

# Syntaktyczno-semantyczny słownik polskich predykatów

## Stan obecny i perspektywy rozwoju



Elżbieta Hajnicz, Agnieszka Patejuk, Adam Przepiórkowski, Marcin Woliński  
Instytut Podstaw Informatyki PAN

### WSTĘP

#### Walenty:

- ▶ **słownik walencyjny** dla ludzi i komputerów,
- ▶ nie tylko **czasowniki**, lecz także **rzeczowniki, przymiotniki, przysłówki**,
- ▶ opraco(wy)wany w ramach kolejnych projektów (obecnie **CLARIN-PL**),
- ▶ **publicznie dostępny** na licencji swobodnej (CC SA),
- ▶ głównie **składniowy** (rozpoczęte prace nad **semantyką**),
- ▶ bogaty komponent **frazeologiczny**,
- ▶ schematy walencyjne **ilustrowane przykładami** (głównie z NKJP),
- ▶ **szczegółowy opis** formatu słownika w pracach cytowanych poniżej.

#### Istotne cechy formalizmu:

- ▶ pozycje składniowe określone na podstawie **testu koordynacji**,
- ▶ uwzględnienie **podmiotów nienominalnych**,
- ▶ uwzględnienie **niestabilności przypadku**,
- ▶ dopełnienie zdefiniowane na podstawie testu **strony biernej**,
- ▶ semantyczne zróżnicowanie typów **fraz tzw. przysłówkowych**,
- ▶ oddanie relacji **kontroli i podnoszenia**,
- ▶ **zaawansowany opis frazeologiczny**.

### PRZYKŁADY

#### Prosty – ADRESOWAĆ:

- ▶ *Dąbrowski adresował listy do Marii.*
- ▶ adresować: \_: imperf:  
subj{np(str)} + obj{np(str)} + {prepn(do,gen)}
- ▶ niedokonany **aspekt** (imperf), dowolna wartość **zanegowania** (-),
- ▶ trzy argumenty:
  - ▶ **podmiot** (subj) rzeczownikowy (np) w **przypadku strukturalnym** (str; zwykle mianownik, ale np. dopełniacz przy gerundiach),
  - ▶ **dopełnienie** (obj) rzeczownikowe w przypadku strukturalnym (zwykle biernik, ale np. dopełniacz negacji),
  - ▶ **argument przyimkowo-rzeczownikowy** (prepn) z przyimkiem DO i frazą rzeczownikową w dopełniaczu (gen).

#### Z różnorodną realizacją argumentu – BAĆ SIĘ:

- ▶ *Boisz się bezrobocia i że zabraknie Ci środków na utrzymanie?*
- ▶ bać się: \_: imperf:  
subj{np(str)} + {np(gen); cp(że)}
- ▶ SIĘ jako część lematu (inaczej w wypadku MYĆ itp.),
- ▶ dwa argumenty:
  - ▶ **podmiot** w przypadku strukturalnym,
  - ▶ niepodmiotowa **pozycja składniowa** wypełniona przez:
    - ▶ **frazę rzeczownikową w dopełniaczu** (np(gen)) lub
    - ▶ **frazę zdaniową** (cp) wprowadzoną przez spójnik podrzędny typu ŻE lub
    - ▶ **ich koordynację**.

### PRZYKŁADY (cd.)

#### Frazeologiczny – PRZYJĄĆ Z OTWARTYMI RAMIONAMI:

- ▶ *Jej przyjaciele przyjęli ją z otwartymi ramionami.*
- ▶ przyjąć: \_: perf:  
subj{np(str)} + obj{np(str)} +  
{lex(prepn(z,inst),pl,'ramię',  
ratr({lex(adjp(agr),agr,agr,'otwarty',  
atr({lex(advp(misc),'szeroko',natr)}))}))}
- ▶ trzy argumenty, w tym **frazeologiczny** (lex):
  - ▶ **fraza przyimkowo-rzeczownikowa** z przyimkiem Z, frazą rzeczownikową w narzędniku (inst), w liczbie mnogiej (pl) i z **elementem głównym RAMIĘ**,
  - ▶ **obowiązkowo modyfikowana** (ratr) przez uzgadniającą się (agr) **frazę przymiotnikową** (adjp) o elemencie głównym OTWARTY,
  - ▶ który z kolei jest **fakultatywnie modyfikowany** (atr) przez **frazę przysłówkową** (advp) o elemencie głównym SZEROKO,
  - ▶ który z kolei **nie może być modyfikowany** (natr).

### WIELKOŚĆ, DOSTĘPNOŚĆ, PRZYSZŁOŚĆ

#### Połowa września 2014:

- ▶ **prawie 11 400 lematów**:
  - ▶ 10 969 czasownikowych,
  - ▶ 217 rzeczownikowych (prace niedawno rozpoczęte),
  - ▶ 178 przymiotnikowych (prace niedawno rozpoczęte),
- ▶ **ponad 50 000 schematów walencyjnych**,
- ▶ tj. średnio prawie 4,4 schematu na lemat,
- ▶ **prawie 160 000 argumentów**,
- ▶ tj. średnio prawie 3,2 argumentu na schemat,
- ▶ ponad 7% schematów zawiera **zleksykalizowane specyfikacje argumentów wielowyrazowych** (przyimki złożone, frazeologizmy).

<http://zil.ipipan.waw.pl/Walenty>:

- ▶ **kolejne wersje**, w tym z września 2014,
- ▶ **w różnych formatach** (PDF, czysty tekst),
- ▶ **poniższe artykuły**.

#### Plany:

- ▶ **Do końca 2015 roku**:
  - ▶ **15 000 lematów**,
  - ▶ w tym co najmniej **3 000** nieczasownikowych,
  - ▶ warstwa semantyczna:
    - ▶ **role semantyczne** (agens, patiens itd.),
    - ▶ **ograniczenia selekcyjne**.
- ▶ **W dalszej perspektywie** (projekty złożone do NCN-u):
  - ▶ rzutowanie na **słowniki walencyjne czeskie i angielskie**,
  - ▶ znacznie **bogatszy opis lingwistyczny**, szczególnie frazeologizmów.

• Przepiórkowski A., Hajnicz E., Patejuk A., Woliński M. 2014a: Extended phraseological information in a valence dictionary for NLP applications, [w:] Proceedings of the Workshop on Lexical and Grammatical Resources for Language Processing (LG-LP 2014), Dublin, Ireland, s. 83–91. • Przepiórkowski A., Skwarski F., Hajnicz E., Patejuk A., Świdziński M., Woliński M. 2014b: Modelowanie własności składniowych czasowników w nowym słowniku walencyjnym języka polskiego, Polonica XXXIII, s. 159–178. • Przepiórkowski A., Hajnicz E., Patejuk A., Woliński M., Skwarski F., Świdziński M. 2014c: Walenty: Towards a comprehensive valence dictionary of Polish, [w:] N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, i S. Piperidis (red.): Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavík, Iceland, s. 2785–2792.