Results of Emotion Annotation in German Drama from 1650-1815

Schmidt, Thomas

thomas.schmidt@ur.de
Media Informatics Group, University of Regensburg, Germany

Dennerlein, Katrin

katrin.dennerlein@uni-wuerzburg.de
Institute for German Philology, Julius-Maximilians Universität Würzburg, Germany

Wolff, Christian

christian.wolff@ur.de
Media Informatics Group, University of Regensburg, Germany

Introduction

We present results of the research project *Emotions in Drama* (EmoDrama) [1] which is a collaboration between literary scholars and computer scientists. The goal of the project is to perform computational emotion prediction to analyze historical German plays from 1650-1815 and to advance previous research in the field of sentiment and emotion analysis on dramatic texts (Alm/Sproat 2005; Mohammad 2011; Nalisnick/Baird 2013; Schmidt/Burghardt, 2018; Schmidt et al., 2018; 2019; Yavuz, 2021). In the following chapters, we elaborate on results of the annotation task of our project: A corpus of 18 German plays with over 21,000 emotion annotations. [2]

Annotation

The annotation process and scheme were derived in an iterative process and are focused on literary emotion concepts important for the time. Our annotation scheme is hierarchical and consists of 6 main emotion classes and 13 sub emotions with positive or negative polarity (figure 1). We annotated 18 plays that are representative for our time period in style and genre (figure 2).

- Emotions of affection / Zuneigung
  - desire / Lust (+)
  - love / Liebe (+)
  - friendship / Freundschaft (+)
  - admiration, reverence / Verehrung, Bewunderung (+)
- Emotions of pleasure / Freude und Glück
  - joy / Freude (+)
  - Schadenfreude (+)
- Emotions of anxiety / Angst und Sorge
  - fear / Angst (-)
  - despair / Verzweiflung (-)
- Emotions of rejection / Ablehnung
  - anger / Ärger (-)
  - abhorrence / Abscheu, Wut, Hass (-)
- Emotions of suffering and empathy / Leid
  - suffering / Leid (-)
  - compassion / Mitleid (-)
- no main class
  - being moved / emotionale Bewegtheit (undetermined)

Figure 1. Emotion scheme.

- Ibrahim Sultan (1649), tragedy
- Andreas Gryphius: Katharina von Georgien (1657), tragedy
- Christian Weise: Ein wunderliches Schau-Spiel vom niederländischen Bauer (1669), comedy
- Der Welt Erschröckende Attila (nach 1682), play, libretto
- Christian Weise: Massaniello (1683), tragedy
- Die getreue Sclavin Doris (1720), play, libretto
- Luise Adelgunde Victorie Gottsched: Das Testament (1745), comedy
- Johann Elias Schlegel: Canut (1746), tragedy
- Christian Fürchtegott Gellert: Die zärtlichen Schwestern (1747), comedy
- Johann Gottlieb Benjamin Pfeil: Lucie Woodvil (1757), tragedy
- Joachim Wilhelm von Brawe: Der Freigeist (1758), tragedy
- Gotthold Ephraim Lessing: Minna von Barnhelm, oder das Soldatenglück (1767), comedy
- Cornelius von Ayrenhoff: Der Postzug (1769), comedy
- Friedrich Schiller: Kabale und Liebe. Ein bürgerliches Trauerspiel in fünf Aufzügen (1784), tragedy
- Ferdinand Eberl: Kasperl' der Mandolettikrämer (1789), comedy, libretto
- August von Kotzebue: Menschenhass und Reue (1790), comedy
- Friedrich Schiller: Wallensteins Lager (1800), tragedy
- Johann Wolfgang Goethe: Faust. Eine Tragödie (1807), tragedy

Figure 2. Annotated plays.

Plays were acquired via Fischer et al. (2019) and other online sources. [3] Each play was annotated by two annotators (6 annotators overall) independent from each other with the CATMA tool (Gius et al. 2020). Annotators were trained in annotation sessions and used an annotation manual specifically developed for this task (Dennerlein et al. 2022c). The annotation was performed content- as well as context-sensitive: Annotators annotated what characters of the play felt as expressed in the text (example: figure 3). The annotation spans could be chosen as the annotators saw fit including overlapping emotion annotations since this proved to be beneficial for our research goals. More information about the annotation process can be found in previous research (Schmidt et al., 2021a; Dennerlein et al., 2022c).

Zweyter Auftritt.

Estrithe, Canut, Godewin.

ESTRITHE.

Mein König, deine Huld, die du mir wiedergiebst,

Beschämt mich, da sie mir bezeigt, wie du mich liebst.

