

# Cambridge update

Ann Copestake

Computer Laboratory, University of Cambridge

July 2013

## DELPH-IN related activities in Cambridge

- ▶ MPhil projects
  - ▶ Using Distributional Semantics to Improve Parse Ranking: Guy Emerson (Tues pm)
  - ▶ A user-extensible Kyrgyz-English MT demonstrator: Nargiza Sarkulova
  - ▶ Incremental Semantic Disambiguation: Shisen Fan
  - ▶ Extraction of Distributions from a Semantically Annotated Corpus: Hao Li
- ▶ Project with IBM (David Mott's talk, Weds am).
- ▶ Ongoing distributional work (partly with Aurélie Herbelot).
- ▶ PARSEME COST action on MWEs (first meeting in September).

## A user-extensible Kyrgyz-English MT demonstrator

- ▶ Objective: build a web-based system for MT on simple sentences that non-linguist users can add lexical entries to.
  1. Use Matrix to build initial grammar of Kyrgyz.
  2. Extend grammar.
  3. Use web demo code to build an interactive MT system.
  4. Develop templates for lexical entry classes.
  5. Test with three users.

Many thanks to oe and Emily for advice and help.

## Extraction of Distributions from a Semantically Annotated Corpus

- ▶ Objective: construct a range of distributional models on Wikiwoods and evaluate.
  1. DMRS version of Wikiwoods.
  2. Construct distributions with various parameters (no dependency, unlabelled dependency (various lengths), labelled dependency. Formalise similarly to Pado and Lapata.
  3. Use standard similarity test set (R & G) for parameter setting.
  4. Test on other similarity data (M & C), plus TOEFL synonym.
  5. Unlabelled 1-step dependencies perform significantly better than no dependency baseline.
  6. Results competitive with other corpus-based approaches (considerable improvement on Pado and Lapata).