Modeling Eco-Poetics and Eco-Politics in 20th Century Anglophone Climate Fiction:
Toxic Water

Miller, Dez Mary

dezmmiller@gmail.com
Emory University, United States of America

Wermer-Colan, Henry Alexander

alex.wermer-colan@temple.edu
Temple University, United States of America

Stefan, SaraGrace

saragrace.stefan@temple.edu
Temple University, United States of America

Kane, Megan

megan.kane@temple.edu
Temple University, United States of America

How do concerns about environmental pollution manifest in literature? In this collaborative cross-university study, we develop a cultural analytics methodology for exploring representations of climate change and pollution in speculative fiction by modeling semantic associations between water and toxicity in three comparison corpora housed in Hathi Trust. In the process, we explore the potential of machine learning methods for natural language processing, topic modeling and vector space modeling to analyze differences in ecopoetics across post-45 Anglophone science fiction genres and subgenres.

Our study takes a closer look at water flows and toxicity given the relevance of these themes during this historical period. Results from initial topic modeling experiments further indicated waterways were a significant subject for post-45 climate fiction. The Clean Water Act of 1972 in the United States was the result of decades of activism fighting widespread trash, sewage, and chemical pollution of U.S. waterways, and activism has continued since that time. Our project explores how literature responds to these changes in historical discourses and material realities.

The centerpiece of this analysis will be a climate fiction ("cli-fi") corpus: an approximately 200-volume selection of post-1945 mass-market science/speculative fiction ("SF") novels which have been recently digitized at Temple University Libraries (Wermer-Colan and Kopacewski, 2022). We compare the Temple cli-fi corpus to two other Hathi Trust recommended worksets: a 4509-volume post-45 subset of the 5161-volume 20th Century Speculative Fiction workset curated by David Mimno and Laure Thompson, and a 849-volume post-45 subset of the 2,730-volume NovelTM Dataset for English-Language Fiction curated by Patrick Kimutis, Ted Underwood, and Jessica Witte, which itself is a randomized subset of a larger 138,164 volume list. By analyzing works by such authors as Ursula K. Le Guin, Samuel R. Delany, Philip K. Dick, and many less well known, this study explores the development of cli-fi as a subgenre of SF and its unique modes of representing environmental concerns and themes.

Our cultural analytics methodology proceeds in the following steps, with each stage of machine learning involving iterative process of close reading relevant passages and narratives . 1) We use Gensim and BERTopic to build topic models of each corpus, filtering and clustering the diverse corpora into subsets of represented ecosystems, enhancing generic distinctions in existing Library of Congress metadata. 2) Using BookNLP and related entity recognition algorithms, we develop a lexicon for rivers and all words correlated with rivers, from water to brook, and a lexicon for toxicity and related terminology for pollution. 3) We perform Gensim word2vec and BERT contextual embedding modeling of each corpus, exploring the associations of water, toxicity, and their intersections in each corpus. 4) We run the Gensim word2vec and BERT embedding modeling again, this time on subsets of each corpora determined via the initial topic modeling to be associated with water- and river-related topics. Final outputs from the project include: 1) a complete disaggregated data set of extracted features for the Temple cli-fi corpus, as well as machine learning models generated from all three corpora and made publicly available online; 2) several Google Colab Jupyter Notebooks with Python code accompanying the cli-fi data that would enable others to do topic modeling and word2vec analyses and visualizations of these corpora and/or other corpora available in HathiTrust's Research Center; 3) a full-length research paper, the short version of which will be presented at ADHO 2023.

At the heart of this project are questions about the potential and limitations of computational text analysis for literary criticism within the environmental humanities. Other recent computational text analysis research investigating the role of nature in culture use machine learning to analyze trends in hard counts of references to natural elements, depicting these as representations of "societal awareness" toward biodiversity (Langer et al, 2021), of the amount of biodiversity in literature (Piper, 2022), or of "social connection" to nature (Kesebir and Kesebir, 2017). Our study aims to contribute to these conversations within the burgeoning field of digital environmental humanities by studying semantic dynamics within carefully curated corpora housed in Hathi Trust.

Bibliography

Kesebir, Selin, and Kesebir, Pelin. (2017): "A Growing Disconnection From Nature Is Evident in Cultural Products," in *Perspectives on Psychological Science* 12.2 (2017): 258–269.

Langer, Lars, et al. (2021): "The Rise and Fall of Biodiversity in Literature: A Comprehensive Quantification of Historical Changes in the Use of Vernacular Labels for Biological Taxa in Western Creative Literature," in *People and Nature*, vol. 3, no. 5.

Piper, Andrew. (2022): "Biodiversity Is Not Declining in Fiction." in *Journal of Cultural Analytics*, vol. 7, no. 3. doi:10.22148/001c.38739 .

Underwood, Ted, Patrick Kimutis, and Jessica Witte. (2020): "NovelTM Datasets for English-Language Fiction, 1700-2009." in *Journal of Cultural Analytics* 5 (2). https://doi.org/10.22148/001c.13147 .

Alex Wermer-Colan and James Kopaczewski. (2022): "The New Wave of Digital Collections: Speculating on the Future of Library Curation," in *Transactions of the American Philosophical Society*, vol. 110, no. 3, pp. 211–41.