

Maximum Entropy Translation Model in Dependency-Based MT Framework

David Mareček, Martin Popel, Zdeněk Žabokrtský

Charles University in Prague, Institute of Formal and Applied Linguistics

Aim

- to improve translation quality by using context sensitive translation model (TM),
- exemplified on English-to-Czech MT.

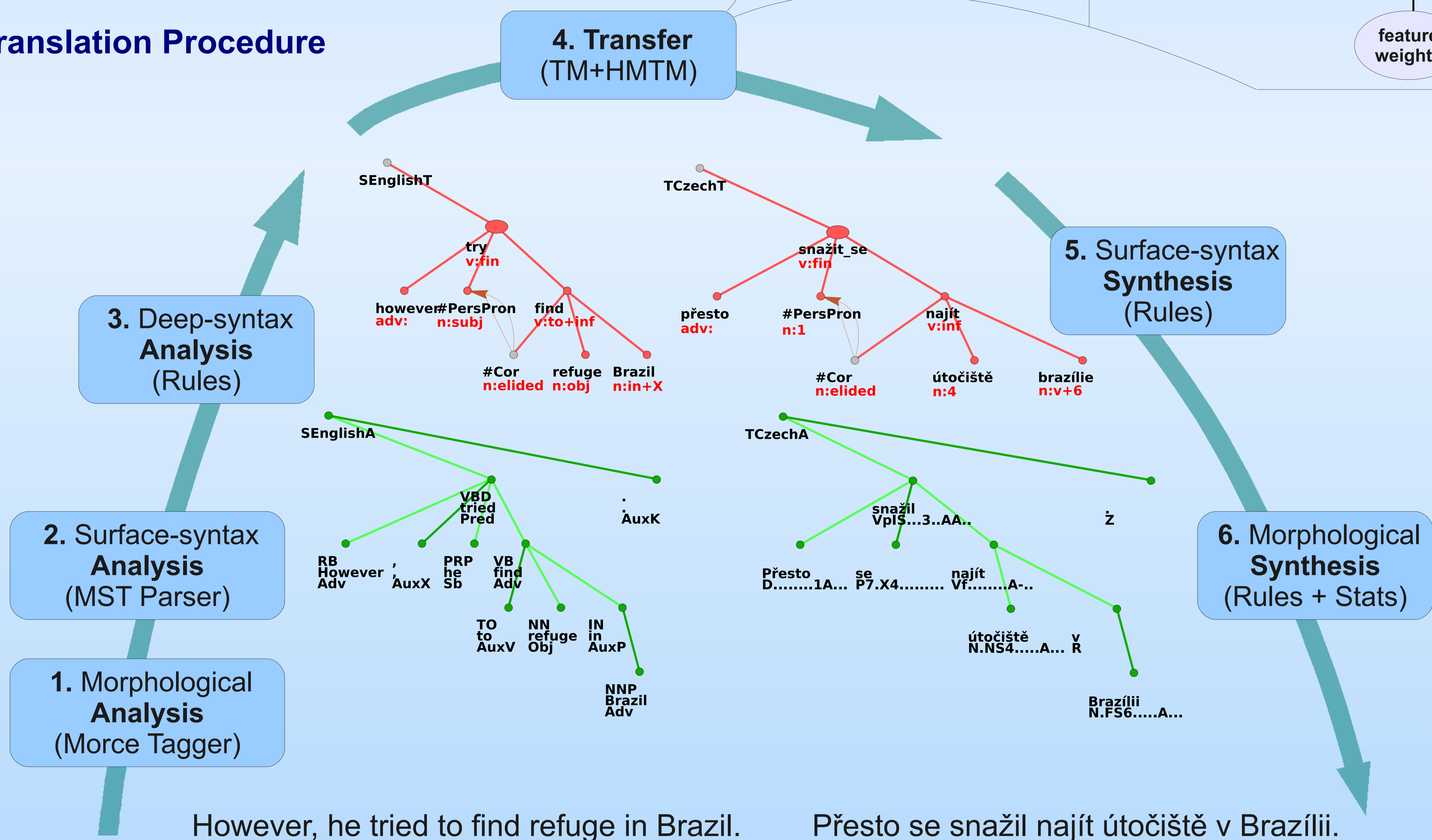
Framework

- Analysis-Transfer-Synthesis**
- Layers of language description adopted from the Prague Dependency Treebank:
 - Morphology
 - Surface dependency syntax
 - Deep dependency syntax
- Implemented in TectoMT

Resources

- 800 MW Czech monolingual corpus
- 60 MW parallel corpus CzEng 0.9 (parsed and aligned at the deep-syntactic level)

Translation Procedure



Evaluation

- WMT 2010 test set (2489 sentences)
- Baseline and MaxEnt TMs also combined with Target Language Tree Model (TreeLM) using Hidden Markov Tree Model approach.
- Confirmed overlap between the contribution of MaxEnt TM and TreeLM.

BLEU scores	No LM (just simple compatibility rules)	TreeLM (using Hidden Tree Markov Model)
Baseline TM	10.44	11.77
MaxEnt TM	11.77	12.58
improvement	+1.33	+0.81

Conclusion

MaxEnt-based translation models help to improve English-to-Czech translation quality in our dependency-based MT system.