

# Evaluation of the system for extraction of multi-word expressions and prediction of their translations

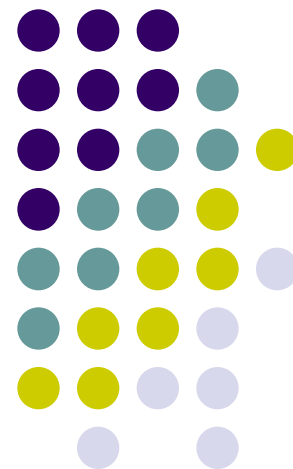
---

**Katerina Zdravkova**

University Sts Cyril and Methodius, Skopje

**Aleksandar Petrovski**

International Slavic University, Sveti Nikole



# System for extraction of MWEs and their translations



- Extraction
  - 3500 candidate MWEs, including some useless:
    - тоа би ја / *toa bi ja*, instead of тоа би ја усреќило / *toa bi ja usrekjilo* = that will make her happy
    - рече тој со / *reche toj so*, instead of рече тој со недоверба / *reche toj so nedoverba* = he said with mistrust;
- Syntactical filtering
  - Less than 500 phrases, sometimes inflections of the same phrase:
    - атомската бомба / *atomskata bomb* = the atomic bomb, атомски бомби / *atomski bombi* = atomic bombs
    - обичен човек / *obichen chovek* = an ordinary man, обичните луѓе / *obichnite lugje* = the ordinary men or the ordinary people,
    - шаховска табла / *shahovska tabla* = a chess board, шаховската табла / *shahovskata tabla* = the chess board
- Translation and cross evaluation
  - Less than 1000 candidate MWEs
- Evaluation of the results
  - 968 English candidate MWEs and their translations

# The crucial problems



- Existence of two candidate MWEs in two mutually aligned sentences
  - “a comb and a piece of toilet paper”
    - чешел и / *cheshel i* = “comb and”
    - парче тоалетна хартија / *parche toaletna hartija* = “a piece of paper”;
  - “guilty of the crimes they were charged with”
    - виновни за / *vinovni za* = “guilty of”
    - за кои беа обвинети / *za koi bea obvineti* = they were charged with
- Inconsistent manual translation
  - Found in 102 MWEs
    - “with the tips of his fingers”, “true feelings towards big brother” or “sweet summer air”.
  - “thieves bandits”
    - растурачи на дрога / *rasturachi na droga* = “drug dillers”
    - крадци бандити / *kradci banditi* = “thieves bandits”
  - “their hands crossed on their knees”
    - затворениците седеа / *zatvorenicite sedea* = “the prisoners sat”
    - неподвижно со / *nepodvizhno so* = “immobile with”
- Inconsistency measured with Herfindahl-Hirschman Index (HHI) measure
- We propose an index of completeness as its upgrading, to express the degree of correct translation of a complete MWE