

TEXT MINING THE “TALK OF LITERATURE” SURROUNDING THE INGEBORG-BACHMANN-PREIS

A CASE STUDY IN DOMAIN ADAPTATION FOR GERMAN-LANGUAGE ASPECT-BASED SENTIMENT ANALYSIS

MOTIVATION

Professional and layperson literary criticism surrounding the *Ingeborg-Bachmann-Preis* and the *Tage der deutschsprachigen Literatur* (TDDL)

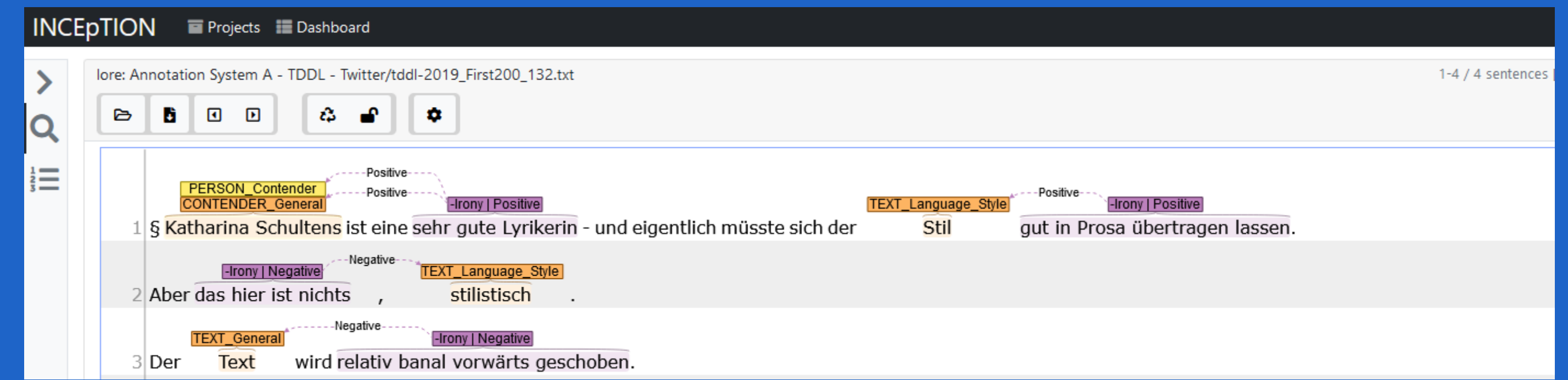
- Discern the evaluative criteria to differentiate between 'good' and 'bad' literature
- Gain insight in the differences in evaluation practices across platforms and media
- Use text mining to examine the role of peer-to-peer recommendation systems
- Role of layperson critics as new literary gatekeepers and cultural transmitters.

Goal:

- Construct literary value through evaluative diction (ABSA)
- Describe the preliminary conditions for training model



CORPUS ANNOTATION



Method: Aspect-Based Sentiment Analysis (ABSA)

Three Layers:

1. Aspects or Feature Expressions (FE):
 - 7 main categories: “Text”, “Reading”, “Contender”, “Jury”, “Onsite Audience”, “Meta” & “Allo-References”
 - 53 subcategories: e.g. “Characters”, “Form”, “General Content or Plot”, “Language or Style”...
2. Named Entities (NE): “Event”, “Organisation”, “Person” & “Product”
3. Polarity or Sentiment Expressions (SE): “positive”, “neutral”, “negative”

EXPERIMENTS

Experimental corpus

4521 German tweets with TDDL-related queries

Task

- Pre-trained German BERT fine-tuned for downstream tasks
- 80% training data, 20% test data
- ABSA subtasks:
 1. Aspect Category Classification
 2. Aspect Polarity Classification
 - Aspect embeddings only vs. context window
 - Irony detection vs. no irony detection

Future Tasks

- Aspect Term Extraction
- Automised Aspect & Polarity Detection and Classification
- Expand experimental corpus

