

Computational Rhetoric in Social Media and Law

Dr. Jelena Mitrović

Leader of Research Group CAROLL, University of Passau

NLP Team Lead, Institute for AI R&D, Serbia

Overview of the talk

- Trends in NLP
- Computational treatment of Rhetoric
- Neural language models for abusive language detection – C-BiGRU and HateBERT
- Legal Knowledge Graphs and Legal Argumentation Mining

Language Technology

mostly solved

Spam detection

Let's go to Agra! ✓

Buy V1AGRA ... ✗

Part-of-speech (POS) tagging

ADJ ADJ NOUN VERB ADV

Colorless green ideas sleep furiously.


Named entity recognition (NER)


PERSON ORG LOC

Einstein met with UN officials in Princeton

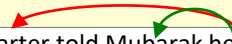
making good progress

Sentiment analysis

Best roast chicken in San Francisco! 

The waiter ignored us for 20 minutes. 


Coreference resolution

Carter told Mubarak he shouldn't run again. 


Word sense disambiguation WSD

I need new batteries for my *mouse*. 

Parsing


I can see Alcatraz from the window! 

Machine translation (MT)

第13届上海国际电影节开幕... 

The 13th Shanghai International Film Festival...

Information extraction (IE)

You're invited to our dinner party, Friday May 27 at 8:30 

Party
May 27
add 

still hard to do well

Question answering (QA)

Q. How effective is ibuprofen in reducing fever in patients with acute febrile illness?

Paraphrase

XYZ acquired ABC yesterday

ABC has been taken over by XYZ

Summarization

The Dow Jones is up

The S&P500 jumped

Housing prices rose



Economy is good

Dialog

Where is Citizen Kane playing in SF? 



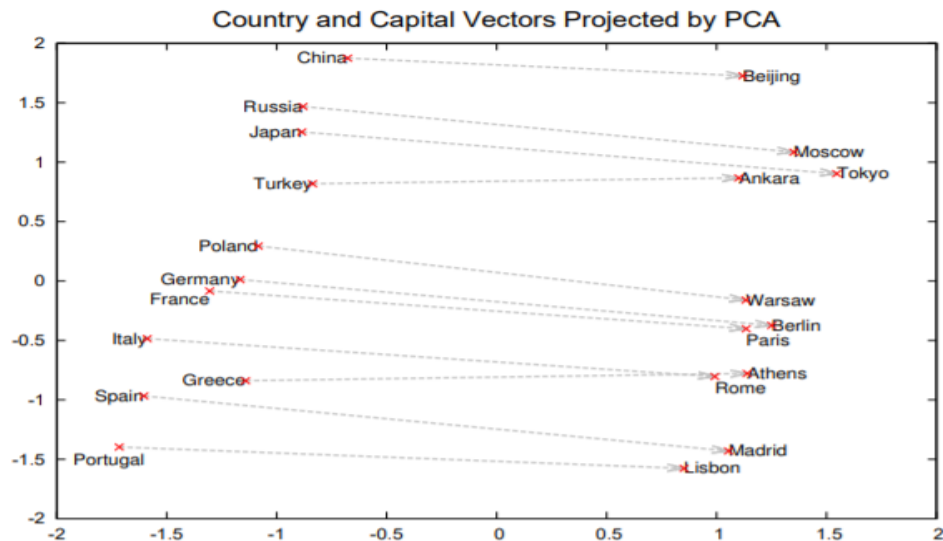
Castro Theatre at 7:30. Do you want a ticket?

* Adapted from slide by Prof. Manning

Recent Trends in NLP

Distributional semantics representations

- “You shall know the word by the company it keeps” (Firth 1957).
- **Word2Vec*** automatically organizes concepts and learns the relationships between them implicitly.



My BERT is better than your BERT*

- *Pre-trained* language models
- *Fine-tuned* depending on the task
- *Fast and dependable*



*Mikolov, T., Sutskever, I., Chen, K., Corrado, G. and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. NIPS.

**Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL-HLT*.

Universal solution to all NLP problems?

Language models is NOT all you need — retrieval is back – Insights from ACL 2020



Knowledge from language models is incomplete and inaccurate.

- LMs are **insensitive to negation** and are easily confused by **misprimed probes*** or related but incorrect answers**.
- Language models alone cannot deal with the **complexity of some NLP tasks!*****
- Hybrid approaches and **deeper understanding** of language.
- Going **back to the basics** – Incorporating elements of **Logic and Cognitive science** into our NLP systems.

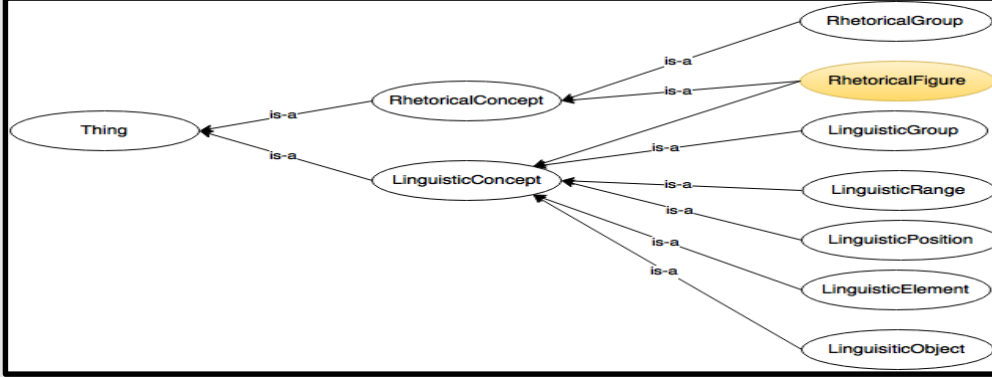
* Kassner, N. and Schutze, H. Negated and Misprimed Probes for Pretrained Language Models: Birds Can Talk, But Cannot Fly. In Proc. of ACL 2020

**Ettinger, A.. What BERT Is Not: Lessons from a New Suite of Psycholinguistic Diagnostics for Language Models. In Proc. of ACL 2020

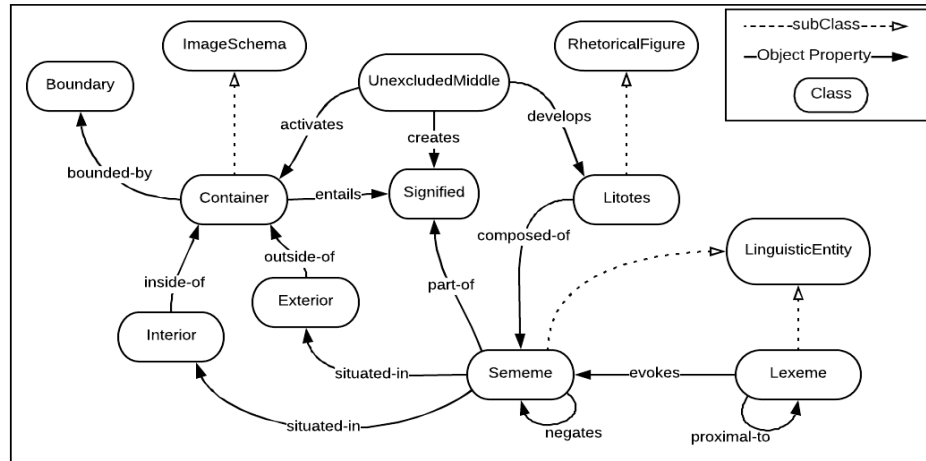
***Mladenović, M. and Mitrović, J. Ontology of Rhetorical Figures for Serbian. In LNAI. 8082/Springer-Verlag Berlin Heidelberg, 2013.

