

germanistische Linguistik 46(2), 300-308. DOI: 10.1515/zgl-2018-0017.

Lemnitzer, Lothar / Zinsmeister, Heike (2010): *Korpuslinguistik*. Tübingen: Gunter Narr Verlag.

Schiller, Anne / Teufel, Simone / Stöckert, Christine / Thielen, Christine (1999): *Guidelines für das Tagging deutscher Textcorpora mit STTS* (Kleines und großes Tagset) <http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf> [zuletzt abgerufen 13. Oktober 2018].

Telljohann, Heike / Versley, Yannick / Beck, Kathrin / Hinrichs, Erhard / Zastrow, Thomas (2013): *STTS als Part-of-Speech-Tagset in Tübinger Baumbanken* in: **Zinsmeister, Heike / Heid, Ulrich / Beck, Kathrin (eds.):** *Das Stuttgart-Tübingen Wortarten-Tagset – Stand und Perspektiven*. Journal for Language Technology and Computational Linguistics 28, 1/2013. 1-15.

Wandruszka, Mario (1979): *Die Mehrsprachigkeit des Menschen*. München: Piper.

Klassifikation von Titelfiguren in deutschsprachigen Dramen und Evaluation am Beispiel von Lessings „Emilia Galotti“

Krautter, Benjamin

Benjamin.Krautter@ilw.uni-stuttgart.de
Universität Stuttgart, Deutschland

Pagel, Janis

janis.pagel@ims.uni-stuttgart.de
Universität Stuttgart, Deutschland

Einführung:

In seiner Studie zu Gotthold Ephraim Lessings bürgerlichem Trauerspiel *Emilia Galotti* (1772) formuliert Gisbert Ter-Nedden: „Der erste Akt war der Selbstdarstellung des Prinzen und der höfischen Partei gewidmet; der zweite gehört den Galottis, und zwar zunächst Odoardo als dem Gegenspieler des Prinzen: nach dem Machthaber betritt der Tugendheld die Bühne“ (Ter-Nedden 1986: 189). Die Titelfigur Emilia geht in seiner Darstellung im Kollektiv der Galottis unter. Prinz und Vater Odoardo werden hingegen als die beiden Gegenspieler hervorgehoben. Sind sie also gemäß Ter-Nedden Protagonist und Antagonist, die Helden und Hauptfiguren des Stücks?

Zur Einteilung und Abstufung des Dramenpersonals mithilfe quantitativ erfassbarer Kriterien führt Manfred Pfister in den späten 1970er Jahren die Terminologie der „quantitative[n] Dominanzrelationen“ (Pfister 2001 [1977]: 226) ein. Er benennt hierfür zwei Kriterien, die Haupt- von Nebenfiguren

unterscheiden sollen: die „Dauer der Bühnenpräsenz einer Figur“ (ebd.) und den Anteil der Figurenrede am Haupttext. Laut Pfister fehle es allerdings an einer differenzierten Handlungsgrammatik, die auch funktionale Relationen – etwa die aktiven Handlungsschritte von Figuren – operationalisieren könne (vgl. ebd.: 227). Er wirbt letztlich für einen nicht weiter explizierten multidimensionalen Ansatz, der die quantitative Einteilung des Personals zuverlässiger und feingliedriger machen soll.

