Sentiment Classification of Historical Danish and Norwegian Literary Texts

Abstract

Sentiment classification is valuable for literary analysis, as sentiment is crucial in literary narratives. It can, for example, be used to investigate a hypothesis in the literary analysis of 19th-century Scandinavian novels that the writing of female authors in this period was characterized by negative sentiment, as this paper shows. In order to enable a data-driven analysis of this hypothesis, we create a manually annotated dataset of sentence-level sentiment annotations for novels from this period and use it to train and evaluate various sentiment classification methods. We find that pre-trained multilingual language models outperform models trained on modern Danish, as well as classifiers based on lexical resources. Finally, in classifier-assisted corpus analysis, we confirm the literary hypothesis regarding the author's gender and further shed light on the temporal development of the trend. Our dataset and trained models will be useful for future analysis of historical Danish and Norwegian literary texts.

1 Introduction

Sentiment analysis, the computational study of emotions, opinions, and evaluations expressed in text, has become an important tool in natural language processing. However, its application in literary studies has been limited. Literary texts, such as novels, poems, and plays, provide a unique cultural window into past attitudes, beliefs, and emotions. By analyzing the sentiment expressed in these texts, researchers can gain a deeper understanding of the cultural and societal attitudes of the past and how they have shaped our present understanding of the world. However, traditional sentiment analysis techniques in the investigation of literary texts may be less effective due to the use of an archaic vocabulary, literary ambiguity, and figurative language, as well as the limited training data, the difficulties of generalizing models trained on modern text, and the challenges of annotation. Therefore, there is a need for new methods and approaches to analyze sentiment in historical and literary texts to uncover the valuable insights they can provide.

In this paper, we (1) introduce an annotated sentiment dataset of historical literary Danish and Norwegian, (2) evaluate various classification models on the dataset, and (3) use an accurate classifier to analyze a large historical literary corpus and provide initial evidence for a literary hypothesis (see §2). Our code and data will be publicly available upon publication.

2 Literary Context: The Unhappy Texts

In our experiments, we focus on literature from the so-called 'Modern Breakthrough,' which denotes the literary currents embedded in the new social realist and naturalist literature which blossomed in Scandinavia during the 1870s and the 1880s as well as a cultural and societal transformation, encompassing politics, morals, gender roles, etc (Ahlström, 1947). On the whole, the historiography of the breakthrough, largely concentrated on a few canonical male authors, does not reflect the diversity of the period's literary production, especially in terms of gender. Women were most noticeable among the new groups of authors stepping forward from 1870-1900. As a rule, female authors were not recognized by contemporary (male) criticism; famous critic Georg Brandes introduced the term breakthrough in a book grouping only male Scandinavian authors, 'The Men of Modern Breakthrough (Brandes, 1883).' and, a generation later, the influential literary historian Vilhelm Andersen dubbed the period's female literary production as an 'odd province' of literature

(Hansen, 1886). While Andersen recognized the period's literary production by Danish female authors, a much later literary handbook *Hovedsporet* does not. Its chapter on the debates on gender in the period's literature has the heading 'Women's literature, written by men' (Jørgensen, 2005).
In fact, for many decades of the 20th century, the female breakthrough authors were overlooked and mostly left out of literary histories, but in the 1970's and 80's, as a part of a greater feminist wave in Scandinavian literary scholarship, scholars made a great effort to reinterpret the period. A result of this was a prevalent hypothesis about the female authorships from the period characterized as 'unhappy texts' (Dalager and Mai, 1982; Jensen et al., 1993). The hypothesis and argument are unfolded in greatest nuance in literary scholar Pil Dahlerup's doctoral dissertation, the women of the modern breakthrough (Dahlerup, 1984). Dahlerup argues that the patriarchal society of the period and its restraints on women resulted in female authorships being centered around a disillusionment with women's actual conditions of life and a hereby derived depressive sentiment (Dahlerup, 1984).
Our investigation will revisit this hypothesis with two points of departure, one regarding empirical selection and scale, and another regarding methodology. The hypothesis of the unhappy texts is based solely on texts by female authors and often on relatively few texts. The first point of departure is thus to revisit the hypothesis with a quantitative perspective, also including the works authored by men. Our second point of departure, in revisiting the hypothesis of the unhappy texts, is based on a theoretical framework from the field of gender studies and feminist literary studies. Central to our inquiry are new insights from affect theory and the notion of sentiment and affects as something humans 'do' rather than an underlying *a priori* structure as is the case in the psychoanalytic assumptions that the hypothesis of the unhappy texts is based on. We thus set out to find new methodologies to test the hypothesis of unhappy texts. In this experimental methodological phase, we find that sentiment analysis is a meaningful starting point because of its juxtaposition of affect theoretical and quantitative perspectives.

