Computational emotion classification for genre corpora of German tragedies and comedies from 17th to early 19th century

Katrin Dennerlein 1,*, Thomas Schmidt 2, Christian Wolff 2

1Institute for German Philology, University of Würzburg, Würzburg, Germany
2Media Informatics Group, University of Regensburg, Regensburg, Germany
*Correspondence: Katrin Dennerlein, E-mail: katrin.dennerlein@uni-wuerzburg.de

Abstract

This article presents a method of emotion analysis for German drama from the 17th to the 19th century that significantly goes beyond previous research approaches in computational literary studies. It is based on annotations of 17 dramatic texts resulting in 11,939 annotations which were used as training material to fine-tune a German language BERT model that achieves an average accuracy of 73% for the single-label emotion classification of fourteen emotion types in cross-validation. We apply the emotion classification on a corpus of 141 comedies and 92 tragedies to compare these genres. For tragedies, the mean proportion percentages of ‘suffering’ and ‘abhorrence’ are higher than for comedies. Inversely, mean percentages of ‘anger’ and ‘joy’ are higher for comedies than for tragedies. A new finding is the surprisingly high proportion of ‘anger’ in comedies. Emotion distribution of the last scenes in dramatic texts also proves the quality of the classified data in terms of literary studies. In addition, the emotion distribution for several subgenres of comedy is investigated including non-canonical works of wide circulation which reached the recipients directly through the depicted emotions in the Kasperl Plays. Comedies from 1740 to 1770 are characterized by a pairing of higher amounts of ‘friendship’ and ‘love’. Satirical comedies from the same period stand out due to high rates of ‘anger’ as well as ‘suffering’. The very successful Kasperl plays turn out to be characterized by a comparatively large percentage of ‘schadenfreude’ and ‘joy’.

1 Introduction

Concepts of emotions see a quick rise to become fundamental categories of thought during the 17th century in Germany, backed by the theories of Descartes, Spinoza, and Thomasius (see Grimm, 2010, p. 29 ff.). The arts, especially music and theatre, are considered as the best media to demonstrate these concepts in action. The rise of the German-language drama to the most popular literary genre in the period from 1650 to the early 19th century (Krause, 1980; Schulte-Sasse, 1980; Zeman, 1985; Brenner, 1999; Meid, 2009, p. 327–501) is closely linked to its development to a ‘school of affects’ (Rotermund, 1972, p. 25). This means that drama should help to feel desired emotions and to deal appropriately with unwanted ones such as fear, envy, or suffering (Schings, 1971; Wiegmann, 1987; Schulz, 1988; Zeller, 2005; Schonlau, 2017). Therefore, in this period, emotions are also crucial for the differentiation of existing genres and the emergence of new ones. In the genre of tragedy, for example, forms emerge in which the final catastrophe is shown as an effect of emotions (Grimm, 2010, p. 26). In the genre of comedy, the plot begins to be structured by intrigues that are often shaped and motivated by emotions (Kraft, 2011, pp. 64–66); in the emerging genre of German opera, love becomes the central emotion. Subgenres of drama also develop along the emotions of admiration, pity, and moving (Schings, 1980; Meier, 1993). In traditional literary studies, this knowledge about the relationship between genre and emotion is gathered by scholars by examining in detail a few dramatic texts as examples for one specific emotion. Consequently, little is known about which emotions are used in dramatic texts at a larger scale of representation.

In this article, we use state-of-the-art language models to investigate emotion representation in dramatic texts from 1650 to 1815 and will thereby considerably exceed previous research approaches. We aim to cover several hundreds of texts and up to fourteen classes (thirteen emotions and a no emotion class). We first define a meta-linguistic concept of emotion that is appropriate to our object of study, the intended character emotions.

© The Author(s) 2023. Published by Oxford University Press on behalf of EADH. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Then we define and outline our set of thirteen emotions. We also report the annotation results of seventeen German dramatic texts as well as the selection and evaluation of computational emotion classification methods. The first goal of the article is to compare the results of the computational emotion classification on 233 dramatic texts with existing knowledge from historical literary studies. For this sake, we investigate whether there are anticipated differences in terms of character’s emotions between tragedies and comedies, and what new insights can be gathered. We distinguish between the emotion profiles of the texts as a whole and those of the final scenes. The second goal of the article is to gain new insights into emotion distributions of several subgenres of comedy in the 18th century. For this purpose, we also chose non-canonical works of wide circulation to compare them to better known comedy subgenres of the Enlightenment period. Here we take a step in the direction to the expansion of the very narrow canon of dramatic texts treated in literary studies of our period under investigation (Meyer, 1986 ff.; Krämer, 1998; Müller-Kampel, 2003; Dennerlein 2021).

2 Related work

Sentiment analysis, also often called opinion mining, deals with computational methods to predict and analyse sentiment in written text (Liu, 2015). Sentiment analysis is often regarded as a classification task for valence or polarity categories, meaning if a text is rather positive, negative, neutral, or mixed (Mäntylä, Graziotin, Kuutila, 2018). The neighbouring area of emotion classification, which is the focus of this article, deals with the analysis and prediction of multiple emotion categories like anger, joy, or sadness. Among the most common application areas for these methods are social media and product reviews (cf. Mäntylä, Graziotin, Kuutila, 2018).

However, both methods have also received a lot of attention in digital humanities and computational literary studies for various narrative text genres in recent years. As early as 2005, Alm and Sproat explored the annotation of emotion and the application of sentiment and emotion analysis in fairy tales. Scholars also investigated sentiment and emotion progressions and distributions (Mohammad, 2011; Kim, Padó, Klinger, 2017) as well as creative visualization methods (Kakkonen and Galic Kakkonen, 2011) in novels and fairy tales. Zehe et al. (2016) evaluated the possibility to use emotion analysis to predict happy ends in novels. Reagan, Mitchel, Kiley, et al. (2016) used sentiment analysis on more than 1,000 fictional texts to identify six reoccurring basic shapes for emotional arcs. Recently, researchers also used sentiment and emotion analysis to inspect the various facets of the genre of online ‘fan fiction’ (Kim and Klinger, 2019b), not just focusing on the prose text but also on comments (Pianzola, Rebora, Lauer, et al., 2020).

