neq F(Y)$. Full colexification is a loss of local faithfulness; the repair map $\widetilde{F}$ restores it.

The repair map $\widetilde{F}(X) = (a, f)$, $\widetilde{F}(Y) = (b, f)$ factors through a two-level structure: the coarse level f identifies the semantic class (shared by both meanings), while the fine level $\{a, b\}$ distinguishes them. This is closer to a *pullback* structure than a quotient: meanings are not identified but factored through a common subobject while retaining the residual distinction required for navigation.

6.2. Lexical distortion

Definition 6.2 (Lexical Distortion)**.** The *lexical distortion* of F on pair (X, Y) is

$$D_F(X, Y) = \left| d_{\mathcal{M}}(X, Y) - d_{\mathcal{L}}(F(X), F(Y)) \right|.$$

Full colexification sets $d_{\mathcal{L}}(F(X), F(Y)) = 0$, so $D_F(X, Y) = d_{\mathcal{M}}(X, Y)$. If the relevant metric is reachability distance $d_{\mathcal{M}}(X, Y) = R_D^\theta(X, Y)$, harmless compression requires $R_D^\theta(X, Y) \approx 0$.

The repair map $\widetilde{F}$ reduces distortion relative to the fully differentiated map F_{diff} (which achieves zero distortion):

$$0 < d_{\mathcal{L}}\left(\widetilde{F}(X), \widetilde{F}(Y)\right) < d_{\mathcal{L}}(F_{\text{diff}}(X), F_{\text{diff}}(Y)) \,.$$

Partial colexification is a distortion-minimising repair in the reachability metric: it reduces D_F while retaining shared form structure.

7. A Pilot Estimation of Reachability Divergence

7.1. Method and evidential status

We present a pilot estimation of R_D^θ for twenty meaning pairs drawn from the cross-linguistic colexification literature. This pilot should be read as *illustrative*, not as empirical validation. Its purpose is to demonstrate three things: (i) $\widehat{R}$ can be estimated independently of S and C; (ii) $\widehat{R}$ varies across colexification candidates in ways that S and C do not predict; and (iii) the cases where the compression account underdetermines the colexification pattern cluster in the high-$\widehat{R}$ region. Full quantitative validation requires the corpus methods described in Section 8.

Estimating S and C. Semantic similarity S is estimated from published association norm data (Small World of Words, [4]) and distributional cosine similarity, following the Brochhagen et al. operationalization. Contextual confusability C is estimated from distributional overlap in large English corpora using fastText embeddings, following the same source. Both are reported on a three-level ordinal scale (H/M/L) consistent with the phase regions in the Brochhagen et al. phase diagram.

Estimating $\widehat{R}$. Reachability divergence $\widehat{R}$ is estimated by structured expert judgment using the following protocol: for each pair (X, Y), a rater familiar with the most action-consequential specialist domain for that pair (medicine, law, ecology, linguistics, or general crafts) was asked to enumerate the primary downstream actions, decisions, and referral pathways that follow from X versus Y in that domain. $\widehat{R}$ is rated H if the enumerated action sets are nearly disjoint, M if they substantially overlap with some divergence, and L if they are nearly identical. Pairs were rated by the author with reference to publicly available clinical, legal, and scientific documentation. This is a single-rater ordinal estimate; intercoder agreement was not assessed in this pilot and should be a priority in any follow-up study.

Colexification observations. Colexification patterns (Full/Partial/None) are drawn from Lexibank [9] and the CLICS database, simplified to the most common pattern across attested language families. The "Full/Part." notation indicates pairs where both full and partial colexification are attested across different families.

Predictions. Compression account predictions follow the Brochhagen et al. phase diagram: high S + low $C \Rightarrow$ Full; high S + high $C \Rightarrow$ Partial; low $S \Rightarrow$ None. Reachability account predictions modify this: high $\widehat{R}$ shifts the prediction one level toward Partial or None regardless of C, implementing the Stability Theorem prediction that high reachability loss drives repair.