3 Related Work

In this section, we present sentiment-related works in both literature and historical corpora.

3.1 Sentiment Analysis on Literature

The study of emotions in literature has become an integral part of literary analysis with the emergence of digital humanities. This field of research focuses on using computational methods to understand emotions in literature. This encompasses a wide range of topics, from tracking changes in the plot to analyzing the emotional content of texts (Kim and Klinger, 2019).

Emotion and sentiment classification involves classifying text into predefined classes based on emotions/sentiment. This task is applied in the literature to group literary texts based on their emotional properties. Some studies have focused on classifying the emotions in works of Francisco de Quevedo's poems (Barros et al., 2013), American poetry (Reed, 2018), and early American novels (Yu, 2008). Volkova et al. (2010) annotated in fairy tales, while Ashok et al. (2013) used sentiment polarity to predict the success of a book. Zehe et al. (2016) used sentiment classification to classify 212 German novels into happy or non-happy endings. They showed encouraging results using support vector machines. While some studies approach the task as a classification problem, others focus on the structural changes of sentiment and emotions. Heuser et al. (2016) explore the relationship between emotions and geographical locations and Taboada et al. (2006) track sentiment and emotions towards certain groups.

3.2 Sentiment Analysis on Historical Text

Identifying and tracking the sentiment of text over time is challenging due to language variation, the dynamic nature of sentiment, and the scarcity of historical corpora. Sentiment can change depending on a variety of factors, such as context, culture, and time. Additionally, historical and diachronic data may have different characteristics than contemporary data, such as changes in language use and writing styles. Along with the above challenges, the main focus of research nowadays on temporal corpora in the news (Souma et al., 2019), and social media (Hazimeh et al., 2019), while little attention has been devoted to the historical domain (Sprugnoli et al., 2016). Several techniques are used for sentiment analysis on diachronic and historical corpora. Schmidt and Burghardt (2018) used sentiment analysis on a German drama text corpus and evaluated the performance of different German sentiment lexicons using a manually

annotated gold standard of 200 speeches. This study created an annotated corpus for the sentiment analysis of historical texts and revealed key issues related to the annotation and pre-processing of historical texts. Sprugnoli et al. (2016) analyzed an Italian corpus of writings of Alcide De Gasperi and developed a new lexical resource for sentiment analysis. The study found that crowdsourcing was more effective for sentiment analysis of historical texts than using a sentiment lexicon. Hills et al. (2019) analyzed national subjective well-being using millions of digitized books in six different languages and countries. They found that GDP and life expectancy have a strong positive effect on well-being, while conflict has a negative effect. These studies demonstrate the challenges and potential of sentiment analysis of historical texts and the importance of manual annotation and crowd-sourcing for accurate analysis.

Schmidt et al. (2021) analyzed emotional expression in historical German plays using various methods, including lexicon-based, traditional machine learning, word embeddings, and pre-trained and fine-tuned language models. The latter achieved state-of-the-art results, while lexicon-based and traditional machine learning consistently outperformed. However, performance decreases significantly with multiple sub-emotions.

To summarize, the challenges in sentiment analysis of historical texts, including lack of native speakers, limited data, unusual textual genres, and historical language, call for innovative and robust methods (Sprugnoli, 2021). Pre-trained language models can address these challenges by being trained on large corpora, fine-tuned on small datasets, and able to identify sentiment in diverse texts (§5).

4 Dataset of Historical Literary Text

This section describes our main corpus of historical literary text and a sub-corpus annotated for sentiment.

4.1 Main Corpus

We rely on the MEMO corpus (Bjerring-Hansen et al., 2022), comprising more than 800 Danish and Norwegian novels spanning the last 30 years of the 19th century and including more than 50 million words in total. We refer to this corpus as the 'main corpus'. The corpus is a rich and diverse collection of texts that will provide valuable insights into the registered sentiments and emotions of the period under investigation. Table 1 shows statistical information about the corpus.

