

Augmenting Stylometric Features to Improve Detection of Propaganda and Manipulation

Radoslav Sabol, Aleš Horák,
xsabol@fi.muni.cz, hales@fi.muni.cz

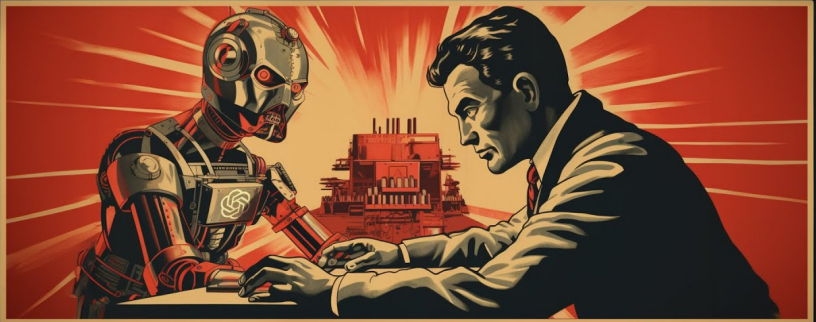
NLP Centre, Faculty of Informatics, Masaryk University

December 8, 2023

Outline

1. **The Current State of Stylometric Features**
2. **Newly Added Feature Extractors:**
 - 2.1 Syntactic Features
 - 2.2 Readability Features
 - 2.3 Morphological features
3. **Evaluation**
 - 3.1 Comparison with the previous feature set
 - 3.2 Feature Importance

COUNTERCLOUD



Motivation

- attempt to **improve the performance** on the *Propaganda* dataset using style analysis
- success of stylometry is mostly based on **feature engineering techniques**
 - need a **diverse, high-quality set of secondary text features**
 - the ones that can be **carried over to other tasks and languages**
 - authorship attribution, author profiling...

Currently Implemented Stylometric Features

Overview

- Distribution of word lengths
- Distribution of sentence lengths
- Word repetition
- Word class n-grams
- Morphological tags n-grams
- Presence of letter-casing in sentences
- Dynamic stopwords
- Punctuation
- Dynamic Typography
- Emoticons
- Character n-grams
- Dependency Tree analysis
- Readability metrics
- Word suffixes
- Word richness

Legend: unchanged, minor changes, reworked, brand new

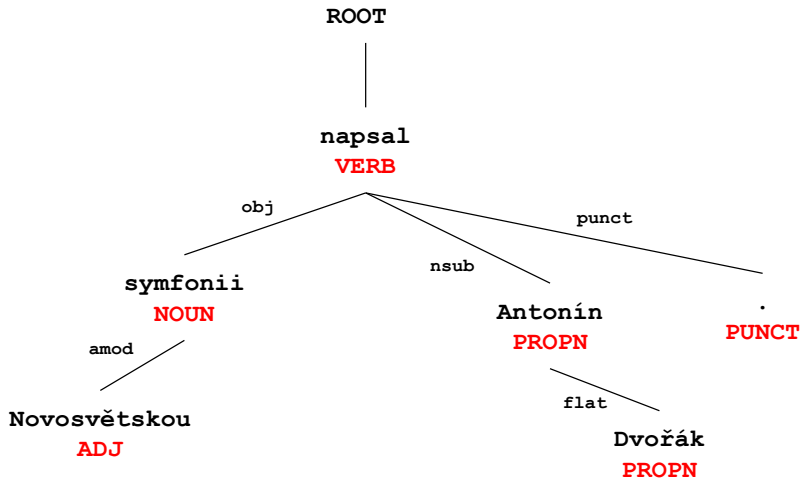
Readability Metrics (4 features)

- statistical measures based on **sentence length**, **number of words** and **syllables**
 - *Flesch Reading Ease*
 - *Gunning Fog Index*
 - *McLaughlin's SMOG formula*
 - *FORCAST formula*
- splitting into syllables is a language-dependent task

Dependency Trees - Intro

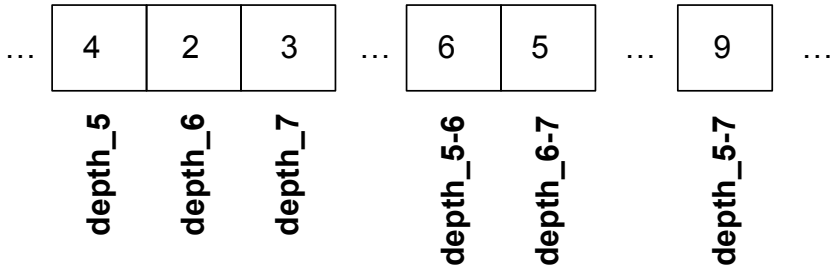
- the previous writeprint features no tree-based syntactic features
- added new support object - **UDPipe** dependency trees for each sentence in the document
- projected into a vector via two kinds of numerical features:
 - *structural tree features*
 - *relation-based features*

