

## **Decoding Historical Dialects: Employing Large Language Models in Historical Sociolinguistic Research**

Christa Schneider (Bern)

Historical sociolinguistics often faces the challenge of uncovering dialectal elements buried within large, unannotated corpora. As the evolution of dialects over time offers unique insights into language change, regional (linguistic) identity, and societal shifts, this is an issue worth further exploration. In this talk, I will introduce innovative approaches that employ Large Language Models (LLMs) to extract and analyse dialectal features from historical texts. My case study focuses on a corpus of early modern witch trial papers from Bern (Switzerland), written in an official administrative language interspersed with dialectal elements. As this study is a work in progress, I welcome any ideas, comments and concerns.

To address this challenge, I will explore different computational strategies for identifying dialectal elements in historical texts. These approaches could include e.g. methods that rely on dialectal features previously extracted from the corpus, machine learning techniques that identify patterns without the need for annotated data, or data-driven strategies designed to face the challenges posed by sparse data.

These approaches will help track dialect evolution over time while integrating new findings into historical sociolinguistic theories. Ultimately, this research demonstrates how cutting-edge computational tools can open new perspectives for historical sociolinguistics, offering deeper insights into dialect evolution over time.