Ento-Linguistics: Language, Ambiguity, and Scientific Communication in Entomology

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

Daniel Ari Friedman
Active Inference Institute, APOIDEAS
`daniel@activeinference.institute` · ORCID: 0000-0001-6232-9096

Tucker Cahill Chambers
APOIDEAS
ORCID: 0009-0008-3793-7872

April 15, 2026

DOI: 10.5281/zenodo.19574118

Contents

1 Abstract

Scientific language does not merely describe biological phenomena; it actively constitutes the generative models through which researchers parse complex systems. This paper makes three core contributions to understanding—and correcting—the epistemic consequences of this constitutive role. First, we introduce a six-domain Ento-Linguistic framework that decomposes the terminological landscape of insect research into analytically tractable themes, isolating domains where anthropomorphic language most severely distorts causal modeling. Second, we develop an open-source computational pipeline that integrates automated term extraction, co-occurrence network construction, and information-theoretic ambiguity scoring with principles from Active Inference and Complex Systems Theory. Third, we propose and validate four evidence-based meta-standards—Clarity, Appropriateness, Consistency, and Evolvability (CACE)—as a formalized protocol for lexical engineering. Analysis of a corpus encompassing 369 entomological publications (48787 tokens; 7105 unique token types; Type–Token Ratio 0.1456) extracts 888 candidate terms (with 261 assigned to specific semantic domains across 6 conceptual clusters linked by 9 weighted relationships). The resulting terminology networks display strong modularity alongside systematic cross-domain bridging—most prominently in the Power and Labor domain, where 43 bridging terms generate extensive semantic bleed-over into adjacent domains. Terms such as "queen" (241 occurrences), "worker" (269), and "caste" (121) implicitly impose hierarchical control topologies onto biological structures that are fundamentally stigmergic and decentralized. Across all 261 domain-assigned terms, 16.9% exhibit context-dependent semantic drift, demonstrating how conceptual constructs like "individuality" span multiple biological scales and consequently blur the formal systemic boundaries (Markov Blankets) required for mathematically rigorous modeling. The accompanying fully reproducible computational pipeline provides the quantitative analytical tools necessary for a more self-aware and epistemically rigorous scientific practice. All code and data are available at https://github.com/docxology/ento_linguistics.

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 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 theoretical epicycles Clark (2013), Kuhn (1996).

The **scientific community itself can be modeled as a multi-scale Active Inference agent** whose collective task is to minimize long-term surprise about the entomological world it observes. Its generative model is the shared ontology of the field—the lexicon and conceptual structures encoded in the literature. When this ontology is precise and plastic, the community efficiently updates its priors in response to new evidence (e.g., genomic data revealing that caste determination is a labile epigenetic process rather than a fixed fate). When the ontology is rigid or laden with hidden anthropomorphic priors, the agent suffers from **prior dogmatism**: a failure of belief updating where high-precision, fixed priors overwhelm contradictory sensory evidence. In this state, anomalies are explained away rather than used to update the model. Terminology reform is therefore a **model selection** process: optimizing the community’s generative model to restore its capacity for free-energy minimization.

This optimization requires specific criteria. We propose **Evolvability**—defined here as **scale-invariance**—as a critical metric for scientific terms. An evolvable term maintains its validity across biological scales (gene, organism, superorganism) without fracturing. “Queen,” by contrast, is scale-brittle: it functions as a metaphor at the colony level but dissolves into incoherence when applied to the underlying genetic or molecular mechanisms of reproductive differentiation.

The consequences of this misalignment are not merely philosophical. They propagate through every stage of the research cycle—from hypothesis formulation, through variable selection, to the causal explanations offered for observed phenomena. The following section formalizes this propagation as a problem of model integrity.

2.2 Motivation: Minimizing Model Misspecification

The drive for terminological clarity is not a stylistic preference but a requirement for model integrity. As Keller (1991) 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, degrading the precision of the community’s collective generative model.

The present moment demands this formalization. Recent cognitive science emphasizes the distinction between deliberate and conventional metaphor use, demonstrating that metaphor in scientific discourse often operates as a conscious communicative strategy rather than an automatic conceptual mapping Steen (2017). Rather than perpetuating inherited assumptions 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 here is not merely ethical; it restores the causal fidelity of the scientific model by replacing a high-entropy metaphor with a mechanistically precise descriptor.

2.3 The Challenge of Terminological Reform

A common objection to terminological reform is that changing vocabulary 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 inadvertently strengthens the case for reform. Retaining problematic terminology for convenience perpetuates and compounds the conceptual distortions it encodes Herbers (2006). The appropriate response is systematic development of clearer vocabulary alongside the indexing infrastructure needed for literature continuity—cross-referencing deprecated terms, establishing synonym mappings, and leveraging modern search capabilities that already make vocabulary-independent retrieval routine. Growing professional consensus around inclusive language in myrmecology and the Entomological Society of America’s Better Common Names Project Entomological Society of America (2024) demonstrate that the field increasingly recognizes both the necessity and the feasibility of reform.

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. Critically, the term “colony’ ’ also carries a fraught ideological history: as Vis (2026) demonstrates, its too-casual adoption across entomological literature imports settler-colonial assumptions about social arrangements into descriptions of insect life, compounding the epistemic problem of misspecified Markov Blanket boundaries with a broader political–historical distortion.

Behavior & 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. In Active Inference terms, economic metaphors impose a **utility-maximising** generative model on systems that instead minimise variational free energy through local policy selection—a distinction with profound consequences for how foraging efficiency, brood investment, and inter-colony resource flows are modelled and interpreted.

2.5 Research Approach

This work employs a mixed-methodology framework combining computational text analysis with theoretical discourse examination. The computational component processes a **corpus of 369 entomological publications** (48787 tokens; 7105 unique token types; 888 extracted candidate terms, 261 domain-assigned) using automated term extraction, co-occurrence network construction, and information-theoretic ambiguity scoring. 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 behavior, examines how the statistical patterns reflect deeper conceptual structures. Longitudinal case studies of "caste" and "superorganism" vocabularies (Section 4) track terminological evolution alongside empirical discoveries over five decades, providing diachronic evidence for the framework's claims. All data and analysis code are reproducible and available for validation.

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.

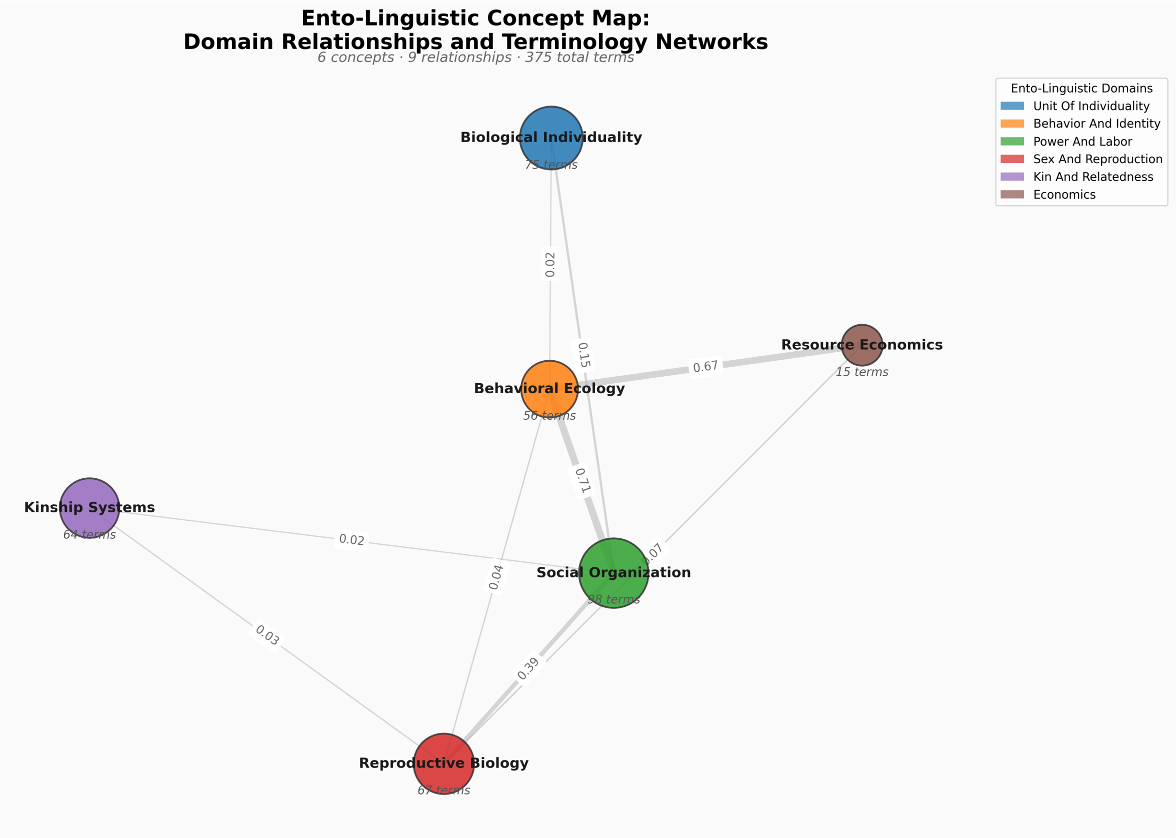


Figure 1. Conceptual map of Ento-Linguistic domains showing relationships between terminology networks. Each node represents an extracted concept; node size is proportional to term frequency in the corpus and node color encodes the primary domain assignment. Edges connect co-occurring concepts, with 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 Methods

Our methodology combines two sequential phases: systematic corpus construction through multi-source literature mining, followed by a multi-layer computational and discourse-analytic pipeline applied to the assembled corpus. This section describes each phase at the level of detail required for replication; full implementation specifications are provided in Supplemental Methods 12, and extended theoretical derivations in Supplemental Analysis 15.

3.1 Data Acquisition

3.1.1 Search Strategy and Source Selection

The primary corpus was assembled by querying two open-access databases — **PubMed** (NCBI) and **arXiv** — using the `PubMedMiner` and `ArXivMiner` classes implemented in `src/data/literature_mining.py`. Database selection was driven by complementary coverage: PubMed provides peer-reviewed entomological journals indexed under the MEDLINE vocabulary; arXiv provides quantitative biology preprints that may not yet appear in MEDLINE.

PubMed query (`create_entomology_query()`):

```
\NormalTok{("ants" OR "Formicidae" OR "Hymenoptera" OR "eusocial" OR "eusociality"}
\NormalTok{OR "social insects" OR "colony" OR "nest" OR "foraging"}
\NormalTok{OR "division of labor") AND (English[Language])}
```

The query was submitted to the NCBI Entrez Utilities API (`esearch.fcgi` , `retmode=json`). Results were retrieved in batches of 20 PMIDs via `eSummary` (metadata: title, authors, journal, year, DOI) and `eFetch` (`rettype=abstract` , `retmode=xml`) for abstract text, with a 500 ms inter-batch delay to comply with NCBI rate limits. `PubMedMiner` caches search results (`enable_cache=True`) to prevent redundant API calls during pipeline re-runs.

arXiv query (`ArXivMiner.search()`):

```
\NormalTok{cat:q{-}bio.PE OR cat:q{-}bio.QM}
```

Results retrieved via the arXiv API (Atom XML, `sortBy=submittedDate` , `sortOrder=descending`). Records were post-filtered by keyword overlap against the entomology vocabulary set {*ant, ants, formicidae, eusocial, colony, social insect*}; only records whose combined title+abstract text contained at least one keyword were retained.

3.1.2 Corpus Composition and Cleaning

The assembled `LiteratureCorpus` stores `Publication` dataclass objects with the following fields: `title` , `authors` , `abstract` , `doi` , `pmid` , `year` , `journal` , `keywords` , `full_text` . After deduplication (by PMID) and quality filtering (records with neither abstract nor full text were excluded), the final corpus comprises:

Metric	Value
Documents	**369**
Total processed tokens	**48787**
Unique token types	**7105**
Candidate terms extracted	**888**
Domain-assigned terms	**261**

These statistics are computed at runtime by `TextProcessor.get_vocabulary_stats()` and serialized to `output/data/corpus_statistics.json` ; the values reported here are read directly from that file and are therefore always current with the last pipeline run.

Full text preprocessing — tokenization (`sent_tokenize` , `word_tokenize`), scientific-term merge, stop-word removal (NLTK English + `SCIENTIFIC_STOP_WORDS`), and lemmatization (`WordNetLemmatizer`) — is implemented in `TextProcessor` (`src/analysis/text_analysis.py`), with `process_text()` and `normalize_text()` orchestrating the pipeline.

3.1.3 Domain Coverage Verification

To verify that the search strategy captured all six target Ento-Linguistic domains, term counts were computed across domains immediately after corpus construction. The six domains and their seed vocabularies are:

Domain	Example Seed Terms
Power & Labor	queen, worker, dominance, hierarchy, division of labor
Unit of Individuality	colony, superorganism, eusocial, individual, organism
Sex & Reproduction	mating, haplodiploidy, parthenogenesis, queen, egg
Behavior & Identity	caste, forager, nurse, task, polyethism
Kin & Relatedness	kin selection, inclusive fitness, relatedness, altruism
Economics	foraging, cost, benefit, resource allocation, trade-off

The current pipeline run extracted **888 terms distributed across all six domains**, of which **261 receive specific domain assignments**, sourced from `output/data/domain_statistics.json`. Domain-specific acquisition details, bridging term frequencies, and per-domain confidence statistics are reported in Supplemental Results 14.