In the context of historical dramatic texts, Mohammad (2011), Nalisnick and Baird (2013), and Yavuz (2021) explored the dramatic texts of Shakespeare to investigate character relationships and their developments. Schmidt and Burghardt (2018) evaluated sentiment analysis approaches on dramatic texts of the German playwright Lessing. For a more in-depth overview of this research field, see Kim and Klinger (2019a). Examining the state of this field shows that most of the research is focused on the analysis of valence or polarity and mostly on individual authors (Mohammad, 2011; Nalisnick and Baird, 2013; Schmidt and Burghardt, 2018). Kim and Klinger (2019a) further point out (1) that the majority of the applied methods are lexicon-based (working with predefined lists of annotated words), which are regarded as outdated in the natural language processing (NLP) community and (2) that there is a lack of annotated corpora to perform more advanced machine learning (ML)-based approaches and evaluate the applied methods. In the following article, we address several of these problems and gaps in German historical dramatic texts by annotating a large corpus to train transformer-based models for emotion classification and applying this to a larger number of texts.

3 Emotion definition

There are quite a few different terms for emotions in the 17th and 18th centuries. The German terminus ‘Affekt’ (or ‘affectus’, ‘affectio’, ‘Gemüthsbewegung’) stems from philosophical as well as medical theories, meaning strong and quick alterations of emotional states which are often linked to the idea of their effects (see Grimm 2010, p. 29). It has to be distinguished from the term ‘passio’, or ‘Leidenschaft’ meaning longer lasting emotional states related to characteristics of a person as the definition of Kant fixes it in 1772. The term ‘Gefühl’ first appears at the end of the 17th century and is discussed during the 18th century (see Ewert 1974, p. 82). It is located in philosophical, psychological, and cultural contexts and it is not until 1800 that it is established as the generic term for emotional states in general (see Kellner, 2007, p. 23). While we take into account that we will find various different terms for emotional states in the period under study we will use the term ‘emotion’ as a metalinguistic term to cover the following range of phenomena that we find in dramatic texts:

- Emotions are intended emotions experienced by and attributed to characters in dramatic texts. We assume that characters in dramatic texts are created in a way that an inner experience is attributed to them.
- Emotions are changing states, which are experienced as mental and physical condition at the same
time and which are evaluated positively or negatively. Both cause and directionality play a role in the processing of emotions. Characters may be aware of emotions to varying degrees.

- Emotions are understood as ego-related throughout the period under study, but only very late as completely individual states. Before that, emotions are understood as determined by class and gender, but also by religious beliefs, political views, and other cultural aspects (Grimm 2010). In addition, dramatic texts are strongly shaped by rhetorical conventions, which only become weaker in the last 3rd of the 18th century (Arnold, 2012; Schonlau, 2017, pp. 100–127).

In our research we do not consider the emotions of authors or recipients. While recipient emotions are a crucial element in the poetics of drama and determine the composition of dramatic texts, a computational literary studies approach to these emotions is only reasonable after understanding the emotions on a representational level.

4 Set of emotions

To discover and explore literary change with NLP methods for literary studies purposes, we defined a set of emotions:

- 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/An 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 (−)
- being moved/emotionale Bewegtheit (undetermined).

In the period from 1650 to 1815, there are several historical category systems for emotions and related phenomena, which often show a mixture of virtues and emotions (Grimm, 2010). Especially in philosophy, the true number and nature of emotions has been reflected upon a lot (Zeller, 2005, p. 692). To grasp the change in the weighting of emotions in our time period, it is necessary to abstract from these concepts in such a way that they can be applied to the entire period as well as to represent reweightings and changes. The main criterion for the selection of emotion categories for us was to enable the representation of changes in literary history and differences in genre. This meant that we would have to include emotions that hardly play a role in certain periods of time, but then suddenly become central categories (like, for example, ‘compassion’, cf. Schings, 1980). Therefore, we have chosen to use the superordinate categories of ‘affection’, ‘pleasure’, ‘anxiety’, ‘rejection’, and ‘suffering/empathy’, and to further subtype them. The following examples illustrate the necessity of differentiation: The category of ‘suffering/empathy’, which is undoubtedly central to the entire European drama, must be complemented by the newly emerging category of ‘compassion’, which, with the author Gotthold Ephraim Lessing, becomes the central category from the middle of the 18th century (Schings, 1980). Similarly, ‘love’ is to be supplemented by two categories. On the one hand, that of ‘desire’, which plays a central role in the devaluation of characters, for example, in Baroque drama or later in the drama of the Storm and Stress period. On the other hand, a hyperonymous category is needed here, titled ‘affection’, to include friendship. The latter is central to Enlightenment anthropology and, together with compassion and the absence of desire, is probably characteristic of the epochal shift from Baroque to Enlightenment drama (Sauer, 1980; Lukas, 2005). Added to this was the category ‘rejection’ that includes ‘anger’ and ‘abhorrence’ because they are crucial for a basic understanding of the plot. Finally, the category ‘being moved’ is necessary in order to be able to distinguish strong and unspecified emotional agitation. It occurs fairly often that the characters do not know what they (should) feel and/or that they vacillate between several feelings.

With this scheme—especially with the division into main and sub-emotion classes—we deviate from emotion category systems commonly used in NLP. Most often they rely on psychological emotion category systems (Plutchik, 1980; Wood et al., 2018a,b). In such models, ‘fear’, ‘anger’, ‘joy’, and ‘sadness’, which we also use, are central categories, but also some others that we have not adopted as main emotion labels, such as ‘confidence’, ‘anticipation’, ‘surprise’, and ‘disgust’. These emotions can also be annotated in our schema, but they are integrated into other concepts. ‘Anticipation’ is integrated with ‘joy’, ‘surprise’ with
‘being moved’, and ‘disgust’ with ‘abhorrence’ in the sense of strong rejection.

