

Annotation of MWEs and NEs in the Serbian extension of ELEXIS-WSD: comparisons, solutions and open questions

WG1&2

cvetana@jerteh.rs, ranka@rgf.rs, aleksandra.markovic@isj.sanu.ac.rs

1. The extension of ELEXIS-WSD

- automatically translated SS, checked, proofread; automatically tokenized, lemmatized, POS-tagged, manually corrected.

To do:

- annotation (MWEs, NEs & syntactic)
- linking with the sense repository.

2. All MWEs & NEs from the WSD

automatically translated into SR.

- *lingua franca* (6 lang. sets)
- 'Greece' the most frequent NE

3. The comparison of MWEs & NEs across languages

- automatic translation of MWEs sometimes imprecise: *uzdizam se* (BG) *nastati* 'become' (**ustati* (SR) 'get up')
- sometimes the translation of MWEs was good, but not annotated in SSS (*društvena mreža* 'social network')
- the lack of annotation is not unusual in other sets: *heavy water* (EN) & *тежка вода* (BG) weren't annotated neither.