Mining and Modeling Perspectives in NLP

Valerio Basile

CLASP seminar
January 31st, 2024

Outline

- Background and case study
  - (undesirable language)
- The Perspectivist Approach:
  - Mining
  - Modeling
  - Evaluating
- More case studies

Sentiment and Hate in NL

EVALITA

Evaluation of NLP and Speech Tools for Italian

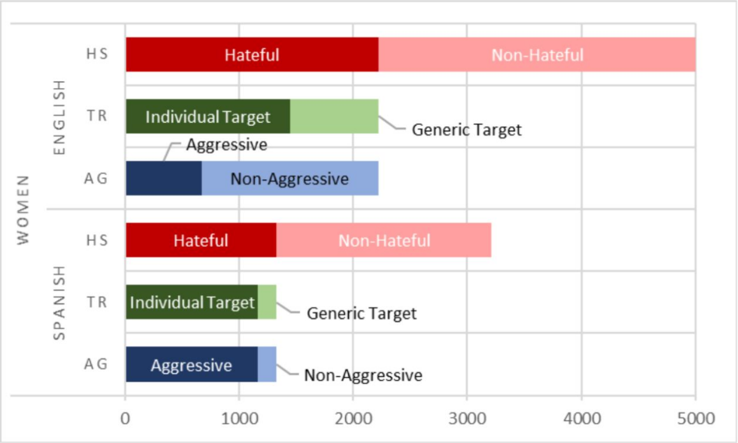
sentipolc / sentiment polarity classification

HOME DATA CONTACT

sentipolc@evalita 2016

SemEval-2019 Task 5: Multilingual Detection of Hate Speech Against Immigrants and Women in Twitter

Valerio Basile♢ Cristina Bosco♢ Elisabetta Fersini♡
Debora Nozza♡ Viviana Patti♢ Francisco Rangel♠♣
Paolo Rosso♣ Manuela Sanguinetti♢



Sentiment and Hate in NL

[HTML] Towards multidomain and multilingual abusive language detection: a survey
EW Pamungkas, V Basile, V Patti - Personal and Ubiquitous Computing, 2023 - Springer
… This study also outlines several **challenges** and open … **challenges** and frontiers in abusive content detection. They outlined several **challenges** of the abusive content detection **task** from …
Save Cite Cited by 24 Related articles All 7 versions

[HTML] A joint learning approach with knowledge injection for zero-shot cross-lingual hate speech detection
EW Pamungkas, V Basile, V Patti - Information Processing & Management, 2021 - Elsevier
… a number of **challenges** and issues in this particular **task**. One of the main **challenges** is … languages in the specific hate speech detection **task** also remain an open problem. However, …
Save Cite Cited by 60 Related articles All 3 versions Web of Science: 19

[HTML] A review on abusive content automatic detection: approaches, **challenges** and opportunities
B Alrashidi, A Jamal, I Khan, A Alkhathlan - PeerJ Computer Science, 2022 - peerj.com
… The automatic detection of abusive content is a **challenging task** due to disagreements on different abusive content definitions. Moreover, some content might be hateful to some …
Save Cite Cited by 2 Related articles All 8 versions

[HTML] Combating hate speech using an adaptive ensemble learning model with a case study on COVID-19
S Agarwal, CR Chowdary - Expert Systems with Applications, 2021 - Elsevier
… media platforms is an essential **task** that has not been solved efficiently despite multiple attempts by various researchers. It is a **challenging task** that involves identifying hateful content …
Save Cite Cited by 24 Related articles All 7 versions Web of Science: 13

[PDF] Hate Speech Detection through AlBERTo Italian Language Understanding Model.
M Polignano, P Basile, M De Gemmis, G Semeraro - NL4AI@ AI* IA, 2019 - academia.edu
… **task**, including the research areas of natural language processing (NLP), psychology, law, social sciences, and many more. The hate speech detection is a **challenging task** that gains …
Save Cite Cited by 25 Related articles All 4 versions

OK
we get it

Hate speech criteria: A modular approach to task-specific **hate speech definitions**
U Khurana, I Vermeulen, E Nalisnick… - arXiv preprint arXiv …, 2022 - arxiv.org
… procedure to **define hate speech**. As outlined above, we follow the view that **hate speech** is … propose the following criteria, represented in Figure 1, to **define** the scope of **hate speech**: …
Save Cite Cited by 9 Related articles All 8 versions

Latent **hatred**: A benchmark for understanding implicit **hate speech**
M ElSherief, C Ziems, D Muchlinski, V Anupindi… - arXiv preprint arXiv …, 2021 - arxiv.org
… We **define** implicit **hate speech** as outlined in the paper and ground this **definition** in a quote from Lee Atwater on how discourse can appeal to racists without sounding racist: "You start …
Save Cite Cited by 86 Related articles All 7 versions

Toxic, **hateful**, offensive or abusive? what are we really classifying? an empiric analysis of **hate speech** datasets
P Fortuna, J Soler, L Wanner - … of the 12th language resources and …, 2020 - aclanthology.org
… or incites violence, but limiting our **definition** only to such cases would exclude a large proportion of **hate speech**. Importantly, our **definition** does not include all instances of offensive …
Save Cite Cited by 90 Related articles All 5 versions

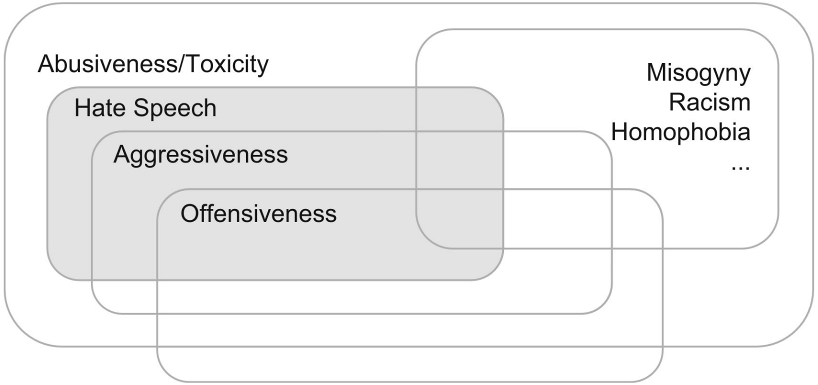
Online **hate speech**
AA Siegel - Social media and democracy: The state of the field …, 2020 - books.google.com
… scientific literature on how to **define** online **hate speech**. Legal **definitions** of **hate speech** are similarly murky. Governments are increasingly defining **hate speech** in their criminal codes …
Save Cite Cited by 109 Related articles All 7 versions

[HTML] Challenges of **hate speech** detection in social media: Data scarcity, and leveraging external resources
G Kovács, P Alonso, R Saini - SN Computer Science, 2021 - Springer
… One benefit of a universally agreed upon productive **definition** for **hate speech** could be important for more reliable annotation, with higher inter-annotator agreement [71]. For example …
Save Cite Cited by 83 Related articles All 5 versions

