Towards a Corpus of Historical German Plays with Emotion Annotations

Thomas Schmidt
Media Informatics Group, University of Regensburg, Germany

Katrin Dennerlein
German Literary Studies and Computational Literary Studies, University of Würzburg, Germany

Christian Wolff
Media Informatics Group, University of Regensburg, Germany

Abstract

In this paper, we present first work-in-progress annotation results of a project investigating computational methods of emotion analysis for historical German plays around 1800. We report on the development of an annotation scheme focussing on the annotation of emotions that are important from a literary studies perspective for this time span as well as on the annotation process we have developed. We annotate emotions expressed or attributed by characters of the plays in the written texts. The scheme consists of 13 hierarchically structured emotion concepts as well as the source (who experiences or attributes the emotion) and target (who or what is the emotion directed towards). We have conducted the annotation of five example plays of our corpus with two annotators per play and report on annotation distributions and agreement statistics. We were able to collect over 6,500 emotion annotations and identified a fair agreement for most concepts around a κ-value of 0.4. We discuss how we plan to improve annotator consistency and continue our work. The results also have implications for similar projects in the context of Digital Humanities.

2012 ACM Subject Classification Applied computing → Arts and humanities; Computing methodologies → Machine learning

Keywords and phrases Emotion, Annotation, Digital Humanities, Computational Literary Studies, German Drama, Sentiment Analysis, Emotion Analysis, Corpus

Digital Object Identifier 10.4230/OASIcs.LDK.2021.9

Funding This research is part of the project "Emotions in Drama" (Emotionen im Drama), is funded by the German Research Foundation (DFG) and part of the priority programme SPP 2207 Computational Literary Studies (CLS).

Acknowledgements We want to thank the following student annotators for their contributions to this project: Viola Hipler, Julia Jäger, Emma Ruß, and Leon Sautter.

1 Introduction

Emotions in dramatic texts are central for the dramaturgy, the characterization of characters, the intended effect on the reader as well as for the propagation of anthropological ideas. Emotions are a frequent and important subject in German literary studies of the 17th and 18th century. For example, literary scholars investigated the intended emotional effect [18, 39] or single emotions in plays of that time [2, 37]. We want to expand this hermeneutical research focused mostly on canonical texts. That is why we are applying computational emotion analysis on larger data sets of historical German plays around 1800. We are aiming at a more holistic view of emotion usage, progression and distribution in the plays of that time.

Computational emotion prediction in Natural Language Processing (NLP) describes the task of predicting the expressed emotion, predominantly in written text. Sentiment analysis, its neighbouring field, is focused on the prediction of the valence/polarity of text (if a text unit is rather positive or negative) while emotion prediction deals with more complex

emotion categories like anger, joy, or surprise [16]. Both methods have gained a lot of interest in Digital Humanities (DH) and Computational Literary Studies (CLS) (cf. [9]) and are applied to analyze emotions and sentiment in historical plays [12, 17, 23, 25, 26, 27, 29, 40], novels [6, 12, 21], fairy tales [1, 12], political texts [38], or online forums [14, 35]. DH projects also explore more modern literary genres like fan fictions [8, 7], original creative works on the web [19], subtitles of movies [5, 42] or song lyrics [24]. From a methodological point of view, many of these projects employ lexicon and rule-based methods to perform the sentiment and emotion analysis [1, 12, 17, 19, 23, 24, 25, 38, 40] leading to the development of lexicon-based sentiment analysis tools specifically designed for the DH-community [33]. However, these methods are outperformed by modern machine learning approaches [16]. The reason for the application of lexicon-based methods is the lack of well-annotated corpora of the particular domains that are necessary to train machine learning algorithms [31]. Currently, however, many projects work towards closing this gap and create first corpora of emotion annotated literary texts to explore deep learning based emotion analysis [7, 8]. Annotation of emotions and sentiments can be a challenging task [13, 41]. The task has been shown as even more problematic for historical and poetic texts [1, 28, 30, 36, 32, 38]. While the application of large-scale crowd-sourcing is common for many text types in NLP (cf. [13]), researchers rather refer to expert-based annotation for historical and poetic texts because of the challenges in language and interpretation [1, 28, 30, 36, 32, 38]. Furthermore, due to the high level of subjectivity and complexity of these texts, agreement statistics among expert or common annotators are oftentimes rather low [1, 28, 30, 36, 32, 38] which poses challenges to creating a valid gold standard. Recent research explores the development of tools with gamification elements to improve upon these problems [34, 42].

