Lexicalized vs. Delexicalized Parsing in Low-Resource Scenarios

Agnieszka Falenska    Özlem Çetinoğlu

Institute for Natural Language Processing

Universität Stuttgart
Table of Contents

Motivation

Methodology

Experiments and Analysis
- **EXTERNALPOS** setting
- **TREEBANKPOS** setting

Application to Test Languages

Summary

Backup slides
Origins

- It all started with the CoNLL-ST. (Zeman et al., 2017)
- Particularly interesting: low-resource scenarios.
Low-resource scenarios of the CoNLL-ST

- Languages with very small training treebanks.

<table>
<thead>
<tr>
<th>Surprise languages</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>bxr Buryat</td>
<td>19 sent.</td>
</tr>
<tr>
<td>kmr Kurmanji</td>
<td>20 sent.</td>
</tr>
<tr>
<td>sme North Sami</td>
<td>20 sent.</td>
</tr>
<tr>
<td>hsb Upper Sorbian</td>
<td>23 sent.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Small languages</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>kk Kazakh</td>
<td>31 sent.</td>
</tr>
<tr>
<td>ug Uyghur</td>
<td>100 sent.</td>
</tr>
<tr>
<td>ga Irish</td>
<td>566 sent.</td>
</tr>
<tr>
<td>uk Ukrainian</td>
<td>863 sent.</td>
</tr>
<tr>
<td>la Latin</td>
<td>1334 sent.</td>
</tr>
</tbody>
</table>
Why are they interesting?

It is not obvious how to approach them.

- When you have a big treebank — train a parser.
- When there is no treebank — apply cross-lingual methods.

What if you have a treebank but it is very small?
Why are they interesting?

It is not obvious how to approach them.

▶ When you have a big treebank – train a parser.
▶ When there is no treebank – apply cross-lingual methods.

What if you have a treebank but it is very small?

Source: English  
Target: Polish
Research questions

Given a language with a small treebank:
- monolingual or cross-lingual method?
- how to decide?
  - depending on the treebank size?
  - depending on the POS accuracy?
  - depending on the language itself?

Given a new language without a treebank:
- can we achieve useful results with cross-lingual methods …
- … or the only way to parse it is to annotate new trees?
Research questions

Given a language with a small treebank:

► monolingual or cross-lingual method?
► how to decide?
  ▶ depending on the treebank size?
  ▶ depending on the POS accuracy?
  ▶ depending on the language itself?

Given a new language without a treebank:

► can we achieve useful results with cross-lingual methods …
► … or the only way to parse it is to annotate new trees?
Research questions

Given a language with a small treebank:

▶ monolingual or cross-lingual method?
▶ how to decide?
  ▶ depending on the treebank size?
  ▶ depending on the POS accuracy?
  ▶ depending on the language itself?

Given a new language without a treebank:

▶ can we achieve useful results with cross-lingual methods …
▶ … or the only way to parse it is to annotate new trees?
Table of Contents

Motivation

Methodology

Experiments and Analysis
  \textsc{ExternalPOS} setting
  \textsc{TreebankPOS} setting

Application to Test Languages

Summary

Backup slides
## Low-resource scenarios of the CoNLL-ST

<table>
<thead>
<tr>
<th></th>
<th>Surprise languages</th>
<th>Small languages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training treebank</strong></td>
<td>very small</td>
<td>small</td>
</tr>
<tr>
<td><strong>Parallel data</strong></td>
<td>no</td>
<td>no/small</td>
</tr>
</tbody>
</table>

POS gold data:  
- **external**\(^1\) for Surprise languages
- **treebank** for Small languages

POS accuracy:  
- **good** for Surprise languages
- **varies** for Small languages

\(^1\)simulated for the CoNLL-ST
### Low-resource scenarios of the CoNLL-ST

<table>
<thead>
<tr>
<th></th>
<th>Surprise languages</th>
<th>Small languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training treebank</td>
<td>very small</td>
<td>small</td>
</tr>
<tr>
<td>Parallel data</td>
<td>no</td>
<td>no/small</td>
</tr>
</tbody>
</table>

| POS gold data:           | external\(^1\)     | treebank        |
| POS accuracy:            | good               | varies          |

