

Toward a Tool for Sentiment Analysis for German Historic Plays

Thomas Schmidt, M. Sc. (Media Informatics Group, Regensburg University)

Dr. Manuel Burghardt (Computational Humanities Department, Leipzig University)



Universität Regensburg



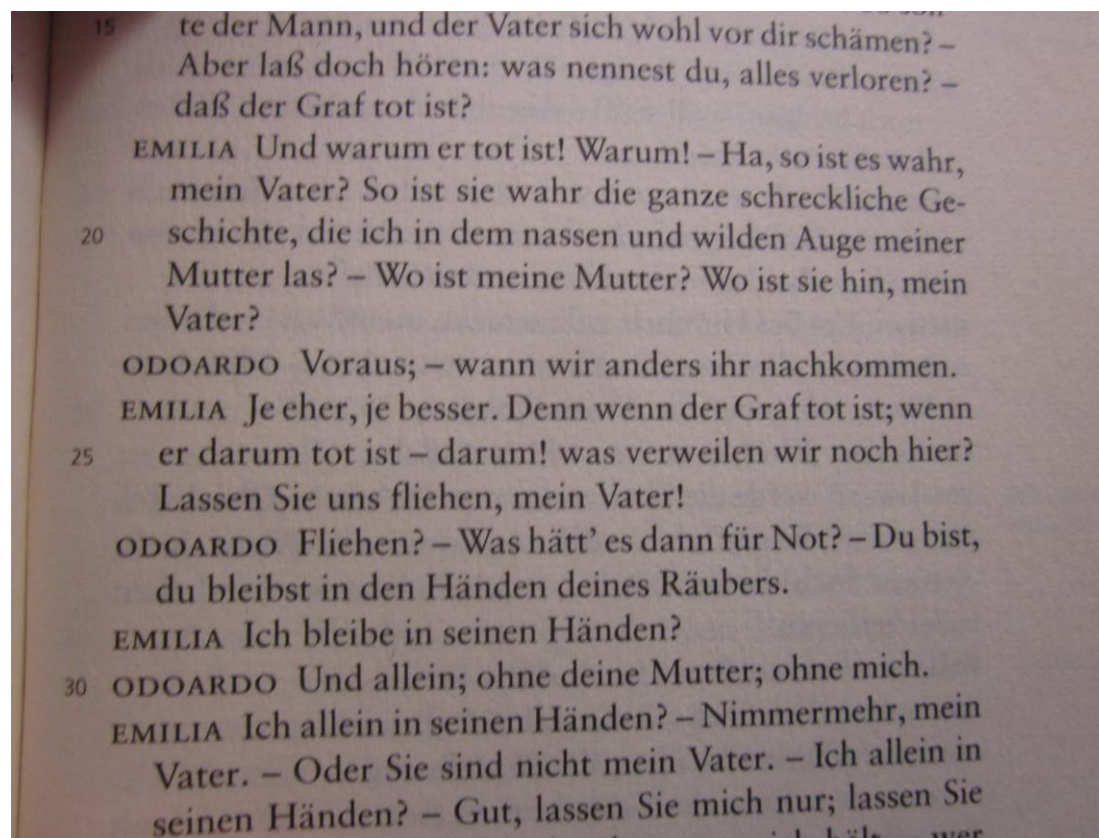
UNIVERSITÄT
LEIPZIG

Sentiment Analysis



Sentiment analysis and drama texts

Why drama texts?



- Analysis of emotion distributions in Shakespeare's plays (Mohammad, 2011)
- Analysis of sentiments of character relations (Nalisnick & Baird, 2013)

Corpus (Plays by G. E. Lessing)

Title	Year	Genre	Speeches	Words
Damon oder die wahre Freundschaft	1747	Comedy	183	7385
Der Freigeist	1749	Comedy	893	20338
Der junge Gelehrte	1748	Comedy	1038	24876
Der Misogyn	1748	Comedy	477	12369
Der Schatz	1750	Comedy	612	11596
Die alte Jungfer	1748	Comedy	470	9266
Die Juden	1749	Comedy	380	9755
Emilia Galotti	1772	Bourgeois tragedy	835	19342
Minna von Barnhelm	1767	Comedy	1134	23538
Miss Sara Sampson	1755	Bourgeois tragedy	690	25395
Nathan der Weise	1779	Dramatic poem	1331	27884
Philotas	1759	Tragedy	181	6867



