DEEP LEARNING AND LANGUAGE ADAPTATION

Sergey Sharoff (S.Sharoff@leeds.ac.uk)

University of Leeds, UK

Many lesser-resourced languages are related to languages, which have better resources. For example, the Universal Dependency treebank has about 2 MW of training resources for Czech, more than 1 MW for Russian, while only 950 words for Ukrainian and nothing for Belorussian, Bosnian or Macedonian. Similarly, the Autodesk Machine Translation dataset only covers three Slavonic languages (Czech, Polish and Russian). In this talk I will discuss a general approach, which can be called Language Adaptation, similarly to Domain Adaptation. In this approach language models can be adapted from a better-resourced (donor) language to a lesser-resourced (recipient) language. In my talk I will discuss examples of a Deep Learning architecture for Language Adaptation, which is based on creating a shared representation across related languages. Three case studies will be presented: Part-Of-Speech tagging, Named Entity Recognition and Translation Quality Estimation. I will also discuss the importance of the typological distance between the donor and the recipient.