Exploring Multimodal Sentiment Analysis in Plays: A Case Study for a Theater Recording of Emilia Galotti

Thomas Schmidt[1], Christian Wolff[2]

[1]*Media Informatics Group, University of Regensburg, D-93040, Regensburg, Germany*
[2]*Media Informatics Group, University of Regensburg, D-93040, Regensburg, Germany*

Abstract

We present first results of an exploratory study about sentiment analysis via different media channels on a German historical play. We propose the exploration of other media channels than text for sentiment analysis on plays since the auditory and visual channel might offer important cues for sentiment analysis. We perform a case study and investigate how textual, auditory (voice-based), and visual (face-based) sentiment analysis perform compared to human annotations and how these approaches differ from each other. As use case we chose *Emilia Galotti* by the famous German playwright Gotthold Ephraim Lessing. We acquired a video recording of a 2002 theater performance of the play at the "Wiener Burgtheater". We evaluate textual lexicon-based sentiment analysis and two state-of-the-art audio and video sentiment analysis tools. As gold standard we use speech-based annotations of three expert annotators. We found that the audio and video sentiment analysis do not perform better than the textual sentiment analysis and that the presentation of the video channel did not improve annotation statistics. We discuss the reasons for this negative result and limitations of the approaches. We also outline how we plan to further investigate the possibilities of multimodal sentiment analysis.

Keywords

sentiment analysis, computational literary studies, video, annotation, multimodality

1. Introduction

Sentiments and emotions are an important part of qualitative and hermeneutical analysis in literary studies and are important cues for the understanding and interpretation of narrative art (cf. [49, 22, 23, 53]). The computational method of predicting and analyzing sentiment, predominantly in written text, is referred to as sentiment analysis and has a long tradition in the computational analysis of social media and user generated content on the web [20]. In general, sentiment analysis regards sentiment (also often referred to as opinion, polarity or valence) as a class-based phenomenon describing the connotation of a text unit as either *positive*, *negative* or *neutral*. The prediction and analysis of more complex categories like anger, sadness or joy (e.g. in a multi-class setting) is called computational emotion analysis [20]. While more complex emotions are of great interest for literary studies, we focus on sentiment solely for our first explorations of sentiment analysis on multiple media channels.

There is a growing interest for sentiment and emotion analysis applications in Digital Humanities (DH), especially in Computational Literary Studies (CLS). Researchers explore these

methods on text genres like fairy tales [1, 24], novels [31, 14, 17, 12], fan fictions [29, 16], movie subtitles [27, 13, 42] and social media content [44, 46]. In the context of historical plays, scholars apply sentiment and emotion analysis to investigate visualization possibilities [24, 34, 37], relationships among characters [34, 26] or evaluate the performance compared to human-based expert annotations [35]. For an in-depth analysis of the current state of sentiment analysis in DH and CLS see [15].

While these projects are very promising, there are several problems concerning the current state of this field: The annotation process has been shown to be rather tedious and challenging, often leading to the need of (expensive) experts that understand context and language of the material. Furthermore, agreement among annotators is low due to the inherent subjectivity of narrative and poetic texts but also due to problems in the interpretation because of historical or vague language [1, 27, 50, 36, 38, 48, 41]. The annotation problems pose challenges to the creation of valid corpora that are necessary for modern machine learning. While some research intends to address this problem by developing user-centered annotation tools [45], the problems still persist. Thus, sentiment and emotion analysis is predominantly performed with rather primitive rule-based methods (lexicon-based methods, cf. [15]) and achieves prediction accuracies on annotated literary corpora between 40 and 80% [35, 2], which is low compared with other text genres like social media or product reviews [20, 52]. As main problems, researchers report the historical and poetic language, as well as the usage of irony and metaphors which pose challenges for rule-based methods that are dependent of a fixed set of contemporary vocabulary annotated with sentiment information (see chapter 4.1).

We argue that the fact that the majority of sentiment analysis in CLS is performed on the textual level might be a reason for the resulting limitations. We propose the exploration of other media channels to advance this area of research. Indeed, multimodal sentiment analysis has proven to be successful in several application areas that offer next to text a video channel [30] and is used in various contexts in human-computer interaction [28, 11, 10]. While many literary texts are at their core texts written for a reading audience, we argue that plays are similar to movie scripts and mostly not intended to be read, but to be performed on stage. In addition one might argue that many narrative forms have their origin in a communicative situation based on spoken language [7] as text is a fairly late invention. Indeed, the important role of the performance and problems with the sole focus on the text for plays have been a major argument and point of discussion throughout the history of theater studies (cf. [51]). Especially concerning attributes like sentiment and emotions, one can argue that they are more so communicated by the actors in a live performance using voice, facial expressions as well as gait and gesture, than via the written text. Furthermore, theater recordings of canonical plays are nowadays easier to access, for example online or via library services.