Sentiment and Hate in NL

Source	Definition
Code of Conduct, between EU and companies	"All conduct publicly inciting to violence or hatred directed against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or national or ethnic" [79]
ILGA	"Hate speech is public expressions which spread, incite, promote or justify hatred, discrimination or hostility toward a specific group. They contribute to a general climate of intolerance which in turn makes attacks more probable against those given groups." [42]
Nobata et al.	"Language which attacks or demeans a group based on race, ethnic origin, religion, disability, gender, age, disability, or sexual orientation/gender identity." [58]
Facebook	"Content that attacks people based on their actual or perceived race, ethnicity, national origin, religion, sex, gender or gender identity, sexual orientation, disability or disease is not allowed. We do, however, allow clear attempts at humor or satire that might otherwise be considered a possible threat or attack. This includes content that many people may find to be in bad taste (ex: jokes, stand-up comedy, popular song lyrics, etc.)." [28]
YouTube	"Hate speech refers to content that promotes violence or hatred against individuals or groups based on certain attributes, such as race or ethnic origin, religion, disability, gender, age, veteran status and sexual orientation/gender identity. There is a fine line between what is and what is not considered to be hate speech. For instance, it is generally okay to criticize a nation-state, but not okay to post malicious hateful comments about a group of people solely based on their ethnicity." [82]
Twitter	"Hateful conduct: You may not promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or disease." [72]

Survey by Fortuna & Nunes, 2018



Poletto et al. 2020

Sentiment and Hate in NL

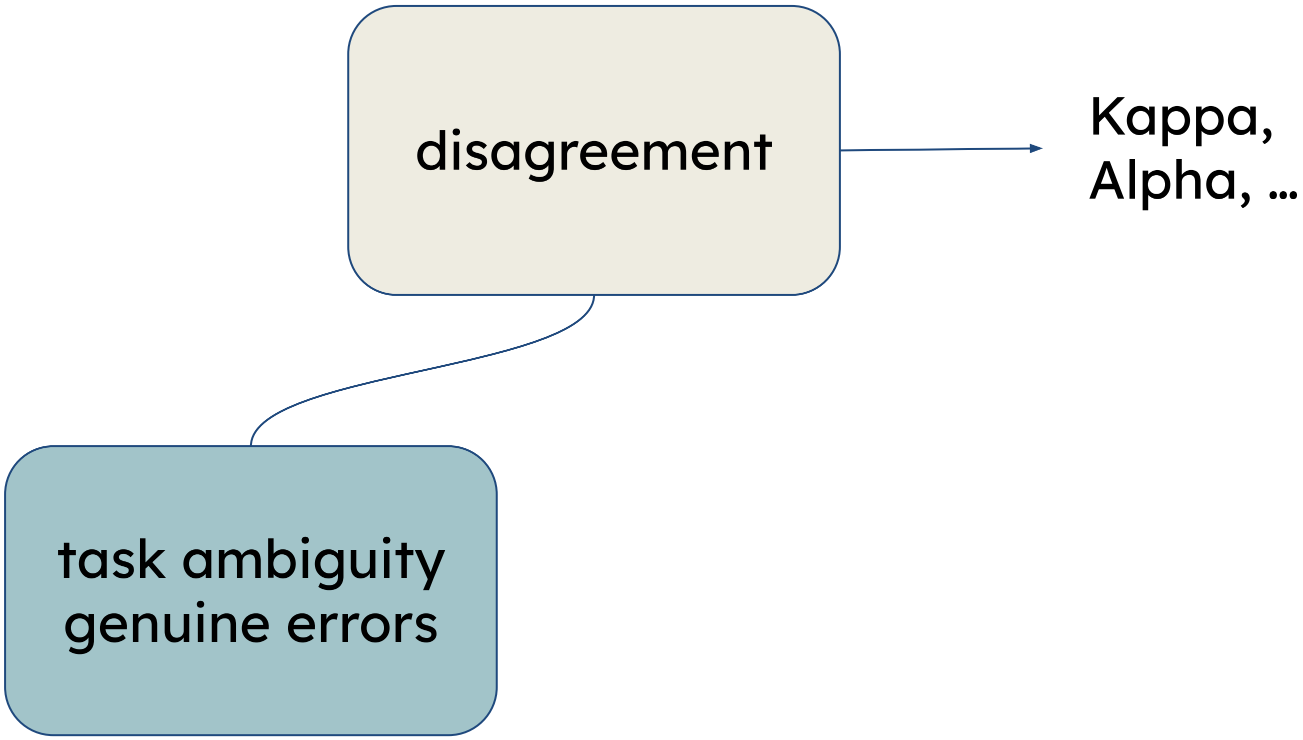
Annotating sentiment and irony is also difficult

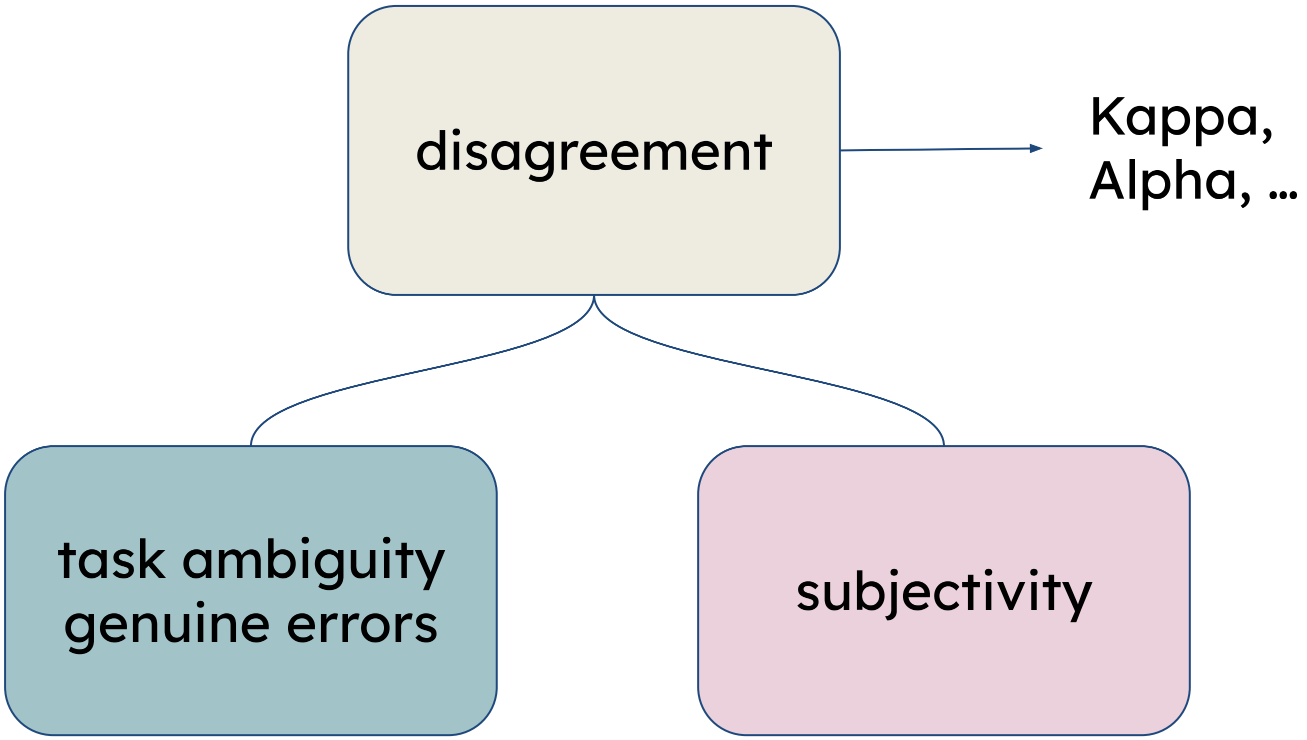
"Valerio, who is an expert on sarcasm?"
– P. Rosso
(who also heard me yell on Skype)