We present first results of a collaborative project between computer scientists and literary scholars exploring computational emotion analysis on German plays around 1800. Our main corpus currently consists of over 200 plays of that time and we performed our first annotation study on five representative plays of this corpus. We report on annotation results and how we address the challenges of emotion annotation in this field. We developed annotation schemes and processes that are more directed towards the literary scholar's perspective and goals than previous annotation schemes in NLP. Our experience with the annotation and overall results have implications for similar projects designing annotation schemes and performing emotion annotation in the context of CLS.

2 Annotation

In the following, we present the annotation scheme and process we developed. Please note that both the process as well as the scheme have been developed in an iterative process (cf. [22]) of pilot annotations on various scenes and plays of our corpus.

2.1 Annotation Scheme

We define emotion as a generic term for a character's state of mind of distinguishable quality at a given time that is expressed, among other channels, through written language. We annotate emotions experienced by the characters and attributed to them as they are represented as text. Please note that we are interested in the "real" intention and meaning of the expressions of the characters in the context of the entire play. For example, in the case of an ironic expression we annotate the intention of the character in this specific context and not what the text would mean independent of the content and context of the play.

We started the annotation scheme with a list of categorical emotions collected from various established systems of psychology (e.g. [20]) which is rather common in emotion prediction in NLP (cf. [13, 16]). However, we realized that these emotion concepts are missing core emotion and affect ideas important to capture the concept of "emotion" in plays of that time. These are important for the research of literary scholars, however. Therefore, we deviate from established psychological concepts of emotion and integrate concepts such as love and friendship which are not regarded as emotions in many psychological definitions (cf. [15]) but important for this type of literature. We continued with some pilot annotations with a very large scheme containing various emotional concepts important throughout literary history. However, the sheer size and complexity hindered the annotation process. The set was filtered on the most important concepts for literary studies for this time and genre. The final annotation scheme for emotions consists of the 13 concepts mentioned below. In brackets we include the German original terms since we do annotate in German. We translated them to English to the best of our knowledge, but semantic details might get skewed.

- Emotions of affection (*Emotionen der Zuneigung*)
  - Desire (*Lust*) (−)
  - Love (*Liebe*) (+)
  - Friendship (*Freundschaft*) (+)
  - Adoration (*Verehrung*) (+)
- Emotions of joy (*Emotionen der Freude*)
  - Joy (*Freude*) (+)
  - Schadenfreude (+)
- Emotions of fear (*Emotionen der Furcht*)
  - Fear (*Angst*) (−)
  - Despair (*Verzweiflung*) (−)
- Emotions of suffering (*Emotionen des Leids*)
  - Suffering (*Leid*) (−)
  - Compassion (*Mitleid*) (−)
  - Anger (*Ärger*) (−)
- Other
  - Hate (*Abscheu*) (−)
  - Emotional movement (*Emotionale Bewegtheit*)

We defined the set in a hierarchical order to deal with the imbalance problem or too few annotations in the later computational emotion prediction by mapping the emotions to the four main classes and two special types (*hate*, *emotional movement*). Emotional movement is used to annotate unspecific emotional arousal (that cannot be described with the other concepts) as well as astonishment. In the highest hierarchical order, emotions are represented by the two classes *positive* and *negative* (valence). We established a default valence for each emotion concept (marked as + and − in the above list) but annotators can also choose to deviate from this or mark an emotion as mixed via an attribute attached with each emotion annotation. *Schadenfreude*, although ambivalent, is assigned as positive per default since most of the time the emotion is perceived as positive by the experiencer in our texts. As with all emotions, annotators can however deviate from this assignment.

While this set of emotions is still not sufficient to fully capture emotional representations in the literature of that time, it is a compromise between the larger interest of literary scholars and the pragmatic limitations for the computational perspective as well as for the annotation process. Annotators annotate speeches (single utterances of a character separated

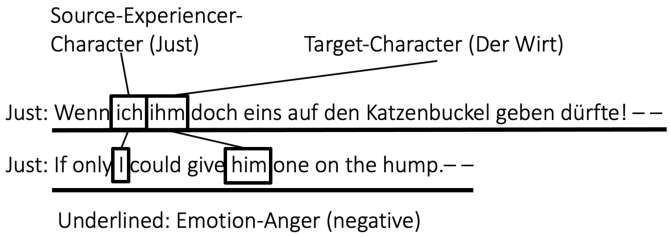

Figure 1 Illustration of an example annotation from Lessing's *Minna von Barnhelm* (Act 1, Scene 3): First line is the German original, second line an English translation. The entire sentence is annotated with anger. "Ich" (I) is the source, "ihm" (him) the target of the emotion. "Just" and "Der Wirt" are the names of the specific characters.

