

# Evaluation and Error Analysis of Rule-based Paraphrase Generation for Czech

Veronika Burgerová Aleš Horák

Natural Language Processing Centre  
Faculty of Informatics, Masaryk University  
Botanická 68a, 602 00, Brno, Czech republic  
{xburger,hales}@fi.muni.cz

RASLAN 2019

# Overview

- 1 Introduction
- 2 The Watsonson Project
- 3 Experiments and Evaluation
- 4 Future Work

- **Paraphrase generation** - useful for several NLP applications
  - Machine translation
  - Text summarization
  - ...
- Rule-based and machine learning methods
- **Evaluation** of rule-based paraphrasing system for Czech
- Working with **individual modules**

# The Watsonson Project

- A **game with a purpose** by Zuzana Nevěřilová (CICLING 2014, RASLAN 2013)
- Generates **paraphrases and entailments** from a given sentence

# The Watsonson Project

## Shenlock Holmer a Dr. Watsonson pátrají společně

Drahý Holmere, jaký případ právě řešíte?



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

Hra



Úroveň: 0  
Body: 0  
Detektiv: Jméno

Watsonsonův  
mozek  
Paměť



Hypotézy

Dedukční schopnosti



# 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

- Wordnet
- Wordhyp
- Word order
- Aara
- Verbinfer

# Wordnet and Wordhyp

- **Wordnet** - uses data from the Czech Wordnet for **synonym** replacement
- Recursively - **wordhyp** module for **hypernym** replacement
- No word sense identification

## Example

Input: Pejscí si chtěli hrát s **děťmi**, ale žádné děti venku nenašli.  
(Dogs wanted to play with **children** but they found no children outside.)

Wordnet: Pejscí si chtěli hrát s **děčky**.  
(Dogs wanted to play with **kids**.)

Wordhyp: Pejscí si chtěli hrát s **lidmi**.  
(Dogs wanted to play with **people**.)



- **Reorders** the sentence **constituents**
- Generates phrases with **all possible orders**

## Example

Input: Pejsci si chtěli hrát s dětmi, ale žádné děti venku nenašli.  
(Dogs wanted to play with children but they found no children outside.)

Wordorder: Nenašli **oni** žádné děti venku.  
(**They** did not find any children outside.)

- Partial **anaphora resolution**
- Replaces **pronomial objects** or subjects by their **antecedents**
- Resolves **zero subjects**

## Example

Input: Pejsci si chtěli hrát s dětmi, ale žádné děti venku nenašli.

(Dogs wanted to play with children but **they** found no children outside.)

Aara: **Pejsci** nenašli žádné děti venku.

(**Dogs** found no children outside.)

- Using Czech verb valency lexicon [VerbaLex](#)
- **Verb frame inference** of 3 types:
  - equality
  - effect
  - precondition
- Can result in new facts

## Example

Input: Pejscí si chtěli hrát s dětmi, ale žádné děti venku **nenašli**.  
(Dogs wanted to play with children but they **found** no children outside.)

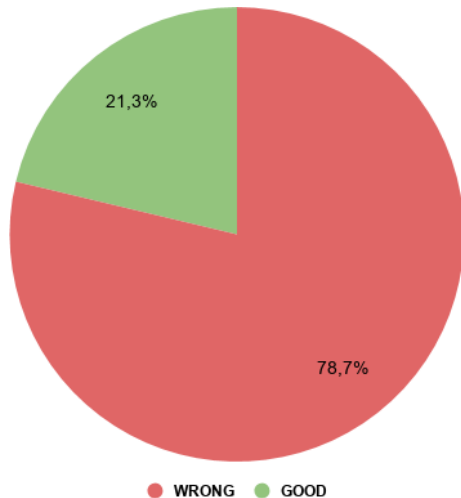
Verbinfer: **Neobjevili** žádné děti venku.  
(They **discovered** no children outside.)

# Input dataset

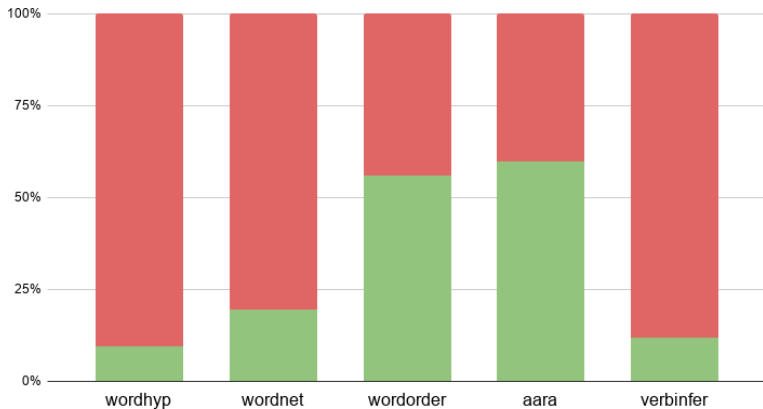
- 40 sentences of various length
- Simple made up phrases or extracted from Czech children tales
- Input examples:
  - *Máma upekla koláč.*
  - *Pejsek s kočičkou se měli moc rádi a bydleli spolu už spoustu let.*
  - *Do paneláku na sídlišti se přistěhoval nový kluk Jakub.*
- Processed by morphological analyzer/generator, morphological tagger and syntactic parser
- Stored in JSON format - input to each evaluated module

# Evaluation

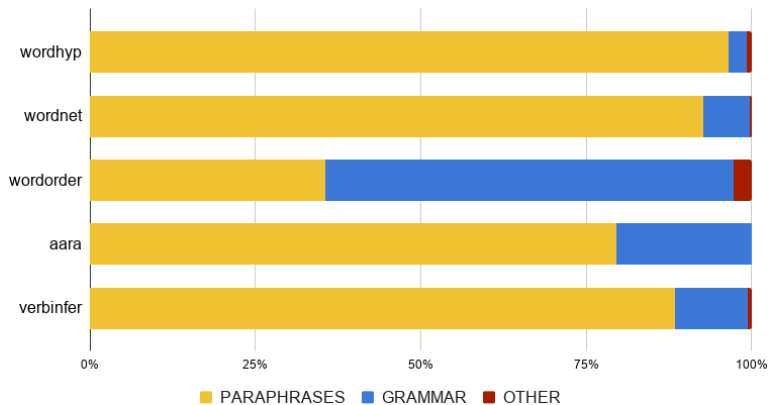
<b>Module</b>	<b>Paraphrases</b>
wordhyp	557
wordnet	488
wordorder	166
aara	97
verbinfer	206
<b>All</b>	<b>1,514</b>



# Evaluation - Individual Scores



# Error Analysis



- Most of the incorrect paraphrases - word replacement
- Repeated across all modules:
  - 1 Prepositions

## Example

- Foxlík slyšel někom nedobřím. (Foxlík heard someone bad.)
- Krteček přespával hluboko na zemi. (The mole slept deep on the ground.)

## 2 Dependencies

## Example

Original meaning: Drak má sedm hlav. (Dragon has seven heads.)  
Paraphrase: Ten pejsek sedm hlav má. (The dog has seven heads.)

- Primary source - syntactic parsing phase



# Future Work

- Improvement of modules and preprocessing
- New types of paraphrasing methods