

ABSTRACT: The main objective of the paper is to verify the assumption of the urban-rural dichotomy, i.e., the attribution of differing values and emotions to urban and rural life, in literary works, as adopted in research, especially concerning the literature of the second half of the 19th century and the turn of the 20th century. Drawing on the studies by Heuser et al. (2016) and Herrmann et al. (2020), we will examine whether there is a difference in how sentiments and emotions are depicted in relation to cities and villages and show how these differences changed over time. The analysis will focus on the context of mentions of towns and villages in Polish literature from the mid-19th century until World War II. The paper will introduce the workflow and the preliminary results of our study.

The background to our research is the axiomatically simplified contrast between literary representations of rural and urban life as positive and negative, respectively. In Polish literary research, this tendency is secondary to the sequence of literary periods. Transition from the Realism (Polish: *Pozytywizm*) into the Young Poland period is related to the growing popularity of the construction of the virtue of the rural and the vice of the urban. On the contrary, the interwar period accentuates the rise of an increasingly urbanised modernity.

Those assumptions were based on the analysis of limited numbers of literary and critical texts from the era, often selected arbitrarily, and were rarely supported by systematic research. Empirical approaches are faced with a scarcity and fragmentation of the data at hand, therefore in our paper, we will test the thesis using a large corpus of texts and employing a distant reading perspective.

The corpus compiled for the study will comprise Polish-language literary prose texts published between 1863 and 1939. The Polish novel corpus for the ELTeC, the European Literary Text Collection, produced by the COST Action Distant Reading for European Literary History, will serve as a starting point. It includes 100 novels from 1840 to 1920, selected according to strict eligibility and balancing criteria. The ELTeC corpus will be supplemented with novels obtained from two digital libraries: the Polona digital library which offers OCR-derived texts and the Wolne Lektury library which offers electronic editions of books. We will only include books of Polish origin and aim to provide a balanced representation of the three literary periods.

We propose an experimental automatic procedure for detecting emotion in relation to representations of places that employs existing Natural Language Processing, i.e. Sentiment Analysis (SA) and Named Entity Recognition (NER), resources in combination with Linked Open Data (LOD). First, we will employ a deep learning-based NER tagger to extract mentions of geo-entities from the corpus. Since the tagger was trained on a corpus in modern Polish, we will apply a diachronic normalizer. Next, we will disambiguate (i.a. distinguish between fictional and real places) and classify the recognized entities as cities or villages, using LOD and dictionary-based approaches. To classify fictional places, we will compile a comprehensive controlled vocabulary of literary and imaginary places. Whereas for the classification of real places, we will reuse and harmonise historical data, i.e. official geostatistics, and census register. To examine the presence of emotions (valence and discrete emotions) and their strength (arousal) related to urban and rural entities, we will employ different SA tools for the Polish language, both dictionary- and machine learning-based. We will analyse and compare sentiment and emotions, either in text spans of +/- 50 words from the entities, or of paragraph size.