Projektbericht zum Projekt „Sentiment Analysis for *Emilia Galotti* based on a Speech-level Sentiment Classifier“

Seminartitel: Sentimentanalyse für literarische Texte - 02-25-1074-ue
Dozentin: Svenja Guhr
Semester: WS 2019-20
Teilnehmerinnen des Projekts: Alba Ciliberti, Jingying Wang

TECHNISCHE UNIVERSITÄT DARMSTADT

Autorin:	Jingying, Wang
Matrikelnummer:	2797667
Emailadresse:	wangjingying1994@outlook.com
Studiengang:	Linguistic and Literary Computing
Fachbereich:	FB02 Gesellschafts- und Geschichtswissenschaften
Abgabedatum:	28.03.2020

Table of contents

1. Introduction

As one of the most famous drama of the Enlightenment, *Emilia Galotti* has been analysed by countless people and from various perspectives. Some studies concentrate on the core concept of this play, namely “Tugend (virtue)” and show how this is obsessed by the bourgeois family and leads to tragedy (Riley and Fischer-Lichte, 2002; Robertson, 2009). Some studies reveal the characteristics of different social hierarchy, e.g. the passivity of the bourgeois (Heitner, 1953), the sham and hypocrisy of the people in power (Haag, 2014) etc. based on thorough analysis for related characters. There are also studies concentrating on the relationships between characters, the father-daughter relationship (Milkova, 2013), for example. However, most of these studies are based on qualitative analysis and traditional interpretation methods. In this project, we set out to apply machine learning methods to perform quantitative analysis for *Emilia Galotti* with the focus on the characters’ sentimental aspects.

Although *Emilia Galotti* belongs to the Age of Enlightenment, it is also featured with sentimentalism (Empfindsamkeit). Most of the characters in this play show excessive emotion and according to Robertson (2009), Odoardo, Appiani and Emilia are representatives of different versions of “Schwärmerei”. The prince in this play is also characterized by his emotionality and is “given to emotional self-indulgence almost in the manner of Werther.” (Robertson, 2009: 51). In comparison to the characters who are guided by their emotion, Marinelli seems to be guided by his ambition for power and is capable of making rational or even cold decisions so as to manipulate others. So the first hypothesis is that Marinelli is the most “unempfindsam” character in this play.

Marinelli is the prince’s chamberlain and acts like the prince’s closest friend who tries hard to fulfill the prince’s wishes. However, Marinelli is not always obedient to the prince. He murdered Count Appiani without the prince’s direct permission, even though he is aware that the prince appreciates this honorable man and even wants to connect with him (*1. Act 6. Scene*). And according to De Doncker and Kanz (2012: 78), almost all characters in *Emilia Galotti*, including Marinelli and the prince, pursue their own benefits and regard others as tools for achieving their own goal. So the second hypothesis is that Marinelli does not share the prince’s emotion / feelings even when they are in the same conversation.

In order to test these two hypotheses with quantitative approaches and “speeches are typically the smallest meaningful unit of analysis in quantitative approaches to the study of drama.” (Schmidt and Burghardt, 2018b), these hypotheses need to be translated respectively into: 1. In comparison to other main characters, the percentage of non-emotional speeches for Marinelli is the highest. 2. When the prince and Marinelli are speaking to each other, their speeches contain significantly different emotional valence. Therefore, the sub-goal of this project is to train a speech-level sentiment classifier that can assign sentiment tags to every speech in this drama automatically.

2. Theoretical Aspects of the Project Plan

Sentiment analysis has received more and more attention in the field of literary studies in recent years. Many of these studies are based on sentiment lexicons. In the project done by Eric T. Nalisnick and Henry S. Baird (2013), character-to-character sentiment are computed by summing the valence values provided by the AFINN sentiment lexicon and the emotion dynamics between characters in Shakespeare's plays are visualized based on the calculated scores. Zehe et al. (2016) have tried different sentiment lexicons like NRC and SentiWS to derive feature vectors and applied machine learning methods to train a classifier that predicts happy endings in German novels. Kim et al. (2017) have employed the information for eight basic emotions in NRC lexicon to investigate the relationship between emotional development and literary genres.

