

## CHAPTER 4

# Showing the Potential of Computational Analysis to Support Environmental Narrative Research

Ross S. Purves

Department of Geography; URPP Language and Space, University of Zurich,  
Switzerland

Olga Koblet

Department of Geography, University of Zurich, Switzerland

Benjamin Adams

Department of Computer Science and Software Engineering, University of  
Canterbury, Christchurch, New Zealand

In the second half of this book, we present seven case studies where multidisciplinary research groups explore research questions by examining environmental narratives. Six of the case studies were the result of the workshop organised to initiate the process of writing this book. At the workshop, groups of authors with diverse backgrounds came together and identified both research questions and potential methods to address these questions. These case studies exemplify how research questions, such as those introduced in Chapter 2, can be integrated

---

### How to cite this book chapter:

Purves, Ross S., Olga Koblet, and Benjamin Adams (2022). "Showing the Potential of Computational Analysis to Support Environmental Narrative Research." In: *Unlocking Environmental Narratives: Towards Understanding Human Environment Interactions through Computational Text Analysis*. Ed. by Ross S. Purves, Olga Koblet, and Benjamin Adams. London: Ubiquity Press, pp. 87–92.  
DOI: <https://doi.org/10.5334/bcs.d>. License: CC-BY 4.0

with computational methods, as introduced in Chapter 3. The final case study was carried out by a single author, a Masters student participating in a course on geographical analysis of text.

All of the studies described in the subsequent chapters combine micro- and macro-readings of text, with a wide range of computational text analysis methods being used in concert with theory and insights from different disciplinary backgrounds. Methodologically, the resulting back and forth between computational methods and more detailed interpretation can be seen as an example of the hermeneutic circle, where the (re)interpretation of material can occur at different stages of the research. It can manifest itself anywhere, for example in identifying collections and queries to create corpora, in the generation of hypotheses, and most obviously in the analysis and interpretation of results to identify supporting evidence in answering the research questions. The goal of these exemplars is thus not to introduce a single recipe for the computational analysis of texts describing the environment. Rather, it is to show, by example, how the questions that we can ask are dependent on the domain of interest, availability and access to appropriate data, the match between data and computational tools, and not least the methodological and thematic expertise of the research team working on a particular problem.

## 4.1 Case Study Summaries

### *Glacial narratives: How can they be captured?*

Glaciers are prominent geographic features that have infused popular imagination in different ways over time. Where once they were seen as threats to livelihood, in contemporary times they are a symbol of what is being lost due to climate change while also tied to recreation and tourism. In our first contribution, Katrín Anna Lund, Ludovic Moncla and Gabriel Viehhauser look at how narrative writing about glaciers has changed over time through an investigation of texts about glaciers derived from three sources: articles from *Der Spiegel* newspaper, debates from the British parliament and an existing corpus focused on mountain-related texts, Text+Berg. After filtering the documents by matching different keyword forms of the word 'glacier' in English and German, they use a statistical test to examine how discourse has changed over time and find some evidence that supports the hypothesis, particularly with respect to language discussing the disappearance of glaciers. They also highlight some important limitations in the computational approach, for example, through the use of glacier as a metaphor.

Contributions: Lund did the literature review and wrote the first part of the chapter. Viehhauser and Moncla collected the data, analysed and wrote

about the experiments and added to the draft. In collaboration, Lund and Viehhauser wrote the conclusions.

***Greening a Post-Industrial City: Applying keyword extractor methods to monitor a fast-changing environmental narrative***

As we saw already with narratives about glaciers, the ways in which people write about environments reflects not only the physical features of places but also the social context from which the writing stems. In our second contribution, Sarah Luria and Ricardo Campos ask the question of whether an unsupervised keyword extraction tool called YAKE! can give new insight into the complex discourse that surrounds ‘revitalization’ efforts in post-industrial urban areas. In this study, text documents are collected to give a diversity of perspectives on a specific location, the Canal District of Worcester, Massachusetts, and then keywords are extracted. The automatically generated keywords for the texts are then compared against a close reading interpretation, which shows in what ways the method can help to support environmental history. The authors also highlight some lessons learned from working in an interdisciplinary way.

Contributions: Luria provided the initial question, framing and data set for the chapter. Campos pre-processed all the documents collected by Luria, from text files, images, word documents and pdfs (through Optical Character Recognition [OCR]). Campos processed all of the data through YAKE and produced the results as tables and word clouds. Campos elaborated a Python notebook to allow for the reproducibility of the results. Campos gave guidance on how to read YAKE’s results, which Luria then analysed and interpreted. Luria led the writing of the essay and its revision with Campos writing the sections pertaining to YAKE and its functions, and offering comments on the essay as a whole.