	Main Corpus	**Sub-corpus**
Total Novels	839	
Total sentences	3,229,137	2,707
Total Words	52,724,457	55,333
Average sentences per novel	3,849	
Average words per novel	62,842	
Average words per sentence	16	20

Table 1: Corpus statistics for the main corpus and sub-corpus used in sentiment analysis.

4.2 Annotated Sub-Corpus

To effectively implement and evaluate sentiment classifiers, it is necessary to annotate a representative subset of these novels. Our main corpus currently comprises 900 novels from 1870-1900, specifically Danish and Norwegian historical novels. For our initial annotation study, we carefully select a representative sample of 2,700 sentences spanning all novels in the main corpus to be annotated. This was done to ensure that the annotations accurately reflect the sentiment present in the entirety of the corpus and to provide a solid foundation for further research in sentiment analysis on historical novels.

4.3 Annotation Process

The annotation was carried out by three trained literary scholars (a master student, a graduate student, and an associate professor), annotating 648, 1029, and 1029 text segments from the corpus, respectively. The annotators shared domain knowledge in 19th century Scandinavian literature, with overarching and intertwining areas of expertise, but also particular, vested interests (gender, history of ideas, and cultural history).

Guidelines. Following the principle that clear and simple instructions are crucial for obtaining high-quality annotations (Mohammad, 2016), and acknowledging the array of intricacies and baffleements, which an emotional analysis of literary texts based on small fragmentary segments raises, the guidelines were rigorously minimalist and pragmatic.

1. The text segments were to be labeled either 'Positive', 'Negative' or 'Neutral'.

2. Only the segment in question should be considered. Contextualisation and 'guessing' on what might go on before or after the segment were ruled out. In cases of doubt the label should be 'Neutral'.

3. Attention should be paid to the historical fluctuations language and semantic change. As an example the adjective 'besynderlig', today means 'weird', while it in the 19th century also had the meaning 'special' or 'curious'. This means that the following sample should be labeled as positive, rather than, and contrary to an anachronistic reading, negative:

   > 'Hun blev strax noget fortumlet over disse uforberedte Kjærtegn ; men da hun laa i hans Arme , saa' hun op paa ham med et besynderligt, ikke fornærmet Blik' (English translation: 'She was at once somewhat taken aback by these unprepared caresses; but as she lay in his arms, she looked up at him with a curious, not offended, look')

4. Finally, pragmatism was to be deployed by the interpreters. Since the the segments are short and heterogeneous (containing both dialogue, non-dialogue/description or a mixture), no extra weight could be given to particular word classes (such as verbs, adjectives or nouns).

Challenges. A fundamental challenge in the annotation work reflects a key question in narratology (i.e., the study of narrative structure in texts): Who is speaking? In this context, distinctions are made between the text's (implied) author, narrator, and characters as well as between 'mood' and 'voice' as levels in literary texts (Genette, 1983). Since the narratological structure is often unclear, the annotation can not aim at deciding whether the dialogue or text voices reflect the author's sentiment. At the risk of not taking into account literary devices such as irony and unreliability, the annotation can only address the dominant sentiment in the segment. In addition, particular attention was paid to the following issues:

Lack of context. Often, segments with clear indications of emotion are ambiguous or vague due to a lack of context. We decided to label such instances as 'Neutral', even though they in reality, might have formed parts of negative vis-à-vis positive discourses. Example (translated): 'When the letter, finally, was finished, she folded it and went with a beating heart to the nobleman's door'. Here the valence is clear (something emotional is going on), but the polarity is unclear (is it something good or bad?).

(Unconscious) Contextualization. We tried our best to rely our judgments on close reading combined with our familiarity with historical modes of morality and sensibility, including standards of courtesy and decorum, as expressed in speech as well as gesture and action. However, it is difficult not to contextualize or to rule out the role of (unconscious) contextualization. For instance, our knowledge of individual authors or texts and their distinctive traits (e.g., the Norwegian author Amalie Skram, famous for her coarse and unsentimental naturalistic style, permeated with negativity) or paratextual effects such as connotations of a book title (e.g., the novel *En Krise* (1892), 'A Crisis' by Johanne Schjørring, inevitably giving the reader a negative vibe towards the text).

