A computational HPSG grammar of Polish NPs and PPs

Adam Przepiórkowski (Polish Academy of Sciences)

The aim of this paper is to present an implemented grammar based on one of the generative theories of language, Head-driven Phrase Structure Grammar (Pollard and Sag 1994). The coverage of the grammar is rather narrow: it includes nominal phrases (NPs) and prepositional phrases (PPs). The grammar is implemented as a parser which, given a naturally occurring Polish text, aims at finding all NPs and PPs in this text.

The grammar is implemented as part of a larger project (project number 3 T11C 003 28, financed by the State Committee for Scientific Research, carried out at the Institute of Computer Science, Polish Academy of Sciences) whose aim is to develop algorithms for automatic valence acquisition, i.e., methods of learning about possible arguments of verbs (and other predicates) automatically, on the basis of a corpus of texts. The two main stages of such methods are: 1) the identification of basic phrase types in texts, i.e., of potential arguments, by means of a computational grammar like the one presented here, and 2) collecting the information about the observed co-occurrences of certain verbs with certain sets of arguments and the application of inferential statistics to infer which of these observed subcategorisation frames are actual frames of given verbs, and which result from 'noise' in the data (i.e., reflect errors in the computational grammar, include adjuncts, miss arguments because of ellipsis, etc.). The results of this project will be published as a valence dictionary of Polish.

The HPSG grammar presented here is a shallow grammar: it does not aim at constructing the full structure of a sentence, but rather at finding all NPs and PPs occurring in a text. Moreover, in case of complex NPs containing relative clauses and further recursively embedded NPs, the grammar does not necessarily find out the full structure of such NPs; in such cases it only identifies the basic subconstituents of such NPs.

The grammar is implemented using the unification-based system developed at the Univerisity of Toronto, University of Tuebingen and Ohio State University, TRALE (Penn et al. 2003).

References:

Gerald Penn, Detmar Meurers, Kordula De Kuthy, Mohammad Haji-Abdolhosseini, Venessa Metcalf, Stefan Mueller, Holger Wunsch. (2003). "TRALE Milca Environment v. 2.5.0.".

Carl Pollard and Ivan A. Sag. (1994). "Head-driven Phrase Structure Grammar". Chicago University Press / CSLI, Chicago, IL.