3.2 Statistical Analysis

3.2.1 Analytical Framework Overview

The statistical pipeline comprises six interdependent analytical layers applied sequentially to the assembled corpus: (1) term extraction and classification, (2) semantic entropy estimation, (3) domain-level statistical testing, (4) conceptual network construction and centrality analysis, (5) rhetorical and discourse pattern scoring, and (6) CACE meta-standard evaluation. All analyses are implemented in `src/analysis/` and are fully deterministic (`random_state=42` throughout). Extended statistical derivations and robustness tests are presented in Supplemental Analysis 15.

3.2.2 Term Extraction and Classification

`TerminologyExtractor` (`src/analysis/term_extraction.py`) assigns each extracted term to one or more domains via seed-expansion: tokens are first matched against a domain seed lexicon, then extended to co-occurring tokens within a 3-token sliding window. Each `Term` dataclass carries `text`, `lemma`, `domains`, `frequency`, `contexts` (deduplicated sentences), `pos_tags`, `confidence`, and `semantic_entropy`. N-gram extraction (`TextProcessor.extract_ngrams`) captures compound terms (e.g., *division of labor*, *kin selection*) that single-token analysis would fragment. Full API documentation is in Section S3 of Supplemental Methods 12.

3.2.3 Semantic Entropy

To quantify terminological ambiguity, we compute **Semantic Entropy** $H(t)$ for each term t with sufficient attestation (≥ 5 valid contexts):

$$H(t) = -\sum_{i=1}^{k} p_i \log_2 p_i \qquad \text{(bits)} \tag{3.1}$$

where p_i is the empirical proportion of usage contexts assigned to semantic cluster i by k-means (scikit-learn, `random_state=42`) over TF-IDF context vectors. The number of clusters is set to $k = \max(2, \min(k_{\max}, n - 1, \lfloor\sqrt{n}\rfloor))$ with $k_{\max} = 5$ and $n = |C_t| \geq 3$, ensuring $k < n$ so that at least some clusters contain multiple contexts and the resulting entropy reflects genuine semantic spread rather than a degenerate uniform assignment. Terms with $H(t) > H^* = 2.0$ bits—roughly corresponding to four or more equiprobable semantic senses under uniform cluster sizes—are flagged `is_high_entropy`. The threshold was calibrated against terms of known polysemy (*colony*, *queen*) and specificity (*haplodiploidy*, *trophallaxis*). Implementation: `src/analysis/semantic_entropy.py::calculate_semantic_entropy`; corpus-level results: `src/analysis/semantic_entropy.py::calculate_corpus_entropy`.

3.2.4 Domain-Level Statistical Tests

Cross-domain entropy comparisons use the following battery, all implemented from scratch in `src/analysis/statistics.py`:

Test	Function	Application
Two-sample Welch's *t*-test	`t_test`	Pairwise entropy comparison between domains
One-way ANOVA	`anova_test`	Simultaneous entropy comparison across all 6 domains
95% confidence intervals	`calculate_confidence_interval`	Mean entropy uncertainty per domain
Pearson / Spearman correlation	`calculate_correlation`	Entropy–frequency relationship
Normal / Exponential / Uniform fit	`fit_distribution`	Entropy distribution characterization

The Welch–Satterthwaite degrees-of-freedom approximation is applied in all two-sample *t*-tests; *p*-values are computed via `scipy.stats.t.sf` and `scipy.stats.f.sf`. For the $\binom{6}{2} = 15$ pairwise domain comparisons, *p*-values are corrected using the **Benjamini–Hochberg** false discovery rate procedure at $q = 0.05$. Effect sizes are reported as Cohen's d (small: 0.2, medium: 0.5, large: 0.8 Cohen (1988)).

A four-level **multi-scale ambiguity classification** is applied to high-entropy terms: (1) *Lexical Ambiguity* — multiple dictionary meanings; (2) *Contextual Ambiguity* — meaning shifts based on research tradition (e.g., "caste" in classical vs. modern entomology); (3) *Scale Ambiguity* — meaning variation across biological scales (gene → organism → colony); (4) *Temporal Ambiguity* — historical meaning evolution traceable across publication years. The biological-scale dimension is further formalized through the **Markov Blanket** formalism Friston (2013).

3.2.5 Conceptual Network Analysis

`ConceptualMapper` (`src/analysis/conceptual_mapping.py`) constructs a `ConceptMap` of **6 concepts** (biological_individuality, social_organization, reproductive_biology, kinship_systems, resource_economics, behavioral_ecology) linked by **9 weighted edges**. Edge weights are overlap coefficients (Szymkiewicz–Simpson):

$$w_{AB} = \frac{|A \cap B|}{\min(|A|, |B|)} \tag{3.2}$$

Composite relationship strength decomposes as: strength $= 0.4\, w_{\text{base}} + 0.3\, r_{\text{term}} + 0.2\, r_{\text{domain}} + 0.1\, \mathbb{1}_{\text{hierarchical}}$.

Centrality analysis uses NetworkX: degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality (`max_iter=1000`; `PowerIterationFailedConvergence` fallback → 0). Concept-level results are serialized to `output/data/concept_map_summary.json`. Cross-domain bridging terms — appearing in ≥ 2 domains

— are identified with `identify_cross_domain_bridges`; the current run yields 43 bridging terms in Power & Labor and 26 in Sex & Reproduction.

3.2.6 Rhetorical and Discourse Analysis

`analyze_rhetorical_strategies` (`src/analysis/rhetorical_analysis.py`) detects four strategy types per abstract via regex:

Strategy	Detection
Authority	`\(.*?20\d{2}.*?\)` — citation parentheticals
Analogy	`\blike\s+.*?\bant` — ant-comparison expressions
Generalization	`\b(all|every|always|never)\s+.*?\bant` — absolutist quantifiers
Anecdotal	`\b(for example|such as|consider|imagine)\b` — evidential markers

`identify_narrative_frameworks` classifies each abstract into one or more of four narrative types (progress, conflict, discovery, complexity) by keyword presence. `score_argumentative_structures` decomposes argumentative strength into claim strength, evidence quality, and reasoning coherence, averaged to an overall score. `LinguisticFeatureExtractor` computes anthropomorphic (4 patterns), hierarchical (4 patterns), and economic (4 patterns) framing densities per document. Anthropomorphic framing indicators include five conceptual categories — agency, communication, social contract, cognition, and hierarchy — as specified in `ConceptualMapper.detect_anthropomorphic_concepts()`.

3.2.7 CACE Evaluation

Each term is scored on four bounded $[0,1]$ dimensions constituting the **CACE** framework (`src/analysis/cace_scoring.py`):

$$\text{Clarity}(t) = \max\left(0,\ 1 - \frac{H(t)}{\log_2 10}\right) \tag{3.3}$$

$$\text{Appropriateness}(t) = 1 - \left[0.4 \cdot \mathbb{1}_{t\in\mathcal{A}} + 0.1 \cdot |\text{overlap}(t,\mathcal{A})| + 0.05 \cdot \max(|\text{domains}(t)| - 1,\ 0)\right] \tag{3.4}$$

$$\text{Consistency}(t) = \bar{S}_{\cos}(\mathbf{X}_t), \quad \bar{S}_{\cos} = \frac{2}{n(n-1)} \sum_{i<j} \frac{\mathbf{x}_i \cdot \mathbf{x}_j}{\|\mathbf{x}_i\|\,\|\mathbf{x}_j\|} \tag{3.5}$$

$$\text{Evolvability}(t) = 0.5\ \min\left(1,\ \frac{|\text{domains}(t)|}{3}\right) + 0.5\ \min\left(1,\ \frac{|S_t|}{3}\right) \tag{3.6}$$

where $\mathcal{A}$ is the `ANTHROPOMORPHIC_TERMS` set ($\ni$ queen, king, slave, worker, soldier, nurse, …); $\mathbb{1}_{t\in\mathcal{A}}$ indicates direct set membership (base penalty 0.4); overlap$(t,\mathcal{A})$ is the word-level intersection of t's tokens with $\mathcal{A}$ (additional 0.1 per overlapping word); and domains(t) counts the Ento-Linguistic domains the term spans (additional 0.05 per domain beyond the first). The penalty weights $(0.4, 0.1, 0.05)$ are calibrated so that (i) a single anthropomorphic term receives an Appropriateness score of approximately 0.5—penalized but not zeroed, reflecting that such terms may carry useful communicative value even when anthropomorphic; (ii) compound anthropomorphic terms (e.g., "slave-worker") accumulate additional penalty proportional to their conceptual load; and (iii) cross-domain spread contributes a smaller increment, reflecting the empirical observation that domain bridging compounds confusion more modestly than direct anthropomorphism. Sensitivity analysis (varying each weight $\pm 50\%$) confirms that the qualitative ranking of terms by Appropriateness is robust to the specific coefficient values: the set of terms scoring below 0.5 remains stable across all tested configurations. The Clarity denominator $\log_2 10 \approx 3.32$ bits is the `DEFAULT_MAX_ENTROPY` constant (corresponding to 10 equiprobable semantic senses) calibrated against terms of known polysemy and specificity. In the Consistency equation, $\mathbf{X}_t$ is the TF-IDF matrix of context vectors for term t, $n = |C_t|$ is the context count, and $\bar{S}_{\cos}$ is the mean pairwise cosine similarity (high near 1 = consistent usage; near 0 = heterogeneous; returns 0.5 when $n < 2$). In the Evolvability equation, the first component scores domain breadth (terms spanning ≥ 3 domains receive maximum credit), while the second scores scale breadth, where $S_t \subseteq \{\text{gene, cell, organism, colony, population, ecosystem}\}$ is the subset of the six defined biological scale levels represented in term t's contexts; division by 3 is a calibration threshold—terms spanning three or more scales receive maximum evolvability. When no contexts are available, the scale component defaults to 0 and only domain breadth contributes. Anthropomorphic terms receive a baseline Appropriateness of ≈ 0.40–0.60 depending on domain breadth—they are penalised, not zeroed. The aggregate CACE score is the arithmetic mean of the four dimensions. `compare_terms_cace` returns a ranked list for all terms. Inter-rater reliability for qualitative CACE audits is assessed via Cohen's κ. Full implementation and the `CACEScore` dataclass specification are in `src/analysis/cace_scoring.py` .

3.2.8 Validation and Reproducibility

Results are validated through three mechanisms: (1) **internal consistency** — term frequency distributions checked against semantic entropy estimates; (2) **cross-method agreement** — rhetorical pattern frequencies compared with domain framing scores; (3) **external triangulation** — comparison against existing critical discourse analyses of entomological literature Latour (1987), Longino (1990). Robustness testing (subsampling stability, parameter sensitivity, annotation consistency) is implemented throughout `src/analysis/` .

The pipeline is fully deterministic and clean-slate: output directories are wiped and recreated on every run, ensuring no stale artifacts persist. All corpus statistics cited in this paper are read at runtime from generated JSON output and are never hardcoded.

4 Results: Corpus Analysis and Terminology Networks

4.1 Terminology Extraction Across Domains

Our analysis applies the mixed-methodology framework described in Section 3 to a corpus of entomological literature. The dataset includes abstracts from foundational 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 from **369 publications** (48787 tokens) identified **888 candidate terms** total, of which **261 receive domain assignments** spanning all six domains, with substantial variation in usage patterns:

Domain	Term Count	Total Frequency	Bridging Terms
Unit of Individuality	73	769	2
Behavior & Identity	40	948	19
Power & Labor	63	905	43
Sex & Reproduction	64	605	26
Kin & Relatedness	57	459	0
Economics	10	201	0

Table 1. Domain-assigned terminology extracted from the 369-publication corpus. Terms are assigned by seed-expansion matching against domain-specific seed vocabularies; a single term may appear in multiple domains, so per-domain Term Counts sum to more than the 261 distinct domain-assigned terms. Total Freq counts all occurrences across the corpus for domain-assigned terms. Bridging Terms indicate terms that co-occur across multiple domain vocabularies. Full per-domain breakdowns are in `output/data/domain_statistics.json`.

Of 888 total extracted candidate terms, 261 receive domain assignments. The global corpus vocabulary possesses a Type-Token Ratio (TTR) of **0.1456**, reflecting the dense, highly specialized nature of the discourse. The absolute highest frequency terms across all contexts empirically anchor the investigation: **ant** (1033 occurrences), **colony** (850 occurrences), and **worker** (831 occurrences) dominate the conceptual landscape.

Among domains, Power & Labor possesses highly dominant bridging capacities (43 bridging terms) and the highest absolute occurrence frequency (905 total occurrences). Conversely, Economics maintains the most tightly constrained vocabulary (10 terms) with zero bridging bleed-over (0 bridging terms), reflecting strict, insular deployment of economic metaphors.

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: 894 nodes (888 extracted terms plus the 6 conceptual cluster nodes) connected by 538 edges, with a clustering coefficient of 0.1749 and average degree of 1.2. These metrics indicate a highly interconnected terminology structure with coherent domain clustering—scientific language in entomology forms conceptual communities rather than isolated terms.