Thus, we perform a case study about how the inclusion of the auditory (voice-based) and the visual (face-based) channel influences (1) the sentiment annotation process and (2) the computational sentiment prediction. The presentation of the video channel might facilitate the annotation process and improve agreements since annotators do not have to solely rely on the difficult language for interpretation. As case study we select the play *Emilia Galotti* (1772) by Gotthold Ephraim Lessing (1729-1781), one of the most famous German playwrights. We describe the material in more detail in chapter 2. Afterwards, we discuss the creation of an annotated gold standard (chapter 3) and the applied sentiment analysis methods for the three media channels: text, audio and video (chapter 4). In chapter 5 we compare the three methods to each other and investigate the performance on the gold standard before discussing the results in chapter 6.

2. Material

As material for this case study we use the play *Emilia Galotti* by G. E. Lessing. Lessing is one of the most famous German playwrights and *Emilia Galotti* one of his most important plays. The play has already been explored in the context of audio sentiment analysis [39]. For our analysis, we use a recorded theater performance. The performance dates from 2002 and was performed at the "Wiener Burgtheater" in German language.[1]

The recording has decent audio and picture quality and meets the necessary quality requirements as demanded by our sentiment analysis tools. The video format is mp4 with a resolution of 640 x 360 and 25 frames per second. The audio was extracted as a wav-file with a sampling rate of 44.1 kHz stereo. Current research on quantitative drama analysis is focused on speeches as a basic structural unit. A speech is usually a single utterance of a character separated by utterances of other characters beforehand and afterwards. Emilia Galotti consists of 835 speeches. However, it is common that the real stage production of a play does not completely adhere to the published text and order of the original material. Thus, we deviate from the textual speeches offered by the written original and acquire the text as actually spoken by the actors during staging to enable correct comparisons among all sentiment analysis methods.

To acquire the text of this performance we used the *Google Clouds Speech to Text-API* (GCST) for German.[2] GCST is considered state-of-the-art for speech-to-text-tasks and produces text structured by units separated when longer breaks occur during the utterances. In the following, we will refer to these units as "speeches". The API produces 672 of these textual units and is therefore quite different from the original speeches of the textual source material. Note that these text units sometimes consist of utterances of multiple speakers or are separated during a speech or a sentence (depending on when a break appears). Furthermore, GCST is not intended for the usage for video recordings of theater performances and while we did not perform an exact evaluation, we were able to identify that GCST produces various mistakes that are quite substantial for some passages. Thus, in a subsequent step, we corrected the mistakes in the output by listening to the play and transcribing certain passages from scratch. GCST also delivers precise time stamps for the 672 units and we separated the audio- and video-file according to these timestamps to obtain 672 comparable units for every modality.

3. Annotation Process and Results

The speeches were annotated by two annotators who are familiar with Lessing's work and the specific play. Annotators assigned one of three polarity classes to every speech: negative, neutral and positive. The instruction was to assign the class the speech is most connoted with depending on the overall sentiment the characters express. Annotators were shown the entire video of a speech as well as the text (meaning all modalities) via a table and a video player. The annotators were given one week to finish the annotation and were compensated monetarily. We conducted short interviews about the annotation process afterwards, which we discuss briefly in chapter 6.

The annotation results are as follows: The annotators agree upon 348 of the 672 speeches (52%) with a *Cohen's κ*-value of 0.233 (fair agreement according to [18]). These are rather low

---

[1]More information about the recording: https://www.amazon.de/Emilia-Galotti-Lessing-Wiener-Burgtheater/dp/B0038QGXOK

