**Semantic Domains in Akkadian Texts**

Tero Alstola, University of Helsinki

Akkadian is a Semitic language that was spoken and written in ancient Mesopotamia from the third millennium BCE until the beginning of the Common Era. Written on durable clay tablets, hundreds of thousands of Akkadian texts have been unearthed from the soil of present-day Iraq and Syria since the 1850s. During the last twenty years, the corpus of Akkadian texts has become increasingly available in an electronic format, which offers exciting opportunities for research in digital humanities. For example, the semantic fields of Akkadian lexemes have traditionally been studied qualitatively, and the meaning of a lexeme has been defined by scholars on the basis of textual attestations they have gathered. Quantitative methods could ease this task and provide additional information on the diachronic developments of the language and dialectical differences in its vocabulary.

The 'Semantic Domains in Akkadian Texts' project (funded by the Academy of Finland in 2016–2020) aims at enhancing the understanding of the Akkadian language by generating contextual semantic domains for Akkadian lexemes. Quantitative methods from language technology, especially word sense induction, are used to analyse an electronic corpus of Akkadian texts. The text corpus used in the project is primarily derived from Oracc (The Open Richly Annotated Cuneiform Corpus), the number of available texts being roughly ten thousand. The project also aims at developing language technological methods and software tools which can be employed to other extinct languages or small and fragmented text corpora. The methods and software tools developed in the project will be made available online.

The project has started its work by analysing the text corpus and testing several language technological methods on the data. An early outcome of this work is a forthcoming article on the semantic fields of the Akkadian lexemes 'horse', 'to speak', and 'power'. These lexemes were analysed by using two existing language technological methods, Pointwise mutual information (PMI) and Word2vec. The results of this study show that these language technological methods can improve our understanding of an ancient language and its concepts when combined with careful philological work on the primary sources.