Machine Learning for Deep-syntactic MT

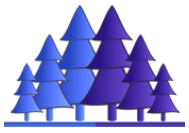
Martin Popel

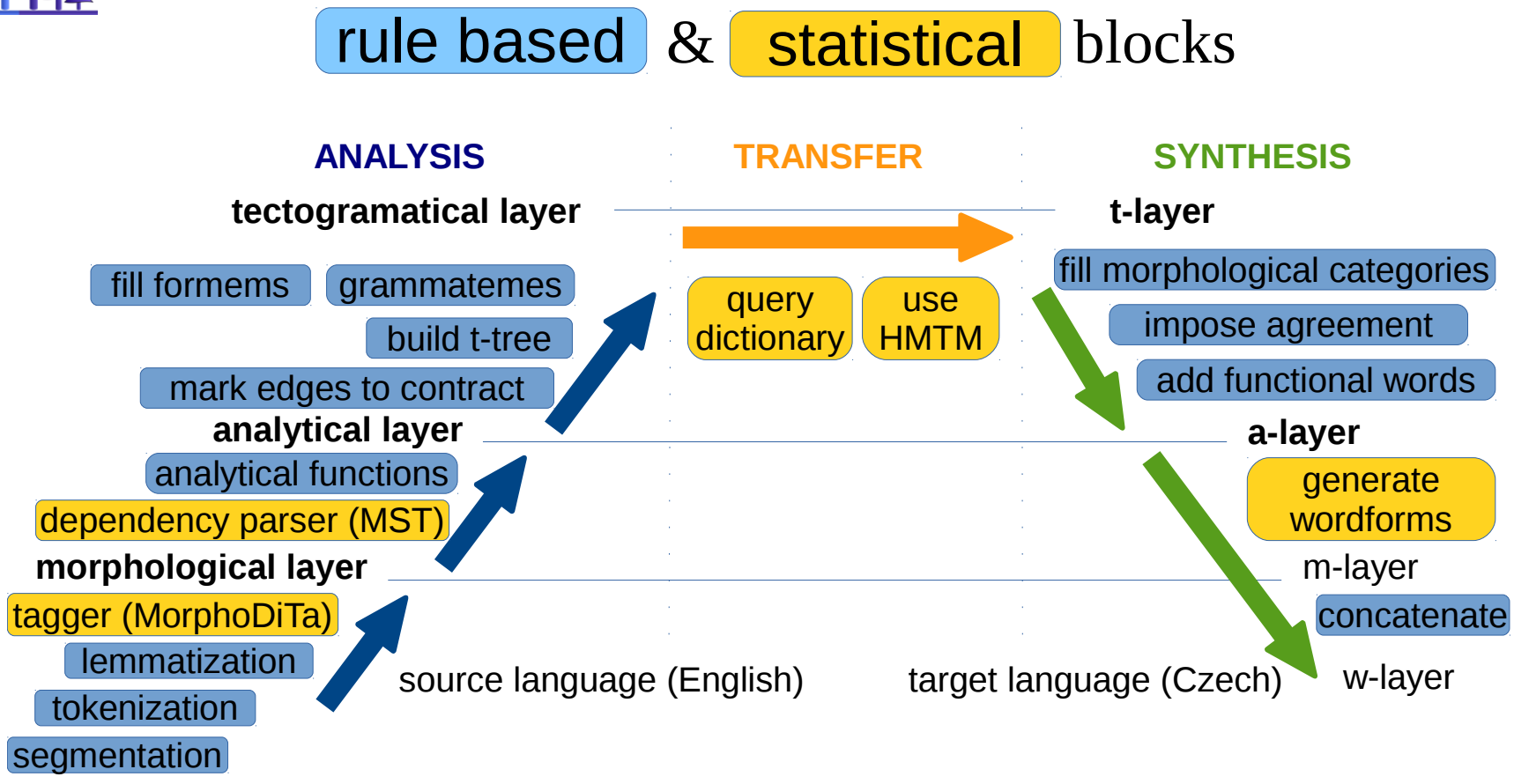
ÚFAL (Institute of Formal and Applied Linguistics)
Charles University in Prague