Semantics

Model Rhetorical Figures

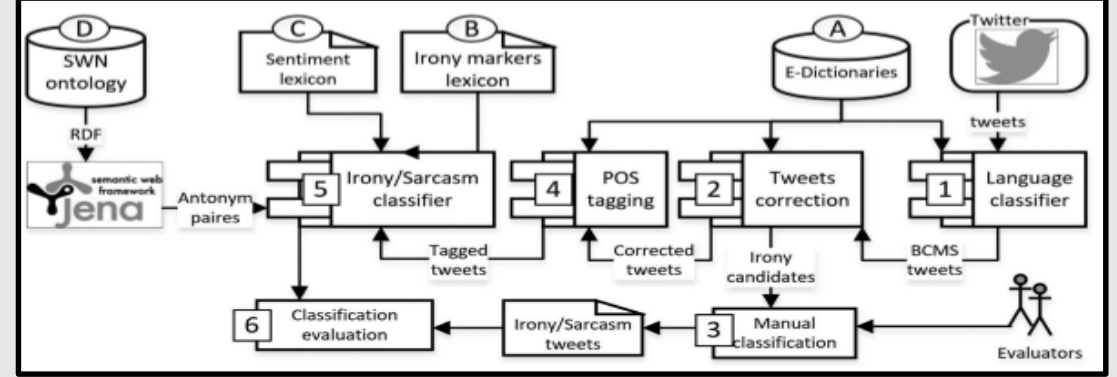


Rhetoric in Hate-Speech Detection

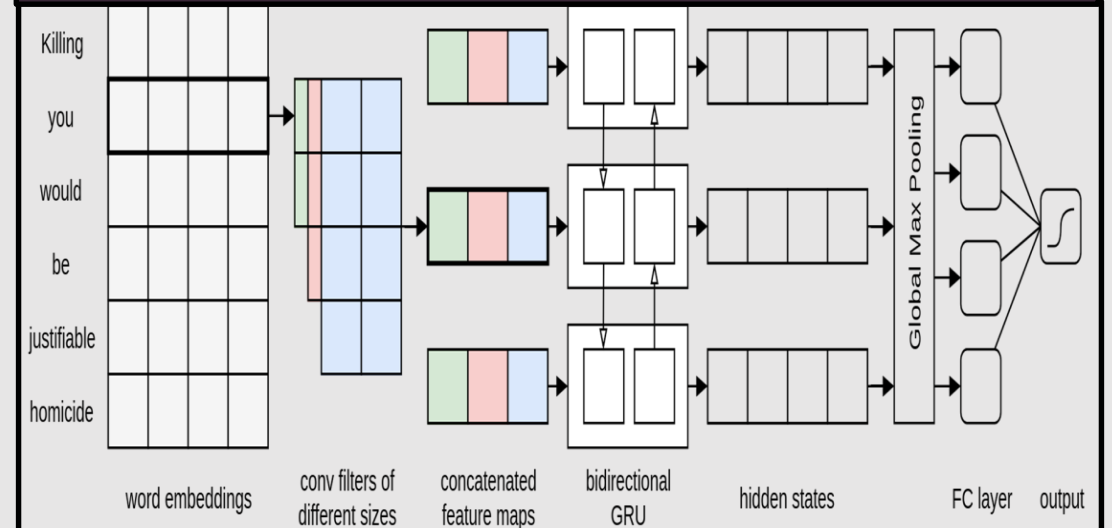


Statistics

Improve ML through Rhet. Modelling



DL for Hate-speech detection



What is Rhetoric?

“Faculty of discovering all the available means of persuasion in any given situation.”

Aristotle (Cooper 1960)

- Rhetoric is **omnipresent** – natural language is highly **nuanced**.
- Issues of **structure, style, and argumentation** – central importance to rhetorical theory.
- **Rhetorical figures** – particular importance for computational applications.

Ontology of Rhetorical Figures (RetFig)

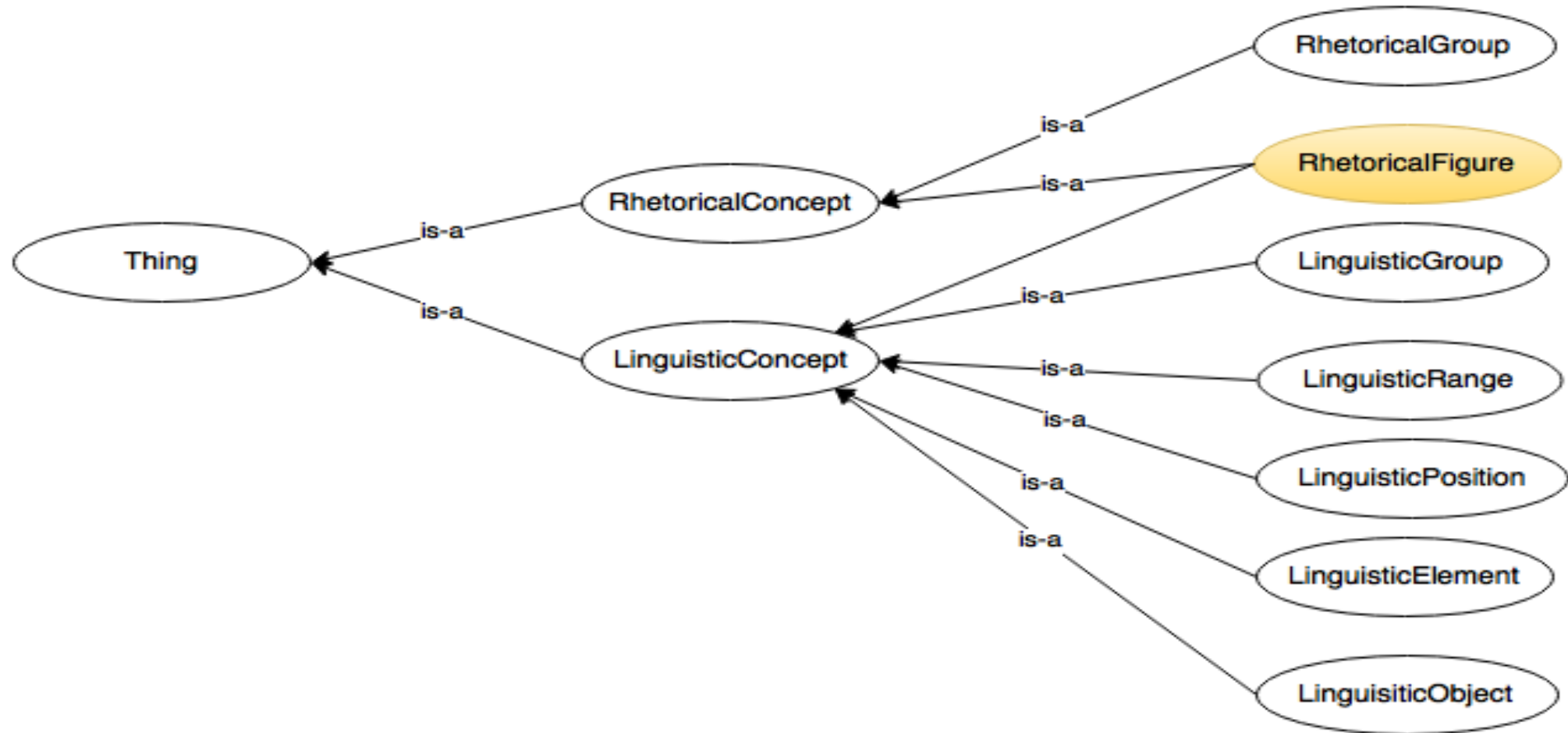
Separating semantic from syntactic similarity in word representations* a thorny problem.

- **RetFig** was built to overcome this problem by **formally modeling rhetorical figures**.
- **Linguistic domain**, descriptive, formal ontology for rhetorical figures in Serbian.
- **Unambiguously** describes and defines rhetorical figures.
- **OWL 2** using top-down modelling – formal representation of **98 Rhetorical figures***.

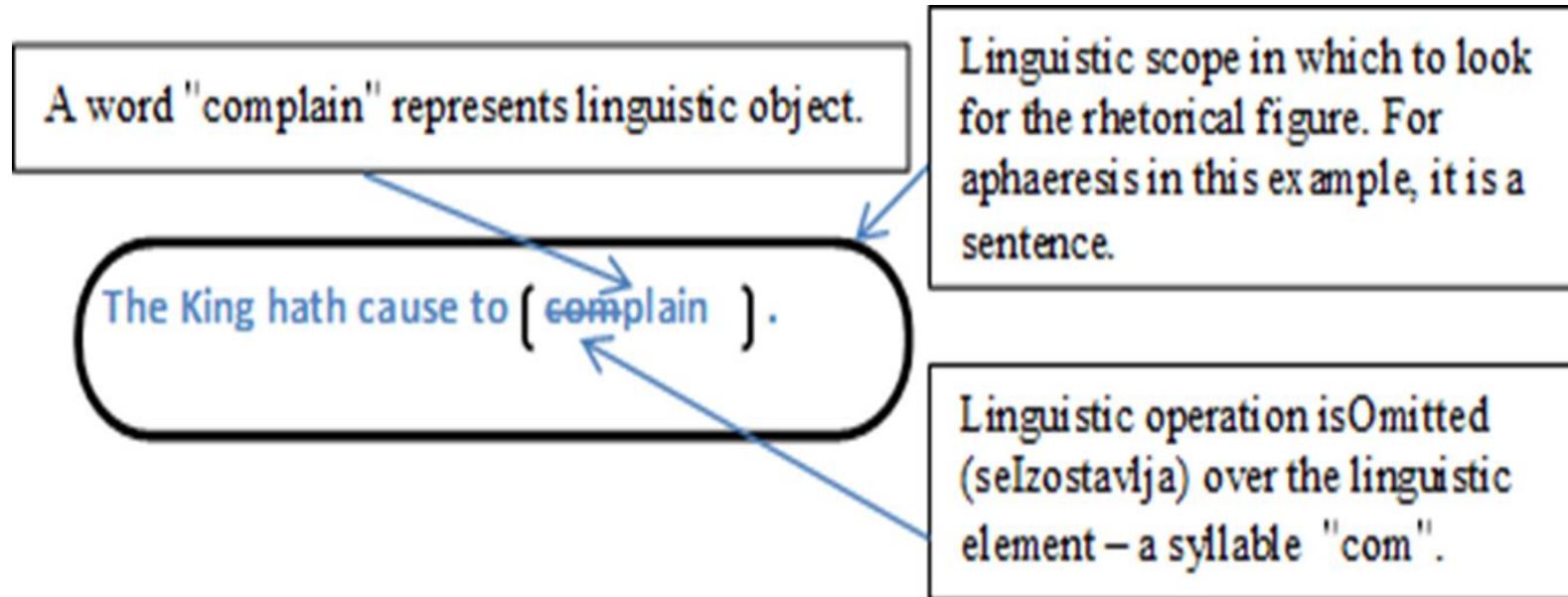
*J. Bjerva, R. Ostling, M. Han Veiga, J. Tiedemann and I. Augenstein, "What do Language Representations Really Represent?," *In CL*. arXiv:1901.02646, 2019.

