

WG4

Representation of MWEs in the Lithuanian Dependency Treebank

Jolanta Kovalevskaitė, Erika Rimkutė, Loïc Boizou

Centre of Computational Linguistics, Vytautas Magnus University (Kaunas, Lithuania)

Lithuanian Dependency Treebank (LDT) *Alksnis* (work in progress)

- 2300 sentences of several text types (newspapers, journals, fiction, legal texts)
- **Morphological** annotation according to *MULTEXT-East format*.
- **Syntactic** annotation: dependency model adapted from the *Prague Dependency Treebank* analytical layer.
- The syntactic analysis is produced by a *rule-based parser*.
- Both morphological and syntactic annotations are performed automatically, but revised manually.

Types of MWEs in Lithuanian:

idioms, collocations, named entities, proverbs, MWEs of grammatical nature.

In LDT, we annotate all words of different MWE types separately (Fig. 2), except for those of grammatical nature (Fig. 1).

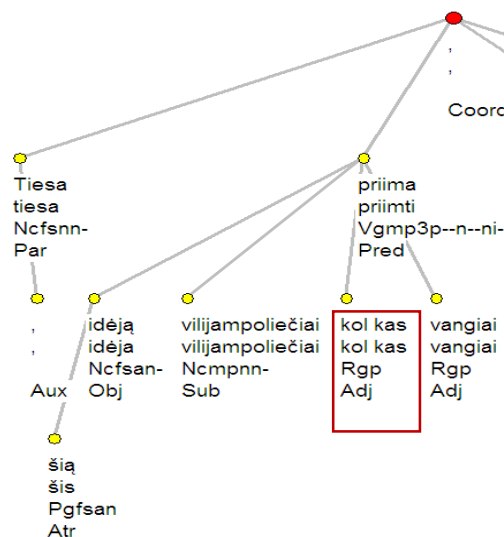


Fig. 1.

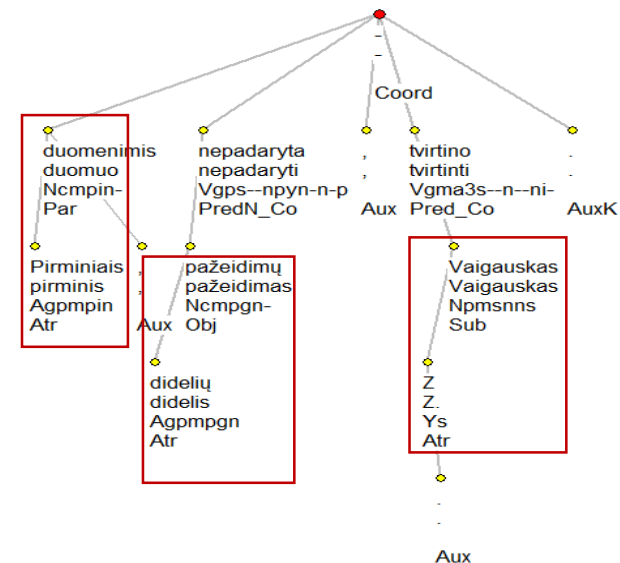


Fig. 2.

The **solutions for the representation** of each type of MWEs are still under development. Possible options:

- to apply more than one principle of annotation in respect to a particular MWE type;
- to have a special label for MWEs.