An diese Idee der quantitativen und zugleich multidimensionalen Einteilung dramatischer Figuren versuchen wir im Folgenden mittels digitaler Analysetechniken anzuschließen (vgl. Fischer u.a. 2018). Dazu fassen wir das Problem der Figureneinteilung als Klassifikationsaufgabe. Zielsetzung ist es, titelgebende Dramenfiguren mit maschinellen Lernverfahren automatisch auszuzeichnen. Dadurch sind wir in der Lage, die genauen Einflussfaktoren zu prüfen und die Ergebnisse transparent zu evaluieren. Nach unserem Dafürhalten sind es zumindest drei Gründe, die titelgebende Dramenfiguren zu einer geeigneten Zielkategorie der Klassifikation machen. Den möglichen alternativen Konzepten mangelt es erstens an einer konsensfähigen Definition und Differenzierung. Das gilt insbesondere für die Begriffe ‚HeldIn‘ und ‚ProtagonistIn‘, die teilweise synonym verwendet werden (vgl. etwa Plett 2002: 21f., Jannidis 2004: 90, 104f.). Überdies ist gerade das Heldenkonzept stark von literaturgeschichtlichen Entwicklungen geprägt und somit historisch variabel (vgl. etwa Alt 1994: 167f., Platz-Waury 2007 [1997]: 591, Martus 2011: 15). Die intersubjektive Annotation der Figurenkategorien bereitet zweitens Schwierigkeiten, vor allem dann, will man ProtagonistInnen oder HeldInnen mit Blick auf ihre Bedeutung für die Handlung bzw. den zentralen Konflikt des Dramas bestimmen.¹ Zieht man dagegen die gegebenen Titelfiguren der Texte heran, entfällt dieses Annotationsproblem. Drittens lässt sich *a priori* annehmen, dass titelgebende Figuren Eigenschaften verkörpern und im Text repräsentieren, die sie in vielen Fällen auch zu Hauptfiguren der Handlung machen.²

Stand der Forschung:

In der aktuellen Forschung folgen mehrere Studien der von Wladimir Propp (1986 [1928]) am russischen Volksmärchen eingeführten Figurentypologisierung, um literarische Figuren auf formaler oder automatischer Basis (sub-)klassifizieren zu können (etwa Declerck / Koleva / Krieger 2012 oder Finlayson 2017). Moretti (2011 und 2013) nutzt indessen Netzwerkdarstellungen von Shakespeares *Hamlet* für eine Neukontextualisierung der dramatischen Figureninteraktion. Mit seiner Formalisierung argumentiert er gegen die Dichotomie zwischen ProtagonistInnen und Nebenfiguren und für netzwerkanalytische Zentralitätsmaße.³ Frank Fischer u.a. (2018) konzentrieren sich auf die dramengeschichtliche Einordnung von Protagonisten in deutschsprachigen Stücken (vgl. auch Algee-Hewitt 2017). Sie nutzen einen multidimensionalen Ansatz, der Netzwerkmetriken mit zählbasierten Maßen kombiniert. Im Gegensatz dazu versuchen Jannidis u.a. (2016) Hauptfiguren in deutschsprachigen Romanen automatisch zu klassifizieren. Als Goldstandard verwenden sie annotierte Zusammenfassungen der Romane.

Versuchsaufbau:

Das untersuchte Korpus umfasst 38 Dramen mit mindestens einer Titelfigur, deren Veröffentlichung sich von der Mitte des 18. bis ins frühe 20. Jahrhundert erstreckt.⁴ *Tabelle 1* gibt einen Überblick über das Verhältnis titelgebender und nicht-titelgebender Figuren. Demnach werden lediglich drei Prozent aller Figuren durch die Titel der Stücke repräsentiert. Bei der Auswahl haben wir darauf geachtet, sowohl ein breites Spektrum literarischer Epochen als auch verschiedene dramatische Strukturen zu berücksichtigen. August Wilhelm Schlegels *Ion* kommt etwa mit sechs Figuren aus, Grabbes *Napoleon oder Die hundert Tage* sieht 183 Figuren vor. Schillers *Die Verschwörung des Fiesko zu Genua* umfasst insgesamt 75 Szenen, Hofmannsthals *Elektra* lediglich fünf.

#Dramen	#Titelfigur (%)	#Nicht-Titelfigur (%)	#Figuren Total
38	42 (3)	1166 (97)	1208

Tabelle 1. Zahl der Titelfiguren und nicht-Titelfiguren.

Für das maschinelle Klassifikationsverfahren nutzen wir mit *Random Forest* (Ho 1995, Breiman 2001) einen Algorithmus, der Entscheidungsbäume zu einem Ensemble fügt und die Parameter mittels mathematischer Regression berechnet.⁵ Da die einzelnen erlernten Entscheidungsbäume eingesehen werden können, ist *Random Forest* auch dafür geeignet, die entscheidungstragenden Merkmale näher zu untersuchen. Vorausgehende Experimente mit *Support Vector Machines* (SVM) erzielten zudem schlechtere Ergebnisse. *Random Forest* scheint besser geeignet zu sein, um viele und potentiell korrelierende Features zu verarbeiten.

