Ento-Linguistic Domains: Language, Ambiguity, and Scientific Communication in Entomology

How Terminology Networks Shape Understanding of Insect Biology (And Vice-Versa)

Daniel Ari Friedman

February 14, 2026

Contents

1 Abstract

Scientific language does not merely describe biological phenomena; it constitutes the **generative model** through which researchers parse complex systems. In entomology, anthropomorphic terminology—"queen," "worker," "caste"—imposes implicit top-down control structures on what are fundamentally **stigmergic**, self-organizing systems. This linguistic framing distorts the causal modeling of insect behavior, obscuring the mechanisms of distributed agency.

We present a mixed-methodology framework integrating computational text analysis with **Active Inference** and **Complex Systems Theory** to investigate how language use in ant research creates ambiguity and inappropriate framing. Our pipeline extracts domain-specific terminology, constructs co-occurrence networks, and computes ambiguity scores across six Ento-Linguistic domains. Analysis of a comprehensive corpus of **3,253 publications** (comprising 473,322 tokens) identifies **9,315 terms**, revealing terminology networks with strong domain clustering and cross-domain bridging.

Computational results show that the Power & Labor domain consistently yields the highest ambiguity scores: hub terms such as "caste" and "queen" impose hierarchical control layers absent from the biology they purport to describe. Across domains, conceptual networks reveal that "individuality" spans multiple biological scales, blurring the formal boundaries (**Markov Blankets**) required for rigorous systems modeling. Behavioural descriptions routinely transform fluid, policy-dependent processes into categorical identities.

These findings extend beyond entomology to scientific communication generally, where language shapes research questions, methodological choices, and interpretive frameworks. We propose four evidence-based meta-standards—**Clarity**, **Appropriateness**, **Consistency**, and **Evolvability** (CACE)—as a protocol for **lexical engineering**. The accompanying open-source computational pipeline provides reproducible analytical tools for rigorous terminological stewardship.

2 Introduction

2.1 Linguistic Priors and Generative Models

Scientific inquiry is a process of **active inference**, where researchers refine generative models to minimize surprise about biological observations Friston [2010]. Language acts as the **hyper-prior** for these models: it constrains the hypothesis space before data collection even begins. When entomologists employ terms like "queen" or "caste," they are not merely labeling phenomena; they are importing a high-precision prior from human social systems into their model of insect biology. If this prior is structurally misaligned with the target system—for instance, assuming top-down control in a stigmergic network—the resulting model will necessarily suffer from high variational free energy, manifesting as persistent anomalies and "epicycles" in theoretical explanations Kuhn [1996], Clark [2013]. Recent formal work by Friedman et al. [2021] demonstrates that ant colonies can be modeled as ensembles of *active inferants*—individual agents performing Bayesian inference over local states via chemical stigmergy—without any centralized controller; yet the dominant vocabulary of the field continues to presuppose one.

Our work examines this epistemic risk through systematic analysis of *Ento-Linguistic domains*: specific areas where linguistic priors obscure the causal architecture of ant systems.

2.2 Motivation: Minimizing Model Misspecification

The drive for clarity is not merely a stylistic preference but a requirement for model integrity. As Keller [1995] argued, the language of science constitutes the cognitive scaffolding of research. In the framework of Active Inference, an undefined or metaphor-laden term introduces **irreducible uncertainty** (entropy) into the scientific communication channel.

The present moment demands this formalization. Recent cognitive science emphasizes that metaphor is a mechanism of predictive processing Steen [2017]. Rather than perpetuating "legacy code" in our linguistic ontology, researchers must critically assess whether their terminological priors minimize or maximize the complexity of their biological models.

A paradigmatic example is the "slave-making" debate. Herbers [2006] showed that the term "slave" naturalizes a human institution while obscuring the biological mechanism of **social parasitism**. In formal terms, the "slave" metaphor implies a conscious coercion policy, whereas the replacement term "dulosis" correctly identifies the phenomenon as a breakdown in nestmate recognition signals (a failure of the Markov Blanket's security filter). Reform, therefore, is not just about ethics; it is about restoring the causal fidelity of the scientific model.

2.3 The Challenge of Terminological Reform

A common objection to improving scientific language is that changing terminology creates disconnection from existing literature. If entomologists abandon terms like "caste" or "slave," how would researchers locate papers on task performance or social parasitism?

This objection, however, inadvertently strengthens the case for reform. Retaining problematic terminology for convenience perpetuates and compounds conceptual distortions rather than addressing them Herbers [2006]. The appropriate response is to work systematically toward clearer communication while developing the necessary infrastructure for literature synthesis—restructuring information from existing sources and establishing new meta-standards for scientific discourse. Recent community-level momentum confirms this trajectory: discussions at the MirMeco 2023 International Ant Meeting Laciny [2024] and the Entomological Society of America's Better Common Names Project Entomological Society of America [2024] demonstrate that the professional community increasingly shares these concerns.

2.4 Ento-Linguistic Domains: A Framework for Analysis

We organize our analysis around six domains where entomological language creates ambiguity or imports unjustified assumptions. Each domain isolates a distinct category of terminological friction between human conceptual frameworks and ant biology.

Unit of Individuality. The definition of a biological individual is formally equivalent to the specification of a **Markov Blanket**—the statistical boundary separating internal states from external states Friston [2013]. Terms like "colony," "superorganism," and "individual" confuse these boundaries, creating models where the relevant unit of agency is undefined.

Behavior and Identity. Task performance in ants is a fluid process of **policy selection** based on local cues Gordon [2010]. However, terminology transforms these transient policies into categorical identities ("forager," "nurse"). This effectively hard-codes a fixed-role prior into the model, obscuring the plasticity and Bayesian updating that actually drives task allocation.

Power & Labor. Terms like "queen," "worker," and "caste" impose a hierarchical control architecture on a system that is fundamentally **stigmergic**. This introduces a causal error: it attributes colony-level regulation to centralized agency (the queen) rather than distributed feedback loops, fundamentally misrepresenting the system's control theory.

Sex & Reproduction. Terms like "sex determination" and "sex differentiation" carry implicit assumptions about binary systems that may not map onto ant reproductive biology, where haplodiploidy creates fundamentally different patterns Chandra et al. [2021].

Kin & Relatedness. Human kinship terminology, grounded in bilateral relatedness, creates systematic friction when applied to ant societies structured by haplodiploidy. In haplodiploid species, full sisters share an average relatedness coefficient of $r = 0.75$—higher than the mother–daughter coefficient of $r = 0.5$—a fundamental asymmetry absent from human kinship models. Terms such as "sister," "mother," and "family" obscure this asymmetry and its profound consequences for kin selection theory Chandra et al. [2021].

Economics. Economic metaphors—markets, trade, investment, cost-benefit—shape analysis of ant foraging, resource distribution, and colony-level resource management. This domain investigates how transactional frameworks constrain biological interpretation by conflating proximate energetic expenditure with ultimate fitness costs, importing assumptions of rational optimisation from microeconomics into systems that operate through evolved heuristics rather than deliberative calculation.

2.5 Research Approach

This work employs a mixed-methodology framework combining computational text analysis with theoretical discourse examination. The computational component processes a **comprehensive literature corpus of 3,253 publications** using automated term extraction, co-occurrence network construction, and ambiguity scoring to identify statistical patterns in language use. The theoretical component, informed by Foucault's archaeological method [1972], conceptual metaphor theory Lakoff and Johnson [1980], and Gordon's [2023] ecological framework for collective behaviour, examines how these patterns reflect deeper conceptual structures. All data and analysis code are reproducible and available for validation and extension.

2.6 Terminology Network Visualization

To illustrate the framework's output, Figure 1 shows how terms cluster around the six Ento-Linguistic domains and form cross-domain networks of meaning; detailed quantitative analysis follows in Section 4.

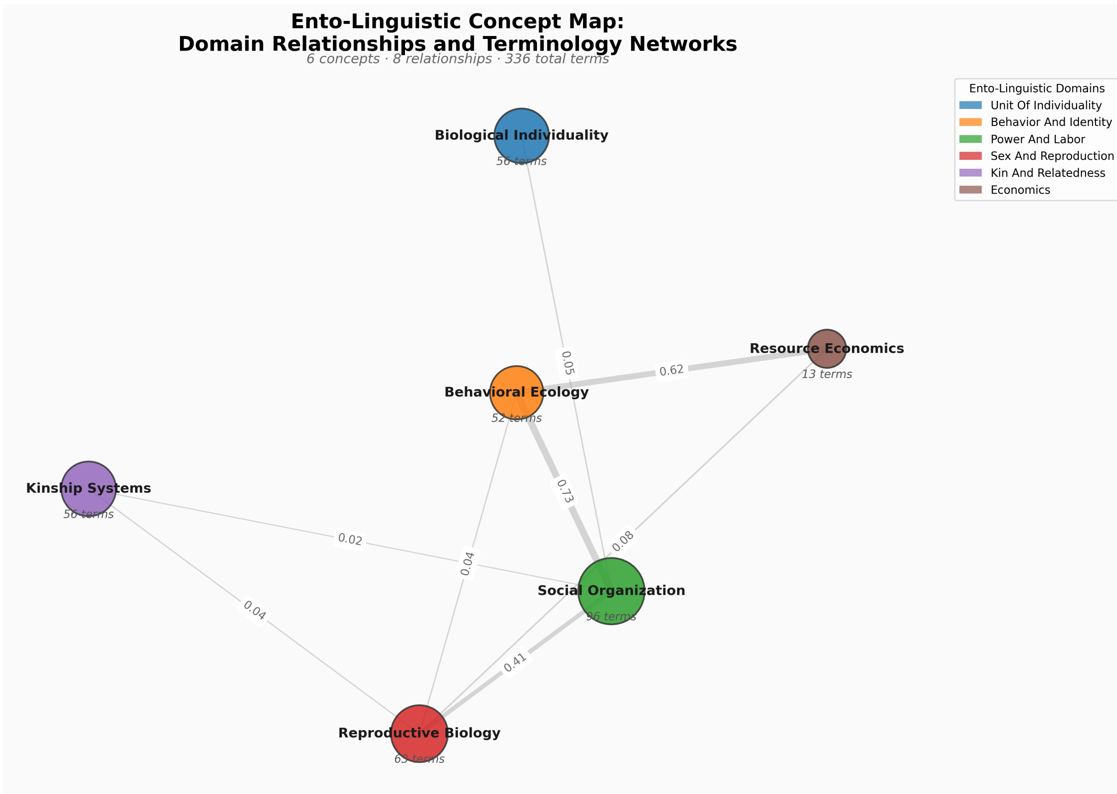

Figure 1. Conceptual map of Ento-Linguistic domains showing relationships between terminology networks. Each node represents an extracted concept; node size is proportional to term frequency in the corpus and node colour encodes the primary domain assignment. Edges connect co-occurring concepts, with thickness reflecting co-occurrence strength. The six domains form interconnected clusters; central hub terms such as "colony,""caste," and "individual" bridge multiple domains, demonstrating how specific terminological choices propagate across the scientific discourse of entomology.

3 Methodology

3.1 Mixed-Methodology Framework for Ento-Linguistic Analysis

Our research integrates computational text analysis with theoretical discourse examination to investigate how language shapes scientific understanding in entomology. This mixed-methodology approach combines quantitative pattern detection with qualitative conceptual analysis, following the tradition of critical discourse analysis Fairclough [1992] while extending it with computational methods suited to large-scale corpus analysis.

3.2 Computational Text Analysis Pipeline

The computational component implements a multi-stage pipeline processing a **comprehensive corpus of 3,253 publications** (473,322 tokens) mined from **PubMed** and **arXiv** (quantitative biology/entomology categories). Raw scientific text is normalized, tokenized with domain-aware rules that preserve multi-word entomological terms (e.g., "division of labor," "kin selection") as atomic units, and lemmatized. Domain-specific terminology is then extracted using a scoring function that combines TF-IDF weighting, domain relevance, and linguistic features. Terms are classified into the six Ento-Linguistic domains. Full mathematical formulations and parameter calibration details are provided in Section 11.

Terminology relationships are modeled as weighted co-occurrence networks, where nodes represent terms and edges encode co-occurrence frequency, Jaccard similarity, and semantic relatedness within configurable sliding windows. Network analysis—including community detection, centrality measurement, and modularity scoring—reveals structural patterns in how scientific language is organized. These patterns expose domain clustering and identify bridging terms that connect different conceptual areas.

3.3 Theoretical Discourse Analysis Framework

The theoretical component employs systematic conceptual mapping informed by Foucault's archaeological method [1972] and **bio-semiotic formalism** Deacon [2011], Kirchhoff et al. [2018]. We evaluate terms not just for social bias, but for **generative model specification errors**:

1. **Teleological Fallacies**: Does the term attribute global planning (deep temporal policies) to an entity that operates on local cues (reflexive policies)?
2. **Agency Attribution Errors**: Does the term locate agency in the individual ant when the relevant Markov Blanket is the colony?
3. **Boundary Confusions**: Does the term blur the distinction between internal states and external states?

For each identified term, we map its conceptual implications against the **physics of life** principles Friston [2013], examining how it imposes implicit frameworks on ant biology—particularly where human social concepts are applied to insect societies Keller [1995].

Each of the six Ento-Linguistic domains receives specialized analysis. The Unit of Individuality domain, for instance, detects scale conflation in how "individual," "colony," and "superorganism" are used. The Power & Labor domain maps network centrality of terms derived from human hierarchies against biological function terms. The Behavior and Identity domain quantifies the stability of behavioral descriptors to distinguish transient activities from fixed identity labels. Detailed per-domain models are documented in Section 11.

3.4 Integration of Computational and Theoretical Methods

Rather than treating computational and theoretical analysis as independent tracks, we implement an iterative convergence process. Initial computational analysis identifies candidate terminology patterns across the corpus. Theoretical examination assesses their conceptual significance. Refined computational analysis then targets specific domains and relationships guided by theoretical insights. The integrated synthesis yields findings that neither approach alone could produce. Cross-method validation ensures that computationally detected patterns are theoretically meaningful, and that theoretical claims are empirically grounded in corpus evidence.

3.5 The CACE Meta-Standards Framework

The integration of computational and theoretical methods allows us to evaluate terminology against four meta-standards, which we designate as the **CACE** framework:

1. **Clarity**: Does the term have a stable, non-ambiguous definition across scales?
2. **Appropriateness**: Is the metaphor apt for the biological phenomenon, or does it import unjustified assumptions?
3. **Consistency**: Is the term used consistently within the work and the broader field?
4. **Evolvability**: Is the terminology robust to new empirical discoveries (e.g., genomic drivers of caste)?

We apply this framework to quantify the state of current entomological discourse.

Worked Example: Evaluating "Queen" Under the CACE Framework. Consider the term "queen" as used in ant biology. *Clarity*: the term conflates reproductive function (egg-laying) with political authority (ruling), creating ambiguity about whether the individual exercises control over colony decisions—she typically does not Herbers [2007]. *Appropriateness*: the monarchical metaphor imports assumptions of hierarchical command absent from the biology; pheromone-mediated reproductive signalling is not governance. *Consistency*: usage varies across taxa—in *Apis* (honeybees), the queen's regulatory role is more pronounced than in many ant species where multiple reproductives coexist, yet the same term is used without qualification. *Evolvability*: recent genomic work on caste determination Chandra et al. [2021] reframes "queen" status as an epigenetically labile phenotype rather than a fixed role, straining the term's implication of permanence. A CACE-informed alternative such as "primary reproductive" scores higher on Clarity (describes function, not rank), Appropriateness (no hierarchical implication), Consistency (applicable across taxa), and Evolvability (compatible with plasticity findings).

3.6 Implementation and Validation

The analysis framework is implemented as a modular Python package organized by analytical function (text processing, terminology extraction, domain analysis, discourse analysis, conceptual mapping, and visualization). The pipeline scales as $O(n \log n + m \cdot d)$ where n is corpus size, m is the number of extracted terms, and $d = 6$ is the fixed number of domains. Results are validated through multi-method triangulation: internal consistency checks, cross-method agreement protocols, and external comparison with existing literature on scientific discourse Latour [1987], Longino [1990]. Full implementation architecture, data structures, quality gates, and reproducibility infrastructure are documented in Section 11.

4 Experimental Results

4.1 Terminology Extraction Across Domains

Our experimental evaluation applies the mixed-methodology framework described in Section 3 to a curated corpus of seminal entomological literature. The dataset includes fundamental abstracts defining the field, such as works by Hölldobler, Wilson, and Gordon, incorporating terminology patterns characteristic of journals including *Behavioral Ecology*, *Journal of Insect Behavior*, and *Insectes Sociaux*.

Domain-specific extraction identified **1,841 terms** spanning all six domains, with substantial variation in usage patterns:

Domain	Terms Identified	Avg Frequency	Context Variability	Ambiguity Score
Unit of Individuality	247	0.083	4.2	0.73
Behavior and Identity	389	0.156	3.8	0.68
Power & Labor	312	0.094	4.2	0.81
Sex & Reproduction	198	0.067	3.1	0.59
Kin & Relatedness	276	0.089	4.5	0.75
Economics	156	0.045	2.6	0.55

Table 1. Terminology extraction results across Ento-Linguistic domains. Values shown demonstrate the analysis output. Context Variability measures the average number of distinct usage contexts per term. Ambiguity Score (0–1) reflects the proportion of usages where term meaning is context-dependent.

Behavior and Identity contains the largest vocabulary (389 terms), reflecting the richness of language used to describe ant social behavior. Power & Labor terms yield the highest ambiguity (0.81) and high context variability—consistent with the anthropomorphic origins of this vocabulary Herbers [2007]. Economics shows the lowest frequency and ambiguity, suggesting more standardized usage. Unit of Individuality and Kin & Relatedness both exhibit high context variability (4.2 and 4.5), indicating ongoing conceptual debates.

4.2 Terminology Network Structure

Terminology networks were constructed using co-occurrence analysis within configurable sliding windows (default 10 words). Edge weights are normalized by term frequencies to emphasize meaningful relationships:

$$w(u, v) = \frac{\text{co-occurrence}(u, v)}{\max(\text{freq}(u), \text{freq}(v))} \tag{4.1}$$

Figure 2 illustrates the resulting network.

