Detecting Protagonists in German Plays around 1800 as a Classification Task

Abstract

In this paper, we aim at identifying protagonists in plays automatically. To this end, we train a classifier using various features and investigate the importance of each feature. A challenging aspect here is that the number of spoken words for a character is a very strong baseline. We can show, however, that a) the stage presence of characters and b) topics used in their speech can help to detect protagonists even above the baseline.

Introduction

In his pioneer study on drama, Manfred Pfister strongly demands a way to identify protagonists: "At the current level of research it is not possible to grade the figures of the *dramatis personae* precisely according to their importance for the development of the plot [... M]ore subtle distinctions such as those between 'major figures' [and] 'minor figures' [...] can only be guessed intuitively and not defined operationally" (Pfister 1988, 166). In this paper, we explore such a way to identify protagonists in German plays. Casting the problem as a classification task allows to inspect the influential factors for the decision, and at the same time offers a clear evaluation method. Diverging definitions of "protagonist" and "hero" circulate in literary studies, whereby different aspects of characters are mentioned as defining criteria. In certain aspects, the definitions of "protagonist" and "hero" clearly overlap, e.g. their plot relevance. Other aspects are distinctive: According to ancient understanding, the only hero of a play must be of good nature, but not flawless (Aristotle 1982, 7), whereas later interpretations of Aristotle allow heroes which can have mixed (Lessing, Hamburgische Dramaturgie, 86. Stück and Martus (2011), 15) or even overly negative characteristics (Asmuth 1997, 94). Since we want to identify those characters automatically that are the most relevant for the plot in a given drama (Pfister 1988, 234), including negative heroes like Woyzeck and Macbeth, a value-based evaluation of a single hero is not feasible. Instead, we opted for value-neutral protagonist-detection and define protagonist as *characters that have a central scope of action either by acting themselves or by triggering the action*. Consequently, more than one character can be a protagonist, and we make no assumptions on polarity.

There are already some publications that focus on sub-classifying literary characters on a formal basis or automatically. Several research projects follow the character classification proposed by Propp (1958) (e.g., Declerck, Koleva, and Krieger (2012), Finlayson (2015)). It distinguishes seven character types in folk tales by their plot function. Moretti (2011, 2013) describes experiments to formalize the notion of protagonists, but makes use of network-based features only. Algee-Hewitt (2017) and Fischer et al. (2018) concentrate on the development of protagonists throughout literary history using different network metrics for corpus analyses of English and German plays. In contrast, Jannidis et al. (2016) try to automatically detect main characters in German novels. For the classification, they use summaries of the novels as a gold standard.

Detecting protagonists in plays

Corpus

Table 1: Corpus overview

Sub corpus	Characters (Ø)	Protagonists (%)
SD	20.8	15.1
BT	12.0	25.7
WK	35.8	7.6

The corpus we employ for this research consists of 16 plays, distributed over the three literary periods/genres *Sturm und Drang* (SD), *bourgeois tragedy* (Bürgerliches Trauerspiel, BT), and *weimar classicism* (Weimarer Klassik, WK). Protagonists have been identified manually according to the above definition. The labeling has been conducted by two authors of this paper in parallel. We did not measure agreement quantitatively, but the perceived agreement was well enough. In total, 40 out of 355 characters are annotated as protagonists (11.3%). The average number of protagonists per play differs significantly between the three sub corpora (cf. Table 1). Therefore, we also test for the influence of a play's affiliation to a specific sub genre.

Features

Table 2: Features and their description

Feature	Info. source	Description
Tokens	Text amount	The normalized number of tokens a character speaks during the whole play
Degree	Social relations	The degree of the node representing a character in a co-presence network based on scenes
Active scenes	Stage presence	The number of scenes in which a character speaks, normalized for the play
Passive scenes	Stage presence	The number of scenes in which a character is mentioned by name, normalized
T1-T20	Speech content	The probability of topic 1-20 given the speech of a character

Table 2 shows the entire set of features used in our experiments. The core idea behind this selection is that the different features represent different information sources: The number of tokens represents the active contribution of a character to a play. The degree represents (in an abstract way) the social relations a character has. A character with a high degree is co-present with more *different* characters than a character with a low degree.[1] Active and passive scenes represent the presence of a character over the course of the play, including their passive presence. Finally, the features T1-T20 are probabilities for topics given the speech of a character and represent the thematic content of the character speech. The topic model has been trained on a corpus of 2,735 German language fictional prose texts provided by Fischer and Strötgen (2015). In the following experiments, we compare performances for different information sources.

