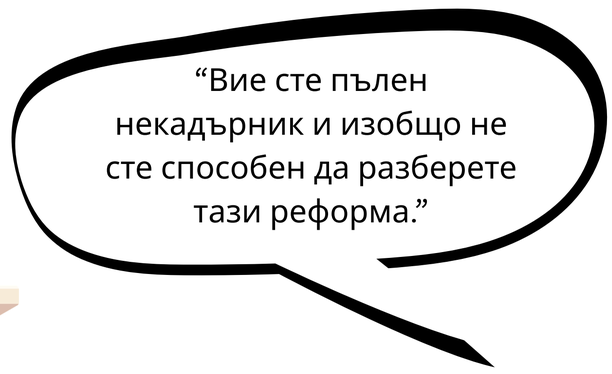
Syntactic Patterns in Political Speech: Dependency-Based Ad Hominem Detection in Bulgarian Social Media

Nevena Grigorova (Gate Institute), Stanislav Penkov (Gate Institute), Keith Kiely (Gate Institute),
Dobromir Tsolyov (Gate Institute), Iglika Ivanova (Gate Institute, Sofia University)



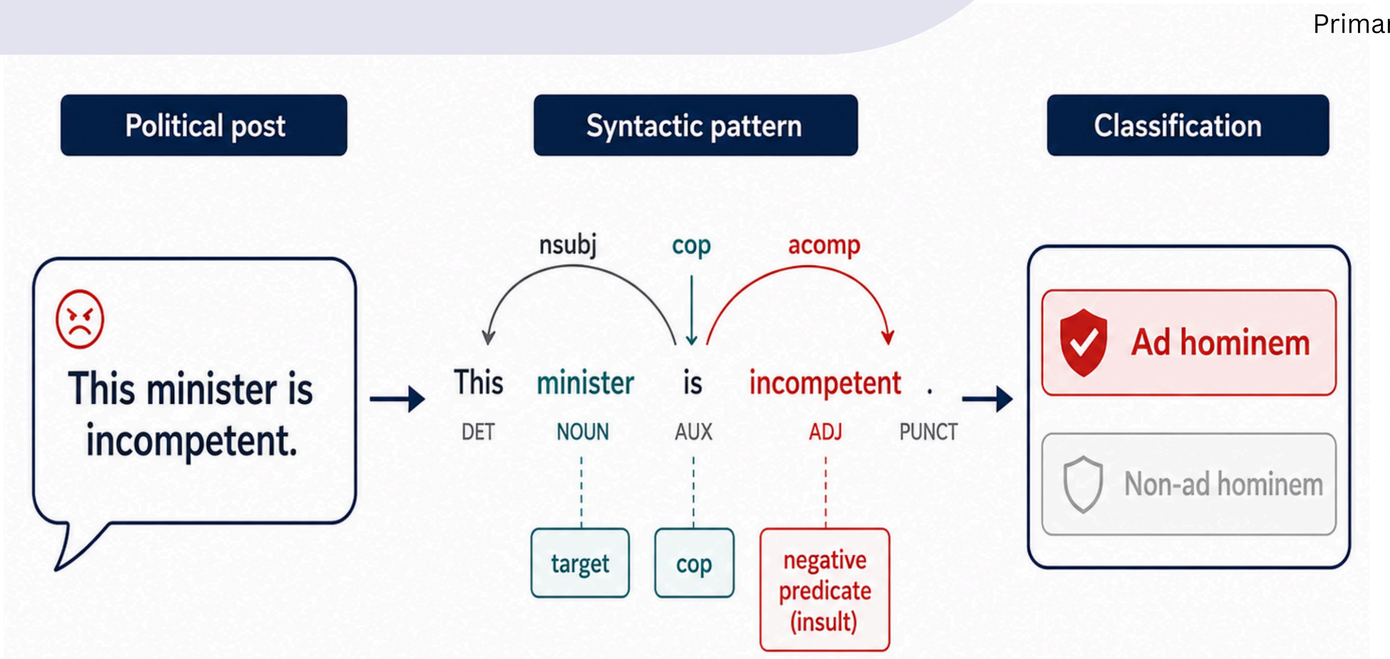
AUTOMATIC DETECTION OF AD HOMINEM ATTACKS IN BULGARIAN POLITICAL SOCIAL MEDIA USING DEPENDENCY-BASED STRUCTURAL FEATURES

