

Artikel aus:  
Zeitschrift für digitale Geisteswissenschaften

Titel:  
»[E]in Vater, dächte ich, ist doch immer ein Vater«. Figurentypen im Drama und ihre Operationalisierung

---

Autor/in:  
Benjamin Krautter

Kontakt: [Benjamin.Krautter@gs.uni-heidelberg.de](mailto:Benjamin.Krautter@gs.uni-heidelberg.de)  
Institution: Universität Heidelberg, Germanistisches Seminar  
GND: [1201330157](#) ORCID: [0000-0003-1804-7520](#)

---

Autor/in:  
Janis Pagel

Kontakt: [janis.pagel@ims.uni-stuttgart.de](mailto:janis.pagel@ims.uni-stuttgart.de)  
Institution: Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung  
GND: [1221232525](#) ORCID: [0000-0003-4370-1483](#)

---

Autor/in:  
Nils Reiter

Kontakt: [nils.reiter@ims.uni-stuttgart.de](mailto:nils.reiter@ims.uni-stuttgart.de)  
Institution: Universität zu Köln, Institut für Digital Humanities  
GND: [1020288329](#) ORCID: [0000-0003-3193-6170](#)

---

Autor/in:  
Marcus Willand

Kontakt: [marcus.willand@gs.uni-heidelberg.de](mailto:marcus.willand@gs.uni-heidelberg.de)  
Institution: Universität Heidelberg, Germanistisches Seminar  
GND: [1050009495](#) ORCID: [0000-0002-3107-311X](#)

---

DOI des Artikels:  
[10.17175/2020\\_007](https://doi.org/10.17175/2020_007)

Nachweis im OPAC der Herzog August Bibliothek:  
[1726147096](#)

Erstveröffentlichung:  
31.12.2020

Lizenz:

Sofern nicht anders angegeben 

Medienlizenzen:  
Medienrechte liegen bei den Autoren

Letzte Überprüfung aller Verweise:  
01.12.2020

GND-Verschlagwortung:  
[Drama](#) | [Literarische Gestalt](#) | [Literaturwissenschaft](#) | [Maschinelles Lernen](#) | [Operationalisierung](#) |

Zitierweise:  
Benjamin Krautter, Janis Pagel, Nils Reiter, Marcus Willand: »[E]in Vater, dächte ich, ist doch immer ein Vater«. Figurentypen im Drama und ihre Operationalisierung. In: Zeitschrift für digitale Geisteswissenschaften. Wolfenbüttel 2020. PDF Format ohne Paginierung. Als text/html abrufbar unter DOI: [10.17175/2020\\_007](https://doi.org/10.17175/2020_007).

Benjamin Krautter, Janis Pagel, Nils Reiter, Marcus Willand

»[E]in Vater, dächte ich, ist doch immer ein Vater«. Figurentypen im Drama und ihre Operationalisierung

## Abstracts

Dieser Artikel behandelt die Operationalisierung von Figurentypen im deutschsprachigen Drama. Ausgehend von der dramen- und theatergeschichtlichen Forschung werden Figuren bestimmt, die einem der drei Figurentypen ›Intrigant\*in‹, ›tugendhafte Tochter‹ und ›zärtlicher Vater‹ entsprechen. Für die Figuren wurden eigenschaftsbasierte Datensätze erstellt, die zu ihrer automatischen Klassifikation herangezogen werden. Neben die inhaltliche Komplexität von Figuren und die theoretische Bestimmung von Figurentypen tritt die methodische Herausforderung, ausgehend von einer kleinen Menge an Annotationen zu generalisieren. Unsere Experimente zeigen, dass sich die gewählten Typen jeweils innerhalb einer Grundgesamtheit als abgrenzbare Teilmengen herausbilden.

This article deals with the operationalization of character types in German-language drama. Based on literary and theatre history, we identify characters that belong to one of the three types: ›schemer‹, ›virtuous daughter‹ and ›tender father‹. For the characters in our corpus, property-based data sets were established, which are then used in turn for their automatic classification. In addition to the complexity of characters and the theoretical determination of character types, we discuss the methodological challenge of generalizing from a small set of annotations. Our experiments show that the selected types emerge as definable subsets within a population.

## 1. Einleitung

Mit einer Bitte wendet sich William Sampson am Beginn des dritten Akts von Gotthold Ephraim Lessings bürgerlichem Trauerspiel *Miß Sara Sampson* (1755)<sup>1</sup> an seinen Bediensteten Waitwell: Er möge Williams Tochter und Titelheldin Sara eine Mitteilung überbringen. Es sei »der Brief eines zärtlichen Vaters, der sich über nichts, als ihre Abwesenheit beklaget.«<sup>2</sup> Diese der Figurenrede Sir Williams zu entnehmende direkte Selbstcharakterisierung<sup>3</sup> wird in der Folgeszene nicht nur aufgegriffen, sondern dient als Anknüpfungspunkt für die weitere Charakterisierung sowohl der Vater- als auch der Tochterfigur. Bei der Briefübergabe wiederholt Diener Waitwell die auffällige Formulierung und Selbstzuschreibung Sir Williams als ›zärtlicher Vater‹<sup>4</sup> auch gegenüber Sara. Er überträgt dabei die Eigenschaft der Vaterfigur auf die Tochter: Er »ist noch immer der zärtliche Vater, so wie sein Sarchen noch immer die zärtliche Tochter ist.«<sup>5</sup> Der aktualisierende Wortlaut »noch immer« deutet über die gegenwärtige Wahrnehmung Saras zum Zeitpunkt der Briefszene hinaus, verdeutlicht er für Sara doch, dass Vater William ihr verziehen wird beziehungsweise schon verziehen hat. Somit suggeriert die Wendung ein überzeitlich stabiles Verhältnis von Vater und Tochter: »Die Bindung« bleibe, so Wilfried Barner, »auch dann erhalten, nachdem Sara die Vater-Tochter-Beziehung gestört hat.«<sup>6</sup> Denn Sara entscheidet sich für ihre Liebe zu Mellefont und »gegen die väterliche Autorität.«<sup>7</sup> Dieser »Fehltritt«, den sie angesichts der väterlichen Vergebung später als »Irrtum«<sup>9</sup> bezeichnen wird, bleibt also vermeintlich folgenlos.<sup>10</sup> Es scheint fast so, als könnten Vater und Tochter aufgrund ihrer als *natürlich* wahrgenommenen »Gefühle[] und Empfindungen« niemals dauerhaft »vom Pfad der ›Tugend‹«<sup>11</sup> abkommen. Den Inhalt des Briefes antizipierend, wendet Sara das Versöhnungsangebot ihres Vaters aber gegen sich selbst. Dass der Vater ihr längst vergeben hat, ist Anlass für sie, ihren Schuldgefühlen stärker Ausdruck zu verleihen:

»Er ist noch der zärtliche Vater? So liebt er mich ja noch? So muß er mich ja beklagen? Nein, nein, das tut er nicht; das kann er nicht tun! Siehst du denn nicht, wie unendlich jeder Seufzer, den er um mich verlöre, meine Verbrechen vergrößern würde? Müßte mir nicht die Gerechtigkeit des Himmels jede seiner Tränen, die ich ihm auspreßte, so anrechnen, als ob ich bei jeder

<sup>1</sup> Während die Erstveröffentlichung den Untertitel einem bürgerlichen Trauerspiel, in fünf Aufzügen trägt, verzichtet der Neudruck von 1772 auf den Zusatz *bürgerlich*. Zum Begriff *bürgerlich* und seiner Entwicklung vgl. kursorisch Eibl 1971, S. 140–142 und Vollhardt 2018, S. 117–120. Zum bürgerlichen Trauerspiel und seiner Entstehung vgl. Michelsen 1981, passim und Mönch 1993, passim. Mönch untersucht die Gattungen des bürgerlichen Trauerspiels anhand von 242 Primärtexten, um anders als »die Mehrheit der literaturwissenschaftlichen Arbeiten zum bürgerlichen Trauerspiel« nicht nur den »etablierten Kanon« zu betrachten. Mönch 1993, S. 2.

<sup>2</sup> Lessing 2003, S. 467.

<sup>3</sup> Pfister 2001, S. 251 spricht mit Blick auf die dramatische Informationsvergabe vom *Eigenkommentar* als explizit-figurale Charakterisierungstechnik.

<sup>4</sup> Zum zärtlichen Liebeskonzept im bürgerlichen Trauerspiel vgl. Pikulik 1981, S. 19–36.

<sup>5</sup> Lessing 2003, S. 471.

<sup>6</sup> Barner 1998, S. 175. Siehe dazu auch Ter-Nedden 2016, S. 139: Der Effekt der von Sir William »nachdrücklich betonten Bedingungs- und Grenzenlosigkeit der Vergebungsbereitschaft liegt darin, dass sie an die Bedingung der Gegenliebe gebunden ist«.

<sup>7</sup> Stockinger 2001, S. 206.

<sup>8</sup> Lessing 2003, S. 442.

<sup>9</sup> Lessing 2003, S. 508.

<sup>10</sup> Dazu konstatiert Eibl 1971, S. 157: »Ihr ›Verbrechen‹ ist zwar vom Vater verziehen worden, aber es bleibt der ›Irrtum‹, es bleibt die Sündenfolge des ›peccatum originans‹, das ›Wissen um Gut und Böse‹ und die Fallibilität des ›Herzens‹«.

<sup>11</sup> Fick 2016, S. 133.

derselben mein Laster und meinen Undank wiederholte? Ich erstarre über diesen Gedanken. Tränen koste ich ihm? Tränen? Und es sind andre Tränen, als Tränen der Freude? – Widerspruch mir doch, Waitwell!«<sup>12</sup>

Der Ausschnitt aus ihrer Figurenrede veranschaulicht, warum Sara von der Forschung als »Typ des gefühlvollen Menschen« betrachtet wird, dessen Moral vor allem dem Herzen folge.<sup>13</sup> »Zärtlichkeit« und »Herz«, das seien entsprechend auch die maßgeblichen Begriffe, die die Beziehung von Vater und Tochter bestimmten und auf denen die Dramenhandlung gründe.<sup>14</sup> Hieran anknüpfend wird der dramatische Konflikt des Stückes »auf einen Fehler des »zärtlichen Herzens«, also der Liebesfähigkeit der Titelheldin«, zurückgeführt.<sup>15</sup> Noch in der Sterbeszene Saras wird die Charakterisierung William Sampsons als zärtlicher Vater ein letztes Mal wiederholt. Hier ist es seine dem Tode nahe Tochter, die erst sich selbst als »eine schuldige, eine reuende, eine gestrafte Tochter« und in Kontrast dazu ihren Vater als »beleidigt[]«, »großmüthig[]« und »zärtlich[]«<sup>16</sup> bezeichnet. Lessings Figuren, so ließe sich also folgern, betreiben durch ihre sprachliche Selbst- und Fremdcharakterisierung eine Form der innerfiktionalen Typenbildung, die sich in den poetologischen und theaterpraktischen Bezeichnungen der Rollenfächer wiederfindet.

Als »zärtlicher Vater« steht William Sampson exemplarisch für die zunehmende Abstufung und Unterscheidung einzelner Rollenfächer – hier der Vaterfigur –, die mit dem bürgerlichen Theater einhergeht.<sup>17</sup> Das Rollenfach gilt als »zentrale[s] Strukturprinzip der europäischen Theaterpraxis«. <sup>18</sup> Dabei organisiert es einerseits die Rollenverteilung: Dramenfiguren werden einem Rollenfach zugeordnet, das von bestimmten Schauspieler\*innen bedient wird. Dadurch garantiert es andererseits einen »konkreten künstlerischen Wirkungsradius des jeweiligen Darstellers«. <sup>19</sup> Das Rollenfach zielt also auf theaterpraktische Überlegungen ab, basiert dabei aber auf literarischen Typen, die Autor\*innen in ihren dramatischen Texten anlegen. <sup>20</sup> Mit Manfred Pfister lässt sich ein Figurentypus als eine Gruppe von Figuren bestimmen, die einen gewissen »Satz von Eigenschaften« teilt und somit eine Abstraktion des Individuellen darstellt.<sup>21</sup>

Die skizzierte Verschränkung von Rollenfächern und Figurentypen machen wir uns im Folgenden zunutze, wenn wir versuchen, Figurentypen mithilfe von computergestützten Analysen zu identifizieren und deren Eigenschaften näher zu ergründen. Neben dem von Sir William verkörperten Typus des »zärtlichen Vaters« rücken wir zwei weitere Figurentypen in den Fokus – nämlich die »tugendhafte Tochter« und den/die »Intrigant\*in«. <sup>22</sup> Während sich der »zärtliche Vater« und die »tugendhafte Tochter«, wie wir es anhand von Lessings Stück umrissen haben, hauptsächlich durch die Vater-Tochter-Relation definieren, sind »Intrigant\*innen« durch ihr ambivalentes Changieren zwischen der Rolle des handelnden Akteurs und der des planvollen Beobachters bestimmt. <sup>23</sup> Schon hieran wird deutlich, dass die Zuweisung von Figur zu Figurentypus ein komplexes mehrschichtiges Problem darstellt.

Die Analyse dramatischer Figurentypen kann auf Grundlagen aufbauen, die im Bereich der *Computational Literary Studies* seit geraumer Zeit diskutiert werden, prominent etwa am Beispiel der Erkennung von Protagonisten in literarischen Texten. <sup>24</sup> Bereits in den 1970er Jahren hat Manfred Pfister unter dem Begriff »[q]uantitative Dominanzrelationen« <sup>25</sup> eine erste Operationalisierung für Dramen vorgeschlagen, die Haupt- von Nebenfiguren unterscheiden sollte. Seine beiden Kriterien, die »Dauer der Bühnenpräsenz einer Figur« sowie ihr »Anteil[] am Haupttext«, <sup>26</sup> müssten jedoch eher als zu kontextualisierende Indizien denn vollkommen zuverlässige Kennzeichen bewertet werden. Für eine präzisere Einteilung des Bühnenpersonals, etwa nach der Handlungsfunktion der Figuren, mangle es – so Pfister – zudem an »differenzierteren handlungsgrammatischen Vorarbeiten«. <sup>27</sup>

In unserem Beitrag werden wir die bestehenden computergestützten Ansätze erweitern und damit die von Pfister geforderte Differenzierung von Bühnenfiguren mittels quantitativer Analysen anstreben. Ziel ist es, automatisiert Figuren zu erkennen, die einem bestimmten Figurentypus zugehörig sind. Mithilfe von Methoden der maschinellen Sprachverarbeitung sollen also

<sup>12</sup> Lessing 2003, S. 471f.