Lexicon based sentiment analysis

1	Wort	Polarität
2	Beschwerde	NEG=0.7
3	bettlägerig	NEG=1
4	arbeitsunfähig	NEG=1
5	Unfall	NEG=1
6	Erdölkrise	NEG=0.7
7	vollstopfen	NEG=0.7
8	Schweinegrippe	NEG=1
9	abwenden	POS=1
10	gelingen	POS=1
11	Erfolg	POS=1
12	Problem	NEG=1
13	drakonisch	NEG=1
14	Zahlungsunfähig	NEG=0.7
15	Steuereinnahme	POS=0.7
16	Plünderung	NEG=1
17	goldig	POS=0.7
18	Bedrohlichkeit	NEG=1

Sentiment bearing word
(sbw)

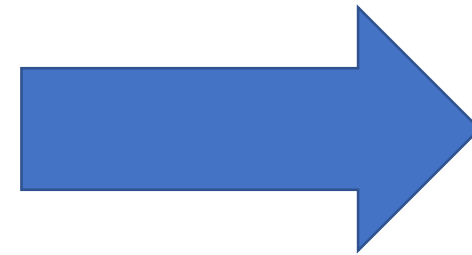
Polarity

Polarity strengths

Snippet by Clematide & Klenner
(2010)

Parameters

- Sentiment lexicon
- Lemmatization
- Stop words/Very frequent words
- Extension with historical linguistic variants
- Case-sensitivity



Evaluation

Gold Standard creation

- 5 annotators
- 200 speeches

ADRAST:

Noch ein Wort! Drohend.

THEOPHAN:

Nunmehr darf ich die Bitte um eine nähere Erklärung doch wohl wiederholen? Ich weiß sie mir selbst nicht zu geben.

ADRAST:

Erklären Sie sich denn gerne näher, Theophan?