Mladenović, M. and **Mitrović, J. Ontology of Rhetorical Figures for Serbian. In *Lecture Notes in Computer Science and Artificial Intelligence* 8082/Springer-Verlag Berlin Heidelberg, 2013, pp. 386-393

RetFig Structure

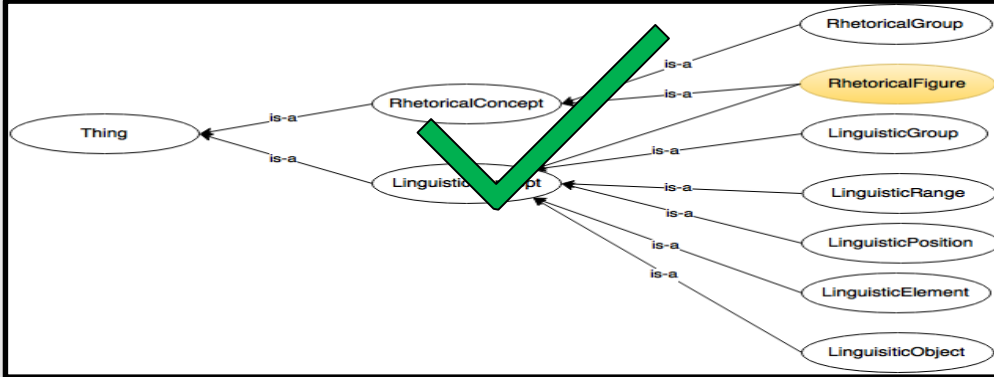


Detection of Aphaeresis



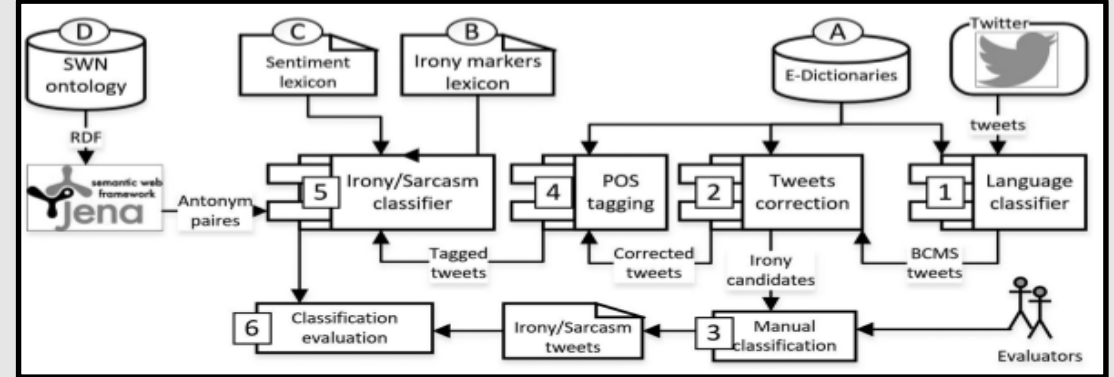
Semantics

Model Rhetorical Figures

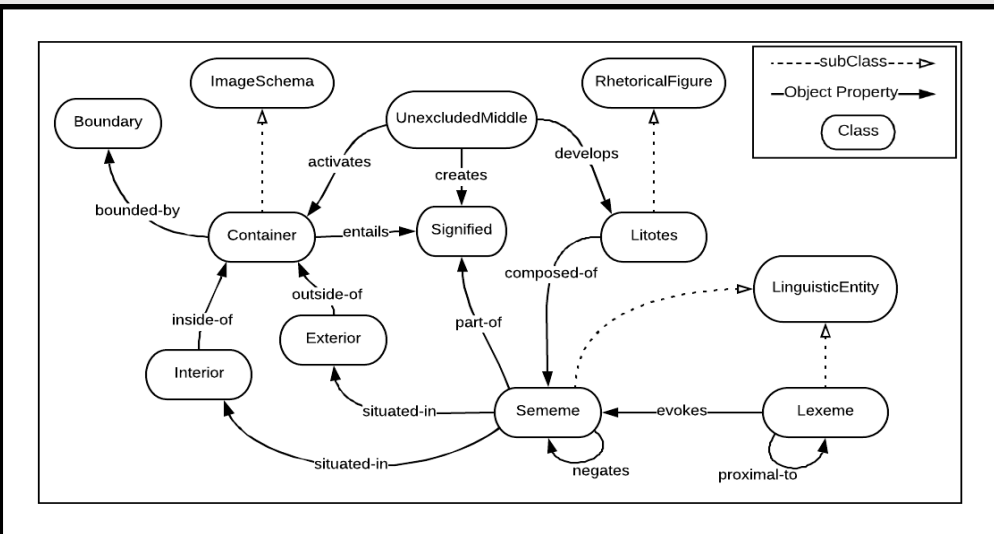


Statistics

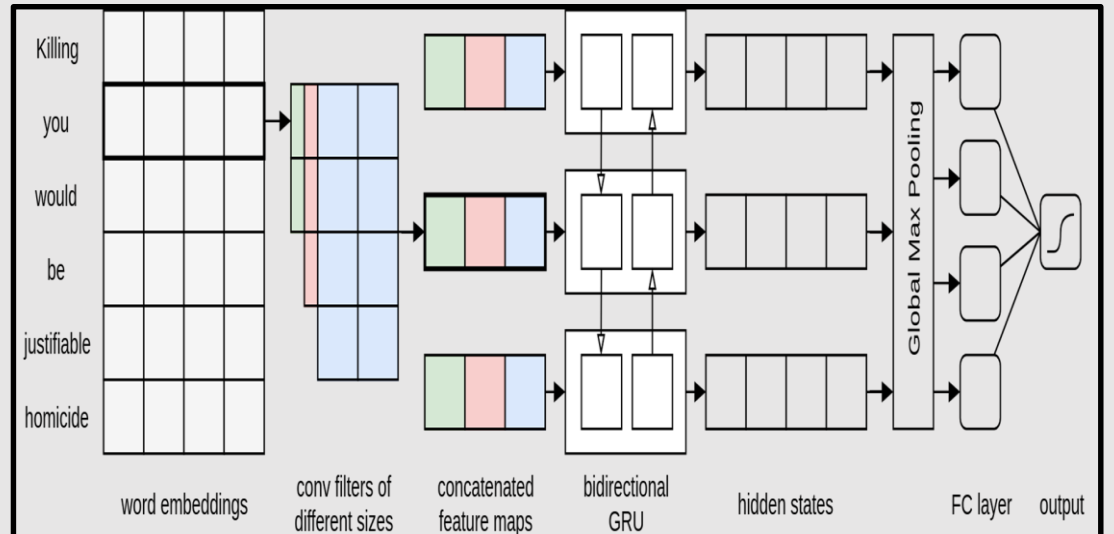
Improve ML through Rhet. Modelling



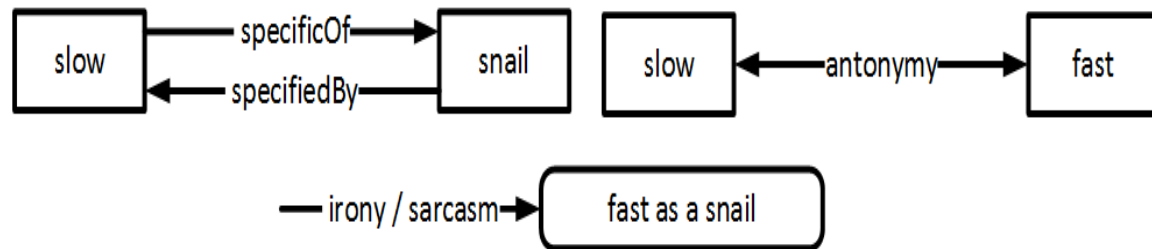
Rhetoric in Hate-Speech Detection



DL for Hate-speech detection



Detection of Irony and Sarcasm

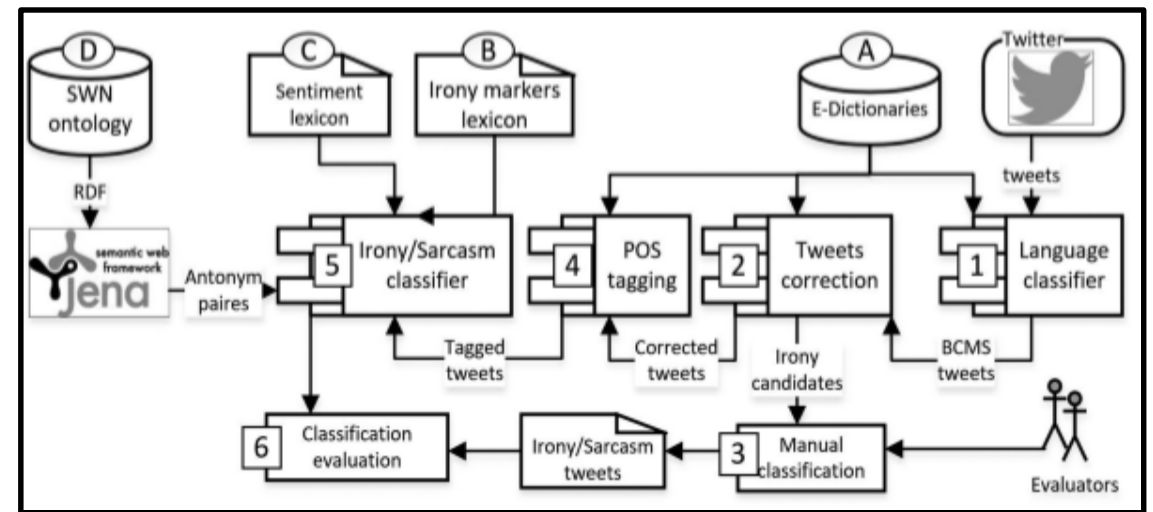


Slow as a Snail

Slow is *SpecificOf* Snail

Snail is *SpecifiedBy* Slow

Fast as a snail



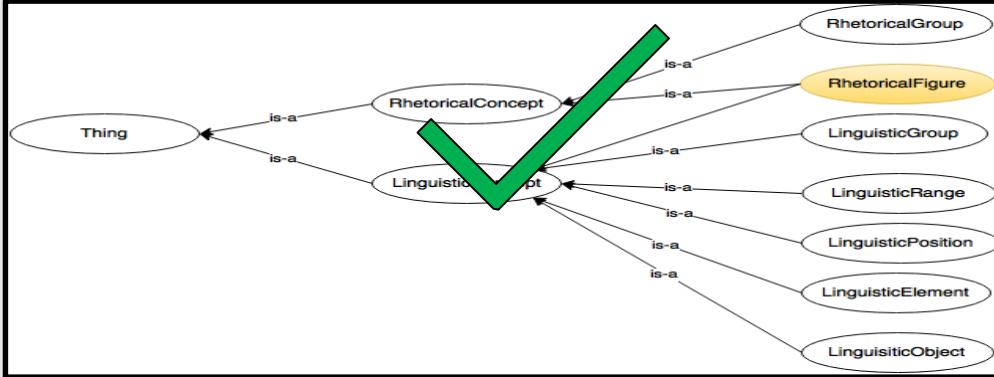
SWNOnTo – serialized from Serbian WordNet.

- **antonymous** pairs obtained using the reasoning rules over SWNOnTo
- **antonymous** pairs in which one member has positive sentiment polarity (PPR)
- polarity of positive sentiment words (PSP)
- ordered sequence of sentiment tags (OSA)
- Part-of-Speech tags of words (POS)
- and irony markers (punctuation marks etc.) (M)

	feature set	P	R	F1	acc
FS1	POS, OSA, M	0.504	0.530	0.517	0.817
FS2	R, OSA, M	0.605	0.486	0.539	0.845
FS3	PSP, POS, OSA, M	0.616	0.473	0.535	0.849
FS4	R, PSP, POS, OSA, M	0.658	0.458	0.540	0.856
FS5	PPR, POS, OSA, M	0.670	0.458	0.544	0.858
FS6	PPR, PSP, POS, OSA, M	0.686	0.451	0.544	0.861