<sup>13</sup> Barner 1998, S. 171.

<sup>14</sup> Schönenborn 2004, S. 120.

<sup>15</sup> Vollhardt 2018, S. 122. Vollhardt weist jedoch darauf hin, dass auch die Vergebung William Sampsons nicht frei von Eigeninteressen sei, betrachtet er Sara doch als seine Stütze im Alter. Vgl. Vollhardt 2018, S. 127f. und Lessing 2003, S. 434.

<sup>16</sup> Lessing 2003, S. 521.

<sup>17</sup> Vgl. Schonlau 2017, S. 45.

<sup>18</sup> Detken / Schonlau 2014, S. 7.

<sup>19</sup> Maurer-Schmoock 1982, S. 158f.

<sup>20</sup> Vgl. Diebold 1978, S. 9f.

<sup>21</sup> Pfister 2001, S. 245.

<sup>22</sup> Begründet ist diese Auswahl durch unser Annotationsverfahren, das sich auf Forschungsliteratur zu den gewählten Figurentypen stützt. Vgl. exemplarisch das Kapitel »Der zärtliche Vater und seine tugendhafte Tochter« in Fischer-Lichte 1999, S. 300–309.

<sup>23</sup> Vgl. Alt 2004, S. 2.

<sup>24</sup> Vgl. Moretti 2011, passim; Fischer et al. 2018, passim; Jannidis et al. 2016, passim; Reiter et al. 2018, passim; Krautter et al. 2018, passim; Krautter et al. 2019, passim.

<sup>25</sup> Pfister 2001, S. 226.

<sup>26</sup> Pfister 2001, S. 226.

<sup>27</sup> Pfister 2001, S. 227.

literaturwissenschaftlich aufschlussreiche Befunde über die drei Figurentypen und deren konstitutive Textmerkmale gewonnen werden. Dabei verschränken wir quantitative Methoden der digitalen Dramenanalyse mit qualitativen, indem maschinelle Lernverfahren und dramenhistorische Erkenntnisse zusammengeführt werden. Neben der Klassifikation und Analyse einzelner Figuren und Figurentypen ermöglichen wir durch diesen skalierten Blick der Korpusanalyse auch die zukünftige Analyse umgreifender Entwicklungslinien: Wie verändert sich das dramatische Personal hinsichtlich der dargestellten Figurentypen? Welche Gattungs- und Epochenunterschiede lassen sich beobachten?

Als Datengrundlage verwenden wir deutschsprachige Dramen, die im **GerDraCor-Korpus**<sup>28</sup> maschinenlesbar aufbereitet wurden, wobei wir uns auf den Zeitraum zwischen 1730 und 1850 konzentrieren. Dass damit nur eine relativ kurze Zeitspanne abgedeckt wird, ist zwei Gründen geschuldet: Zum einen stand das Drama von etwa 1850 bis 1880 deutlich im Schatten der erfolgreichen Prosagattungen des Realismus.<sup>29</sup> Zum anderen veränderte sich das Rollenfach und dessen Einfluss auf die Theaterpraxis im Laufe des 19. Jahrhunderts nachhaltig: Neue, am modernen Individuum orientierte Rollen entstanden und die Grenzen der einzelnen Fächer verschwammen zusehends. Zudem stieg die Zahl der Schauspieler\*innen eines Ensembles mit dem weiter anwachsenden Repertoire der Theater.<sup>30</sup>

Um die oben skizzierten Fragen behandeln zu können, werden wir in einem ersten Schritt einflussreiche theaterpraktische Kontexte für das deutschsprachige Drama des 18. und 19. Jahrhunderts ausführen (Abschnitt 2). Im Zentrum stehen hier die sogenannten Rollenfelder, die einen konzeptionellen Rahmen sowohl für die theoretische Bestimmung als auch für die Auswahl der Figurentypen bilden. Daran anschließend explizieren wir unsere methodischen Überlegungen, die ausgehend von Franco Morettis Konzept des *Distant Reading* in einer mehrstufigen Operationalisierung der Figurentypen münden (Abschnitt 3).<sup>31</sup> Es handelt sich hierbei um eine Operationalisierung für Menschen, die anhand von Annotationsbeispielen veranschaulicht wird. Die quantitative Vermessung der verschiedenen Figurentypen erfordert eine zweite, andere Art der Operationalisierung: Die Operationalisierung für den Computer, die die Eigenschaften literarischer Figuren durch Methoden der quantitativen Textanalyse in messbare Indikatoren überträgt (Abschnitt 4).<sup>32</sup> Diese dienen als *Features* für die computergestützten Auswertungen. In einer Reihe von Experimenten versuchen wir daraufhin, die drei genannten Figurentypen näher zu betrachten und die annotierten Figuren automatisch zuzuweisen. Die Auswertung und Gegenüberstellung einzelner Features und Figuren verspricht aufschlussreiche Einblicke in die Konzeption dramatischer Figuren (Abschnitt 5).

## 2. Theaterpraktische Kontexte: Lessing und das Rollenfach

Früh entwickelte der junge Student Gotthold Ephraim Lessing – durch seinen Freund Christlob Mylius und seinen Schriftstellerkollegen Christian Felix Weiße in die Welt des Theaters eingeführt – ein ausgeprägtes Interesse für die praktische Seite der Bühnenstücke. Er besuchte Proben, stand in engem Kontakt mit der Schauspieltruppe von Friederike Caroline Neuber und übersetzte sogar fremdsprachige Stücke für ihre Inszenierungen.<sup>33</sup> In den späteren 1760er Jahren diskutierte und reflektierte Lessing als festangestellter Kritiker des Hamburger Nationaltheaters die aktuelle Theaterpraxis in wöchentlichen Kritiken, die schließlich – aufgrund von finanzieller Probleme und mangelndem Publikumsinteresse am regelmäßigen Turnus – unter gleichem Namen als *Hamburgische Dramaturgie* (1767–1769) in einem Sammelband erschienen sind.<sup>34</sup> In einem Brief an seinen Bruder Karl vom 26. September 1776 bat Lessing, inzwischen mit dem Nationaltheater in Mannheim in Kontakt stehend, um Hilfe bei der Suche nach Schauspieler\*innen. Der Brief erwähnt explizit die Rollen des Vaters, der Mutter, des männlichen Liebhabers, der weiblichen Liebhaberin, der Dienerin und eines Mädchens, die es zu besetzen gelte.<sup>35</sup> Die angeführten Beispiele, so unspektakulär sie aus heutiger Sicht wirken, zeichnen Lessing als Kenner der zeitgenössischen Theaterpraxis aus. Als Dramatiker habe er Stücke für das ihm bekannte Theater geschrieben und dabei Rollen für die ihm bekannten Schauspieler konzipiert.<sup>36</sup> Dementsprechend stellt Edward P. Harris fest, dass »Lessings fundamentales Verständnis der Besetzungspraxis sowie individueller und ensemble-spezifischer Fähigkeiten [...] ein Schlüsselement seiner Dramaturgie«<sup>37</sup> seien.

---

<sup>28</sup> Vgl. dazu Fischer et al. 2019, passim.

<sup>29</sup> Vgl. Begemann 2007, S. 10.

<sup>30</sup> Vgl. dazu Doerry 1926, S. 33–38.

<sup>31</sup> Operationalisierung verstehen wir als das Messbarmachen theoretischer Konzepte. Vgl. dazu die Pionierarbeit des Physikers Bridgman 1927, passim sowie Chang 2019, passim. Für den Zusammenhang von Operationalisierung und algorithmischer Textanalyse vgl. Pichler / Reiter 2020, passim.

<sup>32</sup> Der von Moretti pointierte Dreischritt von der Theorie durch Quantifizierung zu (einer möglichst großen Zahl von) Texten fällt abhängig von der Forschungsfrage also deutlich komplexer aus, als es seine Ausführungen annehmen lassen. Vgl. Moretti 2013, passim.

<sup>33</sup> Vgl. Barner 1980, S. 108f.

<sup>34</sup> Vgl. Barner 1998, S. 185 und Harris 1992, S. 229.

<sup>35</sup> Vgl. Lessing 1988, S. 842.

<sup>36</sup> Vgl. Rilla 1977, S. 270f.

<sup>37</sup> Harris 1992, S. 232.

Als wichtiges Instrument der deutschen Theaterpraxis etablierte sich im 18. Jahrhundert das sogenannte Rollenfach.<sup>38</sup> Bernhard Diebold erläutert in einer noch immer einflussreichen, aber wenig trennscharfen Definition, dass ein Rollenfach aus »einer Gesamtheit von – in gewisser Beziehung – ähnlichen Rollen besteht«. <sup>39</sup> Diese Ähnlichkeit könne insofern literarisch sein, als dass sie auf »den Rollen zugrunde liegenden, vom Dichter geschaffenen Typen beruht und somit Fächer wie diejenigen der Helden, Väter, Intriganten [...] bedingt«. <sup>40</sup> Die Ähnlichkeit der Rollen bezieht sich aber auch auf die Mittel ihrer künstlerischen Darstellung, die den/die individuelle/n Schauspieler\*in auszeichnen. <sup>41</sup> Nicht selten spezialisierten sich Schauspieler\*innen auf einzelne Rollenfächer, die sie immer wieder bedienten. <sup>42</sup> In vielen Fällen decken sich die literarischen Typen – zum Beispiel verschiedene Bedienstete – mit den Rollen des Schauspielers. <sup>43</sup> Rollenfächer scheinen also einerseits mit dem Konzept von Figurentypen vergleichbar zu sein, das schematisch beziehungsweise abstrakt angelegte Figuren zu kategorisieren versucht. <sup>44</sup> Andererseits ist das Rollenfachsystem aber in seinen Grenzen durchaus flexibel und historisch variabel. <sup>45</sup> Das System umfasst neben sozial klar definierten Typen beispielsweise auch die sogenannten Charakterrollen – ein Fach, das speziell auf dynamische, runde, komplexe oder individuelle Figuren zugeschnitten ist und damit Figuren umfasst, die den Grundgedanken des Fächersystems in Teilen unterlaufen. <sup>46</sup> Die einzelnen Fächer sollten daher nicht als allzu streng definierte Kategorien mit starren Grenzen betrachtet werden. <sup>47</sup> Die Spannung zwischen Individualität und typisch Abstraktem, die David Fishelov ganz grundsätzlich in jeder literarischen Figur als gegeben sieht, scheint durchaus auf die verschiedenen Rollenfächer übertragbar zu sein. <sup>48</sup>

Die Ursprünge des Rollenfachs sind in den Improvisationstruppen der Commedia dell'arte zu finden, die dem deutschen Wandertheater im 17. und frühen 18. Jahrhundert als Vorbild dienten. <sup>49</sup> Noch in Lessings *Emilia Galotti* (1772), so konstatiert Wilfried Barner, schimmerten »Techniken der commedia dell'arte [...] überall durch«. <sup>50</sup> Die Umstellung des Repertoires von Improvisationsstücken hin zu klassischen französischen und später zu regulären deutschen Stücken habe dann die Entwicklung des recht festen theaterpraktischen Systems der Rollenfächer nach sich gezogen. Denn eine ortsgebundene Theatergruppe musste ihr Repertoire um ein Vielfaches erweitern, um das Besucherinteresse weiterhin aufrechtzuerhalten. <sup>51</sup> Infolgedessen kristallisierten sich gewisse Konventionen heraus, die die verschiedenen Rollenfächer definierten und differenzierten. Während kleinere Theatertruppen mit lediglich vier Schauspielerinnen und vier Schauspielern zurechtkommen mussten – die sowohl tragische als auch komödiantische Rollen zu spielen hatten –, <sup>52</sup> konnten wirtschaftlich besser gestellte Häuser mit 16 und mehr Schauspieler\*innen und daher auch mit differenzierteren Rollentypen planen. <sup>53</sup>

| Brandes »Promemoria«   | Schauspieler*in        |
|------------------------|------------------------|
| zärtlicher Alter       | Meyer                  |
| komischer Alter        | Beil                   |
| Raisonneur             | Herter                 |
| erster Liebhaber       | Böck                   |
| zweiter Liebhaber      | Zuccarini, Beck        |
| Petitmaitre, Fats usw. | Beck, Beil             |
| erster Bedienter       | Backhaus               |
| zweiter Bedienter      | Beil                   |
| Charakterrolle         | Böck, Brandes, Iffland |
| zärtliche Mutter       | –                      |

<sup>38</sup> Vgl. Maurer-Schmooch 1982, S. 157 und Detken / Schonlau 2014, S. 9.

<sup>39</sup> Diebold 1978, S. 9.

<sup>40</sup> Diebold 1978, S. 10.

<sup>41</sup> Vgl. Diebold 1978, S. 10 und Doerry 1926, S. 2, 5, 9. Doerry 1926, S. 2 spricht etwa von der auf »verschiedenen Zusammensetzung[en] der körperlichen und geistigen Fähigkeiten beruhende[n] Einteilung der schauspielerischen Talente«.

<sup>42</sup> Dazu Maurer-Schmooch 1982, S. 92: »Im Allgemeinen hatten die Schauspieler je ein Fach der Tragödie und der Komödie zu übernehmen«.