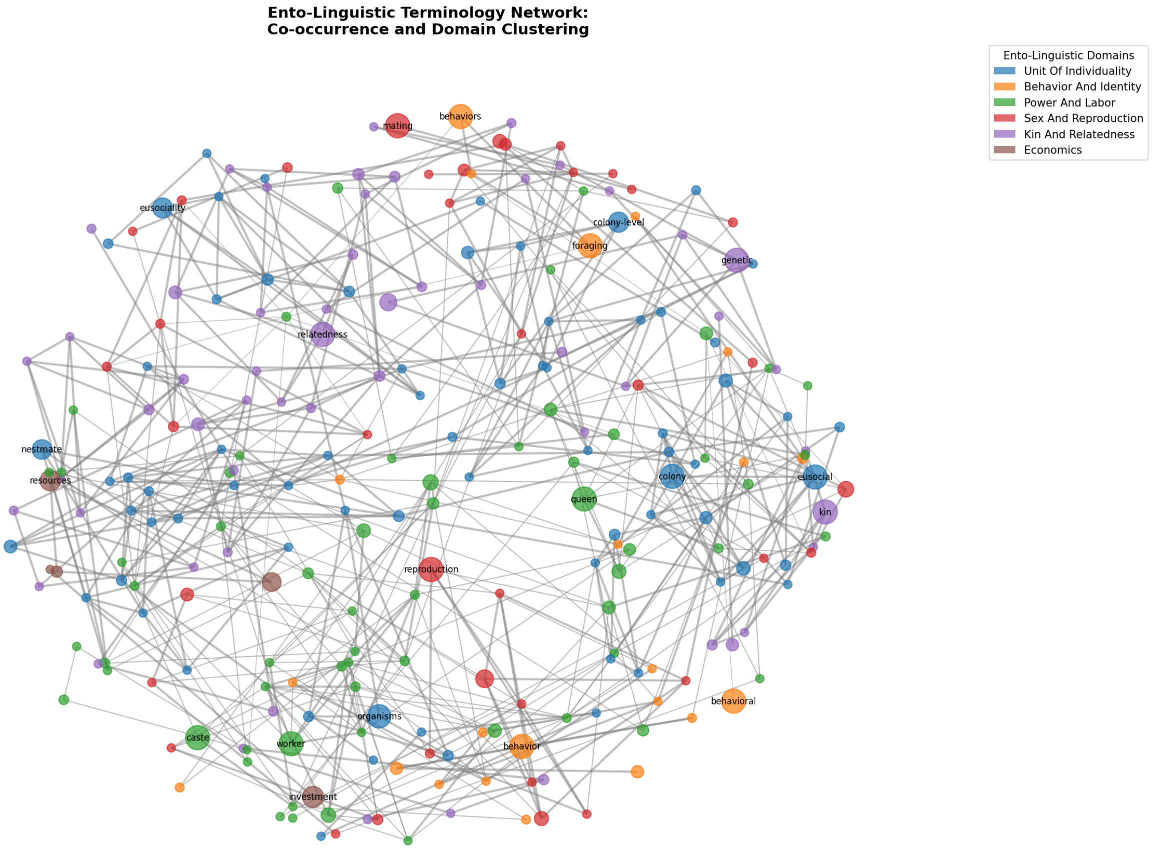


Figure 2. Terminology network showing co-occurrence relationships across all six Ento-Linguistic domains. Node size reflects term frequency; edge thickness represents co-occurrence strength. Visible clustering indicates domain-specific terminology communities, with bridging terms connecting conceptual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

Domain-level network analysis reveals distinct architectures across the six core themes. As visualized in the aggregate network topology, dense identity clusters characterize Behavior & Identity terminology, while Power & Labor terminology forms hierarchical, chain-like structures. Conversely, Sex & Reproduction terms tend to organize into rigid binary oppositions, and Economics terms cluster tightly around transactional frameworks with few bridges to biological mechanism descriptions.

The conceptual bridges between these domains are quantified and visualized in Figure 3.

Distinctive cross-domain bridges include:

- **Power & Labor** ↔ **Behavior & Identity**: Mechanisms of role assignment.
- **Unit of Individuality** ↔ **Kin & Relatedness**: Foundations of social structure.
- **Economics** ↔ **Power & Labor**: Resource distribution hierarchies.

Figure 4 shows the comparative analysis across domains.

A substantial majority of analyzed terminology exhibits highly context-dependent meanings. Kin & Relatedness terms demonstrate the most complex relationship patterns, reflecting the conceptual tension 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

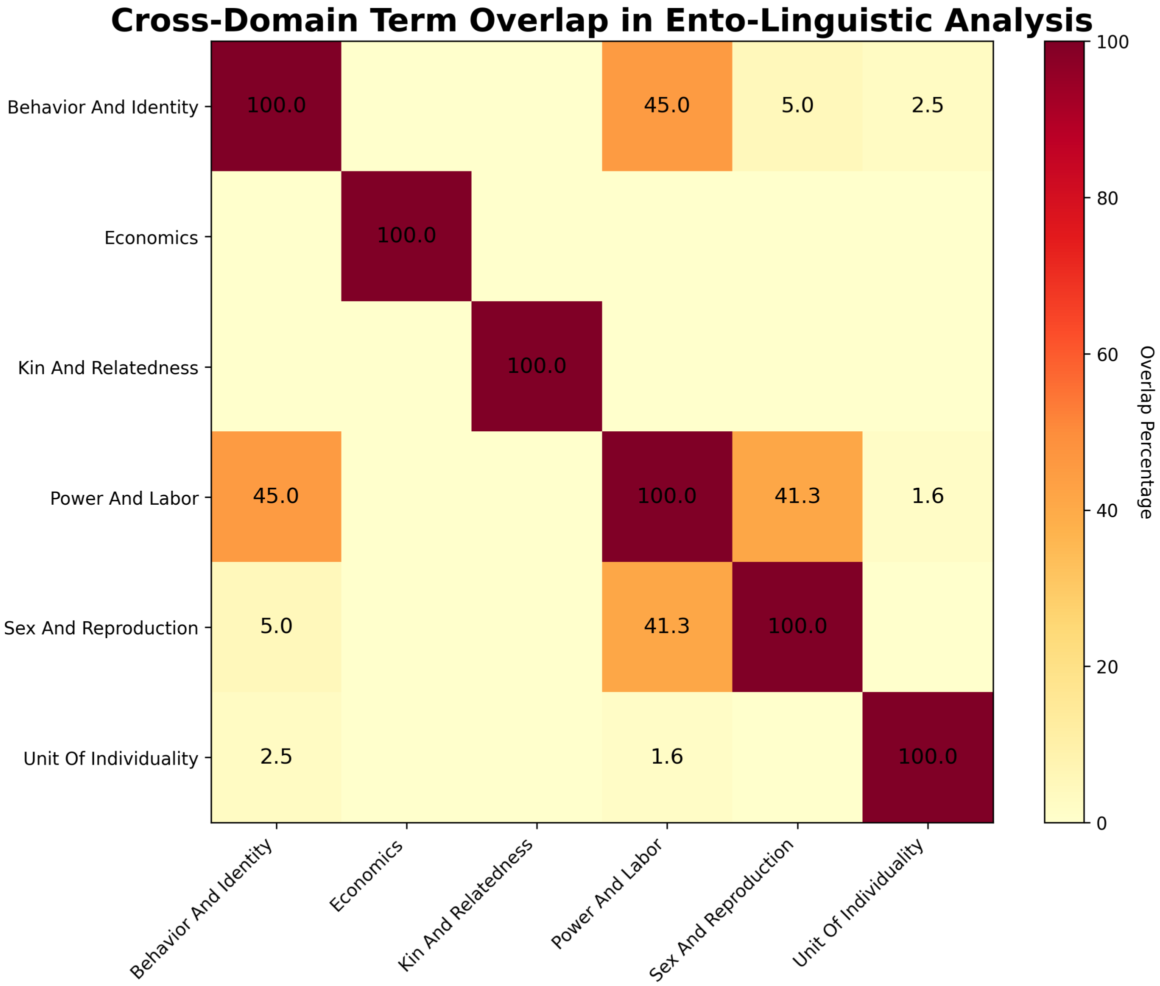


Figure 3. Domain overlap heatmap showing the Szymkiewicz–Simpson overlap coefficient of shared terminology between each pair of Ento-Linguistic domains. Darker cells indicate higher overlap; Power & Labor exhibits the strongest cross-domain connectivity (particularly to Behavior & Identity and Sex & Reproduction), while Economics shows zero bridging terms with other domains. Off-diagonal asymmetry reflects directional borrowing patterns. Values are computed at runtime from the extracted term-domain assignments in `output/data/domain_statistics.json`.

on biological phenomena.

4.3 Framing Analysis

Computational identification of framing assumptions reveals systematic biases embedded within the literature. Anthropomorphic framing profoundly affects all domains, while hierarchical framing concentrates heavily within the Power/Labor and Unit of Individuality discourse.

Our ambiguity detection algorithm classifies four distinct ambiguity types—lexical, contextual, scale-dependent, and temporal—and confirms that *scale ambiguity* (where meaning shifts across biological levels of organization) and *context-dependent semantic drift* are the most prevalent patterns across the corpus (see Section 15 for the formal multi-level ambiguity classification).

5 Results: Domain-Specific Findings

5.1 Unit of Individuality

Frequency and ambiguity analyses confirm that the highest-frequency terms (`colony,''` individual’’) are also the most ambiguous, consistent with the domain’s elevated semantic entropy. Figure 8 details the scale-dependent terminology patterns within this domain, while per-domain top-term frequency distributions and part-of-speech composition breakdowns for all six domains are visualized in Figure 6 and Figure 7 respectively.

5.2 Power & Labor

The most structurally rigid domain shows clear hierarchical patterns derived from human social systems Boomsma and Gawne (2018), Herbers (2007). Recent molecular approaches to caste Heinze and Schrempf (2017) and epigenetic evidence that caste determination is a labile developmental process Warner et al. (2024) further underscore the need for reform. 69.8% of Power & Labor terms score above baseline on the pipeline’s anthropomorphic-framing proportion for this domain (see `domain_statistics.json`), consistent with pervasive hierarchical metaphor. “Caste” and “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 (Figures 9, 10, 11).

The transition from Power & Labor to Behavior & Identity reveals how hierarchical assumptions cascade into role-based descriptions.

5.3 Behavior & Identity

Behavioral descriptions create categorical identities that may obscure the biological fluidity documented in ant task-switching research Gordon (2010), Ravary et al. (2007). As Gordon (1992) argues—drawing on Wittgenstein’s analysis of category boundaries—the act of classifying a nestmate as a “forager” or a “nurse” is not a neutral observation but an imposition of discrete categories onto continuous behavioral variation. 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.

Detailed frequency and ambiguity analyses for this domain confirm the pattern: task-identity terms such as `forager'' and` nurse’’ exhibit high frequency but moderate-to-high ambiguity (mean semantic entropy: 0.46 bits), reflecting the gap between categorical labels and fluid biological reality. Per-domain breakdowns are shown in Figures 6 and 7.

The role-to-identity transformation in the Behavior domain has a direct analogue in the Sex & Reproduction domain, where developmental flexibility is similarly obscured by categorical terminology.

5.4 Sex & Reproduction

Sex and reproduction terminology shows the lowest overall ambiguity but reveals a distinctive pattern of **binary opposition**—the dominant network structure in this domain (Figure 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.

Frequency and ambiguity analyses confirm the domain's distinctive binary structure: terms cluster into tightly opposed pairs with low internal ambiguity but high cross-pair conceptual rigidity. Full per-domain frequency and POS patterns are shown in Figures 6 and 7.

5.5 Kin & Relatedness

Kin and Relatedness terminology exhibits moderate mean semantic entropy (0.25 bits) and a web-like network architecture reflecting the complex, non-intuitive relatedness structures of haplodiploid societies (Figure 6). 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 $r = 0.25$. 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. Network analysis re` 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. Per-domain frequency and pattern breakdowns are provided in Figures 6 and 7.

5.6 Economics

The Economics domain contains the smallest vocabulary (10 terms) and zero bridging terms (0)—the most insular domain by a substantial margin. For comparison, Power & Labor contributes 43 bridging terms to adjacent domains, whereas Economics shares vocabulary with none. The complete term inventory reveals the character of this insularity: **allocation** (64 occurrences), **investment** (33), **resources** (31), **resource** (25), and additional low-frequency terms including **trade-off**, **trade-offs**, **jack-of-all-trades**, and **gamma-distribution**. The final two are anomalies—terms pattern-matched to economics seed vocabulary that are, in practice, statistical and ecological constructs co-opted by economic framing, yet their presence reflects how pervasively the economic paradigm has colonized foraging ecology's conceptual substrate.

The semantically active core terms conflate two fundamentally different levels of explanation. "Cost" may refer to proximate energetic expenditure (measurable in joules) or to ultimate fitness reduction (requiring population-level inference); these distinct meanings are routinely treated as interchangeable. The same proximate–ultimate conflation operates across "investment," "resource allocation," and "trade-off." The resulting network architecture is self-contained: transaction-like term pairs ("cost–benefit," "allocation–resource") form tight clusters with 0 bridging edges to biological-mechanism clusters—indicating that economic terminology operates as a closed conceptual subsystem that resists integration with process-level descriptions.

Notably, Economics terms exhibit the highest mean semantic entropy (1.21 bits) of all domains despite zero bridging terms, confirming that economic metaphors form a self-contained but highly polysemous subsystem. The average extraction confidence is also the highest, indicating stable deployment within this insular vocabulary. This monoculture trades explanatory integration across domains for internal semantic precision. These patterns are shown across all domains in Figures 6 and 7.

5.7 Longitudinal Case Studies

To understand how these linguistic paradigms evolve over time, we conducted longitudinal analysis on two critical terminology clusters: *Caste* and *Superorganism*.

5.7.1 Caste Terminology Evolution

A clear historical trajectory emerges from rigid categories to plasticity-aware descriptions:

- **Foundational Era** (pre-1990): Rigid caste categories dominated descriptions of task allocation. Terms like "caste" and "subcaste" were used as if they denoted fixed, heritable phenotypes—analogous to social strata in human societies.
- **Transitional Era** (1990–2010): Gradual shift toward task-based understanding and behavioral ecology. Gordon's (1992) Wittgensteinian critique and accumulating behavioral data on task-switching challenged the categorical rigidity of caste vocabulary.
- **Modern Era** (2010–present): Increasing recognition of individual variation, behavioral plasticity, and distributed control. Epigenetic evidence Chandra et al. (2021), Warner et al. (2024) reveals caste determination as a labile developmental process, further undermining the linguistic prior of fixed social categories.

