Dependency Trees in Automatic Inflection of Multi Word Expressions in Polish

Ryszard Tuora  Łukasz Kobyliński

Institute of Computer Science
Polish Academy of Sciences
Warsaw, Poland

September 21, 2021
Motivation

- Natural language generation still uses template based methods.
Motivation

- Natural language generation still uses template based methods.
  
  You have a message from ...
Motivation

- Natural language generation still uses template based methods.
  
  You have a message from ...

- Template based methods are harder to implement in morphologically rich languages.
Motivation

- Natural language generation still uses template based methods.
  
  You have a message from ...

- Template based methods are harder to implement in morphologically rich languages.
  
  Masz wiadomość od ...

A genitive case is needed.
Motivation

- Natural language generation still uses template based methods.
  You have a message from ... 
- Template based methods are harder to implement in morphologically rich languages.
  Masz wiadomość od ... 
  A genitive case is needed.
- In many use cases Multi Word Expressions (MWEs) are more prevalent.
Motivation

- Natural language generation still uses template based methods.
  You have a message from ...

- Template based methods are harder to implement in morphologically rich languages.
  Masz wiadomość od ...
  A genitive case is needed.

- In many use cases Multi Word Expressions (MWEs) are more prevalent.
  *Masz wiadomość od Anna Kowalska.
Motivation

- Natural language generation still uses template based methods.
  
  You have a message from ...

- Template based methods are harder to implement in morphologically rich languages.
  
  Masz wiadomość od ...
  A genitive case is needed.

- In many use cases Multi Word Expressions (MWEs) are more prevalent.
  
  *Masz wiadomość od Anna Kowalska.
  vs.
  
  Masz wiadomość od Anny Kowalskiej.
Morphological agreement

Biała flaga z papieru $\xrightarrow{pl.dat}$ Białym flagom z papieru

\text{WHITE FLAG OF PAPER}
Morphological agreement

Biała flaga z papieru $\xrightarrow{pl.dat}$ Białym flagom z papieru

- *Biała* modifies *flaga*, and so must retain agreement under inflection
Morphological agreement

- *Biała* modifies *flaga*, and so must retain agreement under inflection
- *papieru* is a nominal modifier governed by *z*, and so it does not change under inflection.
We consider two methods of inflecting individual words:

- Dictionary-based inflection
- Neural inflection
Single Word Inflection

We consider two methods of inflecting individual words:

- Dictionary-based inflection
- Neural inflection

Then we extend this to MWE’s, using dependency trees.

Biała flaga z papieru
Dependency relations as a proxy for agreement relations

A rule has the form:

\[ \text{dep} \rightarrow \text{attrs} \]

where \text{dep} corresponds to the dependency label, and \text{attr} represents the set of attributes, for which agreement between the head, and dependent occurs.
Dependency relations as a proxy for agreement relations

A rule has the form:

\[ \text{dep} \rightarrow \text{attrs} \]

where \text{dep} corresponds to the dependency label, and \text{attr} represents the set of attributes, for which agreement between the head, and dependent occurs.

\[ \text{amod} \rightarrow \text{number.case.gender} \]

represents agreement between an adjective modifier and its head noun.
## Agreement statistics

<table>
<thead>
<tr>
<th>Deprel</th>
<th>number</th>
<th>gender</th>
<th>person</th>
<th>case</th>
</tr>
</thead>
<tbody>
<tr>
<td>amod</td>
<td>99.78</td>
<td>99.81</td>
<td></td>
<td>99.48</td>
</tr>
<tr>
<td>appos</td>
<td>94.33</td>
<td>82.57</td>
<td></td>
<td>88.30</td>
</tr>
<tr>
<td>conj</td>
<td>83.62</td>
<td>60.57</td>
<td>88.02</td>
<td>95.97</td>
</tr>
<tr>
<td>det</td>
<td>99.36</td>
<td>98.59</td>
<td></td>
<td>98.31</td>
</tr>
<tr>
<td>flat</td>
<td>98.00</td>
<td>92.92</td>
<td></td>
<td>92.77</td>
</tr>
<tr>
<td>nummod</td>
<td>99.59</td>
<td>95.44</td>
<td></td>
<td>97.18</td>
</tr>
<tr>
<td>obj</td>
<td>62.85</td>
<td>20.61</td>
<td>36.59</td>
<td>40.59</td>
</tr>
<tr>
<td>obl</td>
<td>65.39</td>
<td>25.34</td>
<td>59.79</td>
<td>11.27</td>
</tr>
</tbody>
</table>

**Table.** Frequency of agreement (in percent) for a selection of morphological attributes, between the dependency head and its children, given the dependency label, PDB treebank.
Propagating features

The features are propagated along dependency arcs, if the ruleset allows it.
The features are propagated along dependency arcs, if the ruleset allows it

→ Białym flagom z papieru
### Table

Form accuracy for two methods of inflecting MWEs, as evaluated on 50k examples from SEJF.

<table>
<thead>
<tr>
<th></th>
<th>dictionary based</th>
<th>neural based</th>
</tr>
</thead>
<tbody>
<tr>
<td>inflection accuracy</td>
<td>90.54</td>
<td>85.21</td>
</tr>
<tr>
<td>lemmatization accuracy</td>
<td>79.97</td>
<td>78.96</td>
</tr>
</tbody>
</table>
Thank you for your attention!