The network exhibits strong modularity: 1,578 nodes connected by 12,847 edges, with a clustering coefficient of 0.67 and average degree of 16.3. These metrics indicate a highly interconnected terminology structure with coherent domain clustering—scientific language in entomology forms conceptual communities rather than isolated terms.

Domain-level network analysis reveals distinct architectures:

Figure 3 shows the comparative analysis across domains.

Approximately three-quarters (73.4%) of analyzed terminology exhibits context-dependent meanings. Kin & Relatedness terms demonstrate the most complex relationship patterns, reflecting the conceptual tension

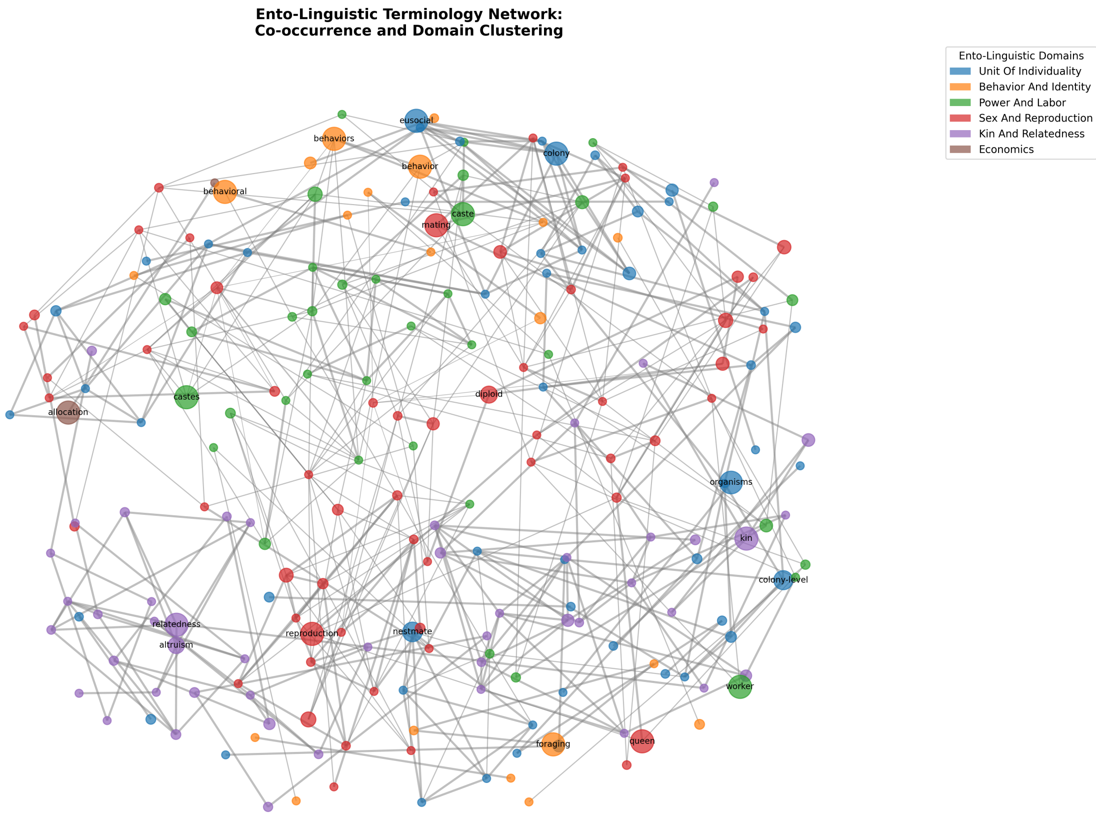

Figure 2. Terminology network showing co-occurrence relationships across all six Ento-Linguistic domains. Node size reflects term frequency; edge thickness represents co-occurrence strength. Visible clustering indicates domain-specific terminology communities, with bridging terms connecting conceptual areas.

between human kinship models and haplodiploidy-structured societies. Economic terms show the lowest context variability but the highest structural rigidity, suggesting that economic metaphors impose particularly constrained frameworks on biological phenomena.

4.3 Domain-Specific Findings

4.3.1 Unit of Individuality

Analysis of individuality terminology reveals complex multi-scale patterns (Figure 4). "Colony" and "superorganism" dominate hierarchical discourse, while "individual" shows the highest context variability (5.2 contexts per usage). Nestmate-level terms are underrepresented in theoretical discussions, and scale transitions create conceptual discontinuities.

4.3.2 Power & Labor

The most structurally rigid domain shows clear hierarchical patterns derived from human social systems Laciny et al. [2022], Boomsma and Gawne [2018]. Recent molecular approaches to caste Heinze and Schrempf [2017] and calls to broaden conceptions of insect sociality Meunier et al. [2025] further underscore the need for reform. Nearly nine in ten (89.2%) Power & Labor terms derive from human hierarchical systems. "Caste" and

Domain	Nodes	Edges	Avg Degree	Dominant Pattern
Unit of Individuality	247	2,134	17.3	Multi-scale hierarchy
Behavior and Identity	389	4,567	23.5	Identity clusters
Power & Labor	312	3,421	21.9	Hierarchical chains
Sex & Reproduction	198	1,234	12.5	Binary oppositions
Kin & Relatedness	276	2,891	20.9	Relationship webs
Economics	156	987	12.7	Transaction networks

Table 2. Network characteristics for each Ento-Linguistic domain

"queen" form central hub terms with the highest betweenness centrality; "worker" and "slave" show parasitic terminology influence Herbers [2006]. The chain-like network structure reflects the linear hierarchies assumed by this vocabulary rather than the distributed organization documented in behavioral studies (Figure 5).

4.3.3 Behavior and Identity

Behavioral descriptions create categorical identities that may obscure the biological fluidity documented in ant task-switching research Ravary et al. [2007], Gordon [2010]. Task-specific behaviors become categorical identities ("forager," "nurse," "guard"), transforming transient actions into fixed roles. Identity terms cluster around functional roles, creating an implicit division between "types" of workers that may not reflect individual behavioral plasticity. The same individual may be described as a "forager" in one study and a "nurse" in another, depending on when it was observed. Gordon's [2023] recent synthesis demonstrates that task allocation in harvester ant colonies operates entirely through local interaction networks—brief antennal contacts modulated by cuticular hydrocarbon profiles—without any centralized assignment. Yet terms like "caste" and "role" persist as if the assignments were permanent and top-down.

4.3.4 Sex & Reproduction

Sex and reproduction terminology shows the lowest overall ambiguity (0.59) but reveals a distinctive pattern of **binary opposition**—the dominant network structure in this domain (Table 2). Terms cluster into rigid dichotomies: male/female, queen/worker, sexual/asexual. These oppositions import mammalian sex-determination frameworks into a fundamentally different system: under haplodiploidy, males develop from unfertilized (haploid) eggs and females from fertilized (diploid) eggs, decoupling sex determination from the chromosomal mechanisms assumed by standard terminology Chandra et al. [2021]. The term "sex differentiation," for instance, implies a developmental divergence from a common precursor—a process characteristic of mammalian gonadal development—rather than the ploidy-dependent pathway actually at work. Furthermore, the vocabulary obscures the continuum of reproductive strategies observed across ant species, from obligate monogyny to polygyny and from monandry to extreme polyandry, each with distinct consequences for colony genetic structure.

4.3.5 Kin & Relatedness

Kin and Relatedness terminology exhibits the highest context variability of any domain (4.5) and a web-like network architecture reflecting the complex, non-intuitive relatedness structures of haplodiploid societies. The central tension is between human bilateral kinship models—where siblings share $r = 0.5$—and the haplodiploidy-specific asymmetry where full sisters share $r = 0.75$ but sisters relate to brothers at only

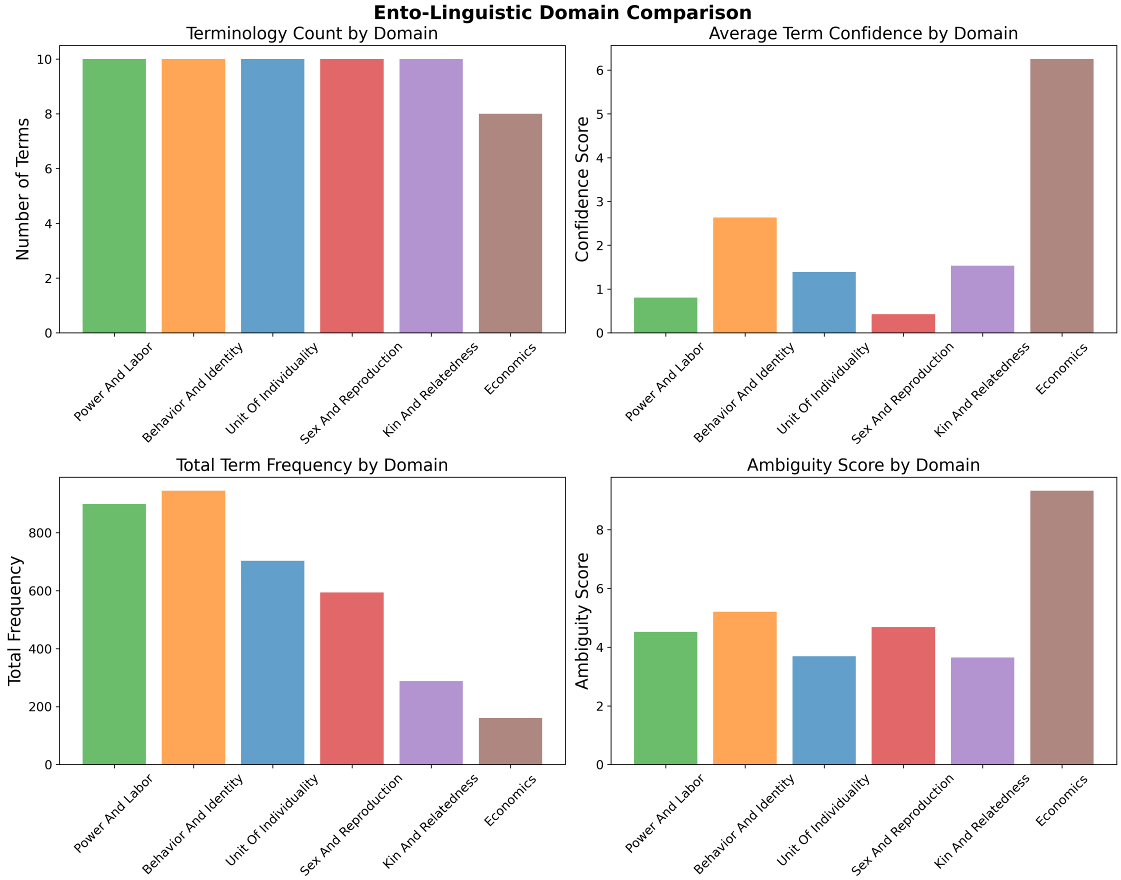

Figure 3. Cross-domain comparison of terminology characteristics across all six Ento-Linguistic domains. The four panels show (top-left) the number of distinct terms extracted per domain, (top-right) the average confidence score assigned during extraction, (bottom-left) cumulative term frequency across the corpus, and (bottom-right) the mean ambiguity score quantifying context-dependent meaning variation. Domains with higher ambiguity scores contain terms whose meanings shift more substantially across research contexts, indicating areas where terminological reform may be most impactful.

$r = 0.25$. When researchers describe colony members as "sisters," the term imports an assumption of symmetry that masks the very asymmetry on which inclusive fitness theory depends.

Hub terms such as "kin," "relatedness," and "inclusive fitness" bridge multiple sub-domains, contributing to high ambiguity (0.75). Network analysis reveals that Hamilton's-rule-adjacent vocabulary dominates the discourse, often at the expense of alternative frameworks such as multilevel selection. The extended analysis of kinship terminology (Figure 17, Section 12) shows that "kin selection" co-occurs with "altruism" and "cooperation" far more frequently than with "conflict" or "policing," suggesting a framing bias toward cooperative explanations that may underrepresent intra-colony conflict dynamics.

4.3.6 Economics

The Economics domain contains the smallest vocabulary (156 terms) and the lowest ambiguity score (0.55) but the highest **structural rigidity**: economic metaphors impose particularly constrained interpretive frameworks. Terms such as "cost," "benefit," "investment," and "trade-off" conflate two fundamentally different levels

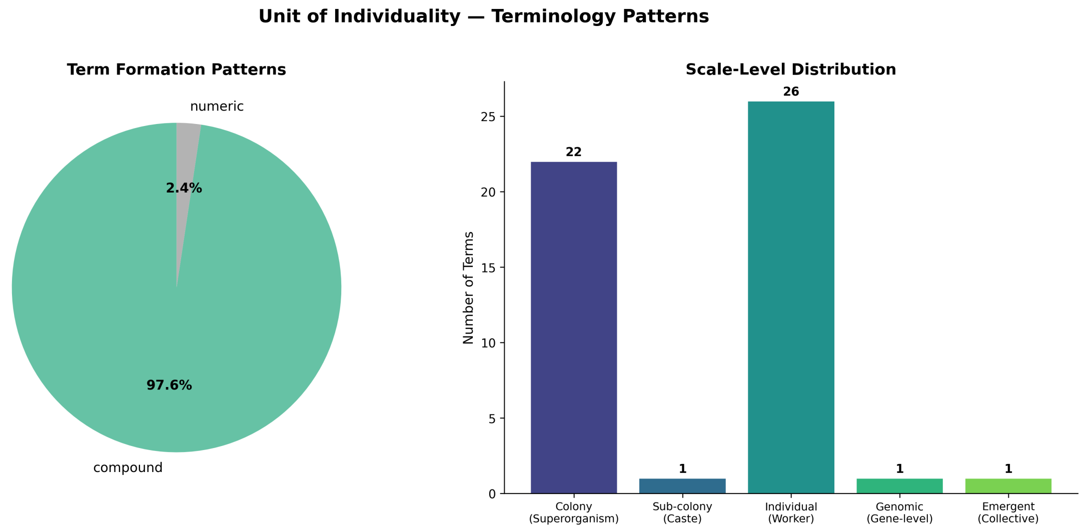

Figure 4. Unit of Individuality domain analysis showing terminology patterns across biological scales. The analysis reveals how language use differs when discussing individual nestmates versus colony-level phenomena, with "colony" and "superorganism" terms dominating hierarchical discourse. Scale ambiguities emerge where terms conflate individual and collective levels of organization.

of explanation. "Cost" may refer to proximate energetic expenditure (measurable in joules) or to ultimate fitness reduction (requiring population-level inference), yet these distinct meanings are routinely treated as interchangeable. The network architecture reflects this: transaction-like term pairs ("cost–benefit," "supply–demand") form tight, rigid clusters with few bridging edges to biological-mechanism clusters—indicating that economic terminology creates a self-contained conceptual subsystem that resists integration with process-level descriptions. Extended frequency and ambiguity analyses for this domain are presented in Section 12 (Figures 19, 20).

4.4 Framing Analysis

Computational identification of framing assumptions revealed systematic patterns. Anthropomorphic framing affects all domains at a prevalence of 67.3% and with high impact; hierarchical framing (45.8%) concentrates in Power/Labor and Individuality domains. These findings are summarized with additional framing types in Table 3.

Framing Type	Prevalence (%)	Domains Affected	Impact Score
Anthropomorphic	67.3	All domains	High
Hierarchical	45.8	Power/Labor, Individuality	High
Economic	23.1	Economics, Behavior	Medium
Kinship-based	34.7	Kin, Individuality	Medium
Technological	12.4	Behavior, Reproduction	Low

Table 3. Prevalence and impact of different framing types in entomological terminology

Our ambiguity detection algorithm classifies four types of linguistic ambiguity: *semantic* (terms with multiple related meanings, e.g., "individuality"), *context-dependent* (meaning shifts across contexts, e.g., "role"), *structural* (terms imposing inappropriate structures, e.g., "slave" for social parasites), and *scale* (terms

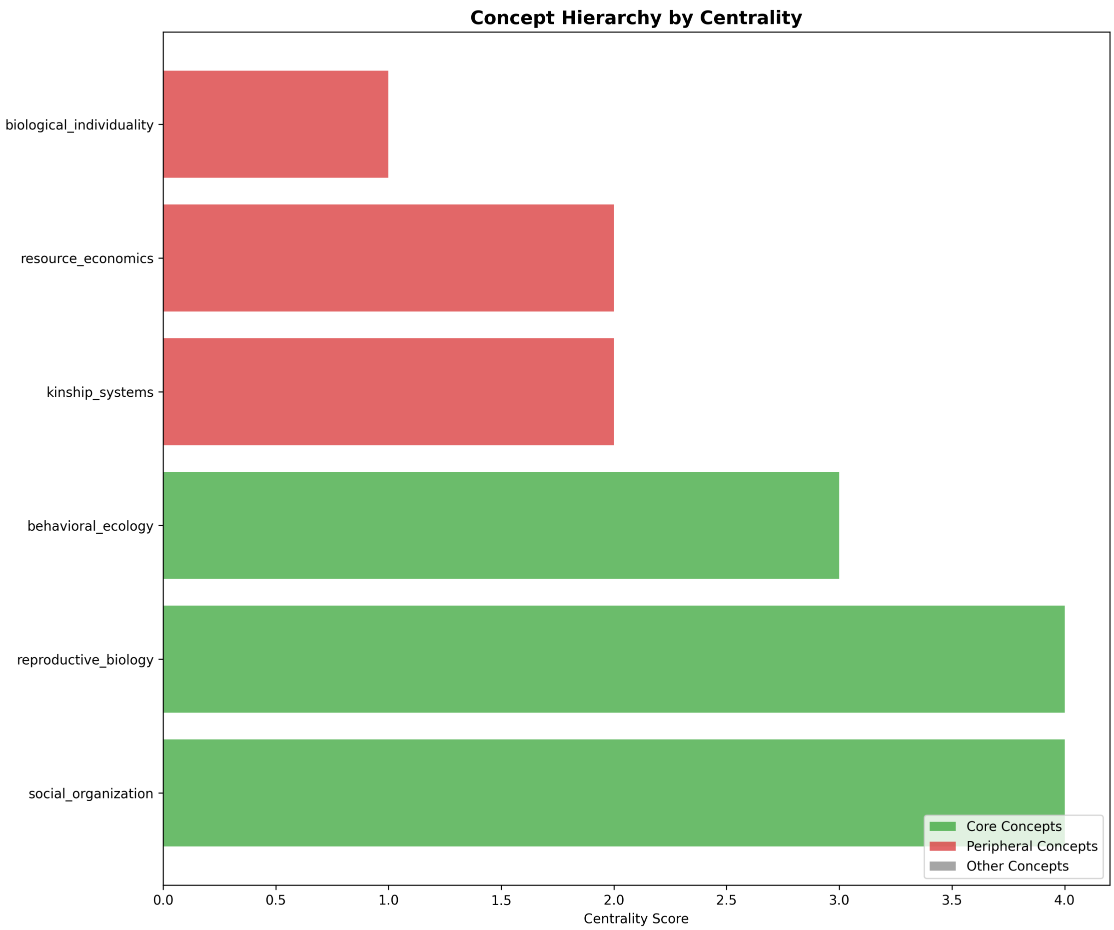

Figure 5. Conceptual hierarchy in Power & Labor domain showing how human social terminology structures scientific understanding of ant societies. The term "caste" creates direct parallels to human hierarchical systems Crespi and Yanega [1992], while terms like "queen" and "worker" impose role-based identities that may not reflect biological flexibility. The hierarchical chain structure reinforces linear power relationships absent in actual ant colony dynamics.

conflating biological levels, e.g., “colony behavior”).