<sup>43</sup> Vgl. Jannidis 2007, S. 712.

<sup>44</sup> Vgl. dazu Maurer-Schmooch 1982, S. 159 und Kretz 2012, S. 106–108.

<sup>45</sup> Vgl. Mehlin 1969, S. 351–356 und Detken / Schonlau 2014, S. 19f.

<sup>46</sup> Dazu im Allgemeinen Theater-Lexikon von 1839: Charakter-Rollen »werden in der Theatersprache alle Rollen genannt, welche die Aufgabe haben einen Charakter in seiner vollständigen Entwicklung zu veranschaulichen«, Storch 1839, S. 122. Vgl. zur Unterscheidung von *flat* und *round characters* Fishelov 1990, S. 422–426.

<sup>47</sup> Vgl. Diebold 1978, S. 70.

<sup>48</sup> Fishelov 1990, S. 422.

<sup>49</sup> Vgl. Hinck 1965, S. 74–80 und Winter 2014, S. 33–38.

<sup>50</sup> Barner 2017, S. 138.

<sup>51</sup> Vgl. Harris 1992, S. 222–225.

<sup>52</sup> Vgl. Harris 1992, S. 227.

<sup>53</sup> Vgl. Diebold 1978, S. 57–61 und Barner 1998, S. 82. Maurer-Schmooch gibt die durchschnittliche Schauspieler\*innenzahl im 18. Jahrhundert mit etwa 20 Mitglieder an. Vgl. Maurer-Schmooch 1982, S. 93.

|                                     |                              |
|-------------------------------------|------------------------------|
| komische Mutter                     | Syler, Pöschel               |
| erste Liebhaberin                   | Wallenstein, Pöschel         |
| zweyte Liebhaberin                  | Seyler, Brands               |
| dritte Liebhaberin zu naiven Rollen | Toscani                      |
| erste und zweyte Soubrette          | Toscani, Kummerfeld, Pöschel |

Tab. 1: Brandes' Aufflistung der Rollenfächer. [Harris 1992, S. 231]

Tabelle 1 veranschaulicht das Ensemble des Mannheimer Nationaltheaters von 1778 bis 1780 unter der Leitung von Wolfgang von Dalberg.<sup>54</sup> Die Tabelle kombiniert eine Aufstellung der Schauspieler\*innen des Ensembles mit einer typischen Liste von Rollenfächern, die Johann Christian Brandes bei seiner Bewerbung um die Leitung des Nationaltheaters im Jahr 1779 vorlegte.<sup>55</sup> Sie enthält auch zwei von Lessings Stücken, Emilia Galotti und das Lustspiel Minna von Barnhelm, oder das Soldatenglück (1767), wobei das letztere der Forschung heute als Paradebeispiel für eine gängige Rollenfach-Besetzung gilt.<sup>56</sup>

Lessing, so urteilt Edward P. Harris, folgt den Konventionen des Rollenfach-Systems und akzeptiert dessen Limitierungen, um die Aufführbarkeit seiner Stücke zu gewährleisten. Für Emilia Galotti und Minna von Barnhelm etwa habe er typische deutsche Figuren konzipiert, die das übliche Rollenspektrum eines Theaterensembles abdecken, ohne dadurch aber »künstlerische Ausschmückung, Verfeinerung und Wirksamkeit [...] einzuschränken«.<sup>57</sup>

### 3. Von der Figur zum Figurentypus: Operationalisierung für Menschen

#### 3.1 Methodische Vorüberlegungen

In unserem ersten, noch rein manuellen Schritt der quantitativen Analyse von dramatischen Figuren orientieren wir uns an Franco Morettis ursprünglicher Konzeption von *Distant Reading*,<sup>58</sup> die er erstmals in seinem Essay *Conjectures on World Literature* 2000 konturierte. Moretti bestimmt Distant Reading dort als »second-hand« criticism«, als »patchwork of other people's research, without a single direct textual reading«.<sup>60</sup> In Anbetracht seines ambitionierten Fokus auf Weltliteratur – er nimmt sich hierfür Goethe sowie Marx und Engels zum Vorbild –<sup>61</sup> schlägt er mit einiger Polemik vor, auf die Lektüre von Primärliteratur zugunsten von Sekundärliteratur, also bereits existierender Expertise, die es nur mehr zusammenzutragen gelte, zu verzichten. Er versucht damit eine geeignete Methode für den Umgang mit literaturgeschichtlichen Fragestellungen zu finden, die sich nicht nur auf kanonische Texte beschränken, da diese nach Moretti »not even one per cent of published literature«<sup>62</sup> ausmachen würden.

Obwohl methodisch unterspezifiziert bietet Morettis frühes distant reading-Verständnis für unsere Analyse der Dramenfiguren eine fruchtbare Perspektive. Denn die computationelle Untersuchung und automatische Identifizierung verschiedener Figurentypen erfordert einerseits einen theoretischen Unterbau, der die Typologie plausibilisiert. Andererseits werden aber auch annotierte Daten benötigt, die Eigenschaften und Merkmale der Figuren auszeichnen, um auf Basis derer eine Zuordnung zu einem Figurentypus zu ermöglichen. Die manuelle Klassifizierung literarischer Figuren lässt sich demnach als Operationalisierungsproblem für Menschen bezeichnen.

Sowohl für mögliche Typologien als auch für die zu betrachtenden Eigenschaften der Figuren wurden in der Forschung bereits verschiedene Vorschläge zur Diskussion gestellt. In seiner grundlegenden *Morphologie des Märchens* (1928) schlägt Vladimir Propp eine Klassifikation vor, die auf »seven areas of action«<sup>63</sup> basiert, also die Handlungsmuster der Figuren in den Vordergrund rückt.<sup>64</sup> Edward Morgan Forster unterscheidet 1927 ganz basal zwischen »runden« und »flachen« Figuren:

<sup>54</sup> Vgl. auch Barner 1998, S. 82. Siehe dazu Ifflands Wunsche, an die er eine Anstellung in Mannheim knüpfte: »Am 9. April 1779 teilte Iffland seine Forderungen in acht Punkten mit, vor allem geht es um die Rollenfächer (zärtliche und komische Alte, Liebhaber, Ausschluss von niedrig-komischen Charakteren), das Gehalt (752 Gulden jährlich) und die Erklärung, bis Monatsende kein anderes Engagement anzunehmen«, Košenina 2017, S. 135f.

<sup>55</sup> Vgl. Diebold 1978, S. 58f. und Maurer-Schmooch 1982, S. 163.

<sup>56</sup> Vgl. Schonlau 2010, S. 83.

<sup>57</sup> Vgl. Harris 1992, S. 232, S. 234f. Zitat S. 235.

<sup>58</sup> Inzwischen gilt *Distant Reading* als Inbegriff quantitativer Analysen im Bereich digitaler Geisteswissenschaften. Vgl. Weitin et al. 2017, S. 162.

<sup>59</sup> Moretti 2000b.

<sup>60</sup> Moretti 2000b, S. 57.

<sup>61</sup> Vgl. Moretti 2000b, S. 54.

<sup>62</sup> Moretti 2000b, S. 55. Vgl. dazu auch Moretti 2000a, S. 226.

<sup>63</sup> Jannidis 2014, S. 35.

<sup>64</sup> Vgl. Propp 1972, S. 79–83. Das Handeln von Figuren ist jedoch nur eine Möglichkeit, um zwischen verschiedenen Figurengruppen zu unterscheiden. Weitere, potentiell unterscheidungstragende Eigenschaften könnten beispielsweise das Geschlecht, die Nationalität, der soziale Status oder das Alter der Figuren sein. Vgl. Detken / Schonlau 2014, S. 11 und Schonlau 2010, S. 82.

»Flat characters were called ›humorous‹ in the seventeenth century, and are sometimes called types, and sometimes caricatures. [...] In their purest form, they are constructed round a single idea or quality: when there is more than one factor in them, we get the beginning of the curve towards the round.«<sup>65</sup>

Diese Dichotomie wurde jedoch als stark reduktiv empfunden, da sie die »degrees and nuances found in actual works of narrative fiction«<sup>66</sup> vernachlässige oder sogar tilge. Angesichts dieser Kritik entstanden differenziertere Taxonomien. Baruch Hochman etwa warb 1985 für acht dichotome Kategorien wie »Wholeness vs. Fragmentariness«, »Dynamism vs. Staticism« oder »Stylization vs. Naturalism«.<sup>67</sup> In der Dramengeschichte sind es vor allem drei Konzepte, die zur Kategorisierung von ein- und mehrdimensionalen Figuren herangezogen werden: Personifikation, Typ und Individuum. Während die Personifikation einer »realen Figur gegenüber am abstraktesten ist«,<sup>68</sup> verberge sich »hinter einer als Individuum konzipierten Figur die Intention, das Einmalige und Unwiederholbare hervorzukehren«.<sup>69</sup> Anders als Figurentypen, die vom Individuellen abstrahieren, »um ein überindividuelles Allgemeines repräsentieren zu können«, müsse die individuelle Figur über ihre »soziale, psychologische und ideologische Typik hinaus spezifiziert«<sup>70</sup> werden.

### 3.2 Annotation als Second-Hand Criticism

Als Input für die computergestützten Analysen fertigen wir manuelle Annotationsdaten an, in denen jede Figur unseres Dramenkorpus einer in der Figurentypologie verankerten Figurengruppe zugeordnet wird. Dafür folgen wir Morettis Vorschlag, betreiben also zunächst *Second-Hand Criticism*. Um Eigenschaften und Charakterzüge möglichst vieler dramatischer Figuren bestimmen zu können – das schließt die eingehende Betrachtung jedes einzelnen Dramas im Korpus aus –, haben wir in einer ersten Annotation von der literaturwissenschaftlichen Forschung akkumuliertes, idealiter konsensuales (Handbuch-)Wissen konzentriert. Literatur- und Dramengeschichten, Handbücher, Lexika usw. wurden auf Zuschreibungen an die Figuren ausgewertet, die in den Stücken unseres Analysekorpus auftreten.<sup>71</sup> Ergebnis dieses Distant Reading ist ein Katalog von Charaktereigenschaften und Merkmalen, der mit Informationen zu Geschlecht oder Familienrelationen kombiniert werden kann, die oftmals bereits durch das zu Beginn eines Dramas gelistete Figurenverzeichnis zu erschließen sind.

Anhand von Lessings 1755 uraufgeführter Miß Sara Sampson veranschaulichen wir diesen Eigenschaftenkatalog nachfolgend. Wir konzentrieren uns dabei exemplarisch auf die titelgebende Figur des Stücks. Sara ist, wie einleitend bereits ausgeführt, die Tochter von Sir William Sampson, einem britischen Landadeligen. Tabelle 2 zeigt ihre annotierten Eigenschaften anhand von sechs Überblicksdarstellungen der Sekundärliteratur. Obwohl diese verschieden ausführlich in ihrer Figurencharakterisierung und -beschreibung sind, zeichnen sie ein recht konsistentes Bild. Sara wird von der Forschung als tugendhaft und empfindsam beschrieben, ihr Handeln sei durch eine ihr eigene Gefühlsmoral bestimmt und ihre Liebe zu Mellefont übersteigt sogar die Autorität des Vaters.

| Literaturangabe                                  | Zugeschriebene Eigenschaften   |
|--|--|
| Figurenverzeichnis <sup>72</sup>                 | Tochter von William Sampson  |
| Dramenlexikon des 18. Jahrhunderts <sup>73</sup> | »dem englischen Landadel entstammend[]«; Geliebte, Tochter; »Entscheidung für die Liebe (d. h. gegen die väterliche Autorität)«; ist nicht »nur ›verführte Unschuld‹«; aufgrund von Schuldgefühlen unfähig, »das Verzeihen des Vaters anzunehmen«; »Tugendhochmut«; verkennt Situation mit Marwood; Marwoods Rivalin |
| Geschichte des Dramas <sup>74</sup>              | Vater-Tochter-Beziehung ist von »gegenseitiger zärtlicher Liebe  |

<sup>65</sup> Forster 1927, S. 103f.

<sup>66</sup> Rimmon-Kenan 2002, S. 40f.

<sup>67</sup> Vgl. Hochman 1985, S. 89, zitiert nach Fishelov 1990, S. 424.

<sup>68</sup> Pfister 2001, S. 244.

<sup>69</sup> Pfister 2001, S. 245.

<sup>70</sup> Pfister 2001, S. 245. Figuren zeichnen sich zudem durch je verschiedene Konzeptionen aus: *geschlossen* vs. *offen* und *transpsychologisch* vs. *psychologisch*. Vgl. Pfister 2001, S. 246–249.

<sup>71</sup> Mit Jörg Schönert gesprochen haben wir also versucht, möglichst *kodifiziertes* Wissen, das also »geprüft«, »akzeptiert« und weitgehend »entproblematisiert« ist, zusammenzuführen. Schönert 2015, S. 324f.

<sup>72</sup> Vgl. Lessing 2003, S. 432.

<sup>73</sup> Vgl. Stockinger 2001, S. 205–207.