Um dem Vorsatz eines multidimensionalen Modells gerecht zu werden, kombinieren wir als Features zählbasierte Metriken (*Tokens*: Anzahl der gesprochenen Tokens normalisiert nach Textlänge), Netzwerkrelationen (*Degree*, *weighted Degree*, *betweenness Centrality*, *closeness Centrality*, *eigenvector Centrality*), die Bühnenpräsenz der Figuren (*Active*: Figur spricht innerhalb einer Szene; *Passive*: Figur wird innerhalb einer Szene namentlich genannt), *Topic Modeling* und Metadaten (etwa Zahl der Dramenfiguren, Epochenzugehörigkeit).⁶

	Precision (TF)	Recall (TF)	F1 (TF)	Precision (C)	Recall (C)	F1 (C)	MCC
Majority BL	-	0	-	0.97	1.00	0.98	0.00
Tokens BL	0.34	1	0.51	1.00	0.93	0.96	0.56
alle Features	0.45	1	0.62	1.00	0.96	0.98	0.66

Tabelle 2. Klassifikationsergebnisse von vier Modellen für Titelfiguren (TF) und nicht-Titelfiguren (C): Majority Baseline, Tokens Baseline und alle Features.

Diskussion der Klassifikationsergebnisse:

Die *F1-Scores* der Klassifikation in *Tabelle 2* zeigen, dass das gelernte Modell (alle Features) in der Lage ist, sinnvolle Vorhersagen zu treffen, die über die Heuristiken der

Majority und Tokens Baselines hinausgehen. Die verwendeten Features scheinen also tatsächlich Annäherungen an die Eigenschaften titelgebender Figuren wiederzugeben. Die Ergebnisse verdeutlichen aber auch, dass das Modell übergeneralisiert. Zwar wird jede titelgebende Figur als solche erkannt (*Recall* [alle Features] 1.00), das System tendiert jedoch dazu, auch nicht-titelgebende Figuren als Titelfigur zu identifizieren (*Precision* [alle Features] 0.45). Das deutet darauf hin, dass die exponierte Stellung der Figuren im Titel des Dramas zwar durchaus im Text Realisierung findet, andere (Haupt-)Figuren die spezifischen Eigenschaften aber dennoch teilen können. Nur in elf der 38 Dramen werden alle Figuren richtig ausgezeichnet.⁷ *Abbildung 1* und *Abbildung 2* zeigen die *Feature Importance* und die Verteilung der Features. Im Entscheidungsprozess des Algorithmus sind die *Tokens* das wichtigste Feature. Die Leistungsfähigkeit der Tokens Baseline ließ das bereits erahnen. Aber auch einige *Topics*, Netzwerkmaße und die Bühnenpräsenz tragen zu den Klassifikationsergebnissen bei. Das erlernte *Topic Model* erreicht beispielsweise auch ohne die Zugabe weiterer Features Klassifikationsergebnisse, die im Bereich der Tokens Baseline liegen (vgl. Reiter u.a. 2018: 3). Um einen besseren direkten Vergleich der Modelle zu ermöglichen, geben wir auch Werte für den *Matthews Correlation Coefficient* (MCC) an. Der MCC ist weder auf eine einzelne Klasse fokussiert noch wird er von starken Klassenunterschieden beeinflusst. Auch hier zeigt sich, dass das Modell mit allen Features den beiden Baselines überlegen ist.