There is also a sentiment lexicon-based project which is designed for doing quantitative sentiment analysis for corpus of Lessing's plays (Schmidt and Burghardt, 2018a), including *Emilia Galotti*. The analysis tool developed in the project can be used directly online.[1] However, as our project is designed to use speeches as the analysis unit, this lexicon-based sentiment analysis approach might not be suitable. According to the evaluation project performed by the same researchers, if assigning sentiment tags only by counting sentiment bearing words (SBW) based on sentiment lexicon and some simple NLP techniques like lemmatization or lowercasing, the accuracy can reach 0.705. This score appears to be fine. However, it should be noted that the corpus established for evaluation has certain limitations, since this corpus only contains speeches longer than 18 words. (Schmidt and Burghardt, 2018b: 140) Considering the fact that very short speeches that may contain no SBWs and thus no sentiment scores can be calculated for this kind of speeches based on sentiment lexicon, the real performance of this technique for the whole drama might be lower than 0.705. For *Emilia Galotti,* there are about 66.35% of speeches that are not longer than 18 words[2] and many of them contain no SBW, which could result in classifying the majority of speeches as neutral, as the score remains at 0. This could furthermore distort the analysis result for this project, as the first hypothesis depends on the data for neutral speeches. Therefore, instead of using term counting methods to determine the sentiment of a speech, we decided to apply machine learning methods to train a speech-level sentiment classifier based on a train/test dataset that represents the structural characteristics of *Emilia Galotti*.

[1] http://lauchblatt.github.io/QuantitativeDramenanalyseDH2015/FrontEnd/sa_selection.html
[2] This textual analysis is based on the structured digital text from https://textgridrep.org/browse/-/browse/rksr_0

3. Methodological procedure

3.1 Developing a Speech-Level Sentiment Classifier[3]

As explained above, in order to test the two hypotheses for this project, a speech-level sentiment classifier that has a decent performance is needed. The workflow for developing the classifier are as follows: 1. Pre-process the text to select speeches for manual annotation. 2. Manual annotation (establish gold standard by three annotators) 3. Apply machine learning methods to train a classification model. 4. Fine-tune the model. 5. Evaluate the performance of the classifier. (Repeat the last 2 steps to select the best model). The detailed information about these steps are as follows:

Selection of Speeches: To establish a train/test corpus that can represent the structural features of *Emilia Galotti*, a well-formed XML file[4] with TEI standard for this play is employed. After analysing this file with a python XML parsing module[5], it is found that there are 493 short speeches with an average length of 7.5 words and 343 long speeches with an average length of 38 words in this play. Based on this information and the different lengths of the five acts, 200 speeches (cf. Schmidt and Burghardt, 2018a) are selected according to the following principles[6]: 1. The ratio between short and long speeches (493:343) should be maintained. 2. The number of speeches selected in each act corresponds to the length of the act.

Manual Annotation: Manual annotation is crucial for applying machine learning methods to develop a classifier, as the gold labels for train/test data are determined in this process. After establishing the initial tag set ("*pos", "neg", "neutral"* and "*mix"*) and guidelines[7], we followed the standard annotation procedure of MAMA-Cycle (Model-Annotate-Model-Annotate). Cohen's kappa statistic is employed to evaluate the quality of the manual annotation.

Bag-of-Words Model: Although extracting features based on sentiment lexicon is a common practice for sentiment classifier training, some bias in the lexicon might be transferred into the classifier (cf. Kiritchenko and Mohammad, 2018: 50). Take the word "Prinz" as an example, this word is usually annotated as positive in sentiment lexicons and the classifier based on this kind of lexicon has higher possibility of assigning "pos" tags to speeches that contain the word "Prinz", but in *Emilia Galotti*, this word actually often occurs in negative contexts. In comparison, the BoW Model, which takes the occurrence of each word from the training corpus into consideration, can select "corpus-specific" positive/negative words as features. If a word like "Prinz" frequently occurs in negative speeches, it can be treated as an informative feature for predicting negative speeches for

