From Corpus to Chronology: Scalable Readings of Narrative Interviews

Often, numerous interesting topics lie dormant in sources that were originally collected or produced for an entirely different research interest – as in the case of Oral History: the multifaceted content of narrative interviews generally extends far beyond the focus of the original inquiry, thus making them valuable research data for secondary analyses. Within the DFG-funded project *Oral-History.Digital*, there is a multitude of such interviews from various collections that have been digitized and are accessible online. However, the search for multifaceted and often implicit themes such as *migration* and *home* is very complex. The larger the corpus under investigation, the more difficult it becomes to manually access the interviews solely through metadata filters, keyword searches, and transcript viewing. Natural Language Processing techniques, such as Topic Modeling, enable a distant reading to obtain initial information about the sources and the themes they contain, and to use this information for accessing the interviews. This reveals patterns of commonalities and differences in the collections that are barely perceptible to the naked eye.

With the chances of platforms like *Oral-History.Digital* also come risks. For example, the full-text search enables to navigate precisely to interview segments containing the search term. This may speed up the research but can contribute to a decontextualization and fragmentation of the complex narrative structure of biographical interviews. In order to address this problem, a chronology visualization based on the computed topics provides an overview of the themes discussed in the interview, their sequential relation (from a horizontal perspective) and their correlations (from a vertical perspective). This allows to contextualize found segments instantly and reconstruct narratives by following their traces visually.

An interesting subject of investigation is the handling of one's own identity and the associated understanding of home. How is contact maintained with the abandoned homeland and how do the brought along and encountered ways of life influence each other? To what extent does the understanding and attachment to home change over time, especially when people decide to stay permanently? The presentation illustrates the path from the quantitative exploration of an interview corpus to the qualitative examination of the interviews by using Topic Modeling and visualization in a scalable reading.