by the next utterance) and stage directions of the plays. They can annotate as much or little text as necessary but not spanning multiple speeches. Therefore, annotators can annotate single words, parts of sentences or multiple sentences. Text units can also consist of multiple or partially overlapping emotion annotations. We have decided to employ this variable and free annotation process since it is in line with the usual annotation work of literary scholars.

Following ideas of aspect-based sentiment analysis [7] we also annotate the source (the character experiencing or attributing an annotated emotion) and the target (the instance an emotion is directed towards). Similar to the emotion set, the set for source and target was adjusted and developed throughout multiple pilot annotation iterations and deemed important for the literary studies perspective since the emotional interaction of the characters are the main aspects of these plays. Source and target consist of the following sub types and possible attributes:

- Source
  - Experiencer: characters of the play, the author, impersonal, unknown
  - Attributing instance: characters of the play, the author, impersonal, unknown
- Target
  - Character: characters of the play, impersonal, unknown
  - Non-Character: animal, state, event or object

Impersonal is a mark for addressing the general public while *unknown* points to characters that are not in the original character list of the play, which is the standard selection of which the annotators can select the characters of the play by their name. The annotation window is as variable as with the emotion annotation. Annotators mark each explicit mention of source and target in the annotated text. In certain cases, it is however possible that an emotion annotation consists of neither source nor target. Figure 1 illustrates the annotation of one example speech of our corpus consisting of an emotion annotation and an explicit annotation of source and target of this annotation.

2.2 Corpus

To start the annotation we decided to annotate five plays of different genres and authors of our main corpus. Plays are annotated in their entirety since we are interested in context and content dependent annotations that need thorough interpretation of the entire plot. While this poses challenges to later generalization processes on the computational side, this is in line with the focus of this project on literary criticism. One aspect to deal with this problem is to annotate plays that are representative concerning content and language of

clusters of the 200 plays corpus. Most plays are taken from the *GerDracor* corpus [3], one play was taken from a free repository.[1] The following five plays have been annotated: *Minna von Barnhelm* (1767) by Lessing (comedy), *Kasperl' der Mandolettikrämer* (1789) by Eberl (comedy), *Kabale und Liebe* (1784) by Schiller (tragedy), *Menschenhass und Reue* (1790) by Kotzebue (comedy), *Faust. Eine Tragödie* (1807) by Goethe (tragedy).

2.3 Annotation Process

Since the annotation of the plays is dependent on deeper knowledge of the language and the content of the plays (as we perform context-aware annotation), crowd-sourcing annotations was not a viable option. In similar projects, annotations are performed by experts and semi-experts with a specific training [1, 28, 30, 31, 38]. In our setting, each play was annotated independently from each other by two students of German literary studies who are compensated monetarily for the annotations and who are employed in the research project. For the annotation of this corpus, we employed three annotators; each play was annotated in different combinations of annotator pairs. The students were introduced to the annotation guidelines by a literary scholar during multiple annotation training sessions and they were offered support during the annotation process. The students participated in the pilot annotation studies to determine the annotation scheme, as well. They had access to an annotation guidelines document consisting of a description of the scheme and multiple examples. The annotation was performed with the tool CATMA [4] for which we created the annotation scheme as described. The annotators were assigned to a play and had a specific deadline to finish it. Depending on the length of the play, each annotator had one to two weeks time to finish the annotation. On average the entire annotation process was performed throughout multiple days of the set time frame and took around 8–12 hours concerning the absolute duration.

3 Annotation Results

We collected over 6,500 emotion annotations for the five plays. First, we look at annotation distributions among the main and sub categories as well as statistics of the annotation lengths via token statistics (see Table 1).