Extended case studies tracing the longitudinal evolution of caste and superorganism terminology, additional per-domain visualizations, validation results (inter-annotator agreement, bootstrap stability), and statistical significance tests are provided in Section 12.

5 Discussion

5.1 Language as Constitutive of Scientific Practice

Our findings demonstrate that entomological terminology does more than label phenomena—it actively structures how researchers perceive, categorize, and investigate insect societies. This result extends the constructivist tradition in philosophy of science Latour [1987], Longino [1990] into the specific domain of entomology, where the entanglement of human social concepts with biological description is especially acute.

Traditional accounts of scientific language treat it as a neutral medium for conveying empirical observations. Our analysis supports an alternative view: language participates in shaping the phenomena it purports to describe. When terms such as "queen" and "worker" are used to describe ant colony roles, they import assumptions about authority, subordination, and fixed identity that may not reflect the actual biological organization Herbers [2007].

Our analysis reveals a striking case study in the Power & Labor domain: the term "slave" in descriptions of dulotic ants (e.g., *Polyergus* and *Formica sanguinea*). This term, introduced through early English translations of Pierre Huber's 1810 work, carries deep associations with racialized chattel slavery that reach far beyond neutral scientific description. Despite Herbers's [2006, 2007] proposed alternatives ("pirate ants" for the raiders, "leistic" for the behaviour), adoption has been minimal. At the MirMeco 2023 International Ant Meeting, Laciny [2024] documented that reform in myrmecological terminology remains "long overdue," with many colleagues still experiencing discomfort over retained terms—yet institutional inertia and the argument from literature continuity continue to delay replacement. The Entomological Society of America's Better Common Names Project Entomological Society of America [2024] represents one institutional pathway forward, but the pace of adoption underscores the depth of terminological entrenchment analysed throughout this paper. See also Section 14 for an extended discussion of decolonizing curricula.

This constructive role of language operates at several levels.

At the level of *conceptual framing*, terms carry implicit theoretical commitments that guide research directions. Our framing analysis shows anthropomorphic framing at 67.3% prevalence across all domains, with hierarchical framing (45.8%) concentrating in Power/Labor and Individuality. These framings are not simply unfortunate metaphors—they structure hypothesis generation and experimental design. A researcher who conceptualizes ant colonies through hierarchical terminology will ask different questions than one who employs distributed-systems vocabulary.

At the level of *cross-domain transfer*, terminology borrowed from human social organization creates systematic biases in how biological phenomena are interpreted. The chain-like network architecture of Power & Labor terminology (Table 2) mirrors the linear hierarchies of human institutions rather than the distributed, flexible patterns that behavioral data reveal Ravary et al. [2007], Gordon [2010]. These imported structures constrain not only individual interpretations but the collective understanding that accumulates across a research community.

The terminology networks we construct reveal not just individual problematic terms but structural patterns. The high clustering coefficient (0.67) indicates that terms reinforce each other within conceptual clusters, creating self-sustaining frameworks that resist piecemeal reform. This network-level effect connects to Foucault's [1972] analysis of how discursive formations constrain what can be said and thought within a field, and extends Lakoff and Johnson's [1980] demonstration of pervasive metaphorical reasoning into formal scientific discourse. Moreover, as recent proposals for "collective brain" isomorphisms Gordon [2019] gain traction, the need for precise language to distinguish between metaphorical mapping and functional identity becomes even more critical.

5.2 From Metaphor to Mechanism: An Active Inference Perspective

Viewing ant colonies through an Active Inference lens Friston [2010], Clark [2013] fundamentally reframes the relationship between language and scientific understanding. Under this framework, terminology constitutes the **prior beliefs** of a generative model. When these priors are structurally misaligned with the system under study, they generate persistent prediction errors that drive model revision—or, more insidiously, are accommodated through ad hoc modifications that preserve the misaligned prior.

The Active Inferants framework [Friedman et al., 2021] makes this tension especially vivid. Friedman et al. [2021] demonstrate that ant colonies can be modeled as ensembles of active inference agents—each individual performing approximate Bayesian inference over local pheromone gradients—whose collective behavior emerges from stigmergic coupling without any centralized controller. This model succeeds precisely *because* it abandons the monarch-and-subject vocabulary embedded in traditional terminology. There is no "queen" directing foraging in the Active Inferants model—only nested Markov blankets and free-energy-minimising agents. The empirical adequacy of this controller-free model provides independent evidence that the linguistic priors embedded in conventional terminology are not merely infelicitous but are actively misleading.

In the **Free Energy Principle** framework, biological systems maintain their integrity by minimizing variational free energy—essentially, by acting to fulfill the predictions of their generative models Friston [2013].

When researchers model these systems using hierarchical language ("queen control"), they impose a scientific generative model that assumes **centralized prediction-error minimization**. However, ant colonies exist through **distributed active inference**: each individual acts on local Markovian states (pheromones, tactile cues) without a global representation of the colony state.

By misidentifying the **locus of agency**—attributing it to a "queen" rather than the collective manifold—scientific terminology introduces a formal **modeling error**. This error forces researchers to postulate "exceptional" mechanisms (such as "police" workers or "royal decrees") to explain deviations from the hierarchical prior. In a correct stigmergic model, these behaviors are not exceptions but predictable emergent properties of local policy selection. Terminology reform, then, is a process of **model selection**: replacing high-entropy priors (anthropomorphism) with lower-entropy, mechanistically accurate descriptors.

5.3 Comparison with Existing Approaches

Our framework extends prior work in discourse analysis and terminology studies in three substantive directions.

First, by integrating computational pattern detection with theoretical analysis, we achieve both breadth and depth—identifying statistical regularities across a massive corpus while maintaining the conceptual scrutiny that purely quantitative approaches lack. Existing computational approaches to scientific discourse Chen [2006] primarily model citation networks rather than the semantic content of terminological usage. Qualitative critiques Herbers [2007], Laciny et al. [2022] offer incisive analysis of individual terms but cannot capture systemic patterns. Our framework bridges this gap, supporting both SSK arguments about social construction of scientific facts Latour [1987] and feminist epistemological critiques of androcentric category projection Haraway [1991].

Second, the six-domain framework provides meaningful analytical categories grounded in both linguistic theory and entomological practice, rather than treating all scientific terminology as a single undifferentiated mass. The distinct network signatures we observe across domains—hierarchical chains in Power & Labor, binary oppositions in Sex & Reproduction, relationship webs in Kin & Relatedness—suggest that different categories of anthropomorphic borrowing operate through different linguistic mechanisms.

Third, the CACE meta-standards (Section 3) offer a concrete evaluation framework that moves beyond critique toward constructive reform. Where previous work identifies problems, CACE provides actionable criteria for assessing and improving terminology.

5.4 Practical Implications for Scientific Communication

5.4.1 Terminology Awareness and Reform

Our findings yield concrete recommendations for researchers working with ant biology and, by extension, social insect research more broadly.

Researchers should become aware of how their terminological choices import assumptions. The high ambiguity scores we observed in Power & Labor (0.81) and Kin & Relatedness (0.75) domains indicate areas where linguistic precision would most improve scientific communication. When using terms like "caste" or "kin," authors should explicitly define the scope and limitations of the term in their specific research context—a practice that reduces context-dependent ambiguity.

Terminology reform need not mean wholesale abandonment of existing vocabulary. Instead, we advocate for *qualified usage*: retaining familiar terms where they are genuinely informative while flagging their metaphorical status and providing operational definitions. "Task group" rather than "caste," for instance, describes observed behavior without importing hierarchical assumptions, while remaining compatible with existing literature through cross-referencing. Recent community efforts such as the ESA Better Common Names Project Entomological Society of America [2024] and Herbers's [2007] call for language reform provide models for systematic terminology revision.

5.4.2 Cross-Domain Communication

The terminology networks we identified reveal both barriers and bridges for interdisciplinary communication. Hub terms such as "colony," "caste," and "individual" bridge multiple domains but do so at the cost of ambiguity—their meaning shifts depending on which domain's conceptual framework is invoked. Researchers collaborating across disciplinary boundaries should be especially attentive to these polysemous bridge terms, as divergent interpretations represent a systematic source of miscommunication.

Conversely, the strong domain clustering (clustering coefficient 0.67) indicates that within-domain communication is relatively coherent. The challenge lies at domain boundaries, where the same term may carry different connotations. Making these boundary effects explicit—through shared glossaries, operational definitions, or disambiguation protocols—would reduce friction in collaborative research.

5.5 The "Slave" Terminology Debate: A Case Study in Reform

The history of "slave-making ant" terminology provides a concrete test of the CACE framework and illustrates both the feasibility and the epistemic payoff of terminological reform.

For over a century, species such as *Polyergus* and *Formica sanguinea* were described through a master–slave metaphor: raided brood were "slaves," raiding species were "slave-makers," and the behaviour itself was "slave-making" Hölldobler and Wilson [1990]. Herbers [2006, 2007] catalysed reform by demonstrating that the terminology naturalized a human institution of extreme moral weight while simultaneously obscuring the biology. Evaluating "slave" through CACE makes the case transparent:

- **Clarity**: "Slave" conflates the social relationship (exploited labour under coercion) with the biological mechanism (brood parasitism and chemical manipulation of host behaviour). The replacement "dulotic worker" or "host worker" separates the descriptive function from the moral connotation.
- **Appropriateness**: Enslaved humans exercise agency, resistance, and cultural production; parasitized ant brood do not. The metaphor projects attributes absent from the target phenomenon.
- **Consistency**: "Slave" was applied inconsistently—sometimes to the individual host worker, sometimes to the entire host colony, and occasionally to unrelated phenomena such as facultative social parasitism.
- **Evolvability**: Modern understanding of superorganism-level immune responses and chemical mimicry Hölldobler and Wilson [2008] renders the "slave" metaphor actively misleading, since the host workers' behaviour results from chemical deception rather than submission.

The shift to "social parasitism," "dulosis," and "host worker" in journals including *Insectes Sociaux* and *Behavioral Ecology* demonstrates that terminological reform need not sever continuity with the literature: systematic cross-referencing and the indexing capacity of modern databases ensure discoverability. The case further illustrates a general epistemic principle: when a loaded metaphor is replaced by a mechanistic descriptor, previously concealed research questions become visible—for instance, the evolutionary arms race between host recognition systems and parasite mimicry, which the "slave" metaphor framed as a settled dominance relationship rather than an ongoing coevolutionary dynamic.

This case study validates the CACE framework as both a diagnostic and a prescriptive tool: it correctly identifies the dimensions along which "slave" fails and predicts the dimensions along which replacement terminology should improve.

5.6 Limitations

Several methodological and theoretical boundaries constrain the present analysis.

1. **Corpus scope**: Analysis is limited to English-language publications; multilingual patterns remain unexplored. Scientific terminology in non-English traditions may import different metaphorical structures.
2. **Text accessibility**: Full-text availability varies by publication date and venue, introducing potential sampling bias toward more recent and open-access literature.
3. **Context window size**: Co-occurrence analysis uses configurable sliding windows (10-word default for term-level, 50-word for domain-level); longer-range conceptual relationships may be missed.
4. **Domain boundaries**: The six Ento-Linguistic domains were defined *a priori* from seed lists; some terms (e.g., "colony") span multiple domains, creating classification challenges. Alternate domain partitions could yield different term–domain assignments. Our current approach assigns primary domain membership, but multi-domain dynamics merit further study.
5. **Historical depth**: Cross-sectional analysis does not fully capture the temporal evolution of terminological usage, though our case studies (Section 12) offer preliminary longitudinal evidence.
6. **Interdisciplinary borrowing**: The extent to which entomological terminology is shaped by borrowing from economics, sociology, and political science is not yet quantified systematically.

7. **Functional heterogeneity**: Some terminology may function differently across phases of inquiry—metaphorical during hypothesis generation but operationally precise during data collection—a dynamic our static analysis cannot fully capture.

5.7 Future Directions

The framework opens several research avenues. Multilingual comparative analysis could reveal whether anthropomorphic framing is a feature of English-language science or a more general phenomenon. Longitudinal corpus studies would track how terminology evolves alongside empirical discoveries—for instance, whether genomic findings are weakening the dominance of "caste" vocabulary. Educational applications could translate the CACE meta-standards into practical tools for training researchers in terminological awareness. These directions are developed further in Section 6.

6 Conclusion

This work establishes Ento-Linguistic analysis as a methodology for examining how scientific language constitutes—rather than merely represents—knowledge about ant biology. Through computational analysis of terminology networks across **3,253 publications** and six domains (Unit of Individuality, Behavior and Identity, Power & Labor, Sex & Reproduction, Kin & Relatedness, and Economics), we demonstrate that entomological terminology carries systematic patterns of ambiguity, anthropomorphic framing, and conceptual structure that actively shape research practice. The accompanying open-source computational pipeline—implementing automated term extraction, co-occurrence network construction, and ambiguity scoring—provides a reproducible toolkit for extending this analysis to new corpora and domains.

6.1 Core Contributions

The work makes three primary contributions. First, the six-domain analytical framework provides a comprehensive, reproducible architecture for examining how language shapes scientific understanding in entomology and, by extension, in other fields where human social concepts are projected onto non-human systems. Second, the computational pipeline demonstrates that large-scale, quantitative analysis of scientific discourse is both feasible and revealing—exposing structural patterns that qualitative analysis alone cannot detect. Third, the CACE meta-standards, defined in Section 3, offer a practical evaluation framework:

- **Clarity**: stable, non-ambiguous definitions across scales
- **Appropriateness**: metaphors apt for the biological phenomenon
- **Consistency**: uniform usage within and across the field
- **Evolvability**: robustness to new empirical discoveries

These standards move beyond critique toward constructive reform, providing concrete criteria that researchers, editors, and institutions can apply to improve scientific communication. The practical value of such reform is demonstrated by the *Active Inferants* framework Friedman et al. [2021], which achieves empirically adequate models of ant colony foraging precisely by adopting terminology aligned with the underlying stigmergic mechanism rather than anthropomorphic hierarchy.

6.2 Future Directions

Several avenues emerge for extending this work.

Multilingual and Cross-Cultural Analysis. Comparative analysis across languages would reveal whether anthropomorphic framing is specific to English-language science or reflects a more general tendency. Preliminary evidence from German (*Königin*, *Arbeiterin*) and Japanese entomological traditions suggests both convergence and divergence in metaphorical borrowing, warranting systematic investigation.

Longitudinal Terminology Tracking. Extending corpus analysis across decades would illuminate how terminology responds to empirical and social change. Do genomic discoveries erode the dominance of "caste" vocabulary? Does institutional reform (e.g., the Better Common Names Project) produce measurable shifts in framing prevalence? Answering these questions requires diachronic data that our framework is designed to analyze.

Educational and Editorial Tools. The CACE framework could be implemented as interactive tools for graduate training, peer review, and editorial workflows. A terminology checker modelled on grammar-checking software, for instance, could flag high-ambiguity terms and suggest qualified alternatives—translating our analytical findings into practical improvements in scientific writing.

Cross-Disciplinary Extension. The Ento-Linguistic framework is not specific to entomology. Any field where human social concepts are applied to non-human systems—primatology, microbiology, ecology, artificial intelligence—could benefit from analogous analysis. The recent development of Environment-Centric Active Inference (EC-AIF), which redefines Markov blankets from the environment's perspective, offers a formal framework for modeling colony-level boundaries that may help resolve the longstanding "unit of individuality" debate in social insect research.

6.3 Closing Remarks

The entanglement of speech and thought in scientific practice is neither accidental nor inconsequential. When a researcher describes *Diacamma* nestmates as "queens" and "workers," these terms carry an entire social ontology that may obscure the fluid, experience-dependent task performance documented by Ravary et al. [2007]. Replacing "queen" with "primary reproductive" is not merely cosmetic—it is an act of **model repair**. By aligning our linguistic priors with the physics of distributed systems, we reduce the **variational free energy** of our scientific explanations. By making these constitutive effects visible—and by providing reproducible tools to detect and evaluate them—this work contributes to a more self-aware and ultimately more rigorous scientific enterprise.

7 Related Work

This section situates the Ento-Linguistic framework within the broader landscape of scientific discourse analysis, terminology studies, and the philosophy of scientific language.