September 11, 2015
Seminar on the 35th Anniversary of the Cooperation between
Charles University in Prague and Hamburg University

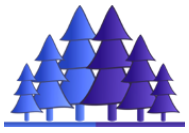
Outline

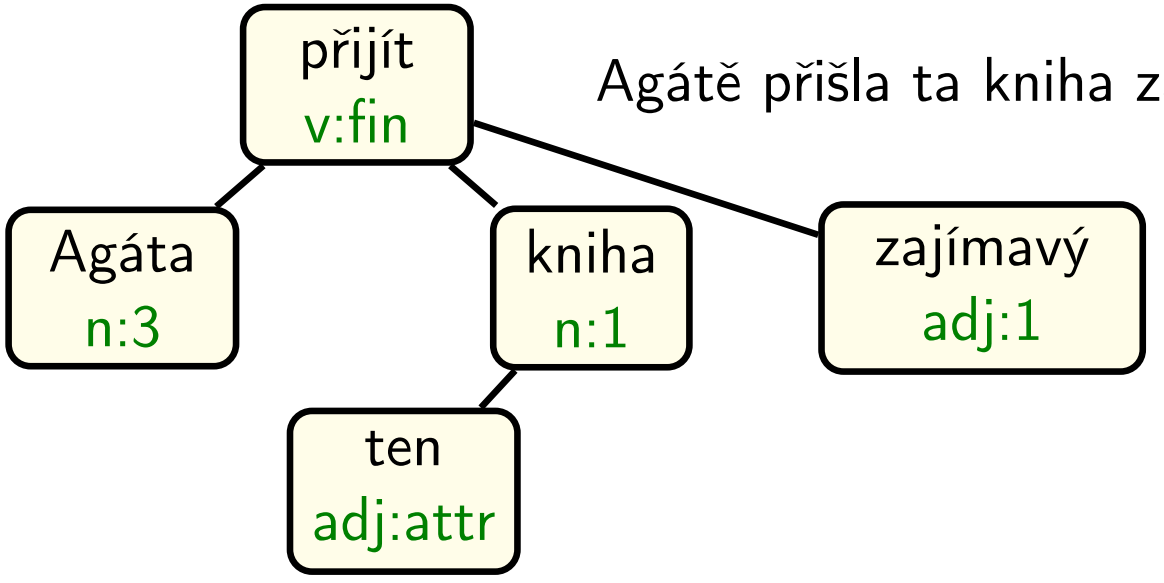