The most frequent annotated emotions are *suffering* (15%) *joy* (13%), *anger* (13%) and *love* (12%). Some emotions are annotated rather rarely in our study e.g. *desire* (1%), *friendship* (2%) and *Schadenfreude* (3%). The main categories themselves are annotated more equally, however with a dominance of more negative emotion categories like the *emotions of suffering* (33%). *Emotional movement* has been proven as an important annotation category (12%). Looking at the size of annotations, all categories have rather similar averages (around 25 tokens) with a large variance ranging from one word annotations to larger paragraphs consisting of over 300 tokens. The dominance of negative emotions is supported by the distribution of the highest hierarchical order valence: 54% of all emotion annotations were either per default *negative* or marked via an attribute as *negative* compared to 34% of *positive* assignments. The remaining annotations in the highest hierarchical order were *emotional movement* annotations (11%). The possibility to select *mixed* as attribute for annotations was rarely used. This attribute has been shown to be rather redundant in our scheme since annotators can assign multiple emotions with differing valence to one text unit.

[1] `http://lithes.uni-graz.at/maezene/eberl_mandolettikraemer.html`

Table 1 Distribution of emotions and corresponding main categories. First, the sub emotions are listed followed by the summed results of the main categories in bold. Percentages are rounded.

Emotion	absolute	%	avg. tokens	min tokens	max tokens	std. tokens
Desire	50	1	23.22	4	83	16.49
Love	783	12	26.16	1	326	33.67
Friendship	127	2	22	1	120	18.66
Adoration	306	5	19.63	1	96	16.36
Emotions of affection	1,266	19	24.05	1	326	28.61
Joy	850	13	22.78	1	223	24.3
Schadenfreude	201	3	25.02	1	121	21.89
Emotions of joy	1,051	16	23.21	1	223	23.86
Fear	424	6	16.87	1	173	17.45
Despair	282	4	30.78	1	206	30.15
Emotions of fear	706	11	22.42	1	206	24.32
Suffering	998	15	26.12	1	302	28.91
Compassion	318	5	21.61	1	156	21.87
Anger	880	13	22.14	1	261	24.35
Emotions of suffering	2,196	33	23.87	1	302	26.27
Hate	614	9	25.05	1	167	26.19
Emotional movement	763	12	24.4	1	313	32.74

Table 2 Agreement statistics per play for the overall valence, the main emotion class and the sub emotions respectively for the text unit of speeches. κ refers to Cohen's κ while % is the proportion of agreed upon speeches among all speeches.

Drama	Valence (κ)	Valence (%)	Class (κ)	Class (%)	Emotion (κ)	Emotion (%)
Faust	0.44	67.853	0.345	59.399	0.342	58.064
Kabale und Liebe	0.382	58.908	0.325	50.313	0.312	47.992
Menschenhass und Reue	0.402	75.28	0.347	72.331	0.347	71.91
Minna von Barnhelm	0.406	74.619	0.377	72.752	0.356	71.23
Kasperl' der Mandolet-tikrämer	0.42	70.83	0.344	65.34	0.312	62.72
Overall	0.41	69.498	0.3476	64.027	0.333	62.383

Table 3 Source and target distributions. The sub categories are listed followed by the summed results of the main categories in bold. The percentages of the sub groups refer to the main class.

Annotation Type	absolute	%	avg. tokens	min tokens	max tokens	std. tokens
Experiencer	6,573	97	1.06	1	7	0.33
Attributing Instance	187	3	1.05	1	3	0.27
Source	6,760	50	1.06	1	7	0.33
Character	5,336	79	1.28	1	14	0.82
Non-Character	1,390	21	3.97	1	26	3.68
Target	6,726	50	1.84	1	26	2.13

Since the annotations are performed on variable text lengths, we decided on the following heuristic to calculate agreement among annotators: We focus on the speech and stage directions as central structural units of plays. They can consist of one word to multiple sentences. For every annotator we assign the specific emotion that is annotated the most (in total token count) for one speech. Thus, if multiple emotions are annotated, we assign the emotion that is annotated the most. We decided for this heuristic in order to be able to apply the traditional agreement metric Cohen's κ and get a first overview of agreement among annotators. We explore possibilities for more fitting fuzzy agreement metrics in future work. If no emotion was annotated the unit is marked as none. None is regarded as additional annotation class in this concept. Table 2 illustrates the agreements. The κ-value according to Cohen's κ is shown as well as the percentage wise agreement. We identified mostly moderate agreement for the valence according to [11] (0.41-0.6) and fair agreement for the main emotion category and the sub emotions (0.21-0.4). Due to the higher number of classes the agreement gets lower for the sub emotions.