Given the large differences between the sub corpora, we first tested the influence of a (one-hot encoded) sub corpus feature by comparing classification performance with and without this information. The small differences were not significant. Hence, we did not include this feature in the following experiments.

As classification algorithm, we employ a support vector machine with linear weights. Since we are primarily interested in comparing the performance of feature sets, we did not tune parameters further. Evaluation has been conducted in a 10 fold cross validation setting. To cancel the class imbalance, we used random oversampling of the minority class in each training fold. Feature values are centered and scaled before processing. Technically, this was achieved with the R package `caret` (Kuhn 2008).

---

[1]While using the weighted degree often yields more insightful results, it reflects a different kind of information. The weighted degree arguably correlates with the number of tokens a character speaks, while the unweighted degree purely represents the number of characters that are co-present, ignoring how often they are co-present.
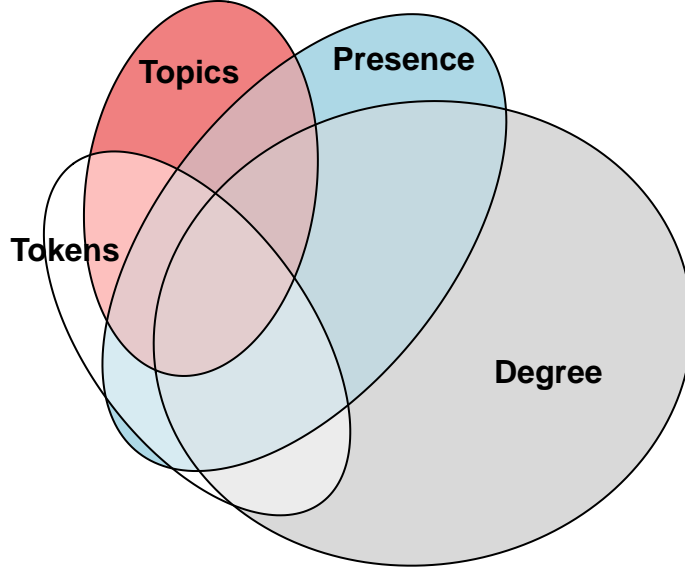
Figure 1: Classification errors by feature set as an Euler chart

Experiment 1: Individual features

Table 3: Evaluation results for Experiment 1, showing precision, recall, and accuracy for individual feature sets: (P) = protagonists, (C) = other characters

	Precision (P)	Recall (P)	Precision (C)	Recall (C)	Accuracy
Majority.Baseline	-	0.00	0.89	1.00	0.89
Tokens	0.59	0.92	0.99	0.92	0.92
Degree	0.30	0.70	0.95	0.79	0.78
Presence	0.47	0.90	0.99	0.87	0.87
Topics	0.57	1.00	1.00	0.90	0.92

Table 3 shows the evaluation results of models trained using the different features as well as results achieved by a majority baseline. Since the data is not evenly distributed, classifying every character as a non-protagonist yields an accuracy of 89%. In addition to accuracy, the classification results are shown in precision and recall for both classes separately. Because the classification is binary, precision errors for protagonists correspond to recall errors for other characters.

Taking a closer look at the protagonist results, we see that all features overgenerate. Only between 30% and 59% of the predicted protagonists are actually protagonists. The recall, however, is already very high for all features. Most of the actual protagonists are detected.