\(^1\)simulated for the CoNLL-ST
Generalized low-resource scenarios beyond CoNLL-ST

<table>
<thead>
<tr>
<th></th>
<th><strong>EXTERNALPOS</strong></th>
<th><strong>TREEBANKPOS</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Training treebank</td>
<td>very small to small</td>
<td>very small to small</td>
</tr>
<tr>
<td>Parallel data</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>POS gold data</td>
<td>external</td>
<td>treebank</td>
</tr>
<tr>
<td>POS accuracy:</td>
<td>good</td>
<td>varies</td>
</tr>
</tbody>
</table>

Test cases:
## Generalized low-resource scenarios beyond CoNLL-ST

<table>
<thead>
<tr>
<th></th>
<th><strong>EXTERNAL POS</strong></th>
<th><strong>TREEBANK POS</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Training treebank</td>
<td>very small to small</td>
<td>very small to small</td>
</tr>
<tr>
<td>Parallel data</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>POS gold data</td>
<td>external</td>
<td>treebank</td>
</tr>
<tr>
<td>POS accuracy:</td>
<td>good</td>
<td>varies</td>
</tr>
<tr>
<td>Test cases:</td>
<td>surprise languages</td>
<td>small languages</td>
</tr>
<tr>
<td>languages</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Simulated low-resource scenarios

- Test languages:
  - surprise/small languages of CoNLL-ST

- Experiments, analysis:
  - simulate low-resource with 41 big CoNLL-ST treebanks
  - artificial training samples: starting with 100 tokens
    - for \texttt{EXTERNALPOS}— sample trees
    - for \texttt{TREEBANKPOS}— trees and POS tags
Simulated low-resource scenarios

- Test languages:
  - surprise/small languages of CoNLL-ST

- Experiments, analysis:
  - simulate low-resource with 41 big CoNLL-ST treebanks
  - artificial training samples: starting with 100 tokens
    - for \textsc{ExternalPOS}—sample trees
    - for \textsc{TreebankPOS}—trees and POS tags
Monolingual method

Method (LEX):

- train a parser on the target treebank (even 10 sentences)

Source: Polish

Target: Polish

Tool:

- transition-based parser
  
  default features and parameters

(Björkelund and Nivre, 2015)
Cross-lingual method

- Most of the cross-lingual techniques need parallel data.

- Alternative: delexicalized parsing (Zeman and Resnik, 2008)
Cross-lingual method

- Most of the cross-lingual techniques need parallel data.

- Alternative: delexicalized parsing (Zeman and Resnik, 2008)
Cross lingual method – DELEX

Method (DELEX):

- Delexicalized multisource model transfer and weighted blending.
  (Sagae and Lavie, 2006), (Rosa and Žabokrtský, 2015), (Agić, 2017)

- Details, colourful plots and animation - backup slides.

- Samples of target trees - similarities between sources and targets.
Cross lingual method – **DELEX**

Method (**DELEX**):  
- Delexicalized multisource model transfer and weighted blending.  
  *(Sagae and Lavie, 2006), (Rosa and Žabokrtský, 2015), (Agić, 2017)*  
- Details, colourful plots and animation - backup slides.  
- Samples of target trees - similarities between sources and targets.
Cross lingual method – DELEX

Method (DELEX):

- Delexicalized multisource model transfer and weighted blending.
  
  (Sagae and Lavie, 2006), (Rosa and Žabokrtský, 2015), (Agić, 2017)

- Details, colourful plots and animation - backup slides.

- Samples of target trees - similarities between sources and targets.
Methodology – few more details

- Universal Dependencies v2.0 treebanks (Nivre et al., 2016)
- Two settings: **EXTERNALPOS** and **TREEBANKPOS**
- 46 source languages
- 41 simulated targets

- Goal: compare **LEX** and **DELEX** methods
Methodology – few more details

- Universal Dependencies v2.0 treebanks (Nivre et al., 2016)
- Two settings: **EXTERNALPOS** and **TREEBANKPOS**
- 46 source languages
- 41 simulated targets

- Goal: compare **LEX** and **DELEX** methods
Table of Contents

Motivation

Methodology

Experiments and Analysis

  EXTERNALPOS setting
  TREEBANKPOS setting

Application to Test Languages

Summary

Backup slides
Table of Contents

Motivation

Methodology

Experiments and Analysis
  \textsc{ExternalPOS} setting
  \textsc{TreebankPOS} setting

Application to Test Languages

Summary

Backup slides
Average results across 41 targets

- **LEX** outperforms when there is only 1100 tokens (avg. 55 sentences)
Average results across 41 targets

Does it mean that if you have more than 55 sentences DELEX does not have sense?
Almost no language falls around 1100 tokens.
For some LEX is always the best choice (even trained on 100 tokens!).
Cutting points for all targets

> **Conclusion**: when you have more tokens than 20k – use LEX.
How to analyze the cutting points?