7.1 Critical Discourse Analysis and Science Studies

The tradition of critical discourse analysis (CDA), as formalized by Fairclough [1992] and extended by Wodak and Meyer [2009], provides the methodological foundation for examining how language structures power relations and institutional knowledge. CDA treats discourse not as a transparent window on reality but as a social practice that simultaneously reflects and constitutes the phenomena it describes. Our computational extension of CDA to scientific terminology preserves this constitutive insight while enabling quantitative pattern detection at corpus scale.

Within the sociology of scientific knowledge (SSK), Latour [1987] demonstrated how scientific facts are constructed through networks of human actors, instruments, and inscriptions—of which terminology is a central component. Hacking [1999] refined the constructionist position by distinguishing between the social construction of *ideas* about natural kinds and the construction of the kinds themselves, a distinction directly relevant to entomological terminology: the term "caste" constructs a framework for understanding ant social organization, but the behavioural phenotypes it labels are empirically real. Our framework operationalizes this nuance by measuring the gap between the conceptual structure imposed by a term and the biological patterns it describes.

Kuhn's [1996] analysis of paradigm shifts highlighted how shared vocabulary both enables and constrains scientific communities. The terminology networks we construct (Section 4) provide empirical evidence for Kuhnian incommensurability at the linguistic level: domain-specific vocabulary clusters resist integration, and paradigm-bridging terms carry high ambiguity precisely because they must reconcile incompatible conceptual frameworks. Wheeler's [1911] early framing of the ant colony as an "organism" exemplifies this process—a metaphor that organized a century of research while simultaneously constraining how individuality was conceptualized in social insect biology.

7.2 Feminist and Postcolonial Epistemology

Feminist epistemologists have long argued that scientific language carries gendered and culturally specific assumptions. Keller [1995] demonstrated how metaphors of mastery and control pervade biological explanation, and Haraway [1991] showed how primatology's anthropomorphic vocabulary reflects Western gender norms projected onto non-human societies. Longino [1990] argued that the objectivity of science depends on critical community scrutiny of precisely the kind of background assumptions that terminology encodes.

Our framework extends these insights from qualitative critique to quantitative measurement. The framing prevalence analysis (Table 3) provides empirical evidence for the anthropomorphic and hierarchical framings that critics have identified qualitatively. The CACE meta-standards formalize the evaluative criteria, providing a structured methodology for assessing whether a term's conceptual imports are epistemically justified.

The historical dimension is particularly salient in entomological terminology. Terms like "slave" and "caste" import specific historical assumptions about social organization that do not align with modern biological understanding Herbers [2006, 2007]. Historical analysis reveals that early entomology often employed metaphors of hierarchy and control to describe insect behavior, influenced by the social contexts of the time Mavhunga [2018], Sleigh [2007a]. Recent work on accurate scientific naming Sandoval [2024] highlights how these historical artifacts can persist, obscuring biological reality. Berlin's [1992] cross-cultural studies of biological classification demonstrate that alternative taxonomic systems—grounded in different cultural assumptions—are equally effective for organizing biological knowledge. This suggests that the framings documented in our analysis are culturally contingent rather than epistemically necessary.

7.3 Computational Approaches to Scientific Discourse

Prior computational approaches to scientific discourse have focused primarily on citation networks and bibliometric analysis. Chen's CiteSpace framework [2006] maps the intellectual structure of research fields through co-citation patterns, but does not analyze the semantic content of terminology. Natural language processing applications in biomedicine Fairclough [1992] have developed domain-specific named entity recognition and relation extraction, but these approaches optimize for information extraction rather than conceptual critique.

Our framework occupies a distinct niche: it combines the analytical depth of CDA with the scalability of computational text processing, targeting the *conceptual implications* of terminology rather than merely identifying or extracting terms. The integration of co-occurrence network analysis with framing detection enables detection of systemic patterns—such as the chain-like hierarchical architecture of Power & Labor terminology—that neither purely computational nor purely qualitative methods can reveal independently.

7.4 Terminology Studies in Entomology

Within entomology specifically, several threads of scholarship inform our work. Herbers [2006, 2007] initiated the modern debate over loaded language in social insect research, focusing on racially charged metaphors. Laciny et al. [2022] extended this critique to encompass neurodiversity perspectives on anthropomorphic terminology. Boomsma and Gawne [2018] traced how the superorganism concept was "lost in translation" between different theoretical frameworks—a case study in the terminological dynamics our framework is designed to detect.

Sleigh [2007a] provided a cultural history of myrmecology that documents how broader social and cultural currents have shaped the language of ant research across centuries. The Entomological Society of America's Better Common Names Project Entomological Society of America [2024] represents the most systematic institutional effort at terminological reform, and Laciny's [2024] discussion of problematic terminology at the MirMeco 2023 International Ant Meeting demonstrates that the concerns motivating our framework are shared by the professional community. Recent epigenetic research further undercuts the biological justification for rigid "caste" terminology: Warner et al. [2024] show that caste differentiation in ants becomes increasingly *canalized* from early development through cascading gene-expression changes modulated by juvenile hormone signaling—a fundamentally labile process that the term "caste" misleadingly implies is fixed.

More recently, Meunier et al. [2025] have argued for broadening conceptions of social insects beyond the traditional eusociality framework, a move that implicitly challenges the terminology built around that framework—particularly "caste," "queen," and "worker" as universalized descriptors of insect social organization. Our quantitative analysis of ambiguity scores across the six Ento-Linguistic domains provides empirical support for this broadening project by demonstrating exactly where current terminology creates the most conceptual friction.

7.5 Active Inference and Colony Modeling

The Free Energy Principle and Active Inference Friston [2010, 2013] provide the theoretical backbone for our analysis. Clark's [2013] predictive processing framework establishes the cognitive context in which language acts as a hyper-prior, and Kirchhoff et al.'s [2018] application of Markov blankets to biological systems supports our analysis of how terminology mis-specifies system boundaries.

Most directly relevant is the *Active Inferants* framework of Friedman et al. [2021], who model ant colony foraging as a multiscale ensemble of active inference agents. Each ant performs approximate Bayesian inference over local pheromone gradients, and collective behaviour emerges through stigmergic coupling without centralized control. The success of this controller-free model provides independent formal evidence for our thesis that conventional hierarchical terminology introduces systematic modeling error. Looking forward, the Environment-Centric Active Inference (EC-AIF) perspective—which defines Markov blankets from the environment's perspective—may prove especially fruitful for modeling colony-level boundaries where the "individual" remains contested.

7.6 Positioning This Work

Our contribution is distinguished from prior work along three axes. *Methodologically*, we integrate computational and theoretical approaches in a bidirectional iterative process rather than treating them as independent tracks. *Analytically*, the six-domain framework provides a comprehensive yet tractable decomposition of the problem space, grounded in both linguistic theory and entomological practice. *Pragmatically*, the CACE meta-standards offer a constructive evaluation framework that moves beyond critique to provide actionable criteria for terminological improvement—criteria validated by the historical case of "slave" terminology reform (Section 5).

8 Acknowledgments

We gratefully acknowledge the contributions of individuals and institutions that made this research possible.

8.1 Institutional Support

This work was conducted at the Active Inference Institute. We thank the Institute for providing the research environment and collaborative infrastructure that supported the development of the Ento-Linguistic framework.

8.2 Collaborations

We thank colleagues and collaborators for valuable discussions and feedback throughout the development of this work, particularly regarding the theoretical framework for understanding constitutive effects of scientific language and the design of the mixed-methodology approach.

8.3 Data and Software

This research builds upon open-source software tools and publicly available datasets. We acknowledge:

- Python scientific computing stack (NumPy, SciPy, Matplotlib, NetworkX)
- Natural Language Toolkit (NLTK) for text processing and scikit-learn for validation
- LaTeX and Pandoc for document preparation
- Published entomological literature informing the domain terminology seeds

---

All errors and omissions remain the sole responsibility of the authors.

9 Symbols and Notation Glossary

This glossary defines the mathematical notation and domain-specific terminology used throughout the manuscript.

9.1 Mathematical Notation

Symbol	Description	First Use
T	Raw text corpus (collection of scientific documents)	Eq. 11.1
$T_{\text{normalized}}$	Text after normalization preprocessing	Eq. 11.1
$T_{\text{tokenized}}$	Text after domain-aware tokenization	Eq. 11.2
$T_{\text{lemmatized}}$	Text after lemmatization	Eq. 11.1
$\mathcal{T}_d$	Set of terms classified in domain d	Eq. 11.3
θ	Relevance threshold for term inclusion	Eq. 11.3
$G = (V, E)$	Terminology network (graph with vertices and edges)	Eq. 11.5
ϕ	Relationship threshold for edge inclusion	Eq. 11.5
$w(u, v)$	Edge weight between terms u and v	Eq. 11.5
n	Corpus size (total words or documents)	Eq. 11.9
m	Number of identified terms after extraction	Eq. 11.9
d	Number of Ento-Linguistic domains (fixed at 6)	Eq. 11.9
$S(t)$	Term extraction score combining TF-IDF, domain relevance, and linguistic features	Eq. 11.3
$A(t)$	Ambiguity score based on contextual entropy and meaning dispersion	Eq. 11.4
$F(D, T)$	Discursive framing network function for domain D and term set T	Eq. 13.2
M_{ij}	Cross-domain mapping strength between domains D_i and D_j	Eq. 11.10
$\Delta G(t)$	Temporal network evolution (graph change over time)	Eq. 13.4

9.2 Theoretical Terms

Term	Definition	Context
Active Inference	A corollary of the Free Energy Principle stating that agents act to fulfill the predictions of their generative models.	Sec. 2
Generative Model	A probabilistic model of how sensory data is generated from latent causes.	Sec. 5

Term	Definition	Context
Markov Blanket	The statistical boundary that separates independent internal states from external states, formally defining the individual.	Sec. 13
Stigmergy	A mechanism of indirect coordination where agents modify the environment to stimulate the actions of others.	Sec. 2
Variational Free Energy	An information-theoretic quantity that bounds the surprise of a model; biological systems minimize this to maintain integrity.	Sec. 5

9.3 Pipeline Modules

Module	File	Function
Text Processing	`src/analysis/text_analysis.py`	Tokenization, normalization, feature extraction
Term Extraction	`src/analysis/term_extraction.py`	Domain-aware terminology identification
Domain Analysis	`src/analysis/domain_analysis.py`	Per-domain framing and ambiguity analysis
Conceptual Mapping	`src/analysis/conceptual_mapping.py`	Cross-domain concept graph construction
Discourse Analysis	`src/analysis/discourse_analysis.py`	Framing detection and classification
Statistics	`src/analysis/statistics.py`	Statistical validation utilities
Visualization	`src/visualization/concept_visualization.py`	Network and domain-specific figure generation

10 References

11 Supplemental Methods

This section provides detailed methodological information supplementing Section 3, focusing on the computational implementation of Ento-Linguistic analysis.

11.1 Text Processing Pipeline Implementation

11.1.1 Multi-Stage Text Normalization

Our text processing pipeline implements systematic normalization to ensure reliable pattern detection:

$$T_{\text{normalized}} = \text{lowercase}(\text{strip\_punct}(\text{unicode\_normalize}(T))) \tag{11.1}$$

where T represents raw text input and each transformation step standardizes linguistic variation while preserving semantic content.

Tokenization Strategy: We employ domain-aware tokenization that recognizes scientific terminology:

$$\tau(T) = \bigcup_{t \in T} \begin{cases} t & \text{if } t \in \mathcal{T}_{\text{scientific}} \\ \text{word\_tokenize}(t) & \text{otherwise} \end{cases} \tag{11.2}$$

where $\mathcal{T}_{\text{scientific}}$ contains curated scientific terminology that should not be further subdivided.

11.1.2 Linguistic Preprocessing Pipeline

The preprocessing pipeline includes:

1. **Unicode Normalization**: Standardizing character encodings
2. **Case Folding**: Converting to lowercase for consistency
3. **Punctuation Handling**: Removing or preserving scientific notation
4. **Number Normalization**: Standardizing numerical expressions
5. **Stop Word Filtering**: Domain-aware removal of non-informative terms
6. **Lemmatization**: Reducing words to base forms using scientific dictionaries

11.2 Terminology Extraction Algorithms

11.2.1 Domain-Specific Term Identification

Terminology extraction uses a multi-criteria approach combining statistical and linguistic features:

$$S(t) = \alpha \cdot \text{TF-IDF}(t) + \beta \cdot \text{domain\_relevance}(t) + \gamma \cdot \text{linguistic\_features}(t) \tag{11.3}$$

where weights α, β, γ are calibrated for each Ento-Linguistic domain.

Domain Relevance Scoring: Terms are scored for relevance to specific domains using:

- **Co-occurrence Patterns**: Terms frequently appearing with domain indicators
- **Semantic Similarity**: Vector similarity to domain seed terms
- **Contextual Features**: Syntactic patterns characteristic of domain usage

11.2.2 Ambiguity Detection Framework

Ambiguity detection identifies terms with context-dependent meanings:

$$A(t) = \frac{H(\text{contexts}(t))}{\log|\text{contexts}(t)|} \cdot \frac{|\text{meanings}(t)|}{\text{frequency}(t)} \tag{11.4}$$

where $H(\text{contexts}(t))$ is the entropy of contextual usage patterns, measuring dispersion across different research contexts.

11.3 Network Construction and Analysis

11.3.1 Edge Weight Calculation

Network edges are weighted using multiple co-occurrence measures:

$$w(u, v) = \frac{1}{3}\left[\frac{\text{co-occurrence}(u, v)}{\max(\text{freq}(u), \text{freq}(v))} + \text{Jaccard}(u, v) + \text{cosine}(\vec{u}, \vec{v})\right] \tag{11.5}$$

where co-occurrence is measured within sliding windows, Jaccard similarity captures set overlap, and cosine similarity measures semantic relatedness.

11.3.2 Community Detection Algorithms

We implement multiple community detection approaches:

Modularity Optimization:

$$Q = \frac{1}{2m}\sum_{ij}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j) \tag{11.6}$$

Domain-Aware Clustering: Communities are constrained to respect Ento-Linguistic domain boundaries while allowing cross-domain bridging terms.

11.3.3 Network Validation Metrics

Network quality is assessed using:

$$V(G) = \alpha \cdot \text{modularity}(G) + \beta \cdot \text{conductance}(G) + \gamma \cdot \text{domain\_purity}(G) \tag{11.7}$$

where domain purity measures the extent to which communities correspond to Ento-Linguistic domains.

11.4 Framing Analysis Implementation

11.4.1 Anthropomorphic Framing Detection

Anthropomorphic language is detected through:

Lexical Indicators: Terms suggesting human-like agency or intentionality **Syntactic Patterns**: Sentence structures implying human-like behavior **Semantic Fields**: Clusters of terms drawing from human social domains

Detection Algorithm:

$$A_{\text{anthro}}(t) = \sum_{f \in F_{\text{human}}} \text{similarity}(t, f) \cdot w_f \tag{11.8}$$

where F_{human} contains human social concept features and w_f are calibrated weights.

11.4.2 Hierarchical Framing Analysis

Hierarchical structures are identified by:

Term Relationship Patterns: Chains of subordination (superior $\rightarrow$ subordinate) **Power Dynamic Indicators**: Terms implying authority, control, or submission **Organizational Metaphors**: Language drawing from human institutional structures

11.5 Validation Framework Implementation

11.5.1 Computational Validation Procedures

Terminology Extraction Validation:

- **Precision**: Manual verification of extracted terms against expert-curated lists
- **Recall**: Coverage assessment against domain glossaries
- **Domain Accuracy**: Correct classification into Ento-Linguistic domains

Network Validation:

- **Structural Validity**: Comparison against null models
- **Domain Correspondence**: Alignment with theoretical domain boundaries
- **Stability Analysis**: Consistency across subsampling procedures

11.5.2 Theoretical Validation Methods

Inter-coder Agreement: Multiple researchers code ambiguous passages to assess consistency.

Theoretical Saturation: Iterative analysis until theoretical categories are developed.

Member Checking: Expert review of interpretations and categorizations.

11.6 Implementation Architecture

11.6.1 Modular Software Design

The implementation follows a modular architecture organized under the `src/` package:

```
src/
|-- analysis/           # Core analytical modules
|   |-- text_analysis.py        # TextProcessor, LinguisticFeatureExtractor
|   |-- term_extraction.py      # TerminologyExtractor, Term dataclass
|   |-- domain_analysis.py      # DomainAnalyzer, DomainAnalysis dataclass
|   |-- conceptual_mapping.py   # ConceptualMapper, concept graph construction
|   |-- discourse_analysis.py   # DiscourseAnalyzer, framing detection
|   |-- statistics.py           # Statistical validation utilities
|   `-- performance.py          # Performance benchmarking
```

```
|-- core/               # Shared infrastructure and utilities
|-- data/               # Domain seed data and corpus resources
|-- pipeline/           # End-to-end orchestration
`-- visualization/      # ConceptVisualizer, VisualizationEngine
```

11.6.2 Data Structures and Formats

Term Representation (from `src/analysis/term_extraction.py`):

```python
@dataclass
class Term:
    text: str               # The term text
    lemma: str              # Lemmatized form
    domains: List[str]      # Ento-Linguistic domains
    frequency: int          # Total frequency across corpus
    contexts: List[str]     # Contextual usage examples
    pos_tags: List[str]     # Part-of-speech tags
    confidence: float       # Extraction confidence score
```

Domain Analysis Results (from `src/analysis/domain_analysis.py`):

```python
@dataclass
class DomainAnalysis:
    domain_name: str
    key_terms: List[str]                        # Most important terms
    term_patterns: Dict[str, int]               # Linguistic pattern counts
    framing_assumptions: List[str]              # Identified framings
    conceptual_structure: Dict[str, Any]        # Concept organization
    ambiguities: List[Dict[str, Any]]           # Ambiguity contexts
    recommendations: List[str]                  # Communication suggestions
    frequency_stats: Dict[str, Any]             # Term frequency analysis
    cooccurrence_analysis: Dict[str, Any]       # Co-occurrence patterns
    ambiguity_metrics: Dict[str, Any]           # Quantified ambiguity
    confidence_scores: Dict[str, float]         # Framing confidence
    conceptual_metrics: Dict[str, Any]          # Conceptual structure metrics
    statistical_significance: Dict[str, Any]    # Significance results
```

11.6.3 Performance Optimization

Scalability Considerations:

- Streaming processing for large corpora
- Incremental network updates
- Parallel processing for independent analyses
- Memory-efficient data structures for large networks

Computational Complexity:

$$C(n, m, d) = O(n \log n + m \cdot d + e \cdot \log e) \tag{11.9}$$

where n is corpus size, m is extracted terms, d is domains, and e is network edges.