Follow-up study on SENTIPOLC (Basile et al., 2021)

"manual correction of part of the dataset has basically no impact on the final evaluation outcome"

disagreement → Kappa, Alpha, …


disagreement
Kappa,
Alpha, …
task ambiguity
genuine errors



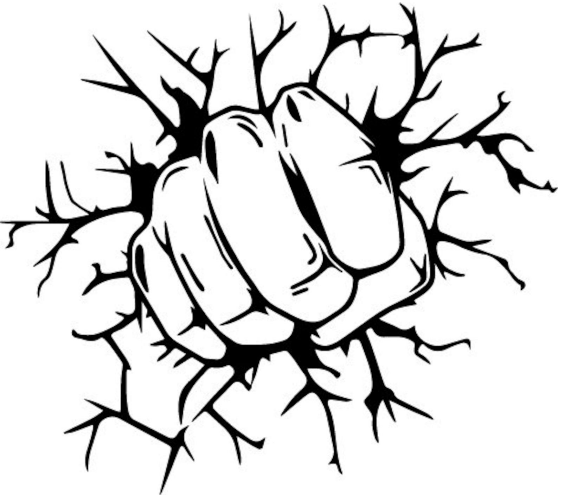
“We need to discuss disagreement in evaluation”
(Basile et al., 2021)

Thesis

Traditional NLP methodologies

do not scale

to subjective phenomena



THE PERSPECTIVIST DATA MANIFESTO

“It’s the End of the Gold Standard as we Know it.”
(Basile, 2020)

(Data) Perspectivism

No perspectivism

Collect annotation

Aggregate

Train & evaluate

Strong perspectivism

Collect annotation

Keep all of them!

Train & evaluate

Bring the extra knowledge all the way through the pipeline

Cabitza, Campagner, Basile (2023)
Toward a Perspectivist Turn in Ground Truthing for Predictive Computing – AAAI-23
(two years on ArXiv)

(Data) Perspectivism

Cabitza, Campagner, Basile (2023)
Toward a Perspectivist Turn in Ground Truthing for Predictive Computing – AAAI-23
(two years on ArXiv)

Strong perspectivism

Collect annotation

Keep all of them!

Train & evaluate

Bring the extra knowledge all the way through the pipeline

Why Perspectivism

- Minority opinions are left out
- Not every label is black and white
- Reasons behind model prediction

Context & Related

- *Aroyo and Welty 2015* → “No one truth”
- *Poesio, Plank, Hovy, et al.* →Soft Loss function
- *Gordon et al. 2021* → disagreement convolution
- *Sommerauer et al. 2020* → coherence-based evaluation
- *Cabitza et al. 2020* → Medical AI
- *Kennedy et al. 2020* → Psychiatry
- *Yun et al. 2021* → Image recognition
- *Dumitrach et al. 2015* → Relation extraction

Perspectives and Disagreement

Perspectives emerge from disagreement

Not all disagreement comes from different perspectives

- Ambiguity, task design, context…

Enough theory, show the numbers

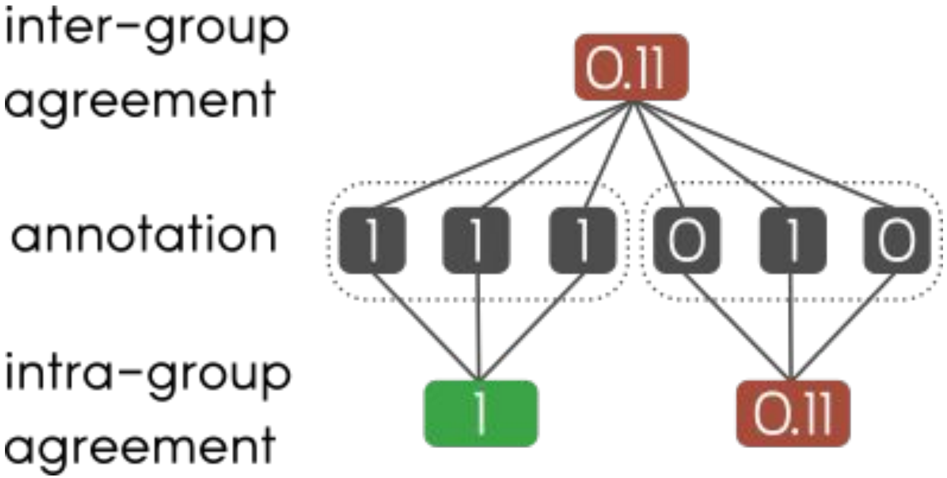
Modeling Annotator Perspectives

- Akhtar et al., AixIA 2019
  - Measure of polarization of annotation
- Akhtar et al., HCOMP 2020
  - Perspective-aware supervised models

PhD thesis on Hate Speech and
the role of victims in its analysis

Data from Twitter on Brexit annotated by
3 muslim immigrants in the UK
+ 3 western background

Polarization Index



	C2	C3	T1	T2	T3
C1	0.6	0.52	0.22	0.23	0.33
C2		0.52	0.16	0.18	0.26
C3			0.24	0.24	0.36
T1				0.69	0.52
T2					0.4

Data from Twitter on Brexit annotated by
3 muslim immigrants in the UK
+ 3 western background

P-based enhancement

- Method
  - Duplicate training instances with high P-index
  - Filter out training instances with low P-index
  - Test sets stay the same
- Data:
  - Sexism+Racism (Waseem et al. 2016)
  - Homophobia in Italian tweets (ACCEPT)

P-based enhancement

Sexism

Classifier	Accuracy	Precision	Recall	F1
SVM	95.11	**87.60**	71.60	78.74
SVM+P-max filter	95.13	86.40	73.01	79.11
SVM+replication	**95.27**	87.01	73.40	79.67
SVM+P-max filter+replication	**95.27**	86.60	**74.01**	**79.83**

Racism

Classifier	Accuracy	Precision	Recall	F1
SVM	98.55	55.40	11.01	18.40
SVM+P-max filter	98.58	59.01	12.01	19.88
SVM+replication	**98.61**	**70.01**	19.60	29.49
SVM+P-max filter+replication	**98.61**	69.80	**19.80**	**29.74**

Homophobia

Classifier	Accuracy	Precision	Recall	F1
SVM	**88.81**	61.01	11.40	19.02
SVM+P-max filter	**88.81**	**63.60**	13.60	22.30
SVM+replication	86.55	50.40	18.40	26.83
SVM+P-max filter+replication	87.63	47.90	**26.20**	**33.67**

