TNA ACTIVITY REPORT

Monolingualism Deconstructed: Modelling Hidden and Invisible Multilingualism in German literature (1790-1890

Author: Jana-Katharina Mende

Current position: postdoc

Affiliation: Martin-Luther-University Halle-Wittenberg

Host institution: Austrian Centre for Digital Humanities and Cultural Heritage, Vienna

Mentor(s): Matej Ďurčo, Vera Maria Charvat

Period of stay: 2024-03-04 until 2024-03-29 and from 2024-06-10 until 2024-06-14

Activity Report: Jana Mende: Monolingualism Deconstructed: Modelling Hidden and Invisible Multilingualism in German literature (1790-1890)

Introduction and research questions

Multilingualism has long been a fundamental aspect of global society and is deeply rooted in history and literature. This project focuses on exploring the often-overlooked history of 19th-century multilingual literature in German-speaking regions. It aims to develop a scalable method to measure and examine the characteristics of this historical multilingual literature, addressing the significant gap where much of the literature remains unread and non-canonical. Despite the increasing focus on 19th-century literature within digital literary scholarship, there is a lack of multilingual approaches and tools, which are predominantly monolingual. The project's methodology involves investigating historical biographical information for references to multilingualism and combining biographical and bibliographical data with geographical coordinates to uncover the multilingual lives and works of lesser-known authors.
This study uses historical biographical data on multilingual competences and everyday use of different languages. Biographical data, including information on languages, multilingualism, and places of residence, can be found in historical biographical dictionaries which are useful sources for investigating multilingual authors even if their main aim is to provide information on German literature. For this

CLS INFRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004984

project, I use two historical dictionaries, Franz Brümmer's (1836–1923*) Lexikon der deutschen Dichter und Prosaisten vom Beginn des 19. Jahrhunderts bis zur Gegenwart* and Sophie Pataky's (1860–1915) *Lexikon der Frauen deutscher Feder*. The advantages of Brümmer's and Pataky's dictionaries compared to others from that time are their focus on contemporary writers in the 19th century and the structure of the articles. The alphabetical, not chronological, order of the dictionaries makes it easier to extract and re-structure biographical, geographical, and bibliographical meta-data semi-automatically.
The research questions for my visit to the ACDH-CH focused on creating an optimal workflow for extracting data from historical dictionaries. Specifically, I aimed to develop a smooth process from text to data to metadata, model the data effectively, and connect this data to other sources, including open linked data and bibliographical information.

Results of the research visit

The questions mentioned above were discussed in several research meetings during my stay in Vienna, where I received invaluable practical feedback and assistance to structure my research. Together with my mentors, Matej Ďurčo and Vera Maria Charvat, and other members of the ACDH-CH, especially Matthias Schlögl, Nina Rastinger, and Martin Anton Müller, we defined work packages and set up a timeline to address these questions during my stay.

Technical Milestones:
As our first task, we created a workflow to extract lists of names from the dictionary and match them with identifiers (GND IDs) to connect the authors' names with existing authority file identifiers wherever possible. This formed the basis for enriching the data using OpenRefine.
The second task involved creating a data model to structure the data acquired during the process. Matej Ďurčo taught me the basics of data modeling and how an RDF data model is structured, using the CLSCor data model as an example. During our meetings, we developed a data model for my research that connects authors, places, works, and languages, allowing me to query the data for connections between language use and literary works of multilingual authors. This data model was implemented as a relational database. Vera Charvat taught me the fundamentals of using relational databases. During my stay, I was introduced to several technical tools, including BaseRow, NodeGoat, and simple Excel tables, to create and structure the data.[1]

Gaining New Knowledge and Skills:

Through mentoring at the ACDH-CH, I gained working knowledge in relational databases (with BaseRow, OpenRefine, NodeGoat, and Excel), an introduction to RDF data models, and general knowledge on structuring and enriching data using linked open data.

[1] Detailed notes of the meetings can be found here: https://docs.google.com/document/d/1Xl3yXoCV_cXBOkhXUgSOMbH2i35D5VXsX1HexvzS0rM/edit?usp=sharing

At the CLS Infra Training School, I learned about programmable corpora, using SPARQL to query CLSCor and Wikidata, and combining these resources. These workshops provided me with essential skills for my research and sparked new research questions, such as the representation of multilingual literature on Wikipedia and the potential to connect my data with Wikidata.

Collaborating and Connecting:

During my research stay, I had the opportunity to collaborate with several researchers in the department for text and literary studies at the ACDH-CH. We discovered that we shared similar research interests, particularly in the areas of data extraction from historical texts, data modeling, and the visualization of geographical data. We dealt with comparable questions and challenges, such as optimizing workflows and ensuring data interoperability. This common ground has laid the foundation for ongoing cooperation.I was introduced to the librarians at the Austrian National Library through the ÖNB Lab, which provided additional research material and also led to an ongoing cooperation to use bibliographical data from the ÖNB. The CLSInfra Training School enabled me to meet other TNA fellows and connect with senior researchers in CLS during the conference that followed the training school, bringing new impulses and directions to my research.

Next steps

The results of the research visit for my project are truly fundamental for my research project. With the invaluable help of my mentors, I developed a robust workflow to structure and enrich data—transforming a list of names from historical sources (two historical dictionaries) into a structured database of multilingual authors. This includes creating a comprehensive data model that enables us to address multiple unique questions in literary history.
Currently, I am continuing to refine the data. I plan to publish a subset—focusing on multilingual authors from Ukraine and the Baltic countries—to evaluate the effectiveness of both the workflow and the data model.