---

[3] For the full python implementation see *SentimentClassifier_SA_Project_JingyingWang.jpynb*.
[4] https://textgridrep.org/browse/-/browse/rksr_0
[5] https://docs.python.org/2/library/xml.etree.elementtree.html
[6] It should be noted that it is the rationale for the initial selection. The final number of selected speeches is 211 (see the next section).
[7] The final version of tag set and guidelines see *SA_Projekt_Gruppe4_Guidelines.pdf*

this specific drama text. Therefore, the BoW training model together with the linear support vector classification (LinearSVC) algorithm are employed in this project.

Applied NLP techniques: As the bag-of-words model derives features from words, applying different NLP techniques for words can lead to different classification results. In this project, 4 processing techniques are tried in different combinations to get the highest performance, namely: 1. Lowercasing, 2. Lemmatization (with spaCy[8]), 3. Stemming (with SnowballStemmer), 4. Normalization of negations (replace "kein", "nicht" etc. with "negation" to treat them equally).

Performance Evaluation: Since the dataset is small, the 5-fold cross-validation is applied in the classifier evaluation process to tackle the problems like selection bias or overfitting. Standard evaluation metrics used are precision, recall, F1-score and accuracy.

3.2 Applying the Classifier on *Emilia Galotti*

After the classifier is developed, it can be applied on the whole *Emilia Galotti* text and assign every speech a sentiment tag. The corresponding speaker can be matched by making use of the XML file mentioned before. To test the first hypothesis, the percentage of neutral speeches for different characters are calculated by the simple formula: $\frac{\sum neutralSpeechOfGivenChar}{\sum allSpeechesOfGivenChar}$

For the second hypothesis, the emotional dynamics for Marinelli and the prince are visualized respectively with the python module *matplotlib.pyplot*[9]. As we did not assign the sentiment tags with any strength values (cf. Schmidt and Burghardt, 2018a: 144), the representation of the emotional dynamics would be too coarse-grained if the dynamic curve only contain 3 different scores respectively for "pos" (+1), "neutral" (0) and "neg" (-1). In order to represent the emotional state for characters in a more fine-grained way, the sentiment scores are calculated as follows: The sentiment of every 3 successive speeches are considered as a unit to get an average score; If the number of speeches are not a multiple of three, then the rest 1 or 2 speeches are combined with the previous unit to get an average score. If the character only produces 1 speech in a whole Scene, the score of this speech will be discarded to avoid a non-average extreme value.

4. Results

After several annotation cycles and discussions, we reached an inter-annotator agreement of 74.9%, which suggests a moderate consistence. The disagreement mainly lies in extremely short speeches that can be interpreted differently (ambiguity); long speeches with varying sentiment objects or even contradicting sentiments; speeches contain ironies or rhetoric questions etc. The speeches that

[8] spaCy is an industrial-strength NLP python library which provides a powerful model for German language: https://spacy.io/models/de
[9] https://matplotlib.org/3.1.1/api/pyplot_summary.html

are still assigned with different sentiment tags are forwarded to the third annotator for final decision. Altogether 103 “neg”, 57 “neutral”, 23 “pos” and 17 “mix” speeches are annotated. Although these statistics may represent the emotional characteristics of this tragedy, they are too imbalanced for training a classifier. To alleviate this imbalance, the “mix” tags are replaced by other sentiment tags according to the general tendency of the speech. After the replacement, the “pos” samples are still scarce. Thus, 11 additional positive speeches (agreed by all annotators) are selected to enlarge and balance the dataset to some degree[10]. The performance of the classifier trained on this dataset and the results based on this classifier for the drama are demonstrated as follows.