TectoMT: analysis, transfer, synthesis



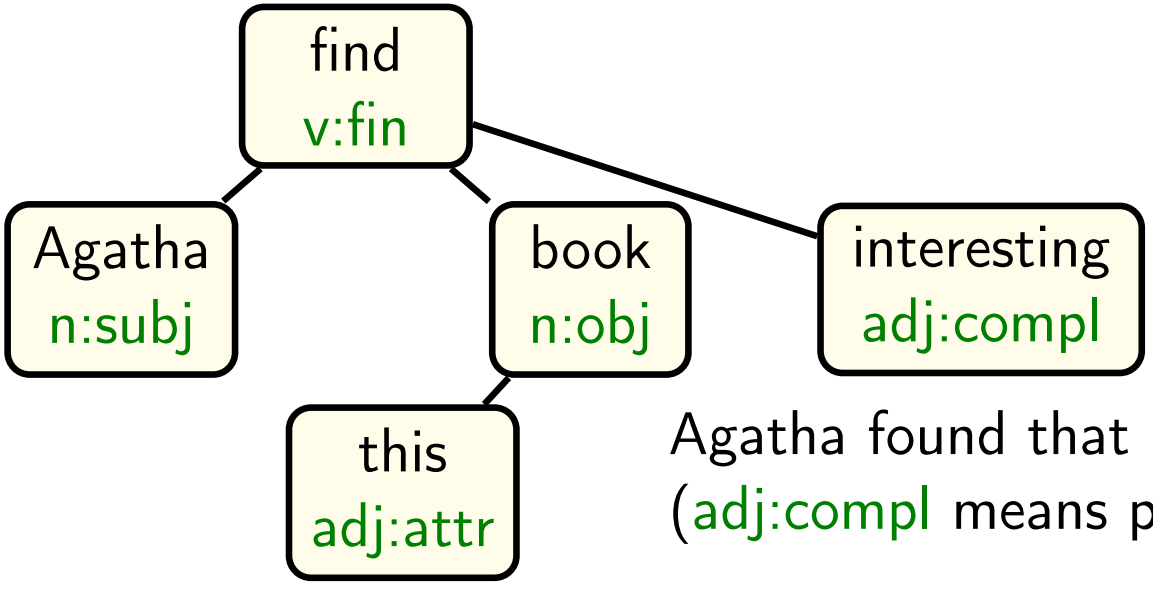


TectoMT: isomorphic transfer (1-1 node mapping)





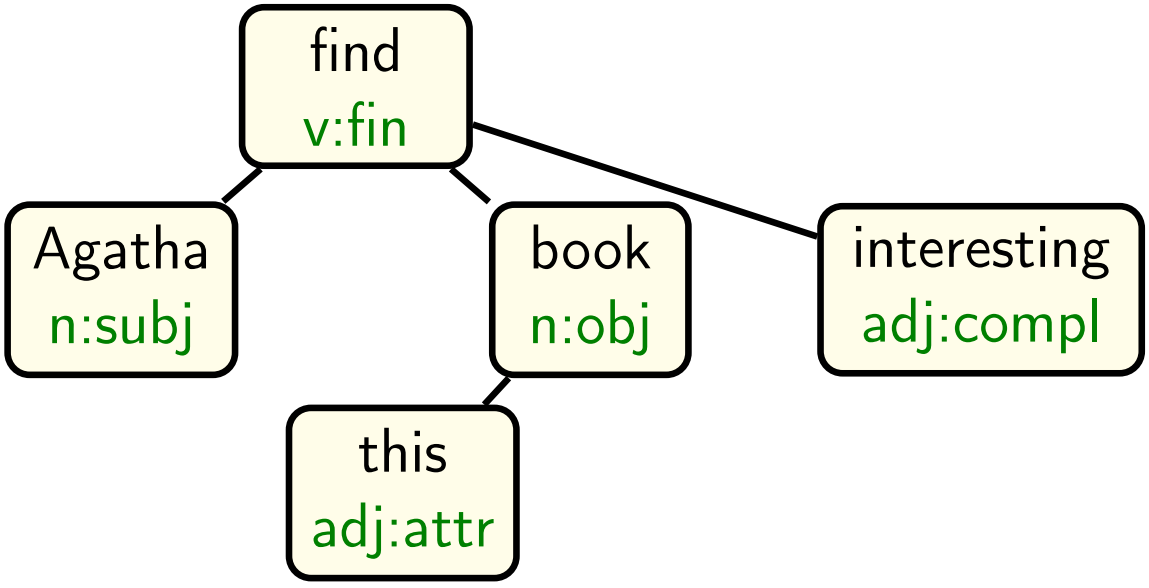
Agátě přišla ta kniha zajímavá.