But there are clear differences between the features: The best performing uses the pure number of spoken tokens, with a precision of (59%), closely followed by topic probabilities 57% and the presence metrics (47%). The degree is the worst predictor for protagonism in our data set. It is an interesting observation that topics and tokens achieve a similar performance, although the features encode so different information.

Error Analysis To gain a better understanding of the classification performances, we compare the mistakes made by the different models. Figure 1 shows *only the classification errors* (false positives and false negatives) as an Euler chart.[2] This allows to see to what extent the models made the same mistakes. As can be seen,

---

[2]Please note that the Euler chart only shows an approximation. The overlapping areas are optimized to be proportional.

the number of errors made by all models is quite small: Only 6 mistakes are made by all models (the area in which all ellipses overlap).[3]

Experiment 2: Combining features

Table 4: Evaluation results for Experiment 2, showing precision, recall and, accuracy for combined feature sets

	Precision (P)	Recall (P)	Precision (C)	Recall (C)	Accuracy
Diverging	0.74	0.98	1	0.96	0.96
All	0.69	1.00	1	0.94	0.95

This observation motivates our final experiment: A combination of the features. If providing different information sources to the classifier leads to increased performance, this is a strong argument that the decision for or against protagonist status for dramatic characters is a multifactorial one. Table 4 shows results for joining all features together as well as the feature sets with clearly diverging mistakes (i.e., all except degree). Taking multiple information sources into account leads to a clear improvement: The precision for detecting protagonists in plays reaches 74%, with an overall accuracy of 96% (for the most diverging features). With Amélie from Holteis *Ein Trauerspiel in Berlin*, only one actual protagonist has not been detected automatically.

Conclusions

We have shown how protagonists in dramatic texts can be identified with high accuracy using features that represent different knowledge sources in the text. A first insight is that criteria for protagonism are stable across sub corpora. Adding sub corpus information does not change the results significantly. A second insight relates to the multifold dimensions of protagonists: While just counting the spoken words already achieves high performance, we were able to improve on that significantly. We take this as an indication that any attempt of identifying protagonists (and potentially other character types as well) **must** take into account different information sources, and not only rely on network or textual features. This is in line with the complex human conceptualization of dramatic characters. Character-focused digital drama analysis (e.g., Moretti (2013), Willand and Reiter (2017), Trilcke et al. (2017)) has often focused on only one dimension, which undoubtedly has pragmatic reasons. However, results achieved in this way have a limited scope.

It is not a new insight that the amount of speech is a very good hint at identifying the protagonist(s) of a play. Although it has been considered uncontroversial since Pfister (1988), we found that there is more to the picture. In contrast to speech, it actually is surprising that topics are doing so well in identifying protagonists. If we assume for a moment that the induced topics actually represent topic-like units, a possible explanation aims at the diverse poetological concepts of characters according to genre and time. It is hard to think of a single protagonist-like topic-identity, but the very combination of multiple topics may constitutes a protagonist. Minor characters are more likely to onedimensionally represent only a single or few of the topics. Given the difficulty of interpreting results achieved by topic modelling (cf. Chang et al. 2009), a closer look at these results will be next on our research agenda.

References

Algee-Hewitt, Mark. 2017. "Distributed Character: Quantitative Models of the English Stage, 1500-1920)." In *Proceedings of the Annual Conference of the Alliance of Digital Humanities Organizations*, 119–21. Montréal, Canada: Alliance of Digital Humanities Organizations.

Aristotle. 1982. *Poetik.* Edited by Manfred Fuhrmann. Stuttgart, Germany.

---

[3]Madame Welldorf (Engel, Johann Jakob: *Eid und Pflicht*), Amalie (Pfeil, Johann Gottlob Benjamin: *Lucie Woodvil*), Betty (Pfeil, Johann Gottlob Benjamin: *Lucie Woodvil*), Illo (Schiller, Friedrich: *Die Piccolomini*), Karl (Schiller, Friedrich: *Die Jungfrau von Orleans*), Gräfin Terzky (Schiller, Friedrich: *Wallensteins Tod*)

Asmuth, Bernhard. 1997. *Einführung in Die Dramenanalyse.* 5th ed. Sammlung Metzler. Stuttgart, Germany: Metzler.