4.1 Performance of the Classifier

As F1-score can be seen as a weighted mean of precision and recall and provide a more plausible view of the performance, we focus more on F1-scores for different tags as well as the accuracy for all tags in this project. Accordingly, the F1-score baselines for each tag are computed by assuming that every speech is assigned with the corresponding tag and the accuracy baseline for all tags are computed by assuming that every speech is assigned with the most frequent tag, namely “neg”. All baselines are shown in Tab. 1. Tab. 2 shows the best performance achieved in this project. According to these 2 tables, although the F1-scores for “pos” and “neutral” tags are not satisfying, F1-scores for all tags and the overall accuracy outperform the corresponding baselines. The overall accuracy of 0.659 does not really meet our expectation, but considering the fact that “Humans typically achieve no greater than 80% accuracy in sentiment classification experiments involving product reviews” (Nalisnick and Baird, 2013: 479), we decided to apply this classifier on *Emilia Galotti* to gain some new insights.

Tab. 1 Baseline

Label Name	F1-score	Accuracy
pos	**0.319**	**0.535**
neg	**0.697**	
neutral	**0.431**	

Tab. 2 Performance of the Classifier

Label Name	Precision	Recall	F1-score	Accuracy
pos	0.95	0.411	**0.507**	**0.659**
neg	0.644	0.893	**0.741**	
neutral	0.625	0.422	**0.497**	

4.2 Results based on the Classifier

For hypothesis 1: After calculating the percentage of neutral speeches respectively for all main characters, the results are displayed as a bar chart (see Fig. 1). According to the chart, the proportion of neutral speeches are quite low among most characters, which corresponds to the feature of sentimentalism. And for Odoardo, Orsina and Emilia these 3 characters, the percentage is even below 10%. In comparison, the proportion of non-emotional speeches for Marinelli (ca. 34%) is

[10] The final version of dataset contains 211 speeches in total.

significantly higher than any of the other characters. The distribution of Marinelli's neutral speeches is displayed in Fig. 2. It is clear that his as "neutral" classified speeches are mainly distributed in *Act 1 Scene 6* (21.1%), *Act 2 Scene 10* (11.8%) and *Act 4 Scene 5* (11.8%).

For hypothesis 2: The emotional dynamics for Marinelli and the prince are displayed in Fig 3. It demonstrates that most of the time when these two characters are talking to each other, they are not in the same emotional status, especially in the early part of *Act 3 Scene 1* and the initial part of *Act 4 Scene 1*. In these two parts, Marinelli has a positive sentiment while the prince shows extremely negative emotion. It is also found that, overall speaking, the emotional valence of Marinelli is usually higher than that of the prince.