<sup>74</sup> Vgl. Fischer-Lichte 1999, S. 300–304.

|   |   |
|---|---|
|   | bestimmt«, bedrohte Tugend, »tugendhafte[] Tochter«; Geliebte; Überführung der Beziehung mit Mellefont in eine Ehe als Wiedergutmachung ihres Fehltritts; empfindsam; »tugendhafte[s] Mädchen«; »innige Liebe«, die »Vater und Tochter verbindet«; Fehler ist verzeihbar; ihre Tugend ist wiederherstellbar; »stirbt als tugendhafte Tochter«; »Engel«, vergibt ihrer Mörderin              |
| Kindlers Literaturlexikon <sup>75</sup>         | Tochter; drängt Mellefont zur Heirat; Sterbende; Tod lässt sich »nicht zwingend aus ihrer Verfehlung ableiten«; »von Gefühlsmoral bestimmte[s] Handeln«; »Mitur aus Märtyrer-Tradition und bürgerlichem Identifikationsmodell«; Entwicklung von einer »starrten und abstrakten Tugendauffassung zu einer gelebten Moral«; »geschwätzig-haarspalterisches Rasonieren über ihre Empfindungen« |
| Geschichte des deutschen Dramas <sup>76</sup>   | wird in »Unglück dieser Welt gerissen«; »Anerkenntnis, daß Gott ihr Schicksal richtig gefügt hat«; »glücklichere Nebenbuhlerin«; »von Gewissensnöten gepeinigt«; leidet; kann den »gefährlichen Verführer« nicht erkennen; Opfer; »dem Tode verfallen«  |
| Neue deutsche Literaturgeschichte <sup>77</sup> | »verfällt dem Verführer Mellefont«; »Abwendung von der Familie«; »weibliche[r] Idealtypus«  |
| Lessing-Handbuch <sup>78</sup>                  | »töchterliche Liebe«; leidet an ihrer Schuld; Rivalin von Marwood; rachsüchtig; schönes und sittliches Mädchen; sinnlich; Nebenbuhlerin; bereit zu vergeben; sanftmütig; Geliebte; »verkörpert die ›Liebe‹«; »unfähig zur Verstellung«, »keine ›reine Unschuld‹«, »›sündige‹ Liebe zu Mellefont«  |

Tab. 2: Eigenschaftskatalog von Sara Sampson. [Krautter et al. 2020]

### 3.3 Bottom-up-Annotation

Das in diesem Eigenschaftskatalog akkumulierte Wissen – also die sozialen, physischen und handlungsrelevanten Merkmale einer Figur – nutzen wir in der Folge für die Ausbildung der Figurentypologie und die Zuordnung der dramatischen Figuren. Wichtige Orientierung für diese Typologie bietet das für das Theater des 18. und frühen 19. Jahrhunderts so einflussreiche Rollenfachsystem. Wir versuchen dabei, die von Diebold angedeuteten Zusammenhänge von literarischem Figurentypus und Rollenfach analytisch zu verbinden.<sup>79</sup> Demnach müsste es möglich sein, konsistente Gruppen von Figuren zu finden, die einerseits demselben Rollenfach entstammen, also eine spezifische Art des Schauspiels erfordern, andererseits aber auch hinsichtlich ihrer textbasierten Eigenschaften als gruppenbildend aufgefasst werden können.<sup>80</sup> Ein Figurentyp in unserem Sinne ließe sich dann als eine Menge ähnlicher Figuren definieren, die bestimmte Eigenschaften und Merkmale teilen, die sie wiederum gegen andere Figuren hin abgrenzbar machen.<sup>81</sup> Die Herausforderungen bei der Ausdifferenzierung verschiedener Figurentypen wie auch der

<sup>75</sup> Grimm 2009, S. 57f.

<sup>76</sup> Mann 1960, S. 163–166.

<sup>77</sup> Brenner 2011, S. 76–78.

<sup>78</sup> Vgl. Fick 2016, S. 133–147.

<sup>79</sup> Vgl. Diebold 1978, S. 10.

<sup>80</sup> Da sich theaterhistorisch beides nicht immer eindeutig trennen lässt, denken wir das Rollenfach hier mit, obgleich für unsere textbasierten Analysen der Figurentypus die zentrale Kategorie ist. Dazu auch Platz-Waury 1999, S. 81: Die Rollenfächer würden daran erinnern, dass »die Tendenz zur Typisierung mit der dramatischen Figurenkonzeption eng verbunden ist«.

<sup>81</sup> Vgl. Eder et al. 2010, S. 38–42.

anschließenden Kategorisierung der Figuren, verdeutlicht Fishelov, der eine Spannung betont zwischen der »individuality of a character and the fact that this very individual is an ›intersection‹ of abstract typical traits«. <sup>82</sup> Das Bestreben einer quantitativen Analyse fügt dieser Spannung eine weitere Dimension hinzu: Eine binäre Unterscheidung zwischen weiblichen und männlichen Figuren allein ist etwa kaum hinreichend, um eine homogene Gruppe herauszubilden. Sie kann aber – kombiniert etwa mit weiteren Merkmalen wie dem Alter der Figuren oder ihrer sozialen Stellung – ein erster Schritt zur Typenbildung sein. Der Anspruch, auch »edle und ernsthafte Väter, polternde Alte, zärtliche, launige Väter, freundliche, gütige, vornehme alte Herren, gefoppte Vormünder, hartherzige, bürgerliche Väter, drollige Hausväter usw.« <sup>83</sup> zu unterscheiden, ist jedoch aufgrund der nur noch marginalen Abweichungen sicherlich zu feingranular, nicht nur mit Blick auf die automatische Klassifizierung und die dafür benötigten Annotationsdaten. Es gilt also einen Skalierungsmodus zu finden, der die Komplexität einzelner literarischer Figuren so gut als möglich achtet, gleichzeitig aber auch die Grenzen ihrer computationellen (und mithin ihrer literaturwissenschaftlichen) Identifikation nicht überschreitet.

Aus diesem Grund verfolgen wir eine Doppelstrategie, die durch unterschiedliche Erschließungstechniken gekennzeichnet ist. Anhand des erarbeiteten Eigenschaftenkatalogs, der die in der Sekundärliteratur gelisteten Charakterzüge zusammenführt, nehmen wir eine erste *Bottom-up*-Verortung vor, die der Erschließung von möglichen Zuschreibungen an die Figuren dient: Aus den Rollenfächern abgeleitete Typisierungen wie ›Liebhaberin‹, ›Vaterfigur‹ oder ›Bedienstete‹ werden dabei um spezifizierende Zuschreibungen ergänzt, wie etwa durch Handlungsfunktionen (›intrigiert‹, ›ist Bote‹) oder durch die Zugehörigkeit zu einem Stand (›Adel‹, ›Klerus‹). Die dadurch entstehende Liste umfasst insgesamt 43 Attribute, von denen jeder Figur bis zu vier zugewiesen wurden. <sup>84</sup> Hiermit sind alle Figuren unseres Korpus zumindest durch eines der 43 Attribute verschlagwortet und damit zugleich einer Gruppe von Figuren zugehörig, die diese Zuschreibungen verbindet. Es handelt sich hierbei also um eine approximative Operationalisierung von Figurenähnlichkeit, die sich an einer Annäherung an Figurentypen versucht. <sup>85</sup> Die annotierten Attribute – ›Mutter‹, ›Liebhaberin‹, ›Adelige‹, ›Militär‹, ›Gegenspieler‹, ›Künstler‹ usw. – sind zwar verwandt mit dem Zielphänomen *Figurentypus*, aber nicht deckungsgleich. Sie bleiben eher abstrakt und tragen deshalb vor allem indirekt zur Figurencharakterisierung bei. Die vorgenommene Formalisierung und Generalisierung kann im Einzelfall sogar zu Fehlern führen, dass wohlbekanntere Informationen über die Figureneigenschaften nicht berücksichtigt werden können, da diese Informationen nicht als Attribut verschlagwortet wurden. Zu diesen unberücksichtigten Informationen gehört teilweise literaturwissenschaftliches Basiswissen, das in hohem Maße relevant für die Textinterpretationen ist, in unserem Ansatz aber nicht beachtet wurde (wie etwa die Frage, ob eine Figur den Freitod wählt, Mörder\*in ist etc.). Unsere Prämisse für die Zuordnung von Attributen war es, sowohl unter Berücksichtigung dramengeschichtlicher Entwicklungen – etwa der Etablierung bürgerlicher Trauerspiele – als auch den innerfiktionalen Realitäten der Dramen, plausible Übereinstimmungen zu ermöglichen. Tabelle 3 listet beispielhaft fünf Figuren und die ihnen zugewiesenen Annotationen.

| Drama                     | Figur         |
|---------------------------|---------------|
| Der sterbende Cato (1732) | Cäsar         |
| Miß Sara Sampson (1755)   | Sara          |
| Der Hofmeister (1774)     | Läufer        |
| Die Räuber (1781)         | Franz Moor    |
| Maria Magdalene (1844)    | Meister Anton |

Tab. 3: Beispiele für die Bottom-up-Annotation. [Krautter et al. 2020]

### 3.4 Top-down-Annotation

Die zweite Erschließungstechnik folgt einer umgekehrten Bewegung (*Top-down*). Ausgangspunkt waren hier nicht die Eigenschaften einzelner Dramenfiguren, sondern die bereits in der Forschung als solche diskutierten Figurentypen. Wir beschränken uns dabei bewusst auf drei stärker rezipierte Typen: den ›zärtlichen Vater‹, die ›tugendhafte Tochter‹ und den/die ›Intrigant\*in‹. Die Zuordnung von Figur zu Figurentyp stützt sich hierbei nur noch in Teilen auf den Eigenschaftenkatalog, sie ist stattdessen zu einem großen Teil durch die Forschung und mitunter die Lektüre der Primärtexte fundiert. Der nachfolgende Abschnitt erläutert die drei gewählten Figurentypen und begründet die Auswahl der zugeordneten dramatischen Figuren. Wir beginnen mit den ›tugendhaften Töchtern‹.

<sup>82</sup> Fishelov 1990, S. 422.

<sup>83</sup> Doerry 1926, S. 11.

<sup>84</sup> Es handelt sich hierbei um keine arbiträr gesetzte theoretische Obergrenze. Die Zahl ergibt sich stattdessen induktiv aus den Annotationsdaten.

<sup>85</sup> Vgl. zur approximativen Operationalisierung das Vorgehen von Reiter / Willand 2018, insbesondere S. 54, 69.

Das Konzept ist Brita Hempels Monografie *Sara, Emilia, Luise: drei tugendhafte Töchter* entlehnt und stellt den Versuch dar, das breite Feld der Töchter und Liebhaberinnen zu strukturieren.<sup>86</sup> Ziel ist es, eine möglichst vergleichbare Gruppe an Figuren zu kondensieren, die eine Mehrheit an Eigenschaften teilt. Hempel kategorisiert Sara Sampson, Emilia Galotti und Luise Miller als »tugendhafte Töchter«, die »in ihrem Verhalten kindliche Liebe und kindlichen Gehorsam mit weiblicher Tugend, mit Häuslichkeit und mit der in der empfindsamen Familie in besonders hohem Maße von den Frauen geforderten Kunst des Ausgleichens und Zurücksteckens vereinen« würden.<sup>87</sup> Dass der Tugendbegriff ein Schlüsselwort für solche Tochterfiguren ist,<sup>88</sup> betont auch Friedrich Vollhardt in seiner *Lessing-Monografie*, wenn er mit Blick auf Sara Sampson urteilt, dass sie im Drama als »Inbegriff tugendhafter Weiblichkeit vorgestellt« werde.<sup>89</sup> Sowohl in den beiden *bürgerlichen* Trauerspielen Lessings als auch in Schillers *Kabale und Liebe* nimmt die Vater-Tochter-Beziehung die zentrale Stellung innerhalb des familiären Zusammenlebens ein.<sup>90</sup> Die Töchter würden »auf die väterliche Zuneigung heftig und beinahe schwärmerisch« reagieren. Bei Emilia steigert sich dieser Eindruck zusätzlich durch ihr »Verhältnis zu ihrem Bräutigam«, das »konventionell und kühl« sei und sich somit deutlich von der Vater-Tochter-Beziehung absetze.<sup>91</sup>

Neben die drei von Hempel geführten Beispiele haben wir zwei weitere Tochterfiguren gestellt, die den größeren Kreis der »tugendhaften Töchter« abstecken und Sara Sampson, Emilia Galotti und Luise Miller literaturgeschichtlich einrahmen. Es handelt sich um Lottchen aus Christian Fürchtgott Gellerts *Die zärtlichen Schwestern* (1747) und die Titelfigur Agnes Bernauer aus Friedrich Hebbels gleichnamigem Trauerspiel. Während Gellerts Rührstück nur kurz vor *Miß Sara Sampson* veröffentlicht wurde, liegt zwischen Lessings Trauerspiel und Hebbels *Agnes Bernauer* (1851) fast ein ganzes Jahrhundert. Zwar wird sie von ihrem späteren Mann Albrecht als »ehr- und tugendsame Augsburgische Bürgertochter«<sup>92</sup> verteidigt und auch die typischen Bestandteile des bürgerlichen Trauerspiels scheinen vordergründig alle vorhanden – die Liebesbindung zwischen Albrecht und Agnes, der sich daraus entspinnde Ständekonflikt, die Vater-Tochter-Beziehung (hier zusätzlich ergänzt um eine Vater-Sohn-Beziehung).<sup>93</sup> Hebbel akzentuiert das Trauerspiel im 5. Akt jedoch so, »daß eigentlich jene aufklärerische Perspektive verlorengelht, die das bürgerliche Trauerspiel in seinen herausragenden Beispielen so wichtig macht«, nämlich dass »die Moralität und das Glück des einzelnen Menschen jenseits jeglicher Standesbindung allmählich zur Basis des Lebens werde[]«. <sup>94</sup> Obwohl sie teilweise als »Hebbels »Luise Millerin«« gelesen wurde,<sup>95</sup> hebt sich Agnes Bernauer also nicht nur in der Entstehungszeit von den übrigen Dramen ab. Der zur Katastrophe führende Konflikt entzündet sich nicht mehr an der Reinheit und Unschuld der Tochter – die Heirat von Agnes und Albrecht ist nämlich längst vollzogen und hat letztlich auch den Segen des Vaters Caspar Bernauer erhalten –, sondern an dem Zerbröckeln von »Differenzen, auf deren Geltung der Staat gegründet ist«. <sup>96</sup>

| Autor                 | Drama                            |
|-----------------------|----------------------------------|
| Gellert, Christian F. | Die zärtlichen Schwestern (1747) |
| Lessing, Gotthold E.  | Miß Sara Sampson (1755)          |
| Lessing, Gotthold E.  | Emilia Galotti (1772)            |
| Schiller, Friedrich   | Kabale und Liebe (1784)          |
| Hebbel, Friedrich     | Agnes Bernauer (1851)            |

Tab. 4: Übersicht der ausgewählten »tugendhaften Töchter«. [Krautter et al. 2020]

Das Adjektiv »zärtlich« wurde »im späten 18. Jahrhundert im Umfeld bürgerlicher Trauerspiele und Familiengemälde zur Ausdifferenzierung von Rollenfächern verwendet«, wie Anja Schonlau in ihrer Monografie *Emotionen im Dramentext* betont.<sup>97</sup> Diese Ausdifferenzierung wird etwa in Figuren des »zärtlichen Vaters« evident, die wir zu Beginn unseres Beitrags anhand Sara

<sup>86</sup> Vgl. dazu auch Fischer-Lichte 1999, S. 300–309.

