

The presented work investigates methods for semi-automatic extraction of lexico-syntactic information from corpora, particularly the information on subcategorization and valency frames. We document that at present time, PDT and CNC corpora are not sufficient for this task. We describe a simple method for a selective extension of corpora based on texts from Internet. We evaluate three parsers available for Czech with respect to the task of extracting verb frames. We have implemented a linguistically motivated filtration of input sentences to identify "very simple sentences", which helps the parsers to achieve better accuracy. The system AX designed in this work is more generic, any kind of linguistic filtration can be employed. The system is also suitable for creating partial or full parsers of natural languages. The thesis also presents a user's guide to the system AX. Furthermore, we compare methods for extraction of subcategorization frames from observed frames. We classify observed frames into a hierarchy suitable for human annotators. Finally, several problems of automatic extraction of valency frames are discussed.