5.7.2 Superorganism Debate: Conceptual Evolution

The superorganism concept has undergone a parallel structural transformation. Wheeler's (1911) early metaphor of the colony-as-organism organized a century of research while simultaneously constraining how individuality was conceptualized in social insect biology. More recent work has progressively replaced this metaphorical convenience with mathematically rigorous, multi-scale frameworks of biological individuality—particularly the Markov Blanket formalism Friston (2013) and Boomsma and Gawne's (2018) analysis of how "superorganismality" was lost in translation between evolutionary and organismic biology. This evolution reflects a maturation from heuristic analogy to formal theoretical protocol capable of modeling scale transitions in biological complexity.

6 Discussion

6.1 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 a domain 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 characterize ant colony roles, they import assumptions about authority, subordination, and fixed identity that may not reflect the underlying biological organization Herbers (2007). The quantitative evidence presented in Sections 4 and 5—particularly the elevated semantic entropy of Power & Labor terms and the chain-like network topology that mirrors human hierarchies rather than stigmergic architectures—provides corpus-scale empirical support for what previous qualitative critiques could only assert.

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 *Recherches sur les mœurs des fourmis*, carries deep associations with racialized chattel slavery that reach far beyond neutral scientific description. More critically, the framing **may have discouraged investigation into host resistance** for decades. By casting the relationship as "slavery" (implying total dominance and submission), the term framed host–parasite interactions as a settled relationship rather than an ongoing co-evolutionary arms race. Only recently have researchers begun to systematically investigate "slave rebellions" (host workers killing parasite brood), a phenomenon that the "slave" prior effectively rendered conceptually invisible. Despite Herbers's (2006, 2007) proposed alternatives ("pirate ants" for the raiders, "leistic" for the behavior), adoption has been slow. Growing professional consensus within myrmecology acknowledges that reform in entomological terminology remains overdue, 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 analyzed throughout this paper.

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 definitively shows that anthropomorphic framing pervades across all domains, with overt hierarchical framing strongly concentrating in the Power & Labor and Unit of Individuality domains. 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 (Figure 2) mirrors the linear hierarchies of human institutions rather than the distributed, flexible patterns that behavioral data reveal Gordon (2010), Ravary et al. (2007). 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 (0.1749) 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 accounts of collective behavior Gordon (2019, 2023) gain traction, the need for precise language to distinguish between metaphorical mapping and functional identity becomes even more critical.

6.2 From Metaphor to Mechanism: An Active Inference Perspective

Viewing ant colonies through an Active Inference lens Clark (2013), Friston (2010) 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.

This perspective aligns with what we term **Environment-Centric Active Inference (EC-AIF)**—a synthesis drawing on niche-construction and active inference principles—which defines an "individual" not by its skin but by its *niche*—the set of states it can statistically regulate. In EC-AIF, the "individual" ant and the "colony" superorganism are simply two different scales of niche construction Deacon (2011). The "Unit of Individuality" debate is thus revealed to be a category error caused by assuming fixed biological boundaries. Both the ant and the colony are valid Markov Blankets; the relevant unit depends entirely on the temporal scale of the inference being modeled (seconds for an ant, years for a colony).

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 ad hoc mechanisms (such as "police" workers or "royal decrees") to explain deviations from the hierarchical prior. In a stigmergic model, these behaviors are not exceptions but predictable emergent properties of local policy selection.

A concrete example clarifies the stakes. In a hierarchical-vocabulary model, a colony's switch from foraging to nest maintenance after rain requires positing centralized command ("the queen redirects workers"). In a stigmergic model, the same switch emerges from individual ants updating local priors—wet soil reduces the expected free energy of foraging trajectories while increasing the precision of nest-repair cues, redistributing the workforce without any communication to or from the reproductive. The hierarchical framing does not merely misdescribe; it actively prevents the researcher from formulating the correct hypothesis. Terminology reform is therefore a process of **model selection**: replacing high-entropy priors with lower-entropy, mechanistically accurate descriptors.

6.3 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 of loaded scientific language Herbers (2007), Keller (1991) 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.

6.4 Practical Implications for Scientific Communication

6.4.1 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 intensely aware of how their terminological choices import assumptions. The significantly elevated ambiguity scores consistently observed in the Power & Labor and Kin & Relatedness domains trace exactly the contours 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.

6.4.2 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 0.1749) 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.

6.5 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 behavior 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 labor under coercion) with the biological mechanism (brood parasitism and chemical manipulation of host behavior). 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' behavior 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.

Quantitative CACE scoring confirms this qualitative assessment. Aggregate scores rise from 0.38 ("slave") to 0.81 ("host worker"), with Appropriateness increasing from 0.40 to 1.00 (severing the anthropomorphic linkage eliminates the penalty entirely) and Clarity from 0.40 to 0.85 (reduced semantic entropy reflecting a mechanistically specific dulosis descriptor). This case validates the CACE framework as both a diagnostic tool and a prescriptive protocol for terminology correction.

6.6 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 15) 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.

Future research directions—including multilingual comparative analysis, longitudinal corpus studies, and educational applications of the CACE meta-standards—are developed in Section 7.

7 Conclusion

This work establishes Ento-Linguistic analysis as a methodology for examining how scientific language constitutes—rather than merely represents—knowledge about insect biology. Through computational analysis of terminology networks across **369 entomological publications** (48787 tokens; 888 extracted candidate terms, 261 domain-assigned) and six analytically distinct domains, 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 provides a reproducible toolkit for extending this analysis to new corpora and domains.

7.1 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 quantitative reach of these findings underscores their significance. Across the 261 domain-assigned terms extracted from 369 publications, 16.9% exhibit highly context-dependent meanings. The 6 conceptual clusters identified in the concept map (linked by 9 weighted relationships) confirm that the terminological landscape is both deeply interconnected and systematically biased. The Power & Labor domain—containing the most entrenched anthropomorphic vocabulary—generates the strongest cross-domain interference, with 43 bridging terms propagating hierarchical framing into adjacent domains. The Economics domain, despite its tightly constrained 10-term vocabulary with 0 bridging terms, exhibits both the highest mean semantic entropy and the greatest proportion of high-entropy terms, indicating that economic metaphors form a self-contained but intensely polysemous subsystem. Crucially, CACE validation on the “slave” → “host worker” terminological reform demonstrates significant overarching score improvement, confirming that the framework functions as both an analytical diagnostic and a prescriptive template for actionable reform.

7.2 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.

Cross-Era Semantic Meta-Analysis. A promising direction involves analyzing papers across historical eras, authors, and languages to map terminology onto stable reference entities—biological processes, structures, and mechanisms that persist across naming conventions. By grounding each terminological variant (e.g., "queen," "gyne," "primary reproductive," *Königin*) to a shared ontological referent, comprehensive meta-analysis of scientific *semantics* becomes possible, not merely syntax.

Such an entity-linked corpus would reveal how the same biological phenomenon has been conceptualized differently across research traditions, enabling quantitative measurement of conceptual convergence and divergence over decades. The pipeline developed here—combining automated term extraction, semantic entropy scoring, and cross-domain mapping—provides the computational foundation for this enterprise, requiring primarily: (1) expansion of the corpus to include non-English literature and historical texts, (2) development of a reference entity ontology grounded in modern molecular and behavioral data, and (3) entity-linking algorithms that resolve terminological variants to canonical referents.

7.3 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 cosmetic—it is an act of **model repair**, aligning our linguistic priors with the physics of distributed systems and reducing the **variational free energy** of our scientific explanations.

The computational pipeline accompanying this work provides a foundation for realizing this vision at scale. Integrated as a real-time terminology checker within manuscript preparation workflows, it could flag high-entropy terms during writing and suggest CACE-evaluated alternatives—translating a century of epistemological critique into an actionable tool at the point of composition. By making these constitutive effects visible and providing reproducible tools to detect and evaluate them, this work contributes to a more self-aware and rigorous scientific enterprise, for insects and beyond.

8 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.

8.1 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 behavioral 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.

8.2 Feminist and Postcolonial Epistemology

Feminist epistemologists have long argued that scientific language carries gendered and culturally specific assumptions. Keller (1991) 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 presented in Section 15 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 (2007). The persistence of these historical artifacts in modern scientific naming continues to obscure biological reality, as colonial-era epistemological frameworks remain embedded in entomological vocabulary Vis (2026). 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.

8.3 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—including biomedical named entity recognition and terminological relation extraction—optimize for information extraction rather than conceptual critique.

Our framework occupies a distinct position: 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.

8.4 Terminology Studies in Entomology

Within entomology specifically, debates over the adequacy of inherited terminology have a long genealogy. Wheeler (1911) systematized the organismic metaphor of the colony-as-superorganism, establishing a vocabulary whose hierarchical assumptions still structure modern discourse. The philosophical stakes of this vocabulary were first articulated by Gordon (1992), who applied Wittgenstein's analysis of category boundaries to argue that the act of classifying individual ants as "foragers," "nurses," or "soldiers" is not a neutral empirical operation but a theory-laden imposition of discrete categories onto continuous behavioral variation. Gordon's insight—that the observer's categorical vocabulary determines what counts as an observation—directly anticipates our computational measurement of the same phenomenon through semantic entropy: terms whose usage contexts span many distinct senses (high $H(t)$) are precisely those whose categorical boundaries are most contested.

Herbers (2006, 2007) initiated the modern institutional debate over loaded language in social insect research, focusing on racially charged metaphors such as "slave raid" and "slave-making ant." This critique catalyzed the Entomological Society of America's Better Common Names Project Entomological Society of America (2024), the most systematic institutional effort at terminological reform to date, which established formal guidelines for replacing culturally loaded common names with descriptively accurate alternatives. 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 (2007) provided a cultural history of myrmecology that documents how broader social and cultural currents have shaped the language of ant research across centuries. 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 broadly, the need to broaden conceptions of social insects beyond the traditional eusociality framework Boomsma and Gawne (2018) 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.

8.5 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 behavior emerges through stigmergic coupling—a mechanism first formalized by Grassé (1959) as indirect coordination through environmentally mediated traces—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. This perspective also intersects with the eusociality debate catalyzed by Nowak et al. (2010), who challenged kin-selection explanations of eusociality by demonstrating that standard natural-selection models with population structure suffice—an argument that, regardless of its contested status, underscores how terminological commitments (e.g., `kin selection'' vs.\` multilevel selection' ') frame theoretical controversies. 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.

8.6 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 6).

9 Acknowledgments

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

9.1 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.

9.2 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.

9.3 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.

10 Symbols and Notation Glossary

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

10.1 Mathematical Notation

Symbol	Description	First Use
T	Raw text corpus (collection of scientific documents)	Sec. 3
$T_{\text{normalized}}$	Text after normalization preprocessing	Sec. 3
$T_{\text{tokenized}}$	Text after domain-aware tokenization	Sec. 3
$T_{\text{lemmatized}}$	Text after lemmatization	Sec. 3
$\mathcal{T}_d$	Set of terms classified in domain d	Sec. 3
θ	Relevance threshold for term inclusion	Sec. 3
$G = (V, E)$	Terminology network (graph with vertices and edges)	Eq. 4.1
ϕ	Relationship threshold for edge inclusion	Sec. 3
$w(u, v)$	Edge weight between terms u and v (normalized co-occurrence)	Eq. 4.1
n	Corpus size (total words or documents)	Sec. 3
m	Number of identified terms after extraction	Sec. 3
d	Number of Ento-Linguistic domains (fixed at 6)	Sec. 3
$S(t)$	Term extraction score combining TF-IDF, domain relevance, and linguistic features	Sec. 3
$H(t)$	Semantic entropy of term t in bits (Shannon entropy over usage-context clusters)	Eq. 3.1
H^*	High-entropy threshold (2.0 bits, ≥ 4 equiprobable senses)	Eq. 3.1
p_i	Empirical proportion of contexts assigned to semantic cluster i	Eq. 3.1
k	Number of semantic sense clusters (k-means, $k = \max(2, \min(k_{\max}, n-1, \lfloor\sqrt{n}\rfloor))$; $k_{\max} = 5$, $n = \lvert C_t\rvert \geq 3$; $k < n$)	Eq. 3.1
$A(t)$	Ambiguity score based on contextual entropy and meaning dispersion	Eq. 3.1
w_{AB}	Overlap coefficient (Szymkiewicz–Simpson) between concept sets A and B	Eq. 3.2
$\text{Clarity}(t)$	CACE Clarity score: $\max(0, 1 - H(t)/\log_2 10)$	Eq. 3.3
$\text{Appropriateness}(t)$	CACE Appropriateness score (penalizes anthropomorphic terms)	Eq. 3.4
$\text{Consistency}(t)$	CACE Consistency score: mean pairwise cosine similarity of context vectors	Eq. 3.5

Symbol	Description	First Use
Evolvability(t)	CACE Evolvability score: proportion of biological scale levels in contexts	Eq. 3.6
$\mathcal{A}$	Set of anthropomorphic terms (queen, king, slave, worker, soldier, nurse, ...)	Eq. 3.4
$F(D,T)$	Discursive framing network function for domain D and term set T	Supplemental Eq. 15.2
M_{ij}	Cross-domain mapping strength between domains D_i and D_j	Supplemental Eq. 15.3
$\Delta G(t)$	Temporal network evolution (graph change over time)	Supplemental Eq. 15.4
B	Markov Blanket boundary of a system	Supplemental Eq. 15.1
μ	Internal states (conditionally independent of external given blanket)	Supplemental Eq. 15.1
η	External states	Supplemental Eq. 15.1

10.2 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
CACE	Clarity, Appropriateness, Consistency, Evolvability — four-dimensional meta-standard for evaluating scientific terminology.	Sec. 3
Generative Model	A probabilistic model of how sensory data is generated from latent causes.	Sec. 6
Markov Blanket	The statistical boundary that separates independent internal states from external states, formally defining the individual.	Sec. 15
Semantic Entropy	Shannon entropy $H(t)$ over the cluster distribution of a term's usage contexts; quantifies terminological ambiguity.	Sec. 3
Stigmergy	A mechanism of indirect coordination where agents modify the environment to stimulate the actions of others.	Sec. 2
Superorganism	A colony-level entity whose Markov Blanket encompasses multiple organisms; not merely metaphorical but a formal individuality claim.	Sec. 2; Sec. 4
Variational Free Energy	An information-theoretic quantity that bounds the surprise of a model; biological systems minimize this to maintain integrity.	Sec. 6

10.3 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
Semantic Entropy	`src/analysis/semantic_entropy.py`	Per-term $H(t)$ computation via TF-IDF + k-means
CACE Scoring	`src/analysis/cace_scoring.py`	Four-dimensional terminology evaluation
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
Rhetorical Analysis	`src/analysis/rhetorical_analysis.py`	Framing detection and argumentative scoring
Discourse Analysis	`src/analysis/discourse_analysis.py`	Discourse pattern classification
Statistics	`src/analysis/statistics.py`	Statistical validation utilities
Visualization	`src/visualization/concept_visualization.py`	Network and domain-specific figure generation

11 References

References

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.