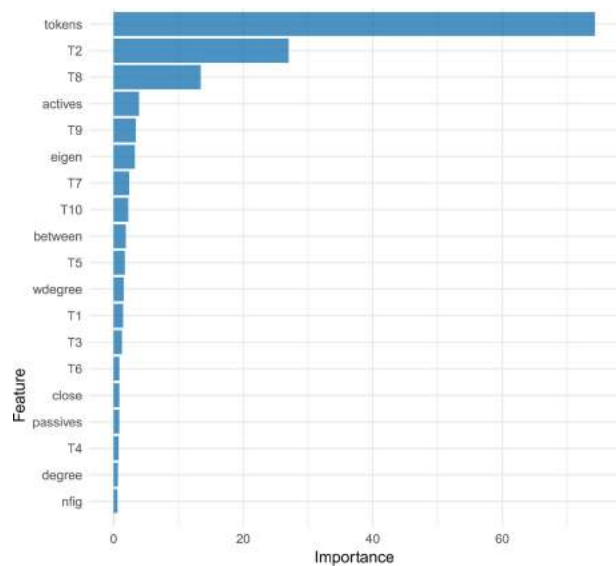


Abbildung 1. Die *Feature Importance* vergleicht die Performanz des Modells, wenn eines der Features entfällt. Die Abnahme an Performanz entspricht der relativen Wichtigkeit des jeweiligen Features für die Klassifikation (Importance). T1-T10 stehen für die zehn erlernten *Topics*.

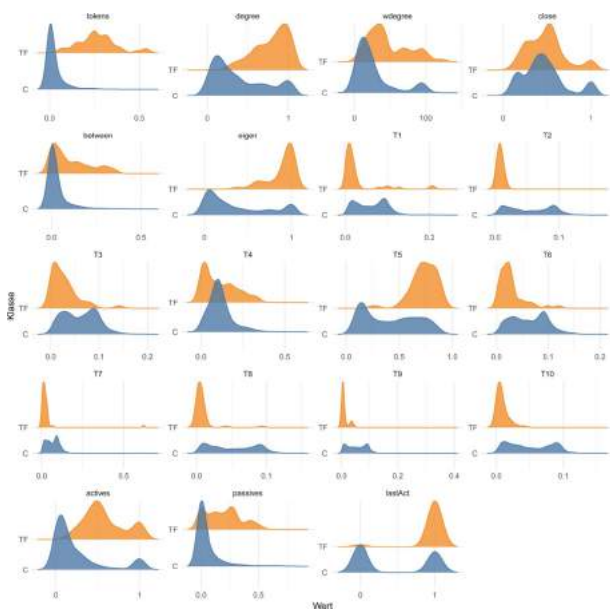


Abbildung 2. Featureverteilung der Klassifikation: Die Skalen auf der x-Achse geben den Wertebereich des jeweiligen Features wieder, die y-Achse zeigt die Verteilung der Feature-Werte auf die jeweiligen Klassen (TF: Titelfigur, C: nicht-Titelfigur).

Emilia Galotti als passiv präsente Titelfigur:

Anhand Lessings *Emilia Galotti* versuchen wir, die aufgeführten Klassifikationsergebnisse genauer nachzuvollziehen. Wir streben hierbei keine Interpretation des Stücks an, sondern wollen auf Grundlage unserer quantitativen Analysen erste Beobachtungen schildern, die sich hauptsächlich auf die Struktur des Dramas beziehen. Zwar beinhaltet die Klassifikation auch *Topic Modeling* Ergebnisse, die erlernten *Topics* sind aber nur begrenzt interpretierbar: sie geben kaum Rückschlüsse auf die Semantiken der Figurenreden. Die Featureverteilung in *Abbildung 2* verdeutlicht, dass lediglich *Topic 5* ein positives Unterscheidungsmerkmal für die Auszeichnung als Titelfigur ist. Unter den zwanzig wahrscheinlichsten Begriffen von *Topic 5* finden sich jedoch ausnahmslos Funktionswörter.⁸ Die Begriffe sind zudem oftmals Teil weiterer *Topics*, die dann zwar auch semantische Ausdrücke beinhalten, sich aber auf Anredeformeln, Angaben zur sozialen Stellung der Figur oder Berufsbezeichnungen beschränken.⁹

Gleich vier Titelfiguren benennt die Klassifikation für Lessings bürgerliches Trauerspiel *Emilia Galotti*. Neben der tatsächlich titelgebenden Emilia werden ihr Vater Odoardo Galotti, der Prinz Hettore Gonzaga und dessen Kammerherr Marinelli als Titelfiguren ausgezeichnet. Betrachtet man ausschließlich die beiden von Pfister angeführten Kriterien quantitativer Dominanzrelationen, würde man Emilia wohl kaum als zentrale Figur des Dramas wahrnehmen. Mit nur 2363 Tokens spricht Emilia nicht nur weniger als der Prinz und Marinelli, die mit jeweils knapp über 5500 Tokens die umfassendsten Redeanteile am Haupttext haben, sie spricht auch weniger als ihr Vater Odoardo und die Gräfin Orsina (

Abbildung 3). Des Weiteren ist sie lediglich in sieben der 43 Dramenszenen aktiv präsent (*Abbildung 4*).