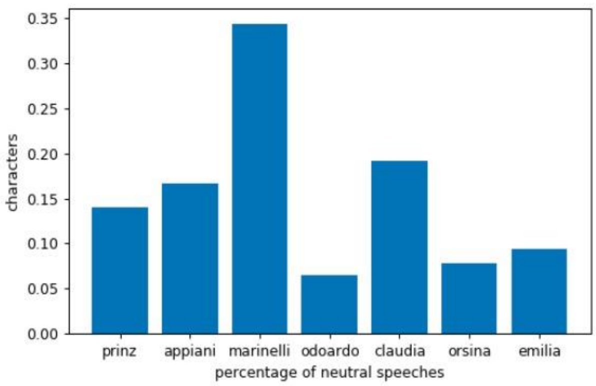

Fig. 1 Percentage of Neutral Speeches for Different Characters

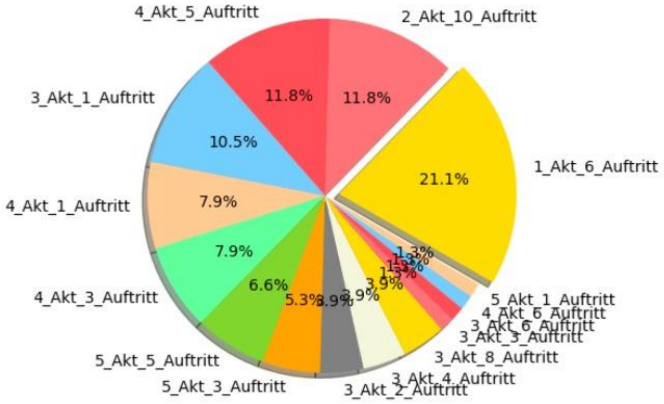

Fig. 2 The Distribution of Marinelli's neutral speeches

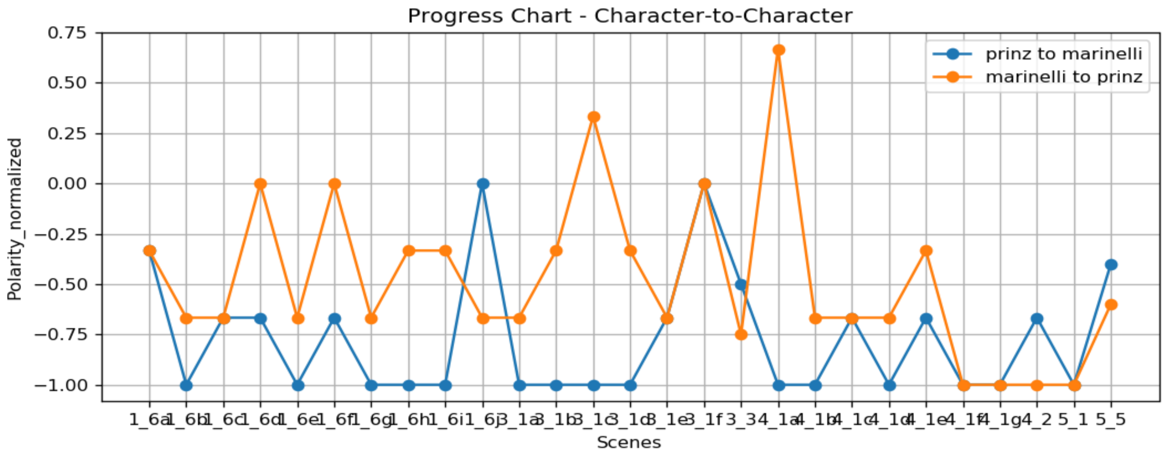

Fig. 3 Sentiment Dynamics - Prinz vs. Marinelli

5. Interpretation and Discussion of the Results

Generally speaking, the results based on the speech-level sentiment classifier demonstrated above support the two hypotheses of this project. Nevertheless, further analysis for specific parts of the drama based on the information provided by the charts above is necessary, in order to find out why Marinelli produces neutral speeches so frequently, or in what kind of situation he expresses no emotion, and why do Marinelli and the prince usually express different feelings when they are in the same conversation.

5.1 Why does Marinelli stay "non-emotional" when facing emotional characters?

According to Fig. 2, Marinelli produces neutral speeches most frequently in *Act 1 Scene 6, Act 2 Scene 10* and *Act 4 Scene 5.* In these 3 scenes, he is respectively in conversation with the prince, Appiani and Orsina. After analysing Marinelli's neutral speeches in the corresponding scenes[11], it is found that most of them are simple answers or very short questions which demonstrate two characteristics about Marinelli:

1. He is uncaring and indifferent: When the prince asks about the news in town, Marinelli reports indifferently: "*<u>Nichts von Belang, das ich wüßte.</u>* – Die Gräfin Orsina ist gestern zur Stadt gekommen."; "*<u>So gut, wie gar nichts.</u>* – Denn daß die Verbindung des Grafen Appiani heute vollzogen wird, – *<u>ist nicht viel mehr, als gar nichts.</u>*" (*Act 1 Scene 6*). And when the prince finds out that the name of Appiani's bride is "Emilia", he becomes extremely upset. He confirms with Marinelli again and again in despair, hoping the bride is not Emilia Galotti, he even says "Sprich dein verdammtes 'Eben die' noch einmal, und stoß mir den Dolch ins Herz!". However, facing the prince's desperation, Marinelli's answer is still "Eben die." with no extra expression of emotion.

