Delexicalized Cross-lingual Transfer of Statistical Syntactic Parsers for Automatic Analysis of Low-resourced Natural Languages
Outline

- Introduction to linguistic analysis
- MSTParser and its delexicalization
- Single-source delexicalized parser transfer
  - $KL_{cpos3}$ language similarity
- Multi-source delexicalized parser transfer
  - treebank concatenation
  - parse tree combination
  - model interpolation
- Future work: lexicalization
Outline

- Introduction to linguistic analysis
- MSTParser and its delexicalization
- Single-source delexicalized parser transfer
  - $KL_{cpos3}$ language similarity
- Multi-source delexicalized parser transfer
  - treebank concatenation
  - parse tree combination
  - model interpolation
- Future work: lexicalization
Introduction to linguistic analysis

- The boy likes travelling by train very much.
Tokenization

- The boy likes travelling by train very much.
Part-of-speech tagging

- The boy likes travelling by train very much.

The boy likes travelling by train very much.
Part-of-speech tagging

- The boy likes travelling by train very much.

```
The DETERMINER boy likes travelling by train
very much .
```
Part-of-speech tagging

- The boy likes travelling by train very much.

- The boy likes travelling by train very much.

- The boy likes travelling by train very much.
Part-of-speech tagging

- The boy likes travelling by train very much.
Part-of-speech tagging

- The boy likes travelling by train very much.

The DETERMINER
boy NOUN
likes VERB
travelling NOUN
by

train
very
much
.
Part-of-speech tagging

- The boy likes travelling by train very much.

The DETERMINER - boy NOUN - likes VERB - travelling NOUN - by PREPOSITION

train - very - much - .
Part-of-speech tagging

- The boy likes travelling by train very much.
## Part-of-speech tagging

- The boy likes travelling by train very much.

<table>
<thead>
<tr>
<th>The</th>
<th>boy</th>
<th>likes</th>
<th>travelling</th>
<th>by</th>
</tr>
</thead>
<tbody>
<tr>
<td>DETERMINER</td>
<td>NOUN</td>
<td>VERB</td>
<td>NOUN</td>
<td>PREPOSITION</td>
</tr>
<tr>
<td>train</td>
<td>very</td>
<td>much</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOUN</td>
<td>ADVERB</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Part-of-speech tagging

- The boy likes travelling by train very much.

```
The  DETERMINER
boy  NOUN
likes VERB
travelling NOUN
by  PREPOSITION
train NOUN
very ADVERB
much ADVERB
.  
```
Part-of-speech tagging

- The boy likes travelling by train very much.

The DETERMINER
boy NOUN
likes VERB
travelling NOUN
by PREPOSITION
train NOUN
very ADVERB
much ADVERB
. PUNCTUATION
Syntactic parsing

The DET boy NOUN likes VERB travelling NOUN by PREP train NOUN very ADV much ADV . PUNCT
Syntactic parsing

The DET boy NOUN likes VERB travelling NOUN by PREP train NOUN very ADV much ADV
Syntactic parsing

The boy likes travelling by train very much.
Syntactic parsing

The DET boy NOUN likes VERB travelling NOUN by PREP train NOUN very ADV much ADV.

#root PUNCT
Syntactic parsing

The boy likes travelling by train very much.

#root

likes VERB

boy NOUN

The DET

. PUNCT

travelling NOUN  by PREP  train NOUN  very ADV  much ADV
Syntactic parsing

The boy likes travelling by train very much.

- The DET
- boy NOUN
- likes VERB
- travelling NOUN
- by PREP
- train NOUN
- very ADV
- much ADV
The boy likes travelling by train. By very much.
The boy likes travelling by train. Very much.
Syntactic parsing

#root

likes VERB

boy NOUN

The DET

by PREP

train NOUN

much ADV

very ADV

by train

very adv

by

much

train

likes

boy

The

#root
The boy likes travelling by train very much.
Outline

- Introduction to linguistic analysis
- **MSTParser** and its delexicalization
- Single-source delexicalized parser transfer
  - $\text{KL}_{\text{cpos3}}$ language similarity
- Multi-source delexicalized parser transfer
  - treebank concatenation
  - parse tree combination
  - model interpolation
- Future work: lexicalization
Maximum Spanning Tree Parser

#root
Rudolf NOUN
likes VERB
trains NOUN
Maximum Spanning Tree Parser

Rudolf Rosa – Cross-lingual Transfer of Syntactic Parsers for Analysis of Low-resourced Natural Languages