Structural Tree Features - Example Tree

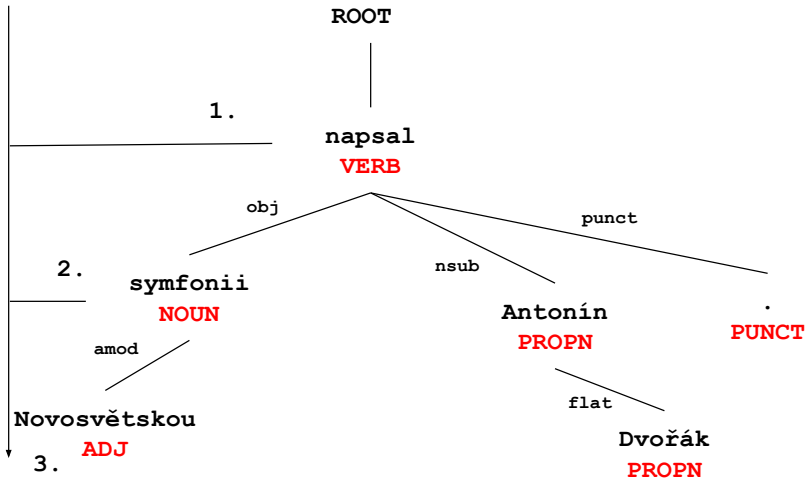


Structural Tree Features (180 features)

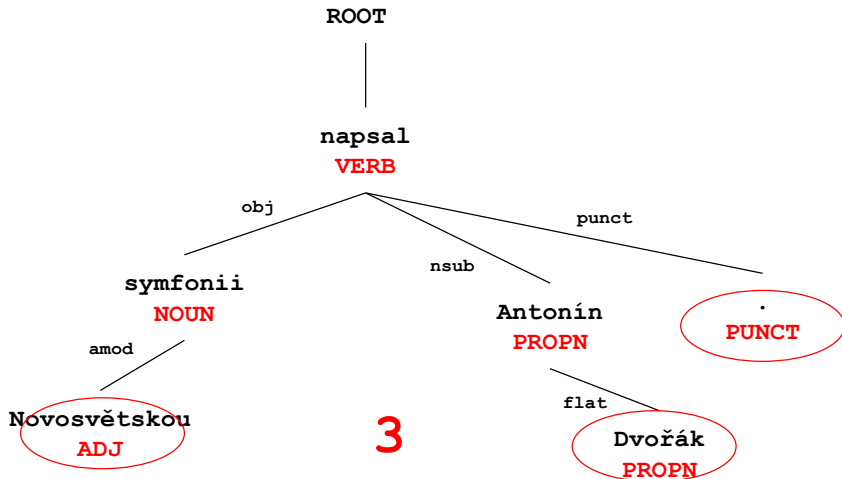
- tree depth, tree width, tree branching factors
- computed as relative frequencies over the entire document
- also contains **binned values**



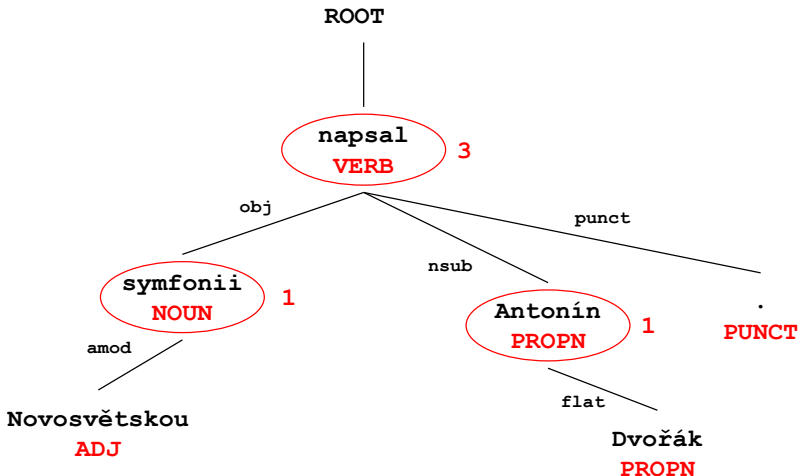
Structural Tree Features - Example Tree Depth