We also gathered over 12,000 *source* and *target* annotations (see Table 3). Both classes are annotated to an equal extent. For sources, characters are mostly marked as *experiencer* of emotions (97%) and rarely as the ones attributing emotions to other characters (3%). Targets of emotions are mostly *characters* (79%). For the sub groups of theses classes, the following findings could be made. *Sources*, being it *experiencer* or *attributing instances*, are for the most part *one character* (94%) or *multiple characters* (2%). The attributes for *unknown* and *impersonal* sources are rarely used (2%). If a character is chosen as a *target*, the distribution is similar with *one to multiple characters* being the most frequent annotation (89%) compared to *unknown* (7%) and *impersonal* (4%). If the *target* is a *non-character*, the attribute assigned most frequently is *event* (61%) followed by *state* (19%), *objects* (16%) and *animals* (4%). Regarding the annotation lengths, source and target annotations are mostly one word annotations like pronouns or character names which points towards token based prediction mechanisms in later computational approaches to predict source and target.

4 Discussion

The annotated corpus will be made publicly available and is currently in the process of preparation.

To validate findings of the annotation analysis, we discussed our results with the annotators after the annotation. The extension of our scheme beyond established categories of psychology has been well received by annotators and we recommend this for similar projects. Concepts such as *love*, *suffering* and *emotional movement* are important parts of literature of that time and genre and have been annotated in large numbers. However, other concepts such as *desire* or *adoration* were rarely annotated. We are discussing the need for these concepts

since any complexity reduction of the scheme is beneficial for annotation speed, consistency and the later prediction. Please note that the annotation of emotions is highly influenced by the plays chosen to be annotated. Concepts such as *desire* and *adoration* are more important for earlier periods which we will investigate in the future and which will likely lead to the collection of more annotations. Looking at the main categories, the distribution becomes more equal. Negative categories are more frequent, although the majority of our chosen corpus consists of comedies. This is in line with previous annotation results in similar contexts [1, 28, 30] showing that negativity is an integral part of the narrative of most plays. The genre assignment comedy just points towards a positive ending, the play itself still consists of conflicts and disputes up until the end. Annotators rarely used the annotation of the attribute *mixed* for emotion. This attribute is redundant in our scheme and will be discarded. Considering *source* and *target*, we identified that annotators mostly annotate them as characters and not as non-characters which is quite intuitive in light of the content of the plays which are driven by emotional interactions of characters. We will reflect upon the question if differentiated sub classes for *non-characters* make sense if this main class is annotated rather rarely. The variable annotation lengths have also been perceived rather positively by the annotators and we also recommend the application of this idea for similar projects in a literary studies context. Emotions were annotated in variable sizes concerning number of words and sentences. This resembles the reality of the emotion expressions in these plays and is also in line with the general annotation behavior of literary scholars. Forcing annotations for a concrete window size would be challenging for decision processes during the annotation and would prolong and complicate the process. We plan to apply heuristics to map annotations on structural units and perform speech, sentence, n-gram and token based multi-label emotion prediction in our computational approaches.

The current agreement results indicate fair to moderate agreement. This is mostly in line with results of projects with similar text types [1, 28, 30, 32, 36, 38] since the material is more subjective and challenging to interpret. Our approach to perform context-sensitive annotation reinforces this aspect. In future work, we plan to explore sentence and token based agreements but also agreements of source and target annotations to get a better overview of the annotation problems. We also see potential in fuzzy agreement scores to represent the agreement in our variable and complex setting in a more fitting way [10] since our heuristic certainly leads to further disagreement in certain instances. Furthermore, we argue that we will reach higher agreements the more experience the annotators gain. To support this process and to find a way to deal with the disagreements among the annotators, we decided to add a subsequent post-annotation phase after the first two independent annotations by the students. This post-annotation phase is performed under the guidance of a literary scholar expert annotator who discusses the annotation with the students and creates a *consensus annotation* during these sessions. Although this might increase the annotation duration, it will improve the understanding of all annotators and might lead to more consistent annotations. Kajava et al. [5] argue that κ-values of 0.6 are acceptable for multi-label emotion annotations to validate the consistency of a scheme. The consensus annotation will also be the material we use to train and evaluate computational emotion analysis based on machine learning. We will adjust the annotation scheme and continue the annotations in the described way.

References

1 Cecilia Ovesdotter Alm and Richard Sproat. Emotional Sequencing and Development in Fairy Tales. In Jianhua Tao, Tieniu Tan, and Rosalind W. Picard, editors, *Affective Computing and Intelligent Interaction*, Lecture Notes in Computer Science, pages 668–674, Berlin, Heidelberg, 2005. Springer. doi:10.1007/11573548_86.