Gender. Finally, and crucially in this context, is the fact that cultural values and norms change over time. Awareness of the issue of cultural value change over time affecting the annotation (gender roles, gendered behavior, then and now). Relying on fundamental insights from social constructivist gender theory, understanding gender as a category subject to change in form and meaning over time, special attention was paid to segments involving gendered behavior and dialogue. A few examples can illustrate, how cultural change poses challenges to inexpressedtion and annotation. First, we have a segment highlighting a male character, whose dominant behavior would be interpreted more negatively today than in the past:

> 'aa ja , men ofte gjorde han sig med Vilje haard , ti saaledes, var det bedst for dem, han havde at gøre med: "du skal!"' (Transl ion: 'oh yes, but often he made himself tough on purpose, for this was the best for those he had to deal with: "you must!"')

In the thematization of female gender roles, a modern understanding could potentially come at odds with the intentions in 19th century texts. This is highlighted in the following segment presenting a female character:

> 'Men allerhelst laa hun dog i Vinteraftenernes Skumring i sin Yndlingsstilling i Armstolen og grublede og drømte og ventede og ventede — ligesom Prinsessen i Eventyret.' (Translation: 'But preferably she would lie in the twilight of the winter evenings in her favourite position in the armchair and pondering and dreaming and waiting and waiting — like the princess in the fairy-tale.'

Sentiment Class	Total Samples	Percentage
Negative	1,117	41.3%
Neutral	788	29.1%
Positive	800	29.6%
Total	2,705	100%

Table 2: Distribution of sub-corpus samples and sentiment categories.

Today, the character's passivity, dreamfulness, and nostalgia are more likely to be perceived as negatives than positives.

4.4 Annotation Results

The sentiment annotations for all sentiment classes are illustrated in Table 2, which displays the distribution of the samples and sentiment classes. Additionally, Table 1 provides further statistical information regarding the sub-corpus.

Agreement. We ues Cohen's Kappa to determine inter-annotator agreement (IAA) on 100 samples annotated by all three experts, resulting in a score of 0.59, which indicates moderate agreement among annotators.

5 Experiments

The dataset is split into three sets in this sentiment classification task: training, validation, and testing. The training set contains 2376 examples, representing approximately 80% of the total dataset. The validation set is used to select the model's hyperparameters. It has 272 samples, representing approximately 10% of the entire dataset. Finally, the testing set is used to evaluate the final performance of the model and contains 100 examples, representing approximately 3% of the total dataset. These are the 100 samples annotated by all three experts, and the final label is decided by a majority vote. For the training and validation sets, each example is annotated only by one expert. We use F1-score as our evaluation metric.

5.1 Models

In this section, we outline the models evaluated in our sentiment classification experiments using both zero-shot and supervised methods. Importantly, all models are based on modern Danish lexica and/or (pre-)training data, with no exposure to historical Danish or Norwegian. In §5.2 and §5.3 we provide detailed information on how these resources were employed to achieve accurate sentiment classification results.

Sentida. Sentida is a lexicon comprised of the existing Danish sentiment lexicon AFINN and a list of new words. It scores sentences based on their words and provides either an average sentiment score or a total score. Sentida takes into account adverb-modifiers, exclamation marks, and negations in its sentiment scoring process (Lauridsen et al., 2019).[1]

TextBlob. A Python package[2] that utilizes a lexicon-based approach in which sentiment is determined by the semantic orientation and the strength of each word in the sentence, using a pre-existing English dictionary that categorizes words as negative, neutral, or positive. We employ a two-step process to ensure accurate sentiment analysis in Danish text. First, we utilize Google Translate[3] to translate Danish text to English. Then, we feed the translated text into TextBlob.

VADER. Valence Aware Dictionary and sEntiment Reasoner is a lexicon and rule-based sentiment analysis tool attuned explicitly to sentiments expressed in social media and works well on texts from other domains (Hutto and Gilbert, 2014).[4] To adjust the VADER sentiment analysis technique for Danish, we use a Danish sentiment lexicon (Nimb et al., 2022; Pedersen et al., 2021)[5] containing a list of words and their associated sentiment scores.

Danish Model BotXO. This BERT model (Devlin et al., 2019) created by Certainly (previously BotXO) has been pre-trained on 1.6 billion Danish words and is freely available.[6] This particular model has not been fine-tuned on sentiment classification.

