



# Solving Three Czech NLP Tasks End to End...

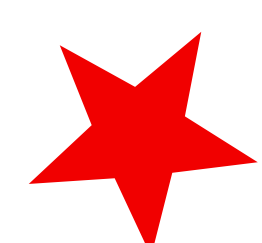
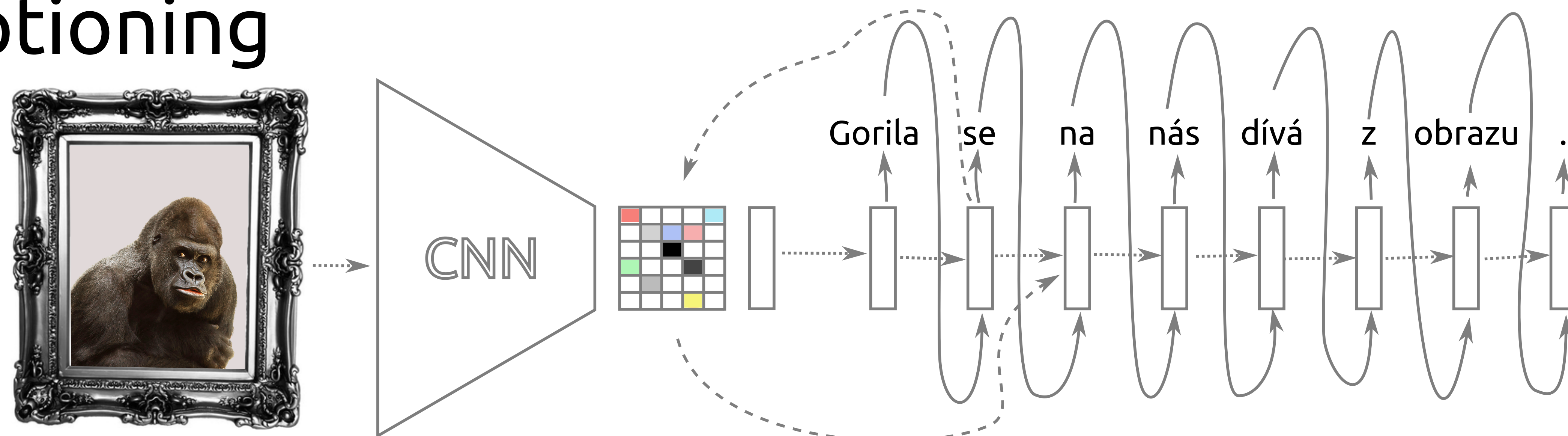
Jindřich Libovický, Rudolf Rosa,  
Jindřich Helcl, and Martin Popel

... with **Neural** Models



Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics

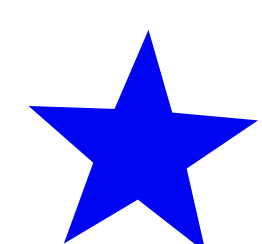
## Image Captioning



Architecture "Show, attend and tell"