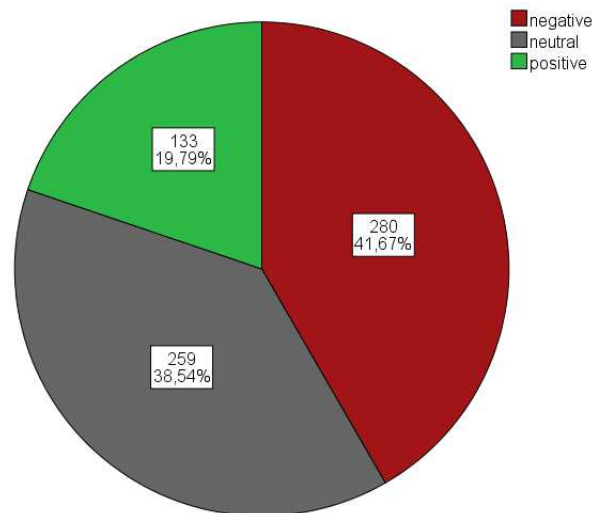[2]https://cloud.google.com/speech-to-text/

Figure 1: Sentiment distribution of gold standard annotations

levels of agreement, which are however in line with previous research concerning annotation of literary or historical texts [1, 27, 42, 50, 36, 38, 48, 41]. We define the gold standard we use for the evaluation of sentiment prediction via the following approach: If annotators agree upon a speech, the speech is assigned the chosen class. Considering speeches the first two annotators did not agree upon, a third expert annotator decided upon the final annotation via the same annotation process as described above. Figure 1 illustrates the distribution of these gold standard annotations.

The majority of annotations are negative, which is in line with previous research considering the annotation of literary texts [1, 42, 36, 38, 48, 41]. The high number of neutral annotations is, according to our analysis, due to the fact that many speeches are very short (e.g. consisting of one word), thus making the assignment of positive or negative sentiment rather difficult.

4. Sentiment Analysis Methods

In this chapter we describe the different sentiment analysis approaches. All approaches were implemented in *Python* with support of various Software Development Kits (SDKs) which we describe more detailed in the upcoming chapters. Statistical analysis was performed in Python or with the *IBM SPSS statistics* software.

4.1. Textual Sentiment Analysis

For the textual sentiment analysis we employ a lexicon-based approach. A sentiment lexicon is a list or table of words annotated concerning sentiment information, e.g. if a word is rather negatively or positively connoted. Due to simple word-based calculations one can infer the sentiment of a text: By summing up the number of positive words and subtracting the number of negative words, one receives an overall value for the sentiment of the text unit which can be regarded as negative if the value is below 0, neutral for 0 and positive if the value is above 0. Oftentimes, sentiment lexicons offer continuous values instead of nominal assignments which

can be used similarly. In research, lexicon-based sentiment analysis is often chosen when machine learning is not possible due to the lack of well annotated corpora and is a common method in sentiment analysis on literary and historical texts [1, 24, 14, 29, 34, 26, 35, 50, 2, 40] or for special social media corpora [44, 46, 25].

We utilize the sentiment lexicon *SentimentWortschatz* (SentiWS) [32], which is one of the most well-known and validated lexicons for German [8] and perform calculations for all speeches as described above. The words in SentiWS are annotated with floating point numbers concerning the polarity on a continuous scale from +3 (very positive) to -3 (very negative). The lexicon consists of 3,469 entries along with their inflections. To address the problem of historical language to some extent we apply the optimizations recommended by Schmidt and Burghardt [35]: lemmatization via *treetagger* [33] and the extension of the lexicon with historical variants. This lexicon-based approach has been shown to be successful in the setting of German historical plays compared to more basic lexicon-based approaches.

4.2. Audio Sentiment Analysis

For the audio sentiment analysis, we use the free developer version of the tool *Vokaturi*.[3] Vokaturi is an emotion recognition tool for spoken language employing machine learning. It is considered to be language independent, is recommended as the best free software for sentiment analysis of spoken language [9] and used in similar comparative research [47]. To implement the analysis, Vokaturi uses machine learning on two larger databases with voice- and audio-based features. We use each of the 672 speeches as input for Vokaturi and receive numerical values on a range from 0 (none) – 1 (a lot) for the five categories *neutrality*, *fear*, *sadness*, *anger* and *happiness*. The value specifies to which degree the corresponding concept is present in the audio file. However, the tool does not report a sentiment/valence score directly. Thus, to map this output to the concept of the sentiment classes (positive/negative/neutral), we apply the following heuristic: we sum up the values for the negative emotions fear, sadness and anger to get an overall value for the negative polarity. We regard the value of happiness as the positive polarity. We then compare these two values and the value for neutrality and assign the maximum of these three values as overall sentiment of the speech. We refer to this method as audio sentiment analysis.

