

# Machine Translation (PV061)

Pavel Rychlý

## Optional Assignments

## Optional Assignments

You can earn extra points for the exam.

# Errors in translations

(2-3 point for a group)

- ▶ find errors in machine translation
- ▶ choose any system (Google, DeepL, ...)
- ▶ translate a document
- ▶ report and describe errors
- ▶ try to find shortest sentence containing that error

# Errors in training data

(2-3 point for a group)

- ▶ look at some training data for translation
- ▶ find errors
  - ▶ wrong alignments
  - ▶ wrong language
  - ▶ non-fluent sentences

# Domain evaluation

- ▶ find a NMT system
- ▶ evaluate on standard test set
- ▶ evaluate on domain-specific data

(5 points for a domain)

# Compare small optimizations

(10 points)

- ▶ find a NMT system
- ▶ train with original setup
- ▶ train with a small optimizaton
  - ▶ different tokenizer (BPE/SentencePiece)
  - ▶ shared/separate embeddings
- ▶ compare results

## Replicate SotA results on own data

- ▶ find a NMT implementation
- ▶ run on own data (Cz-En ficet)
- ▶ compare results