Agatha found that book interesting.
(adj:compl means predicative adjective)

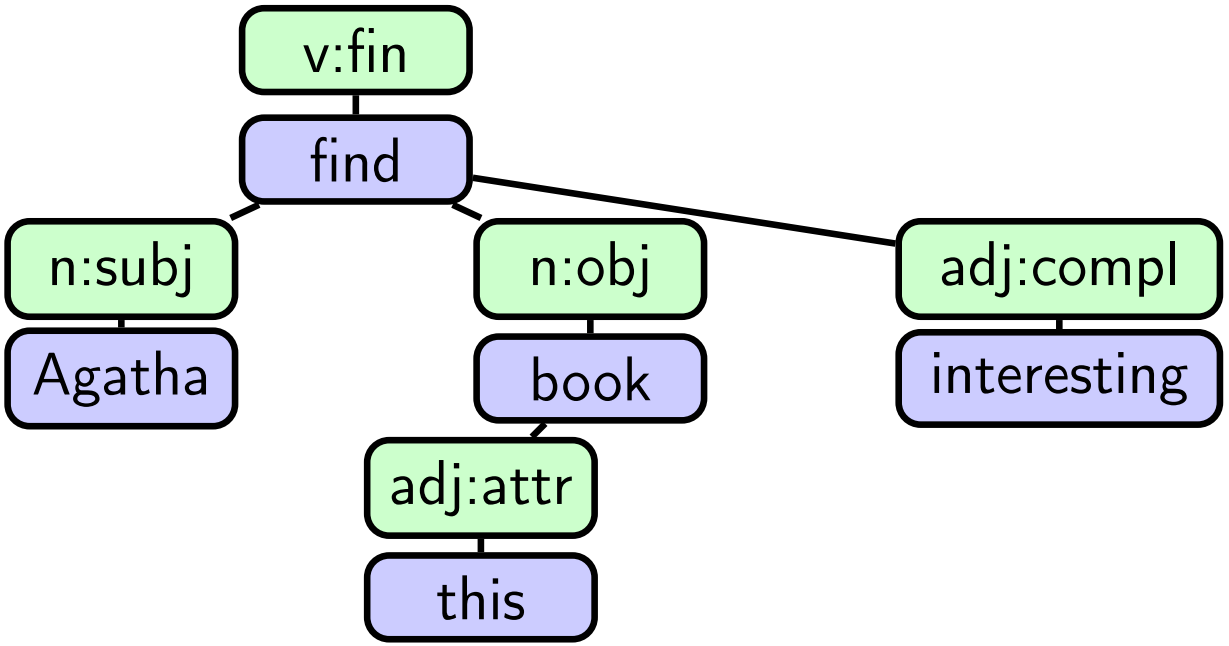
Representation of t-layer

lemma and formeme as two attributes



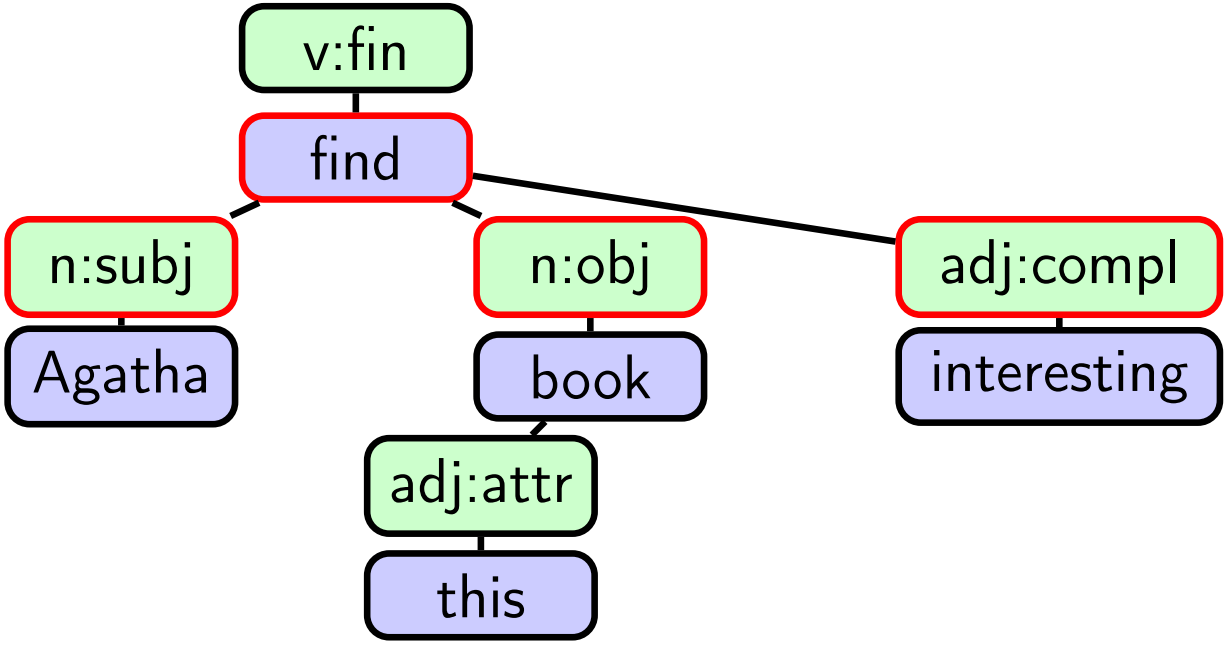