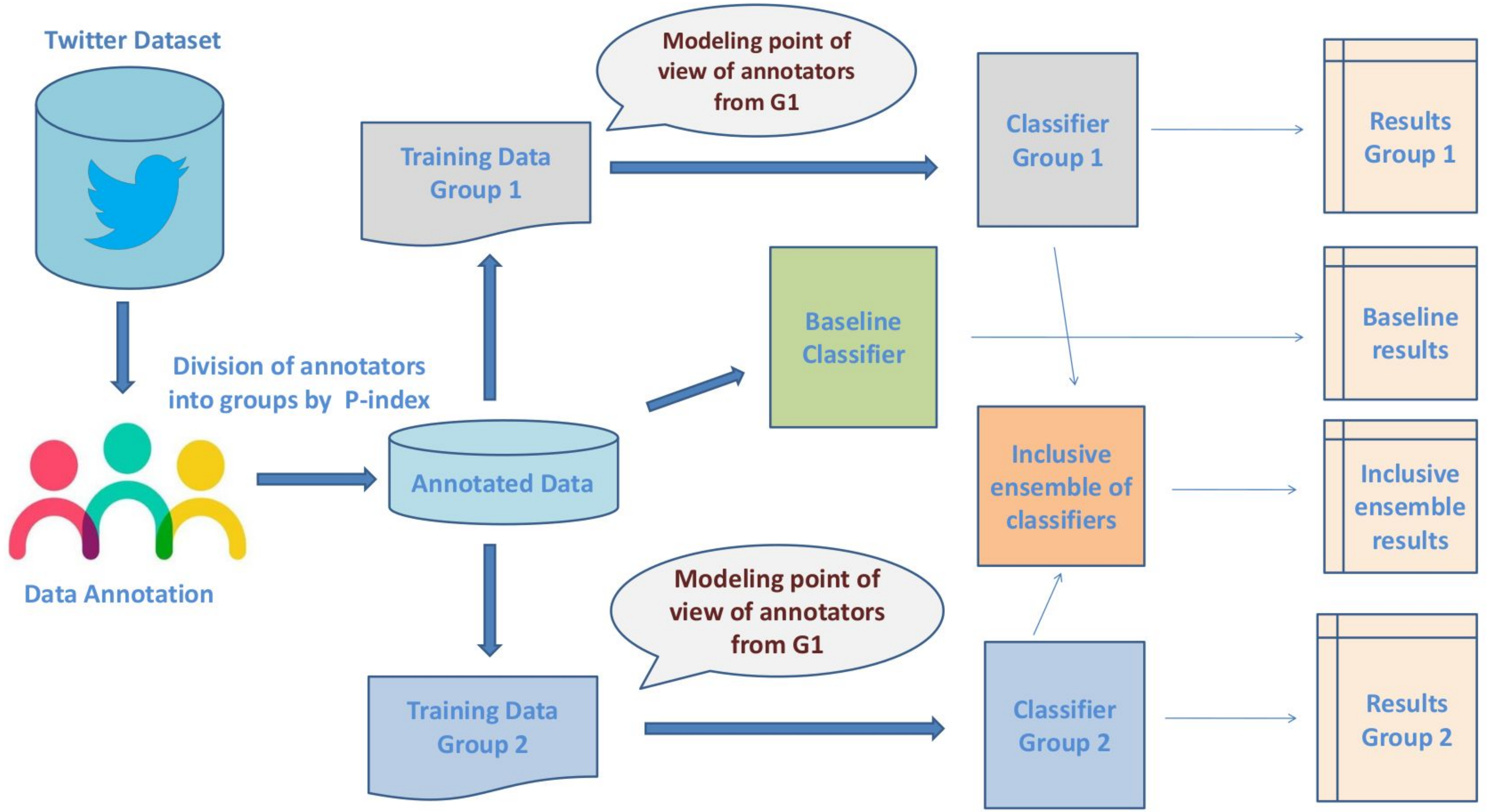
P-based exploration

Ranking the instances of a dataset by P-index, the most polarizing tweets emerge naturally at the top of the list.

- Intersectional issues →race-related remarks in sexist tweets
- Inappropriate jokes
- Polarizing topics → education & LGBT+

Modeling Perspectives

- P-based clustering of annotators
- Compile two different training sets
- Train two perspective-aware models
- Bonus: inclusive ensemble



Twitter Dataset
Division of annotators into groups by P-index
Data Annotation
Annotated Data
Training Data Group 1
Modeling point of view of annotators from G1
Classifier Group 1
Results Group 1
Baseline Classifier
Baseline results
Inclusive ensemble of classifiers
Inclusive ensemble results
Training Data Group 2
Modeling point of view of annotators from G1
Classifier Group 2
Results Group 2

Modeling Perspectives

notice the asymmetry

Sexism

Classifier	Prec. (1)	Rec (1)	F1 (1)
Baseline	**.812** (.034)	.711 (.044)	.756 (.015)
Group 1	.745 (.048)	.764 (.045)	.752 (.008)
Group 2	.720 (.019)	.907 (.018)	**.802** (.008)
Inclusive	.665 (.033)	**.939** (.009)	.778 (.020)

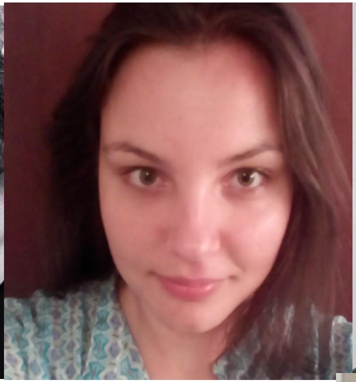
Racism

Classifier	Prec. (1)	Rec. (1)	F1 (1)
Baseline	**.852** (.159)	.194 (.059)	.312 (.085)
Group 1	.654 (.154)	.424 (.140)	.488 (.104)
Group 2	.571 (.175)	.412 (.198)	.419 (.076)
Inclusive	.532 (.141)	**.612** (.136)	**.542** (.091)

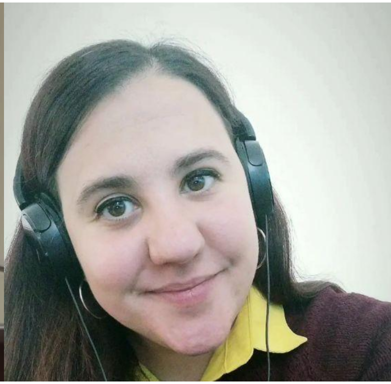
Homophobia

Classifier	Prec. (1)	Rec. (1)	F1 (1)
Baseline	.415 (.146)	.231 (.079)	.273 (.038)
Group 1	.302 (.038)	.471 (.154)	.355 (.040)
Group 2	**.531** (.112)	.178 (.031)	.262 (.033)
Inclusive	.302 (.039)	**.502** (.142)	**.367** (.035)

EPIC: English Perspectivist Irony Corpus





Thanks,
Amazon Alexa!

EPIC: English Perspectivist Irony Corpus

Sources:

→ Reddit

→ Twitter

Time window:

January 2020 - June 2021

~300 text/reply pairs

* 5 varieties

* 2 sources

$\left.\begin{array}{l}\text{~300 text/reply pairs}\\ \text{* 5 varieties}\\ \text{* 2 sources}\end{array}\right\}$ 3,000

Language	**Variety**
English	United Kingdom
	United States
	Ireland
	Australia
	India

Annotation

- ~15 annotators per single variety = 74
  - 200 texts per annotator (with attention-checks)
  - avg of 5 annotations per text
- balanced sets of annotators with respect to:
  - self-declared gender
  - across country of residence of annotators

  → they annotate instances from all varieties of the language, not just the one they speak

Prolific

Annotation Task

Message

My youngest brother and his wife married on Feb 29th. He became my hero.
Today is their fifth anniversary.