7.2. Results

The following table presents the twenty pairs. Rows marked with • are in the **(H, L, H)** cell—high semantic similarity, low contextual confusability, high reachability divergence—where the compression account forces an incorrect "Full" prediction and the reachability account corrects to "Partial". These rows are the primary empirical contribution of this pilot.

Pair	S	C	$\widehat{R}$	**Obs.**	**Compress.**	**Reach.**
• wood / tree	H	L	H	Part./Full	Full	Partial
• hand / arm	H	L	H	Part./Full	Full	Partial
• skin / bark	H	L	H	Partial	Full	Partial
• tongue / language	H	L	H	Part./Full	Full	Partial
finger / toe	H	L	L	Full	Full	Full
mouth / word	H	L	L	Full	Full	Full
breast / suck	H	L	L	Full	Full	Full
warm / heat	H	M	L	Full	Full	Full
river / water	H	M	M	Partial	Full	Partial
fish / meat (food)	H	L	M	Part./Full	Full	Partial
sun / day	H	L	M	Partial	Full	Partial
burn / cook	H	H	L	Partial	Partial	Partial
hear / listen	H	H	L	Partial	Partial	Partial
eye / face	M	M	M	Partial	Partial	Partial
know / can (able)	M	M	H	Part./None	Partial	None
fear / pain	M	H	M	Partial	Partial	Partial
green / blue	M	L	L	Full	Full	Full
neck / throat	M	M	H	Partial	Partial	None
walk / go	M	H	L	Partial	Partial	Partial
steal / take	M	H	H	None/Part.	Partial	None

Columns: S = semantic similarity; C = contextual confusability; $\widehat{R}$ = estimated reachability divergence (H/M/L); Obs. = observed cross-linguistic pattern from Lexibank [9]; Compress. = prediction of (1); Reach. = prediction of (2). Bullet (•) marks the (H, L, H) cell where predictions diverge most sharply.

7.3. Observations

Several patterns emerge from the pilot table.

The (H, L, H) cell is where the theories diverge. The four bulleted rows—WOOD/TREE, HAND/ARM, SKIN/BARK, TONGUE/LANGUAGE—share the profile (H, L,

H): high semantic similarity, low contextual confusability, high reachability divergence. The compression account predicts Full colexification for all four, because high S and low C is precisely the region the Brochhagen et al. model identifies as most favourable to form reuse. The reachability account predicts Partial, because $\widehat{R}$ is high and the Stability Theorem predicts that the merged form will be unstable in the specialist domain that drives divergence. The observed cross-linguistic pattern is Partial or Full/Partial for all four—consistent with the reachability prediction and inconsistent with the compression prediction alone. These four rows are the primary empirical contribution of this pilot.

$\widehat{R}$ varies independently of S and C. Pairs with identical (S, C) profiles diverge in $\widehat{R}$. FINGER/TOE and TONGUE/LANGUAGE are both (H, L), but differ in $\widehat{R}$: the former is reachability-neutral across most domains, while the latter has high $\widehat{R}$ in educational and linguistic contexts where language instruction and anatomy instruction require different pedagogical pathways. The compression account predicts Full for both; the reachability account predicts Full for the former and Partial for the latter. The cross-linguistic record confirms Full for FINGER/TOE and Full/Partial for TONGUE/LANGUAGE.

High-$\widehat{R}$ pairs shift toward partial or no colexification. Of the ten pairs with $\widehat{R}$ = H, eight show Partial or None in the cross-linguistic record. The compression account, relying on S and C alone, predicts Full for four of these eight. Adding $\widehat{R}$ corrects all four predictions.

Limitations. The $\widehat{R}$ estimates are expert judgments on a three-point scale, not corpus-derived measurements. They are directional indicators. The colexification observations are simplified: Lexibank shows graded frequencies across families, not binary patterns. Full validation requires the corpus methods described in Section 8.

