

Ein Schritt zurück: Distinktive Eigenschaften im deutschsprachigen Drama

Krautter, Benjamin

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

Einführung in die Fragestellung

Topic Modeling und soziale Netzwerkanalysen zählen zu den etabliertesten quantitativen Methoden innerhalb der *Computational Literary Studies* (vgl. Du 2019; Trilcke u.a. 2016, Jannidis 2017: 148–161). Beide gelten insbesondere als geeignet, um große literarische Korpora auf Muster – semantischer oder struktureller Art – zu untersuchen, die durch lineares Lesen nicht oder nur schwer zu greifen sind (vgl. Willand 2017: 86). Auf diese Weise versprechen sie literarhistorische Entwicklungslinien aufzuzeigen, die die traditionelle, auf symptomatische Beispiele fußende Literaturgeschichtsschreibung zu übersehen neigt (vgl. Moretti 2017: 6f.; Jockers 2013: 9). Epistemologisch sind die durch *Topics* oder Netzwerkmaße erschlossenen Textmodelle jedoch vom ursprünglichen literarischen Text deutlich zu unterscheiden. So konstatieren Peer Trilcke und Frank Fischer, dass es sich um andere ‚epistemische Dinge‘ handle (vgl. Trilcke, Fischer 2018). Denn beide Methoden reduzieren den literarischen Text auf spezifische Eigenschaften: im Fall der Netzwerkanalyse zumeist auf Figurenbeziehungen, die durch Knoten und Kanten repräsentiert werden; im Fall des *Topic Modeling* auf Wortkollokationen, die als *Topics* interpretiert werden. Positiv ließe sich diese Reduktion als Notwendigkeit der Operationalisierung fassen, die ein – in Abhängigkeit der Fragestellung gewähltes – theoretisches Konzept messbar machen soll (vgl. Moretti 2013). So könnte sich, wie Franco Moretti postuliert, die Handlung eines Textes näherungsweise als die Summe an Interaktionen der Figuren im Netzwerk operationalisieren lassen (vgl. Moretti 2011: 2–4). Die Zentralitätsmaße von Figurennetzwerken ermöglichen dann einen auf quantitativen Werten basierenden Vergleich der literarischen Texte.

Ein mögliches Anwendungsszenario zeigt *Abbildung 1*. Sie stellt das *Average Degree* (durchschnittlicher Grad) von insgesamt 443 Dramennetzwerken dar. *Degree* ist ein simples Zentralitätsmaß, das für jede Dramenfigur bemisst, mit wie vielen anderen sie im Verlauf des Stücks interagiert. Trilcke und Fischer nutzen eine ähnliche Darstellung, um literarhistorische Erkenntnisse zu bestätigen. Basierend auf der Vorstellung, dass *Degree* ein „Indikator für soziale Komplexität“ (Trilcke, Fischer: 2018) sein könnte, formulieren sie die geläufige These, dass

das Drama seit Mitte des 18. Jahrhunderts gesellschaftliche Modernisierungsprozesse widerspiegeln würde.

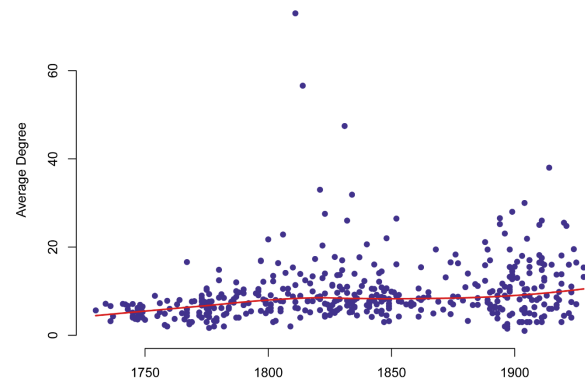


Abbildung 1: Average Degree im historischen Verlauf; LOESS Kurve.