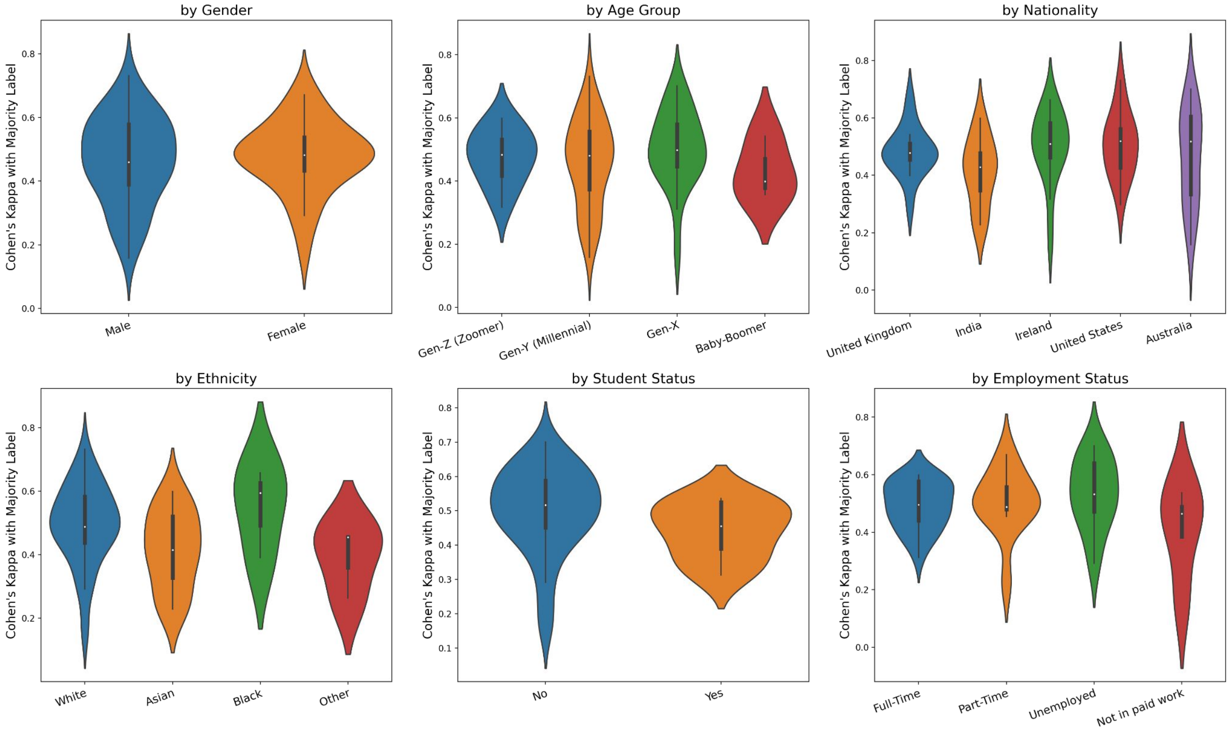
Reply

Means it's been 20 years since their marriage?

Is the **reply** ironic?

Ironic | Not ironic

Distribution of IAA among Perspectives



Does anyone see the irony here?

Frenda et al., NLPerspectives 2023

Modelling Perspectives

model	GOLD TEST SET			PERSPECTIVE-BASED TEST SET				
	F1-score	Confidence		F1-score	Confidence		Δ% Confidence	
		std	*avg*		*std*	*avg*	*std*	*avg*
non-perspectivist	0.681	0.301	0.509	–	–	–		
Fem-persp	0.590	0.239	0.621	0.538	0.234	0.644	-2.09	3.70
Male-persp	0.620	0.274	0.582	0.613	0.267	0.585	-2.55	0.52
Boomers-persp	0.539	0.290	0.502	0.484	0.303	0.532	4.48	5.98
GenX-persp	0.516	0.269	0.603	0.483	0.261	0.612	-2.97	1.49
GenY-persp	0.611	0.265	0.255	0.574	0.259	0.245	-2.26	-3.92
GenZ-persp	0.574	0.234	0.367	0.601	0.240	0.352	2.56	-4.09
Au-persp	0.497	0.173	0.748	0.435	0.165	0.746	-4.62	-0.27
US-persp	0.516	0.259	0.580	0.461	0.262	0.583	1.16	0.52
Ir-persp	0.535	0.273	0.319	0.521	0.293	0.340	7.33	6.58
In-persp	0.466	0.232	0.666	0.432	0.210	0.708	-9.48	6.31
UK-persp	0.507	0.255	0.612	0.533	0.251	0.630	-1.57	2.94

Perspective-aware models take a decision **with less uncertainty** than non-perspectivist models

Perspective-aware models are **more confident** when they are tested on a set representative of their perspective

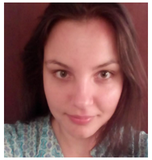
Modelling Perspectives

Confidence-based Perspective-aware Ensemble

Also cross domain (EPIC vs. SemEval)

Also on hate speech (MHS corpus)

		negative class			positive class			macro-average			
LLM	model	prec.	rec.	F1	prec.	rec.	F1	prec.	rec.	F1	Acc.
BERT	aggr	.873	.711	.780	.342	.581	.425	.608	.646	.603	.685
	NPS-vote	.880	.701	.780	.339	.614	.436	.610	.657	.608	.684
	NPS-conf	.873	.737	.799	.350	.568	.432	.611	.652	.615	.703
	RA-vote	.897	.590	.711	.307	.728	.431	.602	.659	.571	.618
	RA-conf	.897	.596	.715	.307	.722	.431	.602	.659	.573	.621
	maj	.868	.745	.801	.349	.543	.423	.608	.644	.612	.705
	vote	.876	.701	.779	.335	.603	.430	.606	.652	.605	.682
	conf	.875	.743	.803	.358	.571	.439	.616	.657	**.621**	.709
DistilBERT	aggr	.894	.658	.757	.332	.685	.447	.613	.671	.602	.663
	RA-vote	.891	.645	.748	.323	.683	.439	.607	.664	.593	.652
	RA-conf	.889	.645	.747	.321	.676	.435	.605	.660	.591	.651
	NPS-vote	.873	.689	.770	.324	.597	.420	.598	.643	.595	.671
	NPS-conf	.889	.645	.748	.321	.676	.436	.605	.661	.592	.652
	maj	.877	.712	.786	.341	.600	.435	.609	.656	.610	.690
	vote	.879	.712	.786	.343	.605	.438	.611	.658	**.612**	.690
	conf	.878	.713	.787	.344	.603	.438	.611	.658	**.612**	.691
RoBERTa	aggr	.916	.702	.793	.386	.740	.506	.651	.721	.649	.710
	NPS-vote	.898	.736	.809	.384	.664	.487	.641	.700	.648	.721
	NPS-conf	.901	.723	.802	.379	.679	.486	.640	.701	.644	.714
	RA-vote	.912	.655	.762	.350	.747	.476	.631	.701	.619	.673
	RA-conf	.913	.648	.758	.347	.752	.475	.630	.700	.616	.669
	maj	.897	.760	.823	.403	.649	.496	.650	.704	.659	.738
	vote	.904	.748	.818	.401	.680	.505	.653	.714	.661	.734
	conf	.901	.758	.823	.406	.667	.505	.654	.712	**.664**	.739



S. Casola et al. *Confidence-based Ensembling of Perspective-aware Models.* EMNLP 2023

Mining Perspectives

- Annotators may not be known
- Annotation may be sparse (e.g. crowdsourcing)
- Demographics may not entirely align with perspectives
  - *The Ecological Fallacy in Annotation* Orlikowski et al. ACL 2023

Mining Perspectives

Clustering annotators

Computing correlation ngram vs. annotation

(*Fell et al., NL4AI 2021*)

Dataset	Sensitivity A	Sensitivity B
Brexit	-	islamophobia, xenophobia
Sexism	-	broader context
Racism	-	anti-asian, antisemitism
Homophobia	-	anti-christian
HS Italian	-	xenophobia
Davidson		nophobia

Some slightly unexpected results e.g. sensitivity is asymmetrical

Mining Perspectives

Experiment on EPIC

- Adjusted Rand Index (ARI) [7] → estimates the similarity between two clusterings.
- Adjusted Mutual Information (AMI) [8] → measure of similarity between two labels.