2 Thomas Anz. Todesszenarien : literarische Techniken zur Evokation von Angst, Trauer und anderen Gefühlen. In Lisanne Ebert, editor, *Emotionale Grenzgänge. Konzeptualisierungen von Liebe, Trauer und Angst in Sprache und Literatur*, pages 113–129. Königshausen & Neumann, Würzburg, 2011.

3 Frank Fischer, Ingo Börner, Mathias Göbel, Angelika Hechtl, Christopher Kittel, Carsten Milling, and Peer Trilcke. Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama, July 2019. Conference Name: Digital Humanities 2019: "Complexities" (DH2019) Publisher: Zenodo. doi:10.5281/zenodo.4284002.

4 Evelyn Gius, Jan Christoph Meister, Marco Petris, Malte Meister, Christian Bruck, Janina Jacke, Mareike Schuhmacher, Marie Flüh, and Jan Horstmann. CATMA, 2020. doi:10.5281/zenodo.4353618.

5 Kaisla Kajava, Emily Öhman, Piao Hui, and Jörg Tiedemann. Emotion Preservation in Translation: Evaluating Datasets for Annotation Projection. In *Proceedings of Digital Humanities in Nordic Countries (DHN 2020)*, pages 38–50. CEUR, 2020.

6 Tuomo Kakkonen and Gordana Galić Kakkonen. SentiProfiler: Creating Comparable Visual Profiles of Sentimental Content in Texts. In *Proceedings of the Workshop on Language Technologies for Digital Humanities and Cultural Heritage*, pages 62–69, Hissar, Bulgaria, 2011. Association for Computational Linguistics.

7 Evgeny Kim and Roman Klinger. Who Feels What and Why? Annotation of a Literature Corpus with Semantic Roles of Emotions. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1345–1359, Santa Fe, New Mexico, USA, 2018. Association for Computational Linguistics. URL: https://www.aclweb.org/anthology/C18-1114.

8 Evgeny Kim and Roman Klinger. An Analysis of Emotion Communication Channels in Fan-Fiction: Towards Emotional Storytelling. In *Proceedings of the Second Workshop on Storytelling*, pages 56–64, Florence, Italy, August 2019. Association for Computational Linguistics.

9 Evgeny Kim and Roman Klinger. A Survey on Sentiment and Emotion Analysis for Computational Literary Studies. *Zeitschrift für digitale Geisteswissenschaften*, 2019. arXiv: 1808.03137. doi:10.17175/2019_008.

10 Andrei P. Kirilenko and Svetlana Stepchenkova. Inter-Coder Agreement in One-to-Many Classification: Fuzzy Kappa. *PLOS ONE*, 11(3):e0149787, 2016. Publisher: Public Library of Science. doi:10.1371/journal.pone.0149787.

11 J. Richard Landis and Gary G. Koch. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–174, 1977. Publisher: [Wiley, International Biometric Society].

12 Saif Mohammad. 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*, pages 105–114, Portland, OR, USA, 2011. Association for Computational Linguistics.

13 Saif Mohammad. A Practical Guide to Sentiment Annotation: Challenges and Solutions. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 174–179, San Diego, California, June 2016. Association for Computational Linguistics. doi:10.18653/v1/W16-0429.

14 Luis Moßburger, Felix Wende, Kay Brinkmann, and Thomas Schmidt. Exploring online depression forums via text mining: A comparison of Reddit and a curated online forum. In *Proceedings of the Fifth Social Media Mining for Health Applications Workshop & Shared Task*, pages 70–81, Barcelona, Spain (Online), December 2020. Association for Computational Linguistics. URL: https://www.aclweb.org/anthology/2020.smm4h-1.11.

15 Kevin Mulligan and Klaus R. Scherer. Toward a Working Definition of Emotion. *Emotion Review*, 4(4):345–357, 2012. Publisher: SAGE Publications. doi:10.1177/1754073912445818.

16 Mika V. Mäntylä, Daniel Graziotin, and Miikka Kuutila. The evolution of sentiment analysis – A review of research topics, venues, and top cited papers. *Computer Science Review*, 27:16–32, February 2018. doi:10.1016/j.cosrev.2017.10.002.

