

Placing M-Phasis on the Plurality of Hate: A Feature-Based Corpus of Hate Online

Dana Ruiter^{1*}, Liane Reiners^{2*}, Ashwin Geet D'Sa³,

Thomas Kleinbauer¹, Dominique Fohr³, Irina Illina³, Dietrich Klakow¹,

Christian Schemer², Angeliki Monnier⁴

{druiter, kleiba, dietrich}@lsv.uni-saarland.de

{liane.reiners, schemer}@uni-mainz.de

{ashwin-geet.dsa, dominique.fohr, irina.illina}@loria.fr

angeliki.monnier@univ-lorraine.fr

* Equal contribution.



Motivation and Concept

Most annotated hate speech (HS) corpora are binary. This ignores the **complex** and **subjective** nature of hate. M-PHISIS is a novel HS corpus, based on a collection of user comments which...

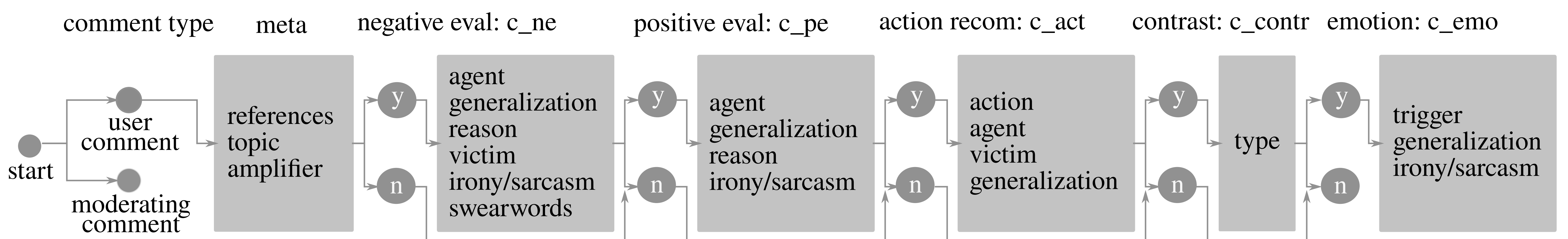
- ▶ are not sampled based on slur-words.
→ capture **implicit** hate.
- ▶ match **migration-related** regular expression keywords.
- ▶ are derived from a **diverse** set of outlets.
- ▶ are in **French and German**.
- ▶ are annotated based on **features**.
→ descriptors of various types of hate.
- ▶ are collected with **context**.

Data Collection

- ▶ Period: January–May 2020
- ▶ Content: News articles and comments from German and French **mainstream** and **fringe** news outlets.