Ziel dieses Beitrags ist es jedoch, einen Schritt hinter solche makroanalytischen Befunde zurück zu treten. In einem vorgelagerten Arbeitsschritt möchte ich erörtern, welche quantitativ erfassbaren Merkmale dramatischer Texte überhaupt geeignet sind, um eine literarhistorische Einordnung und Unterscheidung der Dramen vorzunehmen. Anders formuliert sollen also die Kriterien ermittelt werden, die mit Blick auf die Entstehungszeit der Dramen unterscheidungstragend sind. Zu diesem Zweck dient im vorliegenden Fall eine einfach gehaltene Klassifikationsaufgabe. Ließen sich die dramatischen Texte erfolgreich ihrem Veröffentlichungszeitraum zuweisen, könnte daraus auf die Kriterien rückgeschlossen werden, die den entscheidenden Beitrag zu dieser Klassifikation leisten. Daran anschließend wäre eine Rückübersetzung der ermittelten Merkmale denkbar – analog zur Operationalisierung –, die eine Interpretation anleiten könnten.

Korpus

Die folgenden Untersuchungen konzentrieren sich auf 443 deutschsprachige Dramen zwischen 1730 und 1930 aus dem *German Drama Corpus* (Fischer u.a. 2019). *Abbildung 2* gibt einen Überblick über die zeitliche Verteilung der Dramen. Es handelt sich um ein recht heterogenes Korpus, das auf unterschiedlichen poetologischen Vorstellungen fußt, die wiederum an verschiedene Produktions- und Rezeptionsbedingungen geknüpft sein können. Sowohl versifizierte als auch in Prosa gehaltene Dramen sind enthalten. Sehr kurze Stücke, in denen die Figurenrede weniger als 3000 Tokens umfasst, wurden für die Analysen entfernt, so dass die Länge der Stücke zwischen noch immer kurzen 3017 und längst nicht mehr ungekürzt auf der Bühne darstellbaren 146248 Tokens liegt (median: 22548; Standardabweichung:

13304). Die Zahl der auftretenden Figuren liegt zwischen zwei und 183 (median: 17; Standardabweichung: 19,2). Das Korpus enthält Stücke von insgesamt 166 Autor*Innen, darunter sowohl hochkanonische als auch heutzutage kaum noch wahrgenommene.

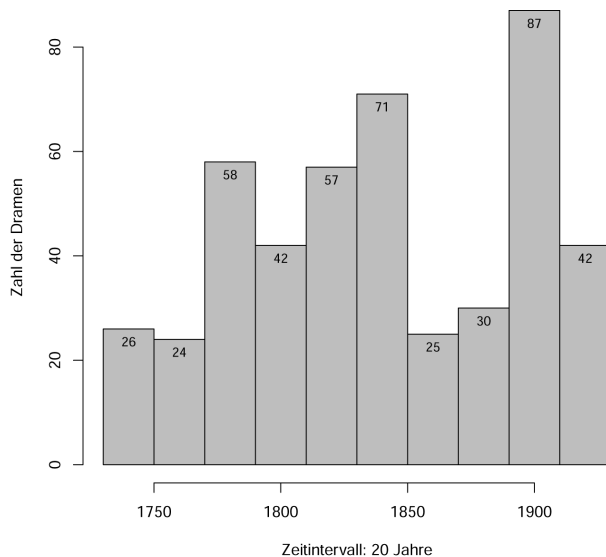


Abbildung 2: Korpusübersicht.