17 Eric T. Nalisnick and Henry S. Baird. Character-to-Character Sentiment Analysis in Shakespeare's Plays. In *Proceedings of the 51st Annual Meeting of the Association for Computa-*

tional Linguistics (Volume 2: Short Papers), pages 479–483, Sofia, Bulgaria, August 2013. Association for Computational Linguistics. URL: https://www.aclweb.org/anthology/P13-2085.

18 Winfried Nolting. *Die Dialektik der Empfindung: Lessings Trauerspiele "Miss Sara Sampson" und "Emilia Galotti": mit einer Einleitung, gemischte Gefühle: zur Problematik eines explikativen Verstehens des Empfindung.* Number 1 in Studien zu einer Geschichte der literarischen Empfindung / Winfried Nolting. F. Steiner Verlag Wiesbaden, Stuttgart, 1986.

19 Federico Pianzola, Simone Rebora, and Gerhard Lauer. Wattpad as a resource for literary studies. Quantitative and qualitative examples of the importance of digital social reading and readers' comments in the margins. *PLOS ONE*, 15(1):e0226708, 2020. Publisher: Public Library of Science. doi:10.1371/journal.pone.0226708.

20 Robert Plutchik. *Emotion, a Psychoevolutionary Synthesis.* Harper & Row, 1980. Google-Books-ID: G5t9AAAAMAAJ.

21 Andrew J. Reagan, Lewis Mitchell, Dilan Kiley, Christopher M. Danforth, and Peter Sheridan Dodds. The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, 5(1):31, 2016. arXiv: 1606.07772. doi:10.1140/epjds/s13688-016-0093-1.

22 Nils Reiter. Anleitung zur Erstellung von Annotationsrichtlinien. In *Reflektierte algorithmische Textanalyse*, pages 193–202. De Gruyter, 2020. doi:10.1515/9783110693973-009.

23 Thomas Schmidt. Distant reading sentiments and emotions in historic german plays. In *Abstract Booklet, DH_Budapest_2019*, pages 57–60. Budapest, Hungary, September 2019. URL: https://epub.uni-regensburg.de/43592/.

24 Thomas Schmidt, Marlene Bauer, Florian Habler, Hannes Heuberger, Florian Pilsl, and Christian Wolff. Der einsatz von distant reading auf einem korpus deutschsprachiger songtexte. In Christof Schöch, editor, *DHd 2020: Spielräume; Digital Humanities zwischen Modellierung und Interpretation. Konferenzabstracts; Universität Paderborn, 02. bis 06. März 2020*, pages 296–300. Paderborn, Germany, 2020. URL: https://epub.uni-regensburg.de/43704/.

25 Thomas Schmidt and Manuel Burghardt. 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*, pages 139–149, Santa Fe, New Mexico, August 2018. Association for Computational Linguistics. URL: https://www.aclweb.org/anthology/W18-4516.

26 Thomas Schmidt and Manuel Burghardt. Toward a Tool for Sentiment Analysis for German Historic Plays. In Michael Piotrowski, editor, *COMHUM 2018: Book of Abstracts for the Workshop on Computational Methods in the Humanities 2018*, pages 46–48, Lausanne, Switzerland, June 2018. Laboratoire laussannois d'informatique et statistique textuelle. URL: https://zenodo.org/record/1312779.

27 Thomas Schmidt, Manuel Burghardt, and Katrin Dennerlein. "Kann man denn auch nicht lachend sehr ernsthaft sein?" – Zum Einsatz von Sentiment Analyse-Verfahren für die quantitative Untersuchung von Lessings Dramen. In Georg Vogeler, editor, *Kritik der digitalen Vernunft. Abstracts zur Jahrestagung des Verbandes Digital Humanities im deutschsprachigen Raum, 26.02. - 02.03.2018 an der Universität zu Köln*, pages 244–249, Köln, 2018. Universitäts- und Stadtbibliothek Köln.

28 Thomas Schmidt, Manuel Burghardt, and Katrin Dennerlein. Sentiment annotation of historic german plays: An empirical study on annotation behavior. In Sandra Kübler and Heike Zinsmeister, editors, *annDH 2018, Proceedings of the Workshop on Annotation in Digital Humanities 2018 (annDH 2018), Sofia, Bulgaria, August 6-10, 2018*, pages 47–52. RWTH Aachen, Aachen, August 2018. URL: https://epub.uni-regensburg.de/43701/.

29 Thomas Schmidt, Manuel Burghardt, Katrin Dennerlein, and Christian Wolff. Katharsis – a tool for computational drametrics. In *Book of Abstracts, Digital Humanities Conference 2019 (DH 2019)*. Utrecht, Netherlands, 2019. URL: https://epub.uni-regensburg.de/43579/.