<sup>87</sup> Hempel 2006, S. 32.

<sup>88</sup> Siehe auch Fischer-Lichte 1999, S. 309: »Die Tugend der Tochter [...] bleibt durchgehend positives Leitbild der bürgerlichen Dramatik des 18. Jahrhunderts«. Erhellend sind hierfür auch die Ausführungen von Cornelia Mönch, die die unterschiedlichen poetologischen Programme von Lessings *Miß Sara Sampson* und Johann Gottlob Benjamin Pfeils *Lucie Woodvil* (1756), die häufig als Schwesterfigur der Sara rezipiert wurde, gegeneinanderstellt. Vgl. Mönch 1993, S. 13–29.

<sup>89</sup> Vollhardt 2018, S. 125. Dazu auch Stephan 1985, S. 8, die postuliert: »Die Tugend ist das Thema, um das alle Gespräche kreisen«.

<sup>90</sup> Siehe dazu auch die vergleichbaren Muster in der Figurenkonstellation von *Emilia Galotti* und *Kabale und Liebe* bei Alt 2000, S. 354: »Louise ist mit ihrer Mischung aus Empfindsamkeit, Furcht und Tatkraft fraglos eine Wahlverwandte von Lessings Titelheldin [Emilia]; in der aufgeklärten Mätresse Lady Milford läßt das Drama die Gräfin Orsina aufleben, daß der Musiker Miller dem biedereren Hausvater Odoardo Galotti, seine Frau wiederum der ehrgeizigen, aber naiven Claudia nachempfunden ist, bleibt ebenso offenkundig wie die Nähe zwischen den windigen Intrigantengestalten Marinelli und Wurm«.

<sup>91</sup> Kaiser 1984, S. 18.

<sup>92</sup> Hebbel 1963a, S. 731.

<sup>93</sup> Vgl. Durzak 2001, S. 84f.

<sup>94</sup> Durzak 2001, S. 85.

<sup>95</sup> Brittnacher 1996, S. 78. Dort heißt es: »Die vom rebellischen Pathos des Dramas eingenommenen Interpreten lasen die Agnes Bernauer, als handele es sich um Hebbels »Luise Millerin«, stand doch die aus dem bürgerlichen Trauerspiel geläufige Opposition von aufklärerisch schwärmerischem Freiheitspathos und abgeklärt zynischer Staatsapologie als vieldeutiger Konflikt auch im Zentrum von Hebbels »deutschem Trauerspiel«.

<sup>96</sup> Brittnacher 1996, S. 92.

<sup>97</sup> Schonlau 2017, S. 45.

Sampsons Vater Sir William exemplarisch konturiert haben. Die Reihe der »zärtlichen Väter« haben wir zusätzlich zu William Sampson um die Figuren Odoardo Galotti (Emilia Galotti), Cleon (Die zärtlichen Schwestern), Musikus Miller (Kabale und Liebe) sowie Lord Berkeley (Sturm und Drang) ergänzt.

Was zeichnet diese Väter nun aus? In seiner Studie zum Rollenfach des 18. Jahrhunderts urteilt Bernhard Diebold, dass »[d]ie »zärtlichen« und gewöhnlich zugleich auch »raisonnierenden« Väter [...] nichts als gut und milde« wären, »das beste wollen und moralisch reden« würden.<sup>98</sup> Das schließe eine gewisse »Verdrießlichkeit« aber keinesfalls aus, so dass der Übergang etwa zum »polternden Vater« fließend sei.<sup>99</sup> So gleicht beispielsweise Sir Williams erste *unerbittliche* Reaktion, die die Flucht von Sara und Mellefont überhaupt erst bedingt, der Rachelogik Marwoods: »er ist bereit, Unrecht mi[t] Unrecht, Verletzung mit Abscheu, Böses mit Bösem zu vergelten«.<sup>100</sup>

Für das bürgerliche Drama sei zudem eine »Emotionalisierung der Väter« zu beobachten, die der »Gefühlsvertiefung der Töchter« entspreche.<sup>101</sup> Die Väter sind auch »empfindsame Väter«:

»So kühl der alte Galotti oder der alte Miller ihren Ehefrauen begegnen, so wenig der alte Miller von der individualisierten Liebe zwischen Mann und Frau versteht oder – in Abwehr Ferdinands – zu verstehen vorgibt, so glühend lieben die beiden Alten ihre herrlichen Töchter.«<sup>102</sup>

Gemein sei den Vaterfiguren auch eine gewisse Schwäche: Sowohl William Sampson als auch Odoardo Galotti würden die Katastrophe dadurch einleiten, dass sie »zu nachgiebig, zu ritterlich und liberal« handelten.<sup>103</sup> Von ihren »normativen Vorstellungen« könnten sie nur schwerlich abrücken.<sup>104</sup> Zu William Sampson hält Peter-André Alt deshalb fest: »Der milde, gerührte Vater war früher ein autoritärer Hausherr, der durch seine fehlende Flexibilität die unheilvollen Ereignisse fast zwangsläufig auslöste«.<sup>105</sup>

| Autor                 | Drama                            |
|-----------------------|----------------------------------|
| Gellert, Christian F. | Die zärtlichen Schwestern (1747) |
| Lessing, Gotthold E.  | Miß Sara Sampson (1755)          |
| Lessing, Gotthold E.  | Emilia Galotti (1772)            |
| Klinger, Friedrich M. | Sturm und Drang (1777)           |
| Schiller, Friedrich   | Kabale und Liebe (1784)          |

Tab. 5: Übersicht der ausgewählten »zärtlichen Väter«. [Krautter et al. 2020]

Die wohl heterogenste der drei angeführten Figurengruppen dürfte die der »Intrigant\*innen« sein.<sup>106</sup> Das mag schon in der Intrige selbst angelegt sein, die als »genuin dramatische[s] Prinzip, das Veränderung bewirkt« zu verstehen ist,<sup>107</sup> deshalb aber auch für vielfältige Erscheinungsformen offen ist.<sup>108</sup> Im literaturgeschichtlichen Zusammenhang ist sie zudem »zunächst moralisch indifferent«.<sup>109</sup> Sie kann ebenso einem guten Zweck dienen, wie die Katastrophe auslösen. So ließe sich zwischen »zerstörende[n] und rettende[n] Intrigen« unterscheiden, wodurch zugleich Charakter und Strategie der Intrigant\*innen abgegrenzt werden könnten.<sup>110</sup> Probates Hilfsmittel intriganter Machenschaften ist häufig der Brief, den Peter Pütz als verlässliches »Vehikel der Intrige« bezeichnet.<sup>111</sup> In der satirischen Typenkomödie werde die Intrige »von den Vernünftigen in der Weise angelegt, daß sie

<sup>98</sup> Diebold 1978, S. 104. Erst mit Gellerts Die zärtlichen Schwestern beginne jedoch »im realistischen bürgerlichen Milieu der zärtliche Vater den Raisonneur zu überwiegen«.

<sup>99</sup> Denn operational voneinander zu trennen, seien die spezifischeren Fächer des Vaters beziehungsweise des Alten – also beispielsweise »launigte«, »polternde«, »zärtliche« oder »edle« Väter – nicht. Vgl. Diebold 1978, S. 103f.

<sup>100</sup> Ter-Nedden 1986, S. 54.

<sup>101</sup> Kaiser 1984, S. 18.

<sup>102</sup> Kaiser 1984, S. 18.

<sup>103</sup> Wittkowski 2013, S. 28. Gisbert Ter-Nedden betont etwa William Sampsons zögerliches Handeln, das ihn »den rechten Zeitpunkt zur Versöhnung versäum[en]« lasse. Ter-Nedden 1986, S. 53.

<sup>104</sup> Kaiser 1984, S. 17.

<sup>105</sup> Alt 1994, S. 196.

<sup>106</sup> Das Rollenfach der Intrigant\*innen bleibt im 19. Jahrhundert länger bestehen als andere Fächer. Das ist auch darin zu begründen, dass es einerseits mit *Charakterrollen* assoziiert wurde, häufig aber auch schlicht den/die Gegenspieler\*in meinte. Vgl. Schonlau 2017, S. 180f.

<sup>107</sup> Alt 2004, S. 1.

<sup>108</sup> Vgl. Asmuth 2016, S. 124f.

<sup>109</sup> Alt 2004, S. 1.

<sup>110</sup> Matt 2002, S. 34.

<sup>111</sup> Pütz 1970, S. 82.

zum Schein auf den Fehler desjenigen eingeht, gegen den sie sich richtet.«<sup>112</sup> Die tragische Intrige verlangt den Überlegungen Aristoteles' folgend nach *Überraschung* und *Glaubwürdigkeit* zugleich.<sup>113</sup> Die Intrige ist dabei im Wesentlichen kein Mittel der Selbstbestimmung des/der Intrigant\*in, sondern dient als destabilisierendes, selektierendes Werkzeug.<sup>114</sup>

Die Gruppe der ›Intrigant\*innen‹ ist in unserem Fall als einzige nicht geschlechtsspezifisch. Alt bestimmt den/die Intrigant\*in als »Figur der Intervention« – exemplarisch zieht er dafür Lessings Marinelli heran.<sup>115</sup> Dieser »verkörper[e] ein dynamisches Prinzip, das mit dem Begriff der ›Störung‹ zu beschreiben« sei.<sup>116</sup> Urheber der Intrige – auch das passt zum gewissenlosen Kammerherren Marinelli – sind nach Karl Eibl häufig Bedienstete.<sup>117</sup> Solche intriganten Höflingsfiguren, »aalglatte, willfährige und zynische Adlati«,<sup>118</sup> dürfen zu der Schaffenszeit Lessings als durchaus etabliert gelten.<sup>119</sup> Neben generell lasterhaftem Verhalten seien es vor allem einzelne Leidenschaften, die als Motivation zur tragischen Intrige dienen: »Neid, Hass und Eifersucht«.<sup>120</sup>

| Autor                  | Drama                                       |
|------------------------|---|
| Gottsched, Johann Chr. | Sterbender Cato (1732)                      |
| Schlegel, Johann E.    | Canut (1746)                                |
| Lessing, Gotthold E.   | Die Juden (1754)                            |
| Lessing, Gotthold E.   | Miß Sara Sampson (1755)                     |
| Lessing, Gotthold E.   | Emilia Galotti (1772)                       |
| Goethe, Johann W.      | Götz von Berlichingen (1773)                |
| Goethe, Johann W.      | Götz von Berlichingen (1773)                |
| Klinger, Friedrich M.  | Die neue Arria (1776)                       |
| Schiller, Friedrich    | Die Räuber (1781)                           |
| Schiller, Friedrich    | Die Verschwörung des Fiesco zu Genua (1783) |
| Schiller, Friedrich    | Kabale und Liebe (1784)                     |
| Iffland, August W.     | Die Jäger (1785)                            |
| Schiller, Friedrich    | Maria Stuart (1800)                         |
| Kleist, Heinrich v.    | Der zerbrochene Krug (1808)                 |
| Nestroy, Johann        | Das Mädl aus der Vorstadt (1841)            |

Tab. 6: Übersicht der ausgewählten ›Intrigant\*innen‹. [Krautter et al. 2020]

## 4. Operationalisierung für den Computer

Nach der inhaltlichen Auseinandersetzung mit Figurentypen und der Beschreibung der annotierten Daten diskutieren wir in diesem Abschnitt die Operationalisierung von Figurentypen für den Computer, so dass eine automatische Erkennung möglich wird.

### 4.1 Begriffe vermessen

Als Operationalisierung für den Computer verstehen wir die Entwicklung von algorithmischen Routinen zur Erkennung von Begriffsinstanzen in Daten. Es geht hier darum, aus einer Menge von Figuren automatisch diejenigen identifizieren zu können, die zu einem bestimmten Figurentyp gehören. Die bisherigen Ausführungen verdeutlichen, dass die manuelle Zuweisung von Figurentypen in Dramen typischerweise über den Gesamttext (und weitere Kontexte, etwa theaterpraktische Überlegungen)

<sup>112</sup> Steinmetz 1978, S. 42. Eibl erkennt hierin ein Muster, das auch auf die Briefszene in *Miß Sara Sampson* zutrefte. Vgl. Eibl 1971, S. 153.

<sup>113</sup> Vgl. Schonlau 2017, S. 177.

<sup>114</sup> Vgl. Alt 2004, S. 3.

<sup>115</sup> Alt 2004, S. 9.

<sup>116</sup> Alt 2004, S. 8.

<sup>117</sup> Vgl. Eibl 1971, S. 153.

<sup>118</sup> Mertens 1995, S. 78.

<sup>119</sup> Vgl. Mertens 1995, S. 70.

<sup>120</sup> Schonlau 2017, S. 182.

erfolgt, also verschiedenste, zumeist semantische Kriterien berücksichtigt werden (müssen). Die Operationalisierung für den Computer macht es nötig, diese Kriterien sowohl möglichst genau zu approximieren als auch durch vergleichbare Informationen anzureichern, oder aber stark korrelierende, inhaltlich unabhängige Indikatoren zu finden.