Methode

Die historische Verortung der dramatischen Texte fasse ich als basal gehaltene Klassifikationsaufgabe. Ziel der Klassifikation ist es, mittels maschineller Lernverfahren näherungsweise den Veröffentlichungszeitraum der Dramen zu bestimmen, um daran anschließend die Einflussfaktoren identifizieren und untersuchen zu können. Dazu greife ich auf Metadaten zurück, die Angaben zur Erstaufführung und zur Erstpublikation umfassen. Diese Metadaten werden genutzt, um jedes Drama einer von vier heuristisch gesetzten Zeitspannen zuzuordnen, die jeweils circa 50 Jahre umfassen: 1730–1785 (93 Dramen), 1786–1832 (116 Dramen), 1833–1881 (105 Dramen) und 1882–1930 (129 Dramen). Die dadurch entstehenden Zeiträume dienen als Zielpunkt der Klassifikation und orientieren sich an wichtigen literaturgeschichtlichen Zäsuren: Aufklärung, Goethezeit, Realismus und literarische Moderne (vgl. etwa Brenner 2011).¹ Als Features der Klassifikation nutze ich verschieden komplexe Netzwerk- und Zentralitätsmetriken sowie durch *Topic Modeling* trainierte *Topics*.

Basis der Netzwerk- und Zentralitätsmetriken sind Netzwerkgraphen, die auf Präsenz- bzw. Adjazenzmatrizen fußen. Knoten und Kanten repräsentieren hierbei Dramenfiguren und deren Interaktion, wobei Interaktion als das gemeinsame Sprechen innerhalb einer Szene operationalisiert ist (vgl. Trilcke 2013: 238f.). Eine Kante zwischen zwei Knoten wird also genau dann instanziiert, wenn die beiden fraglichen Figuren innerhalb

derselben Dramenszene sprechen. Das bedeutet auch, dass verschiedene poetologische Vorstellungen von Akt und Aufzug sowie Szene und Auftritt, die im Verlauf der Dramengeschichte einem Wandel unterliegen, einen Eingang in die Graphen findet (vgl. etwa Vogel 2012). Für die Klassifikation nutze ich die folgenden acht Maße: *Degree*, *Weighted Degree*, *Closeness Centrality*, *Betweenness Centrality*, *Eigenvector Centrality*, *Average Path Length*, *Clustering Coefficient* und *Density*.²

Topic Modeling gilt als Technik, die – in einem weiteren Sinn gefasst – semantische Strukturen in größeren Textkorpora zu identifizieren vermag (vgl. etwa Schöch 2017: 42). Die von mir verwendeten *Topics* wurden auf dem gesamten in *Abbildung 2* dargestellten Dramenkorpus trainiert, wobei die Figurenrede eines jeden Dramas nochmals in Segmente von je 1000 Tokens unterteilt ist. Vorab wurde das Wortmaterial auf Nomen, Adjektive und Vollverben beschränkt. Um die *Topics* zu trainieren, greife ich auf das von David M. Blei, Andrew Y. Ng und Michael I. Jordan (2003) vorgeschlagene probabilistische Modell *Latent Dirichlet Allocation* zurück. Für die maschinelle Klassifikation setze ich ein Modell mit 20 *Topics* (T1–T20) ein, das einen guten Kompromiss bietet zwischen der Interpretierbarkeit einzelner *Topics* und ihrer Eignung, die Texte diachron zu unterscheiden.

Das maschinelle Klassifikationsverfahren selbst nutzt den Algorithmus *Random Forest* (Ho 1995, Breiman 2001). *Random Forest* fügt mehrere unkorrelierte Entscheidungsbäume zusammen und berechnet mittels mathematischer Regression die Parameter. Beim Trainieren wurde das Korpus in zehn Segmente gegliedert (*10-fold cross validation*), wodurch verzerrte Ergebnisse durch die zufällige Verteilung von Trainings- und Testkorpus vermieden werden sollen.³