2. He is good at using rhetoric questions as replies to avoid giving direct answers that might give him away. When Orsina discovered that Appiani is not died of a robbery but murdered, she confronts Marinelli and forces him to tell the truth, but Marinelli shows no sign of panic:

> Orsina: Wissen Sie nicht, was ich denke?
> Marinelli: **Wie kann ich das?**
> Orsina: Haben Sie keinen Anteil daran?
> Marinelli: **Woran?** (*Act 4 Scene 5*)

Marinelli employs very short questions which shows no actual feelings as replies to Orsina's sharp questions. He might be aware that he could not fool Orsina, but he still tries not to leak any further information and acts as if he is innocent. It is not certain whether Marinelli tries to hide his real emotions in this kind of situations or in fact has no feelings at all because of his indifference or even ruthlessness, nevertheless, based on the overall results of the classifier and further analysis above, Marinelli does seem to be the most "unempfindsam" character in this "empfindsam" play.

5.2 Why does Marinelli not share the prince's feelings?

The discussion above can actually explain the second hypothesis to some degree. Unlike the prince who is usually sentimental and rarely controls or hides his emotions especially when he is talking to Marinelli, Marinelli is indifferent and expresses his feelings selectively to achieve his egoistic goals.

[11] The corresponding speeches are printed in python-script: *SentimentClassifier_SA_Project_JingyingWang.jpynb*

Another reason could be that Marinelli has an abnormal or even twisted sense of enjoyment. According to Fig. 3, Marinelli and the prince have the greatest sentiment difference in the initial part of *Act 4 Scene 1*. In this part, Marinelli shows unusual excitement, while the prince is in deep anxiety:

> Der Prinz (als aus dem Zimmer von Emilien kommend). Kommen Sie, Marinelli! Ich muß mich erholen – und muß Licht von Ihnen haben.
>
> Marinelli: O der mütterlichen Wut! Ha! ha! ha!
>
> Der Prinz: Sie lachen?
>
> Marinelli: Wenn Sie gesehen hätten, Prinz, wie toll sich hier, hierim Saale, die Mutter gebärdete – Sie hörten sie ja wohl schreien! – und wie zahm sie auf einmal ward, bei dem ersten Anblicke von Ihnen – – Ha! ha! – Das weiß ich ja wohl, daß keine Mutter einem Prinzen die Augen auskratzt, weil er ihre Tochter schön findet.

This conversation shows that while the prince is extremely anxious and worried that their intrigue might be exposed, just like a normal person, Marinelli is enjoying the “fun” of a mother filled with rage and he does not try to hide this “natural” feeling.

6. Conclusion

In this project, a speech-level sentiment classifier is developed with machine learning methods, in order to analyse Lessing’s bourgeois tragedy *Emilia Galotti* from a quantitative perspective. The dataset used for training/test consists of 211 manually annotated speeches from this drama. The best performance of the classifier achieved in this project is 0.659 (accuracy), which is not good enough, but outperforms the baseline. The results produced by the classifier can support the two hypotheses about this play, namely, Marinelli is the most “unempfindsam” character in this play and he usually does not share the prince’s emotion / feelings. These results are visualized in section 4 and interpreted in section 5 based on textual evidence from the play.

Developing a sentiment classifier is indeed a challenging task, especially for the domain of literary studies. Nevertheless, as is shown in this project, sentiment classifier can not only facilitate quantitative analysis of literary work, but also help researchers identifying plots that might need more intensive analysis. Therefore, sentiment classifier with better performance is worth pursuing. It is worth noticing that according to the problems identified in this project, *speech* might not be the most suitable classification unit for dramas, as many long speeches contain varying sentiment objects or even contradicting sentiments. Segmentation strategies for establishing the training corpus should be further analysed in the future.

7. Reflection

7.1 Reflection - Implementation of the Project

- Plans and Changes

Since the project idea was proposed by me and the other project participant was occupied at the beginning of the project, I started the project by making a plan (task division) first[12] and pre-processed the textual data for the annotation work. This plan worked well until the phase for developing sentiment classifier. Compared with twitter texts, the classifier training process for this literary work went longer and more complicated than I expected. As the main goal we set at the beginning was to develop a sentiment classifier with the best possible performance, I kept fine-tuning the model and was sort of stuck in the "endless" parameter trying. So I decided to consult our docent for her professional opinions. During the discussion with our docent about the problems with the classifier, we were reminded that the goal was supposed to focus more on the text-based literary analysis and not the tool itself, as the focus of this course is not computational linguistics. I realized that my original plan and hypothesis are not appropriate for this course. Thus, the plan was changed from improving the classifier into formulating new hypotheses based on *Emilia Galotti* that have a literary study focus. This became a big challenge[13] and made the second half of the project stressful. I learned from this experience that before implementing a project, it is crucial to make sure that the goal for the project corresponds to the goal of the superior project (in this case the course).

