

Dependency Trees in Automatic Inflection of Multi Word Expressions in Polish

Ryszard Tuora Łukasz Kobylński

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.

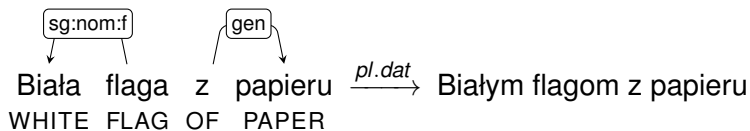
- In many use cases Multi Word Expressions (MWEs) are more prevalent.

*Masz wiadomość od Anna Kowalska.

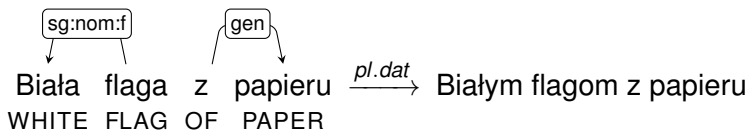
vs.

Masz wiadomość od Anny Kowalskiej.

Morphological agreement

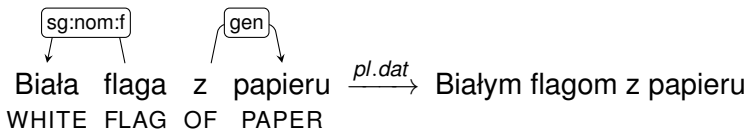


Morphological agreement



- *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

Single Word Inflection

We consider two methods of inflecting individual words:

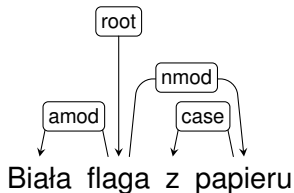
- Dictionary-based inflection
- Neural inflection

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.



Dependency relations as a proxy for agreement relations

A rule has the form:

$$\text{dep} \rightarrow \text{attrs}$$

where `dep` corresponds to the dependency label, and `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 `dep` corresponds to the dependency label, and `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

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

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.

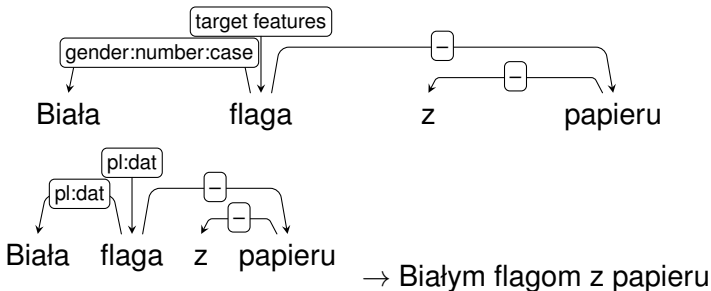
Propagating features

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



Propagating features

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



Evaluation

	dictionary based	neural based
inflection accuracy	90.54	85.21
lemmatization accuracy	79.97	78.96

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

Thank you for your attention!