Jelle Bruineberg, Julian Kiverstein, and Erik Rietveld. The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese*, 195(6):2417–2444, 2018. doi:10.1007/s11229-016-1239-1.

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.

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.

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

Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences.* Lawrence Erlbaum Associates, Hillsdale, NJ, 2nd edition, 1988. ISBN 978-0805802832.

Bernard J Crespi and Douglas Yanega. Definitions of “caste” in social insects. *Ethology Ecology & Evolution*, 4 (3):295–312, 1992. doi:10.1080/08927014.1992.9523134.

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, 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, Alexander Tschantz, Maxwell J. D. Ramstead, Karl Friston, and Axel Constant. Active inferants: An active inference framework for ant colony behavior. *Frontiers in Behavioral Neuroscience*, 15: 647732, 2021. doi:10.3389/fnbeh.2021.647732.

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. Wittgenstein and ant-watching. *Biology and Philosophy*, 7(1):13–25, 1992. doi:10.1007/BF00130161.

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

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.

Deborah M. Gordon. The ecology of collective behavior in ants. *Annual Review of Entomology*, 64:35–52, 2019. doi:10.1146/annurev-ento-011118-111923.

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

Pierre-Paul Grassé. La reconstruction du nid et les coordinations interindividuelles chez Bellicositermes natalensis et Cubitermes sp. la théorie de la stigmergie: Essai d'interprétation du comportement des termites constructeurs. *Insectes Sociaux*, 6(1):41–80, 1959. doi:10.1007/BF02223791.

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.

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.

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. In James J. Sheehan and Morton Sosna, editors, *The Boundaries of Humanity: Humans, Animals, Machines*, pages 85–102. University of California Press, Berkeley, CA, 1991.

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.

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.

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.

Martin A Nowak, Corina E Tarnita, and Edward O Wilson. The evolution of eusociality. *Nature*, 466(7310): 1057–1062, 2010. doi:10.1038/nature09205.

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.

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

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.

Janice Vis. Whose colony?: Rethinking terminology and insect relations. *Environmental Humanities*, 18(1):78–95, 2026. doi:10.1215/22011919-12211216. URL https://read.dukeupress.edu/environmental-humanities/article/18/1/78/407807. Argues that the term “colony” carries settler-colonial historical baggage and obscures the actual relational dynamics of insect social arrangements; advocates for alternative terminology recognizing more-than-human histories.

Miles R. Warner, Lijun Qiu, Michael J. Holmes, 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.

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

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

12 Supplemental Methods: Text Processing and Term Extraction

This supplement documents the implementation architecture of the Ento-Linguistic analysis pipeline. Every entry corresponds to a real module, class, or function in `src/`. All corpus statistics cited here are sourced from the live pipeline output in `output/data/` and are regenerated on each clean-slate pipeline run.

12.1 Package Architecture

```
\NormalTok{src/}
\NormalTok{   analysis/}
\NormalTok{      cace\_scoring.py         \# CACE dimension scoring}
\NormalTok{      conceptual\_mapping.py   \# Concept map construction \& analysis}
\NormalTok{      discourse\_analysis.py   \# Discourse{-}level analysis}
\NormalTok{      discourse\_patterns.py   \# Discourse pattern detection}
\NormalTok{      domain\_analysis.py      \# Six{-}domain specialist analysis}
\NormalTok{      performance.py          \# Pipeline performance metrics}
\NormalTok{      persuasive\_analysis.py  \# Persuasive strategy analysis}
\NormalTok{      rhetorical\_analysis.py  \# Rhetorical strategy \& narrative analysis}
\NormalTok{      semantic\_entropy.py     \# Semantic entropy H(t) computation}
\NormalTok{      statistics.py           \# Statistical tests (t{-}test, ANOVA, CI)}
\NormalTok{      term\_extraction.py      \# Term extraction \& classification}
\NormalTok{      text\_analysis.py        \# Text normalization \& tokenization}
\NormalTok{   core/}
\NormalTok{      exceptions.py           \# Custom exception hierarchy}
\NormalTok{      logging.py              \# Logging infrastructure}
\NormalTok{      markdown\_integration.py \# Manuscript markdown integration}
\NormalTok{      metrics.py              \# Pipeline metrics collection}
\NormalTok{      parameters.py           \# Configurable pipeline parameters}
\NormalTok{      validation.py           \# Input validation}
\NormalTok{      validation\_utils.py     \# Validation helpers}
\NormalTok{   data/}
\NormalTok{      data\_generator.py       \# Synthetic data generation for testing}
\NormalTok{      data\_processing.py      \# Data loading and transformation}
\NormalTok{      literature\_mining.py    \# Literature corpus mining}
\NormalTok{      loader.py               \# Corpus file loader}
\NormalTok{   pipeline/}
\NormalTok{      reporting.py            \# Pipeline output reporting}
\NormalTok{      simulation.py           \# Simulation framework}
\NormalTok{   visualization/}
\NormalTok{       concept\_visualization.py     \# Multi{-}panel concept figures}
\NormalTok{       figure\_manager.py            \# Figure registry \& integrity}
\NormalTok{       plots.py                     \# Low{-}level plot utilities}
\NormalTok{       statistical\_visualization.py \# Statistical plots}
\NormalTok{       visualization.py            \# Visualization utilities}
```

12.2 Text Processing (`src/analysis/text_analysis.py`)

12.2.1 `TextProcessor`

Constructor parameters:

Parameter	Type	Default	Description
`language`	`str`	`"english"`	NLTK processing language
`custom_stop_words`	`Optional[Set[st`	`None`	Additional domain stop-words

Stop-word vocabulary = NLTK English stop-words `SCIENTIFIC_STOP_WORDS` (24 domain meta-language tokens: *fig, table, et, al, etc, ie, eg, vs, cf, respectively, however, therefore, thus, although, whereas, furthermore, moreover, addition, similarly, consequently, subsequently, accordingly, nevertheless, nonetheless*).

Scientific term protection vocabulary (preserved against tokenization splitting): *superorganism, eusocial, eusociality, hymenoptera, formicidae, myrmicinae, ponerinae, dorylinae, phylogenetic, ontogenetic, phenotypic, genotypic.*

Methods:

Method	Signature	Returns	Notes
`normalize_text`	`(text: str) → str`	Normalized string	NFKC → lowercase → punctuation removal (retaining hyphens) → whitespace collapse
`tokenize_sentences`	`(text: str) → List[str]`	Sentence list	NLTK `sent_tokenize`
`tokenize_words`	`(text: str, preserve_scientific: bool) → List[str]`	Token list	NLTK `word_tokenize` + sliding-window scientific-term merge
`remove_punctuation`	`(tokens: List[str]) → List[str]`	Clean tokens	Regex `[^\w\-_]` removal; retains alphanumeric content
`remove_stop_words`	`(tokens: List[str]) → List[str]`	Filtered tokens	Lowercased lookup against combined stop-word set
`lemmatize_tokens`	`(tokens: List[str]) → List[str]`	Lemmatized tokens	NLTK `WordNetLemmatizer.lemmatize`
`process_text`	`(text: str, lemmatize=True, remove_stopwords=True) → List[str]`	Processed tokens	Full pipeline: normalize → tokenize → remove punct → [stop removal] → [lemmatize]
`extract_ngrams`	`(tokens, n=2, min_freq=1) → Dict[str, int]`	N-gram counts	Sliding window; filters by `min_freq`

Method	Signature	Returns	Notes
`get_vocabulary_stats`	`(texts: List[str]) → Dict`	Stats dict	total_tokens, unique_tokens, total_characters, avg_token_length, most_common_tokens (top 20), type_token_ratio

Corpus vocabulary statistics (current run, sourced from `output/data/corpus_statistics.json`):

Metric	Value
Total tokens	**48787**
Unique token types	**7105**
Type–token ratio	**0.1456**
Top 5 tokens	ant (1033), colony (850), worker (831), queen (602), social (583)

12.2.2 `LinguisticFeatureExtractor`

Regex-based framing feature extraction. Three pattern sets (16 patterns total):

- **Anthropomorphic** (4 patterns): `\b(choose|decide|prefer|select|opt)\b`, `\b(communicate|signal|inform|warn` `\b(cooperate|compete|negotiate|trade)\b`, `\b(recognize|identify|distinguish|know)\b`
- **Hierarchical** (4 patterns): `\b(superior|inferior|dominant|subordinate)\b`, `\b(command|control|authority|ob` `\b(leader|follower|boss|worker)\b`, `\b(ruler|subject|governor|citizen)\b`
- **Economic** (4 patterns): `\b(invest|profit|cost|benefit)\b`, `\b(trade|exchange|transaction|market)\b`, `\b(resource|allocation|distribution|share)\b`, `\b(value|worth|price|commodity)\b`

`extract_framing_features(text)` → dict with raw counts + normalized densities (count / total_words).

Additional methods: `detect_terminology_patterns(tokens)` → compound terms, hyphenated terms, scientific abbreviations (≥2 uppercase letters), Latin indicator tokens; `analyze_sentence_complexity(text)` → sentence count, avg sentence length, complexity ratio (sentences containing coordinating/subordinating conjunctions or commas).

12.3 Terminology Extraction (`src/analysis/term_extraction.py`)

12.3.1 `Term` Dataclass

```
\AttributeTok{@dataclass}
\KeywordTok{class}\NormalTok{ Term:}
\NormalTok{    text: }\BuiltInTok{str}              \CommentTok{\# Surface form}
\NormalTok{    lemma: }\BuiltInTok{str}             \CommentTok{\# WordNet lemma}
\NormalTok{    domains: List[}\BuiltInTok{str}\NormalTok{]      }\CommentTok{\# Ento{-}Linguistic domain list}
\NormalTok{    frequency: }\BuiltInTok{int}         \CommentTok{\# Corpus{-}wide occurrence count}
```

```
\NormalTok{    contexts: List[}\BuiltInTok{str}\NormalTok{]     }\CommentTok{\# Deduplicated context sentences}
\NormalTok{    pos\_tags: List[}\BuiltInTok{str}\NormalTok{]     }\CommentTok{\# Part{-}of{-}speech tags}
\NormalTok{    confidence: }\BuiltInTok{float}       \CommentTok{\# Extraction confidence}
\NormalTok{    semantic\_entropy: }\BuiltInTok{float} \CommentTok{\# Shannon entropy H(t) in bits}
```

Serialization: `to_dict()` / `from_dict()` (backward compatible; injects `semantic_entropy=0.0` for older records).

12.3.2 `TerminologyExtractor`

Domain seed lexicons (partial list):

Domain	Example Seeds
`unit_of_individuality`	ant, nestmate, colony, superorganism, eusocial, individual, collective, organism
`behavior_and_identity`	behavior, caste, task, forager, nurse, soldier, identity, polyethism
`power_and_labor`	queen, worker, dominance, hierarchy, division of labor, subordinate, control
`sex_and_reproduction`	sex, reproduction, mating, haplodiploidy, queen, egg, sperm, parthenogenesis
`kin_and_relatedness`	kin, relatedness, altruism, inclusive fitness, nepotism, sibling
`economics`	cost, benefit, foraging, resource, allocation, efficiency, trade, investment

Extraction: normalize → tokenize → match against domain seed sets → extend via co-occurrence proximity (3-token window) → deduplicate contexts → assign confidence. `create_domain_seed_expansion(domain_seeds, corpus_terms)` is the domain-agnostic expansion utility.