Rudolf NOUN likes VERB trains NOUN

#root
Maximum Spanning Tree Parser

- weighting model trained on annotated data
- features – for edge nodes and their neighbours
  - lexical: word form (“likes”), word lemma (“like”)
  - morphological: part-of-speech tag (“VERB”)
  - signed distance of nodes (#root→likes: “+2”)
Maximum Spanning Tree Parser

- Chu-Liu-Edmonds MST algorithm
Outline

- Introduction to linguistic analysis
- MSTParser and its delexicalization
- Single-source delexicalized parser transfer
  - $KL_{cpos3}$ language similarity
- Multi-source delexicalized parser transfer
  - treebank concatenation
  - parse tree combination
  - model interpolation
- Future work: lexicalization
Semi-supervised parsing

- fully supervised dependency parsing
  - requires training data (treebank) or a grammar
Semi-supervised parsing

- fully supervised dependency parsing
  - requires training data (treebank) or a grammar
  - there are ~100 treebanks (manually annotated)
  - there are ~7,000 languages
  - + various domains, language evolution...
Semi-supervised parsing

- fully supervised dependency parsing
  - requires training data (treebank) or a grammar
  - there are ~100 treebanks (manually annotated)
  - there are ~7,000 languages
  - + various domains, language evolution...

- semi-supervised parsing
  - utilize existing resources, avoid new annotations
    - treebanks for other languages (HamleDT: 30 languages)
    - unannotated data (here: part-of-speech tagged)
Lexicalized MSTParser

Rudolf Rosa – Cross-lingual Transfer of Syntactic Parsers for Analysis of Low-resourced Natural Languages
Delexicalized MSTParser

#root

NOUN

VERB

NOUN

7.4 5.2 16.1 -4.3 -1.4 9.7

10.8 2.2 -2.4

16.1

#root

NOUN

VERB

NOUN

10.8 9.7
Outline

- Introduction to linguistic analysis
- MSTParser and its delexicalization
- Single-source delexicalized parser transfer
  - $KL_{cpos3}$ language similarity
- Multi-source delexicalized parser transfer
  - treebank concatenation
  - parse tree combination
  - model interpolation
- Future work: lexicalization
Single-source delex parser transfer

- (Zeman and Resnik, 2008)
- train a delexicalized parser on a source language treebank (e.g. Czech)
- apply it to a target language, without a treebank but with a POS tagger (e.g. Slovak)
Utilizing multiple treebanks

- HamleDT: 30 harmonized treebanks
- How do we choose the source treebank?
- Can we use more/all source treebanks?
Choosing the source treebank

- src should be as similar to tgt as possible
  - World Atlas of Language Structures (WALS)
    - language family, word order properties...
Choosing the source treebank

- src should be as similar to tgt as possible
  - World Atlas of Language Structures (WALS)
    - language family, word order properties...
  - $KL_{cpos^3}(tgt, src)$: Kullback-Leibler divergence of POS trigram distributions

![Graph showing POS trigram distributions for different languages]
\( KL_{cpos^3}(tgt, src) = \sum_{\forall cpos^3 \in tgt} f_{tgt}(cpos^3) \cdot \log \left( \frac{f_{tgt}(cpos^3)}{f_{src}(cpos^3)} \right) \)

\( cpos^3 = \langle cpos_{i-1}, cpos_i, cpos_{i+1} \rangle \)

\( f(cpos^3) = \frac{\text{count}(cpos^3)}{|corpus|} \)
## Sample of results (HamleDT)

<table>
<thead>
<tr>
<th>Target lang.</th>
<th>KL_{cpos 3} selected src lang.</th>
<th>UAS</th>
<th>Oracle (best possible src)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bengali</td>
<td>Telugu</td>
<td>66.7</td>
<td>✓</td>
</tr>
<tr>
<td>Czech</td>
<td>Slovak</td>
<td>65.8</td>
<td>✓</td>
</tr>
<tr>
<td>Danish</td>
<td>Slovenian</td>
<td>42.1</td>
<td>+13.3 English</td>
</tr>
<tr>
<td>German</td>
<td>English</td>
<td>56.8</td>
<td>✓</td>
</tr>
<tr>
<td>Slovak</td>
<td>Slovenian</td>
<td>58.4</td>
<td>+ 3.3 Czech</td>
</tr>
<tr>
<td>Tamil</td>
<td>Turkish</td>
<td>31.1</td>
<td>+22.4 Hindi</td>
</tr>
</tbody>
</table>
Outline