Structural Tree Features - Example Tree Width



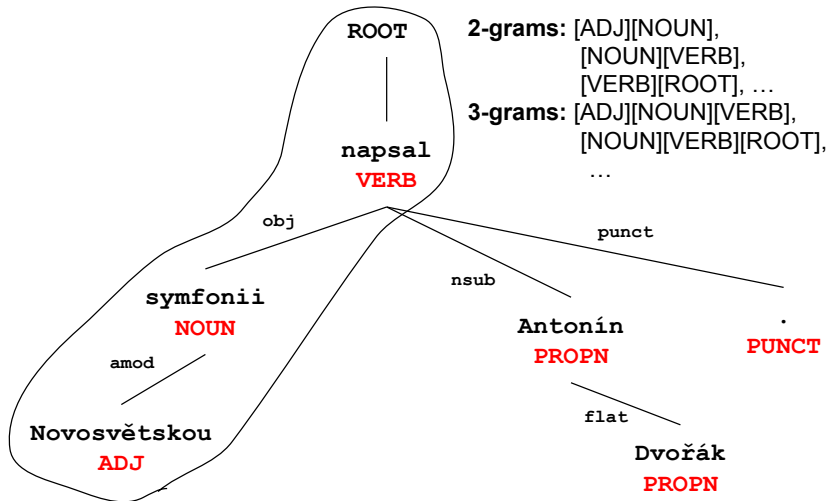
Structural Tree Features - Example Tree Branching Factors



Relation-Based Features (3,267 features)

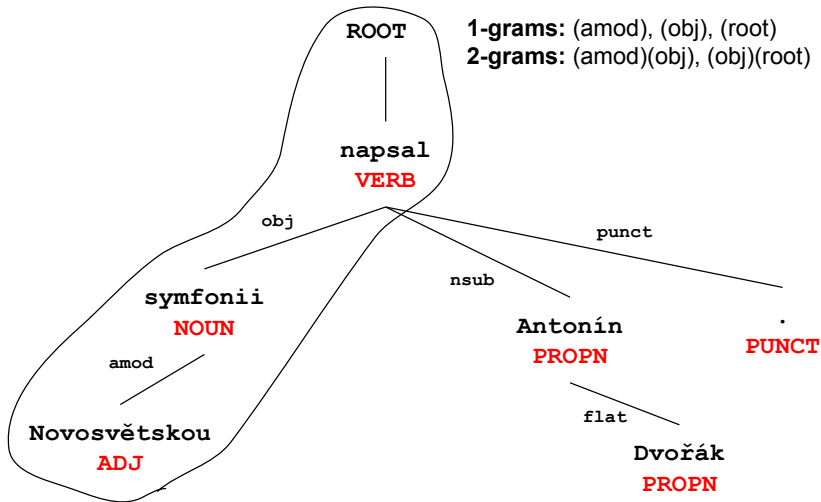
- captures **syntactic relations between words** in the form of n-grams
- three different versions
 - **node labels only** (2-4grams)
 - [NOUN][VERB]
 - **relation types only** (1-4grams)
 - (amod)(obj)
 - **both nodes and relation types** (2-4grams)
 - [NOUN](obj)[VERB]

Node Labels Only - Example

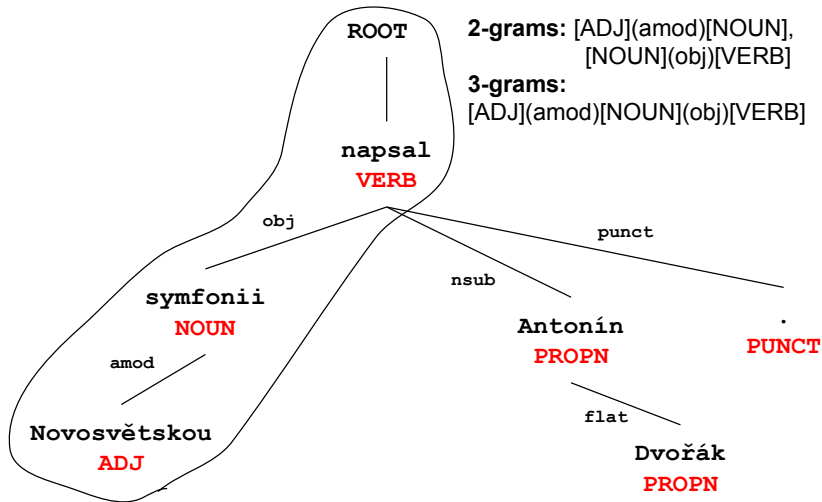


Relation Types Only - Example

1-grams: (amod), (obj), (root)
2-grams: (amod)(obj), (obj)(root)