- we want to analyze the influence of source languages
- we use The World Atlas of Language Structures (WALS)
  - how similar is the best source – color
  - how many good sources – size
The bigger and darker the point the better sources exist.
**Conclusion**: when target belongs to an underrepresented language family – use LEX.
Cutting points with source strength

▶ Conclusion: when there is less trees than 20k but very good sources – use DELEX.
Table of Contents

Motivation

Methodology

Experiments and Analysis

  EXTERNALPOS setting

  TREEBANKPOS setting

Application to Test Languages

Summary

Backup slides
The accuracy of parsers drops due to lower POS accuracy.
Average results across 41 targets

- LEX outperforms DELEX at only 250 tokens.
The color indicates POS tagging accuracy.
Cutting points for all targets

▶ **Conclusion**: when you have more tokens than 3k – use LEX.
Cutting points for all targets

▶ Conclusion: when you have more than 1k tokens:
  ▶ POS accuracy is reasonable (more than 70%)
  ▶ then if you have good sources: use DELEX
Table of Contents

Motivation

Methodology

Experiments and Analysis
  EXTERNALPOS setting
  TREEBANKPOS setting

Application to Test Languages

Summary

Backup slides
Intuition about the real test languages – **EXTERNALPOS**

Rules (conclusions from the plots):

- more tokens than 20k – use **LEX**
- target belongs to an underrepresented family – use **LEX**
- there is less trees than 20k but very good sources – use **DELEX**

<table>
<thead>
<tr>
<th>Language</th>
<th>Token Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buryat (19 sent.)</td>
<td></td>
</tr>
<tr>
<td>Kurmanji (20 sent.)</td>
<td></td>
</tr>
<tr>
<td>North Sami (20 sent.)</td>
<td></td>
</tr>
<tr>
<td>Upper Sorbian (23 sent.)</td>
<td></td>
</tr>
</tbody>
</table>
Intuition about the real test languages – \textsc{ExternalPOS}

Rules (conclusions from the plots):

- more tokens than 20k – use \textsc{LEX}
- target belongs to an underrepresented family – use \textsc{LEX}
- there is less trees than 20k but very good sources – use \textsc{DELEX}

<table>
<thead>
<tr>
<th></th>
<th>\textsc{LEX}</th>
<th>\textsc{DELEX}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buryat (19 sent.)</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Kurmanji (20 sent.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North Sami (20 sent.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper Sorbian (23 sent.)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\rightarrow Buryat – Mongolic language without any close relatives among the source languages.
Intuition about the real test languages – **EXTERNALPOS**

Rules (conclusions from the plots):

- more tokens than 20k – use **LEX**
- **target belongs to an underrepresented family** – use **lex**
- there is less trees than 20k but very good sources – use **DELEX**

<table>
<thead>
<tr>
<th>Language</th>
<th>LEX</th>
<th>DELEX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buryat (19 sent.)</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Kurmanji (20 sent.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North Sami (20 sent.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper Sorbian (23 sent.)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

→ Buryat – Mongolic language without any close relatives among the source languages.
Intuition about the real test languages – ExternalPOS

Rules (conclusions from the plots):

- more tokens than 20k – use LEX
- target belongs to an underrepresented family – use LEX
- there is less trees than 20k but very good sources – use DELEX

<table>
<thead>
<tr>
<th>Language</th>
<th>LEX</th>
<th>DELEX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buryat (19 sent.)</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Kurmanji (20 sent.)</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>North Sami (20 sent.)</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>Upper Sorbian (23 sent.)</td>
<td>✗</td>
<td>✓</td>
</tr>
</tbody>
</table>
Intuition about the real test languages – TreebankPOS