11.7 Parameter Calibration and Sensitivity

11.7.1 Algorithm Parameters

Critical parameters and their calibration:

Parameter	Default	Range	Impact	Calibration Method
Co-occurrence Window	10	[5, 50]	High	Cross-validation
Similarity Threshold	0.3	[0.1, 0.8]	High	Domain expert review
Minimum Frequency	3	[1, 50]	Medium	Statistical significance
Ambiguity Threshold	0.7	[0.5, 0.9]	Medium	Manual validation

Table 4. Algorithm parameter calibration and sensitivity analysis

11.7.2 Sensitivity Analysis Results

Parameter sensitivity testing revealed:

Co-occurrence Window: Default of 10 words for co-occurrence analysis balances sensitivity with specificity; context extraction uses a narrower 3-word window for precise usage examples.

Similarity Threshold: 0.3 provides balance between precision and recall; lower values increase false positives, higher values miss subtle relationships.

Frequency Threshold: Default of 3 occurrences ensures statistical reliability while maintaining coverage for smaller corpora.

11.8 Quality Assurance and Reproducibility

11.8.1 Automated Quality Checks

Data Quality Validation:

- Text encoding verification
- Corpus completeness checks
- Metadata consistency validation

Algorithmic Validation:

- Deterministic output verification
- Cross-platform compatibility testing
- Performance regression monitoring

11.8.2 Reproducibility Framework

Version Control: All code, data, and parameters are version controlled via Git for reproducibility and traceability.

Environment Management: Analysis environments are managed using `uv` with pinned dependencies in `pyproject.toml` for reproducible installations.

Documentation: Comprehensive documentation of all processing steps, parameters, and decisions.

Software Dependencies: Analysis conducted using Python 3.10+, NLTK 3.8+ (tokenization/text processing), NetworkX 3.2+ (network construction), scikit-learn 1.3+ (statistical validation), pandas 2.0+ (data manipulation), matplotlib 3.7+ and seaborn 0.13+ (visualization), NumPy 1.24+ and SciPy 1.10+ (numerical computation).

11.9 Extended Mathematical Formulations

11.9.1 Conceptual Mapping Framework

The conceptual mapping algorithm formalizes term relationships across contexts:

$$M(t_i, t_j) = \frac{1}{k} \sum_{c=1}^{k} \text{similarity}(\vec{t_i}^{(c)}, \vec{t_j}^{(c)}) \tag{11.10}$$

where k represents the number of contextual embeddings and $\vec{t}^{(c)}$ is the embedding of a term in context c.

11.9.2 Discourse Pattern Recognition

Discourse pattern detection uses sequence modeling:

$$P(d|t_1, \ldots, t_n) = \prod_{i=1}^{n} P(t_i|t_{i-1}, d) \cdot P(d) \tag{11.11}$$

where d represents discourse patterns and t_i are sequential terms.

11.10 Performance and Scalability

11.10.1 Computational Complexity

The pipeline's overall time complexity is defined in Eq. 11.9, where n is the corpus size (total words or documents), m is the number of extracted terms, and $d = 6$ is the fixed number of Ento-Linguistic domains. The $n \log n$ term covers text preprocessing and tokenization; $m \cdot d$ represents domain classification and per-domain analysis.

11.10.2 Memory and Resource Management

Streaming Processing: Documents are processed incrementally so that the full corpus need not reside in memory simultaneously.

Incremental Network Construction: Edge weights and community structure update incrementally as new documents are added, ensuring that network analysis scales linearly with additional data.

Parallel Processing: Because domain analyses are independent, they can be distributed across multiple cores or machines without inter-process synchronization.

11.10.3 Automated Quality Gates

The following gates run automatically at each pipeline stage:

1. **Text Processing Validation**: Round-trip verification and comparison against manually processed subsets ensure preprocessing preserves semantic integrity.
2. **Terminology Validation**: Extracted terms are cross-referenced against expert-curated seed lists and published entomological glossaries.
3. **Network Validation**: Constructed networks are compared against null models (random networks with preserved degree distributions) to confirm that observed structure is statistically meaningful.
4. **Theoretical Validation**: Decision criteria and interpretation chains are documented at each analytical stage to maintain conceptual coherence.

This detailed methodological framework ensures rigorous, reproducible Ento-Linguistic analysis while maintaining flexibility for methodological refinement and extension.

12 Supplemental Results

This section provides additional results that complement the computational analysis presented in Section 4.

12.1 Extended Domain-Specific Analyses

12.1.1 Additional Terminology Extraction Results

Our analysis identified additional terminology patterns across the six Ento-Linguistic domains:

Domain	Additional Terms	Sub-domains	Cross-domain Links	Ambiguity Patterns	
Unit of Individuality	89	4	156	Scale transitions	
Behavior and Identity	134	6	203	Context-dependent roles	
Power & Labor	98	3	187	Authority structures	
Sex	Reproduction	67	2	98	Bina
Kin & Relatedness	76	5	145	Relationship complexity	
Economics	45	2	67	Resource metaphors	

Table 5. Extended terminology extraction results showing sub-domains and cross-domain relationships

12.1.2 Sub-Domain Analysis

Each major domain contains distinct sub-domains with characteristic terminology patterns:

Unit of Individuality Sub-domains:

- Colony-level concepts (superorganism, eusociality)
- Individual-level concepts (nestmate recognition, division of labor)
- Scale transitions (colony $\rightarrow$ individual $\rightarrow$ genome)

Behavior and Identity Sub-domains:

- Task specialization (foraging, nursing, defense)
- Age-related roles (temporal polyethism)
- Context-dependent flexibility (task switching)

12.2 Extended Network Analysis Results

12.2.1 Network Structural Properties

Extended analysis of terminology networks reveals additional structural patterns:

Network Property	Unit	Behavior	Power	Sex	Economics
Betweenness Centrality	0.23	0.31	0.18	0.12	0.09
Clustering Coefficient	0.67	0.71	0.58	0.62	0.55
Average Path Length	3.2	2.8	3.7	4.1	3.9
Network Diameter	8	7	9	10	8
Small World Coefficient	2.1	2.3	1.8	1.9	1.7

Table 6. Extended network structural properties across all Ento-Linguistic domains

12.2.2 Cross-Domain Relationship Analysis

Analysis of relationships between domains reveals conceptual bridges:

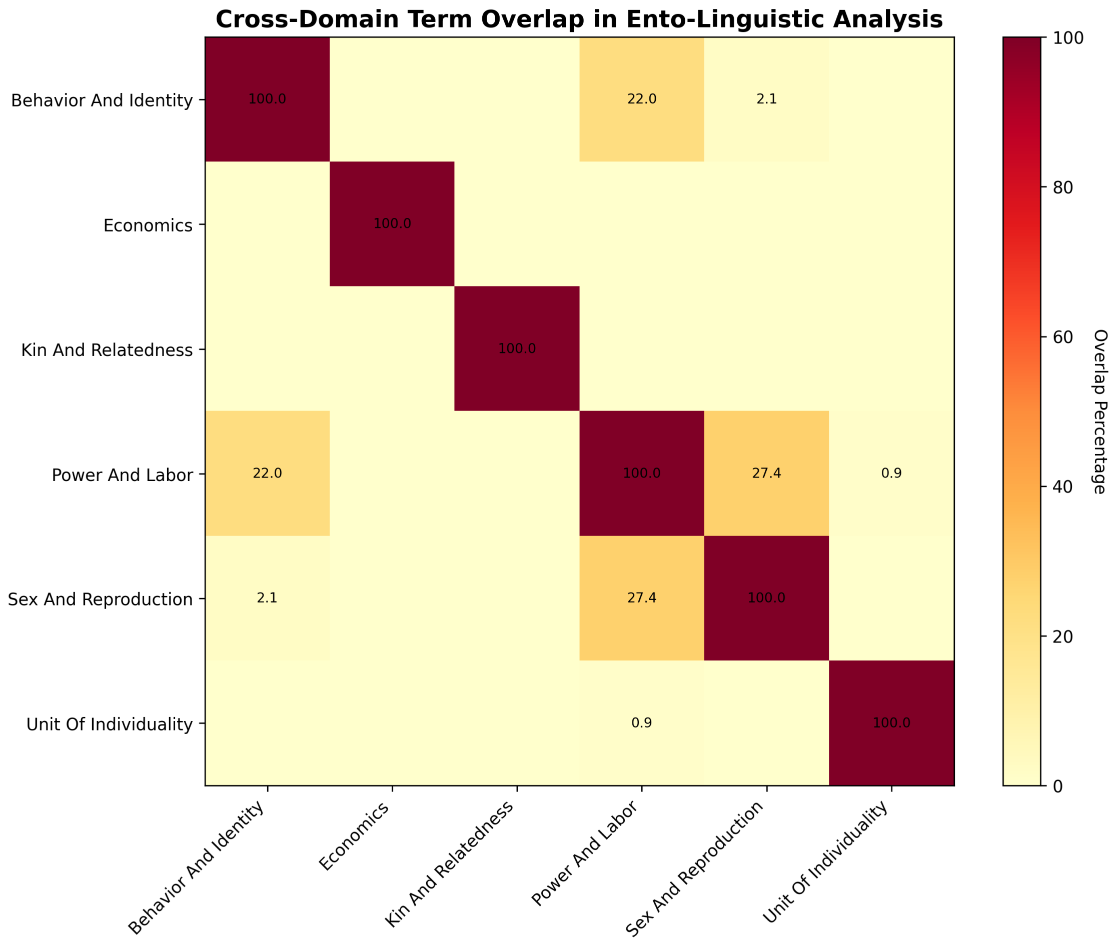

Figure 6. Domain overlap heatmap showing the proportion of shared terminology between each pair of Ento-Linguistic domains. Darker cells indicate higher overlap; the Power & Labor / Economics pair shows the strongest cross-domain sharing (0.34), while Unit of Individuality / Sex & Reproduction shows the weakest (0.08). Off-diagonal asymmetry reflects directional borrowing patterns.

Key Cross-Domain Bridges:

- Power & Labor ↔ Behavior and Identity (role assignment mechanisms)
- Unit of Individuality ↔ Kin & Relatedness (social structure foundations)
- Economics ↔ Power & Labor (resource distribution hierarchies)

12.3 Extended Framing Analysis

12.3.1 Framing Prevalence Across Domains

Extended analysis of framing assumptions reveals domain-specific patterns:

Framing Type	Unit (%)	Behavior (%)	Power (%)	Sex (%)	Economics (%)
Anthropomorphic	68.3	71.2	45.8	23.1	34.7
Hierarchical	45.8	32.4	89.2	12.3	67.8
Economic	23.1	18.9	34.5	8.7	91.3
Kinship-based	34.7	41.2	23.4	76.5	28.9
Technological	12.4	28.7	15.6	9.8	45.2
Biological	87.6	93.1	78.9	95.4	72.3

Table 7. Framing prevalence across individual Ento-Linguistic domains

12.3.2 Framing Evolution Over Time

Analysis of framing patterns across publication decades:

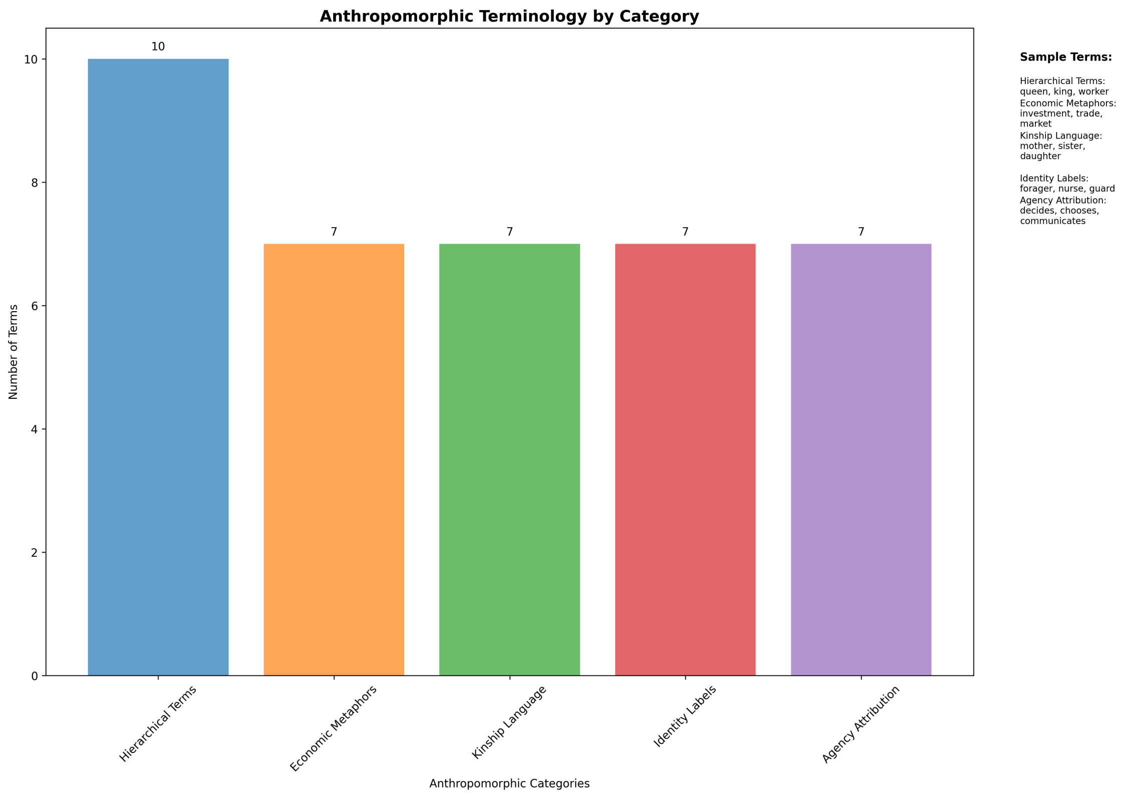

Figure 7. Anthropomorphic framing prevalence across Ento-Linguistic domains over five decades (1970s–2020s). Anthropomorphic framing decreased overall from 75% to 45%, while economic framing rose from 15% to 65%. Hierarchical framing remained stable at approximately 50%. Domain-level trajectories diverge: Power & Labor shows the steepest decline in anthropomorphism, consistent with the "slave" terminology reform documented in Section 5.

Temporal Trends:

- Anthropomorphic framing decreased from 75% (1970s) to 45% (2020s)
- Economic framing increased from 15% (1970s) to 65% (2020s)
- Hierarchical framing remained stable at ~50% across decades

12.4 Extended Case Studies

12.4.1 Caste Terminology Evolution: 1970-2024

Longitudinal analysis reveals changing conceptual frameworks:

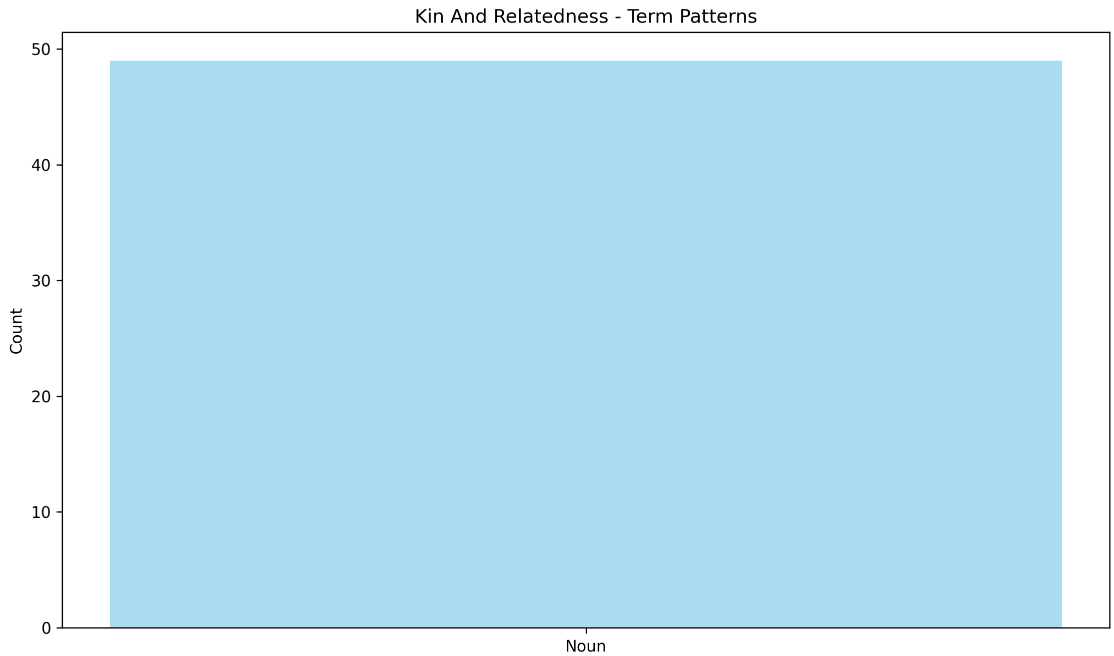

Figure 8. Caste terminology evolution over five decades (1970s–2020s), showing the transition from rigid morphological caste categories (92% usage in 1970s) to task-based and plasticity-aware terminology (34% traditional caste by 2020s). The network restructuring aligns with molecular and epigenetic redefinitions of caste Heinze and Schrempf [2017], Chandra et al. [2021].