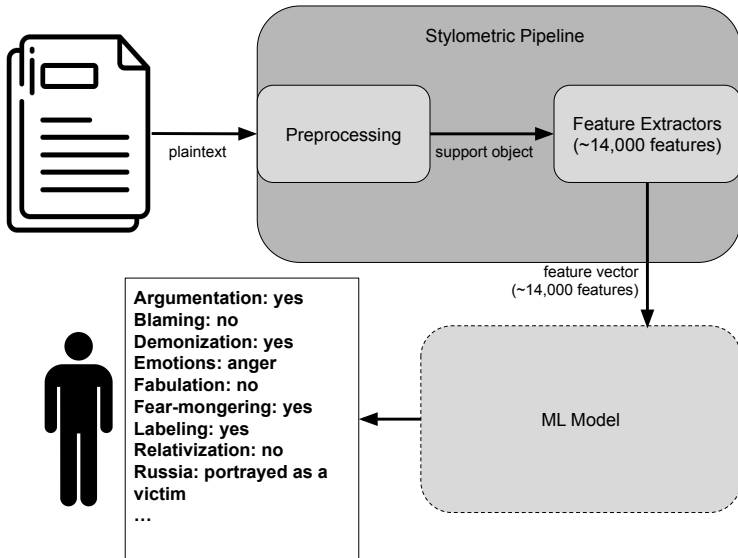
Node and Relation Types - Example



Morphological ngrams Re-visited

- previous extractor was extracting approximately **8,000 different sequences of morphological tags**
 - 1-4grams
 - the amount of extracted features was **determined solely** by the n
 - results in extracting **highly improbable sequences**
- fix: extract only ngrams that occur in **at least 0.05% of the training documents**
 - overall reduction to **1,500 numerical features**

Propaganda Detection Task



Propaganda Dataset

- **benchmark dataset** of 8,644 documents
- annotation of **manipulative techniques**
 - *Argumentation, Blaming, Emotions, Demonization, Fabulation, Fear-mongering, Labelling, Relativizing*
- annotation of **document level attributes**
 - *Genre, Topic, Scope, Location, Overall Sentiment*
- annotation of **other attributes**
 - *Russia, Expert, Source, Opinion*

Propaganda Dataset - Example

Lubomír Man: Skvělý projev zodpovědného prezidenta

[ac24.cz](#) 

✘ odborník

Lubomír Man: Skvělý projev zodpovědného prezidenta

✘ názor

Prezident o této situaci, jež je v tak těžkém rozporu s demokratickým požadavkem rovností možností pro všechny, ve svém inauguračním projevu promluvil, a zdá se téměř neuvěřitelné, že to musel být ze všech našich oficiálních osobností a politiků až on, který tak učinil.

✘ nalepkovani

Ale stalo se, nu a zvěčnělá poslankyně ODS tuhle pravdu prostě neunesla. A nebyla sama, s asi stometrovým odstupem od ní uháněli k východu z Vladislavského sálu další naši demokraté z TOP 09, ODS, KDU-ČSL a STAN, v doprovodu europoslance Pospíšila pokulhával zotavit se z téhle hrůzy na čerstvý vzduch i kníže Schwarzenberg.

Experiment Setup

1. compare the recent changes using **Gradient-Boosted Decision Trees**
 - compare the results between the previous and updated feature detectors (weighted F1)
2. measure the importance of feature extractors
 - feature extractors split into 16 categories (*Character ngrams, Word Lengths, etc.*)
 - **measurement:**
 - 2.1 select a category and remove all underlying feature extractors
 - 2.2 train and evaluate a new model using the remaining subset, then measure the difference in F1

Comparison

Name	old	new	diff
genre	95.47	95.71	0.24
topic	58.21	59.37	1.16
scope	86.65	85.51	-1.14
location	69.42	70.47	1.05
argumentation	68.54	69.04	0.5
emotions	80.69	80.56	-0.13
overall_sentiment	79.45	81.19	1.74
russia	80.61	81.11	0.5
opinion	87.41	87.74	0.33
expert	71.44	73.75	2.31
source	67.07	67.54	0.47
fear_mongering	90.02	90.73	0.71
blaming	71.67	71.72	0.06
labelling	82.13	82.78	0.65
demonization	95.60	95.75	0.00
relativization	92.48	92.56	0.08
fabulation	79.72	80.21	0.49

Comparison (no eyestrain edition)

Name	old	new	diff
topic	58.21	59.37	1.16
scope	86.65	85.51	-1.14
location	69.42	70.47	1.05
overall_sentiment	79.45	81.19	1.74
expert	71.44	73.75	2.31
fear_mongering	90.02	90.73	0.71
labelling	82.13	82.78	0.65

Feature Importance



Conclusion and Future Work

- **New Features:** high potential improvement
 - further tweak the syntactic features
 - add more complex syntactic features
- add more features based on **external corpora**:
 - use feature deviation from non-manipulative texts
 - wordlist frequencies, collocation statistics
- other possible features:
 - sentiment and sentiment contrast
 - *useful idiot, skilled manipulator, beautifully dysfunctional*

MUNI

FACULTY

OF INFORMATICS