

# Augmenting Valence Dictionary of Polish Verbs with Semantic Information

Elżbieta Hajnicz

Institute of Computer Science PAS

Warsaw, 19.04.2006

# Resources

- ▶ The IPI PAN Corpus  
of Polish morphosyntactically annotated texts;
- ▶ The Valence Dictionary of Polish verbs (in preparation);
- ▶ The Polish WordNet (in preparation);
- ▶ Corresponding tools available with the above resources.

# Schedule of tasks

1. Creation and/or adaptation of indispensable linguistic resources;
2. Augmenting the valence dictionary with semantic categories of its arguments;
3. Analysis of argument structure of Polish verbs wrt. diathesis alternation;
4. Determining semantic roles for particular verbs;
5. Elaboration of methods for aggregating verbs into semantic hierarchy;
6. Preparation of a prototype of a FrameNet- or VerbNet-like semantic dictionary of Polish verbs.

# Creation and/or adaptation of linguistic resources

- ▶ Elaboration of formats of input and output data (XML, SQL);
- ▶ Creation (manual or semi-automatic) of small data sets enabling testing programming tools being built and subsequent verification of obtained results;
- ▶ Transformation of sentences annotated syntactically (using tools created in the valence dictionary project) into their simplified forms in which every argument filler is represented by a main element of a source argument filler.

## Examples

[Zdesperowany Jan]<sub>NP:nom</sub> zabrał [swojego ciężko chorego psa]<sub>NP:acc</sub> do [najlepszego weterynarza w mieście]<sub>NP:gen</sub>



[Jan]<sub>N:nom</sub> zabrał [psa]<sub>N:acc</sub> do [weterynarza]<sub>N:gen</sub>

(Driven to despair Jan took his seriously ill dog to the best vet in his town.)

[Robert Sosnowski, burmistrz Pragi Północ i działacz władz krajowych PiS]<sub>NP:nom</sub> załatwił [swojej żonie, radnej i szefowej PiS na Targówku]<sub>NP:dat</sub>, [mieszkanie z zasobów gminnych]<sub>NP:acc</sub>



[Burmistrz]<sub>N:nom</sub> załatwił [żonie]<sub>N:dat</sub> [mieszkanie]<sub>N:acc</sub>

(Robert Sosnowski, the mayor of Northern Praga and [...], granted his wife, a city councillor of Targówek and [...], an apartment from communal resources.)

## Augmenting the valence dictionary with semantic categories of its arguments

- ▶ Classification of nouns into coarse categories such as:  
*personal, animate, institutional, concrete, mass, immovable* etc.;
- ▶ Classification of nouns accordingly to WordNet synsets;
- ▶ Determining classes of WordNet co-hyponyms among all fillers of the particular argument slot for each verb.

# Analysis of argument structure of Polish verbs wrt. diathesis alternation

Diathesis alternations are regular variations in the syntactic expressions of verbal arguments.

- ▶ Detecting alternations for particular verbs with argument slots filled in a way described above;
- ▶ Grouping all alternations of a particular verb into a structure containing mappings between syntactic positions of each argument in corresponding alternations.

# Examples



- ▶ *Jan zaniósł/przyniósł psa do weterynarza.*  
(Jan took (his) dog to a vet.)
- ▶ *Jan \*zaniósł/przyniósł psa od weterynarza.*  
(Jan took (back) (his) dog from a vet.)
- ▶ *Jan zaniósł/przyniósł psa weterynarzowi. (= do)*



- ▶ *Jan zabrał/\*odebrał psa do weterynarza.*
- ▶ *Jan zabrał/odebrał psa od weterynarza.*
- ▶ *Jan zabrał/odebrał psa weterynarzowi. (≈ od)*



# Determining semantic roles for particular verbs

- ▶ Applying machine learning methods using manually prepared training subset;
- ▶ Manual labeling of every set of verbs defined as below.

# Elaboration of methods for aggregating verbs into semantic hierarchy

- ▶ Grouping together verbs with the same syntactic and semantic characteristics;
- ▶ Applying statistical clustering methods on data from various levels of organization;
- ▶ Grouping together verbs accordingly to their WordNet hypo/hyperonym hierarchy;
- ▶ Hybrid methods.