***Best Practice for Forensic Fishing: Combining text processing with an environmental history view of historic travel writing in Loch Lomond, Scotland***

In Chapter 5, Karen Jones, Diana Maynard and Flurina Wartmann, explore the use of text processing on a corpus of travel writing that is geographically and temporally focused on Loch Lomond in the 18th and 19th centuries in order to do historical research on the origin of the region as a touristic landscape. In contrast to the previous two chapters which focused on combining a single computational method with close reading, the approach used here is to consider a number of different computational text analysis methods that are built into the General Architecture for Text Engineering (GATE) toolkit. Tools for information extraction from unstructured text, such as GATE, are now

ubiquitous in a plethora of disciplines, domains and applications, and there are numerous freely available open-source possibilities, of which GATE<sup>1</sup>, Stanford CoreNLP<sup>2</sup> and NLTK<sup>3</sup> are among the most popular due to their ease of use and adaptability. The authors find that the computational methods provide insight into the use of terminology in historical environmental texts, yet close reading is required in order to provide background information that is ‘hidden’ in the text and inaccessible to methods that only analyse the surface text.

**Contributions:** The authors collaborated equally on the scoping of ideas for the chapter, the proposal for integration of different methodologies, and the mechanisms for practical implementation of the study. Maynard and Wartmann led the data assembly, analysis and interpretation of the data, comprising the main body of the chapter. Their findings were calibrated by Jones against a traditional historical/archival interpretation, driving the discussion of challenges in the conclusion, which was written collectively.

### *The wild process: Constructing multi-scalar environmental narratives*

An important class of environmental writing centers on the personal, experiential process of the author interacting with, traveling through, or remembering an environment. Because these narratives present the author’s individual experience, who the author is provides additional context and meaning for researchers to be able to analyse the texts. In this chapter, Joanna Taylor and Benjamin Adams analyse environmental texts through the lens of male and female genders. The goal is to test the hypothesis, put forward by Kathleen Jamie, that popular notions of ‘wildness’ in environmental writing represents a particularly male perspective. Two corpora are constructed: a small set of documents by male and female authors about Rannoch Moor, and a larger corpus of articles from the *Guardian* newspaper’s Country Diary column. They employ a multiscale approach that uses collocation and concordance analysis as well as supervised classification to give insight into general trends in the language in both corpora. Close reading of individual texts provides additional understanding to what the computational results mean allowing them to evaluate the results as evidence for the original hypothesis.

**Contributions:** Adams collected the data for the Guardian Country Diary and led the computational analysis of that collection. Taylor provided the theoretical approaches and close readings of the texts, and led on the chapter’s writing.

---

<sup>1</sup> <http://gate.ac.uk>

<sup>2</sup> <https://stanfordnlp.github.io/CoreNLP/>

<sup>3</sup> <https://www.nltk.org/>

### *Inferring Value: A Multiscalar Analysis of Landscape Character Assessments*

Another domain of environmental writing relates to public policy. Though these documents might not be written from a first-person perspective, social and political contexts still play a role in how they are written. Meladel Mistica, Joanna E. Taylor, Graham Fairclough and Tim Baldwin create a corpus of landscape character assessments – policy documents that are designed to capture the value of landscapes – and explore how topic modelling in combination with close reading can help to untangle the inherent assumptions about ‘value’ adopted in different character assessments. Because the documents that comprise the corpus come from a variety of sources and use different typesetting and design, the first part of the chapter explores a set of automated tools for extracting the texts into a form suitable for computational analysis.

Contributions: Taylor led the writing of this chapter, and also the analysis in the close reading and discussion sections. Mistica was responsible for implementing the computational elements of the chapter, including identifying and testing the PDF parsers, accessing the data and running the topic models. Baldwin led the design and evaluation of the computational analysis. Fairclough’s knowledge about landscape value and Landscape Character Assessments (LCAs) are evident throughout the chapter, with especial contributions to the introduction.

### *Interpreting natural spatial language in a fictional text: Analysing natural and urban landscapes in Mary Shelley’s Frankenstein*

Computational methods for distant reading are often touted for their ability to uncover patterns in large corpora, but they can also be useful for understanding the statistics of language use within single texts. In Chapter 5, Tobias Zuerrer uses concordance analysis on the text of Mary Shelley’s *Frankenstein* to test if there is a dichotomy between how seed terms denoting natural and urban geographic features are described in this Romantic era novel. In a fictional text, the perspective of a character or narrator might be different from the perspective of the author, which adds additional complexity to the analysis.

### *Discovering spatial referencing strategies in environmental narratives*

In the previous six case studies, existing computational methods are applied along with more traditional qualitative analyses to varying success to answer a number of research questions. The last case study by Simon Scheider, Ludovic Moncla and Gabriel Viehhauser puts the focus on the development of a new computational method for analysing spatial language in environmental narratives that aligns closer with how people visually conceptualise what

they read. In particular, they propose a method to explicitly capture spatial frames of reference that are described in natural language. In this case, the two mountaineering texts that are selected are annotated by human readers and then are used to evaluate the efficacy of the new method. The challenges posed in this chapter are indicative of some of the hurdles that researchers face when developing automated natural language processing tools that can approximate some of the more nuanced, close reading tasks performed by domain experts.

Contributions: Scheider contributed the introduction and the parts about spatial reference frames. Viehhauser designed the method for measuring inter-annotator agreement, analysed the results and contributed linguistic background theory. Moncla iteratively designed the parsing rules and ran parsers over all text sources. All authors annotated some texts (in addition to the non-author annotators) and wrote the results discussion together.