Chang, Jonathan, Sean Gerrish, Chong Wang, Jordan L. Boyd-graber, and David M. Blei. 2009. "Reading Tea Leaves: How Humans Interpret Topic Models." In *Advances in Neural Information Processing Systems 22*, edited by Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, 288–96. Curran Associates, Inc. http://papers.nips.cc/paper/3700-reading-tea-leaves-how-humans-interpret-topic-models.pdf.

Declerck, Thierry, Nikolina Koleva, and Hans-Ulrich Krieger. 2012. "Ontology-Based Incremental Annotation of Characters in Folktales." In *Proceedings of the 6th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, 30–34. Avignon, France: Association for Computational Linguistics. http://www.aclweb.org/anthology/W12-1006.

Finlayson, Mark A. 2015. "ProppLearner: Deeply Annotating a Corpus of Russian Folktales to Enable the Machine Learning of a Russian Formalist Theory." *Digital Scholarship in the Humanities*, December. https://doi.org/10.1093/llc/fqv067.

Fischer, Frank, and Jannik Strötgen. 2015. "When Does German Literature Take Place? – On the Analysis of Temporal Expressions in Large Corpora." In *Proceedings of the Annual Conference of the Alliance of Digital Humanities Organizations.* Sydney, Australia: Alliance of Digital Humanities Organizations.

Fischer, Frank, Peer Trilcke, Christopher Kittel, Carsten Milling, and Daniil Skorinkin. 2018. "To Catch a Protagonist: Quantitative Dominance Relations in German-Language Drama (1730-1930)." In *Proceedings of the Annual Conference of the Alliance of Digital Humanities Organizations*, 193–201. Mexico-City, Mexico: Alliance of Digital Humanities Organizations.

Jannidis, Fotis, Isabella Reger, Markus Krug, Lukas Weimar, Luisa Macharowsky, and Frank Puppe. 2016. "Comparison of Methods for the Identification of Main Characters in German Novels)." In *Proceedings of the Annual Conference of the Alliance of Digital Humanities Organizations*, 578–82. Kraków, Poland: Alliance of Digital Humanities Organizations.

Kuhn, Max. 2008. "Building Predictive Models in R Using the Caret Package." *Journal of Statistical Software, Articles* 28 (5): 1–26. https://doi.org/10.18637/jss.v028.i05.

Martus, Steffen. 2011. "Transformationen Des Heroismus. Zum Politischen Wissen Der Tragödie Im 18. Jahrhundert Am Beispiel von J. E. Schlegels *Canut*." Edited by Torsten Burkhard, Markus Hundt, Steffen Martus, and Claus-Michael Ort, 15–42.

Moretti, Franco. 2011. "Network Theory, Plot Analysis." Pamphlets of the Stanford Literary Lab 2. Stanford Literary Lab.

———. 2013. ""Operationalizing": or, the function of measurement in modern literary theory." Pamphlets of the Stanford Literary Lab 6. Stanford Literary Lab.

Pfister, Manfred. 1988. *The Theory and Analysis of Drama.* Translated by John Halliday. European Studies in English Literature. Cambridge University Press.

Propp, Vladimir Yakovlevich. 1958. *Morphology of the Folktale.* 2nd ed. Austin, TX: University of Texas Press.

Trilcke, Peer, Frank Fischer, Mathias Göbel, Dario Kampkaspar, and Christopher Kittel. 2017. "Netzwerkdynamik, Plotanalyse – Zur Visualisierung Und Berechnung Der ›Progressiven Strukturierung‹ Literarischer Texte." In *Book of Abstracts of Dhd 2017.* Bern, Switzerland.

Willand, Marcus, and Nils Reiter. 2017. "Geschlecht und Gattung. Digitale Analysen von Kleists Familie Schroffenstein." *Kleist-Jahrbuch*, Kleist-jahrbuch, 2017: 142–60. https://doi.org/10.1007/978-3-476-04516-4/_16.