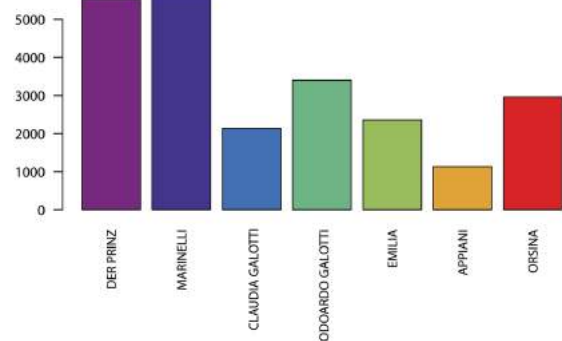


Abbildung 3. Redeanteile in Lessings *Emilia Galotti* gemessen in Tokens. Aufgeführt sind die sieben Figuren mit dem größten Redeanteil am Haupttext.

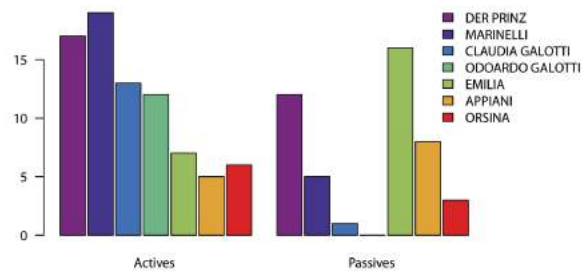


Abbildung 4. Aktive und passive Präsenz in *Emilia Galotti* gemessen an der Zahl der Szenen. Passiv präsent ist eine Figur nur dann, wenn sie in dieser Szene nicht selbst aktiv ist.

Warum also wird Emilia trotzdem als Titelfigur erkannt? Die Featureanalyse in *Abbildung 5* zeigt, dass neben den *Tokens* vor allem die *Topics 8* und *2* sowie die *betweenness Centrality* den größten Einfluss auf die Klassifikation haben. Dass die Zahl der gesprochenen Wörter positiv bewertet wird überrascht, fällt sie doch so deutlich geringer aus als bei Marinelli oder dem Prinzen. Diese Figuren sprechen zwar mehr als Emilia, verglichen mit anderen Dramen aber wohl nicht übermäßig viel mehr. Der hier als gering wahrgenommene Umfang der Figurenrede reicht somit für den Klassifikator aus, um Emilia zumindest in die Riege potentieller Titelfiguren zu heben – auch, da in Lessings Drama insgesamt recht wenige Figuren auftreten.

Für den Leser ist es dagegen wohl eher Emilias durchgängige passive Präsenz in den Dialogen und Monologen anderer Figuren, die sie als Titelfigur kennzeichnet. Exemplarisch dafür steht bereits der erste Akt. Ausgelöst durch eine Bittschrift verliert sich der Prinz in unruhigen Gedanken an Emilia Galotti: „Ich kann doch nicht mehr arbeiten. – Ich war so ruhig, bild' ich mir ein, so ruhig – Auf einmal muß eine arme Bruneschi, Emilia heißen: – weg ist meine Ruhe, und alles! –“ (Lessing 2000 [1772]: 293). Mit wechselnden Gesprächspartnern – Maler Conti, Marinelli und Camillo Rota – wird Emilia immer wieder zum Mittelpunkt der folgenden Dialoge. Das gilt insbesondere für die sechste Szene, als

Marinelli die anstehende Vermählung Emilias mit dem Grafen Appiani preisgibt, woraufhin der eifersüchtige Prinz seinem Kammerherren Marinelli die völlige Handlungsfreiheit in dieser Angelegenheit zugesteht (vgl. ebd., 300–305). Diese passive Präsenz Emilias lässt sich über weite Teile des Dramas nachvollziehen. In 16 Szenen wird in der Figurenrede mit ihrem Namen auf sie referiert, obwohl sie selbst zu diesem Zeitpunkt nicht aktiv am Bühnengeschehen beteiligt ist. Verglichen mit anderen Dramenfiguren ist dieser Wert sehr hoch. Marinellis Name wird beispielsweise nur in fünf Szenen aufgerufen, auf Emilias Mutter Claudia wird ohne ihre Anwesenheit auf der Bühne gar nicht namentlich referiert, wie *Abbildung 4* veranschaulicht.