4.3. Visual Sentiment Analysis

To conduct the video sentiment analysis we utilize the free version of the Emotion SDK of *Affectiva*.[4] The Affectiva Emotion SDK is a cross-platform face-expression recognition toolkit focused on face detection and facial emotion recognition [21] also used in various research fields [19]. According to Affectiva, the analysis is based on a large training database with over 6.5 million faces.

To perform the video sentiment analysis we segment the 672 video parts into frames, one frame per second. Then, we use the facial emotion recognition of the Affectiva Emotion SDK on all of the frames. The SDK produces multiple values relevant for emotion recognition. However, we solely rely on the *valence* value, which is a value produced to describe the overall expression of the face as rather positive or negative. The valence value is positive if predominately positive emotions are recognized and negative for predominately negative emotions. The value can also

---

[3]https://vokaturi.com/
[4]https://www.affectiva.com/

Table 1
Distribution of sentiment classes output per modality approach. # marks the absolute number and % the proportion of the sentiment classes among all speeches.

	negative (#)	negative (%)	neutral (#)	neutral (%)	positive (#)	positive (%)
Textual	313	46.58	187	27.83	172	25.60
Audio (Voice)	420	62.50	62	9.23	190	28.27
Video (Facial)	137	20.39	490	72.92	45	6.70

Table 2
Accuracy results (proportion of correctly predicted speeches) per modality approach. The absolute number of correctly predicted speeches is in brackets.

	Textual	Audio (Voice)	Video (Facial)
Accuracy	46% (311)	40% (264)	44% (295)

be zero if no emotion is apparent or no face can be detected. We sum up all valence values of all frames corresponding to the time-frame of a speech and then assign the sentiment accordingly: positive if the overall valence is positive, negative if the overall valence is negative and neutral for a value of 0. Note, that we configured the SDK to choose the valence of the largest face that is detected on the frame. We will refer to this method as video sentiment analysis in the following.

5. Results

5.1. Comparison of Textual, Audio and Visual Sentiment Analysis

First, we report the general frequency distributions concerning the predicted sentiment of all three modalities: text, audio and video for all 672 speeches (see table 1).

All approaches produce very different results: The textual sentiment analysis predicts the majority of speeches as negative (47%). Neutral predictions are mostly due to short speeches consisting of only few words with no representation in the sentiment lexicon. They are however slightly more frequent (28%) than positive predictions (26%). In contrast, the audio sentiment analysis rarely assigns the neutral class (10%) while negative predictions are dominant (63%). The video sentiment analysis predicts the vast majority of speeches as neutral (73%) and only a small fraction as positive (7%). We identified that the reason for this behavior is that faces are not identified due to difficult angles and camera movements. Thus, no emotion recognition is performed and the frames are regarded as neutral.

5.2. Performance Evaluation

In the following section, we report on the sentiment prediction accuracies of the computational methods using the annotations as gold standard (Table 2). The overall accuracy is the proportion of correctly predicted speeches among all speeches. The random baseline is 33%, the majority baseline is around 42%.

All approaches are above the random baseline and some slightly above the majority baseline. Overall, the accuracies are rather close to each other and no significant differences are identified. The highest accuracy is achieved with the textual approach (46%) followed by the video (44%) and the audio (40%) sentiment analysis. The results are however way below reported accuracies

Figure 2: Frame of a speech correctly predicted as negative by the video/facial sentiment analysis

in similar and different fields: Lexicon-based sentiment analysis on literary texts achieves around 40-80% [35, 2]. Modern deep learning-based approaches in other research areas can achieve up to 95% [20]. In a similar study comparing text to audio of a theater recording, the results are equivalent, however [39].

All data (corpus, annotations and results) is publicly available via a GitHub-repository.[5]

