# Multiword Expressions: Observations from a Parallel BG-EN Newsmedia Corpus

Petya Osenova, Alexander Popov and Kiril Simov (IICT-BAS)

PARSEME Meeting, Malta, 19-20 March 2015

# Summary

- Data: BG-EN phrase-aligned newsmedia corpus
- Types considered: *MWE-to-word*; *MWE-to-MWE*; *MWE-to-phrases*
- Observations on most frequent MWE types :
  - **English**: *Verbal MWE* (phrasal verbs, light constructions and VP idioms); *Nominal MWE* (Nominal compounds); *Other categories MWE*
  - **Bulgarian**: *Verbal MWE* (se-verbs, light constructions, VP idioms); *Nominal MWE* (other types) and *Other categories*
  - **Common**: the *MWE-to-MWE* correspondences and *MWE-to-word* are more frequent. In contrast, the *MWE-to-phrase* correspondence has not shown big distribution.