Figure 3. Example annotation from CATMA. Purple (both lines) is marked as suffering. Blue (last four words) is love. Excerpt from J.E. Schlegel's *Canut*

(Own translation: My king, the grace you return to me is putting me to shame since this shows how you love me).

Annotated corpora

The annotation resulted in the acquisition of 21,609 single annotations which is equivalent to 543,518 annotated tokens (table 1). Of these annotations, the majority are negative emotions (11,706; 54%) followed by positive (8,074; 37%) and the special case 'emotional movement' (1,829; 8%).

Emotion	#	%	# (tokens)	% (tokens)	avg. tokens	median tokens	min tokens	max tokens	std tokens
MC: emotions of affection	5,045	23	11,6380	21.41	23.07	15	1	345	26.11
desire	233	1	6,087	1.12	26.12	19	1	184	25.55
love	2,819	13	63,697	11.72	22.6	15	1	326	26.29
friendship	457	2	9,919	1.82	21.7	17	1	132	18.7
admiration, reverence	1,536	7	36,677	6.75	23.88	16	1	345	27.69
MC: emotions of pleasure	3,029	14	68,211	12.55	22.52	16	1	288	23.79
joy	2,502	12	55,221	10.16	22.07	16	1	277	22.63
Schadenfreude	527	2	12,990	2.39	24.65	18	2	288	28.59
MC: emotions of anxiety	2,208	10	63,370	11.66	28.7	16	1	702	44.82
fear	1,389	6	26,667	4.91	19.2	14	1	173	17.99
despair	819	4	36,703	6.75	44.81	23	1	702	66.75
MC: emotions of rejection	5,156	24	127,942	23.54	24.81	17	1	813	29.57
anger	2,616	12	59,220	10.9	22.64	15	1	813	28.82
abhorrence	2,540	12	68,722	12.64	27.06	18	1	348	30.17
MC: emotions of suffering and empathy	4,342	20	126,627	23.3	29.16	18	1	998	42.23
suffering	3,486	16	105,720	19.45	30.33	18	1	998	45.21
compassion	856	4	20,907	3.85	24.42	16.5	1	263	26.37
being moved	1,796	8	40,247	7.4	22.41	12	1	334	30.56
overall	21,609	100	543,518	100	25.15	17	1	998	31.21

Table 1. Statistics for emotion annotations. Main classes (MC) are followed by their sub emotions.

The most frequent main classes are the emotions of 'rejection' (24%) and 'affection' (23%). For the sub-emotions, 'suffering' (16%) and 'love' (13%) are annotated the most. Token statistics show that the annotation span in the text was on average 25 tokens (around 2 sentences).

Due to the varied and overlapping annotation spans, we calculated agreement by assigning the emotion type that was annotated the most for each annotator (in number of words) for each stage direction and speech. We then analyze the annotation agreement via Cohen's # and percentage-wise agreement (table 2).

Emotion Class	Polarity (κ)	Polarity (%)	Main emotion class (κ)	Main emotion class (%)	Sub emotion (κ)	Sub emotion (%)
Values	0.45	67.34	0.40	61.88	0.36	57.96

Table 2. Agreement statistics.

On average the #-values range from 0.45 to 0.36 depending on the emotion class which points from moderate to fair agreement according to Landis and Koch (1977). Token based agreements show similar results. The overall values are lower than in sentiment analysis projects with contemporary texts (cf. Mäntylä et al. 2018) but in line with annotation studies on literary and historical texts (Alm/Sproat 2005; Sprugnoli et al. 2015; Schmidt et al. 2018; 2019). We identified the problems of agreement as an inherent attribute of the interpretability of literary texts based on reports by the annotators. To construct a final corpus usable for machine learning, we filtered any annotations the two annotators per play disagreed partially or fully on per emotion group. That means, only annotations remained on which both annotators agreed upon or only one annotator made an annotation for a text unit. In the case of sub-emotions this leads to the removal of annotations with 12,701 (59% of all annotations by both annotators) single annotations remaining.

Furthermore, a lot of material is not annotated at all. Thus, in a version of the corpus we include non-annotated text. Table 3 shows distributions and statistics of this corpus filtered by disagreement and extended with non-annotated material for the sub emotions group.

