# Automatic correction of errors in machine translation outputs

Official title: Automatic post-editing of phrase-based machine translation outputs

Rudolf Rosa, ÚFAL MFF UK (Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University in Prague)

**Can improve even Google Translate!**

## Step 0: Input

- an English sentence
- its translation into Czech, produced by a machine translation system (such as Moses or Google Translate)

## Step 2: Error correction

- 28 correction rules (7 shown here) + a statistical model of valency
- a rule changes the morphological categories of a word (number, case...), and then the word form gets regenerated by a morphological generator

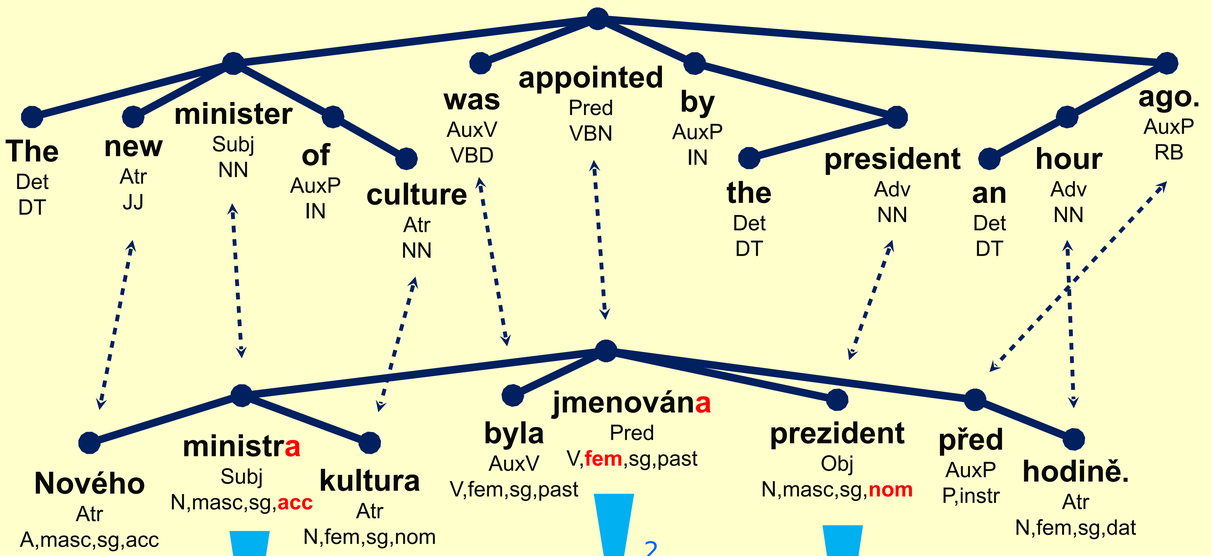
1. **Subject Case**
   subject must be in the nominative (1st case)
2. **Subject – Verb Agreement**
   verb must agree with subject in gender, number, and person
3. **Translation of "by"**
   passive actor must be in the instrumentative (7th case)
4. **Noun – Adjective Agreement**
   adjective must agree with noun in number, gender, and case
5. **Translation of "of"**
   possessor must be in the genitive (2nd case)
6. **Passive – "be" Agreement**
   "be" must agree with passive in gender and number
7. **Preposition – Noun Agreement**
   noun must agree with preposition in case

0. *Input*

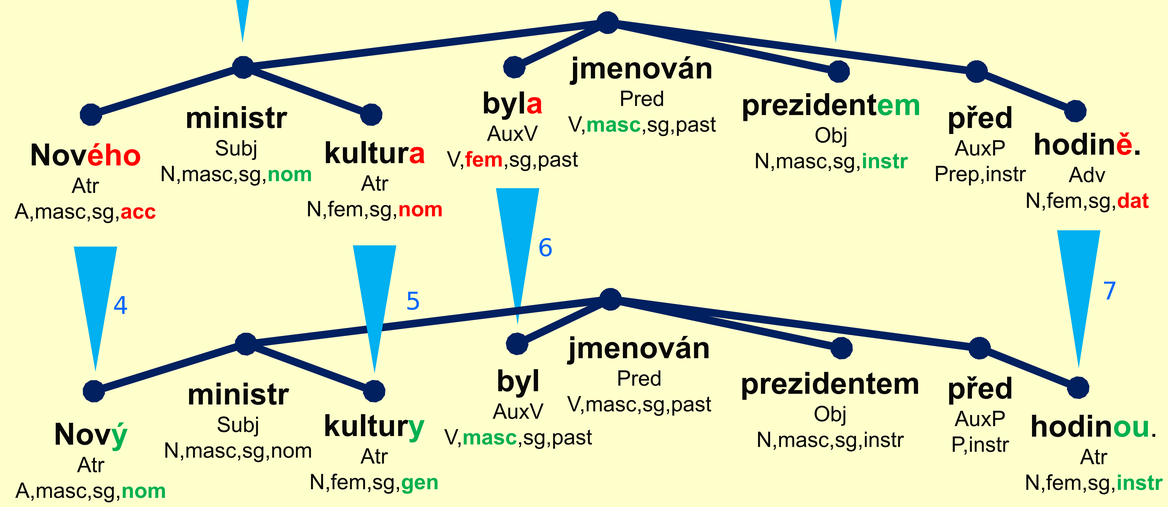
**The new minister of culture was appointed by the president an hour ago.**

**Nového ministra kultura byla jmenována prezident před hodině.**

1. *Analysis of input sentences*




2. *Automatic correction of errors*




*Output*

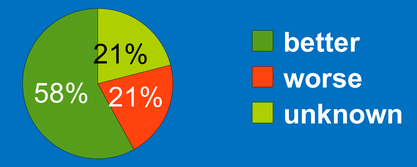
**Nový ministr kultury byl jmenován prezidentem před hodinou.**

## Step 1: Analysis

- morphological tagging
  - part of speech: Noun, Adjective...
  - gender: masculine, feminine...
  - number: singular, plural
  - case: nominative, genitive...
  - …
- word alignment
  - "new" ↔ "nového"
  - "minister" ↔ "ministra"
  - …
- parsing to dependency trees
  - head-modifier structure
  - "new" modifies "minister"...
- assignment of function labels
  - Predicate, Subject, Attribute..

## Evaluation

- manual: on changed sentences




- automatic: standard BLEU metric