5 Corpus

We use a corpus of 296 dramatic texts in our research project so far but will focus on the subset of 233 tragedies and comedies in the following. As far as possible, the dramatic texts are taken from the collection German Drama Corpus (GerDraCor, 222 dramatic texts, Fischer, Börner, Göbel, et al., 2019) because these dramatic texts are prepared for computational analysis, and corrections to the source material were made in the markup and in the full text (Hug, et al., 2021). However, we have also included fifty dramatic texts that are available in the TextGrid Repository (TGRep) and prepared them semi-manually for further usage. The corpus is largely canonical (cf. Alt, 1994; Schulz, 2007). The dramatic texts canonized today for the period under study are not necessarily those that were most frequently performed, printed, or read at the time (cf. Meid, 2009, pp. 327–501; Dennerlein, 2021). Nevertheless, they form an important branch of tradition, which was discussed and worked on. However, in order to map not only this branch of tradition, we also include dramatic texts that are no longer canonized today. Therefore, the corpus contains twenty-four Kasperl plays from the Leopoldstädter Theater in Vienna. These are operettas that had enormous commercial success and were widely distributed throughout the German-speaking area.

To address genres, we manually assigned genre labels to the dramatic texts. The assignment of genre labels was done through careful scholarly research. This was particularly important in the case of non-specific genre designations in the subtitles of the dramatic texts. Roughly speaking, dramatic texts with bad ends were labelled as tragedies, those with good ends and humour were classified as comedies, and those with good ends without humour were classified as ‘Schauspiele’ (‘dramas’). The last category now contains all those dramatic texts which bear genre labels like ‘Schauspiel’, ‘Rührstücke’, ‘Schäferspiel’, ‘Historisches Schauspiel/Drama’, or others. Following this classification, our corpus now comprises 92 tragedies, 141 comedies, and 63 ‘dramas’. See Table 1 for general statistics about the overall corpus. In upcoming chapters we will also analyze comedy sub-genres (Table 2).

6 Annotation

In the following section, we describe the annotation process and the results of the annotation of seventeen dramatic texts.

6.1 Annotation process

The annotation was performed with the CATMA tool (Gius, Meister, Petris, et al., 2020). The annotators were students of German Literary Studies that were compensated monetarily for their work. They had access to detailed annotation guideline documents and were trained for the annotation process via various pilot annotations under the guidance of a literary scholar expert. Annotators annotated representative dramatic texts of our current corpus in their entirety. The reason for this is that annotators are assigned to perform context-sensitive annotations taking into account the entire plot and content of the dramatic text as well as specific attributes of the characters.

The annotation task consisted in indicating every expression of one of the thirteen sub-emotions of the emotion scheme as defined in the annotation guidelines. Annotators could annotate stage directions and speeches (spoken text of characters, separated by the utterances of the other characters) and they could annotate variable text lengths meaning as many or few words as they see fit. They could also annotate the same passage with different classes as well as assign text passages in an overlapping way. While such variable annotation settings are rare in NLP (Wood et al., 2018a, b), we have chosen this approach in an iterative process of pilot annotations since it proved to best fit our annotation goal and the annotation practices of literary scholars.

Annotators had around 2 weeks to finish the annotation for one dramatic text. Each annotator annotated independently from the other annotator without access to the annotation of other annotators. In total, the annotation process for one dramatic text took about 8–30 h depending on its complexity and the training of the annotator.

6.2 Annotated dramatic texts

As the material for the training and evaluation of ML-based emotion classification approaches, we annotated a representative sample of seventeen dramatic texts of our corpus. To our knowledge, it is the first corpus of emotion-annotated historical dramatic texts in German and of adequate size compared to other sentiment and emotion corpora in German (cf. Fehle, Schmidt, Wolff, 2021). The chosen texts are representative of our overall corpus of 296 dramatic texts. Representativeness is not to be understood in purely statistical and quantitative terms but rather denotes a criterion that is gauged in terms of literary studies. The selection of texts includes one example of each of the major subgenres and literary movements of the period we investigate. Care was taken to ensure that a wide range of emotion expressions was present in the dramatic texts in order to best train a language model on emotion.
classification. The following list shows the dramatic texts we annotated and some metadata:

- Catharina von Georgien by Andreas Gryphius (1657, tragedy).
- Der Welt Erschreckende Attila, anonymous (after 1682, drama).
- Massaniello by Christian Weise (1683, tragedy).
- Ein wunderliches Schau-Spiel vom niederländischen Bauer by Christian Weise (1669, comedy).
- Die getreue Sclavin Doris, anonymous (1720, drama).
- Das Testament by Luise Adelgunde Victorie Gottsched (1745, comedy).
- Canut by Johann Elias Schlegel (1746, tragedy).
- Die zärtlichen Schwester by Christian Fürchtegott Gellert (1747, comedy).
- Lucie Woodvill by Johann Gottlieb Benjamin Pfeil (1757, tragedy).
- Der Freigeist by Joachim Wilhelm von Brawe (1758, tragedy).
- Minna von Barnhelm by Gotthold Ephraim Lessing (1767, comedy).
- Der Postzug by Cornelius von Ayrenhoff (1769, comedy).
- Kabale und Liebe by Friedrich Schiller (1784, tragedy).
- Kasperl' der Mandolettikrämer by Ferninand Eberl (1789, tragedy).
- Menschenhass und Reue by August von Kotzebue (1790, comedy).
- Wallenstein's Lager by Friedrich von Schiller (1800, tragedy).
- Faust by Johann Wolfgang von Goethe (1807, tragedy).

Most of the texts stem from the GerDraCor corpus (Fischer, Börner, Göbel, et al., 2019), Catharina von Georgien from the TextGrid repository. Kasperl’ der Mandolettikrämer was acquired from an open web repository,8 Die getreue Sclavin Doris, Der Welt Erschreckende Attila, and Ein wunderliches Schau-Spiel vom niederländischen Bauer from separate editions (Weise, 1986; Noe, 2007). These texts had to be further prepared for the annotation process.

Each dramatic text was annotated by two annotators (independently from each other). Overall, five annotators performed the annotation with switching pairings throughout the annotation process in counterbalanced order.