Representation of t-layer

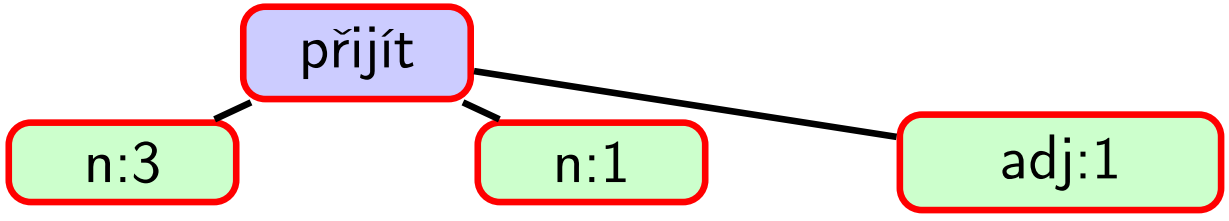
lemma and formeme as interleaved "sub-nodes"



Representation of t-layer

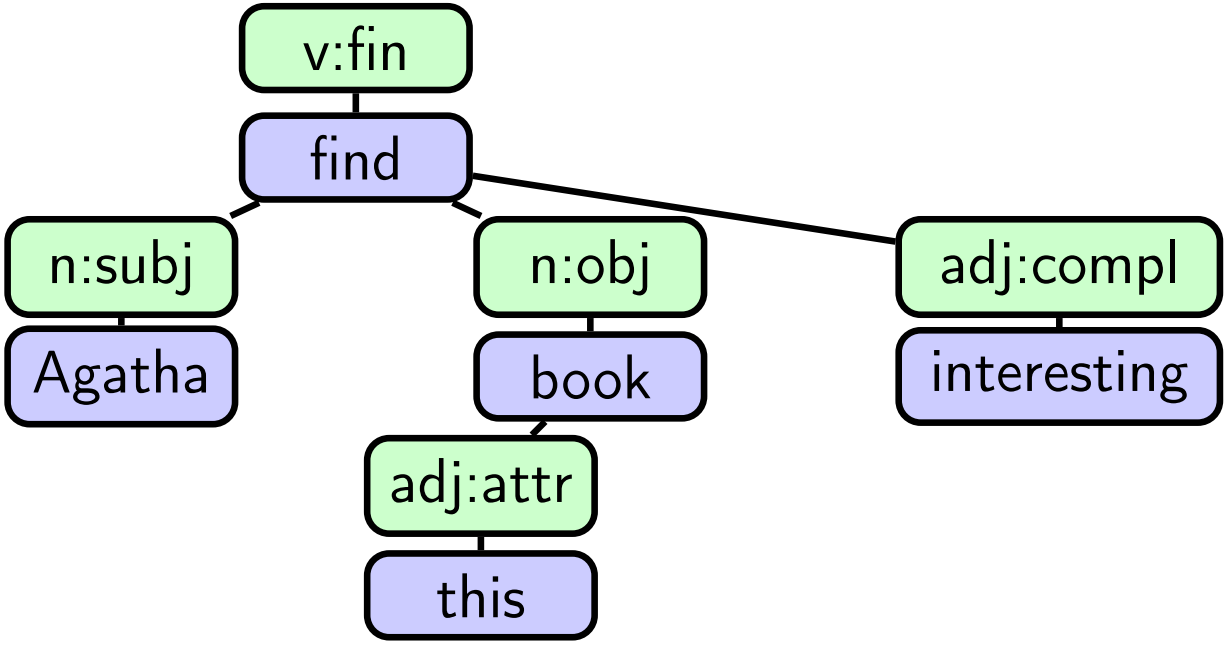
lemma and formeme as interleaved "sub-nodes"