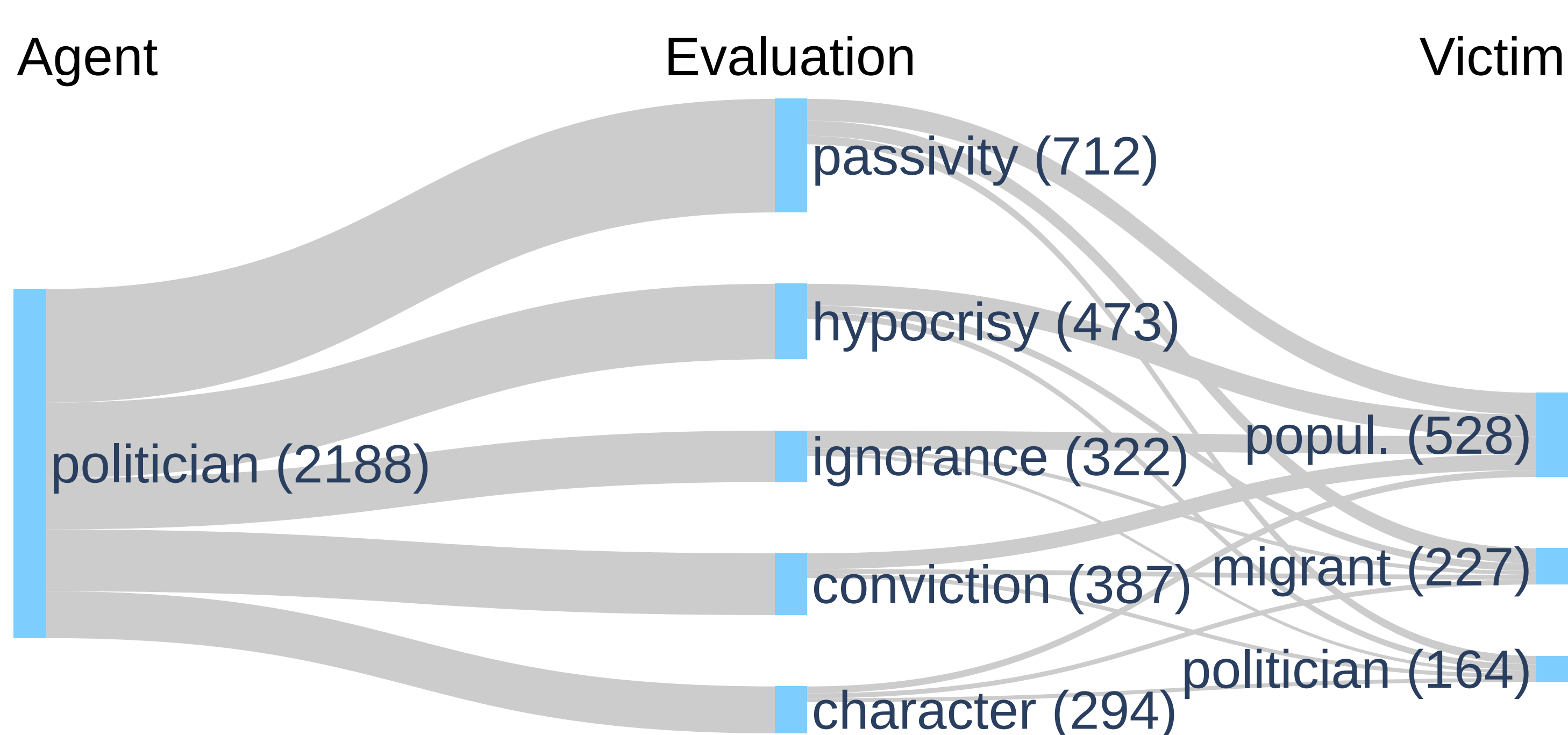
Outlet (<i>fr</i>)	#Art.	#Com.	Outlet (<i>de</i>)	#Art.	#Com.
France Info (<i>fi</i>)	20	618	Tagesschau (<i>ts</i>)	13	1,020
Figaro (<i>ff</i>)	19	1,056	Welt (<i>we</i>)	74	736
Le Monde (<i>lm</i>)	16	554	Zeit (<i>ze</i>)	15	999
Valeurs Actuelles (<i>va</i>)	30	614	Focus (<i>fo</i>)	13	962
AgoraVox (<i>av</i>)	11	369	Compact (<i>co</i>)	12	282
Riposte Laïque (<i>rl</i>)	35	1,435	Epoch Times (<i>et</i>)	2	75
–	–	–	Junge Freiheit (<i>jf</i>)	55	747
–	131	4,646	–	187	4,821

Overview of number of news articles (#Art.) and comments (#Com.) collected from both mainstream (top) and fringe (bottom) media outlets in French (*fr*) and German (*de*) for inclusion in the dataset.



Annotation pipeline for HS features in user comments across five modules (c_ne, c_pe, c_act, c_contr, c_emo). When a comment fulfills the requirement (y) of a module (e.g., contains negative evaluation for c_ne), follow-up categories (gray boxes) are annotated, otherwise (n) we skip to the next module.

Target Analysis



Agent-Evaluation-Victim sankey for the most common agent, *politician*, in the German portion of the dataset. We show the 5 most common evaluations/victims respectively.

→ **Decision makers** most negatively evaluated.

Contributions

- ▶ **Creation** of feature-based HS corpus on French and German news outlet comment.
- ▶ Resulting annotations with high levels of **inter-annotator agreement** ($0.77 \leq \kappa \leq 1$).
- ▶ **Target analysis** showing...
 - ▶ politicians more **criticised** than migrants.
 - ▶ French action recommendations more **radical**.
→ data-driven or actual radicalisation?
- ▶ **Domain analysis** showing...
 - ▶ **clear divide** between mainstream vs. fringe media comments
→ careful of domain-drift when training machine learning models!
- ▶ **Error analysis** and discussion showing limitations of dataset (e.g., some annotation irregularities).

<https://github.com/uds-lsv/mphasis>