6. Discussion

While we identified that all approaches behave rather differently, the accuracy levels for them are below the results of sentiment analysis with other text and media genres applying state-of-the-art machine learning using large training corpora of the fitting domain [20, 52, 30]. In the context of literary texts, the results are in line with the overall – mediocre – accuracies of other studies applying lexicon-based [1, 35] or audio-based methods [39], thus proving again the general difficulty of the task. We could not show that the audio or video sentiment analysis approach outperforms textual sentiment analysis. While problems like historical language might be solved, novel problems occur that decrease the performance of audio and video sentiment analysis, although the applied approaches showed state-of-the-art results with other media material like social media videos [30, 9]. The audio and video approaches perform rather well for extreme emotional expressions (see figure 2 for a correct example). However, the video approach is dependent on the picture quality and has problems with bad lightning, disadvantageous camera positions and when actors express a complex grimace (figure 3). Indeed, facial sentiment analysis is mostly trained on images of people looking directly towards the camera (cf. [5]) which is rarely the case for a live theater recording. Thus, no faces are detected and no emotion recognition can be performed which lead to many false neutral predictions. The audio sentiment analysis detects emotional nuances even in short speeches; however, the annotators tend to rate those speeches as rather neutral. Many problems well known from textual sentiment analysis also remain: how to deal with irony and sarcasm, long speeches or switching between sentiments during a speech.

Despite the mediocre results of the case study, plays are meant to be performed and thus experienced with multiple modalities. Therefore, the application of multimodal sentiment analysis is closer to the artistic experience of the theatergoers as it is intended to be. We want to pursue this idea further by changing the material from theater performances of historical plays to rather simple contemporary movies that might be less challenging considering camera

---

[5]https://github.com/lauchblatt/Video-Emotion-Analysis-for-EmiliaGalotti

Figure 3: Frame of a speech falsely predicted as positive by the video/facial sentiment analysis although negative

angles, performance and audio quality. Thus, we want to investigate to what extent the quality and complexity of the material influences the approaches.

Furthermore, we have focused on ready-to-use sentiment analysis approaches without optimization or domain adaptation for the specific context, which is not uncommon in DH. However, the usage of general-purpose approaches did not prove to be beneficial or at least acceptable. Lexicon-based methods for textual sentiment analysis might very well not be able to deal with the historical language and the nuanced emotional expressions of plays. The audio and video-based models are, of course, trained on contemporary online videos and not on theater recordings. Therefore, we want to explore more sophisticated approaches based on machine learning on the specific material used. On a conceptual level, all three modalities might suffer from the fact that we did not integrate a "mixed"-class in the prediction. Especially for longer passages, change of sentiment can occur and might lead to false interpretations. More sophisticated tools will however enable us to explore the integration of this class in more detail. Another reason for problems concerning the computational predictions but also the corresponding annotations might be that annotators included other cues besides language, face and voice into their interpretation. Indeed, research shows that body cues and body movement might be more important cues for emotional expressions [3], which is something that the applied tools mostly neglect but could be investigated via pose detection [6]. Furthermore, individual differences in the expression of emotions among humans in general and actors specifically might be large, a question that is intensively discussed in psychology [4]. Lastly, our main goal is to fuse the different approaches to a multimodal classification approach that encompasses all modalities as has been successfully applied in other research areas [30].

Considering the annotation, annotators reported that being offered multiple media channels facilitated the annotation and that the image and audio channel helped a lot when the language and the context of the play was unclear. However, this did not show any positive effects for agreement among the annotators. The agreement level remains similarly mediocre as with annotations of literary or historical texts with solely the textual representation [1, 27, 50, 36, 38]. Our assumption that the presentation of the video makes the annotation more clear was proven wrong for this specific use case. The subjectivity in the interpretation of literary material is not affected by the presentation of the video channel. We want to pursue improvements for the annotation process by developing specific video annotation tools, enabling the annotation while watching the movie [42, 43]

While the presented case study can be regarded as negative result we did learn that the application of general-purpose sentiment analysis is not sufficient for our material. Thus, we are

currently conducting larger annotation studies to gather training material for optimized machine learning approaches but also to explore the influence of multimodality on the annotation procedure.

References

[1] C. O. Alm and R. Sproat. "Emotional Sequencing and Development in Fairy Tales". In: *Affective Computing and Intelligent Interaction*. Ed. by J. Tao, T. Tan, and R. W. Picard. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2005, pp. 668–674. DOI: 10.1007/11573548_86.

[2] C. O. Alm and R. Sproat. "Emotional sequencing and development in fairy tales". In: *International Conference on Affective Computing and Intelligent Interaction*. Springer. 2005, pp. 668–674.

[3] H. Aviezer, Y. Trope, and A. Todorov. "Body cues, not facial expressions, discriminate between intense positive and negative emotions". In: *Science* 338.6111 (2012), pp. 1225–1229.

[4] L. F. Barrett. *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt, 2017.

[5] E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang. "Training deep networks for facial expression recognition with crowd-sourced label distribution". In: *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. 2016, pp. 279–283.

