

Standardized multilingualism in the digital humanities? Opportunities and risks of AI-supported automation

German-Czech DH-Conference, 20–21 February 2025, Prague

The availability of AI-driven translation technology is changing the way the humanities work on a practical level. Text sources in different languages, whether scientific literature or primary sources, have become widely available not only by digitization but also by automated translation.

For historians, AI-supported translation technologies revive familiar challenges on a much broader scale with entirely new implications for scientific practice. Historical sciences are in two ways confronted by this:

1. Translation technologies raise historically known topics of cultural transfers that includes not only standardization and domain dependency within one language but also more general questions of the (un)translatability of historical context across language borders. After all, the success and the promise of language processing based on Machine Learning techniques relies on the fact that it can disambiguate data in practice more successful than manually annotated datasets and create a text with hermeneutical meaning.
2. The much broader scale of potentially available data furthermore raises a very immediate issue of how to deal with and how to analyze mass-translated source texts in practice. This is at the same time a great chance for new avenues of research.

For linguistics on the other hand, the question of semantic ambiguities is probably less pressing. Still the question remains, to what degree linguistic corpora are representative of historically embedded meaning and to what degree this transfer works in the practice of other academic disciplines: Taken together, the processing and translation of text sources includes a twofold transfer: a formalized classification of text sources (such as tokenization or entity recognition) and the reconversion of this formalized information into a hermeneutically comprehensible whole.

This raises the question, to what degree the meaningful use of historical source texts still depends on manual document classification and to what degree on automated processes. How can applied linguistics and historical sciences work together on these tasks?

At the German-Czech DH meeting, we want to discuss the interdisciplinary practices. We will place a particular focus on German-Czech history and Bohemistics/Czech studies since the 19th century. This specific domain has a rich and significant history of interconnections in literature, language and politics.

In addition, multilingual historical text data exists that can be used for translation history, reception history, and as a practical example for dealing with semantically enriched data.

We want to discuss the following topics:

1. What is the historical and present-day relevance of national languages and language borders in the practice of the humanities?
2. What can we learn from the historical-political conditions for the standardization of languages, especially in science?
3. What do AI-supported translation technologies achieve in the practice of science and how can the historical sciences in particular benefit from them? What are the limitations and risks involved?
4. What is the state of the enrichment of language data across language boundaries? Do language boundaries play any role at all in automated semantic processes?
5. What is the history of the reception of translations, especially in the German-Czech or Central European context?

The Conference language will be English.

Contributions to these sections can be submitted in the form of an abstract. The abstract should not exceed 1000 words and should be sent by December 1st, 2024 by email to digitalisierung@badw.de.