Automatic acquisition of subcategorisation from large text corpora

Adam Przepiórkowski
Institute of Computer Science, Polish Academy of Sciences

Valence acquisition is a task consisting in the automatic extraction (learning) of subcategorisation – or argument structure – from morphosyntactically annotated corpora. In this talk I will present the general methodology of this task and I will report on the work carried out at the Institute of Computer Science of the Polish Academy of Sciences (http://nlp.ipipan.waw.pl/) aiming at the acquisition of a syntactico-semantic valence dictionary of Polish. A part of the talk will be devoted to the problem of how to evaluate the quality of the resulting valence dictionary, given the multiplicity of approaches to subcategorisation in general and the vagueness of the argument-adjunct distinction in particular.