6.3 Annotation results

The annotations of CATMA were exported and transformed to the JSON format and CSV files via a Python script we developed to calculate various results of the annotation. Because of low to mediocre agreement metrics (see Table 6), we filtered every annotation on which annotators disagreed upon. We have collected 20,297 emotion annotations resulting in 11,939 annotations after removing full and partial disagreements between the annotator pairings. This includes all annotations independent of the length which can on maximum have the length of an entire speech. Extending this with the 11,349 non-annotated text units which we will refer to as no emotion class results in an overall corpus of 23,288 text units (51% emotion annotations, 49% no emotion). Of the 11,939 emotion annotations 50% (6,000) are negative, 40% (4,795) positive, and 10% (1,144) of the class being moved. Table 3 illustrates the annotation distribution for the main classes and sub-emotions only among the emotion annotations filtering the non-annotated material.

The most frequently annotated main classes are the emotions of rejection (24%). Concerning the sub-emotions, we can identify strong class imbalances. The most dominant sub-emotions are ‘suffering’ (15%), ‘anger’ (24%), ‘joy’ (14%), and ‘love’ (13%).9 Sub-emotions that are rarely annotated are ‘desire’, ‘friendship’, and ‘schadenfreude’.

We have analysed token distributions of the annotations via the NLTK Punkt Tokenizer.10 We identified that annotators mark on average twenty-five tokens for a text unit. However, the annotators make significant use of the varied annotations spanning ranging from one-word annotations to multiple sentences (max = 578 tokens; Std = 29.41). Overall, the most common annotation span is an annotation across 2–3 sentences.

To evaluate the validity of the annotation scheme and process, we also look at inter-rater agreement among the two annotators for each dramatic text. Traditional agreement metrics are dependent on fixed annotation sizes. We calculate agreement on speech level since the speech is the major textual unit of a dramatic text. To deal with varied annotation lengths and overlapping annotation, we assign each speech per annotator the emotion category that is annotated the most (measured in number of tokens). While this is a simplification of our annotation process, it enables us...
to calculate traditional agreement metrics like the average Cohen’s $\kappa$ value. Table 4 illustrates the results for each emotion category system.

The agreement ranges from values of 0.5 (polarity) to 0.4 (sub-emotions) which is a moderate level of agreement according to Landis and Koch (1977). The agreement fluctuates from 0.35 to 0.62 depending on the dramatic text and the categorical system. We did not find significant differences in agreement with regard to annotator pairings. Overall, the agreements are rather low compared to other annotation results in the context of sentiment analysis (cf. Mäntylä, Graziotin, Kuutila, et al., 2018) but in line with similar research on historical and narrative text genres (Alm and Sproat, 2005; Sprungnoli et al., 2015; Schmidt et al., 2018; Schmidt et al., 2019a,b). We argue that the inherent subjectivity and the various possibilities of interpretation are reasons for these results. As already outlined, we filtered the disagreeing annotations in the annotation results reported which proved to be beneficial for the subsequent ML approaches.

Please note that more information about the annotation process and results can be found in previous publications (Schmidt et al., 2021a; Dennerlein Schmidt, Wolff, et al., 2022c). More in-depth results considering annotation behaviour (e.g. pair/play-wise agreements) can be found online.11

### 7 Computational emotion classification

We performed multiple evaluation studies for different computational single-label emotion classification methods on varied text lengths with different iterations and settings of the annotations. We differentiate between three classification tasks depending on the hierarchical level: polarity (4 classes: differing between positive, negative, and no emotion), the main emotion classes (7) and the sub-emotions including no emotion (14). We evaluated the methods including several different instances of the annotated material. We found that the best performance can be achieved by filtering the annotations of both annotators per text to just those annotations they agree on fully or partially. That means we removed all full or overlapping annotations of both annotators if they did not belong to the same class depending on the hierarchical system (see Section 6.3). For an in-depth description of these experiments, we refer to previous work (Schmidt et al., 2021b,c, 2022). In the following passage, we summarize the major results of these works and focus on the description of the final selection of the best emotion classification method according to our studies. Please note that all results are achieved with the filtered corpus as presented in Table 3 and Section 6.3. Using an unfiltered corpus that includes disagreeing annotations yields accuracy drops of up to 18%.

Considering the methods, we evaluated baseline methods like lexicon-based approaches (Schmidt and Burghardt, 2018) and traditional machine learning approaches based on bag of words models and support vector machines (SVM). The main focus of our work was on static word embeddings (e.g. fastText by Bojanowski, Grave, Joulin 2017) and transformer-based language models like BERT (Devlin, Chang, Lee, et al., 2019) and ELECTRA (Clark et al., 2020). The latter are currently regarded as state-of-the-art in the

### Table 2. Overview, token, and sentence statistics for subgenres

<table>
<thead>
<tr>
<th>Genre</th>
<th># dramatic texts</th>
<th># sentences</th>
<th># tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kasperl plays</td>
<td>24</td>
<td>40,649</td>
<td>396,996</td>
</tr>
<tr>
<td>Comedies (1740–70)</td>
<td>20</td>
<td>32,369</td>
<td>373,594</td>
</tr>
<tr>
<td>Satirical comedies</td>
<td>10</td>
<td>21,551</td>
<td>230,836</td>
</tr>
<tr>
<td>Overall</td>
<td>54</td>
<td>94,569</td>
<td>1,001,426</td>
</tr>
</tbody>
</table>

---

### Table 3. Distribution of main classes and sub-emotion categories in annotated dramatic texts among emotion annotations after filtering disagreements among annotators