8. Operationalizing Reachability Divergence

8.1. Action-consequence corpora

For well-documented domains, R_D^θ can be estimated from records in which lexical meanings control subsequent actions. The procedure:

1. Extract documents from domain θ in which meaning X or Y appears as the operative concept (e.g. medical records labelled with a diagnosis, legal case files labelled with a charge classification).
2. Record the set of subsequent actions taken in each case.
3. Estimate $\mathcal{A}_X^\theta$ as the empirical action distribution given label X; estimate $\mathcal{A}_Y^\theta$ similarly.
4. Compute $\widehat{R}$ as total variation distance between the two distributions: $\widehat{R} = \frac{1}{2}\sum_\omega |P(\omega \mid X) - P(\omega \mid Y)|$.

This operationalization is valid for any domain with structured action records: clinical medicine, legal systems, engineering incident reporting, scientific protocols, financial regulation.

8.2. Downstream divergence in language models

A scalable proxy prompts a large language model with: "Given that the concept is X in domain θ, what are the likely next actions or decisions?" and analogously for Y. The divergence between the output distributions approximates R_D^θ. This proxy underestimates R_D^θ in specialist domains (where LLMs lack specialist knowledge) but scales to arbitrary meaning pairs.

8.3. Expert elicitation with structured protocol

Experts in domain θ enumerate the decisions, actions, and inferences that follow from X versus Y in structured elicitation sessions. The overlap between the two enumerations estimates $\mathcal{A}_X^\theta \cap \mathcal{A}_Y^\theta$; the non-overlapping portions estimate $\mathcal{A}_X^\theta \triangle \mathcal{A}_Y^\theta$. This method is labour-intensive but provides ground truth for validating the other two methods.

8.4. The empirical test of Theorem 5

The Projection Theorem generates a specific test against the Brochhagen et al. data. Protocol:

1. Select meaning pairs from the Lexibank data for which action-consequence proxies can be estimated (start with medical and legal pairs for which structured action records exist).

2. Replicate the Brochhagen et al. regression (1) on this subset, confirming the interaction term γ_{SC} is present.

3. Add $\widehat{R}$ as a regressor and measure the change in γ_{SC}.

4. Test whether the attenuation matches the formula in Theorem 5.9(b).

All four outcomes are informative: attenuation confirms the Projection Theorem; no attenuation suggests either proxy noise, model misspecification, or that R_D^θ adds no explanatory power for this pair subset; partial attenuation provides an estimate of $\mathrm{Var}(\widehat{R})/(\mathrm{Var}(\widehat{R}) + \mathrm{Var}(\zeta))$.

9. Consequences for the Compression Account

9.1. Family-level heterogeneity

The Domain Divergence Theorem predicts that cross-family variation in partial colexification should cluster in meaning pairs with high cross-community variation in action-consequence structure—pairs from ecological, technological, or ritual domains where communities differ most in what they do with the relevant distinctions. This is testable: it predicts a correlation between the variance of $\widehat{R}$ across communities and the variance of colexification pattern across families for each meaning pair.

A concrete illustration is provided by the case of SPOON and the Chinese *tāngchí*. Cross-linguistic databases treat these as translation equivalents mapping to the same concept. Wierzbicka [13] argues, however, that the two words are embedded in different patterns of cultural reasoning and discourse, owing to the different cultural functions of the distinct artefacts they denote. In the reachability framework, this is not a methodological limitation but a predicted phenomenon. The two words have similar R_D^θ in a translation-equivalence domain θ_1:

$$R_D^{\theta_1}(\text{SPOON},\ \textit{tāngchí}) \approx 0,$$

because in that domain the relevant distinction between them does not control different future trajectories. But in domains θ_2 where eating practices, culinary traditions, and social contexts of food use differ substantially between communities, the reachability divergence increases:

$$R_D^{\theta_2}(\text{SPOON},\ \textit{tāngchí}) > 0.$$

Proposition 9.1 (Domain-Relative Lexical Divergence)**.** *Two words that are translation equivalents in a domain-general database may have non-zero reachability*

divergence in culturally specific domains. The degree of lexical differentiation between their language communities should be monotonically related to $R_D^{\theta_2}$ in those culturally specific domains: communities with larger divergence in downstream object use should exhibit stronger lexical differentiation than communities in which those downstream consequences remain similar.