- Introduction to linguistic analysis
- MSTParser and its delexicalization
- Single-source delexicalized parser transfer
  - $KL_{\text{cpos3}}$ language similarity
- Multi-source delexicalized parser transfer
  - treebank concatenation
  - parse tree combination
  - model interpolation
- Future work: lexicalization
Multi-source delex parser transfer

- treebank concatenation
  - concatenate all source treebanks
  - train one delexicalized parser on the multi-treebank
  - apply the parser to the target text
Multi-source delex parser transfer

- treebank concatenation (baseline)
  - train a parser on concatenation of all treebanks

- parse tree combination
  - train a separate parser for each source treebank
  - separately apply each parser to target text
  - use parser voting and MST algorithm to find the final analysis
Parse tree combination

src 1: #root VERB PREP NOUN

src 2: #root VERB PREP NOUN

src 3: #root VERB PREP NOUN

tgt: #root VERB PREP NOUN
Parse tree combination

src 1:

#root
VERB
PREP
NOUN

src 2:

#root
VERB
PREP
NOUN

src 3:

#root
VERB
PREP
NOUN

tgt:

#root
VERB
PREP
NOUN
Parse tree combination

src 1:

+ src 2:

+ src 3:

= tgt:
Weighted parse tree combination

\[ KL_{\text{cpos3}}^{-4} : \]
\[ x 1.9 \]

\[ x 1.7 \]

\[ x 0.5 \]

\[ = \text{tgt:} \]
\[ #\text{root} \]
\[ \text{VERB} \]
\[ \text{PREP} \]
\[ \text{NOUN} \]

\[ 4.1 \]
\[ 2.4 \]
\[ 2.2 \]
\[ 1.9 \]
\[ 1.7 \]
\[ 1.7 \]
Weighted parse tree combination

\[ KL_{cpos3}^{-4} : \]

src 1:
- #root
- VERB
- PREP
- NOUN

× 1.9

+ src 2:
- #root
- VERB
- PREP
- NOUN

× 1.7

+ src 3:
- #root
- VERB
- PREP
- NOUN

× 0.5

= tgt:
- #root
- VERB
- PREP
- NOUN

4.1

2.4

2.2

1.9

1.7

Rudolf Rosa – Cross-lingual Transfer of Syntactic Parsers for Analysis of Low-resourced Natural Languages
Multi-source delex parser transfer

- treebank concatenation (baseline)
  - train a parser on concatenation of all treebanks
- parse tree combination
  - combine separate src parsers via voting and MST
- parser model interpolation
  - train a parser for each source treebank
  - interpolate the trained models into a combined model
  - apply the parser with the combined model to the target text
**Parser model interpolation**

- **motivation:** maybe the parser is more sure with some edges than other?
- **the score assigned to the edge might show that**
  - MSTParser before running the MST algorithm:
Parser model interpolation

- motivation: maybe the parser is more sure with some edges than other?
- the score assigned to the edge **might** show that
  - MSTParser before running the MST algorithm:
Parser model model interpolation

- score normalization!
Parser model interpolation

\[ \sum = \text{tgt} (\Sigma): \]

- root: 29.9
- VERB: 17.7
- PREP: -1.1
- NOUN: 25.4

Diagram:

- #root to VERB: 2.7
- VERB to PREP: 16.8
- PREP to NOUN: 7.8
- NOUN to #root: 3.5
- VERB to #root: 12.9
Parser model interpolation

tgt:

#root → VERB

VERB → PREP

PREP → NOUN

#root: 29.9 → VERB: 17.7 → PREP: 25.4 → NOUN

2.7 → PREP

12.9 → NOUN

7.8 → NOUN

16.8 → VERB

3.5 → VERB

-1.1
- multiply each edge score with $KL_{cpos^3}^{-4}(tgt, src)$

```
src1:
  #root -> VERB: 17.4
  VERB -> PREP: 14.3
  PREP -> NOUN: 9.7
```

$KL_{cpos^3}^{-4}(tgt, src1) = 0.5$
Weighted parser model interpol.

- multiply each edge score with $KL_{cpos3}^{-4}(tgt, src)$

```
src1: #root 8.7 \rightarrow VERB 7.1 \rightarrow PREP 4.8 \rightarrow NOUN

2.6

3.1

-0.7

5.4

1.1

-1.2

KL_{cpos3}^{-4}(tgt, src1) = 0.5
```
Why “model interpolation”?

- MSTParser edge score = \( w \cdot f \)
Why “model interpolation”?

