map(NLP, LTZ)
Mapping Luxembourgish in the digital domain

Christoph Purschke

Université du Luxembourg

Luxembourgish, as a small language with official status, presents a rewarding object for research in many ways. On the one hand, the language is currently undergoing processes of structural *Ausbau* in which questions of linguistic standardization are intermingled with language policy efforts. On the other hand, the emergence of a new written language can be monitored quasi in real time, largely promoted by digital and social media platforms.

The steadily growing text collections allow interesting conclusions to be drawn about individual variation patterns and the formation of shared writing practices in written Luxembourgish. At the same time, however, there is a lack of language technological tools and pipelines that would make it possible to study these processes at scale. The sheer amount of variation in spelling presents an additional challenge in this regard.

Against this backdrop, the talk will provide insights into current (and future) projects that work at the intersection of NLP for small languages and sociolinguistics. This includes the development of state of the art models and tools for the automatic processing of text corpora. The overall goal, however, is to use computational methods to find answers to sociolinguistic questions, e.g., regarding the ongoing development of Luxembourgish into a standard language.