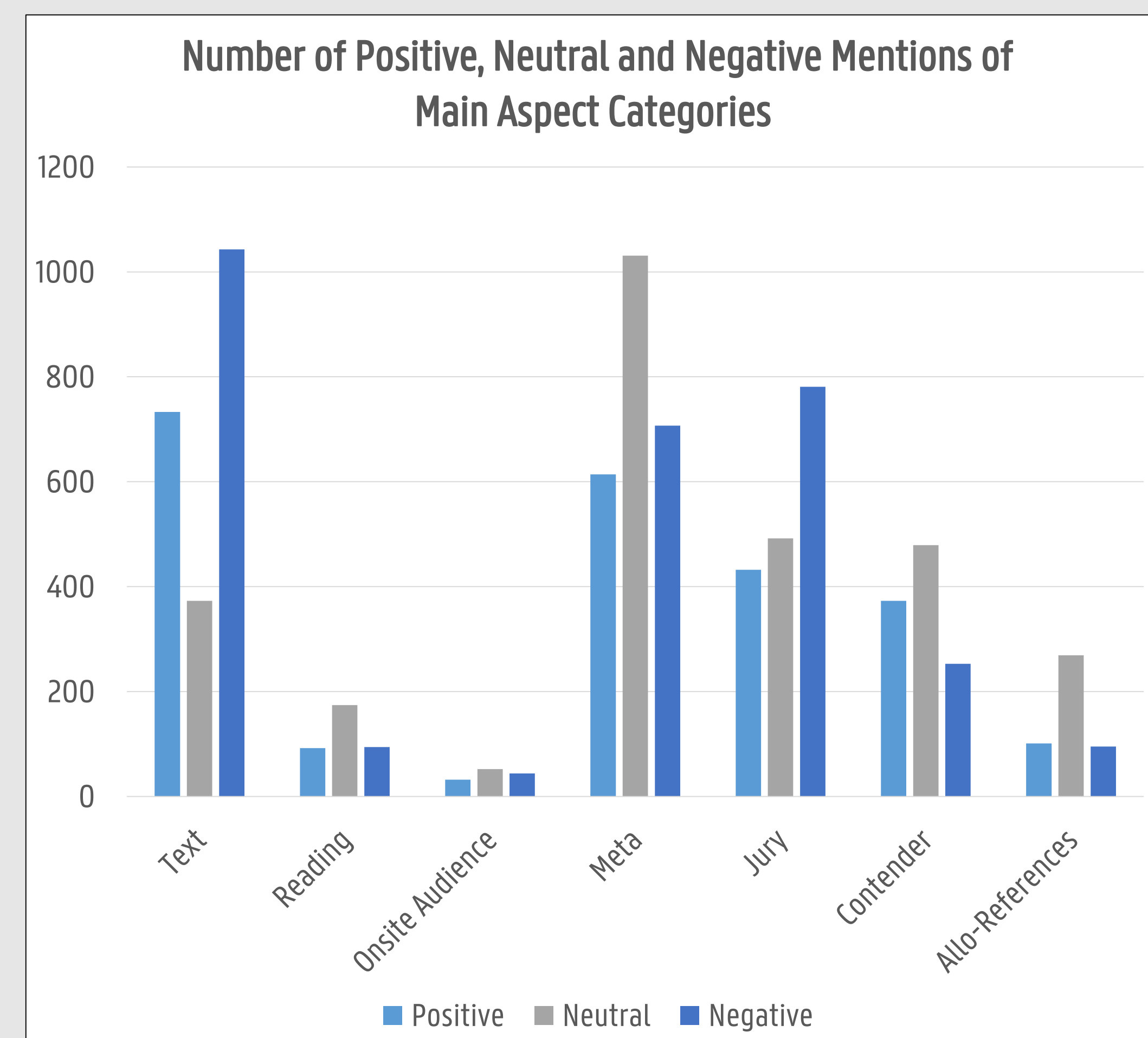
Aspect Category Classification (Main Categories)				
Category	Precision	Recall	F1-score	Support
Accuracy: 0.83				
'TEXT'	0.87	0.83	0.85	408
'JURY'	0.75	0.88	0.81	311
'META'	0.87	0.86	0.86	415
'ALLO-REFERENCES'	0.74	0.68	0.71	76
'ONSITE-AUDIENCE'	0.73	0.67	0.70	24
'READING'	0.78	0.69	0.73	42
'CONTENDER'	0.92	0.82	0.86	198
Macro avg	0.71	0.68	0.69	1475
Weighted avg	0.84	0.83	0.83	1475

Polarity Classification (Context Window = 5 words)				
Category	Precision	Recall	F1-score	Support
With ironic instances				
Accuracy: 0.68				
Positive	0.66	0.64	0.65	467
Negative	0.69	0.72	0.70	537
Macro avg	0.68	0.68	0.68	1004
Weighted avg	0.68	0.68	0.68	1004
Without ironic instances				
Accuracy: 0.71				
Positive	0.71	0.60	0.65	424
Negative	0.71	0.81	0.76	526
Macro avg	0.71	0.70	0.70	950
Weighted avg	0.71	0.71	0.71	950

