Error Correction of PB SMT Outputs with automatic post-editing shown on English to Czech translation
Motivation: Translation of negation
Motivation: Errors in negation

- These are not actually errors.
Motivation: Errors in negation

- These are not actually errors.
  - Moses: Jsou to vlastně chyby.
  - Gloss: These are actually errors.
Motivation: Errors in negation

- These are not actually errors.
  - Moses: Jsou to vlastně chyby.
  - Gloss: These are actually errors.
  - Ref.: Nejsou to vlastně chyby.
Motivation: Errors in negation

- These **are not** actually errors.
  - Moses: **Jsou** to vlastně chyby.
  - Gloss: **These are** actually errors.
  - Ref.: **Nejsou** to vlastně chyby.
- I **would not cheat** on you.
Motivation: Errors in negation

- These are not actually errors.
  - Moses: Jsou to vlastně chyby.
  - Gloss: These are actually errors.
  - Ref.: Nejsou to vlastně chyby.

- I would not cheat on you.
  - Moses: Já bych tě podváděl.
  - Gloss: I would cheat on you.
  - Ref.: Já bych tě nepodváděl.
Motivation: Errors in negation

- These are not actually errors.
  - Moses: Jsou to vlastně chyby.
  - Gloss: These are actually errors.
  - Ref.: Nejsou to vlastně chyby.

- I would not cheat on you.
  - Moses: Já bych tě podváděl.
  - Gloss: I would cheat on you.
  - Ref.: Já bych tě nepodváděl.

- Some phenomena hard to get right with PBSMT
Motivation: Errors in negation

- These **are not** actually errors.
  - Moses: **Jsou** to vlastně chyby.
  - Gloss: These **are** actually errors.
  - Ref.: **Nejsou** to vlastně chyby.
- I **would not cheat** on you.
  - Moses: Já **bych** tě podváděl.
  - Gloss: I **would cheat** on you.
  - Ref.: Já **bych** tě nepodváděl.
- Some phenomena hard to get right with PBSMT
- Even BLEU doesn't complain much here
Simple to fix?

- there is a negation in the source
  - These are not actually errors
Simple to fix?

- there is a negation in the source
  - These are not actually errors
- there is no negation in the target
  - Jsou to vlastně chyby
Simple to fix?

- there is a negation in the source
  - These are **not** actually errors
- there is no negation in the target
  - Jsou to vlastně chyby
Simple to fix?

- there is a negation in the source
  
  - These are not actually errors

- there is no negation in the target
  
  - Jsou to vlastně chyby

- add the negative prefix (ne-) into the target
  
  - Nejsou to vlastně chyby
Simple to fix?

- there is a negation in the source
  - These are **not** actually errors
- there is no negation in the target
  - Jsou to vlastně chyby
- add the negative prefix (ne-) into the target
  - Nejsou to vlastně chyby
- such a simple approach might be sufficient
- but usually useful to use some NLP tools
  - to decide where to put the negative prefix (ne-), to reliably detect the negativeness...
Part-of-speech tagger

- run a POS tagger on the target sentence
  - Jsou to vlastně chyby
  - verb pronoun adverb noun
- a good heuristic: negate the (finite) verb!
  - Nejsou to vlastně chyby
  - verb pronoun adverb noun
- fine-grained tags may even mark the negation
  - jsou VB-P---3P-\text{AA}---
  - nejsou VB-P---3P-\text{NA}---
Dependency parser

- parse both source and target
- project negation through word alignment
dependency parser

- parse both source and target
- project negation through word alignment
Deep syntactic analysis

- auxiliary nodes collapsed into values of attributes on parent nodes
  
  - are
  
  
  - not
  
  - be, neg=1

- abstract from various ways of expressing negation (not, no, un-, in-, ...)
  
  - all marked by neg=1 on the lexical node
Morphological generator

- `form = generate(word, morphological features)`
Morphological generator

- `form = generate(lemma, tag)`
Morphological generator

- \texttt{form = generate(lemma, tag)}
- \textbf{instead of:} \texttt{new\_form = 'ne' + form}
  - \texttt{'nejsou' = 'ne' + 'jsou'}
Morphological generator

- `form = generate(lemma, tag)`

- **instead of:** `new_form = 'ne' + form`
  - `'nejsou' = 'ne' + 'jsou'`

- **use the more sophisticated:** `new_form = generate(lemma(form), negate(tag))`
  - `'nejsou' = generate(lemma('jsou'), negate('VB-P---3P-AA---'))`
once you have set up the pipeline of NLP tools, use it for various error corrections, only adding a few lines of code for each error type:
once you have set up the pipeline of NLP tools, use it for various error corrections, only adding a few lines of code for each error type:

- morphological agreement
- čarodějnic (witches) pl, fem, genitive
- zelených (green) pl, fem, genitive

once you have set up the pipeline of NLP tools, use it for various error corrections, only adding a few lines of code for each error type:

- morphological agreement
- possessives
- verb tenses
- negations
- passives
- ...
Summary

- some things are hard for PB SMT
  - negation, agreement, passives...
- automatic post-editing can often help
  - a handful of simple rules might do the job very well
- NLP tools are typically useful
  - tagger, parser, morphological generator
  - adapt them to SMT outputs if possible
- Rudolf Rosa (2013): Automatic post-editing of phrase-based machine translation outputs