Ergebnisse

Tabelle 1 zeigt die Ergebnisse der Klassifikation anhand von drei Modellen. Zusätzlich zum Gesamtmodell, das Netzwerkanalysen und *Topic Modeling* zusammenführt, wurden die acht Netzwerkmetriken und die 20 *Topics* auch isoliert betrachtet. Die *Baseline* berechnet sich, angelehnt an die einleitenden Überlegungen, anhand des *Average Degree*.⁴ Die Werte in *Tabelle 1* verdeutlichen zweierlei: Einerseits erreicht bereits die *Baseline* annehmbare Ergebnisse. 309 Dramen werden anhand ihres *Average Degree* richtig klassifiziert. Andererseits scheint insbesondere das trainierte *Topic Model* zur Leistungsfähigkeit des gesamten Modells beizutragen. Letzteres zeigt sich mit einem F_1 -Wert von 0.921 angemessen performant – lediglich 34 Dramen werden einem falschen Zeitraum zugewiesen. Wirklich überraschen kann dieser Umstand jedoch nicht, sind die heuristischen Klassifikationszeiträume doch recht groß gewählt. Kleiner gefasste Zeiträume führen das Modell hingegen recht schnell an seine Grenzen. Teilt man

die durch das Textkorpus abgedeckte Dramengeschichte in feingliedrigere Segmente, beispielsweise in zehn Zeiträume von nurmehr 20 Jahren, sinkt der F_1 -Wert auf 0.431 (Precision: 0.585, Recall: 0.451).

Tabelle 1: Modelle trainiert auf 443 Dramen; 10-fold cross validation, SMOTE-sampling.

	Precision	Recall	F_1
Baseline (Avg. Degree)	0.701	0.702	0.698
Netzwerkmetriken	0.829	0.821	0.814
Topic Model	0.899	0.903	0.899
Gesamtes Modell	0.921	0.925	0.921

Die in *Tabelle 1* dargestellten Klassifikationsergebnisse erlauben nun einen Einblick in die Unterscheidungskraft der eingespeisten Features. *Abbildung 3* zeigt die sogenannte *Feature Importance*. Sie vergleicht nach und nach die Leistungsfähigkeit des Modells, wenn jeweils eines der Features nicht beachtet wird. Die Abnahme an Performanz entspricht dann der relativen Wichtigkeit des nicht einbezogenen Features für die Klassifikation. Die Abbildung verdeutlicht, dass das *Topic Model* – insbesondere die *Topics* 8 und 5 – einen starken Einfluss auf die Klassifikation nimmt. Doch bei weitem nicht jedes *Topic* (etwa T17, T7, T10) ist von solch großer Bedeutung. Als gewichtigste Netzwerkmetrik lässt sich die *Betweenness Centrality* identifizieren. Der durchschnittliche Grad (*Average Degree*) ist mittig platziert, wobei der Einfluss der weniger entscheidungstragenden Features insgesamt ähnlich gering ausfällt.

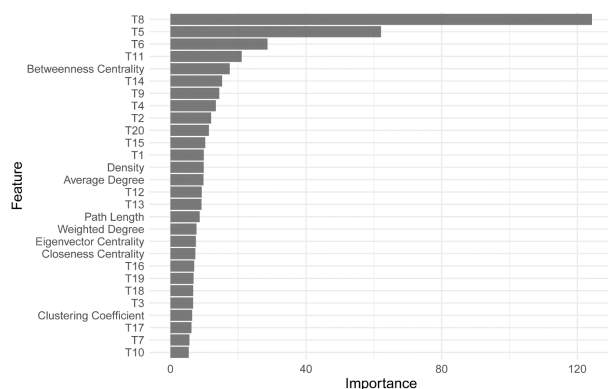


Abbildung 3: Feature Importance des Klassifikationsmodells.

Operationalisierung und Interpretation

Auf Basis dieser Daten lassen sich nun analog zu *Abbildung 1* Werte berechnen und visualisieren, die die zeitliche Entwicklung der als relevant erscheinenden Merkmale darstellen, etwa von *Topic 8* oder der *Betweenness Centrality*. Diese sollten – vertraut man der Klassifikation – einen höheren Aussagegehalt haben, als der zu Beginn diskutierte *Average Degree*.