Danish BERT Tone. The BERT Tone model detects sentiment polarity (positive, neutral or negative) in Danish texts. It was obtained by fine-tuning using a pre-trained Danish BERT model

[1] https://githaub.com/Guscode/Sentida
[2] https://textblob.readthedocs.io/
[3] https://translate.google.com/
[4] https://github.com/cjhutto/vaderSentiment
[5] https://github.com/dsldk/danish-sentiment-lexicon
[6] https://huggingface.co/Maltehb/danish-bert-botxo

Model	Valid.	Test
Zero-shot		
Sentida	**0.64**	**0.63**
TextBlob	0.56	0.52
VADER	0.59	0.62
Danish BERT Tone	0.59	0.62
Danish Sentiment	0.22	0.27
Supervised		
Danish BERT Botxo	0.50	0.52
Danish BERT Tone	0.59	0.70
Danish Sentiment	**0.63**	**0.72**

Table 3: Zero-shot and supervised sentiment classification F1-Score using different methods on validation and testing sets.

(Danish Model BotXO) on Twitter Sentiment and EuroParl sentiment 2.[7]

Danish Sentiment. This model is a fine-tuned version of the multilingual pre-trained model XLM-RoBERTa-base (Conneau et al., 2020) on 198M tweets. It has been fine-tuned using specifically for sentiment analysis.[8]

5.2 Zero-shot classification experiment

Without any additional training or fine-tuning, we evaluate the performance of our sentiment classifiers on the validation and testing split of the subcorpus dataset. We use the Sentida, TextBlob, and VADER tools. Additionally, we use Danish BERT Tone and Danish Sentiment pre-trained language models to perform our zero-shot experiment.

The Sentida classifier achieves the highest results with 64% and 63% on the validation and testing sets, respectively. The results of the other tools and techniques are presented in Table 3. Further analysis reveals that this is due to the high-class imbalance in the other models' predictions. Being the simplest, rule-based classifier, Sentida is the best at transferring to the unseen domain and language variation of historical literary Danish and Norwegian. Despite extensive pre-training and fine-tuning on modern Danish, the pre-trained language models underperform in a zero-shot transfer setting. Fine-tuning on a balanced dataset from our corpus could improve their performance, which is what we proceed to do next.

5.3 Supervised classification experiments

The dataset used in zero-shot classification is also utilized in conducting supervised classification experiments. We fine-tune the three different pre-trained language models and evaluate in our dataset. We use Danish BERT BotXO, Danish BERT Tone, and Danish Sentiment.

In this experiment, we fine-tune the described pre-trained BERT models on a task-specific dataset for sentiment classification. The dataset consists of 2,705 labeled sentences, with a balanced distribution of positive, neutral, and negative sentiments. We use a batch size of 32 and train the model for 10 epochs, using the AdamW optimizer with a learning rate of 1e-3. We evaluate the performance of each model using F1-score. The Danish Sentiment model achieved the highest F1-score of 63% and 72% on the validation and test sets, respectively. Table 3 shows details about the obtained results from each model.

6 Classifier-assisted Corpus Analysis

We perform sentiment prediction on all sentences in the main corpus. We select the best classifier from the above experiments, based on the 'Danish Sentiment' pre-trained language model. We classify high-class sentence levels and align them with the author's gender and the novel's year.

Figure 1 shows the distribution of sentiment distribution relative to the author's gender and the percentage of sentences. On average, female authors exhibit a higher proportion of negative and positive sentiments than male authors.

The analysis of the sentiment values in males and females showed that the average negative sentiment value for males was 34.20%, while the average negative sentiment value for females was 36.33%. On the other hand, the average neutral sentiment value for males was 39.92%, while the average neutral sentiment value for females was 34.74%. Finally, the average positive sentiment value for males was 25.88%, while the average positive sentiment value for females was 28.93%.

