



From Dugongs to Quetzals to Sequoias: How Global Mythology Quantitatively Embeds Humanity's Relationship with Nature

*Autumn Sehy, Capstone Research
Brandeis Computational Linguistics MS Student*



01. Introduction

Imagine sitting through the dullest lecture — how much information would you remember? Before books or the internet, people needed to pass on crucial survival knowledge about the environment. Mythology served as a vivid way to transmit ecological information — ensuring life-saving knowledge could be retained and shared across generations.

02. Objective

To prove flora and fauna are not random, but encode survival knowledge in mythology that is quantifiable with modern computational techniques and comparable with ecological phenomena.

03. Methodology

- Stanza POS tagger + NER + LLM classifiers to differentiate instances like ash (fire substance vs tree), asphodel (flower vs location)
- Pretrained BERT with masking for emotion classification
- CBow word embeddings to compare vectors of species (in progress)
- Dataset of 1000 species of flora/fauna including plurals + ecological features, not limited to: Trophic levels (apex predator vs primary consumer), species family, species wakefulness, species rarity, species size, species affordances to humans
- AllenNLP + Propbank + Verbnet for semantic role labeling (SRL).
Ex: *She [experiencer] --> transformed into --> a dugong [result]*.

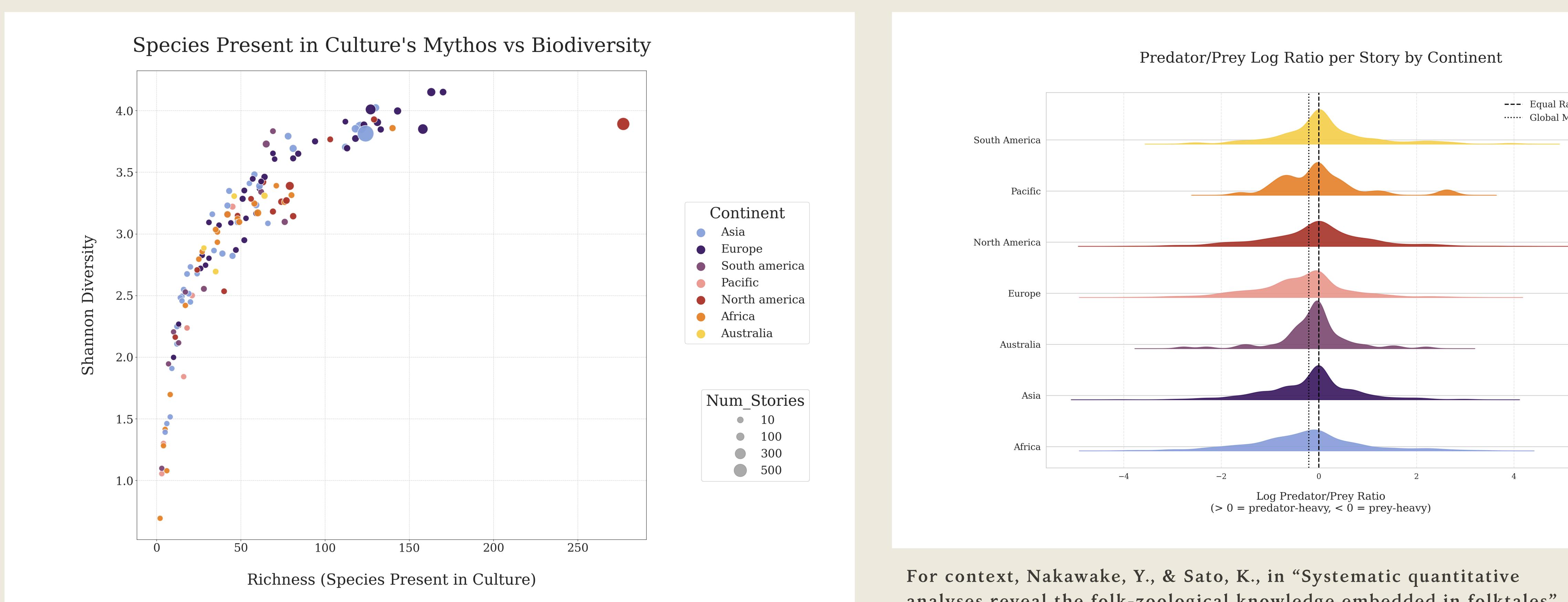
Species Extracted

	73897 Mammals
	21098 Birds
	253 Molluscs
	3773 Small Creatures
	4656 Marine Creatures
	15248 Trees
	22670 Various Flora

Myths Extracted

134 Cultures 6750 stories		
Continent	Cultures	Stories
Asia	36	2573
Africa	23	621
Europe	35	1862
N America	19	1355
S America	12	259
Pacific	6	86
Australia	4	94

04. Results/Findings - See binders for breakdowns on individual species and cultures!



Each dot is a culture

Shannon Diversity = entropy of how many species are in a culture's stories, and how evenly they're mentioned

For context, Nakawake, Y., & Sato, K., in "Systematic quantitative analyses reveal the folk-zoological knowledge embedded in folktales" found predator/prey co-occur often, and alongside the motif of 'deception'

05. Analysis

- Cultures with more stories/species in myths show similar biodiversity regardless of geography, following the ecological principle of Species Area Relationships when applied to large-scale regions
- Humans prefer a predator/prey ratio of 7.2 predators/10 prey - overemphasizing predators to reality, perhaps for dramatic effect or safety, mirroring how human brains adapted for predator recognition
- Charismatic/polarizing/domestic fauna dominating science funding & public appeal are what we anthropomorphize AND are the most popular creatures in myth, following a Zipf's law of distribution
- Trees are omnipresent, often locations - likely water or destination markers, yet deciduous trees are about 25% more agentic than conifers
- Cash crops or staple foods are the most common flora besides trees, yet ChatGPT misses this foundational role, saying plants 'connect realms, alter states of consciousness, or grant knowledge/immortality' - perhaps reflecting a deep bias in the training data/the internet
- Well known 'tricksters' - ravens, spiders, coyotes, foxes- are often neither apex predators nor prey, but are secondary or tertiary consumers. Major exception: rabbits
- LLMs almost never share ecological roles of creatures in myth - perhaps following trends of NLP using Proppian/Campbellian archetypes, which were developed on European stories & use moral roles like 'hero'



06. Conclusion

Myth collection was the most difficult part of this, and improving accessibility is crucial. Furthermore, computational research delving into cultural aspects such as belief systems, magic use, and stereotypes, could provide more context to the dimensions these flora/fauna exist within.