<table>
<thead>
<tr>
<th>Emotion category</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main class (MC): Emotions of affection</td>
<td>2,804</td>
<td>23</td>
</tr>
<tr>
<td>Desire</td>
<td>71</td>
<td>1</td>
</tr>
<tr>
<td>Love</td>
<td>1,569</td>
<td>13</td>
</tr>
<tr>
<td>Friendship</td>
<td>240</td>
<td>2</td>
</tr>
<tr>
<td>Admiration, reverence</td>
<td>924</td>
<td>8</td>
</tr>
<tr>
<td>MC: Emotions of pleasure</td>
<td>1,991</td>
<td>17</td>
</tr>
<tr>
<td>Joy</td>
<td>1,689</td>
<td>14</td>
</tr>
<tr>
<td>Schadenfreude</td>
<td>302</td>
<td>3</td>
</tr>
<tr>
<td>MC: Emotions of anxiety</td>
<td>989</td>
<td>8</td>
</tr>
<tr>
<td>Fear</td>
<td>739</td>
<td>6</td>
</tr>
<tr>
<td>Despair</td>
<td>250</td>
<td>2</td>
</tr>
<tr>
<td>MC: Emotions of rejection</td>
<td>2,856</td>
<td>24</td>
</tr>
<tr>
<td>Anger</td>
<td>1,621</td>
<td>14</td>
</tr>
<tr>
<td>Abhorrence</td>
<td>1,235</td>
<td>10</td>
</tr>
<tr>
<td>MC: Emotions of suffering and empathy</td>
<td>2,155</td>
<td>18</td>
</tr>
<tr>
<td>Suffering</td>
<td>1,760</td>
<td>15</td>
</tr>
<tr>
<td>Compassion</td>
<td>395</td>
<td>3</td>
</tr>
<tr>
<td>Being moved</td>
<td>1,144</td>
<td>10</td>
</tr>
<tr>
<td>Overall</td>
<td>11,939</td>
<td>100</td>
</tr>
</tbody>
</table>

---

### Table 4. Average inter-rater agreement across all dramatic texts

<table>
<thead>
<tr>
<th>Categorical system</th>
<th>Average $\kappa$</th>
<th>Average %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polarity</td>
<td>0.5</td>
<td>68</td>
</tr>
<tr>
<td>Main class (MC)</td>
<td>0.4</td>
<td>62</td>
</tr>
<tr>
<td>Sub-emotion</td>
<td>0.4</td>
<td>58</td>
</tr>
</tbody>
</table>

---

a The sums for the main classes (MC; marked in bold) are listed followed by the sub-emotions. Percentages are rounded.
area of sentiment analysis and emotion classification (Shmueli and Ku, 2019; Cao, Peng, Yin, et al., 2020) and are advantageous for literary studies because the representation of tokens changes with context. We included the most established and well-known German transformer-based models from the platform Hugging Face¹¹ (Wolf et al., 2020) in our analysis. As these models are trained primarily on contemporary language, we also included available models that were trained from scratch (e.g. the German Europeana BERT¹² by Schweter, 2020) or further pre-trained on historical or narrative language (e.g. a German BERT model by Brunner, Tu, Weimer et al., 2020). We also developed models by further training base models with the texts of our corpus since this sort of domain-adaptive pre-training has been shown to be beneficial specifically for text types of special domains (Gururangan, Marasović, Swayamdipta, et al., 2020).

In our final evaluation results, we found, however, that the best-performing model for all four classification tasks is the large German BERT model gbert-large by deepset (Chan et al., 2020). The performance of the model ranges between 0.86 and 0.66 for accuracy depending on the hierarchical classification system. For the results reported in this article, we use a model for sub-emotion classification. The model was trained and evaluated on the filtered corpus in a 5 × 5 stratified fashion and fine-tuned to the specific classification task for four epochs, a batch size of thirty-two, Adam optimizer as optimization algorithm with a Tesla P100 GPU. The model achieves an average accuracy of 73% and weighted f1-score of 72% for the classification task of fourteen classes in a stratified 5 × 5 cross-validation. Based on the sub-emotion classification, we derive the corresponding classes for main emotion category and polarity based on our annotation scheme.

For more precise information about the ML approaches, please refer to previous papers (Schmidt, Dennerlein, Wolff, et al., 2021b,c, 2022). More detailed evaluation results (e.g. class based performance scores) can be found online.¹⁴

While we achieve state-of-the-art results for polarity classification in comparison to other German text genres (Chan, Schweter, Möller, 2020), we were not able to find classification results for German texts with a similar amount of sub-emotion classes. Accuracies of up to 86% have only been reached for the classification of four or more emotion classes in English non-fictional texts in contemporary language (Shmueli and Ku 2019; Yang, Lee, Whang et al., 2019; Cao, Peng, Yin et al., 2020). Thus, while there are possibilities for improvement that we will address in future work, we deem the current results sufficient for first large-scale explorations of classification results.

The concrete classification is performed on sentences of the dramatic texts of the entire corpus. For this, we segmented the stage directions and speeches into sentences using the NLTK Punkt sentence segmentation. The stage directions and speeches were either already annotated in the texts (GerDracor or TextGrid corpus) or added in an additional processing step and manually post-corrected. We chose the sentence as classification text span since it resembles the annotated text spans the most.

8 Emotion classification results for the overall corpus

Regarding the difference between a sentence which contains an emotion and a sentence which contains no emotion in all 526,455 sentences of the overall corpus of 296 dramatic texts, we identified that around 301,253 (57%) of the sentences are classified with no emotion while 225,202 (43%) sentences are classified with an emotion. In the following, we will present and focus on statistics that only include sentences classified as emotions.

The computational emotion classification yields results that are similar to the above-mentioned annotation distributions (cf. Table 5 for the classification results and Table 3 for the annotation). In the annotated corpus the most frequent single emotion is ‘suffering’ (cf. Table 3). In the classified data ‘suffering’ remains the most frequent sub-emotion (cf. Table 5). We see very small deviations in comparison to the percentage-wise distributions in the classified data when comparing them to the percentage-wise distributions of the annotations: +3.1% for ‘suffering’, +3.4% for ‘being moved’, −2.5% for ‘love’, and −1.7% for ‘schadenfreude’. The remarkable difference of 0.9% for ‘desire’ is due to the very few cases in the annotated material.

9 Analysis of classification examples

In the following, we look at some examples of classified sentences and draw conclusions about how to evaluate the prediction results by these samples, bringing distant and close reading together. In preparation of the investigations in comedy subgenres in Section 12 of the article, we chose cases of ‘joy’ and ‘schadenfreude’. For historical emotion definitions, we used three dictionaries from our period of study and compiled the basic aspects from the respective entries for ‘joy’, ‘schadenfreude’, and ‘schadenfreude’ (Zedler, 1732–1754; Adelung, 1793–1801; Grimm and Grimm, 1811).

