

Klassifikation von Figurenauf- und -abtritten in XML-kodierten Dramen

Ehlers, Lena

st161358@stud.uni-stuttgart.de
Universität Stuttgart, Deutschland

Andresen, Melanie

melanie.andresen@ims.uni-stuttgart.de
Universität Stuttgart, Deutschland

In diesem Beitrag wird eine regelbasierte Methode vorgestellt, um Figurenauf- und -abtritte in den Regieanweisungen dramatischer Texte zu klassifizieren. In der Forschung wurde Regieanweisungen meist nur wenig Beachtung geschenkt, etwa weil sie während einer Theateraufführung nicht textuell in Erscheinung treten (Schöbler, 2017, S.3). Als eine der wenigen quantitativen Untersuchungen stellen Trilcke et al. (2020) fest, dass sich die vermutete Episierung des Dramas im Laufe der Jahrhunderte am Korpus GerDraCor bestätigt. Eine Differenzierung nach Funktionen der Regieanweisungen erfolgt nicht.

Dabei ist die (Ko-)Präsenz von Figuren auf der Bühne eine häufig genutzte Grundlage für quantitative Dramenanalysen, insbesondere in der Netzwerkanalyse (z.B. Marcus 1973, Krautter et al. 2018, Fischer et al. 2018). Trilcke et al. (2017) konnten bereits am Beispiel von Lessings Emilia Galotti zeigen, dass in statischen Netzwerken, die das ganze Drama auf einmal darstellen, wichtige Informationen zur Dynamik der Beziehungen zwischen den Figuren verloren gehen können. Unserer Ansicht nach ist außerdem zu berücksichtigen, dass Figuren auch innerhalb von Szenen auf- und abtreten und die Anwesenheit von zwei Figuren in einer Szene nicht zwangsläufig bedeutet, dass diese Figuren auch gleichzeitig auf der Bühne stehen.

Korpus und manuelle Annotation

Das Deutsche Dramenkorpus GerDraCor enthält über 550 deutschsprachige, TEI-kodierte Dramentexte aus dem Zeitraum von 1650 bis 1947 (Fischer et al. 2019). Die Regieanweisungen sind als *stage*-Elemente kodiert, die bisher keine weiteren Attribute besitzen, die Auskunft über den Inhalt der Regieanweisungen, wie z.B. das Auf- oder Abtreten von Figuren, geben würden.

Insgesamt wurden 16 Dramentexte manuell annotiert, wovon vier zur Implementierung¹ und zwölf zur Evaluation genutzt wurden. Die Guidelines für die manuelle Annotation stehen unter <https://doi.org/10.5281/zenodo.6951465> zur Verfügung. Die Annota-

tion erfolgte direkt im XML-Format. Abbildung 1 zeigt, wie ein *stage*-Element mithilfe der Attribute *type* und *who* um die extrahierten Informationen erweitert wird.

```
<sp who="#wagner">
  <speaker>WAGNER.</speaker>
  <lg>
    <l>Ich hätte gern nur immer fortgewacht,</l>
    <l>Um so gelehrt mit Euch mich zu besprechen.</l>
    <l>Doch morgen, als am ersten Ostertage,</l>
    <l>Erlaubt mir ein' und andre Frage.</l>
    <l>Mit Eifer hab' ich mich der Studien beflissen;</l>
    <l>Zwar weiß ich viel, doch möchte' ich alles wissen.</l>
  </lg>
  <stage>Ab.</stage>
</sp>
```

Abbildung 1 zeigt die Erweiterung des *stage*-Elements zu *<stage type="exit" who="#wagner">Ab.</stage>*.

Abbildung 1: Exemplarische Ergänzung eines *stage*-Elements

Regelbasierte Annotation

Das entwickelte Verfahren klassifiziert den unstrukturierten Text innerhalb der *stage*-Elemente mithilfe manueller erstellter Regeln. Diese basieren auf Schlüsselwörtern und -phrasen („treten herein“, „gehen ab“, ...) und nutzen reguläre Ausdrücke. Zusätzlich wird die Position einbezogen, etwa bei Regieanweisungen am Anfang einer Szene, die oftmals ausschließlich die Namen der anwesenden Figuren enthalten.