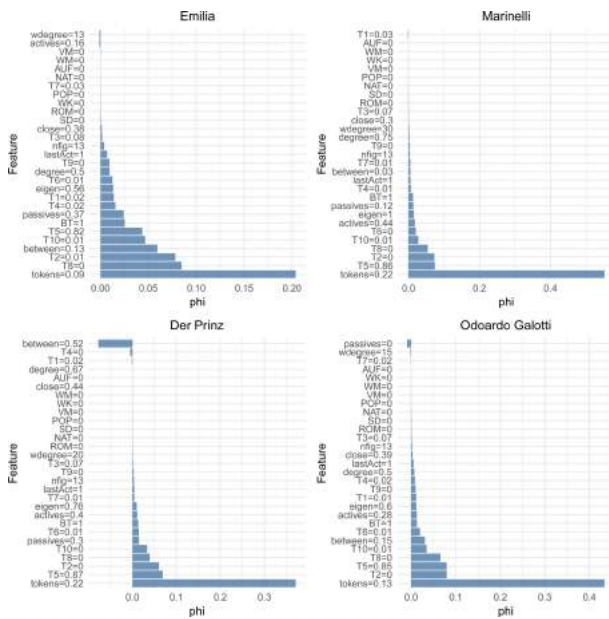


Abbildung 5. Feature Importance der vier als titelgebend klassifizierten Figuren in *Emilia Galotti*.

Fazit und Ausblick:

Wir konnten zeigen, dass unser multidimensionales Modell sinnvolle Ergebnisse für die Klassifikation titelgebender Figuren liefert (MCC 0.66). Titelfiguren werden sehr zuverlässig erkannt (*Recall* 1.00), das Modell neigt jedoch zur Übergeneralisierung. Die Beobachtungen anhand Lessings *Emilia Galotti* lassen aber annehmen, dass die Übergeneralisierung genau die Figuren umfasst, die man als handlungstragende Hauptfiguren des Stücks bezeichnen könnte (vgl. Fick 2000: 337.) Diesen Einzelbefund wollen wir in Zukunft durch die Auswertung und Analyse zusätzlicher Dramen systematisieren. Eine weitere Aufgabe wird es sein, das Erlernen der *Topics* besser nachzuvollziehen und auf die semantische Interpretierbarkeit hin zu optimieren, ohne dabei die Leistungsfähigkeit für die Klassifikation einzuschränken.

Anhang der untersuchten Dramen:

Arnim, L. A. von: Marino Caboga
 Brentano, C.: Ponce de Leon
 Büchner, G.: Dantons Tod
 Büchner, G.: Leonce und Lena
 Büchner, G.: Woyzeck
 Goethe, J. W.: Götz von Berlichingen mit der eisernen Hand
 Goethe, J. W.: Iphigenie auf Tauris
 Goethe, J. W.: Torquato Tasso
 Gottsched, J. Ch.: Der sterbende Cato
 Grabbe, Ch. D.: Don Juan und Faust
 Grabbe, Ch. D.: Hannibal
 Grabbe, Ch. D.: Herzog Theodor von Gothland
 Grabbe, Ch. D.: Napoleon oder Die hundert Tage
 Gutzkow, K.: Richard Savage, Sohn einer Mutter
 Gutzkow, K.: Uriel Acosta
 Hofmannsthal, H. von: Elektra
 Hofmannsthal, H. von: Ödipus und die Sphinx
 Kotzebue, A. von: Die beiden Klingsberg
 Laube, H.: Monaldeschi
 Laube, H.: Struensee
 Lessing, G. E.: Emilia Galotti
 Lessing, G. E.: Miss Sara Sampson
 Pfeil, J. G. B.: Lucie Woodvil
 Romantik Iffland, A. W.: Figaro in Deutschland
 Schiller, F.: Die Jungfrau von Orléans
 Schiller, F.: Die Piccolomini
 Schiller, F.: Die Verschwörung des Fiesco zu Genua
 Schiller, F.: Maria Stuart
 Schiller, F.: Wallensteins Tod
 Schiller, F.: Wilhelm Tell
 Schlaf, J.: Meister Oelze
 Schlegel, A. W.: Alarkos
 Schlegel, A. W.: Ion
 Schlegel, J. E.: Canut
 Schnitzler, A.: Anatol
 Schnitzler, A.: Professor Bernhardt
 Tieck, L.: Der gestiefelte Kater
 Tieck, L.: Prinz Zerbino
 Tieck, L.: Ritter Blaubart
 Uhland, L.: Ludwig der Bayer
 Wieland, Ch. M.: Klementina von Porretta
 Wieland, Ch. M.: Lady Johanna Gray