[6] Q. Dang, J. Yin, B. Wang, and W. Zheng. "Deep learning based 2D human pose estimation: A survey". In: *Tsinghua Science and Technology* 24.6 (2019), pp. 663–676. DOI: 10.26599/tst.2018.9010100.

[7] K. Dautenhahn. "The origins of narrative: In search of the transactional format of narratives in humans and other social animals". In: *International Journal of Cognition and Technology* 1.1 (2002), pp. 97–123. DOI: https://doi.org/10.1075/ijct.1.1.07dau. URL: https://www.jbe-platform.com/content/journals/10.1075/ijct.1.1.07dau.

[8] J. Fehle, T. Schmidt, and C. Wolff. "Lexicon-based Sentiment Analysis in German: Systematic Evaluation of Resources and Preprocessing Techniques". In: *Proceedings of the 17th Conference on Natural Language Processing (KONVENS 2021)*. Düsseldorf, Germany, 2021.

[9] J. M. Garcia-Garcia, V. M. Penichet, and M. D. Lozano. "Emotion detection: a technology review". In: *Proceedings of the XVIII international conference on human computer interaction*. 2017, pp. 1–8.

[10] D. Halbhuber, J. Fehle, A. Kalus, K. Seitz, M. Kocur, T. Schmidt, and C. Wolff. "The Mood Game - How to Use the Player's Affective State in a Shoot'em up Avoiding Frustration and Boredom". In: *Proceedings of Mensch Und Computer 2019*. MuC'19. Hamburg, Germany: Association for Computing Machinery, 2019, pp. 867–870. DOI: 10.1145/3340764.3345369.

[11] P. Hartl, T. Fischer, A. Hilzenthaler, M. Kocur, and T. Schmidt. "AudienceAR - Utilising Augmented Reality and Emotion Tracking to Address Fear of Speech". In: *Proceedings of Mensch Und Computer 2019*. MuC'19. Hamburg, Germany: Association for Computing Machinery, 2019, pp. 913–916. DOI: 10.1145/3340764.3345380.

[12]   F. Jannidis, I. Reger, A. Zehe, M. Becker, L. Hettinger, and A. Hotho. "Analyzing features for the detection of happy endings in german novels". In: *arXiv preprint arXiv:1611.09028* (2016).

[13]   K. Kajava, E. Öhman, P. Hui, and J. Tiedemann. "Emotion Preservation in Translation: Evaluating Datasets for Annotation Projection". In: *Proceedings of Digital Humanities in Nordic Countries (DHN 2020)*. Ceur, 2020, pp. 38–50.

[14]   T. Kakkonen and G. 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*. Hissar, Bulgaria: Association for Computational Linguistics, 2011, pp. 62–69.

[15]   E. Kim and R. Klinger. "A Survey on Sentiment and Emotion Analysis for Computational Literary Studies". In: *Zeitschrift für digitale Geisteswissenschaften* (2019). DOI: 10.17175/2019_008. URL: http://arxiv.org/abs/1808.03137.

[16]   E. Kim and R. Klinger. "An Analysis of Emotion Communication Channels in Fan-Fiction: Towards Emotional Storytelling". In: *Proceedings of the Second Workshop on Storytelling*. Florence, Italy: Association for Computational Linguistics, 2019, pp. 56–64.

[17]   E. Kim, S. Padó, and R. Klinger. "Prototypical Emotion Developments in Literary Genres". In: *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. 2017, pp. 17–26.

[18]   J. R. Landis and G. G. Koch. "The Measurement of Observer Agreement for Categorical Data". In: *Biometrics* 33.1 (1977), pp. 159–174.

[19]   M. Magdin and F. Prikler. "Real time facial expression recognition using webcam and SDK affectiva". In: *Ijimai* 5.1 (2018), pp. 7–15.

[20]   M. V. Mäntylä, D. Graziotin, and M. Kuutila. "The evolution of sentiment analysis–A review of research topics, venues, and top cited papers". In: *Computer Science Review* 27 (2018), pp. 16–32. DOI: 10.1016/j.cosrev.2017.10.002.

[21]   D. McDuff, A. Mahmoud, M. Mavadati, M. Amr, J. Turcot, and R. e. Kaliouby. "AFFDEX SDK: a cross-platform real-time multi-face expression recognition toolkit". In: *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*. 2016, pp. 3723–3726.

