Targeted Paraphrasing on Deep Syntactic Layer for MT Evaluation

Petra Barančíková, **Rudolf Rosa**

Charles University in Prague

Introduction

- Goal: Increase quality of machine translation evaluation
- Human evaluation
    - direct and most reliable
    - slow, expensive, subjective and irreproducible
- Automatic evaluation (BLEU, Meteor...)
    - compare MT output to pre-existing reference translation
    - fast and cheap
    - only approximates human judgment, not considering synonymous expressions, incorrect alignments...
- Approach: Improve automatic evaluation by providing a better reference

Example (WMT2012)

Source	*Banks are testing payment by mobile telephone*					
Hypothesis	*Banky*	*zkoušejí*	*platbu*	*pomocí*	*mobilního*	*telefonu*
	Banks	are testing	payment	with help	mobile	phone
	Banks are testing payment by mobile phone					
Reference	*Banky*	*testují*	*placení*	*mobilem*		
	Banks	are testing	paying	by mobile phone		
	Banks are testing paying by mobile phone					

Related work

- automated metrics with paraphrase support (Meteor)
- paraphrasing reference sentences for MT evaluation

Our approach

- paraphrasing and the translation itself are very similar tasks
- there are many great tools for machine translation
- let's treat paraphrasing as monolingual machine translation
  - "translate" the reference sentence into its paraphrase so that it is more similar to the hypothesis

Data

**Czech paraphrasing tables**

- **WordNet PDT 1.9.**
  - high quality lemmatized paraphrases
  - insufficient amount

- **Meteor**
  - large amount of paraphrases
  - lot of noise, especially among multiword paraphrases – e.g.
    jeho názoru (of his opinion) $\sim$
    šermoval rukama a mlátil neviditelného (waved his arms and beat the invisible one)
  - lemmatization and filtration of one-word paraphrases

Layers
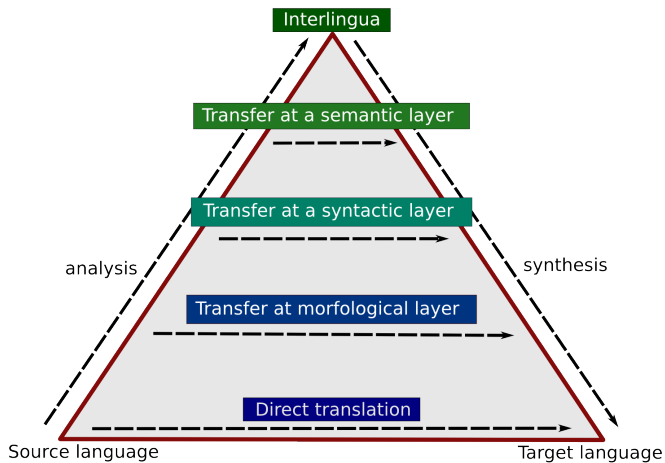
Rule-based systems

- inconclusive results of paraphrasing using direct phrase-based translation by Moses (Barančíková and Tamchyna, 2014)
- advantages of rule-based paraphrasing at deep syntactic level
    - better preservation of meaning and grammaticality
    - more conservative paraphrases
    - no need to create large translation tables
    - adjusting the word order

Treex

- highly modular NLP software system
- developed for machine translation system TectoMT that translates on a deep syntactic layer
- open-source and available on GitHub (https://github.com/ufal/treex)
- stratificational approach to language, adopted from the Functional Generative Description theory and its later extension by the Prague Dependency Treebank
- analysis and synthesis pipeline for Czech and English
    - some support for several other languages

Layers

- **w-layer:** word layer
- **m-layer:** morphological layer
- **a-layer:** shallow-syntax/analytical layer
    - dependency tree over **tokens**
- **t-layer:** deep-syntax/tectogrammatical layer
    - dependency tree over **content** words (lemma)
    - attributes capture information from function words and inflection
   - part-of-speech, case, tense, number, gender...