Comparing the values, female authors had a higher average negative sentiment value and a higher average positive sentiment value than male authors. The male authors, however, had a higher average neutral sentiment value than the female

[7] https://huggingface.co/alexandrainst/da-sentiment-base

[8] https://huggingface.co/vesteinn/danish_sentiment

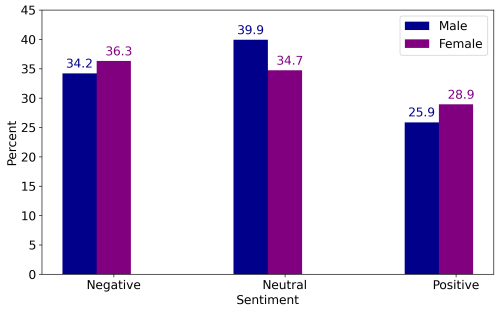

Figure 1: Distribution of sentiment and author's genders. The X-axis is the sentiment class. Y-axis is the percentage of sentences per sentiment category.

authors. These results provide insights into potential differences in sentiment expression between male and female authors in the analyzed data.

Figure 2 provides a detailed analysis of the sentiment distribution over time concerning the author's gender. This figure shows the changes in sentiment tendencies across different gender groups and how these sentiments evolve over a given period. The results provide valuable insights into the dynamic nature of sentiments based on gender and can aid in understanding the evolving nature of sentiments and gender influences. The figure demonstrates that female authors exhibit more negative and positive sentiments over time than male authors.

We calculate the overall sentiment by taking the weighted average of positive, neutral, and negative sentiments, considering sentiment distribution and weights. The sentiment weights are computed as the average of model predictions for each sentiment category. The sentiment of male and female authors is compared, with high scores indicating positive sentiment and low scores indicating negative sentiment (Figure 3).

The significance of our findings was assessed by performing a t-test for each sentiment category, with the null hypothesis that the mean number of sentences with that category across the years is the same for sentences by male and female authors. In statistical analysis, a p-value lower than 0.05 is generally considered statistically significant and enables us to reject the null hypothesis, indicating the counts are statistically different. Table 4 shows small, statistically significant p-values for all sentiment classes.

Sentiment Class	p-value
Negative	5.20×10^{-6}
Neutral	5.13×10^{-8}
Positive	4.57×10^{-10}

Table 4: Comparison of p-values for t-test comparing mean male and female sentiment groups over the years, showcasing significant differences in sentiment trends.

7 Discussion

These initial analyses of the novel corpus with the newly developed sentiment classifier both confirm and contest the thesis of the 'unhappy' female texts. On the face of it, it seems that the female authors in the corpus express not only negative but also positive sentiments. At the same time, the second (implicit) part of the thesis is refuted, as the male authors do not write more positively.

However, further work is needed. In terms of computational interventions, a more in-depth analysis would call for us to employ segmentation of data and comparisons of subcorpora. Are there, for example, certain works or authors that are outliers in terms of gender or valence? Are there any correlations when comparing canonized works and popular literature? By zooming in on either part of the novels (the endings, for example) or parts of the corpus (with regards to time slots or subgroups of producers/individual authors), specific strong trends may stand out.

Also, by working with smaller subsets of the corpus, we could perhaps challenge the essentialist idea of an 'Écriture Féminine', a unique feminine style of writing. Indeed, instead of simply asking whether there is a difference between male and female authorship in the period, we would also be able to explore how this difference is created in the corpus and possibly cultural, historical, gendered, and narratological reasons for it. In this connection, the fact that male authors seemingly write more 'neutrally' while female authors write with a higher degree of valence might be put into perspective through Sara Ahmed's theory of the 'stickiness' of emotions. Inspired by performativity theory, Sara Ahmed is interested in what emotions or affects do, rather than what they are (Ahmed, 2004). Both 'positive' affects, such as (naïve) joy, and 'negative' affects, such as shame,

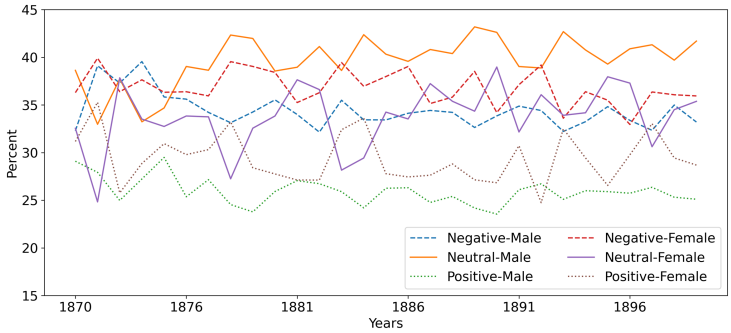

Figure 2: Distribution of sentiment over time. The X-axis is the years. Y-axis is the percentage of sentences of a particular sentiment out of all sentences by authors of the same gender from that year.