Rules (conclusions from the plots):

- more tokens than 3k – use LEX
- more than 1k tokens and very good sources – use DELEX

<table>
<thead>
<tr>
<th>Language</th>
<th>LEX</th>
<th>DELEX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latin (18k tokens)</td>
<td>✔</td>
<td>×</td>
</tr>
<tr>
<td>Irish (13k tokens)</td>
<td>✔</td>
<td>×</td>
</tr>
<tr>
<td>Ukrainian (12k tokens)</td>
<td>✔</td>
<td>×</td>
</tr>
<tr>
<td>Kazakh (529 tokens)</td>
<td>✔</td>
<td>×</td>
</tr>
<tr>
<td>Uyghur (1,5k tokens)</td>
<td>✔</td>
<td>×</td>
</tr>
<tr>
<td>Language</td>
<td>LEX</td>
<td>DELEX</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----</td>
<td>-------</td>
</tr>
<tr>
<td>Buryat</td>
<td>✔</td>
<td>×</td>
</tr>
<tr>
<td>Kurmanji</td>
<td>✔</td>
<td>×</td>
</tr>
<tr>
<td>North Sami</td>
<td>×</td>
<td>✔</td>
</tr>
<tr>
<td>Upper Sorbian</td>
<td>×</td>
<td>✔</td>
</tr>
<tr>
<td>Latin</td>
<td>✔</td>
<td>×</td>
</tr>
<tr>
<td>Irish</td>
<td>✔</td>
<td>×</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>✔</td>
<td>×</td>
</tr>
<tr>
<td>Kazakh</td>
<td>✔</td>
<td>×</td>
</tr>
<tr>
<td>Uyghur</td>
<td>✔</td>
<td>×</td>
</tr>
</tbody>
</table>
### Application to real target languages

<table>
<thead>
<tr>
<th>Language</th>
<th>LEX</th>
<th>DELEX</th>
<th>LEX</th>
<th>DELEX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buryat</td>
<td>✓</td>
<td>×</td>
<td>28.06</td>
<td>32.01</td>
</tr>
<tr>
<td>Kurmanji</td>
<td>✓</td>
<td>×</td>
<td>40.94</td>
<td>38.17</td>
</tr>
<tr>
<td>North Sami</td>
<td>×</td>
<td>✓</td>
<td>29.80</td>
<td>35.80</td>
</tr>
<tr>
<td>Upper Sorbian</td>
<td>×</td>
<td>✓</td>
<td>49.59</td>
<td>58.81</td>
</tr>
<tr>
<td>Latin</td>
<td>✓</td>
<td>×</td>
<td>41.02</td>
<td>35.06</td>
</tr>
<tr>
<td>Irish</td>
<td>✓</td>
<td>×</td>
<td>64.66</td>
<td>44.54</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>✓</td>
<td>×</td>
<td>64.81</td>
<td>63.67</td>
</tr>
<tr>
<td>Kazakh</td>
<td>✓</td>
<td>×</td>
<td>26.55</td>
<td>26.17</td>
</tr>
<tr>
<td>Uyghur</td>
<td>✓</td>
<td>×</td>
<td>34.81</td>
<td>30.11</td>
</tr>
</tbody>
</table>
## Application to real target languages

<table>
<thead>
<tr>
<th>Language</th>
<th>LEX</th>
<th>DELEX</th>
<th>LEX</th>
<th>DELEX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buryat</td>
<td>✔</td>
<td>✗</td>
<td>28.06</td>
<td>32.01</td>
</tr>
<tr>
<td>Kurmanji</td>
<td>✔</td>
<td>✗</td>
<td>40.94</td>
<td>38.17</td>
</tr>
<tr>
<td>North Sami</td>
<td>✗</td>
<td>✔</td>
<td>29.80</td>
<td>35.80</td>
</tr>
<tr>
<td>Upper Sorbian</td>
<td>✗</td>
<td>✔</td>
<td>49.59</td>
<td>58.81</td>
</tr>
<tr>
<td>Latin</td>
<td>✔</td>
<td>✗</td>
<td>41.02</td>
<td>35.06</td>
</tr>
<tr>
<td>Irish</td>
<td>✔</td>
<td>✗</td>
<td>64.66</td>
<td>44.54</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>✔</td>
<td>✗</td>
<td>64.81</td>
<td>63.67</td>
</tr>
<tr>
<td>Kazakh</td>
<td>✔</td>
<td>✗</td>
<td>26.55</td>
<td>26.17</td>
</tr>
<tr>
<td>Uyghur</td>
<td>✔</td>
<td>✗</td>
<td>34.81</td>
<td>30.11</td>
</tr>
</tbody>
</table>
Table of Contents