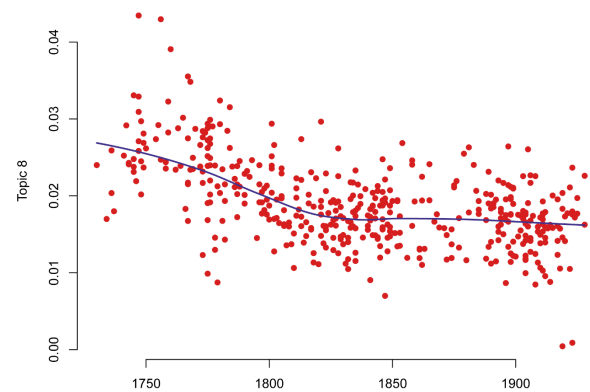


Abbildung 4: Topic 8 im historischen Verlauf, normalisiert nach Dramenlänge; LOESS Kurve.



Abbildung 5: Betweenness Centrality im historischen Verlauf; LOESS Kurve.

Abbildung 4 zeigt die Frequenzen, mit denen sich die Wörter aus *Topic 8* auf die einzelnen Dramen verteilen. Tatsächlich veranschaulicht die Visualisierung eine recht deutliche Entwicklung. Von 1730 ausgehend scheint *Topic 8* bis etwa 1830 recht stark an Einfluss einzubüßen, ehe die Werte fortan auf einem stabilen Niveau bleiben. Betrachtet man die zehn ausschlaggebendsten Wörter des *Topics* – ‚liebe‘, ‚herz‘, ‚machen‘, ‚lassen‘, ‚sagen‘, ‚vater‘, ‚schwester‘, ‚sehen‘ und ‚weiß‘ –, lässt sich dieser

Verlauf auch literaturgeschichtlich plausibilisieren. Zu einem großen Teil können diese Begriffe mit bürgerlichen Trauerspielen und Rührstücken in Verbindung gesetzt werden, die für die Dramengeschichte des 18. Jahrhunderts prägend sind.

Auch der in *Abbildung 5* dargestellte diachrone Verlauf der *Betweenness Centrality* macht eine Entwicklung der Werte sichtbar. Die lokal gewichtete Regression veranschaulicht einen Höhepunkt zwischen 1810 und 1825. Die *Betweenness Centrality* bemisst, in welchem Maß ein Knoten im Netzwerk selbst zum Teil eines Pfades wird, also die indirekte Verbindung von zwei anderen Knoten ermöglicht (vgl. Newman 2010: 185–193). Auf Dramennetzwerke übertragen ließen sich dadurch Figuren identifizieren, die als Brückenfiguren agieren und voneinander getrennte Figurengruppen verbinden. Die Darstellung lässt sich somit als weiterer Indikator für die zu Beginn skizzierte These von Trilcke und Fischer lesen. Die zusehende Abkehr von der Regelpoetik in der zweiten Hälfte des 18. Jahrhunderts scheint eine komplexere Struktur der Dramen nach sich zu ziehen, die sich in den Netzwerkdaten wiederfinden lässt.

Die größte Schwierigkeit bei der Interpretation dieser Daten bleibt jedoch nach wie vor bestehen und ist den hier gezeigten Analysen vorgelagert. Es ist die Operationalisierung der Fragestellung, die zumeist mit einem großen konzeptuellen Aufwand verbunden ist (vgl. Gius 2019: 2f.; Reiter / Willand 2018). Denn unklar bleibt, wie sich Zentralitätsmetriken in einem Figurennetzwerk oder Wahrscheinlichkeitsverteilungen von Worthäufigkeiten zu literaturwissenschaftlichen Kategorien verhalten. Die bemessenen Werte müssten sich konzeptionell so rückübersetzen lassen, dass sie auch mit Blick auf spezifisch literaturwissenschaftliche Fragestellungen interpretiert werden können. Dass etwa die Handlung literarischer Texte nicht einfach durch ein Figurennetzwerk abzubilden ist, muss auch Moretti erkennen, weshalb er die Netzwerktheorie letztlich nurmehr als Vorstufe, als „beginning of the beginning“ (Moretti 2011: 2) zu einer quantifizierbaren Handlung einordnet.

