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:

```
  S (↑ SUBJ) = ↓ a stork
  VP ↑=↓ PECK
  OBJ (↑ OBJ) = ↓ a starling

  S
  DP
  ←SUBJ
  VP
  ←OBJ
  DP
  ↑ SUBJ → OBJ
  TENSE PAST
```
- analyses of diverse languages (English, Warlpiri, Russian, Urdu...)
- LFG grammars may be implemented in XLE
- attempts at commercial use (Bing search engine)

POLFIE
- an LFG grammar of Polish implemented in XLE
- based on previous grammars (DCG, HPSG)
- morphological information from analyser, treebank or corpus
- valence information from converted dictionary
- coverage: parses 32% of sentences of the National Corpus of Polish (NKJP: nkjp.pl)
- structure of the (near) future: adding semantics
- open source, available from: zil.ipipan.waw.pl/LFG

2. WALENTY

About
- valence dictionary developed since 2012, spans 3 projects
- contains 38874 schemata for 8644 verbs (as of 5/03/2014)
- created on the basis of attested data (from NKJP, from the web)
- open source, available from: zil.ipipan.waw.pl/Walenty

Formalism
- syntactic positions (separated by "\(\lambda\)"") are sets (enclosed in "\{\}")
- realisations of the position are members of the relevant set (separated by "/")
- realisations belong to the same set if they may be coordinated

```
\{\obj{np(str)} + \subj{np(str)} + \{\fixed{na kwaśne jabłko}\}\}
```
- fixed expressions:
- cannot be modified in any way, the exact string is given
- fixed(string)
- lexicalised phrases:
- nominal: lexnp(case,number,lemma,mod)
- prepositional: prelexnp(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 adjectival)
- batr: specific modification required (possessive: SWÓJ or WŁASNY, 'own')

Examples
- subst{np(str)} + obj{np(str)} + \{\fixed{na kwaśne jabłko}\}
- Zbił ich na (*bardzo) kwaśne jabłko/*jabłka.
- batr: specific modification required (possessive: SWÓJ or WŁASNY, 'own')

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: prelexnp(preposition,case,number,lemma,mod)
- typical information: case, preposition form
- extra information: number, lemma, modification pattern

Conversion process
- python script (around 1K lines)
- takes entries from Walenty, returns XLE lexical entries
- grammatical function (GF) chosen on the basis of contents of the set corresponding to the relevant position (roughly: on the basis of morphosyntax)

Conversion MWEs into LFG constraints
- number: \(\uparrow\text{GF NUMBER}\) \(\equiv\text{NUM}\)
- lemma: \(\uparrow\text{GF PRED FN}\) \(\equiv\text{LEMA}\)
- modification:
- fixed: same modification constraints as natr
- natr: \(\neg\uparrow\text{GF ADJUNCT}\) \(\neg\uparrow\text{GF POSS}\)
- ratr: no constraint needed (modification allowed but not required)
- batr: \(\uparrow\text{GF ADJUNCT} \in \{\uparrow\text{GF POSS}\}\)
- batr: \(\uparrow\text{GF ADJUNCT} \in \{\uparrow\text{GF POSS}\}\)

5. ISSUES
- not all modification constraints can be expressed in Walenty
- no information about category corresponding to fixed
- semantics: compositional vs non-compositional