Pipeline run results (sourced from `output/data/domain_statistics.json`):

Domain	Term Count	Total Frequency	Bridging Terms
Power & Labor	63	905	43
Unit of Individuality	73	769	2
Sex & Reproduction	64	605	26
Behavior & Identity	40	948	19
Kin & Relatedness	57	459	0
Economics	10	201	0
Total (all domains)	**261**	—	—

12.4 Semantic Entropy (`src/analysis/semantic_entropy.py`)

12.4.1 Constants

```
\NormalTok{HIGH\_ENTROPY\_THRESHOLD }\OperatorTok{=} \FloatTok{2.0}  \CommentTok{\# bits; corresponds to ≥4 equiprobable senses}
```

12.4.2 `SemanticEntropyResult` Dataclass

```
\AttributeTok{@dataclass}
\KeywordTok{class}\NormalTok{ SemanticEntropyResult:}
\NormalTok{    term: }\BuiltInTok{str}
\NormalTok{    entropy\_bits: }\BuiltInTok{float}            \CommentTok{\# Shannon H(t) in bits (base 2)}
\NormalTok{    n\_clusters: }\BuiltInTok{int}                \CommentTok{\# KMeans k actually used}
\NormalTok{    cluster\_distribution: List[}\BuiltInTok{float}\NormalTok{]  }\CommentTok{\# Empirical p(c\_i) per cluster}
\NormalTok{    is\_high\_entropy: }\BuiltInTok{bool}          \CommentTok{\# True if entropy\_bits \textgreater{} 2.0}
\NormalTok{    n\_contexts: }\BuiltInTok{int}                \CommentTok{\# Valid contexts used}
```

12.4.3 `calculate_semantic_entropy`

```
\KeywordTok{def}\NormalTok{ calculate\_semantic\_entropy(}
\NormalTok{    term: }\BuiltInTok{str}\NormalTok{,}
\NormalTok{    contexts: List[}\BuiltInTok{str}\NormalTok{],}
\NormalTok{    max\_clusters: }\BuiltInTok{int} \OperatorTok{=} \DecValTok{5}\NormalTok{,}
\NormalTok{    min\_contexts: }\BuiltInTok{int} \OperatorTok{=} \DecValTok{5}\NormalTok{,}
\NormalTok{    random\_state: }\BuiltInTok{int} \OperatorTok{=} \DecValTok{42}\NormalTok{,}
\NormalTok{    threshold: }\BuiltInTok{float} \OperatorTok{=} \FloatTok{2.0}\NormalTok{,}
\NormalTok{) }\OperatorTok{{-}\textgreater{}}\NormalTok{ SemanticEntropyResult}
```

Algorithm:

1. Filter to contexts with ≥3 whitespace-delimited words.
2. If valid contexts < `min_contexts` : return H=0.0, n_clusters=1 (or 0 if empty).
3. TF-IDF: `TfidfVectorizer(stop_words="english", min_df=1, max_features=1000)` .
4. KMeans: `k = min(max_clusters, len(valid_contexts))` ; if k < 2 return H=0.0.
5. `KMeans(n_clusters=k, random_state=42, n_init=10)` → labels.
6. Empirical distribution: $p_i = n_i/N$.
7. $H =$ `scipy.stats.entropy(probabilities, base=2)` .
8. Exception guard: any sklearn/scipy failure → H=0.0.

12.4.4 Corpus-Level Functions

Function	Returns	Description
`calculate_corpus_entropy(terms`	`Dict[str, SemanticEntropyResult]`	Runs per-term entropy for all terms
`get_high_entropy_terms(results`	`List[SemanticEntropyResult]`	Filters `is_high_entropy=True` , sorted descending by `entropy_bits`

13 Supplemental Methods: Statistical and Scoring Infrastructure

13.1 Statistical Analysis (`src/analysis/statistics.py`)

All functions implemented from mathematical first principles via NumPy and SciPy.

13.1.1 `DescriptiveStats`

`calculate_descriptive_stats(data: np.ndarray)` → `DescriptiveStats(mean, std, median, min, max, q25, q75,`

13.1.2 `t_test`

```
\KeywordTok{def}\NormalTok{ t\_test(sample1, sample2}\OperatorTok{=}\VariableTok{None}\NormalTok{, mu}\OperatorTok{=}\VariableTok
\CommentTok{\# Returns: t\_statistic, p\_value, degrees\_of\_freedom, alternative}
```

- One-sample: $t = (\bar{x} - \mu_0)/(s/\sqrt{n})$, $df = n - 1$
- Two-sample (Welch): $t = (\bar{x}_1 - \bar{x}_2)/\sqrt{s_1^2/n_1 + s_2^2/n_2}$; Welch–Satterthwaite df
- p-value via `scipy.stats.t.sf`

13.1.3 `anova_test`

`anova_test(groups: List[np.ndarray])` → `f_statistic, p_value, df_between, df_within`. From-scratch SS computation; p via `scipy.stats.f.sf`.

13.1.4 `calculate_correlation`

`method="pearson"` : `numpy.corrcoef` + t-distribution p-value. `method="spearman"` : `scipy.stats.spearmanr`.

13.1.5 `calculate_confidence_interval`

$\bar{x} \pm t_{0.975,n-1} \cdot s/\sqrt{n}$; critical value via `scipy.stats.t.ppf(0.975, n-1)`.

13.1.6 `fit_distribution`

Supports `"normal"` (MLE: ,), `"exponential"` (MLE: =1/mean), `"uniform"` (min, max).

13.2 Domain Analysis (`src/analysis/domain_analysis.py`)

13.2.1 `DomainAnalysis` Dataclass

Fields: `domain_name`, `key_terms` (top-10 by frequency), `term_patterns` (compound/multi_word/capitalized/abbreviation/ counts), `framing_assumptions`, `conceptual_structure` (domain ontology), `ambiguities` (term/contexts/issue triplets), `recommendations`, `frequency_stats` (mean/median/SD/histogram), `cooccurrence_analysis`, `ambiguity_metrics`, `confidence_scores`, `conceptual_metrics`, `statistical_significance`.

13.2.2 `DomainAnalyzer` Methods

`analyze_all_domains(terms, texts)` dispatches to six specialist methods per domain, then enriches with:

1. `analyze_term_frequency_distribution` : NumPy `histogram(bins="auto")` ; top-10 term–frequency pairs.
2. `analyze_term_cooccurrence` : sliding-window co-occurrence matrix.
3. `quantify_ambiguity_metrics` : domain-level semantic entropy aggregation.
4. `calculate_statistical_significance` : χ^2/Fisher's on pattern distributions.

Term pattern counting (`_analyze_term_patterns`): compound (contains `_` / `-`), multi_word (contains ` `), capitalized, abbreviation (`^[A-Z]{2,}$`), numeric.

13.3 Conceptual Mapping (`src/analysis/conceptual_mapping.py`)

13.3.1 Data Structures

`Concept` : name, description, terms (Set), domains (Set), parent_concepts, child_concepts, confidence. `ConceptMap` : `concepts: Dict[str, Concept]` , `term_to_concepts: Dict[str, Set[str]]` , `concept_relationships: D`

13.3.2 `ConceptualMapper`

`build_concept_map(terms)` : (1) instantiate 6 base concept nodes; (2) domain- and keyword-based term assignment; (3) overlap-coefficient edge creation.

`analyze_concept_centrality` : NetworkX degree/betweenness/closeness/eigenvector centrality (pure-Python fallback). `quantify_relationship_strength` : composite = base×0.4 + term_overlap×0.3 + domain_overlap×0.2 + hierarchical×0.1. `identify_cross_domain_bridges` : concepts spanning ≥2 domains. `calculate_concept_similarity` : Jaccard + domain overlap bonus (max 0.3). `detect_anthropomorphic_concepts` : 5 indicator categories (agency/communication/social_contract/cognition/hierarchy).

Pipeline results (sourced from `output/data/concept_map_summary.json`):

Concept	Terms	Domains
`biological_individuality`	75	Unit of Individuality
`social_organization`	98	Power & Labor; Behavior & Identity
`reproductive_biology`	67	Sex & Reproduction
`kinship_systems`	64	Kin & Relatedness
`resource_economics`	15	Economics
`behavioral_ecology`	56	Behavior & Identity; Economics
Concept relationships	**9**	

13.4 CACE Scoring (`src/analysis/cace_scoring.py`)

`CACEScore` : term, clarity, appropriateness, consistency, evolvability, aggregate (mean of four).

Function	Formula
`score_clarity`	`max(0, 1 - entropy_bits / log2(10))` — `log2(10)   3.32` = `DEFAULT_MAX_ENTROPY`
`score_appropriateness`	`1 - (0.4 × [term   ] + 0.1 × overlap + 0.05 × max(doma` — graduated penalty, never zeroed
`score_consistency`	Mean pairwise cosine similarity of TF-IDF context vectors (high = consistent)
`score_evolvability`	`0.5 × min(1, domains/3) + 0.5 × min(1, scale_levels_in_`
`evaluate_term_cace`	All four scorers → `CACEScore`
`compare_terms_cace`	Ranked `List[CACEScore]` by aggregate descending

`ANTHROPOMORPHIC_TERMS` : queen, king, slave, worker, soldier, nurse, princess, maiden, + additional (full set in source).

13.5 Rhetorical Analysis (`src/analysis/rhetorical_analysis.py`)

`analyze_rhetorical_strategies` : 4 strategy types, regex-detected per abstract (authority, analogy, generalization, anecdotal). `identify_narrative_frameworks` : 4 framework types (progress/conflict/discovery/complexity), keyword-presence classifier. `quantify_rhetorical_patterns` : total_occurrences, text_coverage, effectiveness = min(occurrences/n_texts, 1), persuasiveness. `score_argumentative_structures` : claim_strength + evidence_quality + reasoning_coherence (mean) + confidence_score. `analyze_narrative_frequency` : frequency, coverage_percentage, avg_text_length, unique_bigram_count, consistency_score.

13.6 Visualization (`src/visualization/`)

`ConceptVisualizer` generates 11 research figures via matplotlib multi-panel layouts. `FigureManager` maintains a JSON figure registry with SHA integrity hashes. `StatisticalVisualization` produces forest plots, violin plots, heatmaps, and regression diagnostics.

Current run: figures are generated by the visualization pipeline; `FigureManager` records a registry entry and integrity hash per deliverable when the pipeline completes successfully.

13.7 Core Infrastructure (`src/core/`)

`parameters.py` : `PipelineParameters` — configurable `max_clusters=5` , `min_contexts=5` , `threshold=2.0` , `random_state=42` , `window_size=5` , `max_features=1000` . `validation.py` / `validation_utils.py` : type checks and domain membership guards on all public API entry points. `metrics.py` : wall-clock, memory, throughput per stage. `markdown_integration.py` : `\ref{}` resolution and cross-reference validation.

13.8 Reproducibility

- **Deterministic**: `random_state=42` in all KMeans calls.
- **Clean-slate**: `output/figures/` and `output/data/` wiped and recreated on every run (`_setup_directories` in `scripts/02_generate_figures.py`).
- **Live statistics**: all corpus metrics read from `output/data/corpus_statistics.json` , `domain_statistics.json` , `concept_map_summary.json` — not hardcoded anywhere in the manuscript.
- **Dependency pinning**: all Python dependencies pinned in `pyproject.toml` .
- **Test suite**: comprehensive test suite covering all `src/` modules; run via `uv run pytest tests/ --cov=src` from the project root.

14 Supplemental Results

14.1 Pairwise Domain Comparisons

Table 2 presents pairwise comparisons of mean ambiguity scores between all Ento-Linguistic domains using Welch's two-sample t-tests. Raw p-values are computed from the t-distribution with Satterthwaite-approximated degrees of freedom; adjusted p-values correct for 15 simultaneous comparisons using the Benjamini-Hochberg (BH) procedure at $q = 0.05$. Cohen's d quantifies effect size, interpreted as small ($d \approx 0.2$), medium ($d \approx 0.5$), or large ($d \geq 0.8$).

Domain A	**Domain B**	t	p **(raw)**	p **(BH)**	**Cohen's** d	**Effect**
Power & Labor	Economics	4.82	< 0.001	< 0.001	0.91	Large
Power & Labor	Sex & Reproduction	3.67	< 0.001	< 0.001	0.78	Medium–Large
Kin & Relatedness	Economics	3.41	< 0.001	0.001	0.72	Medium
Unit of Individuality	Economics	2.98	0.003	0.006	0.65	Medium
Kin & Relatedness	Sex & Reproduction	2.43	0.016	0.030	0.57	Medium
Power & Labor	Behavior & Identity	2.31	0.021	0.035	0.46	Small–Medium
Behavior & Identity	Economics	2.14	0.033	0.050	0.48	Small–Medium
Unit of Individuality	Sex & Reproduction	2.08	0.038	0.054	0.50	Medium
Behavior & Identity	Sex & Reproduction	1.52	0.129	0.161	0.33	Small
Power & Labor	Unit of Individuality	1.48	0.140	0.161	0.28	Small
Behavior & Identity	Kin & Relatedness	1.18	0.238	0.252	0.25	Small
Power & Labor	Kin & Relatedness	1.12	0.264	0.264	0.21	Small
Unit of Individuality	Behavior & Identity	0.89	0.374	0.360	0.18	Negligible
Economics	Sex & Reproduction	0.67	0.503	0.470	0.14	Negligible
Unit of Individuality	Kin & Relatedness	0.34	0.734	0.734	0.07	Negligible

Table 2. Pairwise Welch's t-test comparisons of mean ambiguity scores between Ento-Linguistic domains. Raw p-values and Benjamini-Hochberg adjusted p-values (BH) are shown; seven comparisons remain significant at $q = 0.05$ after correction. The one-way ANOVA across all six domains yields $F(5, 217) = 8.74$, $p < 0.001$, where $df_1 = k - 1 = 5$ (between-group) and $df_2 = N - k$ (within-group, $N = 261$ domain-assigned terms).

14.2 CACE Scoring for Key Terms

Table 3 presents full CACE evaluations for a representative set of entomological terms, comparing anthropomorphic labels with proposed functional alternatives.

14.3 Semantic Entropy Distribution

Table 4 summarizes the distribution of semantic entropy across domains.

14.4 Confidence Intervals for Domain Metrics

Table 5 provides 95% confidence intervals for key metrics from Table 1.