Fazit und Ausblick

Das vorgestellte multidimensionale Modell liefert sinnvolle Ergebnisse und kann den recht weit gefassten Veröffentlichungszeitraum deutschsprachiger Dramen mit angemessener Genauigkeit (F_1 -Wert 0.921) klassifizieren. Da die Entstehung neuer literarischer Epochen und Strömungen zumeist als fließender Prozess zu beschreiben ist, muss die hier vorgestellte Methode aber als Heuristik eingestuft werden, die vor allem zum Ziel hat, die entscheidungstragenden Merkmale der Klassifikation zu identifizieren. Erst dadurch bietet die Klassifikationsaufgabe Anschlusspotential für literarhistorische Studien. Für künftige Arbeiten erscheint es einerseits lohnend, eine metrische Vorhersage der

Veröffentlichungsjahre zu erproben. Dadurch würde die Vorhersage deutlich an Präzision gewinnen. Andererseits würde es sich anbieten, neben *Topic Modeling* und Netzwerkanalysen auch stilometrische Maße in die Klassifikation zu integrieren. So könnte auch der Stil der Stücke – in einem quantitativen und damit weiten Sinn – Teil der Voraussage werden.

Fußnoten

1. Da literarische Epochen und Strömungen eine fließende Entwicklung nehmen, ergeben sich durch diese Setzung zwangsläufig Überschneidungen. So fallen beispielsweise Lessings *Emilia Galotti* (1772) und Schillers *Die Räuber* (1781), zwei schon strukturell sehr verschieden gebaute Stücke unterschiedlicher Strömungen, in die gleiche Klasse.
2. Für eine Übersicht über die verschiedenen Metriken vgl. Newman (2010): 168–204.
3. 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/>. *Caret* bietet vielfältige Optionen für das *Preprocessing* und *Sampling* der Daten. Ich nutze die Methoden *center* und *scale* zur Kalibrierung und *SMOTE-sampling*, um die zahlenmäßige Ungleichverteilung der Dramen in den verschiedenen Zeiträumen auszugleichen (eine genauere Beschreibung eines ähnlichen Versuchsaufbaus findet sich in Krautter / Pagel / Reiter / Willand 2018: 19–29).
4. Die Leistungsfähigkeit einer *Majority Baseline* wäre aufgrund des *multiclass* Klassifikation sehr eingeschränkt (Precision: 0.291).

Bibliographie

- Blei, David M. / Ng, Andrew Y. / Jordan, Michael I.** (2003). „Latent Dirichlet Allocation“, in: *The Journal of machine Learning research* 3: 993–1022.
- Breiman, Leo** (2001): „Random Forests“, in: *Machine Learning* 24 (2): 5–32.
- Brenner, Peter J.** (2011): *Neue deutsche Literaturgeschichte. Von ‚Ackermann‘ zu Günter Grass*. Berlin / New York: De Gruyter.
- Du, Keli** (2019): „A Survey on LDA Topic Modeling in Digital Humanities“, in: *DH 2019. Conference Abstracts*. <https://dev.clariah.nl/files/dh2019/boa/0326.html> [letzter Zugriff 05. Januar 2019].
- Fischer, Frank / Börner, Ingo / Göbel, Mathias / Hecht, Angelika / Kittel, Christopher / Milling, Carsten / Trilcke, Peer** (2019): „Programmable Corpora. Die digitale Literaturwissenschaft zwischen Forschung und Infrastruktur am Beispiel von DraCor“, in: *DHD 2019. Konferenzabstracts*: 194–197. DOI: 10.5281/zenodo.2596094 [letzter Zugriff 05. Januar 2019].

Gius, Evelyn (2019): „Computationelle Textanalysen als fünfdimensionales Problem: Ein Modell zur Beschreibung von Komplexität“, in: *Litlab Pamphlet* 8. <https://www.digitalhumanitiescooperation.de/pamphlet-8-computationelle-textanalysen/> [letzter Zugriff 05. Januar 2019].