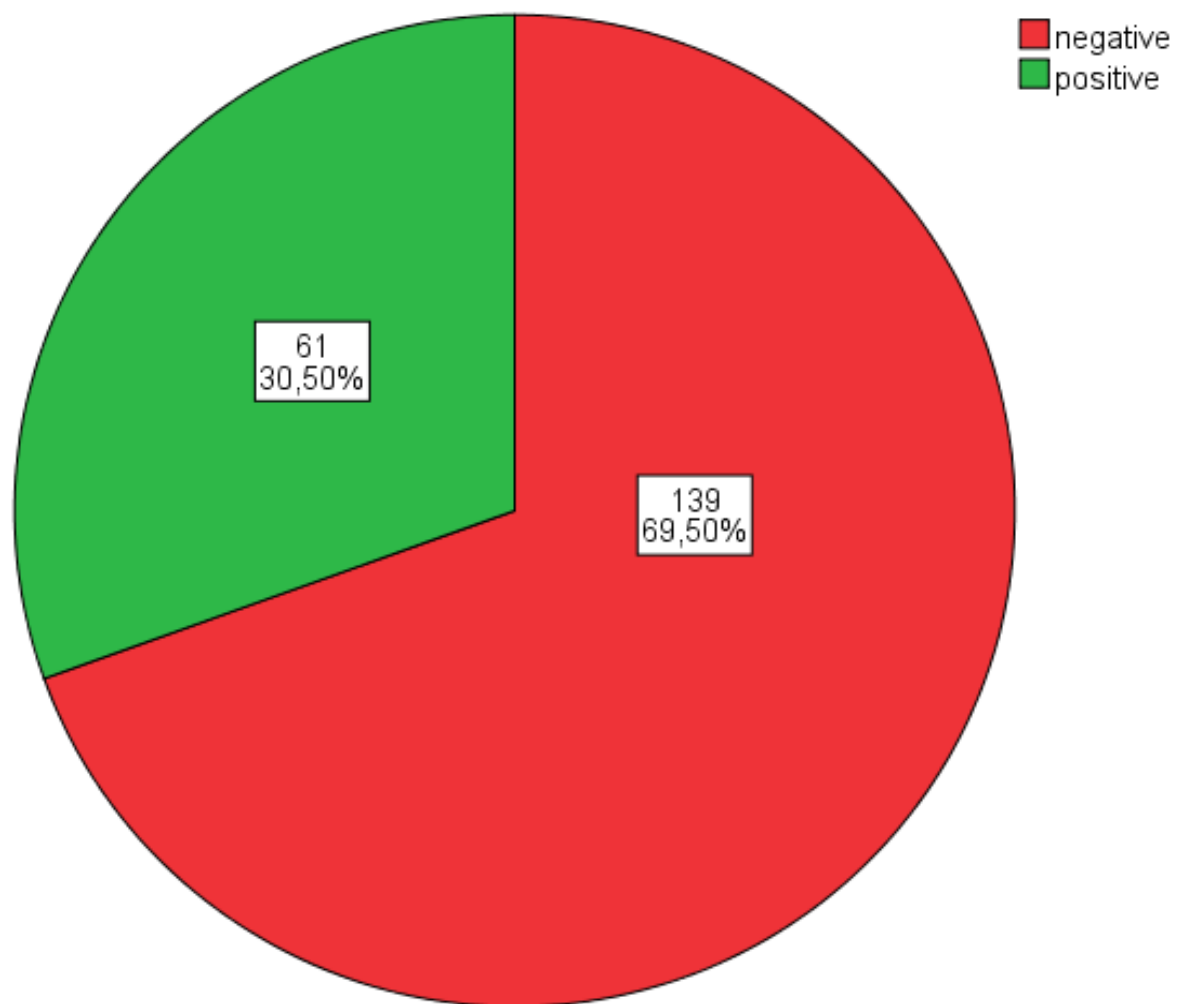
Very Negative	Negative	Neutral	Mixed	Positive	Very Positive

Negative	Positive

Levels of agreement

Annotation	Krippendorff's α	Average percentage of agreement among annotators
Polarity differentiated	0.22	40%
Binary polarity	0.47	77%