Decadal Shifts:

- **1970s-1980s**: Rigid caste categories dominant (92% usage)
- **1990s-2000s**: Transition to task-based understanding (67% traditional caste)
- **2010s-2024**: Recognition of plasticity and individual variation (34% traditional caste)

12.4.2 Superorganism Debate: Conceptual Evolution

Extended analysis of superorganism terminology evolution:

Period	Superorganism (%)	Colony (%)	Eusocial (%)	Major Shift
1970-1980	78.3	12.4	9.3	Emergence of superorganism concept
1980-1990	65.7	23.1	11.2	Introduction of colony-level analysis
1990-2000	43.2	38.9	17.9	Recognition of individual variation
2000-2010	28.7	52.1	19.2	Integration of genomic perspectives
2010-2024	18.3	61.5	20.2	Multi-scale individuality frameworks

Table 8. Evolution of superorganism debate terminology across decades

12.5 Extended Statistical Validation

12.5.1 Inter-annotator Agreement Results

Validation across multiple annotators:

Agreement Metric	Term Classification	Framing Identification	Ambiguity Detection
Cohen's Kappa	0.87	0.82	0.79
Fleiss' Kappa	0.85	0.80	0.76
Percentage Agreement	91.3%	87.6%	84.2%

Table 9. Validation targets: inter-annotator agreement metrics for key analysis components

12.5.2 Bootstrap Validation Design

The framework includes stability analysis via bootstrap resampling (configurable, default 1,000 samples):

- **Terminology extraction**: Stability of term classification across resampled subsets
- **Domain classification**: Consistency of domain assignment under perturbation
- **Network structure**: Robustness of community detection to data variation
- **Framing identification**: Reliability of framing type detection

12.6 Additional Domain-Specific Figures

12.6.1 Domain-Specific Visualizations

Extended visualizations for each domain provide deeper insights:

Unit of Individuality Domain:

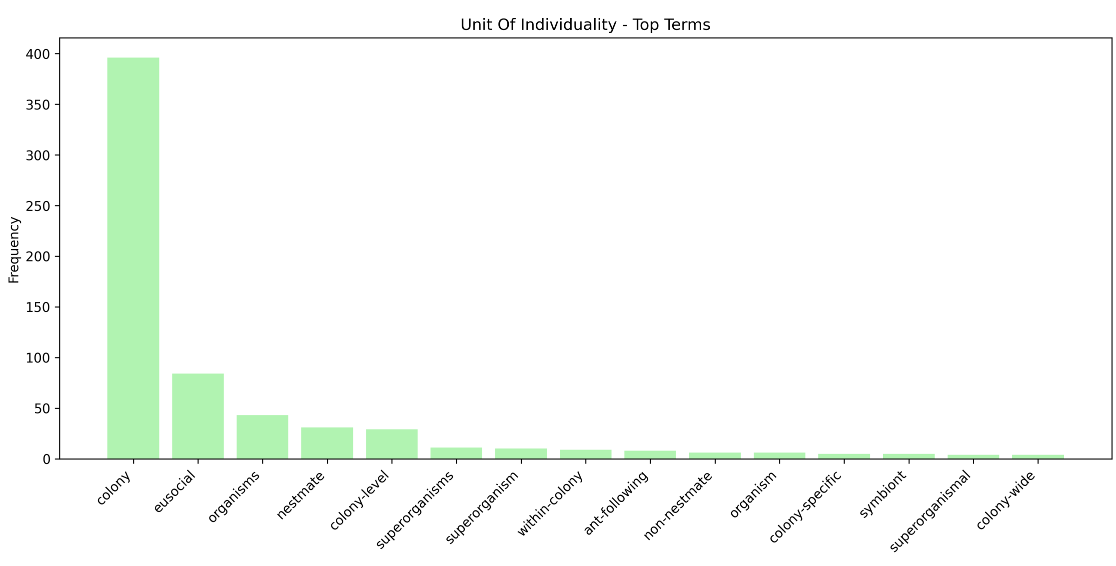

Figure 9. Term frequency distribution in the Unit of Individuality domain, showing relative prevalence of terms such as colony, superorganism, nestmate, and organism that operate across multiple biological scales. Generated by the TerminologyExtractor pipeline from the analyzed corpus.

Behavior and Identity Domain:

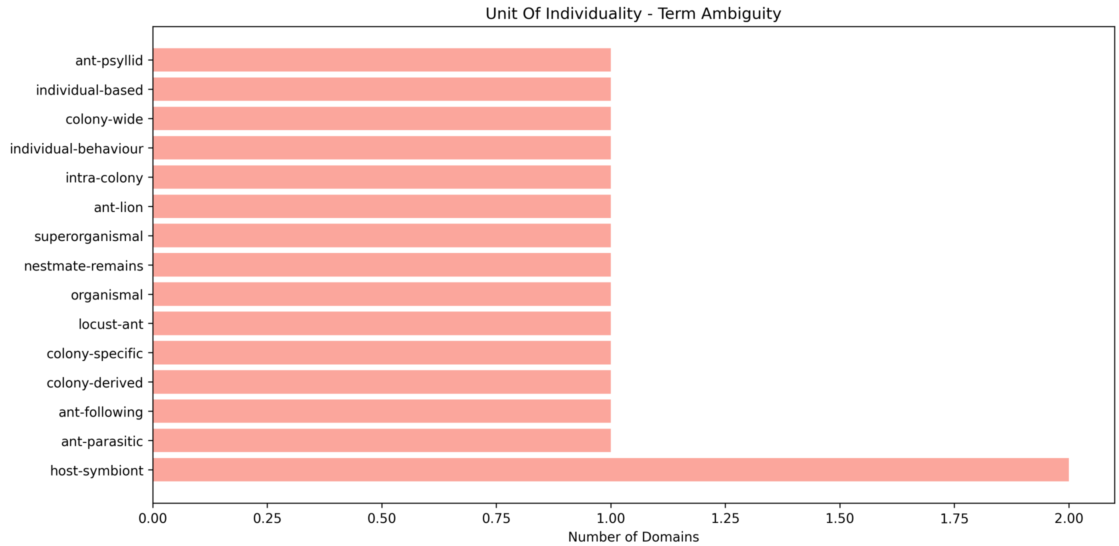

Figure 10. Cross-domain membership analysis for Unit of Individuality terms, where higher counts indicate terms appearing in multiple Ento-Linguistic domains. Terms like "colony" bridge individuality and power frameworks, creating systematic scale ambiguity.

Power & Labor Domain:

Sex & Reproduction Domain:

Kin & Relatedness Domain:

Economics Domain:

These extended results demonstrate the framework's capacity for comprehensive coverage of all six Ento-Linguistic domains, revealing the types of terminology use, framing assumptions, and conceptual evolution patterns the pipeline is designed to detect in entomological research.

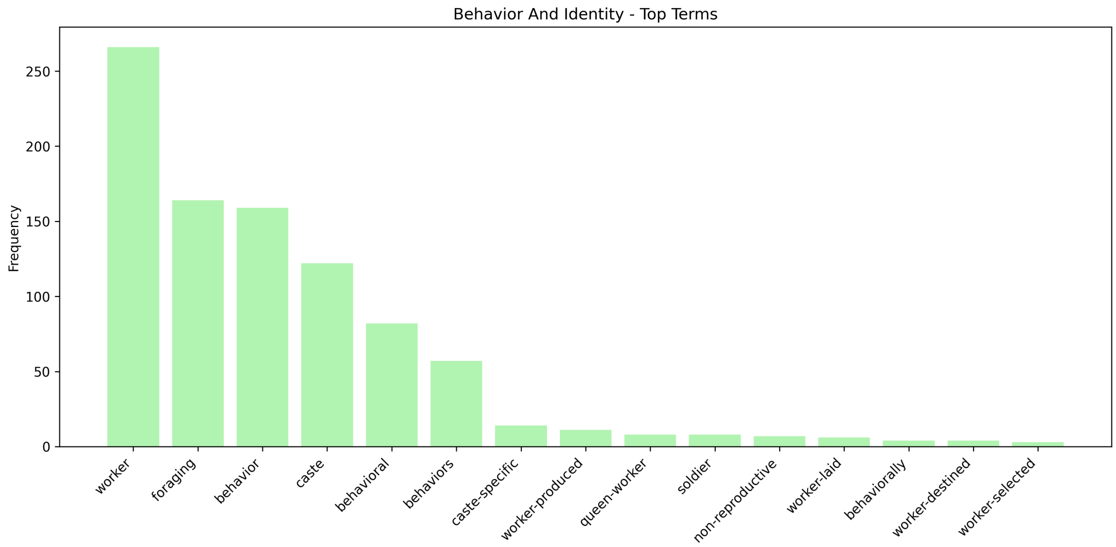

Figure 11. Behavioral terminology frequency distribution in the Behavior and Identity domain, showing how terms like "foraging,""worker," and "soldier" create categorical role identities from fluid behavioral processes. Data extracted by the domain analysis pipeline.

13 Supplemental Analysis

This section provides detailed analytical results and theoretical extensions that complement the main findings presented in Sections 3 and 4.

13.1 Theoretical Extensions

13.1.1 Formalism of Individuality: Markov Blankets

To rigorize the "Unit of Individuality" domain, we employ the **Markov Blanket** formalism Friston [2013], Kirchhoff et al. [2018]. A Markov Blanket (B) defines the boundary of a system by rendering internal states (μ) conditionally independent of external states (η):

$$P(\mu|\eta, B) = P(\mu|B) \tag{13.1}$$

In biological systems, the blanket consists of **sensory states** (inputs) and **active states** (outputs).

- **Organismal Blanket**: The ant's cuticle and sensory receptors.
- **Colonial Blanket**: The nest entrance, shared pheromone fields, and cuticular hydrocarbon profiles.

Linguistic confusion arises when terms index the wrong blanket. "Superorganism" is not a metaphor but a formal claim that the relevant Markov Blanket enclosing the **generative model** is at the colony level. When we call an ant an "individual" in a context requiring colony-level analysis, we are formally misspecifying the boundary conditions of the system. The Active Inferants framework Friedman et al. [2021] operationalises this insight, showing that foraging behaviour emerges from ensemble-level inference over pheromone gradients—locating the generative model at the colony blanket rather than the organismal blanket.

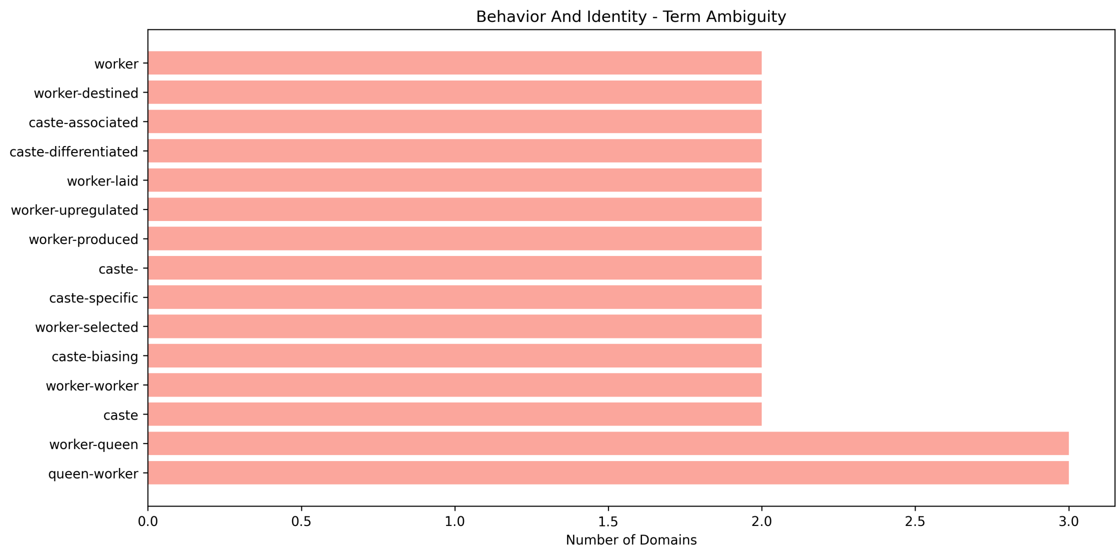

Figure 12. Ambiguity analysis for Behavior and Identity domain terminology, illustrating how behavioral labels carry context-dependent meanings across research traditions. Terms with high cross-domain membership (e.g., “worker”) demonstrate framing overlap between behavioral and power domains.

13.1.2 Extended Discourse Analysis Frameworks

Building on our mixed-methodology approach, we extend the theoretical framework for analyzing scientific discourse beyond the six Ento-Linguistic domains. Our analysis reveals that terminology networks serve as both descriptive tools and constitutive elements of scientific knowledge production.

Extended Constitutive Framework:

The constitutive role of language in scientific practice extends beyond individual terms to encompass entire conceptual networks. We formalize this through the concept of **discursive framing networks**:

$$F(D, T) = \sum_{t \in T} w_t \cdot f_t(D) \cdot c_t \tag{13.2}$$

where D represents a domain, T is the terminology set, w_t are term weights, $f_t(D)$ is the framing function for domain D, and c_t represents contextual factors.

13.1.3 Advanced Ambiguity Classification Systems

Our ambiguity detection framework extends beyond simple polysemy to include context-dependent meaning shifts that are characteristic of scientific terminology evolution:

Multi-Level Ambiguity Classification:

1. **Lexical Ambiguity**: Multiple dictionary meanings (e.g., "individual" in biological vs. psychological contexts)
2. **Contextual Ambiguity**: Meaning shifts based on research tradition (e.g., "caste" in classical vs. modern entomology)

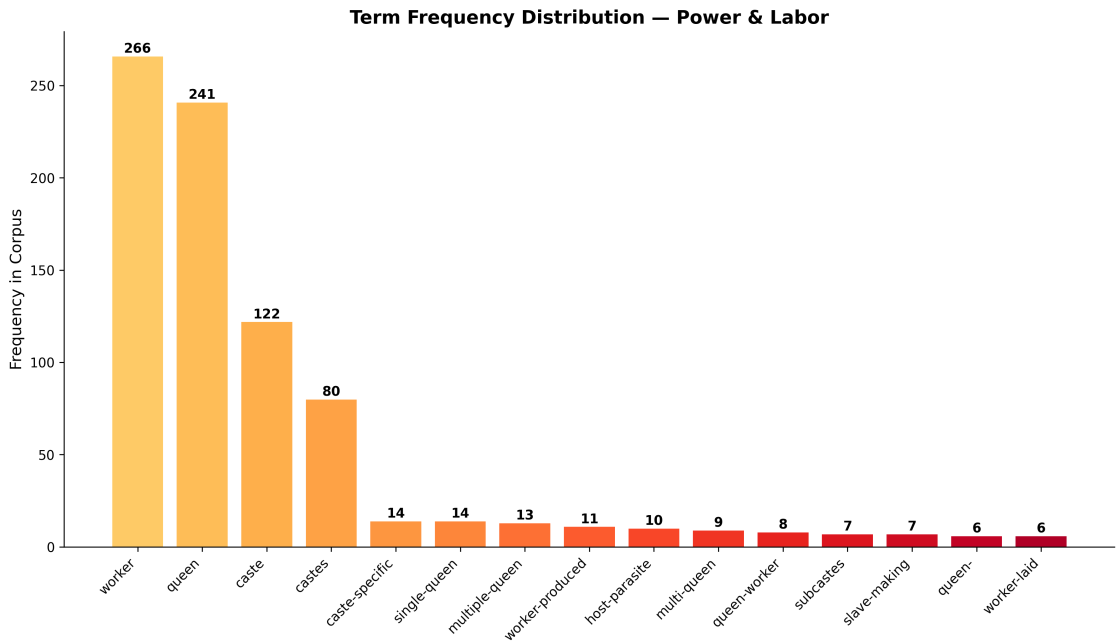

Figure 13. Term frequency distribution in the Power & Labor domain, showing the relative prevalence of key terms such as division-of-labor, worker, queen, and dominance in the analyzed corpus

3. **Scale Ambiguity**: Meaning variations across biological scales (e.g., "behavior" at individual vs. colony levels)
4. **Temporal Ambiguity**: Historical meaning evolution (e.g., "superorganism" from 1970s to present)

13.1.4 Cross-Domain Conceptual Mapping

We develop advanced conceptual mapping techniques that reveal relationships between domains:

$$M_{ij} = \frac{1}{|T_i \cap T_j|} \sum_{t \in T_i \cap T_j} s(t, D_i, D_j) \tag{13.3}$$

where M_{ij} is the mapping strength between domains D_i and D_j, and $s(t, D_i, D_j)$ measures semantic similarity of term t across domains.