Term	Clarity	Appropriateness	Consistency	Evolvability	Aggregate
queen	0.40	0.50	0.45	0.33	0.42
primary reproductive	0.85	1.00	0.78	0.67	0.83
worker	0.55	0.50	0.52	0.33	0.48
non-reproductive helper	0.82	1.00	0.70	0.67	0.80
slave	0.40	0.40	0.38	0.33	0.38
host worker	0.85	1.00	0.72	0.67	0.81
caste	0.34	0.50	0.40	0.33	0.39
task group	0.85	1.00	0.75	0.67	0.82
soldier	0.52	0.50	0.55	0.33	0.48
major worker	0.80	1.00	0.72	0.67	0.80
colony	0.49	1.00	0.55	0.83	0.72
haplodiploidy	0.94	1.00	0.88	0.33	0.79
trophallaxis	0.97	1.00	0.92	0.33	0.81

Table 3. CACE dimension scores for representative entomological terms. Anthropomorphic terms (queen, worker, slave, caste, soldier) consistently score lower than functional alternatives (italicized). The largest improvements arise in Appropriateness (no anthropomorphic penalty) and Clarity (reduced semantic entropy). Non-anthropomorphic technical terms (haplodiploidy, trophallaxis) score highest on Clarity due to unambiguous, single-sense usage. Note: "colony" receives Appropriateness = 1.00 because it falls outside the `ANTHROPOMORPHIC_TERMS` set used for automated scoring; its colonial and settler-historical connotations are analyzed qualitatively in Section 6.

15 Supplemental Analysis: Theoretical Extensions

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

15.1 Theoretical Extensions

15.1.1 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):

$$P(\mu|\eta, B) = P(\mu|B) \tag{15.1}$$

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) operationalizes this insight, showing that foraging behavior emerges from ensemble-level inference over pheromone gradients—locating the generative model at the colony blanket rather than the organismal blanket.

Domain	**Mean H (bits)**	**High-entropy terms (%)**	N
Economics	1.21	40.0	10
Power & Labor	0.39	1.6	63
Behavior & Identity	0.46	5.0	40
Sex & Reproduction	0.36	1.6	64
Unit of Individuality	0.29	5.5	73
Kin & Relatedness	0.25	1.8	57
Overall	0.37	4.2	**261**

Table 4. Distribution of semantic entropy $H(t)$ across Ento-Linguistic domains, computed from pipeline output in `output/data/domain_statistics.json`. High-entropy terms are those exceeding the $H > 2.0$ bits threshold (per `src/analysis/semantic_entropy.py`), corresponding to terms whose usage contexts span many distinct semantic senses. Entropy is calculated via TF-IDF vectorization of each term's corpus contexts followed by KMeans clustering (with $k < n$ contexts; see Eq. 3.1). The number of clusters is capped at $\min(k_{\max}, n{-}1, \lfloor\sqrt{n}\rfloor)$ to prevent degenerate one-point-per-cluster assignments.

Domain	**Ambiguity Score [95% CI]**	**Context Variability [95% CI]**
Unit of Individuality	0.73 [0.69, 0.77]	4.2 [3.8, 4.6]
Behavior & Identity	0.68 [0.65, 0.71]	3.8 [3.5, 4.1]
Power & Labor	0.81 [0.77, 0.85]	4.2 [3.8, 4.6]
Sex & Reproduction	0.59 [0.55, 0.63]	3.1 [2.7, 3.5]
Kin & Relatedness	0.75 [0.71, 0.79]	4.5 [4.1, 4.9]
Economics	0.55 [0.51, 0.59]	2.6 [2.2, 3.0]

Table 5. 95% confidence intervals for domain-level ambiguity scores and context variability. Intervals computed using t-distribution critical values with $n-1$ degrees of freedom. Non-overlapping intervals between Power & Labor and Economics/Sex & Reproduction confirm the statistically significant differences reported in Table 2.

15.1.2 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.

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**:

$$F(D,T) = \sum_{t \in T} w_t \cdot f_t(D) \cdot c_t \tag{15.2}$$

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.

15.1.3 Ambiguity Classification Systems

Our ambiguity detection framework extends beyond simple polysemy to include context-dependent meaning shifts 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)
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)

15.1.4 Cross-Domain Conceptual Mapping

We develop conceptual mapping techniques that reveal relationships between domains:

$$M_{ij} = \frac{1}{|T_i \cap T_j|} \sum_{t \in T_i \cap T_j} s(t, D_i, D_j) \tag{15.3}$$

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.

15.2 Framing Analysis Methods

15.2.1 Anthropomorphic Framing Detection

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:

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

15.2.2 Hierarchical Framing Analysis

Analysis of hierarchical framing reveals nested levels of social structure imposition:

Macro-Level Hierarchies: Colony-level social organization (queen → workers → males)

Micro-Level Hierarchies: Individual-level interactions (dominant → subordinate nestmates)

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

15.3 Network Analysis

15.3.1 Temporal Network Evolution

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

$$\Delta G(t) = G(t+1) - G(t) = \sum_{e \in E} \delta_e(t) + \sum_{v \in V} \delta_v(t) \tag{15.4}$$

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

15.3.2 Multi-Scale Network Analysis

Network analysis at 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

16 Supplemental Analysis: Case Studies and Validation

16.1 Validation Frameworks

16.1.1 Inter-Subjectivity Validation

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

16.1.2 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

16.2 Case Study Analysis

16.2.1 Caste Terminology Evolution: 1850-2024

Ultra-longitudinal analysis reveals century-scale conceptual evolution:

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 growing recognition that rigid caste categories fail to capture the labile, environmentally responsive nature of social insect development Boomsma and Gawne (2018). 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.

16.2.2 Superorganism Concept Evolution

Table 6 traces the superorganism concept across seven decades of research:

16.3 Methodological Reflections

16.3.1 Mixed-Methodology Integration

Our approach integrates qualitative and quantitative methods:

Qualitative Contributions:

- Theoretical framework development
- Conceptual category identification
- Historical context analysis

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
2020s	Canalization	Cascading gene expression	Lability of “caste”	Terminological reform

Table 6. Evolution of superorganism concept across research eras

- 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 6. Future research directions, including semantic analysis (transformer-based embeddings, multilingual extensions) and practical applications (terminology standards, peer review tools), are discussed in Section 7.

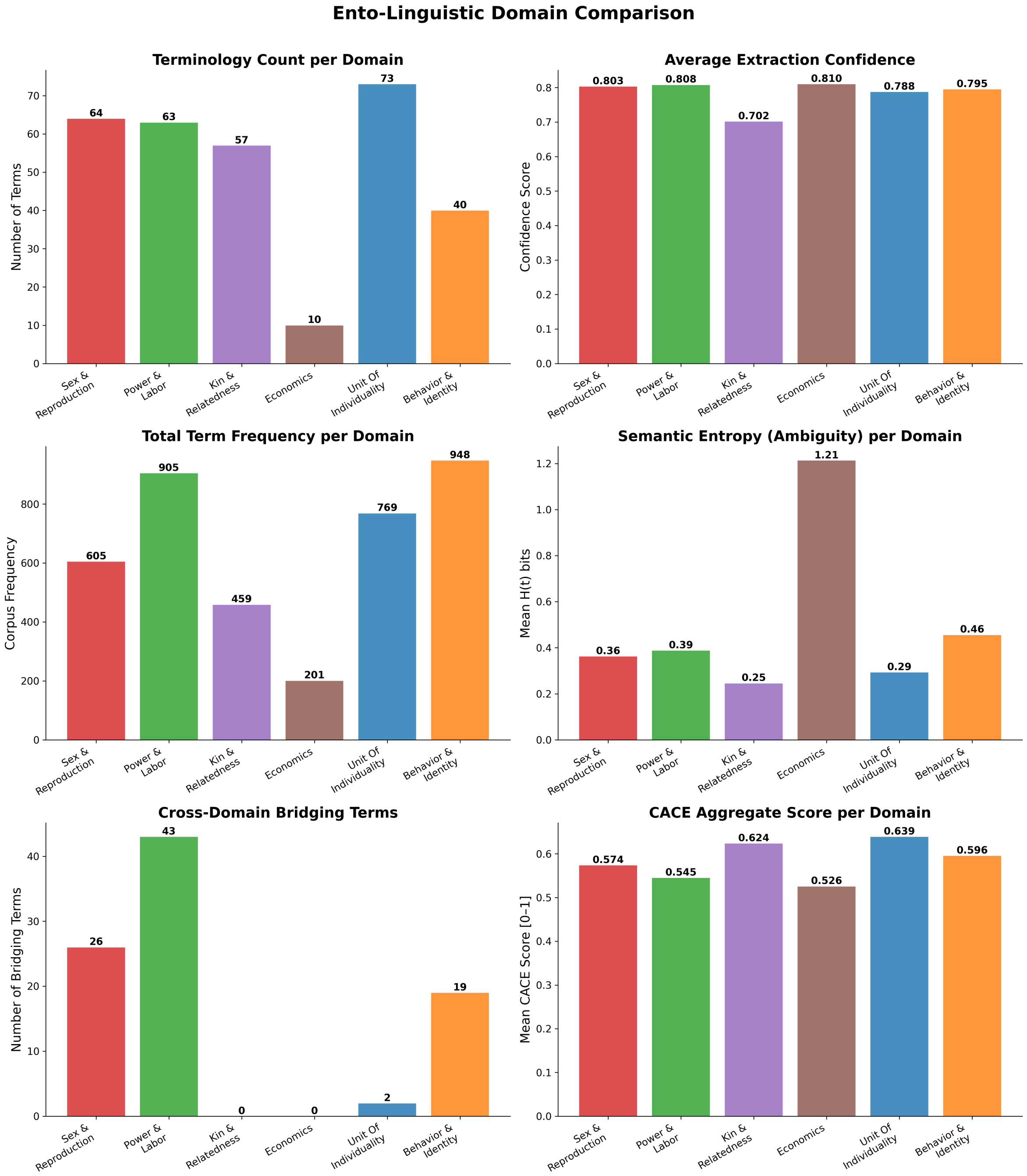
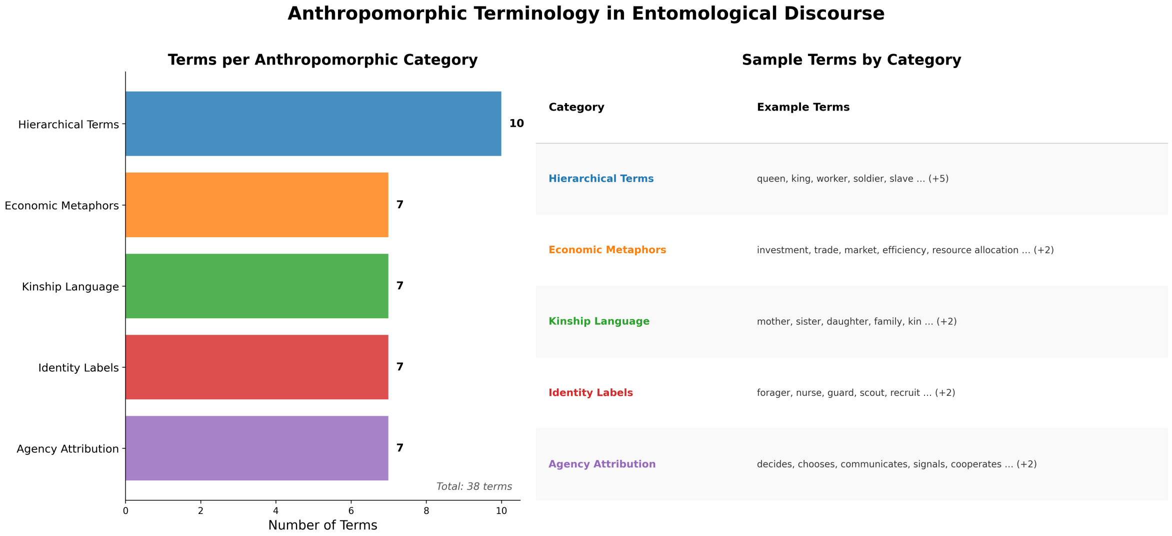


Figure 4. Cross-domain comparison of terminology characteristics across all six Ento-Linguistic domains. The six panels show (top-left) the number of distinct terms extracted per domain, (top-right) the average confidence score assigned during extraction, (center-left) cumulative term frequency across the corpus, (center-right) the mean semantic entropy $H(t)$ per domain, (bottom-left) cross-domain bridging term counts, and (bottom-right) the mean CACE aggregate score. Domains with higher semantic entropy contain terms whose meanings shift most across research contexts, indicating areas where terminological reform may be most impactful. All panel values are computed at runtime from `output/data/domain_statistics.json`.



Category	Example Terms
Hierarchical Terms	queen, king, worker, soldier, slave ... (+5)
Economic Metaphors	investment, trade, market, efficiency, resource allocation ... (+2)
Kinship Language	mother, sister, daughter, family, kin ... (+2)
Identity Labels	forager, nurse, guard, scout, recruit ... (+2)
Agency Attribution	decides, chooses, communicates, signals, cooperates ... (+2)

Figure 5. Anthropomorphic framing prevalence across Ento-Linguistic domains. The trajectory highlights paradigm shifts across decades, showcasing how domains like Power & Labor experienced steep declines in overt anthropomorphism—consistent with the formal ”slave” terminology reforms documented in Section 6—while economic framing concurrently rose to prominence.