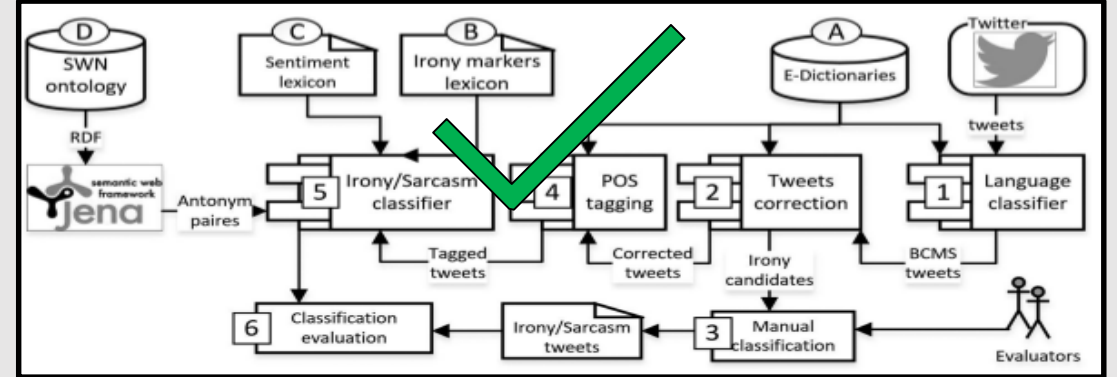
Semantics

Model Rhetorical Figures

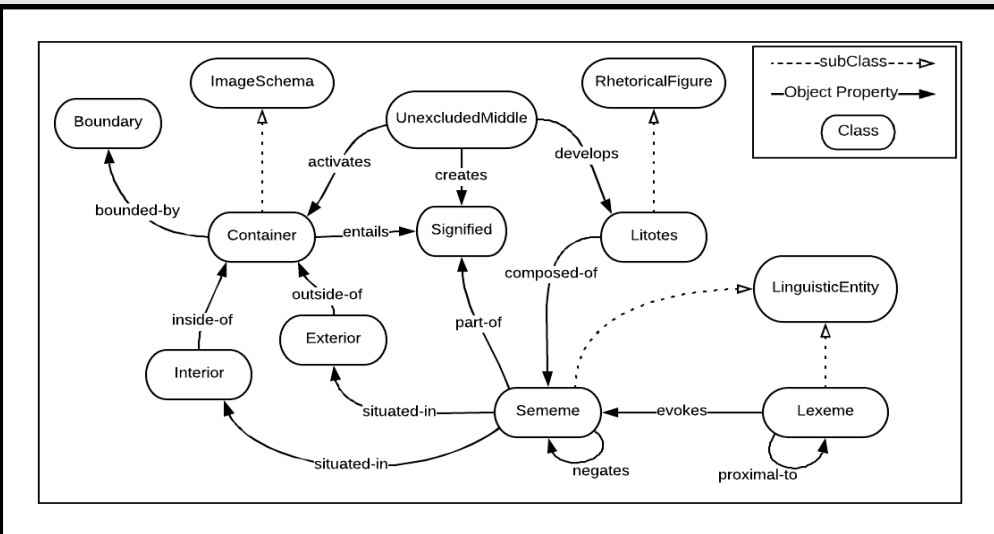


Statistics

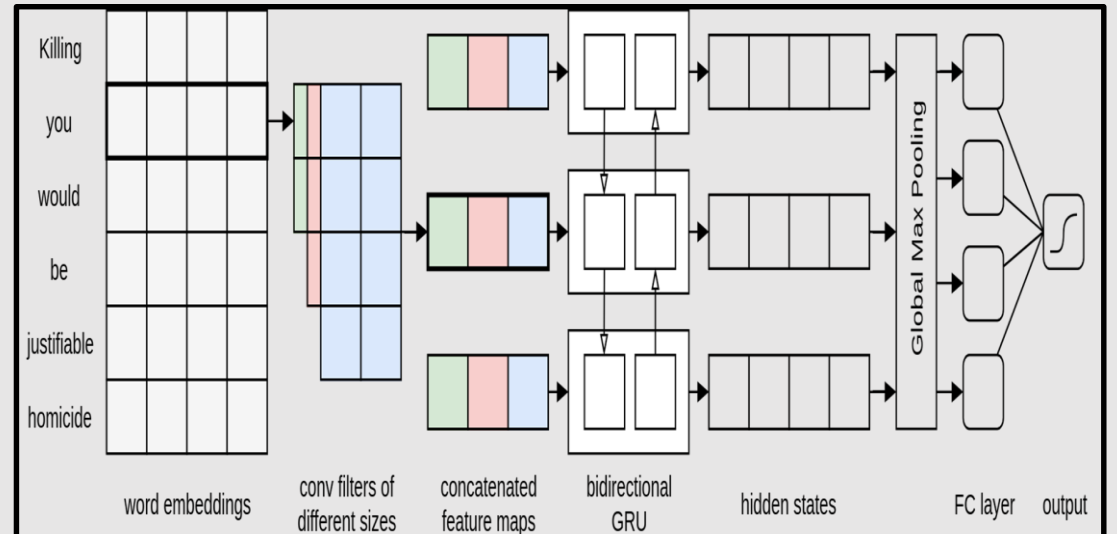
Improve ML through Rhet. Modelling



Rhetoric in Hate-Speech Detection



DL for Hate-speech detection



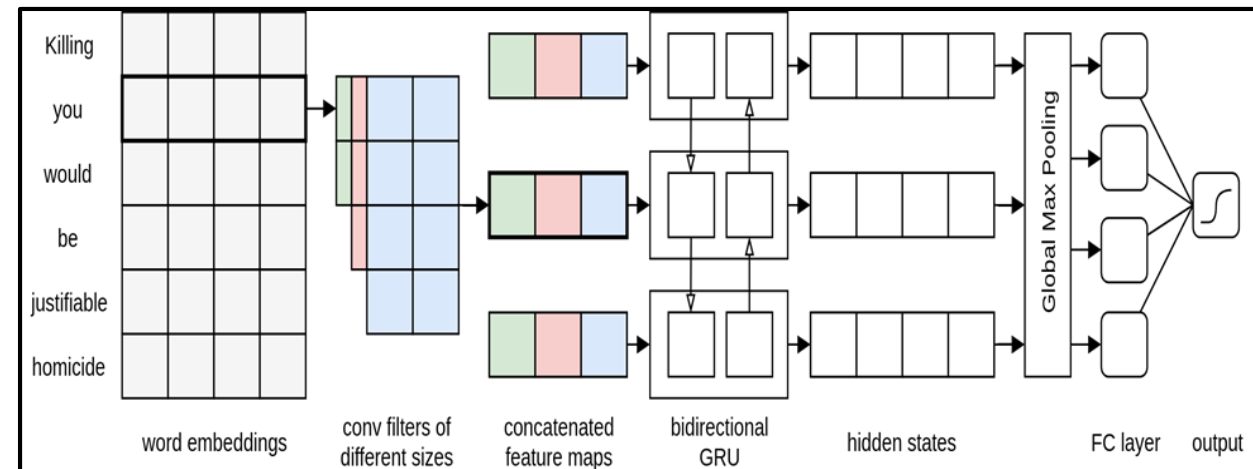
Challenges in Detecting Hate-Speech

- Systems that **deal with, analyze, and possibly prevent** hate-speech online?
- **The ever-changing landscape of hate** – users frequently change the rules of the game:
 - masking the offensive words
 - using different types of spelling
 - using entirely different words

Examples

- *Linux kernel mailing list* – Instead of *the f word*: **go hug yourself**
- “**She is not exactly a beauty queen.**” – D. Trump
- “**You can go and love yourself**” – pop song lyrics

- SemEval 2019 – **C-BiGRU system** – CNN + RNN – bidirectional GRU + word2vec – Scored in top 10%;
- **English, German, Danish, and Turkish**; Good initial results for **Hindi, Greek, Serbian and Hungarian**.



	Baseline		C-BiGRU	
	CV	gold	CV	gold
OLID	70.22%	-	76.28%	79.40%
GermEval	66.61%	66.78%	71.13%	72.41%

OFFENSIVE SPEECH DETECTOR

This is a demo of our system submitted to the [SEMEVAL-2019 Task 6](#) presented at NAACL - HLT 2019. Please find more details in our [paper](#).

Type text or copy a tweet from Twitter

You are not the smartest pea in the pod

Search for a random tweet using a Hashtag

Import a random tweet

Tweet is not offensive

I AGREE

I DON'T AGREE

Your feedback will help us improve our model.

Language

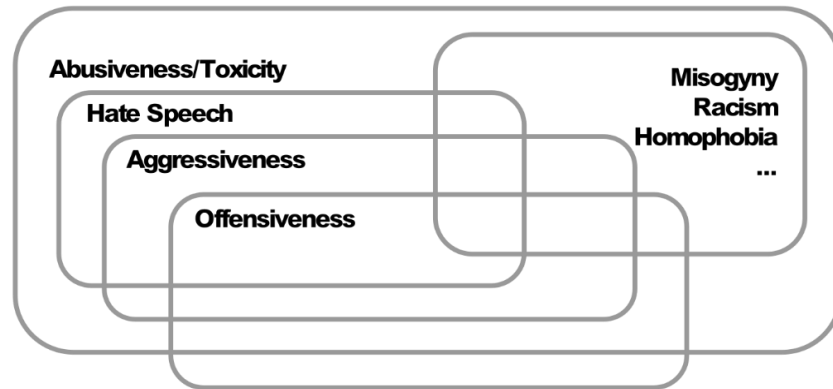
English

Analyze Text

Disclaimer: The contents of this demo are used exclusively for research purposes.

HateBERT

Can we re-train the BERT model to be sensitive to abusive language?



Overlapping phenomena, their treatment, and definitions*

- **HateBERT*** – retraining BERT on a large Reddit dataset (71 M tokens)
- further pre-training is a viable strategy
- **Huggingface***** code on one **Nvidia V100 GPU** – Re-train the English **BERT base-uncased** model – 50 days from March to May 2021