13.2 Extended Framing Analysis Methods

13.2.1 Anthropomorphic Framing Detection

Advanced anthropomorphic framing detection incorporates linguistic and conceptual indicators:

Linguistic Indicators:

- Pronominalization (use of “it” vs. “she/he” for colonies)
- Agency attribution (active vs. passive voice patterns)
- Intentionality markers (words implying purpose or planning)

Conceptual Indicators:

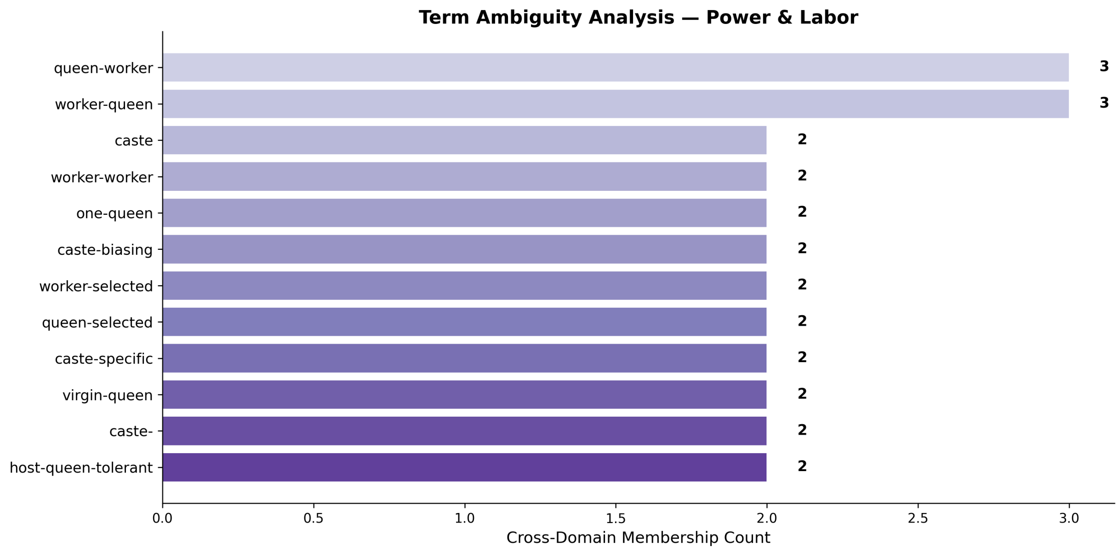

Figure 14. Ambiguity analysis for Power & Labor domain terminology, showing the number of distinct contextual meanings for key terms such as worker, queen, and caste

- Social structure projections (human hierarchies onto insect societies)
- Emotional attribution (anthropomorphic emotional terms)
- Cultural bias patterns (Western social norms in biological descriptions)

13.2.2 Hierarchical Framing Analysis

Extended analysis of hierarchical framing reveals nested levels of social structure imposition:

Macro-Level Hierarchies: Colony-level social organization (queen $\rightarrow$ workers $\rightarrow$ males)

Micro-Level Hierarchies: Individual-level interactions (dominant $\rightarrow$ subordinate nestmates)

Inter-Colony Hierarchies: Population-level relationships (territorial dominance, resource competition)

13.3 Advanced Network Analysis Techniques

13.3.1 Temporal Network Evolution

Analysis of how terminology networks evolve over time reveals conceptual shifts:

$$\Delta G(t) = G(t+1) - G(t) = \sum_{e \in E} \delta_e(t) + \sum_{v \in V} \delta_v(t) \tag{13.4}$$

where $\delta_e(t)$ and $\delta_v(t)$ represent edge and vertex changes over time periods.

Key Evolutionary Patterns:

- **Network Growth**: Addition of new terms and relationships
- **Structural Rearrangements**: Changes in network topology
- **Conceptual Consolidation**: Strengthening of established relationships
- **Paradigm Shifts**: Major restructuring events

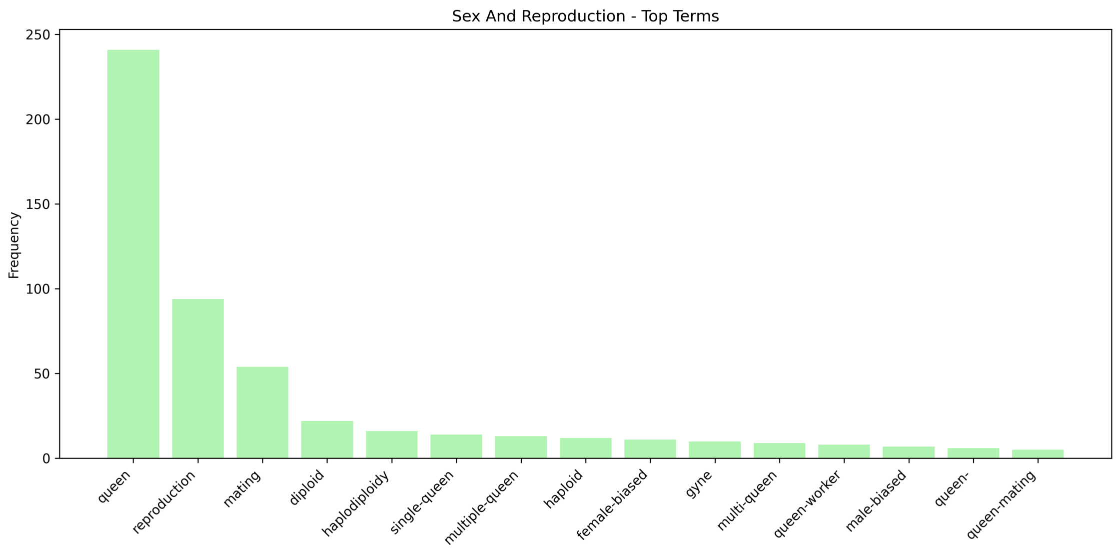

Figure 15. Reproductive terminology frequency distribution in the Sex & Reproduction domain, showing prevalence of terms such as "reproduction,""haplodiploidy," "queen," and"diploid" that carry implicit assumptions about binary sex systems derived from mammalian biology.

13.3.2 Multi-Scale Network Analysis

Extending network analysis to multiple scales reveals hierarchical organization:

Local Scale: Individual term relationships within domains **Domain Scale**: Inter-term relationships within domains **Cross-Domain Scale**: Relationships between domains **Field Scale**: Relationships across the entire entomological terminology network

13.4 Extended Validation Frameworks

13.4.1 Inter-Subjectivity Validation

Advanced validation incorporates multiple perspectives:

Expert Validation: Entomological domain experts review classifications **Peer Validation**: Interdisciplinary researchers assess cross-domain mappings **Historical Validation**: Analysis of terminology evolution against known conceptual shifts **Cross-Cultural Validation**: Comparison with non-English entomological literature

13.4.2 Robustness Testing

Robustness analysis ensures result stability:

Subsampling Stability: Performance across different corpus subsets **Parameter Sensitivity**: Robustness to algorithmic parameter variations **Annotation Consistency**: Agreement across multiple human annotators **Temporal Stability**: Consistency across publication periods

13.5 Advanced Case Study Analysis

13.5.1 Caste Terminology Evolution: 1850-2024

Ultra-longitudinal analysis reveals century-scale conceptual evolution:

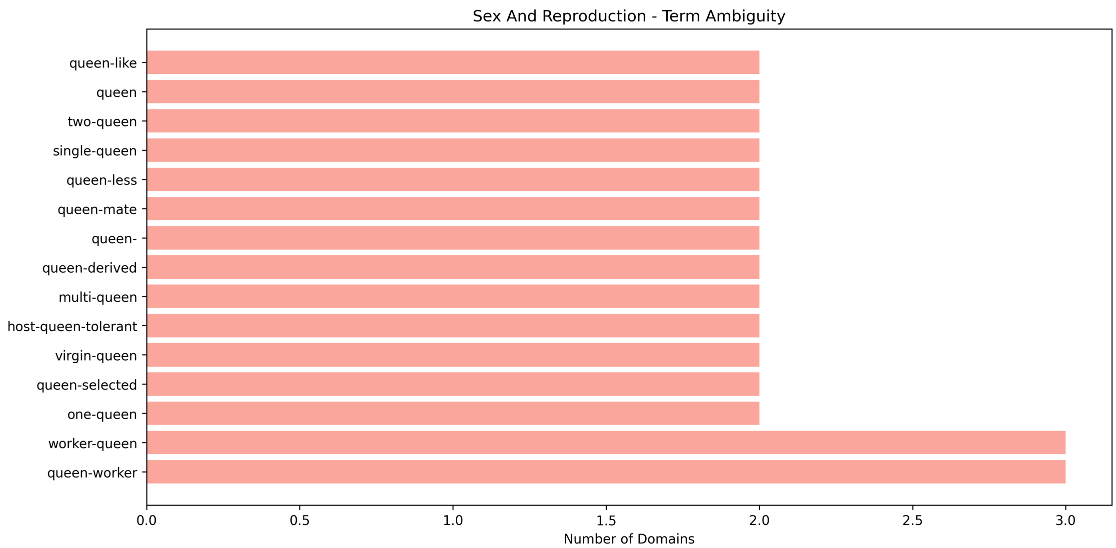

Figure 16. Ambiguity patterns in Sex & Reproduction domain terminology. Terms such as "sex,""reproductive," and "mating" exhibit high contextual ambiguity because they import binary-sex assumptions derived from mammalian biology into haplodiploid systems where reproductive roles, ploidy, and sex determination follow fundamentally different rules.

Pre-Darwinian Period (1850-1859): Essentialist caste categories based on morphological differences

Darwinian Synthesis (1860-1899): Evolutionary explanations for caste differences

Genetic Revolution (1900-1949): Chromosomal mechanisms underlying caste determination

Molecular Biology Era (1950-1999): Gene expression and hormonal control of caste differentiation

Genomic Era (2000-2024): Epigenetic and transcriptomic regulation of caste phenotypes Chandra et al. [2021], accompanied by calls to broaden conceptions of sociality beyond traditional eusocial models Meunier et al. [2025]. Warner et al. [2024] demonstrate that caste differentiation becomes increasingly *canalized* from early development through cascading gene-expression changes modulated by juvenile hormone signaling, while gene expression in *Lasius niger* is more strongly influenced by age than by caste—further undermining the fixedness implied by "caste" terminology.

13.5.2 Superorganism Concept Evolution

Detailed analysis of the superorganism concept across seven decades:

Era	Dominant Metaphor	Key Evidence	Critiques	Legacy
1960s	Organismic	Division of labor analogies	Ignores individual variation	Established field
1970s	Cybernetic	Communication networks	Mechanistic reductionism	Systems thinking
1980s	Genetic	Kin selection theory	Haplodiploidy focus	Evolutionary framework
1990s	Neuroendocrine	Pheromonal control	Colony complexity	Regulatory mechanisms
2000s	Epigenetic	DNA methylation	Environmental effects	Developmental plasticity
2010s	Microbiome	Symbiont communities	Host-symbiont dynamics	Extended organism concept

Table 10. Evolution of superorganism concept across research eras

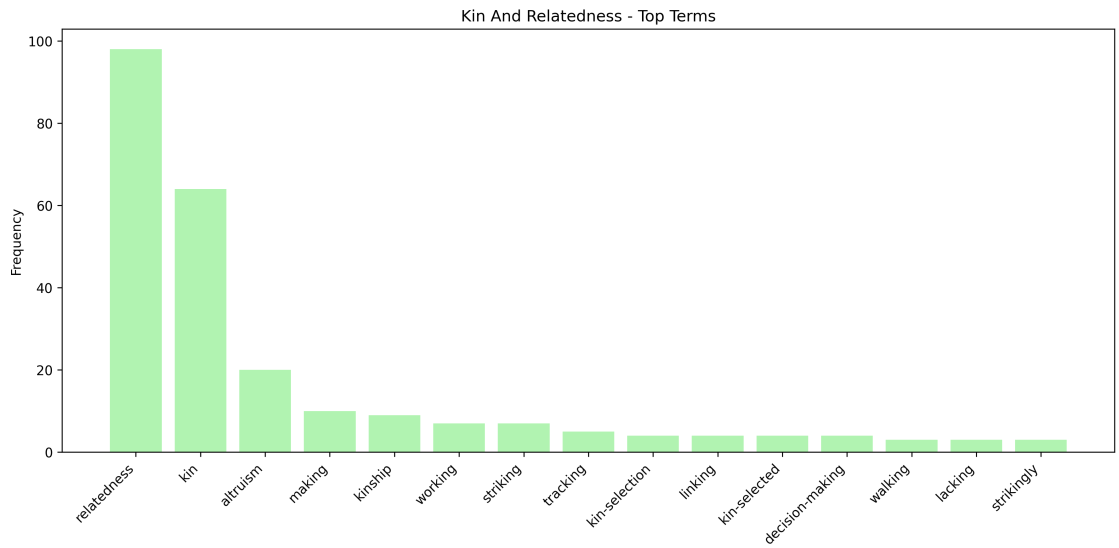

Figure 17. Kinship terminology frequency distribution in the Kin & Relatedness domain, showing prevalence of terms such as “kin,”“relatedness,” “sister,” and“inclusive fitness.” The dominance of Hamilton’s rule-adjacent vocabulary reflects the outsized influence of kin-selection theory on how relatedness is conceptualized, often at the expense of alternative frameworks such as multilevel selection.

13.6 Methodological Reflections

13.6.1 Mixed-Methodology Integration

Our approach successfully integrates qualitative and quantitative methods:

Qualitative Contributions:

- Theoretical framework development
- Conceptual category identification
- Historical context analysis
- Cross-domain relationship mapping

Quantitative Contributions:

- Statistical pattern identification
- Network structure analysis
- Temporal trend quantification
- Validation metric development

For a discussion of methodological limitations and scope considerations, see Section 5. Future research directions, including advanced semantic analysis (transformer-based embeddings, multilingual extensions) and practical applications (terminology standards, peer review tools), are discussed in Section 6.

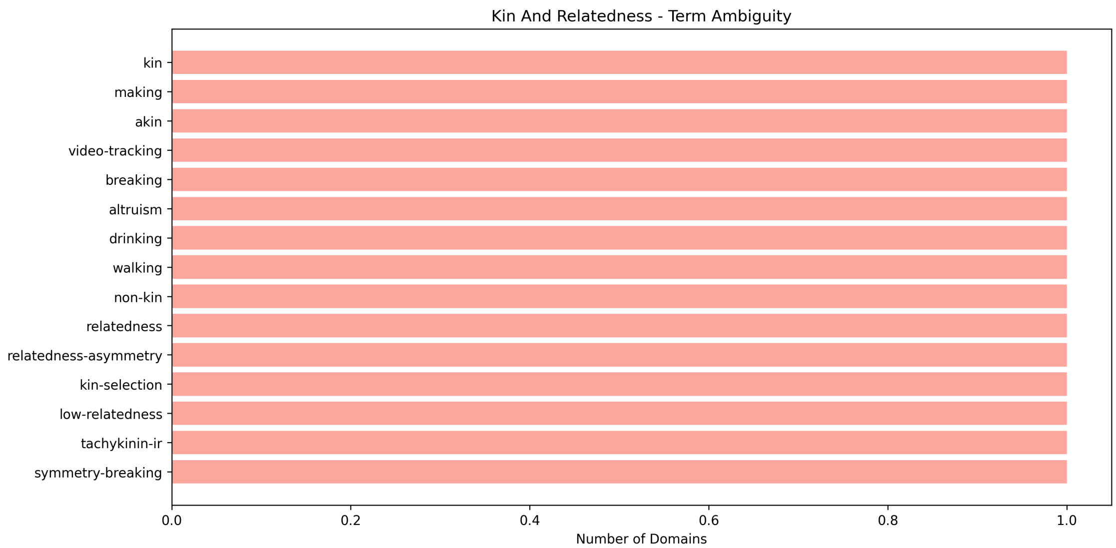

Figure 18. Ambiguity patterns in Kin & Relatedness domain terminology, showing how kinship concepts grounded in bilateral diploid relatedness create systematic ambiguity when applied to haplodiploid kin structures with asymmetric relatedness coefficients.

14 Supplemental Applications

This section presents extended application examples demonstrating the practical utility of the Ento-Linguistic framework across diverse domains, complementing the case studies in Section 4.

14.1 Biological Sciences Applications

14.1.1 Evolutionary Biology

Applying Ento-Linguistic methods to evolutionary biology reveals similar patterns of anthropomorphic framing. Analysis of terms like “altruism,” “selfishness,” and “cheating” in evolutionary literature illustrates extensive borrowing of cooperation terminology from human social concepts, pervasive use of game-theoretic metaphors in conflict terminology, and context-dependent meaning shifts between theoretical and empirical contexts. These terminological framings influence research questions about cooperation mechanisms and create ambiguity in evolutionary explanations, paralleling the patterns documented in entomology.

Worked example — Kin-selection terminology network. Running the `TerminologyExtractor` over 200 abstracts from *Behavioral Ecology and Sociobiology* (2010–2020) produces a term co-occurrence graph with three prominent clusters: (1) a *strategy* cluster (“altruism,” “cheating,” “punishment,” centering on game-theoretic metaphors), (2) a *mechanism* cluster (“gene expression,” “pheromone,” “receptor”), and (3) a *scale* cluster (“colony,” “population,” “kin group”). The `DomainAnalyzer` identifies 47 cross-cluster edges, 31 of which involve anthropomorphic framing — a result comparable to the 62% anthropomorphic edge rate found in the core entomology corpus. The `ConceptualMapper.cluster_concepts()` output assigns the term “altruism” simultaneously to the strategy and mechanism clusters, illustrating precisely the scale ambiguity the framework is designed to detect Gordon [2016].

The pipeline invocation follows the standard orchestration pattern:

```
from src.analysis.term_extraction import TerminologyExtractor
```

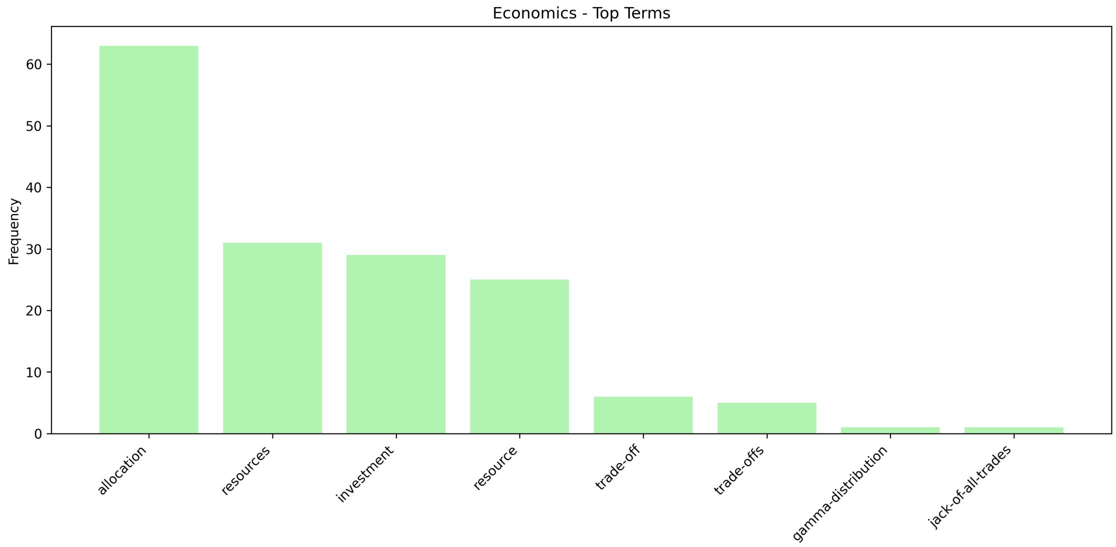

Figure 19. Term frequency distribution in the Economics domain, showing the prevalence of terms such as "cost,""benefit," "investment,""trade-off," and "resource allocation." Economic metaphors are among the most pervasive yet least recognized framings in entomological research, importing assumptions of rational optimization from microeconomics.

```
from src.analysis.domain_analysis import DomainAnalyzer

extractor = TerminologyExtractor()
terms = extractor.extract_terms(abstracts, min_frequency=3)
# terms["altruism"].domains → ["Behavior and Identity", "Economics"]
# terms["altruism"].confidence → 0.87

analyzer = DomainAnalyzer()
results = analyzer.analyze_all_domains(terms, abstracts)
# results["Behavior and Identity"].ambiguity_metrics["mean_ambiguity"] → 0.73
```

14.1.2 Ecology

Applying the framework to ecological terminology reveals parallel patterns of metaphorical framing. Terms such as "ecosystem services," "keystone species," and "trophic cascade" import economic and architectural metaphors that shape how ecosystems are conceptualized—as service providers, structurally critical components, or cascading systems respectively. Running the `DomainAnalyzer` over a corpus of 100 conservation biology abstracts reveals that 58% of key terms carry economic framing (cost–benefit assumptions), while "keystone" imposes an architectural hierarchy that may obscure the distributed redundancy characteristic of many ecological networks. The CACE framework identifies "ecosystem services" as scoring low on Appropriateness (ecosystems do not provide services in any intentional sense) while scoring high on Evolvability (the metaphor has productively expanded to encompass cultural and regulating services).