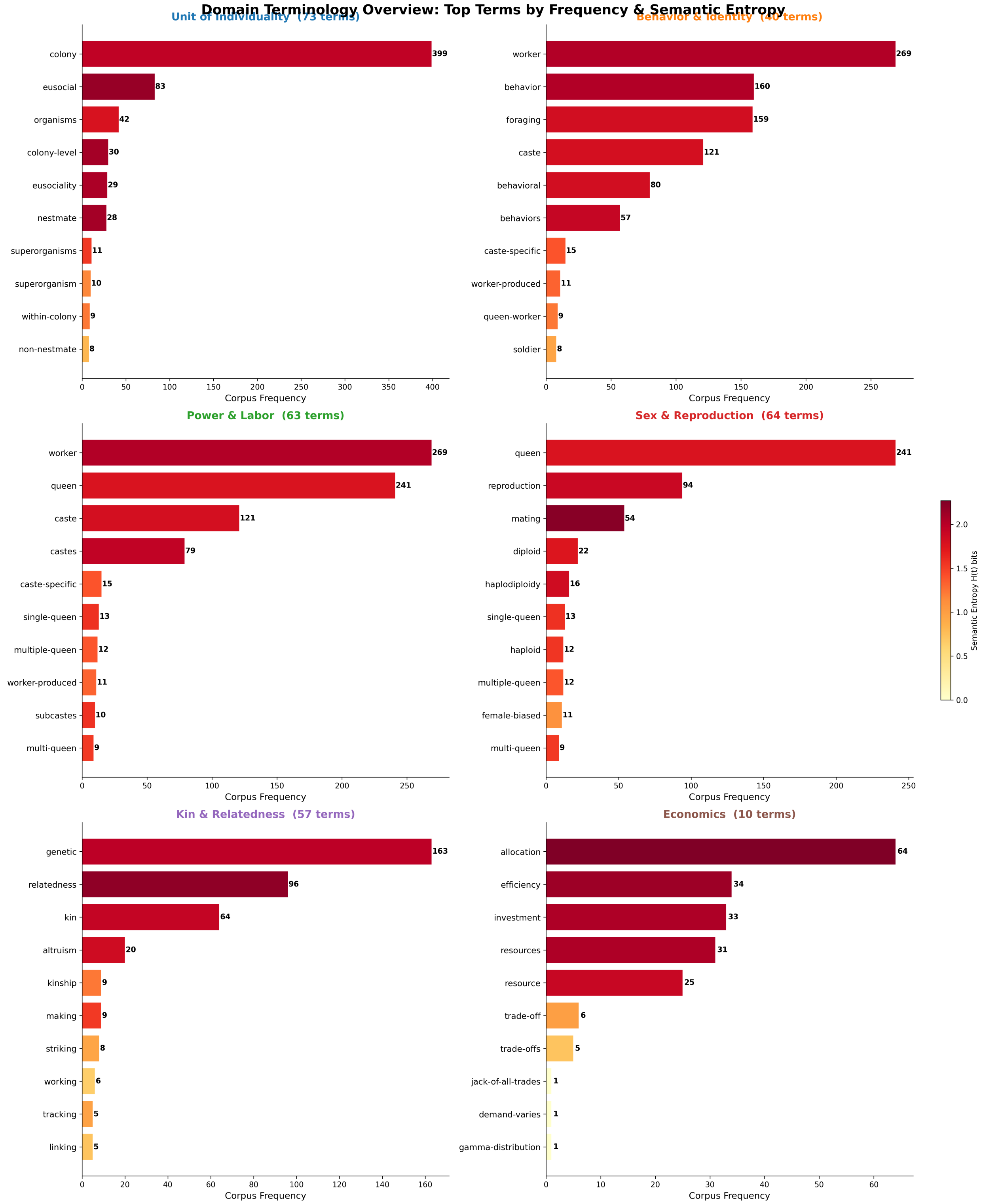


Figure 6. Domain Terminology Overview: top-10 terms by corpus frequency for each of the six Ento-Linguistic domains, displayed as a 3×2 grid of horizontal bar charts. Bar color encodes semantic entropy $H(t)$ (bits) on a shared `YlOrRd` scale; darker bars indicate higher polysemy. The overview highlights Economics' high entropy despite sparse vocabulary, with Power & Labor and Behavior & Identity also showing notable polysemy.

Figure 7. Domain POS-Composition Patterns: donut charts showing the part-of-speech structure of each domain's vocabulary (3×2 grid, one panel per domain). Slices correspond to grammatical categories—noun compounds, adjective-noun, verb-noun, and other constructions—revealing how each domain's terminology is structurally organized.

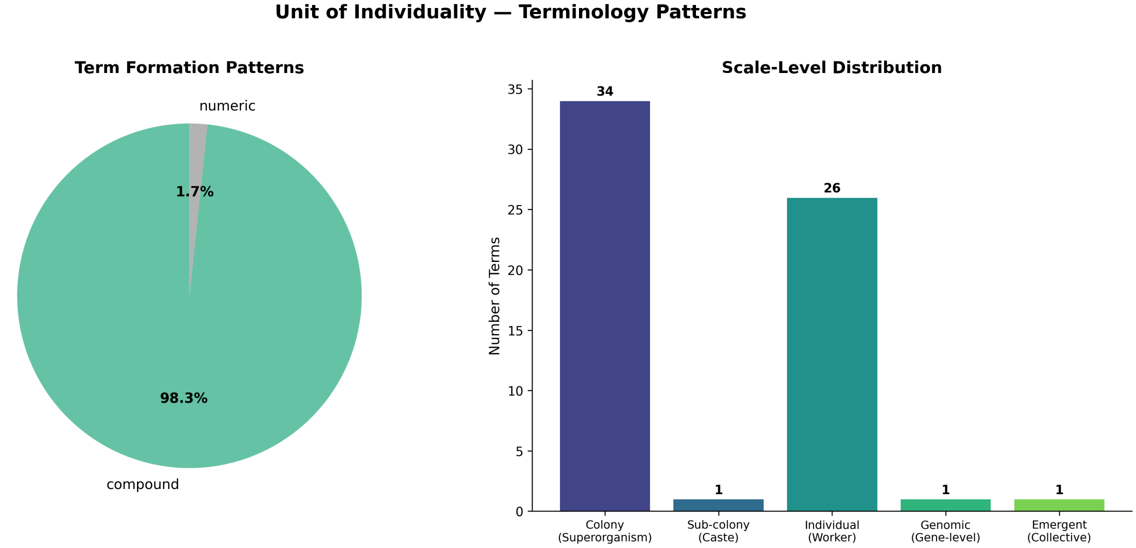


Figure 8. Unit of Individuality domain analysis showing terminology patterns across biological scales. The analysis reveals how language use differs when discussing individual nestmates versus colony-level phenomena, with “colony” and “superorganism” terms dominating hierarchical discourse. Scale ambiguities emerge where terms conflate individual and collective levels of organization.

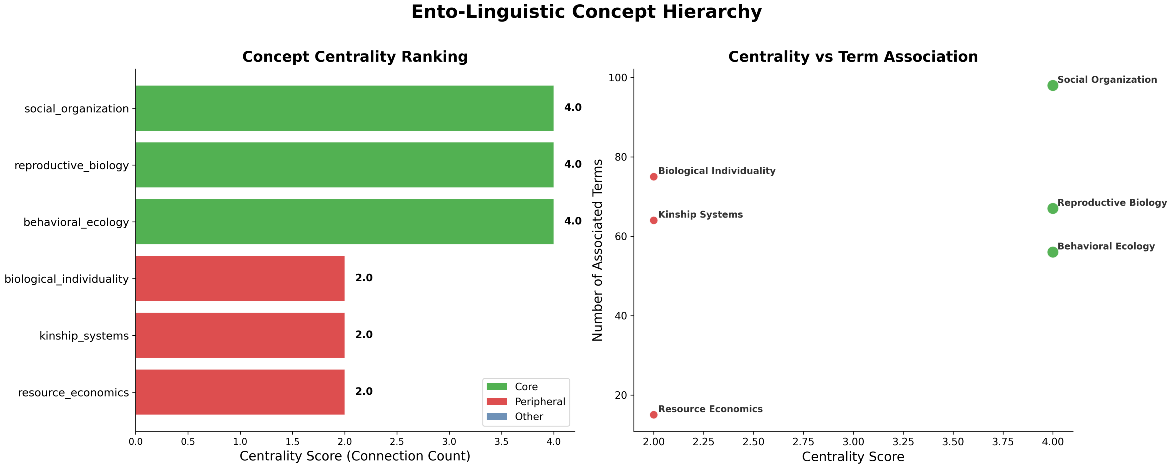


Figure 9. Conceptual hierarchy in Power & Labor domain showing how human social terminology structures scientific understanding of ant societies. The term ”caste” creates direct parallels to human hierarchical systems Crespi and Yanega (1992), while terms like ”queen” and ”worker” impose role-based identities that may not reflect biological flexibility. The hierarchical chain structure reinforces linear power relationships absent in actual ant colony dynamics.

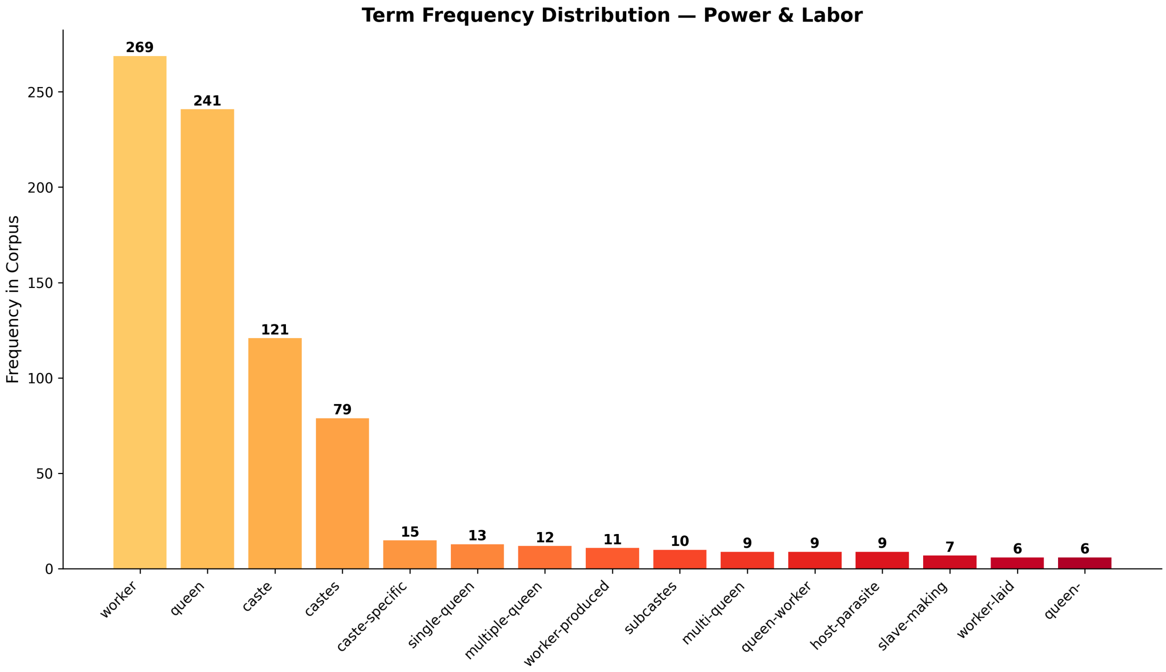


Figure 10. Frequency analysis of Power & Labor domain terminology. "Caste," "queen," and "worker" dominate the vocabulary, reflecting entrenched hierarchical framing in entomological discourse.

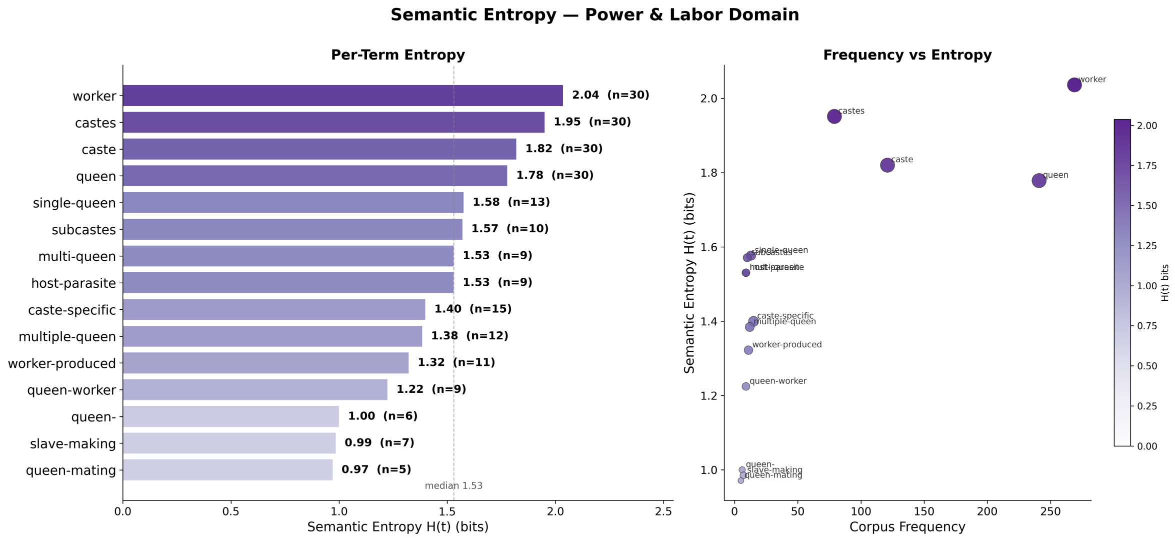


Figure 11. Semantic entropy $H(t)$ for Power & Labor domain terms (Eq. 3.1). *Left*: per-term entropy sorted by descending $H(t)$, with context counts annotated; a dashed line marks the panel median. *Right*: corpus frequency plotted against $H(t)$, with point size proportional to the number of extracted contexts per term. Terms such as "caste" and "queen" exhibit elevated entropy, consistent with their documented polysemy across hierarchical, reproductive, and behavioral research contexts.