

## **Towards the automatic construction and evaluation of valence dictionaries**

Adam Przepiórkowski (Institute of Computer Science, Polish Academy of Sciences)

The aim of this paper is to present the results of the project „Automatic extraction of linguistic knowledge from a large corpus of Polish" (a Ministry of Education and Science grant number 3T11C00328) carried out between 9 March 2005 and 8 September 2007 at the Institute of Computer Science, Polish Academy of Sciences, Warsaw.

In the talk we will sketch the methodology of valence (subcategorisation) acquisition from a morphosyntactically (not syntactically or semantically!) annotated corpus. This task may be conceptually divided into two stages. The first – linguistic – stage consists in the identification of various types of phrases in texts; such phrases are potential arguments of predicates. In the current project a shallow grammar of Polish – to be sketched in the full paper – has been developed for this purpose. In the second – statistical – stage, various statistical inference techniques are applied to deduce which of these observed potential argument structures are almost certainly true argument structures. The second stage is crucial as the results of the linguistic processing are usually very noisy, due to errors in morphosyntactic analysis, errors in shallow parsing, lack of overt distinction between argument and adjuncts, and due to ellipsis.

The main thrust of the paper will, however, concern the issues of evaluation: given an automatically constructed valence dictionary, how do we know how exhaustive it is? The merits and drawbacks of various evaluation methods will be discussed in some depth, including:

1. comparison to existing, manually constructed valence dictionaries,
2. evaluation on the basis of manually syntactically annotated texts,
3. evaluation of the ultimate usefulness in syntactic parsers using the valence dictionary.

We will show that none of these methods is ideal and that a sound evaluation methodology for valence dictionaries is still an open issue. In particular, as already demonstrated in earlier work (Przepiórkowski and Fast 2005), two existing manually constructed valence dictionaries of Polish may be as different from each other as they are from an automatically constructed dictionary, so in the case of the first evaluation method above it is not clear which manually constructed valence dictionary should serve as the gold standard. On the other hand, while the last method mentioned above may measure the practical usefulness of a valence dictionary, it makes the evaluation results depend on a particular processing tool and, hence, rather ephemeral.

The evaluation method that we will adopt here is the second method mentioned above and it consists in finding out how well a given valence dictionary models naturally occurring sentences. While marking phrases which are potential verb's arguments in corpora is still a subjective task, this annotator-dependence may be alleviated by requiring that each sentence be annotated by a number of linguists and providing guidelines which maximise inter-annotator agreement.

### **References**

Adam Przepiórkowski and Jakub Fast. (2005). [Baseline Experiments in the Extraction of Polish Valence Frames](#). In the proceedings of Intelligent Information Systems 2005 (New Trends in Intelligent Information Processing and Web Mining), pp. 511-520.