30 Thomas Schmidt, Manuel Burghardt, Katrin Dennerlein, and Christian Wolff. Sentiment annotation for lessing's plays: Towards a language resource for sentiment analysis on german literary texts. In Thierry Declerck and John P. McCrae, editors, *2nd Conference on Language,*

Data and Knowledge (LDK 2019), pages 45–50. RWTH Aachen, Aachen, May 2019. URL: https://epub.uni-regensburg.de/43569/.

31 Thomas Schmidt, Manuel Burghardt, and Christian Wolff. Herausforderungen für Sentiment Analysis-Verfahren bei literarischen Texten. In Manuel Burghardt and Claudia Müller-Birn, editors, *INF-DH-2018*, Berlin, Germany, September 2018. Gesellschaft für Informatik e.V. doi:10.18420/infdh2018-16.

32 Thomas Schmidt, Manuel Burghardt, and Christian Wolff. Toward Multimodal Sentiment Analysis of Historic Plays: A Case Study with Text and Audio for Lessing's Emilia Galotti. In Costanza Navarretta, Manex Agirrezabal, and Bente Maegaard, editors, *Proceedings of the Digital Humanities in the Nordic Countries 4th Conference*, volume 2364 of *CEUR Workshop Proceedings*, pages 405–414, Copenhagen, Denmark, March 2019. CEUR-WS.org. URL: http://ceur-ws.org/Vol-2364/37_paper.pdf.

33 Thomas Schmidt, Johanna Dangel, and Christian Wolff. Senttext: A tool for lexicon-based sentiment analysis in digital humanities. In Thomas Schmidt and Christian Wolff, editors, *Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021)*, volume 74, pages 156–172. Werner Hülsbusch, Glückstadt, 2021. URL: https://epub.uni-regensburg.de/44943/.

34 Thomas Schmidt, Marco Jakob, and Christian Wolff. Annotator-centered design: Towards a tool for sentiment and emotion annotation. In Claude Draude, Martin Lange, and Bernhard Sick, editors, *INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft (Workshop-Beiträge)*, pages 77–85, Bonn, 2019. Gesellschaft für Informatik e.V. doi:10.18420/inf2019_ws08.

35 Thomas Schmidt, Florian Kaindl, and Christian Wolff. Distant reading of religious online communities: A case study for three religious forums on reddit. In *DHN*, pages 157–172, Riga, Latvia, 2020.

36 Thomas Schmidt, Brigitte Winterl, Milena Maul, Alina Schark, Andrea Vlad, and Christian Wolff. Inter-rater agreement and usability: A comparative evaluation of annotation tools for sentiment annotation. In Claude Draude, Martin Lange, and Bernhard Sick, editors, *INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft (Workshop-Beiträge)*, pages 121–133, Bonn, 2019. Gesellschaft für Informatik e.V. doi:10.18420/inf2019_ws12.

37 Anja Schonlau. *Emotionen im Dramentext: eine methodische Grundlegung mit exemplarischer Analyse zu Neid und Intrige 1750-1800*. Number Band 25 in Deutsche Literatur. De Gruyter, Berlin Boston, 2017. OCLC: 978262308.

38 Rachele Sprugnoli, Sara Tonelli, Alessandro Marchetti, and Giovanni Moretti. Towards sentiment analysis for historical texts. *Digital Scholarship in the Humanities*, 31:762–772, 2015. Publisher: Oxford : Oxford University Press. doi:10.1093/llc/fqv027.

39 Hermann Wiegmann, editor. *Die ästhetische Leidenschaft: Texte zur Affektenlehre im 17. und 18. Jahrhundert*. Number 27 in Germanistische Texte und Studien. Olms, Hildesheim, 1987. OCLC: 15741918.

40 Mehmet Can Yavuz. 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, 2021.

41 Emily Öhman. Challenges in Annotation: Annotator Experiences from a Crowdsourced Emotion Annotation Task. In *Proceedings of the Digital Humanities in the Nordic Countries 5th Conference*, pages 293–301. CEUR Workshop Proceedings, 2020.

42 Emily Sofi Öhman and Kaisla S. A. Kajava. Sentimentator: Gamifying Fine-grained Sentiment Annotation. In *Digital Humanities in the Nordic Countries 2018*, pages 98–110. CEUR Workshop Proceedings, 2018.