[22]   K. Mellmann. "Literaturwissenschaftliche Emotionsforschung". In: *Handbuch Literarische Rhetorik*. De Gruyter, 2015, pp. 173–192.

[23]   B. Meyer-Sickendiek. *Affektpoetik: eine Kulturgeschichte literarischer Emotionen*. Königshausen & Neumann, 2005.

[24]   S. 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*. Portland, OR, USA: Association for Computational Linguistics, 2011, pp. 105–114.

[25]   L. Moßburger, F. Wende, K. Brinkmann, and T. 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*. Barcelona, Spain (Online): Association for Computational Linguistics, 2020, pp. 70–81. URL: https://www.aclweb.org/anthology/2020.smm4h-1.11.

[26] E. T. Nalisnick and H. S. Baird. "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, 2013, pp. 479–483. URL: https://www.aclweb.org/anthology/P13-2085.

[27] E. Öhman. "Challenges in Annotation: Annotator Experiences from a Crowdsourced Emotion Annotation Task". In: *Proceedings of the Digital Humanities in the Nordic Countries 5th Conference*. CEUR Workshop Proceedings, 2020, pp. 293–301.

[28] A.-M. Ortloff, L. Güntner, M. Windl, T. Schmidt, M. Kocur, and C. Wolff. "SentiBooks: Enhancing Audiobooks via Affective Computing and Smart Light Bulbs". In: *Proceedings of Mensch Und Computer 2019*. MuC'19. Hamburg, Germany: Association for Computing Machinery, 2019, pp. 863–866. DOI: 10.1145/3340764.3345368.

[29] F. Pianzola, S. Rebora, and G. 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". In: *Plos One* 15.1 (2020), e0226708. DOI: 10.1371/journal.pone.0226708.

[30] S. Poria, E. Cambria, R. Bajpai, and A. Hussain. "A review of affective computing: From unimodal analysis to multimodal fusion". In: *Information Fusion* 37 (2017), pp. 98–125.

[31] A. J. Reagan, L. Mitchell, D. Kiley, C. M. Danforth, and P. S. Dodds. "The emotional arcs of stories are dominated by six basic shapes". In: *EPJ Data Science* 5.1 (2016), p. 31. DOI: 10.1140/epjds/s13688-016-0093-1.

[32] R. Remus, U. Quasthoff, and G. Heyer. "SentiWS-A Publicly Available German-language Resource for Sentiment Analysis." In: *Lrec*. 2010.

[33] H. Schmid. "Probabilistic part-ofispeech tagging using decision trees". In: *New methods in language processing*. 2013, p. 154.

[34] T. Schmidt. "Distant Reading Sentiments and Emotions in Historic German Plays". In: *Abstract Booklet, DH_Budapest_2019*. Budapest, Hungary, 2019, pp. 57–60. DOI: 10.5283/epub.43592.

[35] T. Schmidt and M. 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*. Santa Fe, New Mexico: Association for Computational Linguistics, 2018, pp. 139–149. URL: https://www.aclweb.org/anthology/W18-4516.

[36] T. Schmidt, M. Burghardt, and K. Dennerlein. "Sentiment Annotation of Historic German Plays: An Empirical Study on Annotation Behavior". In: *Proceedings of the Workshop on Annotation in Digital Humanities 2018 (annDH 2018), Sofia, Bulgaria, August 6-10, 2018*. Ed. by S. Kübler and H. Zinsmeister. 2018, pp. 47–52. URL: https://epub.uni-regensburg.de/43701/.

[37] T. Schmidt, M. Burghardt, K. Dennerlein, and C. 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/.

[38] T. Schmidt, M. Burghardt, K. Dennerlein, and C. Wolff. "Sentiment Annotation for Lessing's Plays: Towards a Language Resource for Sentiment Analysis on German Literary Texts". In: *2nd Conference on Language, Data and Knowledge (LDK 2019)*. Ed. by T. Declerck and J. P. McCrae. 2019, pp. 45–50. URL: https://epub.uni-regensburg.de/43569/.

[39] T. Schmidt, M. Burghardt, and C. Wolff. "Toward Multimodal Sentiment Analysis of Historic Plays: A Case Study with Text and Audio for Lessing's Emilia Galotti". In: *Proceedings of the Digital Humanities in the Nordic Countries 4th Conference*. Ed. by C. Navarretta, M. Agirrezabal, and B. Maegaard. Vol. 2364. CEUR Workshop Proceedings. Copenhagen, Denmark: CEUR-WS.org, 2019, pp. 405–414. URL: http://ceur-ws.org/Vol-2364/37%5C%5Fpaper.pdf.