- We retrained for **100 epochs** - almost **2 million steps**, in **batches of 64 samples**, including **up to 512** sentence piece tokens.
- We used the **Adam optimizer with learning rate 5e-5**.

* F. Poletto, V. Basile, M. Sanguinetti, C. Bosco, and V. Patti. 2020. Resources and Benchmark Corpora for Hate Speech Detection: a Systematic Review. LREC, 54(3):1–47

** T. Caselli, V. Basile, J. Mitrović, M. Granitzer, (2021), HateBERT: Retraining BERT for Abusive Language Detection in English, WOH Workshop @ ACL-IJCNLP 2021

***pre-trained model available via the Huggingface Transformers library - <https://github.com/huggingface/transformers>

Intrinsic Evaluation of HateBERT

- The result is a shifted BERT model, **HateBERT** base-uncased, along two dimensions:
 1. language variety (i.e. social media)
 2. polarity(offense-, abuse-, and hate-oriented model)
- To verify that HateBERT has shifted towards abusive language phenomena we use the Masked Language Model (MLM) on five template sentences of the form

“**[someone]** is a(n)/ are **[MASK]**”

- This template can trigger biases in the model’s representations.

[someone] - “you”, “she”, “he”, “women”, “men”
- HateBERT consistently present profanities or abusive terms as mask fillers, while this very rarely occurs with the generic BERT.

BERT	HateBERT
“women”	
excluded (.075)	stu**d (.188)
encouraged (.032)	du*b (.128)
included (.027)	id***s (.075)

pre-trained model available via the Huggingface Transformers library -
<https://github.com/huggingface/transformers>

How to use from the 🤗/transformers library

 [Read model documentation](#)

```
from transformers import AutoTokenizer, AutoModelForMaskedLM

tokenizer = AutoTokenizer.from_pretrained("GroNLP/hateBERT")

model = AutoModelForMaskedLM.from_pretrained("GroNLP/hateBERT")
```

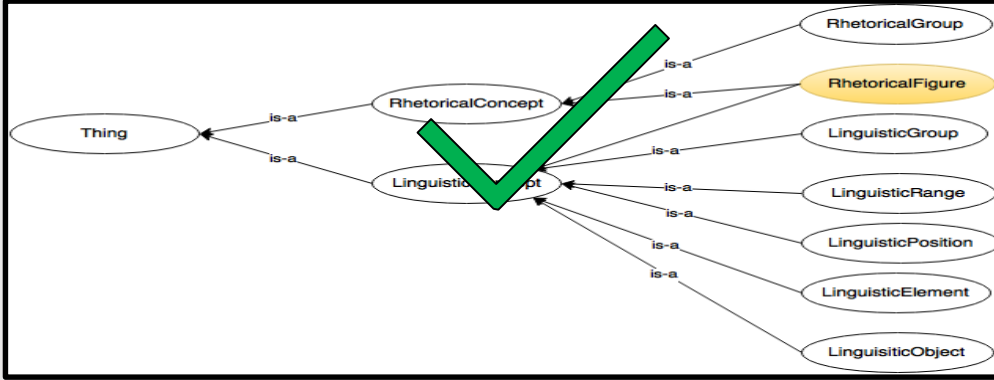
Or just clone the model repo

```
git lfs install
git clone https://huggingface.co/GroNLP/hateBERT

# if you want to clone without large files - just their pointers
# prepend your git clone with the following env var:
GIT_LFS_SKIP_SMUDGE=1
```

