



The Center for Spatial and Textual Analysis Seminar Series presents:

**ROSETTA: ResOurces for Endangered languages Through TranslAted texts**

**Ronald Jenn, Amel Fraisse, Zheng Zhang, Shelley Fisher Fishkin**

**Tuesday, April 9th, 2019 | 12:00PM - 1:00PM | Wallenberg Hall, 4th Floor**



Out of the world's 6000+ languages only a small fraction currently enjoys the benefits of modern language technologies. Languages left behind are called endangered or technologically low-resourced (even though they may have millions of speakers). This collaborative and interdisciplinary digital humanities research project aims to help salvage those languages by combining computational linguistics, American Literature, and Translation Studies. Much as the Rosetta Stone helped decipher the demotic and hieroglyphic scripts thanks to the presence of the Greek translation, our project intends to preserve contemporary endangered languages and assist with their survival through translation. Our project puts to use the extant translated versions of a single fictional text—Mark Twain's *Adventures of Huckleberry Finn*—into a number of low-resourced languages spanning a period of nearly a century and a half. The project relies on the involvement of humans for data collection while natural language processing tools generate language resources (corpora, dictionaries, thesauri, lexicons) for those endangered languages.

**Ronald Jenn** is a Professor at Université de Lille (France) where his research focuses on Translation and ideology, Digital Humanities, and Mark Twain. A recipient of a grant from the MESHS in Lille (2017-18) for the "Global Huck Project," he is now involved in Rosetta which looks at global translations of *Adventures of Huckleberry Finn* using digital tools. **Amel Fraisse** is an Associate Professor at Université de Lille (France) working on Digital Humanities and Natural Language Processing. Her research focuses on: Information Extraction, Knowledge Acquisition and Visualization from Texts, Multilingual information, and under-resourced languages. **Zheng Zhang** is a third year PhD student at Université Paris-Saclay (France). His research interests are graph-based word embeddings learning and parallel sentence extraction from comparable corpora. **Shelley Fisher Fishkin** is the Joseph S. Atha Professor of Humanities, Professor of English, and Director of American Studies at Stanford. She is Founding Editor of the *Journal of Transnational American Studies*, winner of the John Tuckey Award for Lifetime Achievement in Mark Twain Studies, and Co-Director of the Chinese Railroad Workers in North American Project at Stanford.

This seminar is made possible by a grant from the France-Stanford Center.

Visit [cesta.stanford.edu/events](http://cesta.stanford.edu/events) for more information.

Lunch will be served.