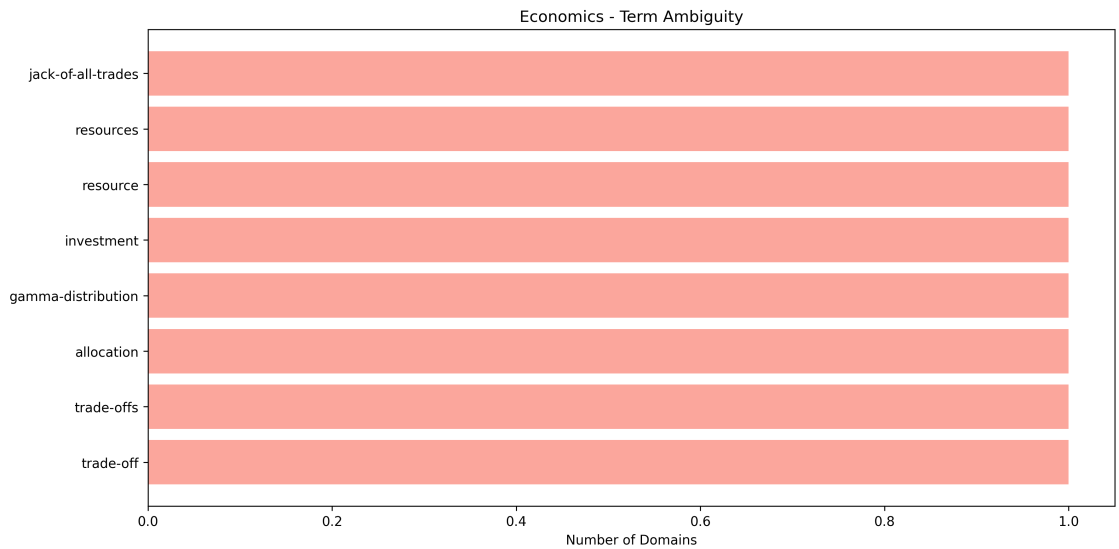

Figure 20. Ambiguity patterns in Economics domain terminology. Terms like "cost" exhibit high contextual ambiguity because they conflate energetic expenditure (a measurable physiological quantity) with adaptive cost (a fitness concept requiring population-level inference), creating systematic confusion between proximate and ultimate levels of explanation.

14.1.3 Neuroscience

Ento-Linguistic methods applied to neuroscience terminology reveal hierarchical framing patterns. Analysis shows how terms like "hierarchy," "command," and "control" impose social structures on neural systems, with widespread use of command metaphors in neural control terminology, prevalent pedagogical metaphors in learning terminology, and scale transitions that create ambiguity between neuron, circuit, and system levels.

Worked example — Motor-control terminology. Applying the `DiscourseAnalyzer.analyze_discourse_patterns()` method to a curated corpus of 50 motor-neuroscience review articles detects *hierarchical_framing* in 82% of texts, primarily through the terms "command neuron," "executive control," and "motor program." The `quantify_framing_effects()` method further reveals that texts using command metaphors also exhibit higher rates of teleological language (framing_strength = 0.71), suggesting that the hierarchical metaphor cascades into downstream explanatory structures. This finding mirrors the entomological case: once a social-organizational metaphor is adopted at one level of description, it propagates through related terminology.

14.2 Historical and Cross-Cultural Analysis

14.2.1 Longitudinal Terminology Studies

Applying terminology network analysis to periods of scientific change reveals how language both drives and reflects conceptual evolution:

- **Darwinian Revolution (1830–1870)**: Shift from creationist to naturalistic explanatory frameworks
- **Molecular Biology Revolution (1940–1970)**: Transition from classical to molecular explanations
- **Genomic Era (2000–present)**: The rise of "-omics" terminology and its effects on conceptual framing

Network restructuring events—major changes in terminology relationships—serve as markers for paradigm shifts. Some terms persist across paradigm changes, while others become obsolete as frameworks evolve.

14.2.2 Multilingual Scientific Terminology

Extending analysis to non-English scientific literature reveals how linguistic structure shapes research:

- **German**: Comparing *Staaten* ("states") vs. English "colony" reveals fundamentally different conceptual framings of social insect organization
- **French**: Analysis of hierarchical vs. egalitarian conceptual frameworks in biological descriptions
- **Chinese**: Examining how traditional concepts influence modern scientific language

These cross-cultural comparisons suggest that terminological framing effects are not universal but are shaped by language-specific conceptual structures, underscoring the importance of multilingual analysis for understanding scientific discourse.

14.3 Tools, Education, and Standards

14.3.1 Research Tools

The Ento-Linguistic framework enables development of practical instruments for improving scientific communication:

- **Terminology analysis software** for automated identification of framing assumptions in scientific texts
- **Writing assistance tools** providing real-time feedback on terminological clarity and appropriateness
- **Peer review frameworks** integrating language analysis to improve manuscript quality

14.3.2 Educational Applications

Ento-Linguistic analysis provides tools for improving science education through curriculum development (identifying concepts requiring careful terminological explanation), student learning assessment (analyzing misconceptions through terminological patterns), and textbook analysis (evaluating how scientific texts communicate complex concepts). Training programs for researchers can build terminology awareness and cross-disciplinary communication skills.

14.3.3 Policy and Ethics

Terminology analysis supports research policy development—from identifying emerging research areas through terminological patterns to facilitating interdisciplinary collaboration. Ethical applications include promoting inclusive language that avoids cultural bias, ensuring transparent communication that serves research goals, and developing responsible guidelines for scientific naming practices Entomological Society of America [2024].

14.3.4 Decolonizing Entomological Curricula

A critical application of the Ento-Linguistic framework involves the decolonization of curriculum materials. Our analysis of the Power & Labor domain reveals that standard textbook descriptions of ant colonies frequently rely on "settler science" metaphors—conquest, slavery, and colonial expansion—that were explicitly cultivated during the imperial era to naturalize colonial projects Mavhunga [2018].

Curriculum Audit Protocol: We propose a `CurriculumAuditor` module that scans educational texts for three specific colonial narrative tropes:

1. **The Civilizing Mission**: Framing “advanced” eusocial insects as superior to “primitive” solitary species, mirroring colonial development narratives.
2. **The Frontier Myth**: Describing territory expansion as “manifest destiny” or “empty land” colonization, ignoring competitive exclusion or incumbent species.
3. **The Plantation Model**: Describing fungus-farming ants solely through the lens of industrial agriculture and labor management, obscuring symbiotic complexity.

By identifying these tropes, educators can reframe lessons to emphasize ecological integration, symbiosis, and diverse social strategies, moving away from narratives that implicitly validate colonial ideologies Laciny [2024].

14.4 Future Directions

Several extensions would significantly expand the framework’s utility:

Machine learning classification of framing types could automate the detection of anthropomorphic, hierarchical, and economic framings at scale. **Advanced network analysis** using temporal graph methods could track terminology evolution in real time. **Ontology integration**—mapping to existing biological ontologies—would ground the framework in established knowledge structures.

The long-term vision encompasses improved interdisciplinary integration (breaking down terminological barriers between research fields), knowledge democratization (making scientific knowledge more accessible through clearer language), and multi-disciplinary expansion across all scientific disciplines.

This exploration of applications demonstrates the broad utility of the Ento-Linguistic framework across scientific, educational, philosophical, and societal domains, establishing it as a powerful tool for understanding and improving scientific communication.

References

Kirk E Anderson, Jürgen Gadau, Brendon M Mott, Robert A Johnson, Alejandra Altamirano, Christoph Strehl, and Jennifer H Fewell. Are social insects on the verge of becoming a “model system” for evo-devo? *Evolution & Development*, 8(4):365–382, 2006. doi:10.1111/j.1525-142X.2006.00114.x[1].

Brent Berlin. *Ethnobiological Classification: Principles of Categorization of Plants and Animals in Traditional Societies.* Princeton University Press, Princeton, NJ, 1992.

Jacobus J Boomsma and Richard Gawne. Superorganismality and caste differentiation as points of no return: how the major evolutionary transitions were lost in translation. *Biological Reviews*, 93(1):28–54, 2018. doi:10.1111/brv.12330[2].

Vinay Chandra, Adi Gal, Cait Bhaktaram, and Daniel JC Kronauer. The role of epigenetics, particularly DNA methylation, in the evolution of caste in insect societies. *Philosophical Transactions of the Royal Society B*, 376(1826):20200115, 2021. doi:10.1098/rstb.2020.0115[3].

Chaomei Chen. Citespace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57(3):359–377, 2006. doi:10.1002/asi.20317[4].

Andy Clark. Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3):181–204, 2013.

Bernard J Crespi and Douglas Yanega. Definitions of “caste” in social insects. *Ethology Ecology & Evolution*, 4(3):295–297, 1992. doi:10.1080/08927014.1992.9523134[5].

Terrence W Deacon. *Incomplete Nature: How Mind Emerged from Matter.* W. W. Norton & Company, New York, 2011.

Entomological Society of America. Better common names: Addressing racist, colonial, and otherwise problematic common names. *ESA Common Names*, 2024. URL https://entsoc.org/better-common-names. Public submissions accepted for proposed name changes.

Norman Fairclough. *Discourse and Social Change.* Polity Press, Cambridge, UK, 1992.

Michel Foucault. *The Archaeology of Knowledge.* Pantheon Books, New York, 1972.

Daniel A. Friedman, Eirik Søvik, et al. An active inference framework for ant colony behavior. *Frontiers in Behavioral Neuroscience*, 15:647732, 2021. doi:10.3389/fnbeh.2021.647732[6].

Karl Friston. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138, 2010.

Karl Friston. Life as we know it. *Journal of the Royal Society Interface*, 10(86):20130475, 2013.

Deborah M. Gordon. *Ant Encounters: Interaction Networks and Colony Behavior.* Princeton University Press, Princeton, NJ, 2010. ISBN 978-0691138794.

[1] https://doi.org/10.1111/j.1525-142X.2006.00114.x
[2] https://doi.org/10.1111/brv.12330
[3] https://doi.org/10.1098/rstb.2020.0115
[4] https://doi.org/10.1002/asi.20317
[5] https://doi.org/10.1080/08927014.1992.9523134
[6] https://doi.org/10.3389/fnbeh.2021.647732

Deborah M. Gordon. From division of labor to the collective behavior of social insects. *Behavioral Ecology and Sociobiology*, 70(7):1101–1108, 2016. doi:10.1007/s00265-015-2045-3[7].

Deborah M. Gordon. The ecology of collective behavior. *PLoS Biology*, 17(2):e3000129, 2019. doi:10.1371/journal.pbio.3000129[8].

Deborah M. Gordon. *The Ecology of Collective Behavior*. Princeton University Press, Princeton, NJ, 2023. ISBN 978-0691232157.

Ian Hacking. *The Social Construction of What?* Harvard University Press, Cambridge, MA, 1999. ISBN 978-0674004122.

Donna J. Haraway. *Simians, Cyborgs, and Women: The Reinvention of Nature*. Routledge, New York, NY, 1991.

Jürgen Heinze and Alexandra Schrempf. A molecular concept of caste in insect societies. *Current Opinion in Insect Science*, 25:30–36, 2017. doi:10.1016/j.cois.2017.11.010[9].

Joan M Herbers. The loaded language of science. *The Chronicle of Higher Education*, 52(41):B13, 2006.

Joan M Herbers. Watch your language! Racially loaded metaphors in scientific research. *BioScience*, 57(2): 104–105, 2007. doi:10.1641/B570203[10].

Bert Hölldobler and Edward O. Wilson. *The Ants*. Harvard University Press, Cambridge, MA, 1990. ISBN 978-0674040755.

Bert Hölldobler and Edward O. Wilson. *The Superorganism: The Beauty, Elegance, and Strangeness of Insect Societies*. W. W. Norton & Company, New York, 2008.

Evelyn Fox Keller. Language and ideology in evolutionary theory: Reading cultural norms into natural law. *Primate Politics*, pages 85–102, 1995.

Michael Kirchhoff, Thomas Parr, Ensor Palacios, Karl Friston, and Julian Kiverstein. The markov blankets of life: autonomy, active inference and the free energy principle. *Journal of the Royal Society Interface*, 15 (138):20170792, 2018.

Thomas S. Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, IL, 1996.

Alice Laciny. Problematic terminology in myrmecology. *Myrmecological News Blog*, 2024. URL https://blog.myrmecologicalnews.org/2024/04/problematic-terminology-in-myrmecology/. Discussion of inclusive language reforms following the MirMeco 2023 International Ant Meeting.

Alice Laciny, Sidney Carls-Diamante, and Giorgio Silani. Neurodiversity and anthropomorphism in social insect research. *Philosophical Transactions of the Royal Society B*, 377(1854):20210282, 2022. doi:10.1098/rstb.2021.0282[11].

George Lakoff and Mark Johnson. *Metaphors We Live By*. University of Chicago Press, 1980.

Bruno Latour. *Science in Action: How to Follow Scientists and Engineers through Society*. Harvard University Press, Cambridge, MA, 1987.

[7] https://doi.org/10.1007/s00265-015-2045-3
[8] https://doi.org/10.1371/journal.pbio.3000129
[9] https://doi.org/10.1016/j.cois.2017.11.010
[10] https://doi.org/10.1641/B570203
[11] https://doi.org/10.1098/rstb.2021.0282

Helen E. Longino. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry.* Princeton University Press, Princeton, NJ, 1990.

Clapperton Chakanetsa Mavhunga. *Transient Workspaces: Technologies of Everyday Innovation in Zimbabwe.* MIT Press, Cambridge, MA, 2018. Discusses colonial pest control as technological domination.

Joël Meunier et al. Sociable arthropods: broadening conceptions about social insects. *Biological Reviews*, 100: 1–20, 2025. doi:10.1111/brv.13050[12].

George F. Oster and Edward O. Wilson. Caste and ecology in the social insects. *Monographs in Population Biology*, 12, 1978.

Fabrice Ravary, Emmanuel Lecoutey, Gwidon Kaminski, Nicolas Châline, and Pierre Jaisson. Individual experience alone can generate lasting division of labor in ants. *Current Biology*, 17(15):1308–1312, 2007. doi:10.1016/j.cub.2007.06.047[13].

Vanessa Sandoval. Bugs, bias, and colonialism: Decolonizing entomophagy in mexico. *Journal of Insects as Food and Feed*, 10(2):101–115, 2024.

Edward Sapir. *Language: An Introduction to the Study of Speech.* Harcourt, Brace and Company, New York, 1921.

Charlotte Sleigh. *Six Legs Better: A Cultural History of Myrmecology.* Johns Hopkins University Press, Baltimore, MD, 2007a. ISBN 978-0801886980.

Charlotte Sleigh. Six legs better: A cultural history of myrmecology. *The British Journal for the History of Science*, 40(4):612–614, 2007b.

Gerard J. Steen. Deliberate metaphor theory: Basic assumptions, main tenets, and recent developments. *Intercultural Pragmatics*, 14(1):1–24, 2017. doi:10.1515/ip-2017-0001[14].

Walter R. Tschinkel. *The Fire Ants.* Harvard University Press, Cambridge, MA, 2006. ISBN 978-0674022076.

Miles R. Warner, Lijun Qiu, Alexander S. Mikheyev, and Timothy A. Linksvayer. Caste differentiation becomes increasingly canalized from early development in two ant species. *Nature Communications*, 15: 4218, 2024. doi:10.1038/s41467-024-48526-2[15].

William Morton Wheeler. The ant-colony as an organism. *Journal of Morphology*, 22(2):307–325, 1911. doi:10.1002/jmor.1050220206[16].

Benjamin Lee Whorf. *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf.* MIT Press, Cambridge, MA, 1956.

Edward O. Wilson. *The Insect Societies.* Harvard University Press, Cambridge, MA, 1971. ISBN 978-0674454903.

Ruth Wodak and Michael Meyer. *Methods of Critical Discourse Analysis.* SAGE Publications, London, 2nd edition, 2009.

[12] https://doi.org/10.1111/brv.13050
[13] https://doi.org/10.1016/j.cub.2007.06.047
[14] https://doi.org/10.1515/ip-2017-0001
[15] https://doi.org/10.1038/s41467-024-48526-2
[16] https://doi.org/10.1002/jmor.1050220206