



# EVALUATING A MODERN POLISH POS-TAGGER ON HISTORICAL DATA

A project by  
Maria irena szawerna for LT2314 Language Technology Resources

# Background

- POS information can be useful in other types of research (e.g. corpus linguistics).
- POS taggers are created for big modern languages or require much data to train.
- Solution: using modern POS taggers on historical data?
- Issues: OOV items, variation in spelling and punctuation, different grammar...
- These problems can be solved, but first they need to be identified.
- Most research has been done on Germanic languages.



# RESEARCH QUESTION:

How well does the GATE Cloud tagger for Polish perform on historical data and what kinds of mistakes does it tend to make?

# Materials and methods

- GATE Cloud tagger for Polish, trained on Universal Dependencies treebanks.
- Manually annotated data from 19<sup>th</sup> century memoirs, 3000+ tokens.
- Code in a Jupyter Notebook file: data preprocessing, tagger annotation, an array of evaluation measures (quantitative and qualitative).
- Manual error analysis.

# RESULTS

POS tag	Capitalized	Lowercase
ADJ	<b>61.51%</b>	52.56%
ADP	95.65%	<b>95.94%</b>
ADV	<b>72.29%</b>	70.11%
AUX	<b>76.92%</b>	73.91%
CCONJ	<b>97.52%</b>	97.50%
DET	98.46%	<b>98.67%</b>
NOUN	<b>84.94%</b>	73.23%
NUM	<b>73.68%</b>	69.44%
PART	82.22%	<b>82.61%</b>
PRON	82.89%	<b>83.12%</b>
PROPN	<b>66.23%</b>	0.00%
PUNCT	100.00%	100.00%
SCONJ	<b>80.43%</b>	77.50%
VERB	<b>84.40%</b>	82.45%
X	<b>64.52%</b>	64.00%

Table 2: Precision per POS tag per trial.

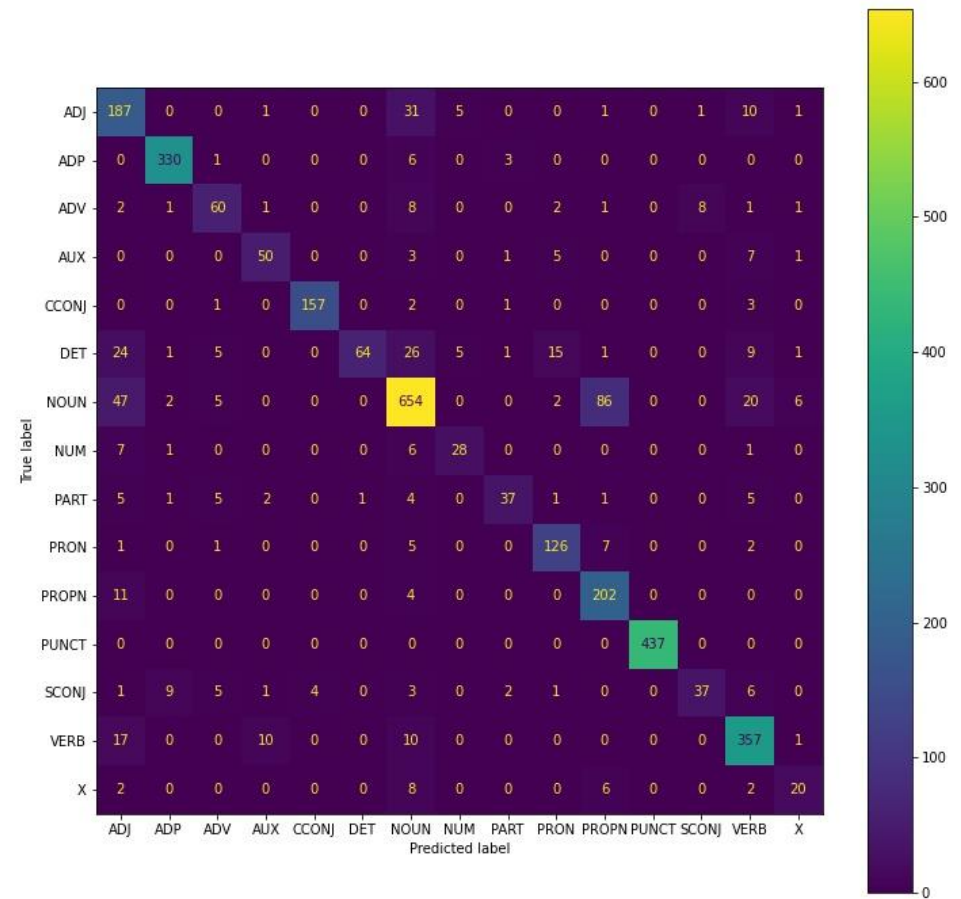
POS tag	Capitalized	Lowercase
ADJ	<b>78.90%</b>	78.06%
ADP	97.06%	<b>97.35%</b>
ADV	70.59%	<b>71.76%</b>
AUX	74.63%	<b>76.12%</b>
CCONJ	<b>95.73%</b>	95.12%
DET	41.11%	<b>48.68%</b>
NOUN	79.56%	<b>89.54%</b>
NUM	<b>65.12%</b>	58.14%
PART	59.68%	<b>61.29%</b>
PRON	88.73%	<b>92.66%</b>
PROPN	<b>93.09%</b>	0.00%
PUNCT	100.00%	100.00%
SCONJ	<b>53.62%</b>	44.93%
VERB	90.38%	90.38%
X	<b>52.63%</b>	42.11%