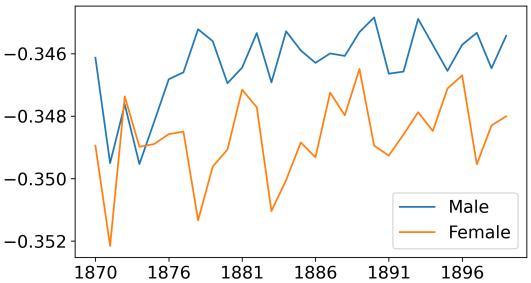

Figure 3: Average sentiment for both male and female authors over time.

have especially in the period but still, today clung to the feminine, while hegemonic masculinity has been and is associated with the rational and unsentimental (Connell, 1995).

In other words, a further refinement of the analytical steps goes hand in hand with critical interactions with the theoretical framework and specific historical contexts concerning the concepts of gender and emotion.

8 Conclusion

In this work, we used the MEMO corpus to create a high-quality human-annotated sentiment dataset for historical literary Danish and Norwegian. Despite multiple challenges, we showed that the task is feasible and that inter-annotator agreement is sufficiently high to warrant the use of the dataset for the training and evaluation of sentiment classifiers. In an extensive evaluation of such models, we found that the best performance is obtained with XLM-RoBERTa fine-tuned on sentiment analysis of modern Danish tweets and then further on the training set from our dataset. Using this model to annotate the whole MEMO corpus automatically, we observe that, as the literary theory predicts, female authors expressed more negative sentiment than male authors in the period.

In future work, we intend to experiment with further improvements of the classifier, including pre-training a language model on the whole corpus as a basis for fine-tuning the sentiment dataset. Furthermore, we intend to complement the classifier with topic models to investigate the evolution of sentiment towards specific topics (cultural change) over time and the evolution of words used to express a sentiment (language change).

From a literary perspective, a sentiment classifier trained on Scandinavian novels from the end of the 19th century holds great potential outside of the context of our specific test case. To give just one example, it would be interesting to study the so-called 'naturalism' from the time of the Modern Breakthrough. This literary movement or current is often said to be characterized by a pronounced pessimistic worldview. One can assume that negative segments predominate naturalist writings, and if that is in fact, the case, the numbers would be interesting to compare with the rest of the literature from the period, i.e., canonical, realistic novels and this period's many forgotten texts.

From a literary theoretical perspective, a classifier such as this can be used to shed light on the importance of narratological layers at a macro level. Although the human reader is far superior to algorithms in separating character, narrator, and (implied) author, the numerical output of the computer is very interesting for existing research in the period. Questions like 'Does the sentiment distribution correspond to the overall assertions of the individual works?' and 'How does sentimental quantity relate to literary expression?' become possible to investigate.

References

Gunnar Ahlström. 1947. *Det moderna genombrottet i Nordens litteratur*. Kooperativa förbundets bokförlag.

Sara Ahmed. 2004. *The cultural politics of emotion*. Routledge.

Vikas Ganjigunte Ashok, Song Feng, and Yejin Choi. 2013. Success with style: Using writing style to predict the success of novels. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1753–1764.

Linda Barros, Pilar Rodriguez, and Alvaro Ortigosa. 2013. Automatic classification of literature pieces by emotion detection: A study on quevedo's poetry. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 141–146. IEEE.

Jens Bjerring-Hansen, Ross Deans Kristensen-McLachlan, Philip Diderichsen, and Dorte Haltrup Hansen. 2022. Mending fractured texts. a heuristic procedure for correcting ocr data.

Georg Brandes. 1883. *Det moderne gjennembruds maend...* Gyldendal.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

R. W. Connell. 1995. *Masculinities*. Polity Press.

Pil. Dahlerup. 1984. *Det moderne gennembruds kvinder.*, 2. edition. Gyldendal.

Stig Dalager and Anne-Marie Mai. 1982. *Danske kvindelige forfattere*, volume 2. Gyldendal.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

G. Genette. 1983. *Narrative Discourse: An Essay in Method*. Cornell UP.

Peter Hansen. 1886. *Illustreret dansk litteraturhistorie*, volume 2. PG Philipsen.