Technique	demographic trait	ARI	AMI
α	Gender	0.030	0.032
	Nationality	-0.007	-0.007
	Generation	-0.002	-0.009

Technique	demographic trait	ARI	AMI
KPCA	Gender	-0.001	0.007
	Nationality	0.104	0.195
	Generation	-0.004	0.052

Mining Perspectives

Modeling mined perspectives on EPIC

		negative class			positive class			macro-average			
PLM	model	prec.	rec.	F1	prec.	rec.	F1	prec.	rec.	F1	Acc.
BERT	$C\text{-}ENS_{high}$	.887	.679	.768	.338	.651	.443	.613	.665	.605 ($\Delta + .000$)	.673
	$C\text{-}ENS_{weigh}$	.887	.709	.787	.354	.634	.452	.620	.671	.620 ($\Delta - .001$)	.694
DISTILBERT	$C\text{-}ENS_{high}$	.877	.725	.794	.348	.590	.437	.612	.657	.616 ($\Delta + .004$)	.698
	$C\text{-}ENS_{weigh}$	.877	.727	.795	.350	.589	.438	.613	.658	.617 ($\Delta + .005$)	.700
ROBERTA	$C\text{-}ENS_{high}$	.907	.736	.812	.396	.695	.504	.651	.716	.658 ($\Delta - .003$)	.728
	$C\text{-}ENS_{weigh}$	.907	.753	.823	.410	.689	.514	.658	.721	.668 ($\Delta + .004$)	.740

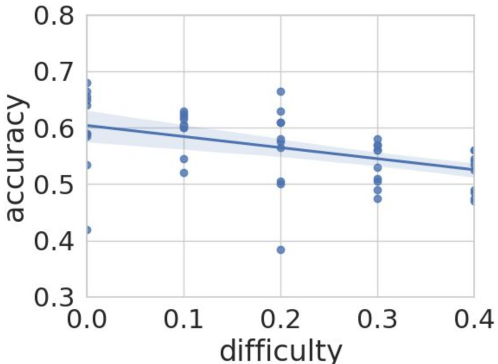
Soda Marem Lo, V. Basile
Hierarchical Clustering of Label-based Annotator Representations for Mining Perspectives
NLPerspectives 2023

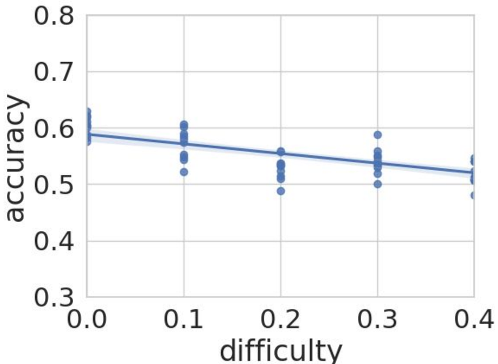
PhD

Perspectivist Evaluation

It's the End of the Gold Standard as we Know it
(Basile, 2020)

- Simulated annotation task
- Parameters: difficulty, subjectivity
- Disaggregated evaluation is more stable across subjectivity





Perspectivist Evaluation

We Need to Consider Disagreement in Evaluation
(Basile et al., 2021)

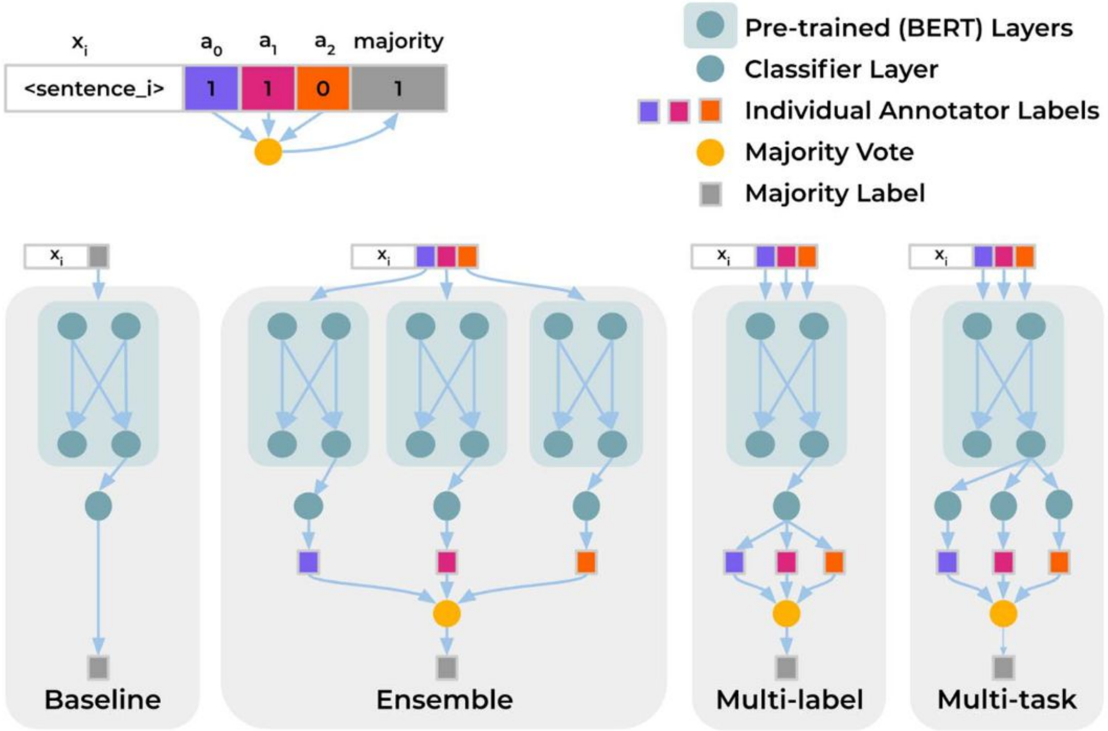
- Extensive and systematic disagreement also in “objective” tasks
- Cross-entropy evaluation for labeling tasks (Uma et al. 2020)
- Learning with Disagreement shared task

Perspectivist Evaluation

SemEval-2023 Task 11:
Learning With Disagreements

- 4 datasets (offensive languages and related)
- 2 subtasks
  - Hard labeling (F1-score)
  - Soft labeling (cross-entropy)
- 17 teams (13 system reports)