Representation of t-layer

lemma and formeme as interleaved "sub-nodes"



grammatemes:

- translated in postprocessing (current approach)
- as subnodes (leaves, children of lemmas)
- encoded within lemma, but only if grammateme changed

Handling non-isomorphic transfer

- preprocessing or postprocessing within transfer (current approach)
- natively in the main transfer algorithm
- convert training data to isomorphic trees [not tried yet]
  - n-1 alignment: add special [delete_node] label to the target side
  - 1-n alignment: encode added nodes (L+F) into the "main" lemma
  - encode topology change: as_child, as_sibling, as_parent

TectoMT transfer over years

year	*BLEUdiff*	*method*
2008		initial baseline
2009	+1.5	HMTM (TreeViterbi, TreeLM)
2010	+0.8	HMTM + MaxEnt
2012	-2.2	TectoMoses
2012	NA	Gibbs sampling treelets
2013	-3.0	Easy-first treelets
2013	-2.0	Interpol treelets
2014	+0.1	VowpalWabbit

TectoMT transfer over years

year	*BLEUdiff*	*method*
2008		initial baseline
2009	+1.5	HMTM (TreeViterbi, TreeLM)
2010	+0.8	HMTM + MaxEnt
2012	-2.2	TectoMoses
2012	NA	Gibbs sampling treelets
2013	-3.0	Easy-first treelets
2013	-2.0	Interpol treelets
2014	+0.1	VowpalWabbit
	+0.9	other improvements in 2010–2014

TectoMT transfer over years

year	*BLEUdiff*	*method*
2008		initial baseline
2009	+1.5	HMTM (TreeViterbi, TreeLM)
2010	+0.8	HMTM + MaxEnt
2012	-2.2	TectoMoses
2012	NA	Gibbs sampling treelets
2013	-3.0	Easy-first treelets
2013	-2.0	Interpol treelets
2014	+0.1	VowpalWabbit
	+0.9	other improvements in 2010–2014
2015	+8.6	QTLeap en→cs in two months