Hussein Hazimeh, Mohammad Harissa, Elena Mugellini, and Omar Abou Khaled. 2019. Temporal sentiment tracking and analysis on large-scale social events. In *Proceedings of the 2019 8th International Conference on Software and Computer Applications*, pages 17–21.

Ryan Heuser, Mark Algee-Hewitt, and ANNALISE LOCKHART. 2016. Mapping the emotions of london in fiction, 1700–1900: A crowdsourcing experiment. In *Literary mapping in the digital age*, pages 43–64. Routledge.

Thomas T Hills, Eugenio Proto, Daniel Sgroi, and Chanuki Illushka Seresinhe. 2019. Historical analysis of national subjective wellbeing using millions of digitized books. *Nature human behaviour*, 3(12):1271–1275.

Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225.

Elisabeth Møller Jensen, Eva Hættner Aurelius, Inger-Lise Hjordt-Vetlesen, Margaretha Fahlgren, Unni Langås, Lone Finnichand, and Anne-Marie Mai. 1993. *Nordisk kvindelitteraturhistorie*, volume 2. Rosinante.

Jens Anker Jørgensen. 2005. *Hovedsporet: Dansk litteraturs historie*. Gyldendal A/S.

Evgeny Kim and Roman Klinger. 2019. A survey on sentiment and emotion analysis for computational literary studies. *Zeitschrift für digitale Geisteswissenschaften*.

Gustav Aarup Lauridsen, Jacob Aarup Dalsgaard, and Lars Kjartan Bacher Svendsen. 2019. Sentida: A new tool for sentiment analysis in danish. *Journal of Language Works-Sprogvidenskabeligt Studentertidsskrift*, 4(1):38–53.

Saif Mohammad. 2016. A practical guide to sentiment annotation: Challenges and solutions. In *Proceedings of the 7th workshop on computational approaches to subjectivity, sentiment and social media analysis*, pages 174–179.

Sanni Nimb, Sussi Olsen, Bolette Sandford Pedersen, and Thomas Troelsgård. 2022. A thesaurus-based sentiment lexicon for danish: The danish sentiment lexicon. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 2826–2832.

Bolette Sandford Pedersen, Sanni Nimb, and Sussi Olsen. 2021. Dansk betydningsinventar i et datalingvistisk perspektiv. *Danske Studier*.

Ethan Reed. 2018. Measured unrest in the poetry of the black arts movement. In *DH*, page 477.

Thomas Schmidt and Manuel Burghardt. 2018. An evaluation of lexicon-based sentiment analysis techniques for the plays of gotthold ephraim lessing. Association for Computational Linguistics.

Thomas Schmidt, Katrin Dennerlein, and Christian Wolff. 2021. Using deep learning for emotion analysis of 18th and 19th century german plays.

Wataru Souma, Irena Vodenska, and Hideaki Aoyama. 2019. Enhanced news sentiment analysis using deep learning methods. *Journal of Computational Social Science*, 2(1):33–46.

Rachele Sprugnoli. 2021. Sentiment analysis for latin: a journey from seneca to dante alighieri. *Sentiment Analysis in Literary Studies*.

Rachele Sprugnoli, Sara Tonelli, Alessandro Marchetti, and Giovanni Moretti. 2016. Towards sentiment analysis for historical texts. *Digital Scholarship in the Humanities*, 31(4):762–772.

Maite Taboada, Mary Ann Gillies, and Paul McFetridge. 2006. Sentiment classification techniques for tracking literary reputation. In *LREC workshop: towards computational models of literary analysis*, pages 36–43.

Ekaterina P Volkova, Betty Mohler, Detmar Meurers, Dale Gerdemann, and Heinrich H Bülthoff. 2010. Emotional perception of fairy tales: achieving agreement in emotion annotation of text. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 98–106.

Bei Yu. 2008. An evaluation of text classification methods for literary study. *Literary and Linguistic Computing*, 23(3):327–343.

Albin Zehe, Martin Becker, Lena Hettinger, Andreas Hotho, Isabella Reger, and Fotis Jannidis. 2016. Prediction of happy endings in german novels based on sentiment information. In *3rd Workshop on Interactions between Data Mining and Natural Language Processing, Riva del Garda, Italy*, pages 9–16.