Paraphrasing on t-layer

- paraphrase the lemma where appropriate
  - if a lemma in the MT hypothesis *can be* correct (based on paraphrase tables and the reference), then assume it *is* correct
- keep the other attributes unchanged

Example (WMT2012)

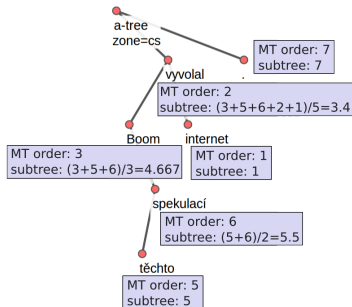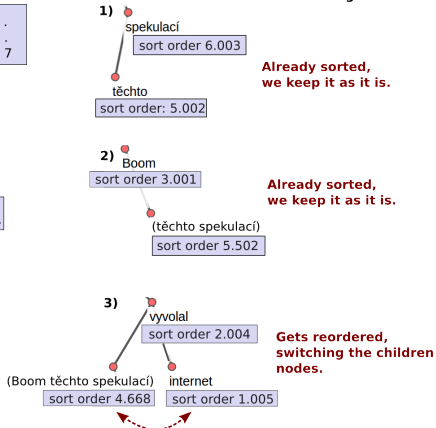| **Source**	The Internet has caused a boom in these speculations.
Hypothesis	Internet vyvolal boom v těchto spekulacích .
	Internet caused boom in these speculations .
	The Internet has caused a boom in these speculations.
Reference	Rozkvět těchto spekulací způsobil internet .
	Boom these speculations caused internet .
	A boom of these speculations was caused by the Internet.

Paraphrasing module
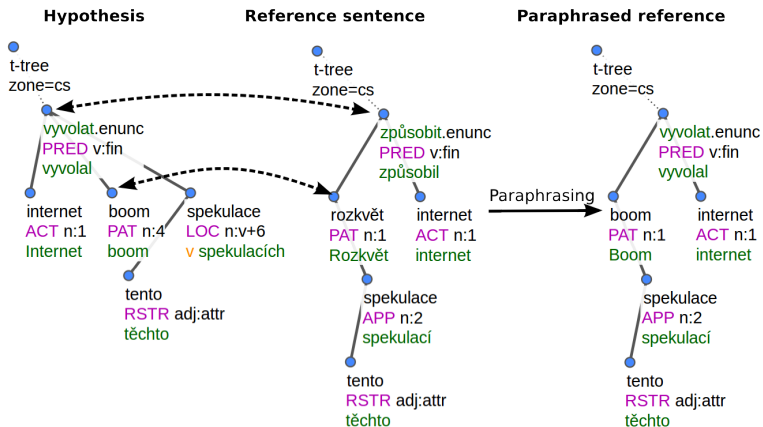
Reordering module

Hypothesis

Lemmas:	Internet	vyvolal	boom	v	těchto	spekulacích		.
	internet	vyvolal	boom	v	ten	spekulace		.
MT order:	1	2	3	4		6		7

Paraphrased reference

a-tree
zone=cs

vyvolal
MT order: 2
subtree: (3+5+6+2+1)/5=3.4

MT order: 7
subtree: 7

Boom
MT order: 3
subtree: (3+5+6)/3=4.667

internet
MT order: 1
subtree: 1

spekulací
MT order: 6
subtree: (5+6)/2=5.5

těchto
MT order: 5
subtree: 5

Recursive reordering

1)
spekulací
sort order 6.003

těchto
sort order: 5.002

Already sorted,
we keep it as it is.

2)
Boom
sort order 3.001

(těchto spekulací)
sort order 5.502

Already sorted,
we keep it as it is.

3)
vyvolal
sort order 2.004

(Boom těchto spekulací)
sort order 4.668

internet
sort order 1.005

Gets reordered,
switching the children
nodes.

Results

- 2x12 English-to-Czech machine translation systems
  - manual evaluation $\rightarrow$ human score
  - automatic evaluation $\rightarrow$ BLEU, Meteor, Exact Meteor
    - original reference
    - paraphrased reference
    - paraphrased & reordered reference

- Pearson correlation of human score and automatic score

	WMT12			**WMT13**		
references	original	para	para&reord	original	para	para&reord
BLEU	0.751	0.783	0.804	0.834	0.850	0.878
Meteor	0.833	0.864	0.868	0.817	0.871	0.870
Ex.Meteor	0.861	0.900	**0.903**	0.848	**0.893**	**0.893**

Future work

- more complex paraphrases (now only single-word)
- syntactic paraphrases (now only lexical)
- more sophisticated reordering (more constraints?)
- more languages (Treex-supported)