CorefMT

Michal Novák, Liane Guillou
Coref in MT

Question: Can coreference help in translating other words than pronouns?

Example: He bought a bread and cut it into pieces.

cut prices -> snížit ceny
cut bread -> krájet chleba
cut forest -> kácet les
cut paper -> rezat/stříhat papír
Baseline

- using TectoMT - a tree-to-tree MT system implemented in Treex framework
- A discriminative model for every English lemma
- Classes are possible Czech translations
- Features extracted from the current node, its parent, children, the preceding and following node in the surface ordering
  - lemma
  - morpho-syntactic pattern
  - POS
  - grammatical features
Progress

- Integrated Stanford CoreNLP coref tool as a Treex block
- Designed and ran 5 experiments:
  - Baseline
  - Replace child’s lemma with antecedent’s lemma
  - Replace child’s/parent’s lemma with antecedent’s lemma
  - Add new feature: lemma of closest antecedent
  - Add new feature: lemmas for all antecedents in coref chain
- we did not retrain the discriminative model
Results

• No change in BLEU :(
• Some translations better, some worse
• Many errors caused by the coreference resolver
Extensions

- Retrain discriminative model
- Extend to NPs - so far just looked at pronouns
- Detailed manual analysis of changes in MT output