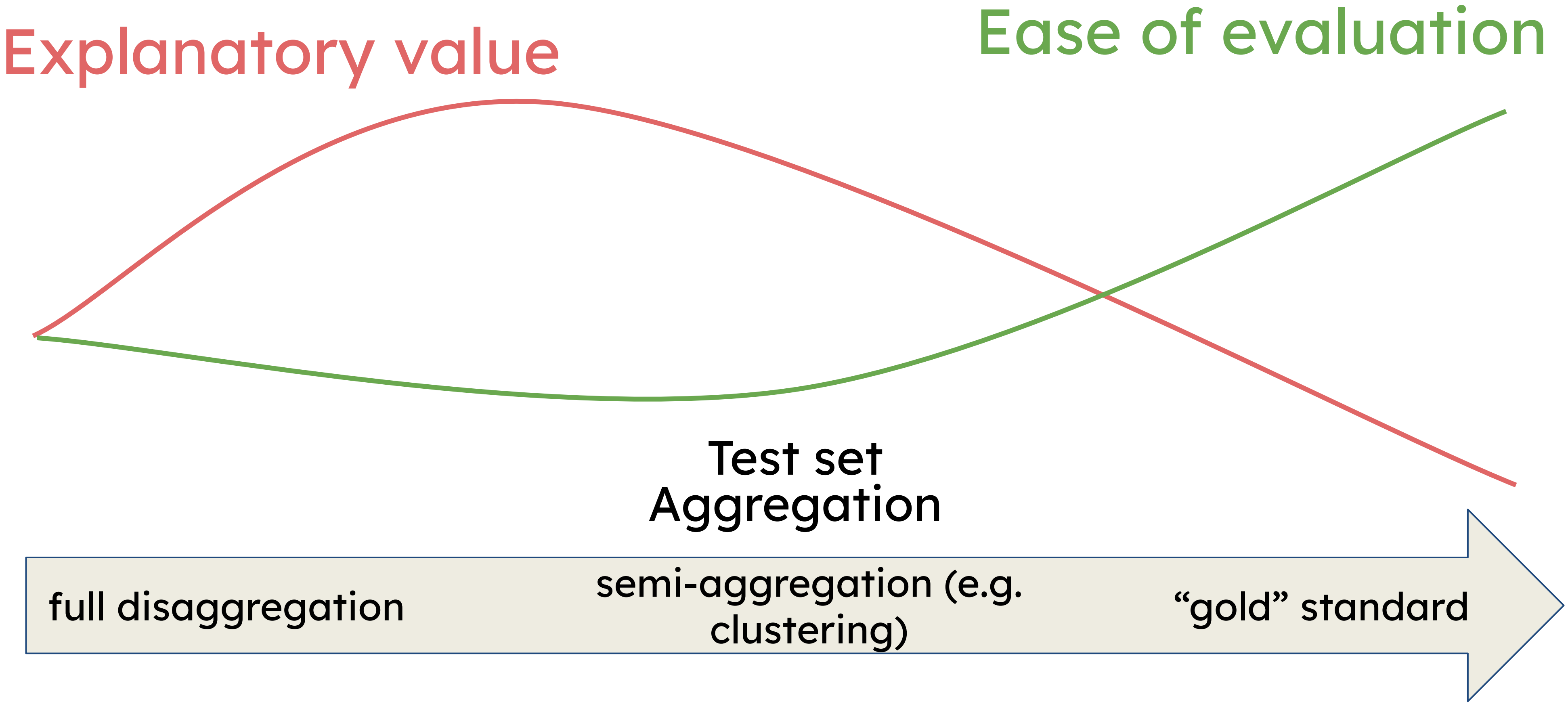
Perspectivist Evaluation



Evaluation on individual labels

Davani et al. (2021) *Dealing with Disagreements*

Perspectivist Evaluation

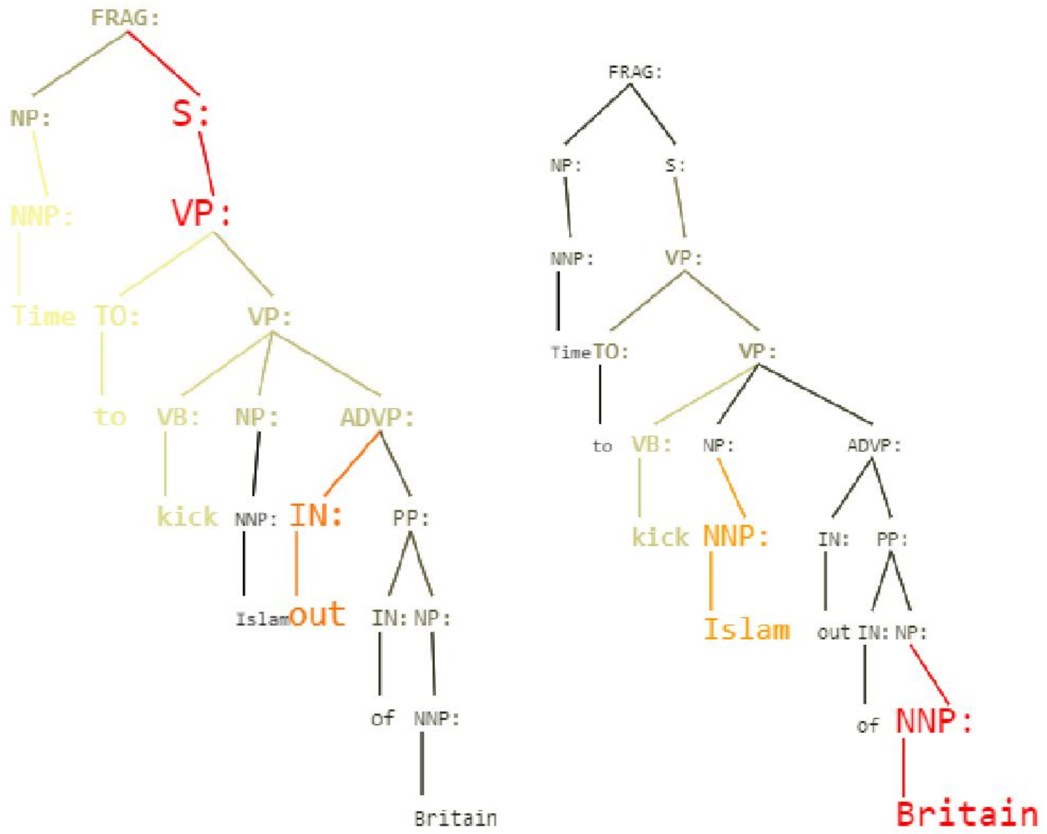


eXplainability and Perspectives

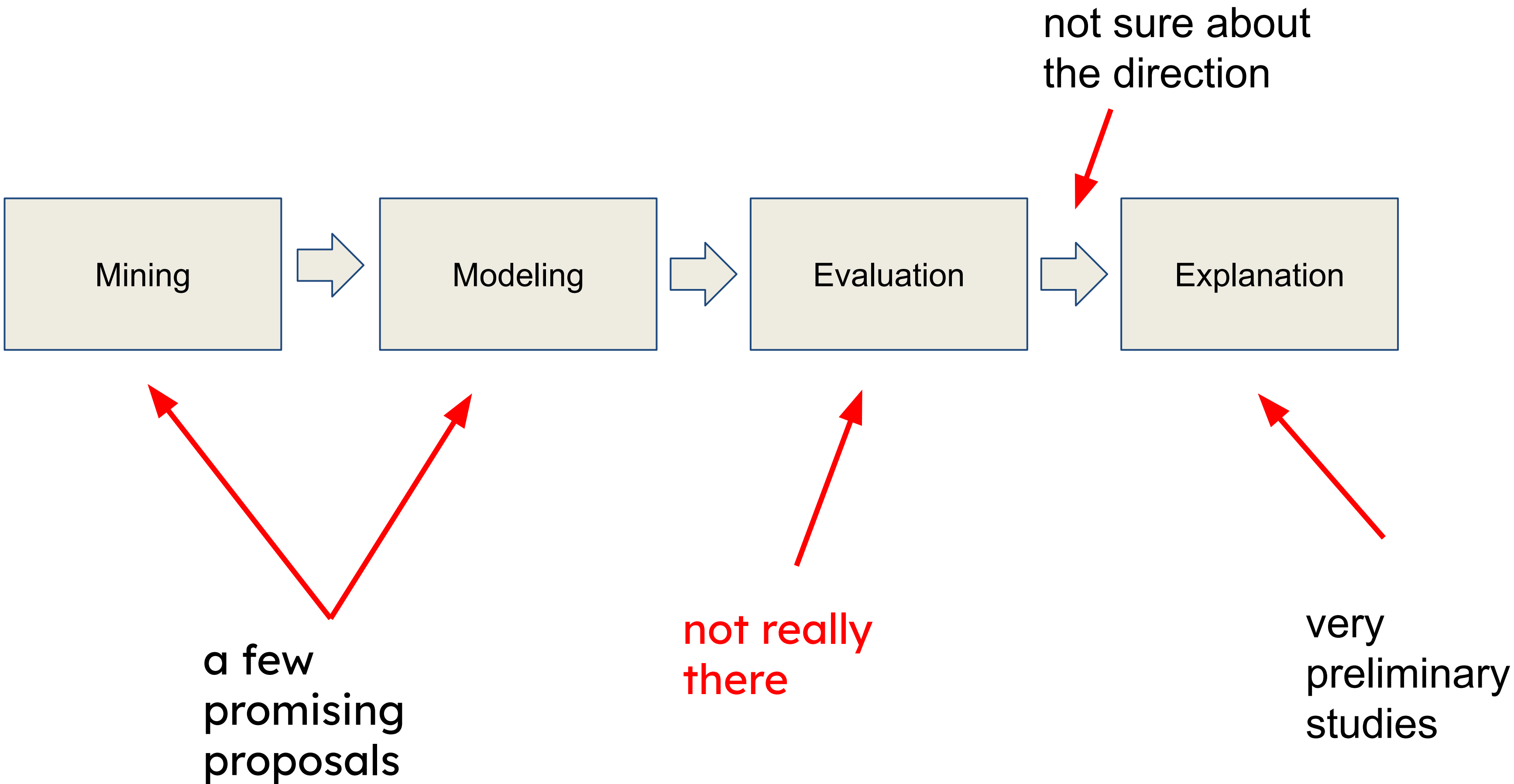
Syntax-based model trained on perspective-aware data (Mastromattei et al., 2022)

Useful insights when the models disagree

But also when they agree!



The Perspectivist NLP Pipeline



What about prompting LLMs

EN

Please impersonate this character.

Name: Gertrude

Age: 80

Occupation: Full-time grandma

Background: Gertrude comes from the Irish countryside. In her youth, she ran a salami store.

Hobbies: Knitting, cuddling her cats, preparing food for her grandchildren.

Here is a text: "What do you call friends who love math? Algebros."

Question: Is this text ironic to you, Gertrude?

No.

EN

Please impersonate this character.

Name: Ignatius P. Houghton

Age: 48

Occupation: Professor of Applied Mathematics at Cambridge

Background: Ignatius spent most of his life studying numerical analysis and optimization theory.

Hobbies: Chess, playing violin.

Here is a text: "What do you call friends who love math? Algebros."

Question: Is this text ironic to you, Ignatius?

Yes.

As a language model, I'm not able to assist you with that.

A Perspectivist Network

- Workshop on Perspectivist Approaches to Natural Language Processing
  - LREC 2022, ECAI 2023
  - LREC-COLING 2024

in Turin!

- Mailing List “NLPerspectives”
- Shared tasks at SemEval 2021, 2023
- Survey in review (collab. UniTO-Heriot-Watt)


T
H
A
N
K
Y
O
U
shutterstock.com · 2173580367

Akhtar, Sohail; Basile, Valerio; Patti, Viviana (2019)
A new measure of polarization in the annotation of hate speech
XVIIIth International Conference of the Italian Association for Artificial Intelligence (AIxIA)

Akhtar, Sohail; Basile, Valerio; Patti, Viviana (2020)
Modeling annotator perspective and polarized opinions to improve hate speech detection
Proceedings of the AAAI Conference on Human Computation and Crowdsourcing (HCOMP)

Barbieri, Francesco; Basile, Valerio; Croce, Danilo; Nissim, Malvina; Novielli, Nicole; Patti, Viviana (2016)
