

Reconstructing and interpreting large datasets on historical translation flows

Dr. Ondřej VIMR

Abstract

In historical research into international literary flows, quantitative bibliographic data are often used for a synthetic analysis (Poupaud, Pym, and Torres Simón 2009; Hedberg 2012; McMartin 2020). This is despite a general lack of reliable bibliographies and databases on translations for many language combinations and historical epochs.

At the same time, the emerging Big Translation History calls for more data-driven research into translation history (Roig Sanz and Fóllica 2021; Vimr 2020), and the Bibliographic Data Science (Lahti *et al.* 2019) provides a framework for processing large bibliographic datasets for book history purposes. How to build an alternative to the more established and traditional bibliographies of translations in the age of Big Data and large-scale processing? What are the challenges and limits of obtaining and processing such data? How reliable are the results? How to use and interpret the new datasets from the perspective of Big Translation History?

This paper explores the challenges of reconstructing and interpreting a bibliographic dataset on the dissemination of Czech literature in 1850-2020 from a large-scale data perspective. Starting with over 2 000 000 bibliographic records on books licensed from the OCLC (WorldCat.org), it examines data processing and cross-enrichment, scrutinises the quality and reliability of the resulting dataset, and finally uses the data to examine the global trajectories of authorships and their clusters throughout almost two centuries. The results may serve as an inspiration for any historian of translation who suffers from a lack of bibliographic data especially in case of smaller national literatures.

Biography

Dr. Ondřej VIMR is a researcher at the Czech Academy of Sciences (Institute of Czech Literature), the scientific secretary of the Czech Literary Bibliography Research Infrastructure and the head of its Global and Digital Literary Studies Lab. His research interests range widely within the history and sociology of translation and publishing, especially focusing on the circulation of “less translated” European literatures. Currently, he is the principal investigator of a Czech Science Foundation project exploring the global trajectories of the dissemination of Czech literature since 1945 from both large-scale and small-scale perspectives.