Joy: reaction to a pleasant situation, the presence or happiness of loved ones or friends, situated in the past (memory) or in the future (hope, expectation). Joy can be—according to Zedler’s definition—morally

---

¹¹ German Europeana BERT

¹² gbert-large by deepset

¹³ endorsement of the platform Hugging Face

¹⁴ More detailed evaluation results can be found online.
reasonable or unreasonable, physically violent or moderate. According to Adelung, joy can also arise from an object or be abstracted and described as an object itself, still meaning the emotion ‘joy’.

These are examples of annotated passages that are in line with the above definition.

Sisigambis: Schaut/Schwestern/welch ein Licht steigt auß so finstern Hölen. Die grosse Sultanin/die Ibrahim verschloß/Weil sie mein Engel war/ist wieder frey und loß.15

Here we see joy as a reaction to a hoped-for situation. In the following speech, Gretchen from Goethe’s Faust is pleased about a Jewellery box:


The subsequent sentences have been correctly classified with ‘joy’ by the computational emotion classification in regard to our definition of ‘joy’:

Dorantino. [...] Es ist alles darauf eingerichtet, Bruder!17

Alle. [...] Ungemein schön18

Cases containing words like ‘pleasure’, ‘cheerful’, ‘joyful’, ‘happy’, etc. were as well classified correctly. This also applies to stage directions describing bodily expressions like ‘Laughs and drinks’ or ‘smiling’.

Cases of ‘joy’ in tragedies that are likewise correctly classified are the following:

Prothoe. Ein sel’ger Augenblick wa¨r dir beschieden, Und in den Staub vielleicht, dir huldigend, Sähst du den Sohn der Go¨ tter niederfallen.19

The classifications become significantly worse the further one goes back in time. This starts around 1730 and deteriorates further back in the 17th century.

In the subsequent example, we find that the second speech is misclassified as ‘joy’:

Pimpinone. In Opera und auf Balleten gehn? Vespeta. Díś thu ich nie.20

Cardenio. [...] Ich lehrt vnd ward gelehrt; vnd klüger vor den Jahren/Manch greisser Bart erstarrt ob meinen gelben Haren/Auch muntert ich den Leib zu allen Ku¨ nsten auff/Sprang auff ein hurtig Pferd/begab mich in den Lauff.21

In large part, this is probably due to the Baroque language, but also to the proliferation of verse plays and vocal parts in dramatic texts and parts of dramatic texts in the period from 1730 downwards. These problems can be addressed via further annotation and fine-tuning of the model.

In the following, we examine the sub-emotion ‘schadenfreude’. We compiled a working definition from historical dictionaries: Schadenfreude: a special kind of joy, namely, joy at the misfortune or harm of another person. Making fun of someone and mocking someone also belongs to the spectrum of ‘schadenfreude’. In our

emotion set presented in Section 4, we have marked ‘schadenfreude’ as a positive emotion because we use the evaluation from the point of view of the characters. For the characters in the dramatic texts, ‘schadenfreude’ is a positive emotion. They rejoice in the harm of others. While it is initially connected to resentment, jealousy, or antipathy towards the person who is the target of schadenfreude, at the moment of its occurrence it is a form of joy. It is based on felt ill-will, jealousy, or antipathy toward the person who is the target of the ‘schadenfreude’. So far we have not discovered that the characters expressing ‘schadenfreude’ in our dramatic texts are aware of the fact that they should not rejoice in the harm of others.

In the annotated corpus, the two cases below are marked as cases of ‘schadenfreude’.

Lieschen. So ist’s ihr endlich recht ergangen. Wie lange hat sie an dem Kerl gehangen!22

A young woman from the neighbourhood taunts Gretchen about her illegitimate pregnancy, which leads to her social ostracism.

Lotte. Ha! ha! ha! mein lieber Herr Bittermann, Sie haben sich ein wenig blamiert.23

The character Lotte here is laughing at Bittermann, delighting in the embarrassment of him. The harm is already inflicted.

The following examples of the classification are all taken from the Kasperl plays, because the last section of this article will give special consideration to ‘schadenfreude’ in these works. First, we look at some correctly classified sentences:

Zickzack. [...] Der Braeutigam, Herr von Knerzel und Frau von Wampel, alle deine Feinde sind dabey—Du könntest sie auf einmal fangen.24

Here we have a correctly classified sentence in which one character ascribes ‘schadenfreude’ to another in the future.

Kasperl. (aefft sie nach) Hi hi hi!25

Kasperl here makes fun of Chara, who, because of her changed appearance, he thinks is an ugly old woman.

Additionally, we present a few examples which were not classified correctly:

Dencker. Danck dir!26

Kilian. Laß er mir zugleich Margreth, und Bonavent herkommen.27

10 Emotion distribution results in tragedies versus comedies

In order to be able to verify not only statistically but also from a literary studies perspective that the classified data provides reliable results, we first ask whether the classified data show clear differences between tragedies and comedies in terms of emotions.28 One would expect that tragedies contain significantly more ‘suffering’ than comedies and that ‘joy’ plays a central role in comedies.29 Table 6 displays the percentage-wise distribution of sub-emotions among the sentences classified with an emotion (which means filtering no emotion classification).30 It shows that tragedies and comedies are well distinguished from each other with regard to the emotion profile. In tragedies, characters are significantly more likely to talk about or feel ‘suffering’ (21.3% versus 14.7% in comedies) and ‘abhorrence’ (12% versus 9.1% in comedies). In comedies ‘anger’ (16.4% versus 9.5% in tragedies) and ‘joy’ (15.5% versus 11.6% in tragedies) are distinguishing emotions compared to tragedies. The high presence of anger in comedies is a counterintuitive result, but it may be explained by the conflict structure of the comedies and the entertainment value of anger as far as ‘anger’ is attributed to ridiculous characters.