Wurde im ersten Schritt ein Auf- oder Abtritt erkannt, folgt als zweiter Schritt die Zuordnung der betroffenen Figuren. Hierfür wird die im XML enthaltene Liste der Namen aller sprechenden Figuren genutzt, die auch die Abbildung auf die Figuren-IDs ermöglicht. Der Auf- oder Abtritt wird entweder einer in der Regieanweisung genannten Figur oder derjenigen Figur, deren Rede die Regieanweisung zugeordnet ist (vgl. Fall in Abb. 1), zugeschrieben.

Evaluation

Die Evaluation erfolgt anhand von zwölf manuell annotierten Texten, die nicht zur Aufstellung der Regeln herangezogen wurden. Evaluert werden 1) die Klassifikation der Regieanweisungen in Figurenauf- und -abtritte und 2) die Zuordnung der betroffenen Figuren. Im zweiten Schritt werden nur die Elemente in die Evaluation einbezogen, die im ersten Schritt korrekt klassifiziert wurden. Tabelle 1 zeigt, dass die durchschnittlichen Werte für Precision, Recall und F1-Score für die Auf-/Abtritterkennung bei 0,85 liegen. Auch die Figurenerkennung liefert gute Ergebnisse (F1 = 0,87). Zwischen den Texten zeigt sich allerdings eine erhebliche Variation in der Qualität.

Tabelle 1: Evaluationsergebnisse

Auto- r*in	Titel	Jahr	Auf- Abtrit- terken- nung	Figur- ren- ken- nung						
			P	R	F1	n	P	R	F1	n
Gottsched	Das Testament	1745	0,94	1	0,97	543	0,95	0,65	0,77	63
Schlegel	Canut	1746	1	1	1	23	1	1	1	23
Gellert	Die zärtlichen Schwes- tern	1747	0,95	0,98	0,97	129	0,99	0,99	0,99	63
Pfeil	Lucie Wood- vil	1756	0,95	0,97	0,96	141	1	0,9	0,95	78
Lenz	Der Hof- meister	1774	0,86	0,83	0,85	315	0,89	0,8	0,84	70
Schiller	Die Räuber	1781	0,8	0,79	0,79	544	0,86	0,74	0,79	75
Goethe	Die natürliche Tochter	1803	0,93	0,97	0,95	101	0,97	0,98	0,98	28
Kleist	Die Familie Schrocken- stein	1803	0,85	0,7	0,76	320	0,91	0,73	0,81	71
Günther	Magie und Schick- sal	1805	0,93	0,95	0,94	84	0,98	0,88	0,93	37
Günther	Udohla	1805	0,94	0,62	0,74	41	1	0,96	0,98	16
Weißenthurn	Das Manuscript	1817	0,7	0,89	0,79	675	0,85	0,72	0,78	64
Hofmannsthal	Der Rosenkavalier	1911	0,39	0,47	0,43	715	0,76	0,41	0,53	38
Mittelwerte		0,85	0,85	0,85		0,81	0,93	0,87		

Aufgrund des regelbasierten Verfahrens schneiden Texte, die stark von den zur Erstellung der Regeln verwendeten Texten abweichen, in der Evaluation schlechter ab. Besonders Texte mit langen Regieanweisungen sorgen dafür, dass viele Schlüsselwörter auch in anderen Kontexten vorkommen. Das zeigt sich insbesondere beim *Rosenkavalier*, dessen Regieanweisungen mit 12 Tokens im Mittel doppelt so lang sind wie der Durchschnitt aller Dramen. Ein weiteres Problem stellen von der Figurenliste abweichende Namen dar, etwa Varianten des Eigennamens oder Appellativa. Ersteres könnte in Zukunft durch die Nutzung von Ähnlichkeitsmaßen adressiert werden, die aber zu mehr Falsch-Positiven führen können.

Analyse

Abbildung 2 zeigt, dass die meisten Auf- und Abtritte tatsächlich innerhalb von Szenen stattfinden (ca. 46%). Etwas weniger erfolgen am Beginn einer Szene (41%) und etwa 13% am Ende. Erwartungsgemäß handelt es sich am Anfang der Szene fast ausschließlich um Auftritte, am Ende um Abtritte. Innerhalb der Szenen komme Auf- und Abtritte zu jeweils gleichen Anteilen vor.