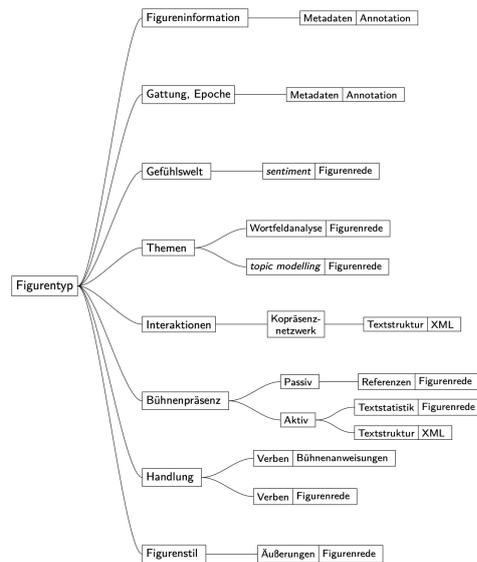


Abb. 1: Hierarchische Operationalisierung von Figurentypen in Dramen. [Krautter et al. 2020]

Abbildung 1 gibt unseren Ansatz zur Operationalisierung schematisch wieder. Da der Zielbegriff »Figurentyp« nicht direkt messbar ist, wird er zunächst durch sieben weitere Teilbegriffe operationalisiert (»Figureninformation«, »Gattung/Epoche«, ...), die für eine Zuordnung relevant sind. Doch auch diese Begriffe lassen sich noch nicht direkt messen, müssen also weiter operationalisiert werden. Daraus ergibt sich eine Hierarchie, an deren Ende direkt messbare Indikatoren stehen (müssen).<sup>121</sup> In Abbildung 1 befinden sich am rechten Ende jeweils zweiteilige Kästen, in denen zum einen in Kurzform die konkrete Umsetzung angegeben ist, und zum anderen die Datenquelle, auf die sie sich bezieht. Metadaten werden hier durch Annotation ermittelt, wobei die Annotationen partiell bereits Teil des Korpus sind. Quelle der Metadaten ist aber in jedem Fall eine manuelle oder manuell überprüfte automatische Zuweisung. Die Textstruktur beschreibt Informationen, die deterministisch und regelbasiert aus einem XML-codierten Korpus extrahiert werden können. Hier sind das vor allem Informationen zur Koprsäsenz von Figuren auf der Bühne. Die meisten Informationen extrahieren wir jedoch aus zwei Textebenen, der Figurenrede und den Bühnenanweisungen der Dramen, wofür verschiedene Verfahren zum Einsatz kommen, die in der Abbildung nur durch Stichworte angegeben sind und an späterer Stelle genauer erläutert werden.<sup>122</sup>

Operationalisierungen wie diese müssen immer im Hinblick auf die Operationalisierbarkeit untergeordneter Begriffe vorgenommen werden, wobei sowohl pragmatische als auch theoretische Aspekte eine Rolle spielen. Damit ist die hier vorgenommene Operationalisierung ebenfalls ein Kompromiss zwischen Anspruch und Machbarkeit. Offensichtlich lässt sich etwa ein Begriff wie »Handlung« nicht allein durch Verben repräsentieren. Auch die Interaktionen einer Figur können nur eingeschränkt durch Koprsäsenznetzwerke abgebildet werden.<sup>123</sup> Zugleich ist die Abgrenzung der beiden Begriffe keinesfalls trennscharf: Handlung entsteht durch handelnde Figuren und deren Interaktion mit anderen Figuren. Gattungs- und Epochenunterschiede sorgen für zusätzliche Komplexität: Unterschiedliche Konventionen z. B. für das Einziehen von Szenengrenzen – entweder nach Auf- oder Abtritt von Figuren oder aber nach Ortswechseln – haben erheblichen Einfluss auf die Figurennetzwerke. Auch am *unteren Ende* der Operationalisierungshierarchie stellen sich weitere Fragen, die dieser Beitrag nur am Rande behandeln kann: So ist etwa die Erkennung von Referenzen in Texten, mit dem sich die Computerlinguistik unter dem Stichwort der Koreferenzauflösung seit vielen Jahren beschäftigt, ein keineswegs triviales Problem.<sup>124</sup>

<sup>121</sup> Was als *direkt messbar* angesehen werden kann, ist im Einzelfall zu entscheiden und bei textbezogenen Begriffen auch vom Forschungsstand in der maschinellen Sprachverarbeitung abhängig.

<sup>122</sup> An einer Reihe dieser Begriffe oder Teilbegriffe wurde in der Forschung bereits gearbeitet: Nalisnick / Baird 2013 modellieren die Gefühlswelt zwischen jeweils zwei Figuren. Karsdorp et al. 2015 fokussieren eine bestimmte Art der Relation zwischen Figuren (nämlich auf Liebe, die sich in unserem Schema auf Handlung, Gefühlswelt und Interaktion verteilen würde). Dramatische Gattungen wurden von Schöch 2017 mithilfe von *Topic Modeling* operationalisiert. Fischer et al. 2018 arbeiten an der Erkennung von Protagonisten und verwenden dafür Merkmale, die sich aus linguistischen/statistischen Eigenschaften der Figurenrede und den Interaktionen der Figuren zusammensetzen. Aktive und passive Bühnenpräsenz wurde in eigenen Arbeiten expliziert, vgl. Willand et al. 2020. Zur Erkennung von Figurenreferenzen in Texten wurde ein Korpus mit annotierten Koreferenzketten veröffentlicht, vgl. Pagel / Reiter 2020.

<sup>123</sup> Was dem Koprsäsenznetzwerk zu einer vollumfänglichen Darstellung von Interaktionen fehlt, ist jegliche Art der Qualifizierung, also der Benennung und Kategorisierung einer Interaktion. Eine Liebesszene auf dem Balkon und ein Schwertkampf in der Arena sind sehr unterschiedliche Arten der Interaktion, führen aber beide zu einer Koprsäsenz.

<sup>124</sup> Vgl. Sukthanker et al. 2020 für einen aktuellen Überblick.

Nachfolgend werden die verwendeten Indikatoren aus einer technischen Perspektive erläutert und anschließend tabellarisch zusammengefasst. Zur klareren Differenzierung verwenden wir für einzelne Indikatoren die englische Bezeichnung »feature«. Die technischen Bezeichnungen der Features, die teilweise in den Abbildungen verwendet werden, sind jeweils in nichtproportionaler Schrift gesetzt.

## 4.2 Verwendete Features

Die Features lassen sich anhand der verwendeten Techniken grob in folgende Gruppen einteilen:<sup>125</sup> Textstatistik, Netzwerkmaße, Bühnenpräsenz, Wortfelder, *Topics*, Handlungsverben, *Sentiment*, Figureninformationen und *Priors*. Im Folgenden beschreiben wir diese Features und ihre konkrete Implementierung näher.

### 4.2.1 Textstatistik

Zu den textstatistischen Features zählen

- *tokens*: die Anzahl an *Tokens*, die eine Figur im Verlauf des Stücks äußert. Die absolute Anzahl an Tokens wird über die Anzahl an Tokens im gesamten Stück normalisiert.<sup>126</sup>
- *utterances*: die Anzahl an Äußerungen, die eine Figur tätigt. Eine Äußerung ist hierbei als die Spanne an Tokens bestimmt, die eine Figur äußert, ohne von einer anderen Figur unterbrochen zu werden.
- *utteranceLengthMean*: die in Tokens gemessene, durchschnittliche Länge der Äußerungen einer Figur.
- *utteranceLengthSd*: die Standardabweichung der Länge der Äußerungen einer Figur.
- *type-token-ratio* (TTR): Die TTR wird gebildet, indem die Anzahl an Types, also der unikal Tokens, geteilt wird durch die Gesamtzahl der Tokens. Da die Höhe der TTR von der Textlänge abhängig ist und die annotierten Figuren unterschiedlich viel sprechen, verwenden wir eine standardisierte TTR, bei der der Text in Segmente gleicher Länge eingeteilt wird. Die TTR wird für jedes dieser Segmente einzeln berechnet und anschließend gemittelt. Wir haben uns hier für eine Segmentlänge von 200 Tokens entschieden.

### 4.2.2 Netzwerkmaße

Netzwerkmaße beschreiben verschiedene Metriken, die eingesetzt werden können, um zentrale Knoten in (formalen) Netzwerken zu ermitteln. Die Netzwerkanalyse ist eine in den Computational Literary Studies häufig genutzte Technik zur Analyse von Figureninteraktionen. Dazu werden zunächst sogenannte Kopräsenzmatrizen erstellt, die die Information enthalten, welche Figuren zu welchem Zeitpunkt gemeinsam auftreten, also *kopräsent* sind. Wir zählen zwei Figuren als *kopräsent*, wenn sie sich in der gleichen Szene durch Figurenrede äußern. Die Knoten im Kopräsenznetzwerk stellen die Figuren dar, die Kanten zwischen Figuren repräsentieren die gemeinsame Präsenz und das Gewicht einer Kante ergibt sich aus der Zahl der Szenen, in denen zwei Figuren *kopräsent* sind.

Auf diesen Netzwerken berechnen wir die Maße:

- *Degree Centrality* (*degree*)
- *Weighted Degree Centrality* (*wdegree*)
- *Betweenness Centrality* (*between*)
- *Closeness Centrality* (*close*)
- *Eigenvector Centrality* (*eigen*).<sup>127</sup>

---

<sup>125</sup> Bei dieser Einteilung spielen technische Aspekte eine stärkere Rolle – sie sind daher etwas anders gruppiert als in Abbildung 1.

<sup>126</sup> Wie in der Sprachverarbeitung üblich, unterscheiden wir zwischen *Tokens* und Wörtern. Die Zählung von *Tokens* umfasst hier insbesondere auch Satzzeichen und Ziffern.

<sup>127</sup> *degree* bezeichnet die Anzahl der an einem Knoten liegenden Kanten, *weighted Degree* die Summe der Kantengewichte der an einem Knoten liegenden Kanten und die drei *centrality*-Varianten kodieren verschiedene Arten, um die Zentralität eines Knotens im Gesamtnetzwerk zu quantifizieren. Vgl. für eine Übersicht zu den verschiedenen Maßen Wassermann 1994, passim.

### 4.2.3 Bühnenpräsenz

Da die beschriebenen Netzwerkmaße auf der gemeinsamen Bühnenpräsenz von Figuren basieren, decken sie nur einen Ausschnitt potentieller Figureninteraktionen ab. Die beiden folgenden Features konzentrieren sich deshalb auf die Präsenz einzelner Figuren. Wir unterteilen diese in eine aktive Präsenz, wenn also Figuren selbst auf der Bühne stehen und sprechen, sowie eine passive Präsenz, wenn in Dialogen über Figuren gesprochen wird, die nicht auf der Bühne stehen:

### 4.2.4 Wortfelder

Um die Semantiken der Figurenrede zu erschließen, nutzen wir sieben verschiedene Wortfelder (WF). Ein Wortfeld enthält lemmatisierte Einträge zu Begriffen einer bestimmten Domäne, die wir als relevant für die annotierte Zeitspanne (1730–1850) erachten. Für jede Figur wird gezählt, wie oft sich das Lemma eines gesprochenen Tokens in einer der Wortlisten wiederfindet und dies wird anschließend durch die Gesamtzahl der von einer Figur gesprochenen Tokens geteilt. Die Wortlisten wurden manuell angelegt und umfassen die Bereiche Familie, Krieg, Liebe, Politik, Ratio, Religion und Wirtschaft.<sup>128</sup>

### 4.2.5 Topics

Neben den händisch angelegten Wortfeldern kann die Figurenrede weitere Themenbereiche enthalten, an die wir uns mittels *Topic Modeling*<sup>129</sup> annähern. Hierzu bedienen wir uns der *Latent Dirichlet Allocation* (LDA) mit *Gibbs Sampling*. Als Eingabe verwenden wir das *GerDraCor-Korpus*<sup>130</sup> und segmentieren die einzelnen Dramen in Abschnitte von je 1000 Tokens. Wir betrachten dabei lediglich diejenigen Tokens, die als Nomen, Verb, Adjektiv oder Adverb ausgezeichnet wurden. Auch Eigennamen werden entfernt. Die Anzahl der Topics wird auf 20 festgelegt, woraus sich die Topics T1–T20 ergeben. Als Feature-Wert erhält jede Figur die *Posterior*-Wahrscheinlichkeit eines Topics, also die Wahrscheinlichkeit einer Figur, Tokens zu einem bestimmten Topic zu äußern. Die Posterior-Wahrscheinlichkeit einer Figur wird auf Grundlage aller Äußerungen dieser Figur ermittelt.

### 4.2.6 Handlungsverben

Sowohl die Wortfelder als auch die Topics sollten zu einem gewissen Grad den Inhalt der Figurenrede wiedergeben und somit indirekt Aufschlüsse über die Funktion einer Figur für die Handlung des Dramas zulassen. Als weiteres Feature, das die Handlungsfunktion der Figuren zu quantifizieren versucht, nutzen wir die Verteilung von Verben im Dramentext. Dazu werden aus dem gesamten Korpus die zehn häufigsten Verben extrahiert – sowohl in den Figurenreden als auch in den Bühnenanweisungen –, die einer bestimmten Figur zugeordnet sind.<sup>131</sup> Modal- und Hilfsverben werden nicht gezählt. Für jede Figur wird anschließend die Häufigkeit der geäußerten Verben ermittelt und durch die gesamte Zahl der Tokens, die eine Figur spricht, geteilt. In Tabelle 7 sind die häufigsten Verben gelistet, aufgeschlüsselt nach Äußerungen und Bühnenanweisungen sowie in ihrer Frequenz absteigend.

| Figurenrede (Präfix: utt.) | Anzahl |
|----------------------------|--------|
| lassen                     | 2,655  |
| sagen                      | 2,600  |
| kommen                     | 2,209  |
| sehen                      | 2,196  |
| machen                     | 2,109  |
| gehen                      | 1,922  |
| geben                      | 1,594  |
| tun                        | 1,144  |

<sup>128</sup> Sie sind auf [GitHub](#) abrufbar. Vgl. zum Vorgehen Willand / Reiter 2017, S. 183–190.