- MSTParser edge score = $w \cdot f$
- unweighted model interpolation
  - edge score = $\sum_{src} (w_{src} \cdot f)$
Why “model interpolation”? 

- MSTParser edge score = \( w \cdot f \)
- unweighted model interpolation
  - edge score = \( \sum_{src} (w_{src} \cdot f) = (\sum_{src} w_{src}) \cdot f \)
Why “model interpolation”? 

- MSTParser edge score = $w \cdot f$
- unweighted model interpolation
  - edge score = $\sum_{src} (w_{src} \cdot f) = (\sum_{src} w_{src}) \cdot f$
  - interpolated model $w_{int} = (\sum_{src} w_{src})$
  - edge score = $w_{int} \cdot f$
Why “model interpolation”? 

- MSTParser edge score = $w \cdot f$
- unweighted model interpolation
  - edge score = $\sum_{src} (w_{src} \cdot f) = (\sum_{src} w_{src}) \cdot f$
  - interpolated model $w_{int} = (\sum_{src} w_{src})$
  - edge score = $w_{int} \cdot f$
- weighted model interpolation: $KL_{cpos3}^{-4}(tgt, src)$
  - edge score = $\sum_{src} (KL_{src} \cdot w_{src} \cdot f) = (\sum_{src} KL_{src} \cdot w_{src}) \cdot f$
  - interpolated model $w_{int} = (\sum_{src} KL_{src} \cdot w_{src})$
Average UAS over 18 test TBs

- **Baseline:** Treebank concatenation
  - UAS: 44.5%

- **Single-source selection**
  - UAS: 48.6%

- **Unweighted combination**
  - UAS: 48.0%

- **Weighted combination**
  - UAS: 52.5%

- **Unweighted interpolation**
  - UAS: 45.6%

- **Weighted interpolation**
  - UAS: 52.8%

- **Upper bound: Oracle src**
  - UAS: 55.9%
Outline

- Introduction to linguistic analysis
- MSTParser and its delexicalization
- Single-source delexicalized parser transfer
  - KL_{cpos3} language similarity
- Multi-source delexicalized parser transfer
  - treebank concatenation
  - parse tree combination
  - model interpolation
- Future work: lexicalization
Future work: lexicalization

- lexical features (words) are important
  - supervised parsers UAS: ~85% lex, ~75% delex
  - but they are language-specific (unlike POS tags)
Machine translation lexicalization

- parse tree transfer
  - translate target sentence to source language
  - parse translated sentence by source parser
  - transfer source parse tree to target sentence

- treebank transfer
  - translate source treebank to target language
  - transfer parse trees from source treebank to translated treebank
  - train target parser on transferred treebank
Machine translation lexicalization

- parse tree transfer
  - transfer translated source parse tree to target sentence
- treebank transfer
  - train target parser on transferred translated source TB
Problem 1: translation

- parse tree transfer
  - transfer translated source parse tree to target sentence
- treebank transfer
  - train target parser on transferred translated source TB
- problems with machine translation
  - high-quality often available only to/from English
  - for low-resourced languages often low quality
    - requires large amounts of bilingual texts
Problem 2: transfer

- parse tree transfer
  - transfer translated source parse tree to target sentence
- treebank transfer
  - train target parser on transferred translated source TB
Solution (to both): morphs?

- morphs could get closer to 1:1 correspondence
  - especially if segmentation and alignment done jointly
- translation via morphs could do with less data
  - split rare complex words into frequent simple morphs
Joint segmentation and alignment?

- given a corpus of bilingual sentences
- morph segmentation and alignment, so that
  - alignment is close to 1:1
  - aligned morphs have similar meaning
- Bayesian approach?
  - maximize probability of bilingual morphs
    - \( P(\text{-em : by}) \sim \text{count}(\text{-em : by}) \)
  - also account for alignment fertility, alignment fluency, lexical roots vs auxiliary morphs distinction
Conclusion

- Parsing of low-resourced natural languages
- Single-/Multi-source delexicalized parser transfer
  - parse tree combination
  - MSTParser model interpolation
  - $KL_{cpos3}$: language similarity for src selection/weighting
- Future work: Lexicalization
  - machine translation
  - morph splitting and alignment
Thank you for your attention

Rudolf Rosa, Zdeněk Žabokrtský
{rosa,zabokrtsky}@ufal.mff.cuni.cz

Delexicalized Cross-lingual Transfer of Statistical Syntactic Parsers for Automatic Analysis of Low-resourced Natural Languages

Charles University in Prague
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics

http://ufal.mff.cuni.cz/rudolf-rosa/