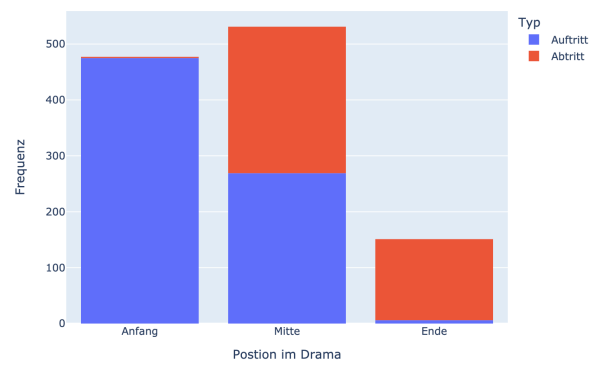


Abbildung 2: Verteilung der annotierten stage-Elemente auf den Beginn, die Mitte und das Ende einer Szene.

Fazit

In diesem Paper haben wir ein mit Figurenauf- und -abtritten annotiertes Teilkorpus zu GerDraCor präsentiert und einen regelbasierten Algorithmus vorgestellt, der diese Annotationen mit einem mittleren F1-Wert von über 0,85 reproduzieren kann. Ein Großteil der annotierten Auf- und Abtritte erfolgt innerhalb von Szenen. Diese Veränderungen in den Figurenkonstellationen werden bei einer szenenweisen Betrachtung der Figurenpräsenz nicht berücksichtigt, haben aber potenziell Auswirkungen auf beispielsweise netzwerkanalytische Arbeiten. Alle Daten und Skripte zu diesem Beitrag sind unter <https://github.com/quadrama/enter-exit> verfügbar.

Fußnoten

1. Es handelt sich um *Emilia Galotti* (Lessing, 1772), *Götz von Berlichingen* (Goethe, 1773), *Iphigenie auf Tauris* (Goethe, 1787) sowie *Maria Stuart* (Schiller, 1800).

Bibliographie

Fischer, Frank, Ingo Börner, Mathias Göbel, Angelika Hecht, Christopher Kittel, Carsten Milling, und Peer Trilcke. 2019. „Programmable Corpora – Die digitale Literaturwissenschaft zwischen Forschung und Infrastruktur am Beispiel von DraCor“. In *DHd 2019 Digital Humanities: multimedial & multimodal. Konferenzabstracts*, 194–97. <https://doi.org/10.5281/zenodo.2596095>.

Fischer, Frank, Peer Trilcke, Christopher Kittel, Carsten Milling, und Daniil Skorinkin. 2018. „To Catch a Protagonist: Quantitative Dominance Relations in German-Language Drama (1730–1930)“. In *DH 2018. Book of Abstracts*, 193–201. Mexiko City.

Krautter, Benjamin, Janis Pagel, Nils Reiter, und Marcus Willand. 2018. „Titelhelden und Protagonisten – interpretierbare Figurenklassifikation in deutschsprachigen Dramen“. *LitLab Pamphlets* 7 (Novem-

ber). https://www.digitalhumanitiescooperation.de/wp-content/uploads/2018/12/p07_krautter_et_al.pdf.

Marcus, Solomon. 1973. Mathematische Poetik. Linguistische Forschungen 13. Frankfurt a.M.: Athenäum.

Schöbner, Franziska. 2017. *Einführung in die Dramenanalyse*. 2. Aufl. J. B. Metzler. <https://link.springer.com/book/10.1007/978-3-476-05285-8>.

Trilcke, Peer, Frank Fischer, Mathias Göbel, Dario Kampkaspar, und Christopher Kittel. 2017. „Netzwerkdynamik, Plotanalyse – Zur Visualisierung und Berechnung der ›progressiven Strukturierung‹ literarischer Texte“. In *4. Jahrestagung des DHd*, 175–80. Bern: Zenodo. <https://doi.org/10.5281/ZENODO.4622799>.

Trilcke, Peer, Christopher Kittel, Nils Reiter, Daria Maximova, und Frank Fischer. 2020. „Opening the Stage – A Quantitative Look at Stage Directions in German Drama“. In *Digital Humanities 2020 - Book of Abstracts*, 422–25. Ottawa, Kanada.