<sup>129</sup> Vgl. dazu Blei et al. 2003, passim.

<sup>130</sup> Vgl. dazu Fischer et al. 2019, passim.

<sup>131</sup> Das können entweder Bühnenanweisungen sein, die auf die Sprecherbezeichnung folgen oder direkt in die Figurenrede eingebettet sind.

|        |     |
|--------|-----|
| wissen | 947 |
| stehen | 925 |

Tab. 7: Die häufigsten zehn Verben in Figurenrede (links) beziehungsweise Bühnenanweisungen (rechts), die als Approximation an Handlung eingesetzt werden. [Krautter et al. 2020]

## 4.2.7 Sentiment

Um Einblicke in das Gefühlsleben einer Figur zu bekommen, bedienen wir uns einer statischen Liste an positiv und negativ konnotierten Wörtern aus dem *SentiWS-Korpus*<sup>132</sup>. Die Liste enthält nach Wortarten sortierte lemmatisierte Einträge sowie Gewichtungen und mögliche Flexionsformen zu den Lemmata. Ähnlich wie für die Wortfelder gleichen wir die von einer Figur gesprochenen, automatisch lemmatisierten Tokens mit beiden Listen ab und summieren bei einem Treffer die Gewichtungen für die Funde. Die beiden daraus resultierenden Endwerte (positiv und negativ) werden anschließend durch die Zahl der gesprochenen Tokens einer Figur geteilt, um Vergleichbarkeit zu gewährleisten. Daraus ergeben sich die beiden Features *posRatio* und *negRatio*.

## 4.2.8 Figureninformationen

Unter Metainformationen fallen die bereits zuvor angesprochenen Eigenschaften einer Figur: Alter (*age*), Geschlecht (*gender*) und sozialer Status (*socialStatus*). Mögliche Werte der Eigenschaften sind jung / mittel / alt, weiblich / männlich, niedrig / mittel / hoch. Die Informationen wurden durch manuelle Annotation gewonnen.<sup>133</sup> Abgesehen von der Unterspezifiziertheit einiger Figuren bezüglich ihres Alters und des relational zu bestimmenden sozialen Status lassen sich diese Annotationen mit hoher Zuverlässigkeit durchführen.<sup>134</sup>

## 4.2.9 Priors

Unter Priors verstehen wir Informationen, die durch die Dramen und ihre Entstehung selbst gegeben und nicht unmittelbar an eine einzelne Figur geknüpft sind. Dennoch können diese Priors als Selektion dienen, um Stücke zu gruppieren und somit Figureneigenschaften zu kontextualisieren. Wir verwenden zwei Priors:

- Jahrzehnt (*decade*): Jeder Figur wird zugeordnet, in welchem Jahrzehnt das Drama, in dem sie auftritt, entstanden ist.
- Prosa / Vers (*prose*): Jeder Figur wird zugeordnet, ob das Drama (überwiegend) in Versform oder Prosa geschrieben ist.

Wir ermitteln dies automatisch über die Anzahl an *<l>*-, *<lg>*- und *<p>*-tags in den GerDraCor-TEI-Daten eines Stücks: eine Mehrzahl an Tags der Form *<l>* (*line*) und *<lg>* (*line group*) steht für Versform, eine Mehrzahl an *<p>*-Tags (*paragraph*) steht für Prosa.

## 4.3 Übersicht

Tabelle 8 zeigt die Features in der Übersicht mit ihren Wertebereichen sowie einer Zuordnung aller Features zu den Teilbegriffen aus *Abbildung 1*. Der Begriff *Figurenstil* wird z. B. über die Indikatoren *type-token-ratio* sowie Durchschnitt und Standardabweichung der Längen der Äußerungen einer Figur operationalisiert.

| Teilbegriff             | Feature      |
|-------------------------|--------------|
| Figuren-<br>information | age          |
|                         | gender       |
|                         | socialStatus |
| Gattung,                | decade       |

<sup>132</sup> In der Version 2.0.132, vgl. Remus et al. 2010, passim.

<sup>133</sup> Hinter diesen Annotationen steht eine separate Operationalisierungsaufgabe, die der Analyse von Figurentypen vorgelagert ist. Dabei wurden die drei Attribute Geschlecht, sozialer Status und Alter den Figuren manuell zugewiesen. Vgl. Krautter et al. 2021 (im Erscheinen), passim.

<sup>134</sup> Abhängig vom übrigen Bühnenpersonal können etwa dem Landadel zugehörige Figuren einen höheren Status einnehmen, werden sie mit bürgerlichen Figuren verglichen, oder einen niedrigeren, wenn Figuren aus dem Hochadel auftreten.

|               |   |
|---------------|---|
| Epoche        |   |
|               | prose   |
| Gefühlswelt   | posRatio  |
|               | negRatio  |
| Themen        | Familie, Krieg, Liebe, Politik, Ratio, Religion, Wirtschaft   |
|               | T1-20   |
| Interaktionen | degree  |
|               | wdegree   |
|               | between   |
|               | close   |
|               | eigen   |
| Bühnenpräsenz | actives   |
|               | passives  |
|               | firstBegin  |
|               | lastEnd   |
|               | Tokens  |
|               | utterances  |
| Handlung      | utt.geben, utt.gehen, utt.kommen, utt.lassen, utt.machen, utt.sagen, utt.sehen, utt.stehen, utt.tun, utt.wissen |
|               | sd.gehen, sd.nehmen, sd.sehen, sd.treten, sd.kommen, sd.stehen, sd.setzen, sd.fallen, sd.ziehen, sd.halten      |
| Figurenstil   | TTR   |
|               | utteranceLengthMean   |
|               | utteranceLengthSd   |

Tab. 8: Auflistung aller im Folgenden verwendeten Features, gruppiert nach den Teilbegriffen wie sie in Abbildung 1 verwendet werden. [Krautter et al. 2020]

## 5. Experimente und Auswertung

Die vorgestellten Features nutzen wir im Folgenden, um mehrere *Machine-Learning*-Modelle zu trainieren. Damit verfolgen wir zwei Ziele: einerseits sollten die Modelle in der Lage sein, Figurentypen vorherzusagen, wenn die Features eine erfolgreiche Operationalisierung der Figuren und ihrer Typen darstellen. Die erzielte *Performance* kann also zu einer Qualitätsabschätzung über die Operationalisierung dienen. Andererseits lässt sich untersuchen, welche Features für die Erkennung besonders relevant waren. Davon erhoffen wir uns Einblicke in die zugrundeliegenden Eigenschaften, die die Figurentypen charakterisieren.

Konkret beschreiben wir zunächst den Datensatz und betrachten daraufhin die Verteilung der Datenpunkte und Merkmale, bevor wir uns der automatischen Erkennung von Figurentypen als solche zuwenden. Anschließend richten wir den Blick auf die Features und ihren jeweiligen Beitrag, gefolgt von einer Visualisierung der Datenpunkte und ihrer Verteilung.

### 5.1 Datensatz

|                               |                   |
|-------------------------------|-------------------|
| Figurentyp                    | Distinkte Figuren |
| Insgesamt                     | 257               |
| ›Potentielle Intrigant*innen‹ | 228               |

|                             |            |
|-----------------------------|------------|
| ›Intrigant*innen‹           | 15 (# = 2) |
| ›Nicht-Intrigant*innen‹     | 213        |
| ›Vater‹                     | 23         |
| ›Zärtlicher Vater‹          | 5          |
| ›Nicht-Zärtlicher Vater‹    | 225        |
| ›Tochter‹                   | 26         |
| ›Tugendhafte Tochter‹       | 5          |
| ›Nicht-Tugendhafte Tochter‹ | 215        |

Tab. 9: Übersicht über den Datensatz und seine Verwendung in den Experimenten. ›SVM‹ bezeichnet das Klassifikationsexperiment (Abschnitt 5.3), ›PCA‹ das Experiment zur Bildung der Teilmengen (Abschnitt 5.4). Die Spalte ›Exkludierte Charakterisierungen‹ beschreibt, welche Eigenschaften eine Figur nicht haben darf, damit sie als Negativbeispiel verwendet werden kann. [Krautter et al. 2020]

Tabelle 9 zeigt den Datensatz in einer Übersicht. Insgesamt umfasst er 257 Figuren aus 43 Stücken (durchschnittlich sechs Figuren pro Stück), die zwischen 1730 und 1850 uraufgeführt oder erstpubliziert wurden. Diese wurden 43 verschiedenen Figurencharakterisierungen zugeordnet, so dass insgesamt 468 Paare aus Figur und Charakterisierung im Datensatz enthalten sind. 25 Figuren ist einer der Typen ›Intrigant\*in‹, ›zärtlicher Vater‹ oder ›tugendhafte Tochter‹ zugewiesen. Hierbei ist zu beachten, dass zwar jeweils eine Teilmenge der Figuren als ›Intrigant\*in‹, eine Teilmenge der Väter als ›zärtlicher Vater‹ und eine Teilmenge der Töchter als ›tugendhafte Tochter‹ markiert ist, wir aber keine Annahmen über diejenigen Figuren treffen können, die nicht in der jeweiligen Teilmenge enthalten sind. Weil eine Vaterfigur nicht als ›zärtlicher Vater‹ annotiert ist, heißt das also nicht, dass sie nicht doch diesem Figurentypus entsprechen kann. Da für das Training von Klassifikationsverfahren die Existenz von Negativbeispielen eine zwingende Voraussetzung ist, haben wir für die folgenden Experimente jeweils verschiedene Figurengruppen gegenübergestellt. Ausgehend von der Menge aller Figuren wurden als Negativbeispiele jeweils andere Figuren gewählt, die nicht mit einer ähnlichen Charakterisierung ausgezeichnet waren. In Tabelle 9 sind diese Eigenschaften in der dritten Spalte aufgeführt. Als Negativbeispiele für die ›tugendhafte Tochter‹ dienen also Figuren, die weder ›Tochter‹, ›Liebhaberin‹, ›unverheiratete Frau‹ noch ›Schwester‹ sind. Damit sind in den Negativbeispielen für die ›tugendhafte Tochter‹ nur Figuren enthalten, die gar keine Töchter sind, was mit Recht kritisiert werden kann. Im vorliegenden Szenario gibt es zwei Möglichkeiten damit umzugehen, die wir hier aus Platzgründen nicht beide diskutieren können. Das erste Szenario würde im Datensatz *Minimalpaare* gegenüberstellen, also Figuren die sich möglichst ähnlich sind, aber in der Zielkategorie unterscheiden. Ein Beispiel hierfür wären zwei Tochterfiguren mit ähnlichen Strukturmerkmalen, von denen eine tugendhaft ist und die andere nicht. Ein maschinelles Lernverfahren könnte dann zielgerichtet auf die Unterschiede zwischen den Figuren schauen. Dieses Verfahren ist allerdings sehr voraussetzungsreich und auf Basis der Literaturgeschichte kaum zu leisten: Hinweise auf die Nicht-Zugehörigkeit zu einem Figurentyp finden sich praktisch gar nicht, ein entsprechender Datensatz wäre also nur im Zuge einer umfangreichen Interpretation einzelner Dramen erstellbar. Die von uns gewählte Alternative besteht darin, möglichst klar nicht zugehörige Figuren als Negativbeispiele zu verwenden. Dabei nehmen wir in Kauf, dass das Lernverfahren Zusammenhänge *lernt*, die offensichtlich sind, z. B. dass eine männliche Figur keine tugendhafte Tochter ist. Wir gehen allerdings davon aus, dass der auf diese Weise verfügbare größere Datensatz diesen Nachteil ausgleicht.

## 5.2 Verteilung der Feature-Werte

Entsprechend der quantitativen Auswertungen erhält jede Figur in unserem Korpus für jedes der beschriebenen Features einen Wert. Auf diese Weise entsteht eine Matrix, in der die einzelnen Zeilen die untersuchten Figuren und die Spalten die zugehörigen Features darstellen. Die quantitativen Eigenschaften der Figuren lassen sich somit anhand der Zusammenschau ihrer Feature-Werte beschreiben. Dadurch ist es auch möglich, Figuren oder Figurengruppen hinsichtlich ihrer Feature-Werte miteinander zu vergleichen # etwa anhand der in Abschnitt 3 beschriebenen Annotation von Attributen und Figurentypen.

Abbildung 2 zeigt die Verteilung einer Auswahl von Features für die drei Figurentypen ›Intrigant\*in‹, ›tugendhafte Tochter‹ und ›zärtlicher Vater‹. Die x-Achse gibt den Wertebereich eines Features an, während die y-Achse die Anzahl an Figuren anzeigt, die diesen Wert für das jeweilige feature annehmen. Die Figurenanzahl ist geglättet dargestellt, sodass sich Trends verfolgen lassen. In der Feature-Verteilung wird deutlich, dass kein einzelnes feature imstande ist, alle drei Typen trennscharf voneinander abzugrenzen. Es lassen sich jedoch bestimmte Features ausfindig machen, die aufschlussreiche Unterschiede in den Verteilungen der Figurengruppen abbilden. Bei Feature  $\tau_{16}$  (also das 16. Topic aus dem Topic Model) etwa haben alle drei Kurven ihren Scheitelpunkt an verschiedenen Stellen – ein Großteil der ›zärtlichen Väter‹ (rote Kurve) hat hier also einen höheren Wert (d. h. eine höhere Wahrscheinlichkeit, Wörter aus dem Topic zu verwenden) als die ›tugendhaften Töchter‹ (orange Kurve) oder die ›Intrigant\*innen‹ (blaue Kurve). Ein ähnliches Bild zeigt sich bei dem Feature *passives*. Hier sind es vor allem die

»tugendhaften Töchter«, bei denen hohe Ausprägungen auszumachen sind. Zur Verdeutlichung sei als Gegenbeispiel auf das Feature `close` (Closeness Centrality) hingewiesen: Hier überlappen sich alle Kurven sehr stark, so dass sich das Feature nicht als Unterscheidungskriterium eignet. In den folgenden Experimenten versuchen wir mit maschinellen Lernverfahren diejenigen Feature-Kombinationen zu finden, die einen Figurentypus am besten charakterisieren und ihn gegen andere Figurengruppen abgrenzen.