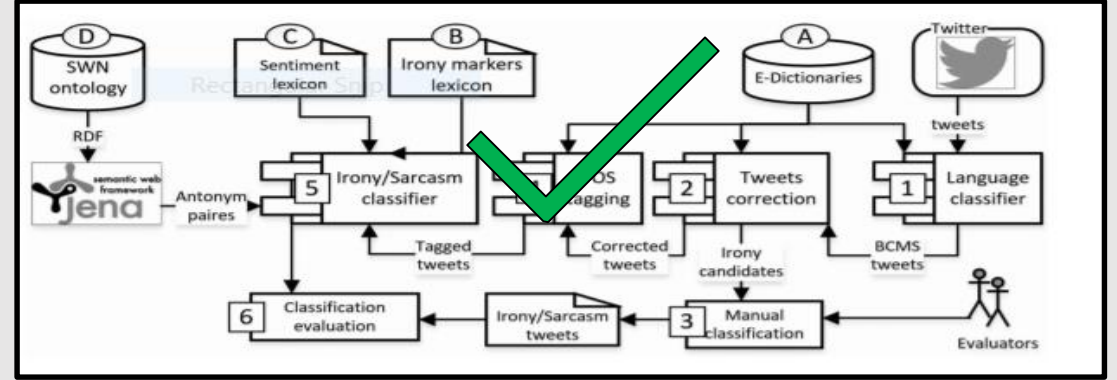
Semantics

Model Rhetorical Figures

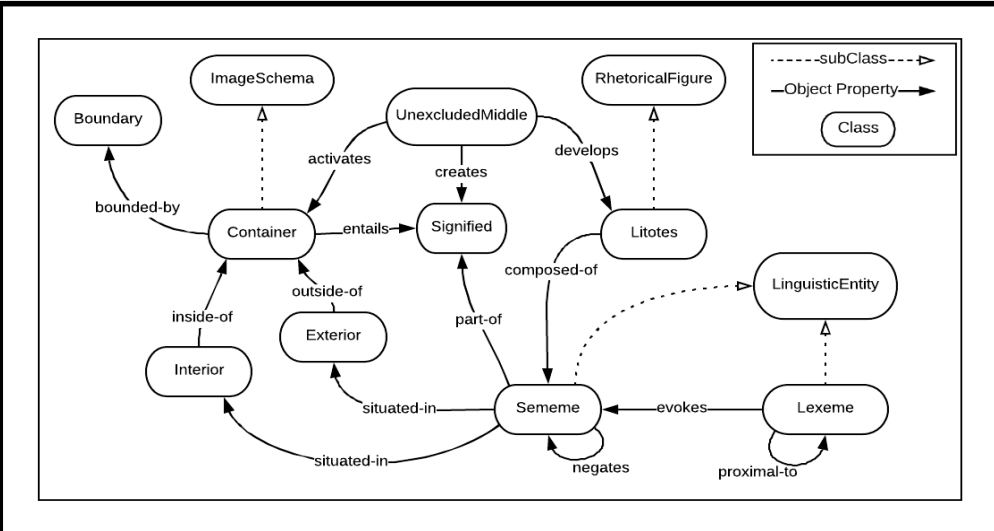


Statistics

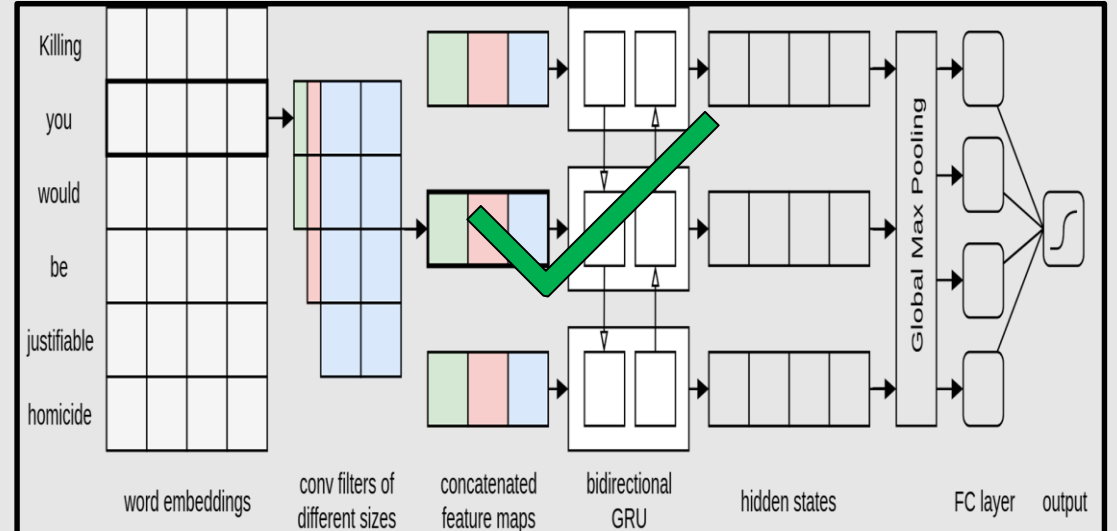
Improve ML through Rhet. Modelling



Rhetoric in Hate-Speech Detection



DL for Hate-speech detection



Ontological Modeling of Litotes

Examples

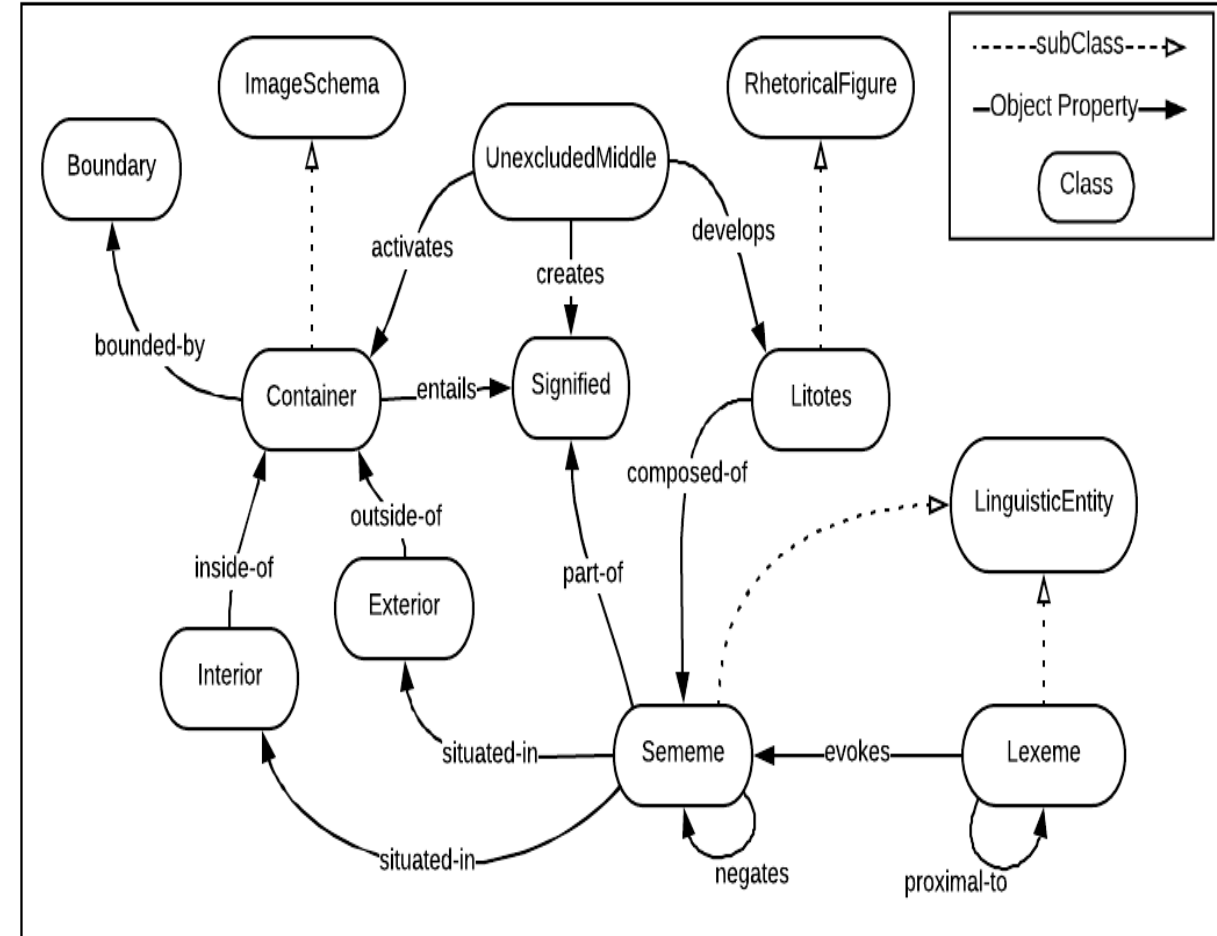
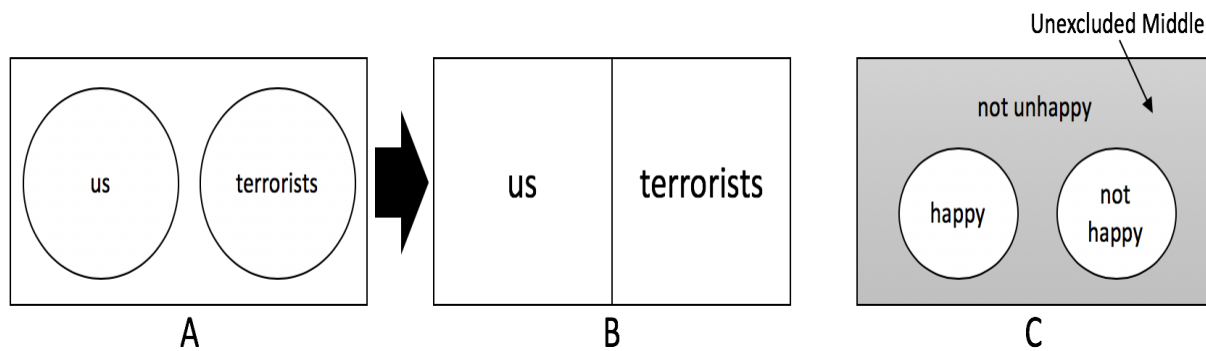
“You are not the smartest pea in the pod”

“Not the sharpest tool in the shed”

“A can short of a six-pack”

The Figure Litotes

- He is not the wisest man in the world (m. He is a fool) Henry Peacham’s *The Garden of Eloquence* (1593)



Composite Rhetorical Figures and Argumentation

“And so, my fellow Americans: ask not what your country can do for you – ask what you can do for your country.” – J.F.Kennedy

- **Antimetabole** – the entire statement
- **Mesodiplosis** – “can do for”
- **Anantithesis** – “Ask not . . .” and “Ask . . .”
- **Hyperbaton** (evoking archaic or biblical phrasing) – “Ask not”
- **Epanaphora** – “Ask not what” and “ask what”
- **Parison** – “what your country can do for you” and “what you can do for your country”

Sept 11th, I’m wearing a shirt that says “All Buildings Matter”.

