

# Paraphrase Generation for Czech

Veronika Burgerová

April 15, 2020

# Overview

- 1 Introduction
- 2 The Watsonson Project
- 3 What's new?

- Paraphrase generation - useful for several NLP applications
  - Machine translation
  - Text summarization
  - ...
- Rule-based and machine learning methods
- Improvement of rule-based paraphrasing system for Czech

# The Watsonson Project

- By Zuzana Nevěřilová
- A game with purpose
- Generates paraphrases and entailments from a given sentence

# The Watsonson Project

## Shenlock Holmer a Dr. Watsonson pátrají společně

Drahý Holmere, jaký případ právě řešíte?



Hra



Úroveň: 0  
Body: 0  
Detektiv: Jméno

Watsonsonův  
mozek  
Paměť



Hypotézy

Dedukční schopnosti



Řekněte Watsonsonovi o novém případě

Řekněte dr. Watsonsonovi, že

- má mluvit dál
- je šikovný
- je nemožný
- mu poradíte
- se vrátíte ke starému případu
- mu popíšete nový případ
- už nemáte čas ani trpělivost

# The Watsonson Project

Pane Holmere, řekl jste: **Drahý Watsonsone, zopakujme si, co o případu víme: Petr stál na zastávce a čekal na autobus. Nudil se a tak si zapálil cigaretu. Autobus přijel s půlhodinovým zpožděním. Petr nastoupil a autobus odjel. Od té doby Petra nikdo neviděl.**

Má Watsonson pravdu, když myslí, že:

- Petr stál na zastávce. ✓ ✗ 😞
- Petr čekal na autobus. ✓ ✗ 😞
- Petr se nudil. ✓ ✗ 😞
- Petr si zapálil cigaretu. ✓ ✗ 😞
- Autobus přijel s půlhodinovým zpožděním. ✓ ✗ 😞
- Petr nastoupil. ✓ ✗ 😞



Pokračovat

# Modules evaluated in past

- synonym replacement
- hypernym replacement
- word reordering
- anaphora resolution
- verb frame inference

# Most common problems

- Incorrect sense in word replacement
- Errors caused by syntactic parser



# What's new? - New module

- New module in progress - Voice transformer
  - Transforming active voice to passive
  - Working for simple sentences so far
  - Generates only one phrase (not always the correct one)
  - Example:
    - Input: "Babička upekla sladkou buchtu. Ten starý muž nehledá ženu."
    - Output: "Babičkou byla upečena sladká buchta. Tím starým mužem není hledána žena."

# Synonym replacement

- Reimplementation of the module for synonym replacement
- Implemented in Python 3
- Generates all possible combinations
- Example
  - Input: Tohle už bohužel bude složitější.
  - Some of the outputs:
    - Tohle už naneštěstí bude složitější
    - Tohle už bohužel bude komplikovanější.
    - Tohle už naneštěstí bude spletitější.
    - ...

## Problem with word sense

- Doing some research about this topic
- Trying bidirectional language model
- Approaches being used:
  - Computing probability of the replaced words in given context
  - Computing probability of the whole sentence
- No results yet

# Future work

- Use the model to get rid of no sense phrases
- Add syntactic parsing to synonym replacer
  - To be able to replace with not grammatically similar words
- Improve the voice transformer
  - Generate all the possible phrases
  - Fix some errors
  - Make it work for more complex sentences
  - Eventually add the passive-to-active part
  - Add more features (reflexive passive, transform only transitive verbs...)