11 Happy versus bad end

We will now further specify the evaluation of the classified data and inquire which emotions are classified with gbert for the drama endings. One would assume that the endings of tragedies and comedies are very clearly separated in terms of emotions. As far as the happy end is concerned, we know two main features (Kraft, 2011). First, the conflicts are supposed to resolve for the sympathetic characters, ideally relatively unmediated. On the other hand, the good end should feature a reunion of one or more couples in love, whose marriage(s) is (are) confirmed or celebrated at the end. The emotions felt and expressed by the characters are expected to be quite different from those observed at the end of tragedies, in which often enough a character dies and death or other suffering are mourned.

In performing the calculations for the drama endings, however, it was not possible to use all the dramatic texts in our corpus, because twenty-six (thirteen tragedies and thirteen comedies) do not contain a subdivision into scenes or are not marked in such a way that they can be meaningfully evaluated. Thus, when comparing the results on the whole texts with those on the last scenes, it should be noted that the database is reduced by 10%.31 In this corpus, we find the following distinguishing sub-emotions (cf. Table 7).

- For tragedies ‘suffering’ is high with 20.2% (versus 13.7% for comedies).
For comedies ‘joy’ is high with 18.8% (versus 12.4% for tragedies). ‘Anger’ also turns out to be a distinguishing category since there is a four percentage points difference between tragedies and comedies regarding this emotion.

Moreover, the last scenes for both genres show ‘being moved’ as second highest sub-emotion (12.9% in tragedies and 15.3% in comedies). This category is defined as follows: ‘being moved’ is a strong emotional agitation, which, however, is not specified more precisely in terms of content. Astonishment, excitement, strong embarrassment, and shame belong to it. In these cases, the character usually does not know what he or she is feeling, suppresses a feeling, or vacillates between feelings. The category was explicitly not annotated as a reinforcement of other, identifiable emotions. Here are some examples from the annotated material:

Wilhelm. Lucie, Lucie, was für einen Sturm haben Sie auf meine Seele getan!32

Faust. indem er hinzustürzt, innehaltend.33

This category is often correctly classified, even in very short exclamations and sentences such as ‘Nein/No!’, ‘Ist er der Vater/Is he the father?’, ‘Aber itzt?/But now?’, and so forth.

For the last scenes of tragedies and comedies, the expected emotions would relate to the well-known plot events in dramatic texts. One would expect for tragedies a dominance of suffering in the face of death and loss or sadness over unrequited love, anger over spurning, and lack of success, for example. For comedies, one would anticipate joy due to the resolution of all conflicts or family reunions and expressions of love in the face of marriage. To explain the high presence of ‘being moved’, a close reading of some final scenes was performed, which showed that it is indeed mostly shortly before the end that some characters still experience strong turbulence because they are unsecure, torn or surprised.

### 12 Comedy subgenres

In this part of the article, we will use the classification data to explore subgenres of comedies. For this sake, we had to assign subgenres to the dramatic texts. This task is not quite easy, because on the one hand, lists of dramatic genres are relatively rare in literary studies of that period. On the other hand, digital copies of the respective subgenres are rare.34 Considerable work remains to be done here, both in the field of metadata and in the domain of full-text digitization. However, from the already existing digitized material of German-language dramatic texts from 1650 to 1815, three groups of comedies could be formed, which, although not quite published at the same time, are nevertheless comparable:

- 10 copies of ‘Sächsische Typenkomödie’, a special type of Enlightenment comedy further named as ‘satiirical comedies’ from the period (1740-1770). The selection follows Steinmetz, 1966.
- 10 Enlightenment comedies from the same time period (1740–70) which do not follow the model of satirical comedies. This group is called ‘comedies 1740-1770’.
- 24 comedies of the Leopoldstädt Theater in Vienna35

Our comparative research reveals some telling differences between the three comedy subgenres concerning the emotions of the characters (Table 8).

- The satirical comedies are characterized by high rates of ‘anger’ (19.5% as opposed to 16.1% and 16.6% for the other comedy subgenres) as well as ‘suffering’ (15.5% as opposed to 14.3% and 12.2% for the other comedy subgenres). Close readings prove that the satirical comedies show characters being annoyed by a main, morally imperfect character, who in turn is angry with most of his fellow human beings (Steinmetz, 1966). Trouble arises both from the obstruction of the marriage envisaged from the beginning and from interactions with the negative protagonist.
• The comedies from 1740 to 1770 stand out by a combination of higher levels of friendship (2.6% while the arithmetic mean is 1.8%) and love (12.2% as opposed to 10.8% and 7.6% for the other comedy subgenres).
• The Kasperl plays are characterized by an unusually high proportion of ‘schadenfreude’ (2.6% as opposed to 1.2% and 1.6% for the two other comedy subgenres) and ‘joy’ (16.3% in contrast to 14.3% and 13.3% in the other two comedy subgenres). With respect to the Kasperl plays, some new insights also come to light here. So far, these comedies have been researched in theatre studies with an emphasis on stage conventions, motives, character types, and topics (Müller-Kampel and Kuzmics, 2010). Based on our results, emotional characteristics come to the fore.

13 Discussion
The comparative results for tragedies and comedies show that even a prediction accuracy of 73% already maps well the expected differences between comedies and tragedies. Compared to comedies, tragedies show higher percentages for ‘suffering’ (21.3% versus 14.7%) and ‘abhorrence’ (12% versus 9.1%). Comedies have high values for ‘anger’ (16.4% versus 9.5%) and ‘joy’ (15.5% versus 11.6%). Based on the surprising result that the scores for ‘anger’ in comedies are so high, we can tentatively derive the following hypotheses, which should be examined in further research:

- Anger is the equivalent action motivator to ‘suffering’ in tragedy. In tragedy, the characters act because they suffer or fear to suffer in the future.
- In comedy, on the other hand, the characters act out of anger at one or more situations that turn out for the better only at the end. They are tricked in their endeavours, betrayed, or are unsuccessful for other reasons and are angry about this.

Emotions profiles of the last scenes of the respective genres confirm this distinguishing quality. For tragedies ‘suffering’ is predominant: 20.2% compared to 13.7% for comedies. For comedies ‘joy’ is predominant with 18.8% (versus 12.4% for tragedies). ‘Anger’ also turns out to be a distinguishing category since there is a four percentage points difference between tragedies and comedies regarding this emotion. A next step will be to use both the progressions and the emotion profiles of the last scenes as textual enrichments, which should allow us to train a BERT model on genre recognition only.