- Group Meetings

Several group meetings were arranged during the annotation process; new hypotheses formulation process and the presentation preparation process. Meetings for annotation discussion helped us identify our disagreement and different understanding, and was helpful for revising annotation guidelines. However, sometimes meetings were not quite productive. For example, when discussing new hypotheses without enough previous independent research or knowledge about specific related studies or papers as basis, the result of the meeting was meaningless.

- Documentation of the Project

After several discussions during the annotation process, I felt that this was going to be a larger project than I thought, so I proactively documented the annotation problems we met; corresponding examples and the decisions we made during the discussion. And then I shared them on Google Docs[14]. This turned out to be very helpful for preparing the presentation slides. On the other hand, due to the fact that I spent quite much time trying different approaches to improve the performance of the classifier, I only made comments for certain code blocks. Until the other participant posed questions about the code, I did not realize the difficulties this caused for her to understand the python-script. However, the presentation date was already quite near at that moment, so I explained the whole script

[12] I put the plan and the timetable in Google Docs, so that the other participant can see it and give feedbacks: https://docs.google.com/document/d/1rrNKU3KmhF67WSsdTHyr1jUyitKlE2W-xvsEjLvqLxM/edit?usp=sharing

[13] This will be explained later in detail in the C*hallenges* part.

[14]https://docs.google.com/document/d/1_ST2P22UxYCSGCs6bRHIK6kVdybuaUePABKuyK5eGiI/edit?usp=sharing

to her verbally. In order to make it easier for others and also myself to understand how the classifier was developed or replicate the analysis, I added more illustrating comments to codes after the presentation.

- Presentation

In order for all project participants to be able to edit the presentation slides simultaneously, I created a basic presentation file on Google Docs[15] and then we decided about who is responsible for which presentation part. After that we firstly worked on the corresponding slides independently and then exchange feedbacks towards the slides produced by the other one, so that we could modify our own slides accordingly. Nevertheless, sometimes I was still not satisfied with the slides already edited / modified by the other participant, e.g. the slides for different versions of the annotation tables or the first hypothesis. I could not help myself providing some solutions for her slides which I found better for demonstration. Fortunately, she agreed and used them. This part was actually exhausting for me, and maybe also for the other participant. I tried to make the presentation as good as possible, but in the meantime was also afraid to be considered as arrogant or doing some unnecessary work. Then I realized why the docent said that this project could be complicated in the first place. Next time, before I dive in, I would thoroughly assess the time and the scope for the project as well as the interest or the ability the other project participant has for the project.

7.2 Reflection - My Role in the Project

From my point of view, I played a proactive role in this project. I was aware that producing new ideas or solutions based on convincing arguments is harder than giving general comments, and some work just had to be done sooner or later. So for the tasks that both of us were supposed to be responsible for, such as revising guidelines and applying the classifier to do analysis, I usually provided my ideas or solutions proactively. On the other hand, I'm not sure whether this resulted in influencing the other participant's thoughts or motivation in a not ideal way. For some trivial things like correcting minor grammatical mistakes or putting references on the slides, I took care of them proactively as well. I also acted as a mediator passing on messages or feedbacks between my partner and the other classmate Ronja (the third annotator); between my partner and the docent.

