Supplemental Methods

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

S1.1 Text Processing Pipeline Implementation

S1.1.1 Multi-Stage Text Normalization

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

$$T_{\text{normalized}} = \text{lowercase}(\text{strip}_p unct \text{strip}_p unct \text{strip}_p unct_{\text{strip}_p unct(}$$

where T represents raw text input and each transformation step standardizes linguistic variation while preserving semantic content.
Tokenization Strategy: We employ domain-aware tokenization that recognizes scientific terminology:

$$\tau(T) = \bigcup_{t \in T} \begin{cases} t & \text{if } t \in \mathcal{T}_{\text{scientific}} \\ \text{word\_tokenize}(t) & \text{otherwise} \end{cases} \tag{2}$$