Verbal Valency in the MT Between Related Languages

Verb 2010, Pisa

Klyueva Natalia and Vladislav Kubon

Institute of Formal and Applied Linguistics

Charles University in Prague

Research Outline

- Languages: Czech and Russian (Slavic languages)
  - Rich morphology: 6 cases in Russian, 7 in Czech
- Valency: Vebrs + surface form of their arguments
  - Frames were extracted from the Ruslan dictionary
  - They were sorted on prepositional and non-prepositional
  - Ex. It demands special attention.Gen-ru/Acc-cz
    - Czech:To vyžaduje zvláštní pozornost.Acc vs.
    - Russian: Это требует особого внимания.Gen
- Machine Translation: evaluation ad-hoc

List of Valency Differences

- Not many verbs with different valency structure
- Different simple frames:
  - CZ:rušit+Acc vs. RU:мешать + Dat   - to disturb
  - CZ:vyžadovat+Acc vs. RU:требовать+Gen - to demand
- Different prepositional frames:
  - záviset   (na(loc) vs. зависеть от (gen)) - to depend on
  - narazit   na(acc) vs. столкнуться с(ins) - to run into
- Others (mixed):
  - docházet (k(dat) vs. возникать(nom) - to happen

Machine Translation

- Ad-hoc evaluation:
  - Rule-Based MT system for Slavic - Česílko
  - Evaluate MT output wrt. valency
  - Say if our list of differences can help or not
- Number of mistakes in valency is not so big
- A little valency dictionary is enough