[40] T. Schmidt, J. Dangel, and C. Wolff. "SentText: A Tool for Lexicon-based Sentiment Analysis in Digital Humanities". In: *Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021)*. Ed. by T. Schmidt and C. Wolff. Vol. 74. Glückstadt: Werner Hülsbusch, 2021, pp. 156–172. DOI: 10.5283/epub.44943. URL: https://epub.uni-regensburg.de/44943/.

[41] T. Schmidt, K. Dennerlein, and C. Wolff. "Towards a Corpus of Historical German Plays with Emotion Annotations". In: *3rd Conference on Language, Data and Knowledge (LDK 2021)*. Ed. by D. Gromann, G. Sérasset, T. Declerck, J. P. McCrae, J. Gracia, J. Bosque-Gil, F. Bobillo, and B. Heinisch. Vol. 93. Open Access Series in Informatics (OASIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021, 9:1–9:11. DOI: 10.4230/OASIcs.LDK.2021.9.

[42] T. Schmidt, I. Engl, D. Halbhuber, and C. Wolff. "Comparing Live Sentiment Annotation of Movies via Arduino and a Slider with Textual Annotation of Subtitles." In: *DHN Post-Proceedings*. 2020, pp. 212–223.

[43] T. Schmidt and D. Halbhuber. "Live Sentiment Annotation of Movies via Arduino and a Slider". In: *Digital Humanities in the Nordic Countries 5th Conference 2020 (DHN 2020). Late Breaking Poster*. 2020.

[44] T. Schmidt, P. Hartl, D. Ramsauer, T. Fischer, A. Hilzenthaler, and C. Wolff. "Acquisition and Analysis of a Meme Corpus to Investigate Web Culture." In: *Digital Humanities Conference 2020 (DH 2020)*. Ottawa, Canada, 2020. DOI: 10.17613/mw0s-0805.

[45] T. Schmidt, M. Jakob, and C. Wolff. "Annotator-Centered Design: Towards a Tool for Sentiment and Emotion Annotation". In: *INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft (Workshop-Beiträge)*. Ed. by C. Draude, M. Lange, and B. Sick. Bonn: Gesellschaft für Informatik e.V., 2019, pp. 77–85. DOI: 10.18420/inf2019_ws08.

[46] T. Schmidt, F. Kaindl, and C. Wolff. "Distant Reading of Religious Online Communities: A Case Study for Three Religious Forums on Reddit." In: *Dhn*. Riga, Latvia, 2020, pp. 157–172.

[47] T. Schmidt, M. Schlindwein, K. Lichtner, and C. Wolff. "Investigating the Relationship Between Emotion Recognition Software and Usability Metrics". In: *i-com* 19.2 (2020), pp. 139–151. DOI: 10.1515/icom-2020-0009.

[48] T. Schmidt, B. Winterl, M. Maul, A. Schark, A. Vlad, and C. Wolff. "Inter-Rater Agreement and Usability: A Comparative Evaluation of Annotation Tools for Sentiment Annotation". In: *INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft (Workshop-Beiträge)*. Ed. by C. Draude, M. Lange, and B. Sick. Bonn: Gesellschaft für Informatik e.V., 2019, pp. 121–133. DOI: 10.18420/inf2019_ws12.

[49] A. Schonlau. *Emotionen im Dramentext: eine methodische Grundlegung mit exemplarischer Analyse zu Neid und Intrige 1750-1800*. Deutsche Literatur Band 25. Berlin Boston: De Gruyter, 2017.

[50] R. Sprugnoli, S. Tonelli, A. Marchetti, and G. Moretti. "Towards sentiment analysis for historical texts". In: *Digital Scholarship in the Humanities* 31 (2015), pp. 762–772. DOI: 10.1093/llc/fqv027.

[51] D. Taylor. *The archive and the repertoire*. Duke University Press, 2003.

[52] G. Vinodhini and R. Chandrasekaran. "Sentiment analysis and opinion mining: a survey". In: *International Journal* 2.6 (2012), pp. 282–292.

[53] S. Winko. *Über Regeln emotionaler Bedeutung in und von literarischen Texten*. De Gruyter, 2011.