Data

- Bulgarian political social media corpus, annotated for **logical fallacies.**
- **6 annotators** with **85% agreement** on the ad hominem label
- Binary sentence classification task (1 = ad_hominem, 0 = control).
- The resulting dataset contains **177 sentences** in total (53 positives, 124 negatives; positive rate: 29.9%).
- The exapmples are parsed **with Stanza,** using its Bulgarian UD pipeline, aligned with the **UD Bulgarian-BTB treebank.**

Five syntactic patterns of AD HOMINEM attacks:

- Attributive Insult (Qualified Subject)
- Predicative Insult
- Coordinated Discrediting
- Passive Discrediting
- Attack within Complement Clauses



Methodology and experiments

- five recurring UD configurations via three structural feature groups (UD relation ratios, UPOS distributions, and template-shape indicators)
- compared them to TF–IDF
- combined representation across LR, DT, and SVM classifiers

Results

Under the primary repeated stratified nested cross-validation protocol, **TF–IDF provides the strongest and most stable baseline**, while dependency-derived structural features remain competitive and capture syntactic regularities consistent with the linguistic patterns.

Feature set	Model	F1	Prec.	Rec.	ROC-AUC
Structural only	Logistic Regression	0.4390	0.3000	0.8182	0.6400
	Decision Tree	0.5185	0.4375	0.6364	0.6455
	SVC	0.5385	0.4667	0.6364	0.6145
TF–IDF only	Logistic Regression	0.2222	0.2857	0.1818	0.6036
	Decision Tree	0.3158	0.3750	0.2727	0.5364
	SVC	0.3158	0.3750	0.2727	0.6255
Combined	Logistic Regression	0.3810	0.4000	0.3636	0.7200
	Decision Tree	0.3810	0.4000	0.3636	0.5764
	SVC	0.5385	0.4667	0.6364	0.6618

Illustrative evaluation: single stratified 80/20 train–test split with 5-fold GridSearchCV on the training set. Results are split-sensitive due to the small test set.

Feature set	Model	F1	Prec.	Rec.	ROC-AUC
Structural only	Logistic Regression	0.489 ± 0.099	0.435 ± 0.098	0.597 ± 0.182	0.655 ± 0.081
	Decision Tree	0.474 ± 0.140	0.468 ± 0.131	0.502 ± 0.181	0.632 ± 0.098
	SVC	0.478 ± 0.091	0.407 ± 0.085	0.615 ± 0.171	0.639 ± 0.082
TF–IDF only	Logistic Regression	0.548 ± 0.074	0.414 ± 0.060	0.818 ± 0.130	0.698 ± 0.093
	Decision Tree	0.418 ± 0.123	0.454 ± 0.124	0.410 ± 0.165	0.604 ± 0.083
	SVC	0.396 ± 0.157	0.350 ± 0.150	0.570 ± 0.340	0.509 ± 0.222
Combined	Logistic Regression	0.491 ± 0.105	0.470 ± 0.121	0.561 ± 0.184	0.688 ± 0.088
	Decision Tree	0.446 ± 0.124	0.455 ± 0.135	0.454 ± 0.154	0.612 ± 0.084
	SVC	0.477 ± 0.120	0.460 ± 0.132	0.538 ± 0.185	0.674 ± 0.097

Primary evaluation: repeated stratified nested cross-validation (5 folds × 20 repeats; mean ± SD over 100 folds). Metrics refer to the positive class (ad hominem).

Limitations

- Small dataset: 177 sentences / 53 ad hominem.
- No discourse context or coreference resolution.
- Social-media language may cause UD parsing errors.
- Structural patterns may produce false positives.
- TF–IDF may capture dataset-specific lexical bias.
- Generalization beyond the corpus is limited.

Future work

- Larger & more diverse Bulgarian political datasets
- **Context-aware modelling** beyond sentence-level classification
- Richer syntactic features: dependency paths and subgraph motifs
- LLMs as complementary models under strict evaluation
- **LLMs for annotation & validation assistance**
- Structured explanations aligned with UD evidence
- Improved error analysis and pattern refinement

This research is part of the GATE project funded by the Horizon 2020 WIDESPREAD- 2018-2020 TEAMING Phase 2 Programme under grant agreement no. 857155, the Programme "Research, Innovation and Digitalization for Smart Transformation"2021- 2027 (PRIDST) under grant agreement no. BG16RFPR002-1.014-0010-C01 and, the project BROD (Bulgarian-Romanian Observatory of Digital Media) funded by the Digital Europe Programme of the European Union under contract number 101226153.