

# Castles, Battlefields, and Continents: A Dataset of Maps from Literature

Axel Bax<sup>1</sup> , David Mimno<sup>1</sup> , and Matthew Wilkens<sup>1</sup> 

<sup>1</sup> Information Science, Cornell University, Ithaca, New York

## Abstract

Maps are not common in novels. It is not obvious that they are necessary at all. Yet maps do appear in some novels. Why and to what ends? To answer these questions, scholars need a large collection of novels that contain maps. We develop a computational system to identify maps from page images and apply it to a large historical corpus of fiction. We deploy a three part workflow using an ensemble of three finetuned EfficientNet convolutional neural network (CNN) classifiers, Contrastive Language-Image Pre-training (CLIP), and human annotation to identify 2,622 maps in over 32 million pages of fiction published 1800–1928. We find that 1) maps are rare, making up 0.008% of all pages (1.7% of novels contain at least one map) 2) “map novels” were most common at the turn of the 20th century, 3) maps mostly appear on endpapers or front matter, 4) only 43% of map novels contain references to maps in their library MARC records, 5) 25% of maps depict fictional settings, 6) 70% of maps represent areas at a regional or larger scale, and 7) map novels contain more spatial language than non-map novels.

**Keywords:** computational humanities, literary cartography, digital humanities, CLIP, EfficientNet, maps

---

Axel Bax, David Mimno, and Matthew Wilkens. “Castles, Battlefields, and Continents: A Dataset of Maps from Literature.” In: *Computational Humanities Research 2025*, ed. by Taylor Arnold, Margherita Fantoli, and Ruben Ros. Vol. 3. Anthology of Computers and the Humanities. 2025, 280–280. <https://doi.org/10.63744/oYbvYsUA743D>.

© 2025 by the authors. Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0).