Final annotation



Evaluation results – main findings

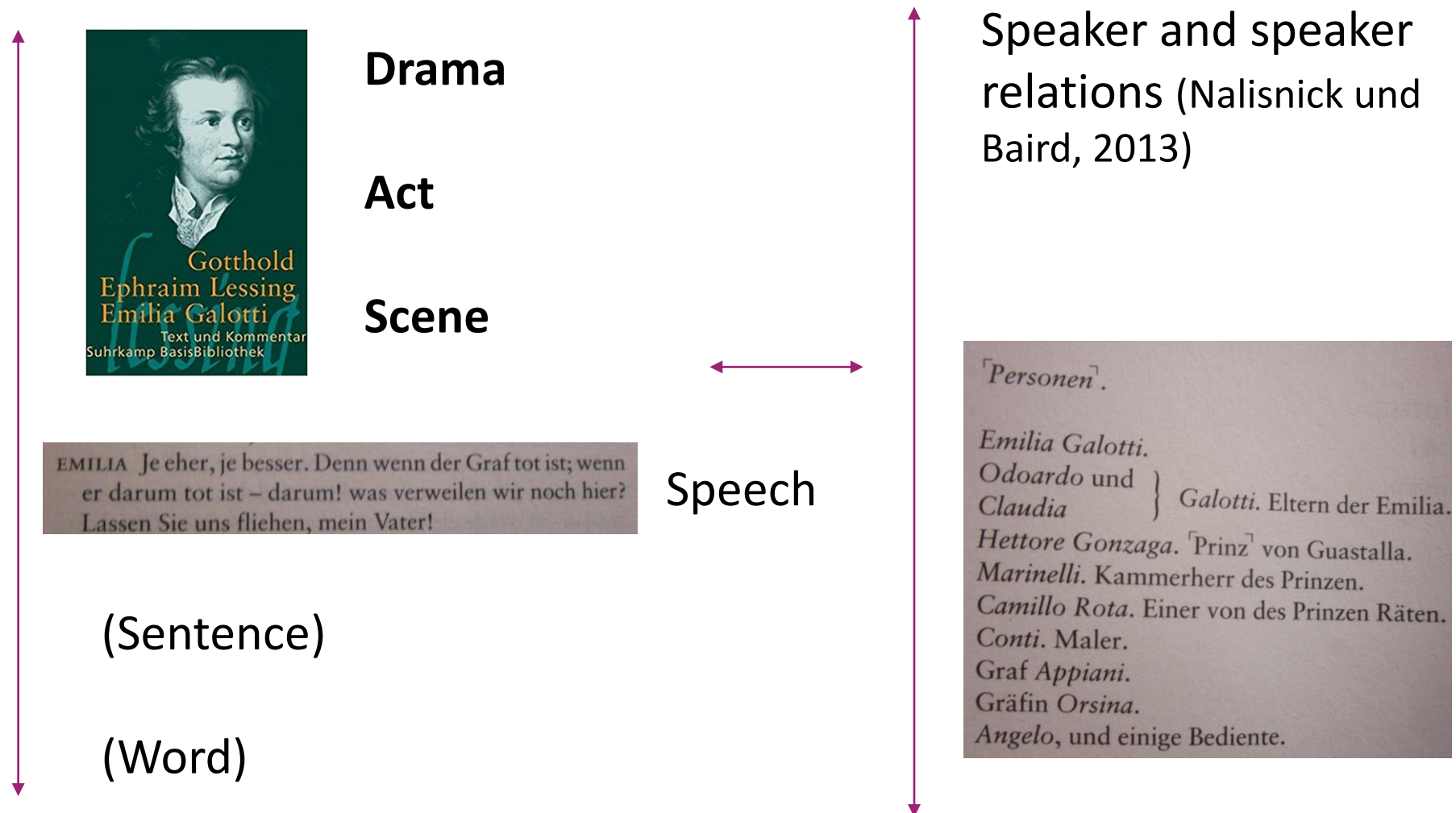
- Sentiment lexicons with polarity strengths outperform other lexicons
- SentiWS (Remus et al., 2010) is the best performing lexicon
- Automatic lemmatization is problematic with historical and poetic language
- Extension of lexicons with historical linguistic variants increases accuracy consistently
- Stop words and very frequent words sometimes lead to over-prediction of one polarity

Best configuration

Accuracy of 70% with:

- SentiWS (Polarity strengths)
- Extended with historical linguistic variants
- Treetagger-lemmatizer
- No stop words list
- Lower casing

Levels of sentiment analysis



Visualization

Goal:

- Online-Tool for Exploration
- Use cases for literary studies
- Requirements of literary scientiests