- **Metonymy x 2 + metaphor** = a beautifully epitomized argument.
- Applying rhetorical insights and figuring out the figures that are **pillars of persuasiveness** can bring forth deeper understanding of the arguments and their interplay.

Stacking of Figures leads to “Semantic Conspiracy”

CAROLL - Computational Rhetoric in Social Media and Law



CAROLL

Use Cases:

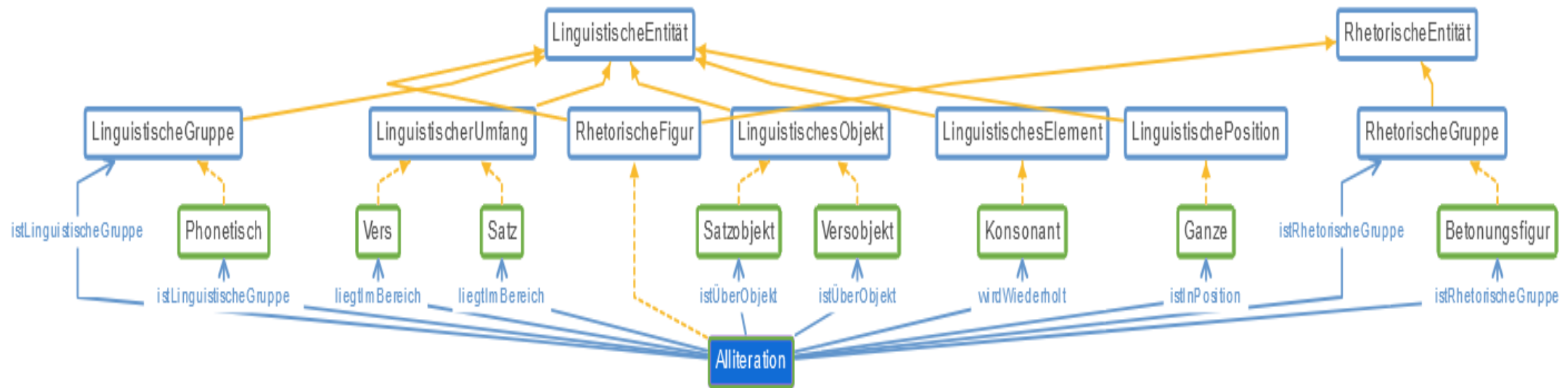
Social Media (**Communication studies**, collaboration with Prof. Hannah Schmid – Petri, **Sociology** – Prof. Anna Henkel)

- Communication patterns and saliency of arguments in the **climate change discourse**, **(anti)vaccination discourse**, **COVID-19 topics**.

Law (Collaboration with Prof. Kramer – Didactics of Law)

- **Law argumentation**, why are some arguments more or less persuasive?
- **Legal Knowledge graphs and ontological reasoning** (SUMO ontology)
- Hybrid approaches – ontological modeling of basic legal concepts + Deep learning

Ontology of Rhetorical Figures for German



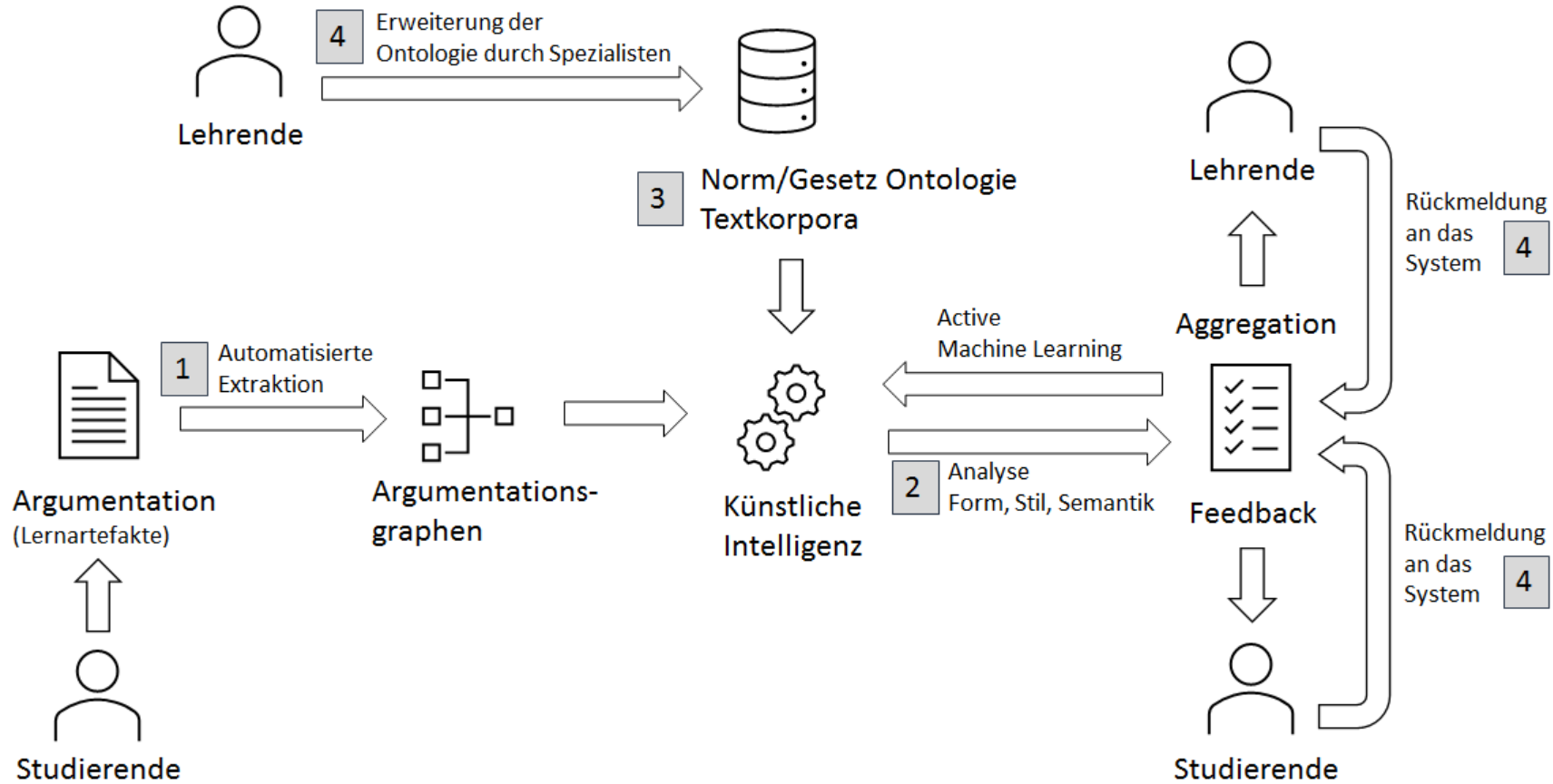
CAROLL + RHETORICON

- University of Waterloo, Canada – AI Institute and Computer Science Department and Composite effects/ Data augmentation techniques/ Political communication analysis
- Database of rhetorical figures in many languages

Deep Write Project

1. **Named Entity Recognition (NER)** using standard models.
2. **Argument annotation**: an annotation schema combined with NER Facts. Basis is Toulmin Schema, extended by didactical aspects.
3. **Entity Linking / Concept Extraction**: For a deeper semantic understanding – glossaries, lists of definitions or law/economy texts for concept extraction.
4. **Relationship Extraction**: deeper semantic dive - semantic graphs and proper reasoning.

