The Challenge of Syntax-Based Machine Translation

Ondřej Bojar

Computational Linguistics

ÚFAL, Charles University, Prague

(visited LT Group, CSSE, Melbourne University)
Syntax-Based Machine Translation

Do people translate structure?

No
Translating Trees

Decompose

Translate

Combine

People

Structure

People

Structure
Results & Discussion

• Tree translation performs poorly:
  – More steps → More errors.
  – Complex structure → More conflicts.

• Future experiments:
  – Use smaller bits (decompose nodes).
  – Improve components of the system.

• Lazy alternative given enough data:

  Překládají lidé  
  Do people translate

  rovnou celé fráze  
  longer phrases at once

  ?  

Maybe