Number of Positive, Neutral and Negative Mentions of Aspect Subcategories						
Category	Total	Neutral	Positive		Negative	
			Total #	+ Irony	Total #	+ Irony
Accuracy Aspect Subcategory Classification: 0.73						
'TEXT_Title'	11	5	2	1	4	0
'TEXT_Quote'	188	68	65	14	55	0
'TEXT_PoV_Narration'	57	3	24	0	30	0
'TEXT_Motifs_Themes'	72	26	20	0	26	0
'TEXT_Language_Style'	293	36	112	6	145	0
'TEXT_General-Content_Plot'	376	63	124	22	189	0
'TEXT_General'	1026	141	413	19	472	1
'TEXT_Form'	39	16	9	1	14	0
'TEXT_Flow_Rhythm_Punctuation'	25	2	11	0	12	0
'TEXT_Characters'	62	13	16	1	33	0
'READING_Pronunciation_Intonation_Understandability'	24	1	6	1	17	0
'READING_General'	316	172	88	5	56	0
'READING_Flow_Rhythm_Punctuation'	20	1	5	1	14	0
'ONSITE-AUDIENCE_General'	33	15	13	0	5	0
'ONSITE-AUDIENCE_Behaviour'	77	31	20	4	26	0
'ONSITE-AUDIENCE_Appearance_Clothing'	14	6	3	0	5	0
'ONSITE-AUDIENCE_Age'	4	0	1	1	3	0
'META_Winner_Award-Ceremony'	210	93	82	5	35	0
'META_Weather'	75	10	15	6	50	0
'META_Voting'	112	64	23	1	25	0
'META_Videoportrait'	84	6	40	2	38	0
'META_Technology_Social-Media'	311	131	91	8	89	0
'META_Side-Event'	19	9	6	0	4	0
'META_Shortlist'	52	26	11	1	15	0
'META_Opening-Speech'	125	57	51	3	17	0
'META_Online-Assessment'	217	162	42	3	13	0
'META_Music'	51	8	25	6	18	0
'META_Montage'	30	11	9	1	10	0
'META_Main-Event'	825	313	234	33	278	2
'META_Longlist'	4	3	0	0	1	0
'META_Location'	121	77	30	0	14	0
'META_Literature_Literary-Prizes'	116	61	26	4	29	0
'JURY_Voice_Language-Use'	41	4	14	6	23	0
'JURY_Quote'	165	75	42	11	48	0
'JURY_General'	276	82	117	9	77	2
'JURY_Discussion_Valuation'	1145	316	317	57	512	1
'JURY_Behaviour'	38	9	12	7	17	1
'JURY_Appearance_Clothing'	37	6	17	1	14	0
'JURY_Age'	3	0	0	0	3	0
'CONTENDER_Voice_Language-Use'	14	2	7	0	5	0
'CONTENDER_Quote'	15	11	2	0	2	0
'CONTENDER_General'	991	437	362	18	192	0
'CONTENDER_Gender'	47	19	5	2	23	0
'CONTENDER_Appearance_Clothing'	34	9	17	1	8	0
'CONTENDER_Age'	4	1	1	0	2	0
'ALLO-REFERENCES_TEXT_Other-Text'	111	67	26	1	18	1
'ALLO-REFERENCES_TEXT_Other-Author'	189	116	36	3	37	4
'ALLO-REFERENCES_SCREEN_Film_Tv'	49	22	19	4	8	0
'ALLO-REFERENCES_SCREEN_Director_Actor'	16	13	2	0	1	0
'ALLO-REFERENCES_OTHER_Person'	8	4	1	0	3	0
'ALLO-REFERENCES_MUSIC_Musician'	16	5	8	0	3	0
'ALLO-REFERENCES_MUSIC_Music'	3	1	1	0	1	0
'ALLO-REFERENCES_General'	73	41	14	4	18	1
Total	8264	2870	2637	273	2757	13

VALORISATION & CONCLUSION

Number of Positive, Neutral and Negative Mentions of Main Aspect Categories



Advantages:

- High accuracy and reliability for course- and fine-grained Aspect Classification (well represented categories)
- ABSA handles “compliment sandwich” in evaluative talk: “mediocre text by a terrific author”
- Use ABSA to open up layperson literary discourse
 - Show discussed topics and focus
 - Trace shifts in time, platforms & groups
 - Examine and expose prejudice

Technical challenges:

- Aspects: Varying span length & Category under-representation
- Polarity: Implicitness, Irony & Proximity...

Contact

Lore De Greve
Ghent University
Lore.DeGreve@UGent.be
www.talklitmining.ugent.be
@DeGreveLore

Project Website



Github