Fußnoten

1. Bei einem entsprechenden Versuch erzielten wir Übereinstimmungen zwischen 0.43 und 0.83 (Cohen's κ).
2. Titelfiguren könnten dann ersatzweise als (Pseudo-)Goldstandard für die Klassifikation von Hauptfiguren genutzt werden, um zu testen, ob Eigenschaften der Titelfiguren auf andere Hauptfiguren übertragbar sind.
3. Probleme von Morettis Vorgehen diskutiert Peer Trilcke (2013: 226–232).
4. Wir nutzen dafür das German Drama Corpus: <https://github.com/dracor-org/gerdracor>.
5. Die Implementierung erfolgt über die Pakete randomForest und Caret für R: <https://cran.r-project.org/>

web/packages/randomForest/index.html/ und <https://cran.r-project.org/web/packages/caret/>. Der vollständige Code zu den Experimenten ist zu finden unter <https://quadrama.github.io/blog/2018/12/12/detect-protagonists.de/>.

6. Eine genauere Beschreibung aller Features findet sich in Krautter, Benjamin / Pagel, Janis / Reiter, Nils und Willand, Marcus (2018): „Titelhelden und Protagonisten – Interpretierbare Figurenklassifikation in deutschsprachigen Dramen“, in: *Litlab Pamphlet 7*.

7. Vollständig korrekt ausgezeichnet wurden die folgenden Dramen: *Die beiden Klingsberg*, *Marino Caboga*, *Die Piccolomini*, *Wallensteins Tod*, *Die Verschwörung des Fiesco zu Genua*, *Ludwig der Bayer*, *Die Jungfrau von Orleans*, *Herzog Theodor von Gothland*, *Dantons Tod*, *Hannibal*, *Napoleon oder Die hundert Tage*.

8. Die 20 wahrscheinlichsten Begriffe von Topic 5 sind: „ich“, „und“, „die“, „nicht“, „der“, „ist“, „zu“, „-“, „das“, „Ich“, „in“, „so“, „den“, „dem“, „es“, „ein“, „mich“, „sie“, „Sie“, „er“.

9. Etwa: „Herr“, „König“, „Königin“, „Vater“, „Narr“, „Graf“, „Majestät“, „Schuldmeister“, „Mutter“, „Tochter“, „Sohn“, „Doktor“.

Bibliographie

Algee-Hewitt, Mark (2018): *Distributed Character: Quantitative Models of the English Stage, 1500-1920*, in: Digital Humanities 2017: Conference Abstracts: 119–121.

Alt, Peter-André (1994): *Tragödie der Aufklärung. Eine Einführung*. Basel / Tübingen: Francke.

Breiman, Leo (2001): *Random Forests*, in: Machine Learning 24(2): 5–32.

Declerck, Thierry / Koleva, Nikolina / Krieger, Hans-Ulrich (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 <http://www.aclweb.org/anthology/W12-1006> [zuletzt abgerufen 13. Oktober 2018].

Fick, Monika (2000): *Emilia Galotti*, in: Lessing Handbuch. Leben-Werk-Wirkung. Stuttgart: Metzler 316–343.

Finlayson, Mark A. (2017): ‚ProppLearner‘: *Deeply Annotating a Corpus of Russian Folktales to Enable the Machine Learning of a Russian Formalist Theory*, in: Digital Scholarship in the Humanities 32(2): 284–300 <https://doi.org/10.1093/lc/fqv067> [letzter Zugriff 13. Oktober 2018].