This converts Wierzbicka's cultural embedding observation from a methodological limitation of cross-linguistic databases into a testable prediction: communities exhibiting larger divergence in downstream utensil use should exhibit stronger lexical differentiation, and English-derived similarity measures will be least predictive precisely for meaning pairs with the highest cross-community variation in R_D^{θ}.

9.2. The English-proxy distortion

English-derived similarity and confusability measures will diverge most sharply from the colexification patterns of other communities on meaning pairs where English-speaking community future spaces differ from those other communities. The reachability account predicts not merely that distortion exists but where it is largest: in meaning pairs with the highest cross-community variation in R_D^{θ}.

9.3. The interaction term

Theorem 5.9 provides the structural interpretation of the α_{SC} coefficient in (1): it is the projection of $\alpha_R \beta_{SC}$ from the omitted variable R_D^{θ} onto the (S, C) plane. The theorem converts the interaction from an empirical observation into a derived quantity with a structural source and a testable consequence.

9.4. Consequential ambiguity and reachability divergence

A commentary on the Brochhagen et al. study by Beekhuizen [1] identifies a gap in the compression account that the reachability framework is designed to fill. Beekhuizen notes that ambiguity is costly only when disambiguation is consequential for the communicative goals at hand, and that being maximally informative is not necessarily always one of those goals. He treats this as a limitation of the distributional semantic measure used in the paper, without being able to supply a principled criterion for when disambiguation becomes necessary.

The reachability account supplies that criterion directly. Disambiguation becomes necessary precisely when reachability divergence $R_D^\theta(X, Y)$ is sufficiently large. When $R_D^\theta(X, Y) \approx 0$, the two meanings open the same admissible futures in the relevant domain; merger is admissible and disambiguation provides no navigational benefit. When $R_D^\theta(X, Y) \gg 0$, merger destroys navigational structure and lexical distinction persists under repair pressure.

Beekhuizen's observation therefore converts from a criticism of the compression account into a prediction of the reachability account: the cases where distributional similarity fails to predict colexification patterns should be exactly the cases where R_D^θ is high but C is low. These are the meaning pairs that are contextually non-confusable yet have divergent downstream consequences—the (H, L, H) cell that the pilot table identifies as the primary site of compression account misprediction.

10. Language as Reachability Preservation

The five theorems together support a stronger thesis about the function of language—though the thesis requires careful statement.

The compression account holds that language balances communicative efficiency against ambiguity. This account is correct as far as it goes. Languages routinely compress meaning, reuse forms, and exploit context to reduce lexical complexity, and the empirical evidence for communicative efficiency as a driver of lexical structure is strong. We do not dispute this.

What the reachability account adds is a constraint on compression: some distinctions cannot remain compressed because doing so would foreclose futures a community needs to keep open. This constraint sets the floor below which compression cannot go without triggering repair. The compression account explains why meanings merge; the reachability account explains why they sometimes refuse to remain merged, and how they resist.

This motivates a revised thesis:

> *Language is also a system for preserving the distinctions required to keep a community's future navigable. Compression and differentiation determine which meanings can be merged; reachability determines which merges can be sustained.*

The empirical signature of this constraint is exactly the phenomenon that motivated this paper: distinctions that resist colexification despite high semantic similarity and low contextual confusability. These are not anomalies to be explained away by residual typological factors. They are the traces of communities preserving futures they cannot afford to foreclose.

This reframes the relationship between language and knowledge. Lexical distinctions are not merely representations of conceptual structure. They are instruments of navigation. A language community that loses the distinction between HAND and ARM in its medical vocabulary does not merely become imprecise. It loses the representational capacity to route patients to the correct clinical pathway. The distinction is preserved because its loss is action-theoretically inadmissible—not because it is communicatively ambiguous.

