About

• First Latvian Wordnet, work in progress
• Goal: high quality core based on corpus evidence
• Mapping to Princeton WordNet
• Lexical data from Tezaurs.lv online dictionary
• Data from Latvian National corpora collection Korpuss.lv

Results

• 2717 fully annotated synsets with examples, semantic relations and external links to Princeton WordNet
• WordNet connects 6432 words (5528 synsets)
• Publicly available at Tezaurs.lv and wordnet.aillab.lv
• Basis for ongoing WSD development for Latvian

Methodology and Tools

Revisiting word sense inventory based on corpus evidence

Relations: synonym sets and gradation sets, hyponymy, meronymy, antonymy, similarity