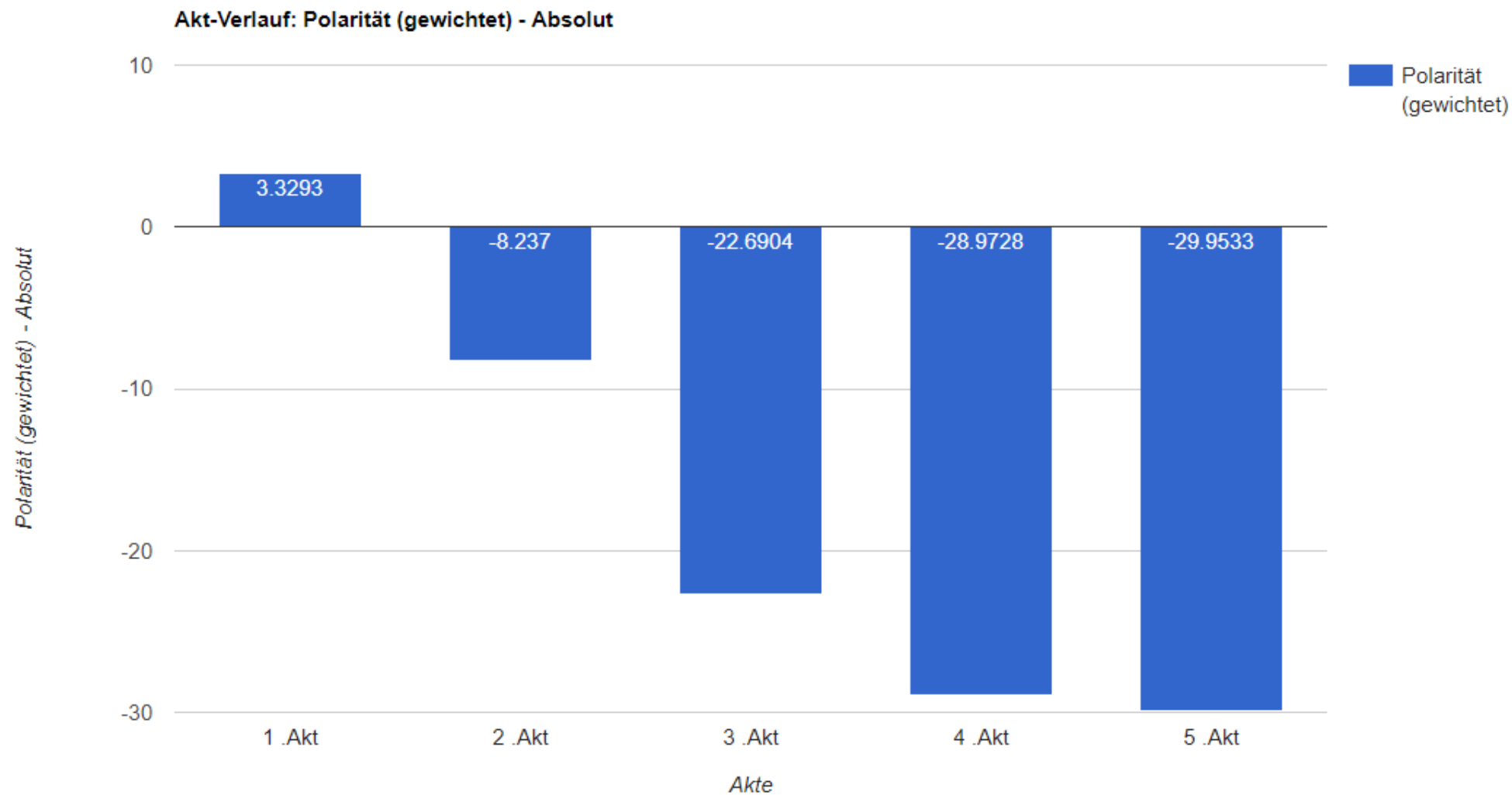
Levels: drama, acts, scenes, speakers, speaker relations

Analysis methods: sentiment progressions, sentiment distributions

Categories: polarity, positive, negative, emotions



Emilia Galotti – Polarity progression per act



Current state and future plans

- Extended annotation project:
 - Trained students and experts
 - More participants
 - More speeches
 - More sophisticated schema
 - Analysis of annotation behavior and requirements
- (Maybe) crowdsourcing
- More authors and more dramas
- → Machine Learning and hybrid methods

Thank you for your attention! Any Questions?

Contact:

thomas.schmidt@ur.de

burghardt@informatik.uni-leipzig.de

Twitter:

@thomasS_UniR

@8urghardt



References

Clematide, S. & Klenner, M. (2010). Evaluation and extension of a polarity lexicon for German. In *Proceedings of the First Workshop on Computational Approaches to Subjectivity and Sentiment Analysis* (pp. 7-13).

Mohammad, S. (2011). From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities* (pp. 105-114). Association for Computational Linguistics.

Nalisnick, E. T. & Baird, H. S. (2013). Character-to-character sentiment analysis in shakespeare's plays. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (pp. 479–483).

Remus, R., Quasthoff, U. & Heyer, G. (2010). SentiWS-A Publicly Available German-language Resource for Sentiment Analysis. In *LREC* (pp. 1168-1171).