2008: baseline TectoMT transfer

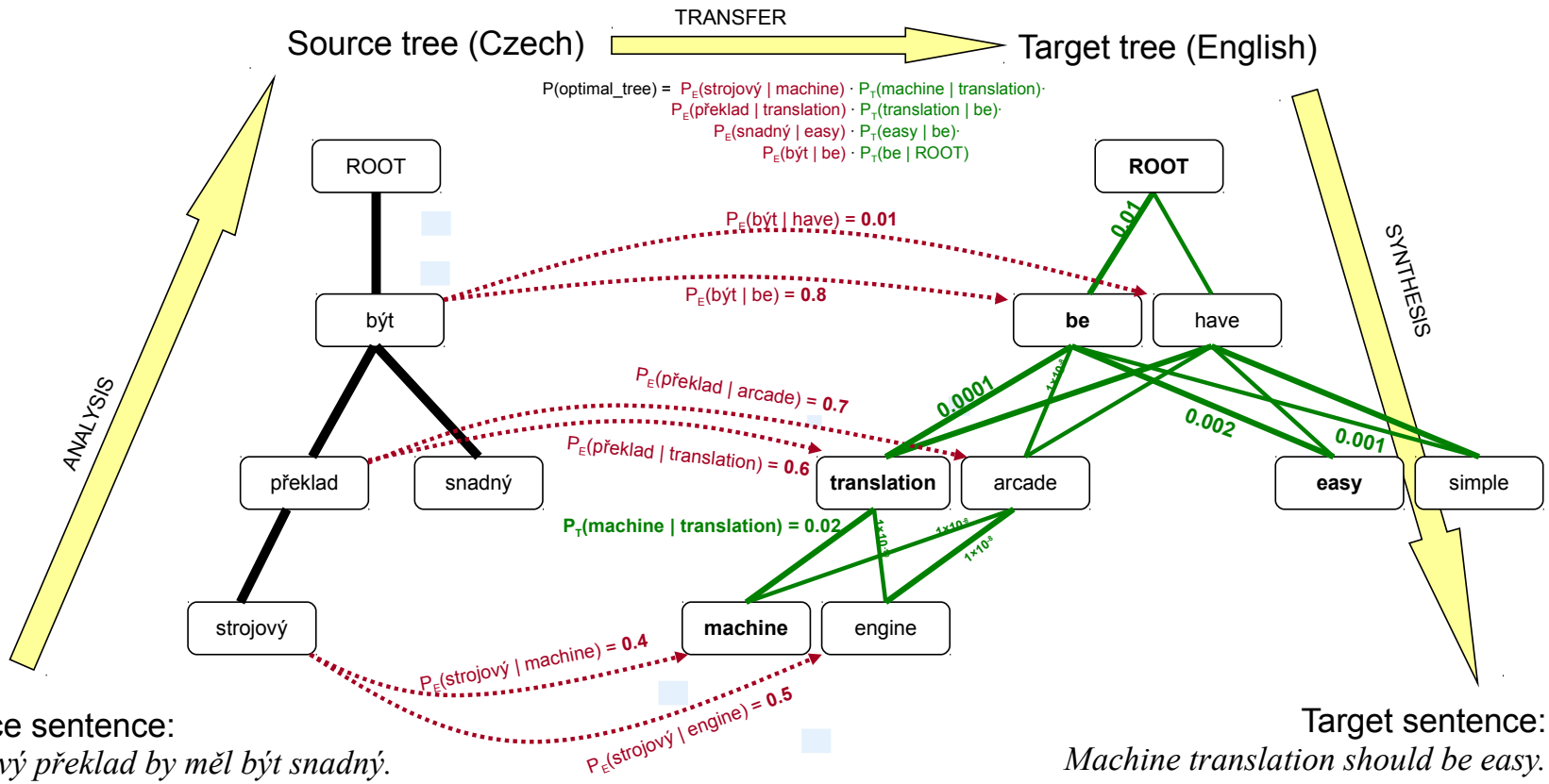
- "static" translation model $P(target|source) = \frac{\#(source,target)}{\#(source)}$
- first translate formemes, then lemmas
- use **only the top variant**

WMT 2009 en→cs results

	BLEU	human score
Moses (CUNI)	14.2	61
Google	13.6	66
Moses (UEdin)	13.5	53
Eurotran XP	9.5	67
PC Translator	9.4	67
TectoMT	7.3	48

2009: Hidden Markov Tree Model (HMTM)

- still using "static" translation models, but also
- TreeLM (target lemma-formeme and parent-child compatibility)
- best labeling is found via HMTM (Tree Viterbi)