Emotion	#	%	# (tokens)	% (tokens)	avg. tokens	median tokens	min tokens	max tokens	std tokens
MC: emotions of affection	3,029	24	58,992	13	19.48	14	1	326	20.08
desire	117	0	2,751	1	23.51	20	2	184	20.39
love	1,672	7	31,697	7	18.96	13	1	326	20.52
friendship	248	1	4,815	1	19.42	16	1	132	16.01
admiration, reverence	992	4	19,729	4	19.89	15	1	266	20.17
MC: emotions of pleasure	2,039	8	40,523	9	19.87	15	1	243	19.90
joy	1,729	7	34,020	8	19.68	15	1	223	19.41
Schadenfreude	310	1	6,503	1	20.98	16	2	243	22.42
MC: emotions of anxiety	1,091	4	22,752	5	20.85	13	1	315	27.04
fear	792	3	12,972	3	16.38	12.5	1	131	13.61
despair	299	1	9,780	2	32.71	16.0	1	315	44.59
MC: emotions of rejection	3,106	13	63,867	15	20.56	15	1	204	19.08
anger	1,688	7	31,483	7	18.65	14	1	143	16.28
abhorrence	1,418	6	32,384	7	22.84	16	1	204	21.73
MC: emotions of suffering and empathy	2,272	9	51,653	12	22.73	15	1	387	23.90
suffering	1,872	8	44,305	10	23.67	16	1	387	25.08
compassion	400	2	7,348	2	18.37	13	1	171	16.70
being moved	1,164	5	18,926	4	16.26	10	1	224	20.67
no annotation	11,901	48	183,534	42	15.42	9	1	891	22.74
overall	24,602	100	440,247	100	12.18	14.5	1	891	21.60

Table 3. Emotion annotation statistics for the filtered corpus including non-annotated material.

While the overall distributions of the emotions after filtering do not change, the addition of the non-annotated material shows new insights. Around half of the material is non-annotated (11,901; 48%) compared to the filtered annotated texts (12,701; 52%).

Discussion

The emotion-annotated corpora we acquired are a unique data source for digital humanities and computational literary studies. To our knowledge, no other corpus for this literary genre exists with a similar amount of context- and content-sensitive emotion annotations. The application areas are multifaceted: We performed annotation analysis investigating emotion developments across plays (Dennerlein et al. 2022a; 2022b) and used the corpora to build emotion classification algorithms (Schmidt et al. 2021b; 2021c; 2022) which we then used to explore genre differences (Dennerlein et al. 2022b; Dennerlein et al. 2023b) or differences in stage directions and character speeches (Dennerlein et al. 2023a). We publish the corpus and additional data as CSV / JSON files. [4]

Notes

2. Corpus/Data available via https://github.com/lauchblatt/Emotions_in_Drama.

3. https://textgrid.de/, http://lithes.uni-graz.at/maezene/maeze-ne_startseite.html
4. https://github.com/lauchblatt/Emotions_in_Drama.

Bibliography

Alm, Cecilia O. / **Sproat, Richard** (2005). "Emotional Sequencing and Development in Fairy Tales", in: Tao, Jianhua / Tan, Tieniu / Picard, Rosalind W. (eds.): *Affective Computing and Intelligent Interaction*. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer: 668–674. DOI: 10.1007/11573548_86.

Dennerlein, Katrin / **Schmidt, Thomas** / **Wolff, Christian** (2022a): "Emotionen im kulturellen Gedächtnis bewahren", in: *DHd 2022 Kulturen des digitalen Gedächtnisses. 8. Tagung des Verbands 'Digital Humanities im deutschsprachigen Raum' (DHd 2022)*. Potsdam, Zenodo: 93–98. DOI: 10.5281/zenodo.6327957.

Dennerlein, Katrin / **Schmidt, Thomas** / **Wolff, Christian** (2022b): "Emotion Courses in German Historical Comedies and Tragedies", in: *Book of Abstracts, Digital Humanities 2022 (DH2022). Responding to Asian Diversity*. Tokyo, Japan: The University of Tokyo: 193–197.

Dennerlein, Katrin / **Schmidt, Thomas** / **Wolff, Christian** (2022c): Figurenemotionen in deutschsprachigen Dramen annotieren. Zenodo. DOI: 10.5281/zenodo.6228152.

Dennerlein, Katrin / **Schmidt, Thomas** / **Wolff, Christian** (2023a; in print): "Emotions in Stage Directions in German Drama of the Early Modern Period: Explorations via Computational Emotion Classification", in: Andresen, Melanie / Reiter, Nils / Krautter, Benjamin / Pagel, Janis (eds.): *Computational Drama Analysis: Achievements and Opportunities*. Berlin/New York: de Gruyter.

Dennerlein, Katrin / **Schmidt, Thomas** / **Wolff, Christian** (2023b; to appear): „Computational Emotion Classification for Genre Corpora of German Tragedies and Comedies from 17th to 19th Century", in: Digital Scholarship in the Humanities.