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

Jannidis, Fotis (2017): „Netzwerke“, in: Jannidis, Fotis / Kohle, Hubertus / Rehbein, Malte (eds.): *Digital Humanities: Eine Einführung*. Stuttgart: J.B. Metzler 148–161.

Jockers, Matthew L. (2013): *Macroanalysis. Digital Methods and Literary History*. Urbana / Chicago / Springfield: University of Illinois Press.

Krautter, Benjamin / Pagel, Janis / Reiter, Nils / Willand, Marcus (2018): „Titelhelden und Protagonisten – Interpretierbare Figurenklassifikation in deutschsprachigen Dramen“, in: *Litlab Pamphlet* 7. https://www.digitalhumanitiescooperation.de/wp-content/uploads/2019/06/p07_krautter_et_al-1.pdf [letzter Zugriff 05. Januar 2019].

Moretti, Franco (2011): „Network Theory, Plot Analysis“, in: *Literary Lab Pamphlet* 2. <https://litlab.stanford.edu/LiteraryLabPamphlet2.pdf> [letzter Zugriff 05. Januar 2019].

Moretti, Franco (2013): „Operationalizing: or, the Function of Measurement in Modern Literary Theory“, in: *Literary Lab Pamphlet* 6. <https://litlab.stanford.edu/LiteraryLabPamphlet6.pdf> [letzter Zugriff 27. September 2019].

Moretti, Franco (2017): „Patterns and Interpretation“, in: *Literary Lab Pamphlet* 15. <https://litlab.stanford.edu/LiteraryLabPamphlet15.pdf> [letzter Zugriff 05. Januar 2019].

Newman, Mark E. J. (2010): *Networks: An Introduction*. Oxford: UP.

Reiter, Nils / Willand, Marcus (2018): „Poetologischer Anspruch und dramatische Wirklichkeit: Indirekte Operationalisierung in der digitalen Dramenanalyse. Shakespeares natürliche Figuren im deutschen Drama des 18. Jahrhunderts, in: Bernhart, Toni / Willand, Marcus / Richter, Sandra / Albrecht, Andrea (eds.): *Quantitative Ansätze in den Literatur- und Geisteswissenschaften: Systematische und historische Perspektiven*. Berlin / Boston: De Gruyter 45–75.

Schöch, Christof (2017): „Gattungen des Kriminalromans: Ein quantitativer, Topic-basierter Zugang“, in: Koch, Corinna / Schmitz, Sabine / Lang, Sandra (eds.): *Dialogische Krimianalysen. Fachdidaktik und Fachwissenschaft untersuchen aktuelle Repräsentationsformen des französischen Krimis*. Frankfurt a. M.: Peter Lang 37–64.

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.

Trilcke, Peer / Fischer, Frank (2018): „Literaturwissenschaft als Hackathon. Zur Praxeologie der Digital Literary Studies und ihren epistemischen Dingen“, in: Huber, Martin / Krämer, Sybille (eds.): Sonderband 3 der ZfdG: Wie Digitalität die Geisteswissenschaften verändert: Neue Forschungsgegenstände und Methoden. http://www.zfdg.de/sb003_003 [letzter Zugriff 05. Januar 2019].

Trilcke, Peer / Fischer, Frank / Göbel, Matthias / Kampkasper, Dario (2016): „Theatre Plays as ‚Small Worlds‘? Network Data on the History and Typology of German Drama, 1730–1930“, in: *DH 2016. Conference Abstracts*: 385–387.

Vogel, Juliane (2012): „Aus dem Takt: Auftrittsstrukturen in Schillers *Don Karlos*“, in: *Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte* 86 (4): 532–546.

Willand, Marcus (2017): „Hermeneutische Interpretation und digitale Analyse: Versuch einer Verhältnisbestimmung“, in: Banki, Luisa / Scheffel, Michael (eds.): *Lektüren. Positionen zeitgenössischer Philologie*. Trier: WVT 77–98.