Motivation

Methodology

Experiments and Analysis
  \textsc{ExternalPOS} setting
  \textsc{TreebankPOS} setting

Application to Test Languages

Summary

Backup slides
Conclusions

Lexicalized vs. delexicalized?

- treebank size
- POS tagging accuracy
- typological relations between languages

We presented a methodology which worked for 8 out of 9 test languages.

Take home messages:

- Monolingual approach is a strong baseline.
- One should not be deceived by a small training treebank size.
- Never believe only your averages.
Conclusions

Lexicalized vs. delexicalized?

- treebank size
- POS tagging accuracy
- typological relations between languages

We presented a methodology which worked for 8 out of 9 test languages.

Take home messages:

- Monolingual approach is a strong baseline.
- One should not be deceived by a small training treebank size.
- Never believe only your averages.
Conclusions

Lexicalized vs. delexicalized?

▶ treebank size
▶ POS tagging accuracy
▶ typological relations between languages

We presented a methodology which worked for 8 out of 9 test languages.

Take home messages:

▶ Monolingual approach is a strong baseline.
▶ One should not be deceived by a small training treebank size.
▶ Never believe only your averages.
Conclusions

Lexicalized vs. delexicalized?
- treebank size
- POS tagging accuracy
- typological relations between languages

We presented a methodology which worked for 8 out of 9 test languages.

Take home messages:
- Monolingual approach is a strong baseline.
- One should not be deceived by a small training treebank size.
- Never believe only your averages.
Conclusions

Lexicalized vs. delexicalized?

- treebank size
- POS tagging accuracy
- typological relations between languages

We presented a methodology which worked for 8 out of 9 test languages.

Take home messages:

- Monolingual approach is a strong baseline.
- One should not be deceived by a small training treebank size.
- Never believe only your averages.
Conclusions

Lexicalized vs. delexicalized?

▶ treebank size
▶ POS tagging accuracy
▶ typological relations between languages

We presented a methodology which worked for 8 out of 9 test languages.

Take home messages:

▶ Monolingual approach is a strong baseline.
▶ One should not be deceived by a small training treebank size.
▶ Never believe only your averages.
Table of Contents

Motivation

Methodology

Experiments and Analysis

EXTERNALPOS setting

TREEBANKPOS setting

Application to Test Languages

Summary

Backup slides
Cross-lingual method – main idea

Delexicalized multisource model transfer and weighted blending.
Cross-lingual method – main idea

Delexicalized multisource **model transfer** and weighted blending.

Source: English  
Target: Polish
Cross-lingual method – main idea

**Delexicalized** multisource model transfer and weighted blending.

Source: English

Target: Polish
Cross-lingual method – main idea

Delexicalized **multisource** model transfer and weighted blending.

**Source:** English

**Source:** French

**Source:** German

**Target:** Polish
Cross-lingual method – main idea

Delexicalized multisource model transfer and weighted **blending**.
Cross-lingual method – main idea

Delexicalized multisource model transfer and **weighted** blending.

Source: English

Source: French

Source: German

Target: Polish
Cross lingual method – weighting methods

Source weighting methods:

- Gold POS tags (Rosa and Žabokrtský, 2015)
- Predicted POS tags, topological features (Agić, 2017)
- LAS on target trees samples

- Here only the last one – DELEX.
Cross lingual method – weighting methods

Source weighting methods:

- Gold POS tags  
  (Rosa and Žabokrtský, 2015)
- Predicted POS tags, topological features  
  (Agić, 2017)
- LAS on target trees samples

- Here only the last one – DELEX.
Cross lingual method – weighting methods

Source weighting methods:

- Gold POS tags (Rosa and Žabokrtský, 2015)
- Predicted POS tags, topological features (Agić, 2017)
- **LAS on target trees samples**

- Here only the last one – DELEX.