Fischer, Frank et al. (2019): "Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama", in: *Book of Abstracts, Digital Humanities Conference 2019 (DH 2019)*. Utrecht, Netherlands. DOI: 10.5281/zenodo.4284002.

Gius, Evelyn et al. (2020): *CATMA*. Zenodo. DOI: 10.5281/zenodo.4353618.

Landis, J. Richard / **Koch, Gary G.** (1977): "The Measurement of Observer Agreement for Categorical Data", in: *Biometrics*, 33(1): 159–174.

Mäntylä, Mika V. / **Graziotin, Daniel** / **Kuutila, Miikka** (2018): "The Evolution of Sentiment Analysis—A Review of Research Topics, Venues, and Top Cited Papers", in: *Computer Science Review*, 27: 16–32. DOI: 10.1016/j.cosrev.2017.10.002.

Mohammad, Saif (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*. Portland, OR, USA: Association for Computational Linguistics: 105–114.

Nalisnick, Eric T. / **Baird, Henry S.** (2013): "Character-to-Character Sentiment Analysis in Shakespeare's Plays", in: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Sofia, Bulgaria: Association for Computational Linguistics: 479–483.

Schmidt, Thomas / **Burghardt, Manuel** (2018): "An Evaluation of Lexicon-Based Sentiment Analysis Techniques for the Plays of Gotthold Ephraim Lessing", in: *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Santa Fe, New Mexico: Association for Computational Linguistics: 139–149.

Schmidt, Thomas / **Burghardt Manuel** / **Dennerlein, Katrin** (2018): "Sentiment Annotation of Historic German Plays: An Empirical Study on Annotation Behavior", in: Kübler, Sandra / Zinsmeister, Heike (eds.): *Proceedings of the Workshop on Annotation in Digital Humanities 2018 (AnnDH 2018)*. Sofia, Bulgaria: RWTH Aachen: 47–52.

Schmidt, Thomas / **Burghardt, Manuel** / **Dennerlein, Katrin** / **Wolff, Christian** (2019): "Sentiment Annotation for Lessing's Plays: Towards a Language Resource for Sentiment Analysis on German Literary Texts", in: Declerck, Thierry / McCrae, John P. (eds.): *2nd Conference on Language, Data and Knowledge (LDK 2019)*. Leipzig, Germany: 45–50.

Schmidt, Thomas / **Dennerlein, Katrin** / **Wolff, Christian** (2021a): "Towards a Corpus of Historical German Plays with Emotion Annotations", in: Gromann, Dagmar et al. (eds.): *3rd Conference on Language, Data and Knowledge (LDK 2021)*. Open Access Series in Informatics (OASIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum Für Informatik: 9:1-9:11. DOI: 10.4230/OASIcs.LDK.2021.9.

Schmidt, Thomas / **Dennerlein, Katrin** / **Wolff, Christian** (2021b): "Using Deep Learning for Emotion Analysis of 18th and 19th Century German Plays", in: Burghardt, Manuel et al. (eds.): *Fabrikation von Erkenntnis: Experimente in den Digital Humanities*. Esch-Sur-Alzette, Luxembourg, Melusina Press. DOI: 10.26298/melusina.8f8w-y749-udlf.

Schmidt, Thomas / **Dennerlein, Katrin** / **Wolff, Christian** (2021c): "Emotion Classification in German Plays with Transformer-Based Language Models Pretrained on Historical and Contemporary Language", in: *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Punta Cana, Dominican Republic (Online): Association for Computational Linguistics: 67–79. DOI: 10.18653/v1/2021.latechclfl-1.8.

Schmidt, Thomas / **Dennerlein, Katrin** / **Wolff, Christian** (2022): "Evaluation computergestützter Verfahren der Emotionsklassifikation für deutschsprachige Dramen um 1800", in: *DHd 2022 Kulturen des digitalen Gedächtnisses. 8. Tagung des Verbands 'Digital Humanities im deutschsprachigen Raum' (DHd 2022)*. Potsdam, Germany: Zenodo: 107–113. DOI: 10.5281/zenodo.6328169.

Sprugnoli, Rachele / **Tonelli, Sara** / **Marchetti, Allessandro** / **Moretti, Giovanni** (2015): "Towards Sentiment Analysis for Historical Texts", in: *Digital Scholarship in the Humanities*, 31: 762–772. DOI: 10.1093/llc/fqv027.

Yavuz, Mehmet C. (2021): "Analyses of Character Emotions in Dramatic Works by Using EmoLex Unigrams", in: *Proceedings of the Seventh Italian Conference on Computational Linguistics, CLiC-It'20*. Bologna, Italy.