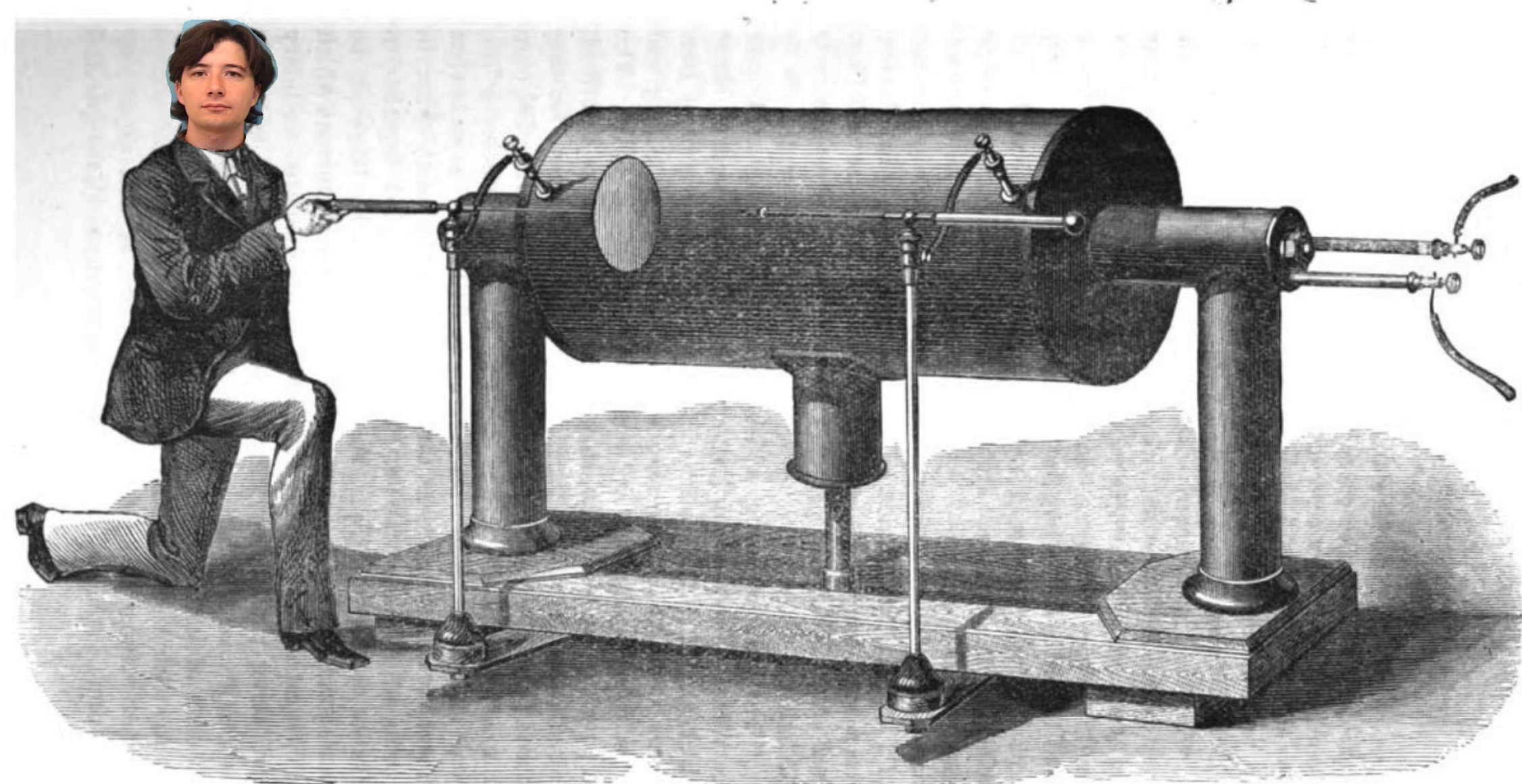


Czech part of the Multi30k dataset for multimodal translation

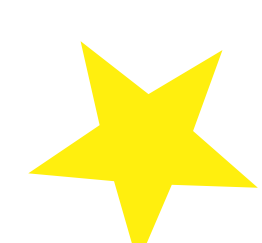


First image captioning model for the Czech language (although the scores are rather poor)

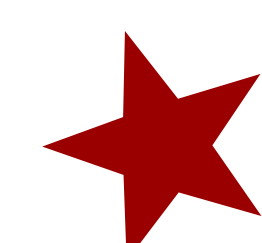
method	BLUE
Xu et al., 2015 EN	19.1
ours EN	19.7
ours CS	2.3