Table 3: Recall per POS tag per trial.

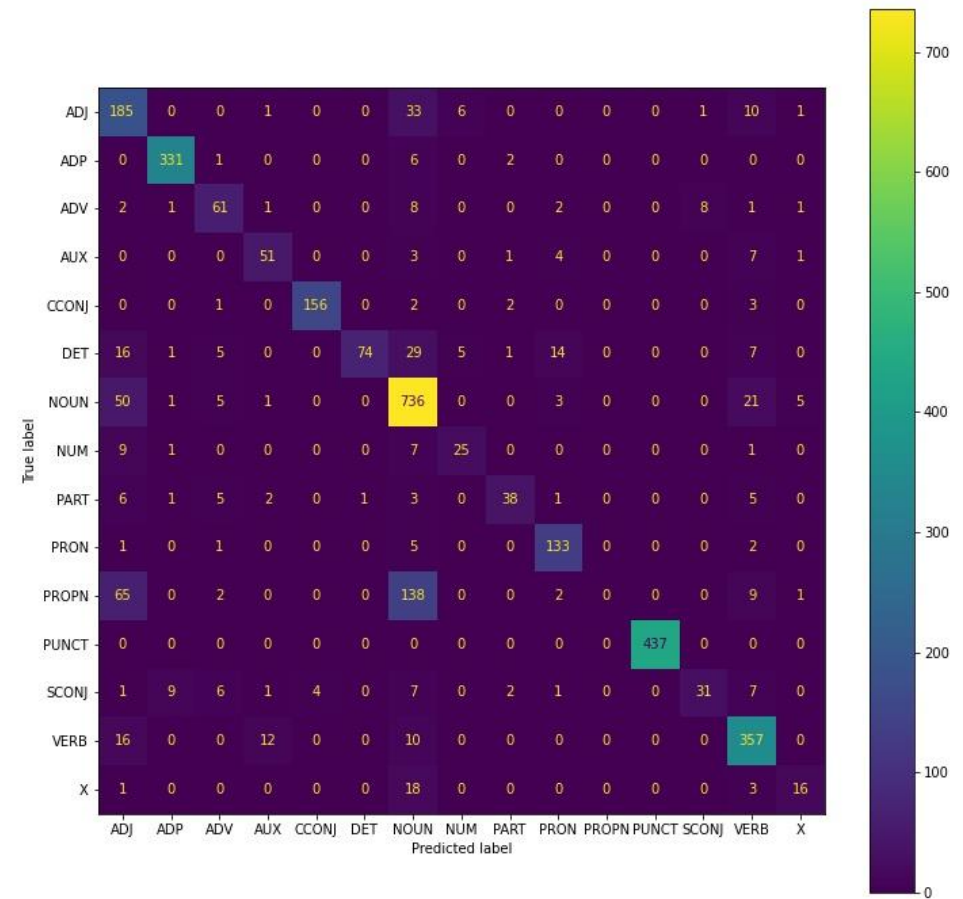
Data capitalization	Accuracy	MCC
Original capitalization	83.98%	81.82%
All lowercase	80.46%	77.80%

Table 1: Accuracy and MCC per trial.

# Confusion matrix (original capitalization)



# Confusion matrix (lowercase)



# Qualitative error analysis

- [Jupyter Notebook](#)
- Capitalization, confusing word endings, variation in spelling, the ability of some words to function as multiple classes, and archaic vocabulary.
- Mysterious cases.



# Discussion

- Performance better than expected (above 80%).
- Issues similar to the ones described in the literature, except the confusing endings.
- Lowercasing the data does not bring obvious benefits, and is detrimental to PROPN detection.

# Future Work

- Influence of punctuation.
- Testing other taggers on the same data.
- Testing this tagger on different kinds of historical data.
- Testing taggers on modern nonstandard data.
- Developing methods for preprocessing the data.



THANK YOU FOR  
YOUR ATTENTION