uOttawa

Translation & Minority 2
Freedom and Difference

Traduction & Minorité 2
Liberté et différence

Overcoming Minor-ness through the Digital Revolution:

Developing a Minority Language Corpus and Translation Tools

Mansur Saykhunov

Kazan Federal University (Russia)

Gulnara Shaydullina

University of Québec in Montreal (Canada)

Our presentation explores how new technologies and digital tools contribute to a greater representation of minority languages taking the Tatar language as an example.

Firstly, we will briefly present the Tatar people and their mother tongue (a Kipchak Turkic language, subfamily of the Altaic languages). Then we will offer an overview of the current state of the Tatar online corpora and the way it is evolving. We will present the Corpus of Written Tatar project: based entirely on a volunteer contribution, it is the result of creativity, enthusiasm, and professionalism of IT and Humanities experts from Kazan, the capital of the Tatar Republic in Russia. Their team built, launched, and has been successfully maintaining the first online Tatar corpora outperforming similar state-funded projects that appeared later. We will examine some of its functions and the impact it has on the representation of the Tatar language online.

Finally, we will present an overview of Apertium, an international project, in which the team behind the Corpus of Written Tatar is also participating. Apertium is a free/open-source rule-based machine translation platform whose goal is to promote digital representation of minor languages through translation. Built and maintained through participatory work, this platform is a machine translation engine that enables language speakers and learners to translate from major languages into smaller ones and vice versa.

Not only do these projects facilitate language acquisition of the Tatar language both as a first and as a second mother tongue, but they also provide valuable applications, free and efficient, for writers, revisers, and translators. Just like any other Digital Humanities tools, these projects call for a joint team efforts and expertise.

Key-words: minority languages; representation of minority languages; minority languages corpus; online collaboration; Digital Humanities

Bios: Mansur Saykhunov holds a PhD in Tatar Philology and is one of the authors and the main developer of the Corpus of Written Tatar. He also participated in the development of the Tatar speech synthesizers. His current research interests are corpus linguistics, speech synthesis, Tatar phonetics. His main projects are: Text Corpus of the modern Tatar language; the Sketch Engine Project : Tatar News Corpus; Leipzig Corpora Collection : Tatar Mixed Corpus; APERTIUM: a free/open-source machine translation platform (contribution in testing and extending the dictionary of morphological analyzer for Tatar language)

Gulnara Shaydullina is a PhD student in Cognitive Informatics (University of Québec in Montreal). Born and raised in Russia, she graduated from the faculty of the Romance-Germanic Philology at the Bashkir State University (BGU, in Ufa, Russia) and worked as an ESL teacher, a translator, and an interpreter. Upon arrival in Canada, she received a Graduate Diploma in Translation (McGill University) followed by MA in Translation (University of Montreal). Her current research interests are crowdsourcing in translation, technology and translation, digital pedagogy, and teaching minority languages.