7.3 Reflection - Challenges

I encountered different kinds of challenges during this project. The first challenge was establishing the training data (gold standard). As a non-native speaker, I found it difficult to understand this German drama from the 18th century. In order to assign sentiment tag as properly as possible, I listened to the audio version of *Emilia Galotti* and also read the corresponding English

[15]https://docs.google.com/presentation/d/1FCLZAkEVJbzh_fJGRQRtlerWLnsfUxrk3_nLWaJuR10/edit?usp=sharing

translation[16] to have a better understanding of the context. As mentioned before, reformulating hypotheses in the late phase of the project was also a big challenge. The first new hypothesis we agreed upon was about the stereotyping of female characters' emotion. However, despite intense search, we still could not find enough supporting literature. It turned out that we spent too much time on a wrong direction. The presentation day was very close and I was upset that no proper hypothesis was formulated yet. I continued doing literature research, but focused more on the character which interested me the most (*Marinelli*). Finally, before the presentation, I was able to formulate a new hypothesis that I found better and did corresponding analysis with the classifier.

Another challenge was condensing this big amount of work into a 25-minute presentation, which we unfortunately failed to accomplish. Next time I would definitely arrange rehearsal time to have a better time management for the actual presentation. There are also some technical challenges for me, e.g. I had to apply new methods and unfamiliar python packages to modify the classification model or visualize the results. Therefore, I read plenty of documentation about the methods provided by the website for *scikit-learn* library and watched tutorial videos about *Matplotlib* to handle these techniques.

[16]https://www.gutenberg.org/files/33435/33435-h/33435-h.htm#div1Ref_Emilia

References

De Doncker, J., & Kanz, C. (2012). Egoismus bzw. Selbstsucht im bürgerlichen Trauerspiel des 18. Jahrhunderts. Eine analyse Lessings Emilia Galotti und Lenz' Die Soldaten aus egoistischer Sicht. .

Heitner, R. (1953). "Emilia Galotti": An Indictment of Bourgeois Passivity. The Journal of English and Germanic Philology, 52(4), 480-490. Retrieved February 20, 2020, from www.jstor.org/stable/27713594

Ingrid Haag, « Über die „Wahrheit" der weiblichen Natur und wie diese auf der Bühne des bürgerlichen Trauerspiels Lügen gestraft wird », Cahiers d'Études Germaniques, 67 | 2014, 123-136. DOI : https://doi.org/10.4000/ceg.1738

Kiritchenko, S., & Mohammad, S.M. (2018). Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems. ArXiv, abs/1805.04508.

Milkova, Maria (2013). Vater-Tochter-Verhältnis in Lessings Emilia Galotti, München, GRIN Verlag, https://www.grin.com/document/263190

Nalisnick, Eric & Baird, Henry. (2013). Character-to-Character Sentiment Analysis in Shakespeare's Plays. ACL 2013 - 51st Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference. 2. 479-483.

Prodinger, Anna (2012), Spannungsfelder in Lessings "Emilia Galotti", München, GRIN Verlag, https://www.grin.com/document/195945

Riley, J., Fischer-Lichte, E. (2002). History of European Drama and Theatre. London: Routledge, https://doi.org/10.4324/9780203450888

Robertson, Ritchie. (2009). Virtue Versus 'Schwärmerei' in Lessing's Emilia Galotti. German Life and Letters. 62. 39 - 52. 10.1111/j.1468-0483.2008.01447.x.

Schmidt, Thomas und Burghardt, Manuel (2018a) An Evaluation of Lexicon-based Sentiment Analysis Techniques for the Plays of Gotthold Ephraim Lessing. In: SIGHUM Workshops on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH-CLfL 2018), 25. Aug. 2018, Santa Fe, USA.

Schmidt, Thomas und Burghardt, Manuel (2018b) Toward a Tool for Sentiment Analysis for German Historic Plays. In: Piotrowski, Michael, (ed.) COMHUM 2018: Book of Abstracts for the Workshop on Computational Methods in the Humanities 2018. Laboratoire laussannois d'informatique et statistique textuelle, Lausanne, Switzerland, S. 46-48.

Jannidis, F., Reger, I., Zehe, A., Becker, M., Hettinger, L. & Hotho, A. (2016). Analyzing Features for the Detection of Happy Endings in German Novels. arXiv preprint arXiv:1611.09028. online: http://ceur-ws.org/Vol-1646/paper2.pdf.