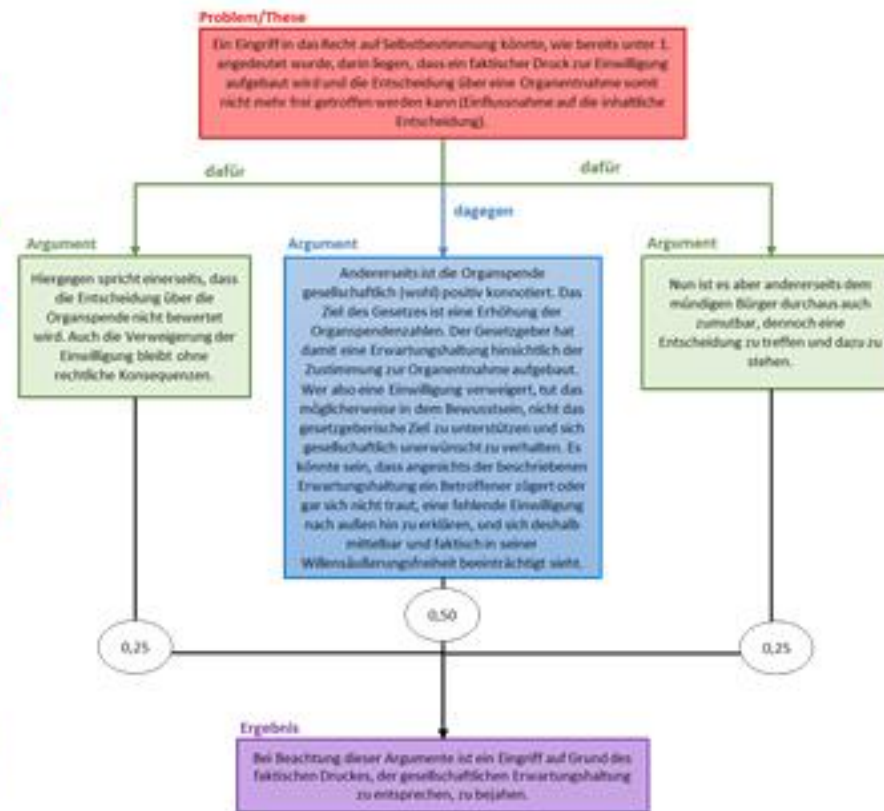
Deep Write Project



Deep Write Project

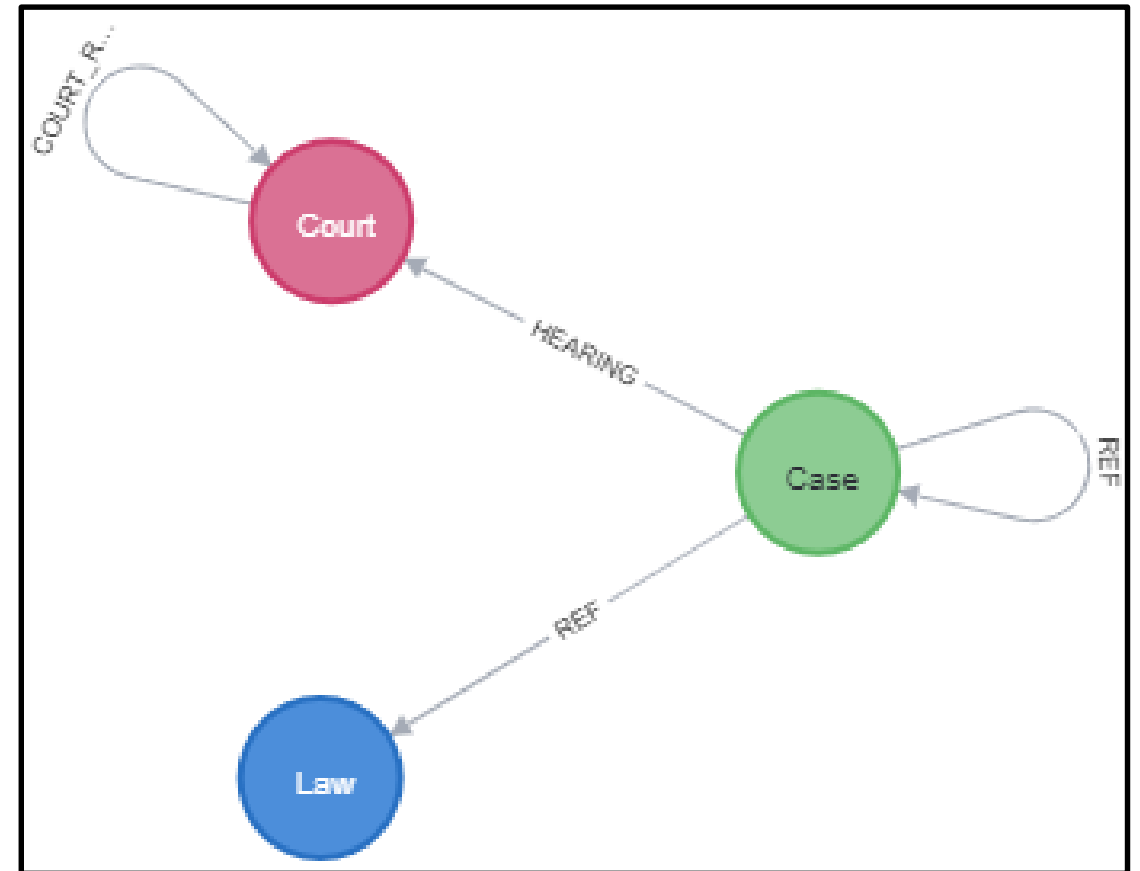
Ein Eingriff in das Recht auf Selbstbestimmung könnte, wie bereits unter 1. angedeutet wurde, darin liegen, dass ein faktischer Druck zur Einwilligung aufgebaut wird und die Entscheidung über eine Organentnahme somit nicht mehr frei getroffen werden kann (Einflussnahme auf die inhaltliche Entscheidung).

Hiergegen spricht einerseits, dass die Entscheidung über die Organspende nicht bewertet wird. Auch die Verweigerung der Einwilligung bleibt ohne rechtliche Konsequenzen. Andererseits ist die Organspende gesellschaftlich (wohl) positiv konnotiert. Das Ziel des Gesetzes ist eine Erhöhung der Organspendenzahlen. Der Gesetzgeber hat damit eine Erwartungshaltung hinsichtlich der Zustimmung zur Organentnahme aufgebaut. Wer also eine Einwilligung verweigert, tut das möglicherweise in dem Bewusstsein, nicht das gesetzgeberische Ziel zu unterstützen und sich gesellschaftlich unerwünscht zu verhalten. Es könnte sein, dass angesichts der beschriebenen Erwartungshaltung ein Betroffener zögert oder gar sich nicht traut, eine fehlende Einwilligung nach außen hin zu erklären, und sich deshalb mittelbar und faktisch in seiner Willensäußerungsfreiheit beeinträchtigt sieht. Nun ist es aber andererseits dem mündigen Bürger durchaus auch zumutbar, dennoch eine Entscheidung zu treffen und dazu zu stehen. Bei Beachtung dieser Argumente ist ein Eingriff auf Grund des faktischen Druckes, der gesellschaftlichen Erwartungshaltung zu entsprechen, zu bejahen.

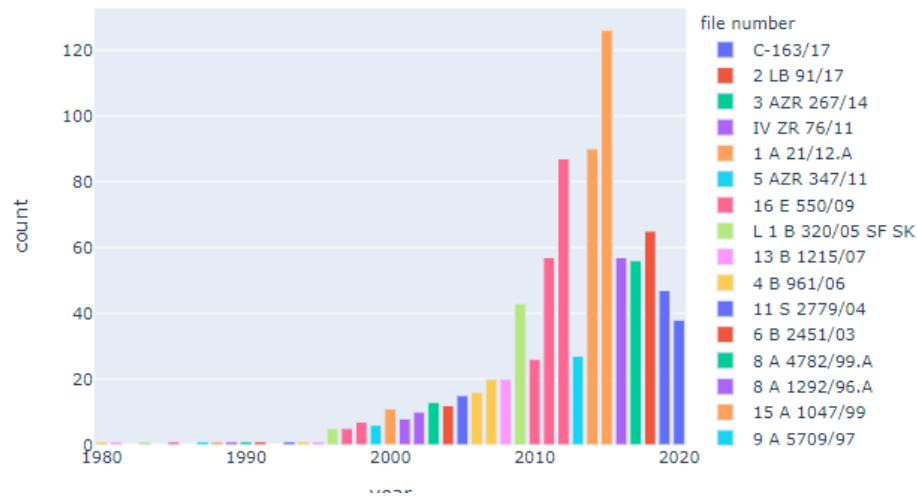


German Legal Citation Network

- 200.000 German court cases from all levels of appeal and jurisdiction, > 50.000 laws.
- References to court decisions and laws extracted from decision text of the court cases and added as links to the network.
- Network-based analysis techniques to support common legal information retrieval tasks - identification of important court decisions and laws and case similarity searches.

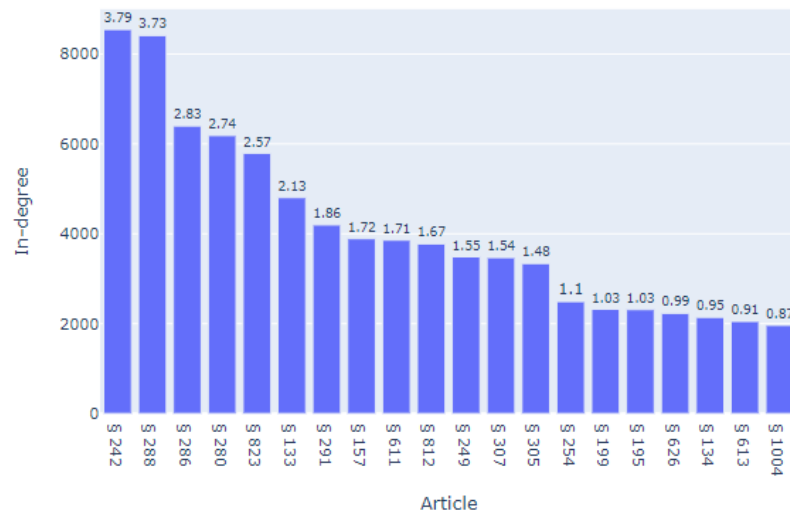


Most Cited Case by Year



- Court decisions with the highest amount of incoming citations per year.
- Some of the top cited cases seem to stay important for a consecutive year, but do not receive the same amount of attention after that

Most Cited Laws (top 20) in BGB



- Most cited laws of the German civil code.
- Article 242 referenced most often receiving 3.79% of all citations towards the civil code (BGB).
- "Performance in good faith: An obligor has a duty to perform according to the requirements of good faith, taking customary practice into consideration."

Thank you for your attention!

jelena.mitrovic@uni-passau.de

jelena.mitrovic.rs