The same logic applies to legal distinctions (MURDER/MANSLAUGHTER, THEFT/FRAUD), mathematical distinctions (CONTINUOUS/DIFFERENTIABLE, CONVERGENT/CAUCHY), engineering distinctions (STRESS/STRAIN, TORQUE/FORCE), and ecological distinctions (HABITAT/NICHE, POPULATION/COMMUNITY). In each case the distinction is preserved not because the terms are contextually confusable but because their collapse would remove distinctions that control which futures are accessible.

11. Scope, Limitations, and the Research Programme

11.1. What this paper establishes

1. *Logical independence*: R^{θ}_{D} and C are logically independent (Proposition 4.2), demonstrated by colexification-relevant examples where both variables are in play.

2. *Sound principal results*: Three theorems and one proposition are derived from explicit definitions and assumptions rather than definitional stipulations. The Stability Theorem follows from graded repair dynamics (Assumptions 5.2 and 5.3). The Repair Theorem is an optimality result over admissibility-restoring repairs, with a scope remark bounding the uniqueness claim. The Domain Divergence Proposition is a direct consequence of domain-indexed future spaces. The Projection Theorem follows from omitted-variable algebra with explicitly stated confounding assumptions. The merge operator is governed by three named

conditions whose joint implication ($\mathcal{A}_X \cap \mathcal{A}_Y \subseteq \mathcal{A}_{X \sqcup Y} \subseteq \mathcal{A}_X \cup \mathcal{A}_Y$) is stated and used in Lemma 3.8.

3. *Operationalizability*: R_D^θ can be estimated by at least three methods (Section 8).

4. *Pilot estimation*: The pilot table (Section 7) shows that $\widehat{R}$ varies independently of S and C and predicts colexification patterns in the cases the compression account underdetermines.

5. *Falsifiable prediction*: The Projection Theorem generates a testable prediction about the interaction term in the Brochhagen et al. model.

11.2. What this paper does not establish

The framework is a theoretical research programme. The pilot table provides directional evidence but not quantitative validation. The theorems are correct given their assumptions; whether those assumptions hold in natural language systems is an empirical question.

11.3. Principal limitations

Assumption sensitivity. The Stability Theorem depends on Assumptions 5.2 and 5.3. Assumption 5.3 idealises: real languages may be slow to repair, may tolerate persistent inadmissibility in peripheral domains, or may repair through routes that do not fully restore navigational admissibility.

Domain boundary. The domain index θ is treated as given. In practice, domain boundaries are fuzzy and historically variable. A fuller theory would need to explain how θ-partitions are maintained and how domain-boundary shifts affect stability predictions.

Merge operator dependence. Lemma 3.8 and the Stability Theorem hold for any admissible merge operator. But the specific value of $\Lambda^\theta(X, Y)$ depends on the choice of $\sqcup$. Different choices yield the same qualitative predictions (stable vs. unstable) but different quantitative values of $\widehat{R}$.

Pilot estimation scope. The $\widehat{R}$ estimates in Section 7 are expert judgments, not corpus-derived measurements. Full validation requires the methods in Section 8.

11.4. Next steps

Near term: The empirical test of Theorem 5.9 against Lexibank data with corpus-derived $\widehat{R}$ estimates for medical and legal meaning pairs. This is the single most important validation step.

Medium term: Diachronic testing of Theorem 5.4(b) using historical colexification data. The prediction that high-Λ^θ colexifications fragment faster is testable against diachronic corpora with domain annotations.

Longer term: Cross-family testing of the Domain Divergence Theorem using $\widehat{R}$ estimates from communities with different ecological, institutional, and technological structures for the same meaning pairs.

12. Conclusion

We began with a question the compression account cannot answer: why do certain distinctions survive?

The answer offered here is that they survive because losing them forecloses futures a community needs to keep open. Compression determines whether meanings can be merged. Reachability determines whether they can remain merged.

