ROSETTA

ResOurseS for Endangered Languages Through TranslAted texts

Out of the world’s 6000+ languages only a small fraction currently enjoys the benefits of modern information technologies. Languages left behind are called technologically low-resourced and endangered (even though they may have millions of speakers). This collaborative and interdisciplinary digital humanities research project aims to help create digital resources in those languages by combining Computational Linguistics, Library and Information Science, 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 help preserve contemporary endangered languages and assist with their survival through translation. Our project puts to use 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 will generate digital language resources for these languages (dictionaries, thesauri, etc.). In the process, scholars of American literature and Translation Studies will also have the opportunity to gain insight into the global circulation of a canonical American novel.