Handling MWEs in Walenty, a new valence dictionary for Polish [WG1]

Agnieszka Patejuk
aep@ipipan.waw.pl

PARSEME meeting, Athens, March 2014
What is this poster about?

- modelling Polish MWEs together with their syntactic structure
- framework: Lexical-Functional Grammar (LFG)
- platform: Xerox Linguistic Environment (XLE)
- Walenty, a valence dictionary of Polish:
  - open source, available from: zil.ipipan.waw.pl/Walenty
  - developed since 2012, spans 3 projects
  - contains 38874 schemata for 8644 verbs
  - created on the basis of attested data
  - can be used by various formalisms (currently: LFG)
  - accounts for coordination (syntactic positions as sets)
  - accounts for MWEs:
    - internal structure (NP/PP, fixed phrase)
    - interactions with syntax (case assignment for NPs)
    - displayed modification pattern
**Intro**

**Aim**
Modelling Polish MWEs together with their internal syntactic structure

**Means**
- framework: Lexical-Functional Grammar (LFG)
- platform: Xerox Linguistic Environment (XLE)
- valence dictionary: Walenty

**1. LFG**

**Formalism**
- constraint-based, highly lexicalised
- parallel levels of representation:

  \[
  \begin{array}{c}
  \text{S} \\
  \text{DP} \\
  \text{VP} \\
  \text{DP} \\
  \text{(↑\hspace{0.3cm}SUBJ)=↓} \\
  \text{↑=↓} \\
  \text{(↑\hspace{0.3cm}OBJ)=↓} \\
  \text{OBJ} \\
  \text{PRED} \\
  \text{PECK} \\
  \text{STORK} \\
  \text{STARLING} \\
  \text{TENSE} \\
  \text{PAST}
  \end{array}
  \]

- analyses of diverse languages (English, Warlpiri, Russian, Urdu...)
- LFG grammars may be implemented in XLE
- attempts at commercial use (Bing search engine)

**Polfie**
- LFG grammar of Polish implemented in XLE

**3. MWES in Walenty**

**MWE Types**
- fixed expressions:
  - cannot be modified in any way, the exact string is given
  - fixed(string)
- lexicalised phrases:
  - nominal: lexnp(case, number, lemma, mod)
  - prepositional: preplexnp(preposition, case, number, lemma, mod)
- typical information: case, preposition form
- extra information: number, lemma, modification pattern

**Modification Patterns**
- natr: modification not allowed
- atr: modification allowed (though not necessary)
- ratr: modification required (often possessive, NP or adjective)
- batr: specific modification required (possessive: SWÓJ or WŁASNY, 'own')

**Examples**
- subj{np(str)} + obj{np(str)} + \{fixed('na kwaśne jabłko')\}
  Zbił ich na (*bardzo) kwaśne jabłko/*jabłka.
- beat then for very sour apple.SG/PL
  'He beat them to a pulp.' (literally: 'He beat them into a sour apple.')
- subj{lexnp(str.sg, 'krew', atr)} + {preplexnp(w,loc,pl,'żyła',ratr)}
  (Gorąca) krew/*krawie płynie/*pływą w *(jej/ Marysi/tých) żyłach/*żyłe.
- hot blood.SG/PL flow.SG/PL in her/Mary's/those veins.PL/SG
  'Hot blood flows in her/Mary's/those veins.'
- subj{np(str)} + \{np(str.sg, 'głowa', natr)}
  Daje (*swój/mały) głowy/*głowy że, przypina.