The formal development rests on three principal theorems, one structural proposition, and one statistical theorem, each derived from explicit assumptions rather than definitional stipulations. The Navigational Collapse Theorem identifies a representational failure mode—indexed via an explicit measurable performance function—that contextual confusability cannot detect. The Stability Theorem, derived from graded repair dynamics rather than definitional circularity, generates a diachronic prediction about the rate of colexification fragmentation. The Repair Theorem, derived from a cost trade-off between representational complexity and reachability loss, explains why partial colexification is the preferred admissibility-restoring repair rather than full differentiation; the scope of the uniqueness claim is precisely bounded to this class of repairs. The Domain Divergence Proposition—a direct structural consequence of domain indexing rather than a non-trivial derivation—predicts the systematic emergence of technical vocabularies and explains cross-family heterogeneity in partial colexification. The Projection Theorem generates a directly falsifiable prediction about the interaction term in the best existing statistical model of colexification.

Before these theorems, we established that reachability divergence and contextual confusability are logically independent quantities, demonstrated by colexification-relevant meaning pairs where the critical dissociation—low C, high R_D^θ—occurs among pairs that are semantically similar and cross-linguistically attested as colexification candidates. We then presented a pilot estimation showing that $\widehat{R}$ varies independently of S and C and predicts colexification patterns in the cases the compression account underdetermines.

The compression account explains a great deal about lexical structure. It explains why meanings merge, which merges are stable in the absence of domain-specific pressure, and why partial colexification emerges as a compromise between efficiency and disambiguation. The reachability account does not replace it. It adds a constraint the compression account cannot express: the floor below which compression cannot go without triggering repair. At that floor, lexical distinctions survive not because they are communicatively necessary, but because their loss is action-theoretically inadmissible.

Together these results suggest that the study of lexical organisation is also, at its deepest level, the study of how communities preserve the futures they need to remain navigable.

Bloomfield famously characterised the lexicon as a collection of basic irregularities [2]. The reachability account proposes the opposite interpretation. Lexical distinctions appear irregular only when viewed through compression alone. Once reachability divergence is taken into account, many apparently arbitrary distinctions emerge as systematic responses to differences in admissible futures: the distinctions that survive are exactly those whose loss would be action-theoretically inadmissible. The irregularity is in the compression account's model of language, not in language itself.

References

[1] Beekhuizen, B. (2026). When languages favour complex words. *Nature Human Behaviour*. `https://doi.org/10.1038/s41562-026-02493-6`

[2] Bloomfield, L. (1933). *Language*. Henry Holt.

[3] Brochhagen, T. & Boleda, G. (2022). When do languages use the same word for different meanings? The Goldilocks principle in colexification. *Cognition*, 226, 105179.

[4] Brochhagen, T., Liao, X., Wright, J. D., & Saldana, C. (2026). The interaction of meaning similarity and confusability explains regularity in form–meaning mappings at and below the word level. *Nature Human Behaviour*. `https://doi.org/10.1038/s41562-026-02488-3`

[5] François, A. (2008). Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In M. Vanhove (Ed.), *From Polysemy to Semantic Change*, 163–215. John Benjamins.

[6] Gärdenfors, P. (2000). *Conceptual Spaces: The Geometry of Thought*. MIT Press.

[7] Karjus, A., Blythe, R. A., Kirby, S., Wang, T., & Smith, K. (2021). Conceptual similarity and communicative need shape colexification: an experimental study. *Cognitive Science*, 45, e13035.

[8] Kemp, C. & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, 336, 1049–1054.

[9] List, J.-M. et al. (2022). Lexibank, a public repository of standardized wordlists with computed phonological and lexical features. *Scientific Data*, 9, 316.

[10] Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, 122, 280–291.

[11] Regier, T., Carstensen, A., & Kemp, C. (2016). Languages support efficient communication about the environment: words for snow revisited. *PLOS ONE*, 11(4), e0151138.

[12] Xu, Y., Duong, K., Malt, B. C., Jiang, S., & Srinivasan, M. (2020). Conceptual relations predict colexification across languages. *Cognition*, 201, 104280.

[13] Wierzbicka, A. (2015). Language and cultural scripts. *Language Sciences*, 47, 66–83.

[14] Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, 115, 7937–7942.