Fischer, Frank / Trilcke, Peer / Milling, Carsten / Skorinkin, Daniil (2018): *To Catch a Protagonist: Quantitative Dominance Relations in German-Language Drama (1730–1930)*, in: Digital Humanities 2018: Conference Abstracts: 193–201.

Ho, Tin Kam (1995): *Random Decision Forests*, in: Proceedings of the 3rd International Conference on Document Analysis and Recognition: 278–282.

Jannidis, Fotis (2004): *Figur und Person. Beitrag zu einer historischen Narratologie*. Berlin: de Gruyter.

Jannidis, Fotis / Reger, Isabella / Krug, Markus / Weimer, Lukas / Macharowsky, Luisa / Puppe, Frank (2016): *Comparison of Methods for the Identification of Main Characters in German Novels*, in: Digital Humanities 2016: Conference Abstracts: 578–582.

Krautter, Benjamin / Pagel, Janis / Reiter, Nils / Willand, Marcus (2018): *Titelhelden und Protagonisten –*

Interpretierbare Figurenklassifikation in deutschsprachigen Dramen, in: Litlab Pamphlet 7.

Lessing, Gotthold Ephraim (2000 [1772]): *Emilia Galotti. Ein Trauerspiel in fünf Aufzügen*, in: **Bohnen, Klaus (ed.):** *Lessing Werke und Briefe*. Bd. 7. Werke 1770–1773. Frankfurt a.M.: Deutscher Klassiker-Verlag 291–371.

Martus, Steffen (2011): *Transformationen des Heroismus. Zum politischen Wissen der Tragödie im 18. Jahrhundert am Beispiel von J. E. Schlegels Canut*, in: **Burkhard, Torsten / Hundt, Markus / Martus, Steffen / Ort, Claus-Michael (eds.):** *Politik – Ethik – Poetik. Diskurse und Medien frühnezeitlichen Wissens*. Berlin: Akademie Verlag 15–42.

Moretti, Franco (2011): *Network Theory, Plot Analysis*, in: Literary Lab Pamphlet 2. <https://litlab.stanford.edu/LiteraryLabPamphlet2.pdf> [zuletzt abgerufen 13. Oktober 2018].

Moretti, Franco (2013): ‚Operationalizing‘: *or, the Function of Measurement in Modern Literary Theory*, in: Literary Lab Pamphlet 6. <https://litlab.stanford.edu/LiteraryLabPamphlet6.pdf> [zuletzt abgerufen 13. Oktober 2018].

Pfister, Manfred (2001 [1977]): *Das Drama. Theorie und Analyse*. München: Fink.

Platz-Waury, Elke (2007 [1997]): *Art. Figurenkonstellation*, in: **Weimar, Klaus / Fricke, Harald / Grubmüller, Klaus / Müller, Jan-Dirk (eds.):** *Reallexikon der deutschen Literaturwissenschaft*. Bd. 1. Berlin / New York: De Gruyter 591–593.

Plett, Bettina (2002): *Problematische Naturen? Held und Heroismus im realistischen Erzählen*. München / Paderborn / Wien / Zürich: Schöningh.

Reiter, Nils / Krautter, Benjamin / Pagel, Janis / Willand, Marcus (2018): *Detecting Protagonists in German Plays around 1800 as a Classification Task*, in: EADH 2018: Conference Abstracts. <http://dx.doi.org/10.18419/opus-10162> [zuletzt abgerufen 22.12.2018].

Ter-Nedden, Gisbert (1986): *Lessings Trauerspiele. Der Ursprung des modernen Dramas aus dem Geist der Kritik*. Stuttgart: Metzler.

Trilcke, Peer (2013): *Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft*, in: **Ajouri, Philip / Mellmann, Katja / Rauen, Christoph (eds.):** *Empirie in der Literaturwissenschaft*. Münster: Mentis 201–247.

Korpuserstellung als literaturwissenschaftliche Aufgabe

Gius, Evelyn

evelyn.gius@uni-hamburg.de
Universität Hamburg, Deutschland

Katharina, Krüger

katharina.krueger@uni-hamburg.de
Universität Hamburg, Deutschland