## Machine Translation



Cutting-edge transformer architecture



Trained on CzEng 1.7, a parallel corpus with 57M sentence pairs

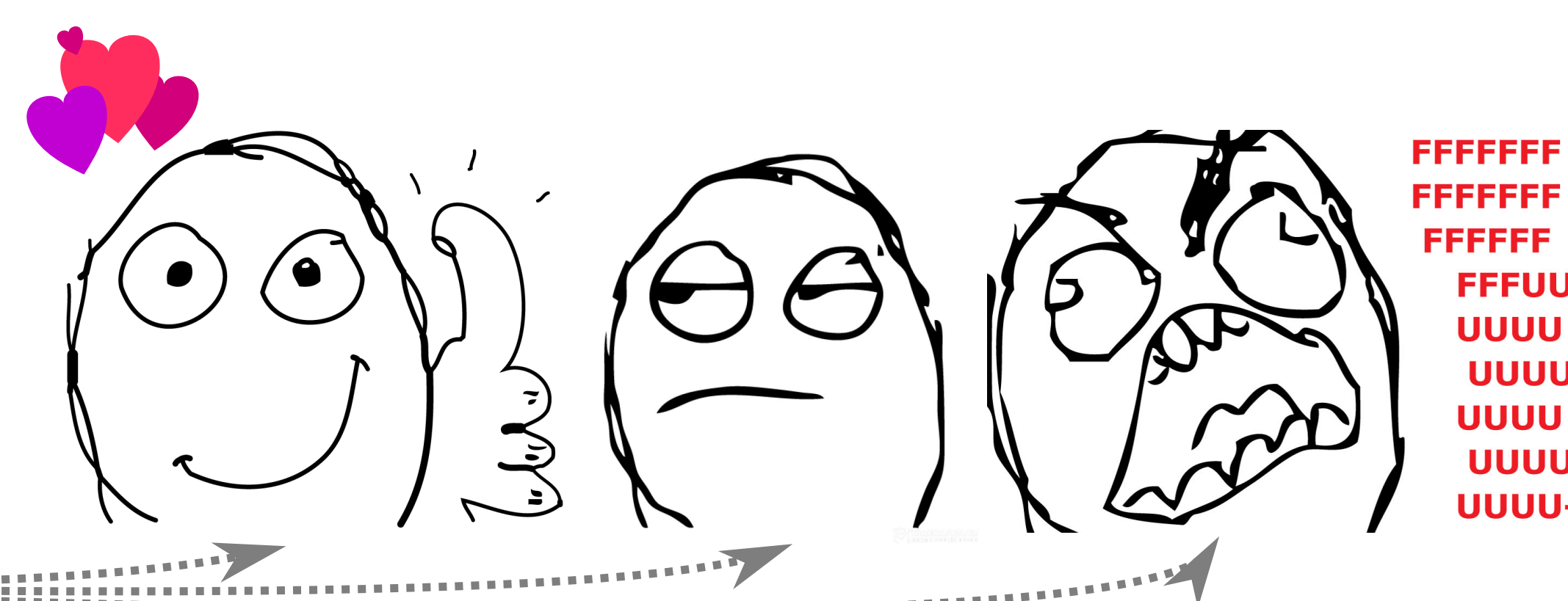
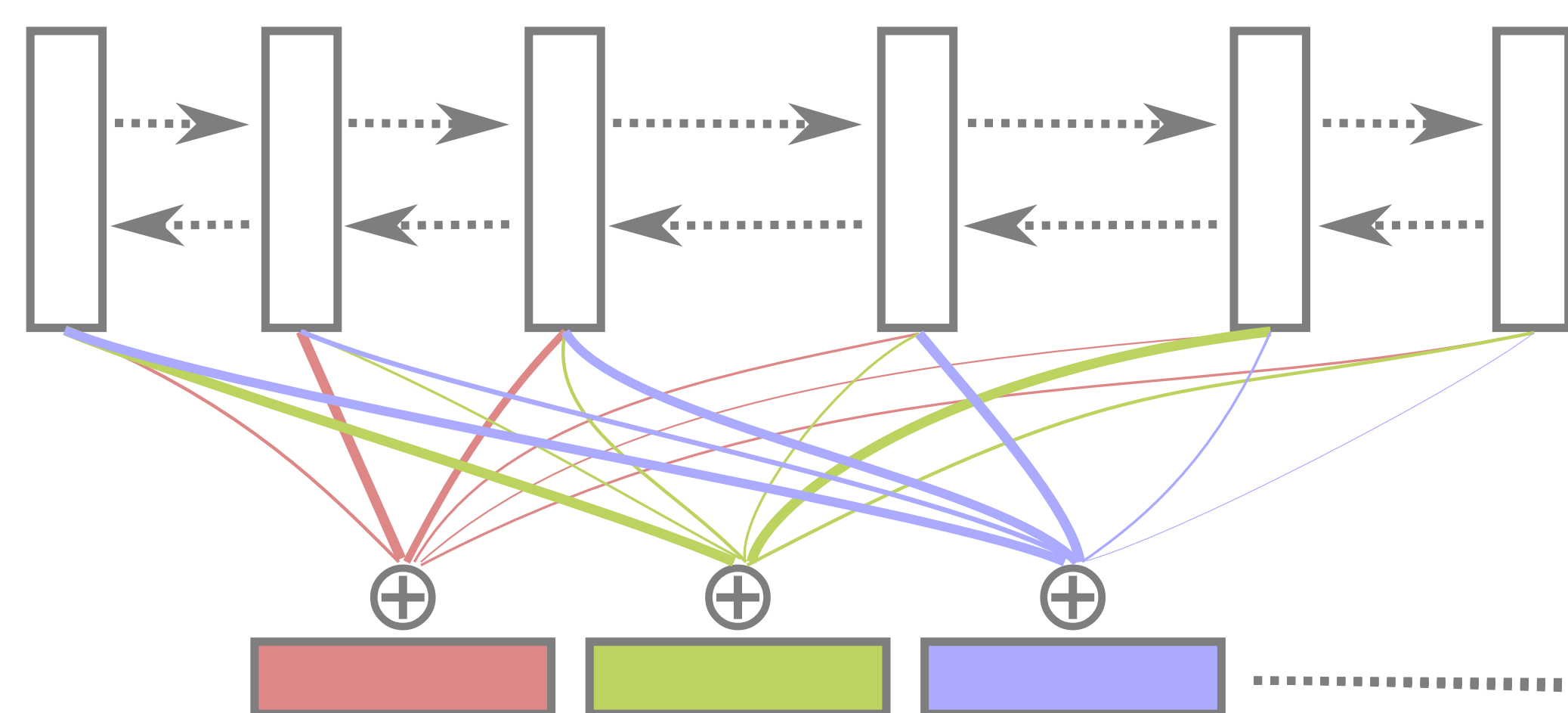


State-of-the-art results on English-to-Czech WMT news translation shared task

method	BLUE
Google Translate	20.8
WMT17 winner	22.8
ours	23.8

## Sentiment Analysis

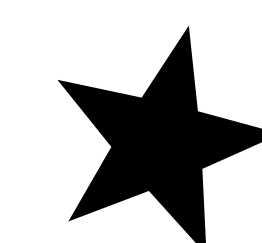
Tohle je moooc pěkná gorilka !



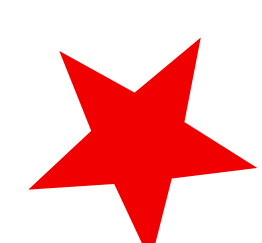
method	accuracy
Lenc et al., 2016	71.0 %
Brychcín and Habernal, 2013	81.5 %
ours	80.8 %



Bidirectional LSTM with self-attentive layers



CSFD CZ dataset, 91k movie reviews from the Czechoslovak film database



Competitive results, though not outperforming the current state of the art



<http://github.com/ufal/neuralmonkey>

<http://quest.ms.mff.cuni.cz/neuralmonkey-czm>