We also used the emotion classification to gain new insights into three subgenres of comedy of that period (including less well-researched ones). The results show that the comedy subgenres can be distinguished very well on the basis of the emotion profiles.

- Comedies from 1740 to 1770 are characterized by a pairing of significantly higher levels of ‘friendship’ (2.6% while the arithmetic mean is 1.8%) and ‘love’ (12.2% as opposed to 10.8% and 7.6% for the other comedy subgenres). This suggests that the

<table>
<thead>
<tr>
<th>Table 8. Percentage-wise distribution of sub-emotion categories among emotion classification in different types of comedies from 1740 to 1800</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Emotion category</strong></td>
</tr>
<tr>
<td>MC: Emotions of affection</td>
</tr>
<tr>
<td>Desire</td>
</tr>
<tr>
<td>Love</td>
</tr>
<tr>
<td>Friendship</td>
</tr>
<tr>
<td>Admiration, reverence</td>
</tr>
<tr>
<td>MC: Emotions of pleasure</td>
</tr>
<tr>
<td>Joy</td>
</tr>
<tr>
<td>Schadenfreude</td>
</tr>
<tr>
<td>MC: Emotions of anxiety</td>
</tr>
<tr>
<td>Fear</td>
</tr>
<tr>
<td>Despair</td>
</tr>
<tr>
<td>MC: Emotions of rejection</td>
</tr>
<tr>
<td>Anger</td>
</tr>
<tr>
<td>Abhorrence</td>
</tr>
<tr>
<td>MC: Emotions of suffering and empathy</td>
</tr>
<tr>
<td>Suffering</td>
</tr>
<tr>
<td>Compassion</td>
</tr>
<tr>
<td>Being moved</td>
</tr>
<tr>
<td>Overall</td>
</tr>
</tbody>
</table>
majority of them are sentimental comedies, which should be checked on the basis of the texts. If this is the case, one would have to examine whether one can identify sentimental comedies solely on the basis of these emotional characteristics.

• Satirical comedies from the same period stand out due to high rates of ‘anger’ (19.5% as opposed to 16.1% and 16.6% for the other comedy subgenres) as well as ‘suffering’ (15.5% as opposed to 14.3% and 12.2% for the other comedy subgenres). This looks like the intention of the satirical comedies is implemented on an emotional level. The characters are annoyed with the laughable main character here and recipients are supposed to share that feeling. The love stories, obligatory for the comedy, seem to be brought down to a minimum in this subgenre presumably so that they do not interfere with the ameliorative intention.

• The Kasperl plays are characterized by a larger percentage of ‘schadenfreude’ (2.6%), which is higher than the arithmetic mean of 1.3%, and ‘joy’ (16.3% in contrast to 14.3% and 13.3% in the other two comedy subgenres). The high level of ‘schadenfreude’ may be explained by the characteristics of the central comic character who always tries to take advantage of others and is happy when they have less money, food, alcohol, or prospects for love affairs than he does. It is interesting to note that the Kasperl plays were the production of a theatre that was able to finance a standing business with this type of play for years. In addition, the Kasperl plays were also a blockbuster for many other German-speaking theatres. Further research should investigate the correlation of ‘schadenfreude’ and success.

On the methodological side, we intend to explore various promising approaches to improve classification accuracies. First, we suffer from class imbalances. Thus, we want to explore advanced methods to deal with class imbalances in our setting (Buda et al., 2018). Second, we want to continue exploring the possibilities of historical language models by further pretraining existing contemporary models with a large number of texts. While we did apply these methods, our training texts were rather small in size and did not contain enough text for our specific historical domain. Other research, however, has shown that this method of domain-adaptive pretraining is beneficial in the use case of historical German texts (Labusch, Neudecker, Zellhöfer, et al., 2019; Schweter and Baiter, 2019; Brunner, Tu, Weimer, et al., 2020). Finally, distant and close reading as methodically different but complementary approaches to analysing literary texts may be interpreted as digital humanities’ specific version of a mixed methods approach (Pereira, 2019).

Funding
Investigating emotion distributions and developments in individual texts, across genres and periods, is the goal of the Emotions in Drama project, which since April 2020 has been identifying and researching emotions in dramatic texts of this period using a combined procedure of manual annotation and their prediction using deep learning-based language models. This project is funded by the DFG (German Research Association) in the priority programme Computational Literary Studies (SPP 2207/1) with two grants (project number 424207618; grants DE 2188/3-1 and WO 835/4–1). https://dfg-spp-cls.github.io/projects_en/2020/01/24/TP-Emotions_in_Drama/

Authors’ contributions
Katrin Dennerlein (Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Writing—original draft, and Writing—review and editing), Thomas Schmidt (Data curation, Investigation, Methodology, Software, Visualization, Writing—original draft, and Writing—review and editing), and Christian Wolff (Conceptualization, Funding acquisition, Project administration, Supervision, and Writing—review and editing).

Data availability statement
Data can be found here: https://github.com/lauchblatt/Emotions_in_Drama.

Notes
1. See for the discussion of this development in poetics of drama Meyer-Sickendiek (2005), Kraft (2011), and Ranke (2009).
2. In the field of emotion research in literary studies, a distinction is drawn between production-, reception-, text-, and context-related approaches (Winko, 2003). For an overview of emotion research in literary studies, see Winko (2019, pp. 397–402).
3. The emotion set was iteratively tested during our annotation work at the beginning of the project and adapted several times, as is common and recommended for the humanities (Reiter, 2020). The plus and minus signs refer to the evaluation of an emotion from the perspective of the feeling character. For the special case of the positive value for ‘schadenfreude’ cf. the definition in Section 9.
4. https://dracor.org/ger
5. https://textgridrep.org/
In the following, we do not distinguish between stage directions and spoken texts. For an analysis of emotions in stage directions see Dennerlein, Schmidt, Wolff, 2023.

References


rather than generators. ICLR. https://openreview.net/pdf?id=r1xMH1BvB (accessed 1 April 2023).