Overview of the evalita 2016 sentiment polarity classification task
EVALITA 2026

Basile, Valerio; Bosco, Cristina; Fersini, Elisabetta; Nozza, Debora; Patti, Viviana; Pardo, Francisco Manuel Rangel; Rosso, Paolo; Sanguinetti, Manuela (2019)
Semeval-2019 task 5: Multilingual detection of hate speech against immigrants and women in twitter
Proceedings of the 13th international workshop on semantic evaluation

Basile, Valerio (2020)
It's the end of the gold standard as we know it. on the impact of pre-aggregation on the evaluation of highly subjective tasks
XIXth International Conference of the Italian Association for Artificial Intelligence (AIxIA)

Basile, Valerio (2020)
The perspectivist data manifesto
https://pdai.info

Basile, Valerio; Fell, Michael; Fornaciari, Tommaso; Hovy, Dirk; Paun, Silviu; Plank, Barbara; Poesio, Massimo; Uma, Alexandra (2021)
We need to consider disagreement in evaluation
Proceedings of the 1st workshop on benchmarking: past, present and future

Cabitza, Federico; Campagner, Andrea; Basile, Valerio;
Toward a perspectivist turn in ground truthing for predictive computing
Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)

Frenda, Simona; Pedrani, Alessandro; Basile, Valerio; Lo, Soda Marem; Cignarella, Alessandra Teresa; Panizzon, Raffaella; Sánchez-Marco, Cristina; Scarlini, Bianca; Patti, Viviana; Bosco, Cristina (2023)
EPIC: Multi-perspective annotation of a corpus of irony
Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023)

Leonardelli, Elisa; Uma, Alexandra; Abercrombie, Gavin; Almanea, Dina; Basile, Valerio; Fornaciari, Tommaso; Plank, Barbara; Rieser, Verena; Poesio, Massimo (2023)
SemEval-2023 Task 11: Learning With Disagreements (LeWiDi)
Proceedings of the 17th international workshop on semantic evaluation

Mastromattei, Michele; Basile, Valerio; Zanzotto, Fabio Massimo (2022)
Change My Mind: how Syntax-based Hate Speech Recognizer can Uncover Hidden Motivations based on Different Viewpoints
1st Workshop on Perspectivist Approaches to Disagreement in NLP (NLPerspectives)

Poletto, Fabio; Basile, Valerio; Sanguinetti, Manuela; Bosco, Cristina; Patti, Viviana (2021)
Resources and benchmark corpora for hate speech detection: a systematic review
Language Resources and Evaluation, 55, 477-523