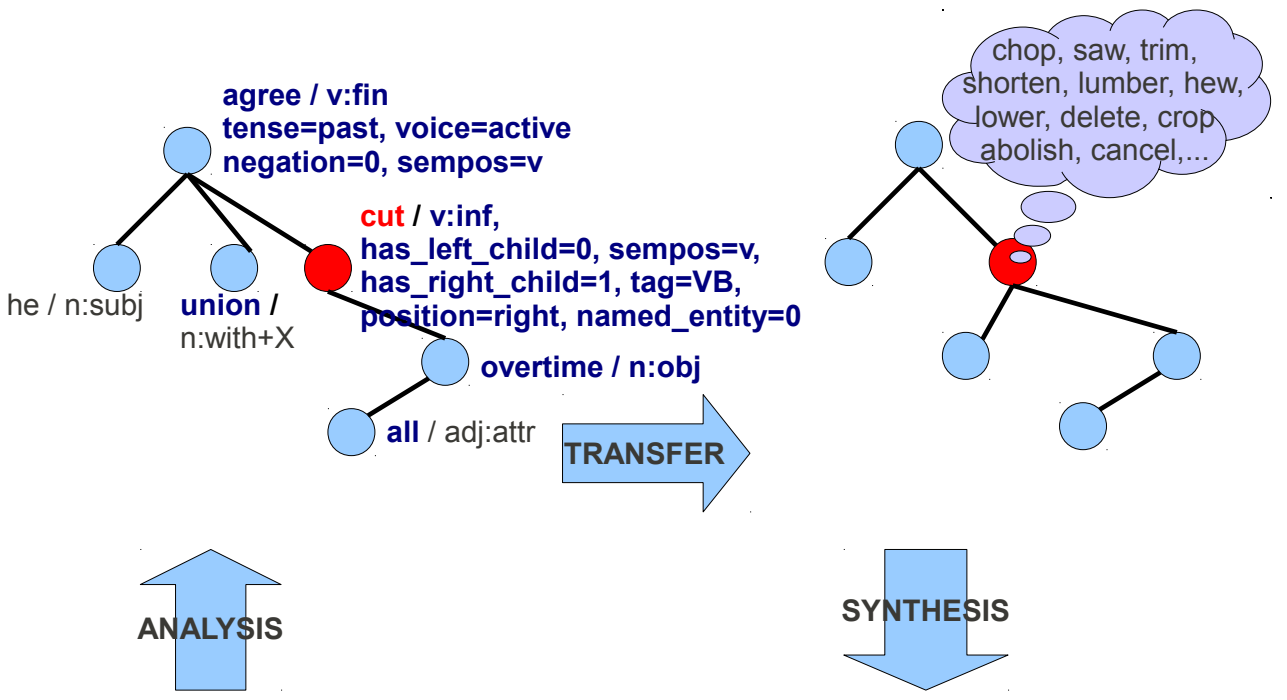


P_E(source | target) … emission probabilities … **translation model**

P_T(dependent | governing) … transition probabilities … **target-language tree model**

2010: Maximum Entropy translation model

- still using HMTM (and generative TreeLM),
- but the "static" model P(lemma | src_lemma) interpolated with
- context-sensitive discriminative (MaxEnt) model P(lemma | src_lemma, **other features**)



2014: VowpalWabbit-based transfer

- VW is an ultra-fast and modular machine learning toolkit
- optimized SGD (AdaGrad, dense+sparse features,...)
- cost-sensitive one-against-all reduction to binary classification
- logistic loss enables probabilistic interpretation (for HMTM)
- all lemmas in one model, fixed memory requirements
- label-dependent features (features shared for more lemmas)

VowpalWabbit Example

training data:

```
shared |S lemma=start formeme=v;for+ger neg=neg1 tag=VBG ...
1:0 _začít#V       |T start^začít#V      |P start^V
2:1 _zahájení#N    |T start^zahájení#N   |P start^N
3:1 _začínat#V     |T start^začínat#V    |P start^V
...
21:1 _spouštění#N |T start^spouštění#N |P start^N
```

training command:

```
vw -d train.data -c -f my.model
--loss_function=logistic --csoaa_ldf=mc -b 29 -qST
--holdout_off --passes 1 -l 3
```

test command:

```
vw -d test.data -c -i my.model -t -r out.predictions
```

Future plans

- non-isomorphic transfer
- experiments with VowpalWabbit
- include word embeddings (word2vec) as features
  - of the translated word (for rare words)
  - of its dependency context (for ambiguous words)
  - plus target-language embeddings of the translation
  - NN with a hidden layer

Thank you