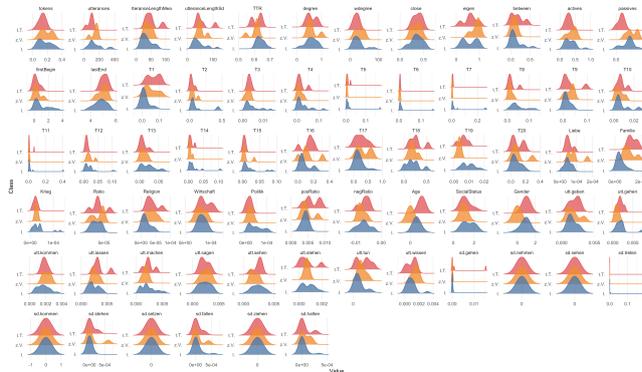


Abb. 2: Verteilung der Feature-Werte für die drei Figurentypen »tugendhafte Tochter« (t.T., orange), »zärtlicher Vater« (z.V., rot) und »Intrigant\*in« (I., blau). Alle Feature-Werte werden hier numerisch repräsentiert. [Krautter et al. 2020]

### 5.3 Klassifikation: Automatische Erkennung von Figurentypen

Für die Klassifikation verwenden wir den Algorithmus *Support Vector Machine* (SVM).<sup>135</sup> Auf einer Teilmenge der Daten, dem Trainingsset, findet der Algorithmus mithilfe der Features diejenigen Parameter, die die Figurentypen unterscheiden. Ein trainiertes SVM-Modell kann dann auf anderen Daten getestet werden, um zu überprüfen, ob auf Grundlage der gesehene Datenpunkte generalisiert werden konnte. Erreicht das Modell auf unbekanntem Testdaten eine hohe Performanz, kann das Training als geglückt angesehen werden und der Algorithmus hat *gelernt*, die Figurentypen auf Basis der Features zu unterscheiden.

Vor dem Training führen wir zwei Vorverarbeitungsschritte aus: Da stark korrelierende Features redundante Informationen beisteuern, werden zunächst paarweise alle Features auf ihre Korrelation überprüft. Bei Paaren mit einem Pearson-Korrelationskoeffizient von mehr als 0,7 wird das Feature entfernt, das mit allen anderen am stärksten korreliert. Da unsere Daten außerdem stark ungleich verteilt sind (es gibt jeweils sehr viel mehr negative als positive Instanzen einer Klasse), verwenden wir SMOTE,<sup>136</sup> um dynamisch neue Trainingsbeispiele zu erzeugen. Ausgehend von *Nearest Neighbors* in den Feature-Werten erzeugt SMOTE künstliche Datenpunkte. Auf diese Weise hilft SMOTE, Ungleichgewichte in den Klassen zu minimieren.

Wir verwenden in allen folgenden Experimenten *10-fold-cross Validation*, um die Daten in zehn unterschiedliche Gruppen aufzuteilen. In jedem *Fold* wird eine andere Aufteilung der Daten in Trainings- und Testdaten vorgenommen, sodass am Ende alle Datenpunkte einmal als Testdaten verwendet werden. Die so entstehenden Trainingsdaten übergeben wir an den SVM-Algorithmus, und erhalten zehn verschiedene Modelle, basierend auf den jeweiligen Trainingsdaten. Wir wenden jedes Modell auf sein jeweiliges Testset an und erhalten ein finales Testergebnis, indem wir den Durchschnitt aller zehn Modelle bilden. Als Evaluationsmetriken verwenden wir *Precision*, *Recall* und *F1-Score*.

Um zu untersuchen, inwieweit welche Features zu einem guten Klassifikationsergebnis beitragen, analysieren wir zunächst Feature-Kombinationen. Dabei orientieren wir uns an den Dimensionen, die in *Abbildung 1* dargestellt wurden. Eine Übersicht über die Qualität der Klassifikationsergebnisse ist in *Tabelle 10* dargestellt. Der Typ »Intrigant\*in« kann dabei mit einem F1-Score von 0,7 am zuverlässigsten erkannt werden.

| Figurentyp          | Precision |
|---------------------|-----------|
| Intrigant*in        | 0,90      |
| Zärtlicher Vater    | 0,50      |
| Tugendhafte Tochter | 0,40      |

<sup>135</sup> Vgl. zu SVM Steinwart / Christmann 2008, passim.

<sup>136</sup> Vgl. dazu Chawla 2002, passim.

Tab. 10: Ergebnisse der automatischen Klassifikation. Gezeigt ist die positive Klasse, als Negativbeispiele dienen jeweils spezifisch ausgewählte Figurengruppen (siehe Abschnitt 5.1). [Krautter et al. 2020]

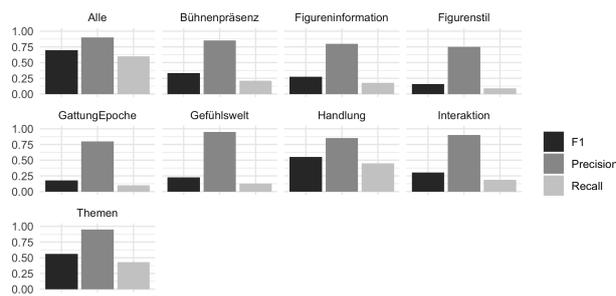


Abb. 3: Klassifikationsergebnisse für die Erkennung von Intrigant\*innen für einzelne Gruppen von Features. Die Einteilung folgt den in Abschnitt 4.1 diskutierten Dimensionen. [Krautter et al. 2020]

Abbildung 3 zeigt die Klassifikationsergebnisse der einzelnen feature-Gruppen für ›Intrigant\*innen‹. Die Verwendung aller Features (Diagramm *Alle*, oben links) erzielt die besten Klassifikationsergebnisse. Die Resultate unterstreichen, dass die Identifikation von dramatischen Figurentypen als multidimensionales Problem aufzufassen ist: Erst im Zusammenwirken der verschiedenen Dimensionen wird eine annehmbare Performanz erreicht. Schwächen zeigen sich vor allem mit Blick auf den recall, während die precision bei den meisten Gruppen höher ist. Das bedeutet, dass zwar nicht alle ›Intrigant\*innen‹ erkannt wurden – die Figuren aber, die das Modell als ›Intrigant\*in‹ auszeichnet, sind dies mit recht hoher Zuverlässigkeit.

Die beiden Dimensionen, die in der Einzelwertung am besten abschneiden, sind Handlung und Themen. Das bestätigt die Annahme, dass ›Intrigant\*innen‹ sich durch ihre Funktion für die Handlung auszeichnen, sich die Eigenschaft aber auch in den Semantiken der Figurenrede niederschlägt. Strukturelle (Präsenz, Interaktionen) und stilistische (Figurenstil) Merkmale scheinen alleine nicht geeignet für eine zuverlässige Erkennung.

Die Messung von *Feature Importance* ist eine Methode, um die Bedeutung einzelner Features für das Klassifikationsergebnis zu ermitteln. Sie vergleicht nach und nach die Performanz des Modells, wenn jeweils eines der Features entfernt wird. Die Abnahme an Performanz entspricht dann der relativen Wichtigkeit des entfernten Features für die Klassifikation.<sup>137</sup> Damit wird ein etwas genauerer Einblick in die Wirkungsweise der Features und auch in die Operationalisierung insgesamt erreicht. Die zehn wichtigsten Features für die Erkennung von Intrigant\*innen sind die folgenden (in Klammern jeweils der Feature-Importance-Wert): T1 (1,31), T5 (1,05), utterances (1,05), T8 (0,89), between (0,79), T11 (0,77), close (0,73), passives (0,71), T9 (0,71), T13 (0,71). Das wichtigste Feature zur Klassifikation von ›Intrigant\*innen‹ ist also Topic 1, das nach manueller Durchsicht hauptsächlich Wörter beinhaltet, die den Themenkreis des familiären Zusammenlebens abdecken. Vor allem familiäre Relationen werden abgebildet.<sup>138</sup> Dazu passt es, dass das Wortfeld *Familie* ebenfalls zu den vergleichsweise wichtigen Features zählt. Zu beachten ist hier jedoch, dass eine hohe Feature Importance nicht identisch ist mit hohen Feature-Werten. Die Wichtigkeit des Features kann auch dadurch zustandekommen, dass alle ›Intrigant\*innen‹ besonders geringe Ausprägungen dieser Features aufweisen. In jedem Fall sind die Merkmale mit hoher Importance besser geeignet, die Figuren zu unterscheiden, als diejenigen mit niedrigerer Importance.

## 5.4 Bildung von Teilmengen

Da für die Figurengruppen der ›tugendhaften Töchter‹ und der ›zärtlichen Väter‹ jeweils nur fünf Figuren annotiert wurden, verspricht eine automatische Klassifikation kaum Einsichten. Um diese Klassen näher zu untersuchen, vergleichen wir deshalb ›tugendhafte Töchter‹ mit anderen Töchtern in unserem Datensatz. Das gleiche nehmen wir auch für die Gruppe der ›zärtlichen Väter‹ vor. Dabei verfolgen wir nicht mehr die oben skizzierte Strategie der zielgerichteten Zusammenstellung von Negativbeispielen, sondern versuchen nachvollziehbar zu machen, inwiefern sich die spezifischen Figurentypen auf Grundlage der Feature-Werte als Teilmenge eines größeren Typs manifestieren.

<sup>137</sup> Für die Feature Importance wurde die Klassifikation mit einem *Random Forest* vorgenommen, und die Feature Importance über den gemittelten Zuwachs an *Impurity* in den Klassen gemessen. Als Implementierung kam das R-Paket *Ranger* zum Einsatz.

<sup>138</sup> Die 20 wichtigsten Terme des Topics sind: Vater, Mutter, Sohn, Kind, Tochter, Haus, alt, Gott, arm, Knabe, Bruder, Jahr, Eltern, Segen, Alter, Sohne, segnen, weinen, hart, Schwester.

Hierzu verwenden wir eine *Principal Component Analysis* (PCA).<sup>139</sup> PCA ist ein mathematisches Verfahren, um hochdimensionale Daten mit geringstmöglichem Informationsverlust in eine niedrigere Zahl an Dimensionen zu überführen, um z. B. zweidimensionale Visualisierungen zu ermöglichen. Die auf diese Weise erzeugten Dimensionen werden *Principle Components* (Hauptkomponenten) genannt und nach der in ihnen enthaltenen Varianz absteigend sortiert. Die erste Hauptkomponente enthält also den größten Anteil der Streuung im Datensatz. Die nachfolgenden Abbildungen zeigen das Ergebnis einer solchen PCA für Töchter (Abbildung 4) beziehungsweise Väter (Abbildung 5), und zwar als paarweise Gegenüberstellungen der ersten drei Hauptkomponenten.<sup>140</sup> Es wird jeweils die Teilmenge der tugendhaften beziehungsweise zärtlichen Figuren mit der Grundmenge an Tochter- beziehungsweise Vaterfiguren verglichen. Auf die sonst übliche Darstellung der Variablen als Vektoren wurde aus Gründen der Lesbarkeit verzichtet.

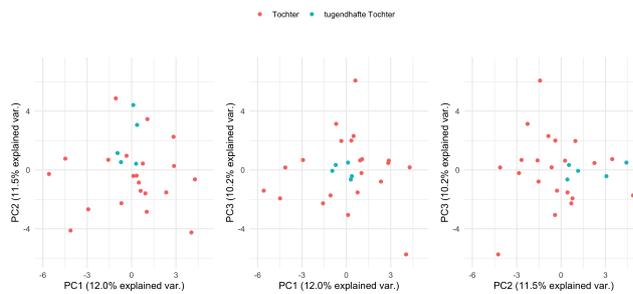


Abb. 4: Visualisierung der Principle component analysis für den Figurentyp »tugendhafte Tochter«. Gezeigt sind die X-Y-Diagramme für jeweils zwei Hauptkomponenten (PC1–3). [Krautter et al. 2020]

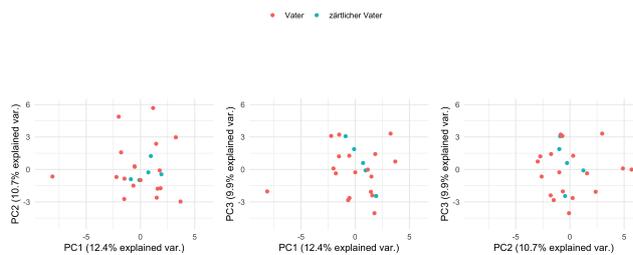


Abb. 5: Visualisierung der principle Component Analysis für den Figurentyp »zärtlicher Vater«. Gezeigt sind die X-Y-Diagramme für jeweils zwei Hauptkomponenten (PC1–3). [Krautter et al. 2020]

Die Abbildungen zeigen, dass sich sowohl die »tugendhaften Töchter« als auch die »zärtlichen Väter« (*blaue Punkte*) als eigene Gruppen innerhalb der Gesamtheit der »Töchter« und »Väter« etablieren (*rote Punkte*). Während die roten Punkte in allen Abbildungen im gesamten Darstellungsbereich verteilt sind, konzentrieren sich die blauen Punkte in vergleichsweise engen Bereichen. Ins Auge fällt hier insbesondere das mittlere Diagramm in *Abbildung 4*, das die Verteilung nach erster und dritter Hauptkomponente zeigt. Hier bilden die »Töchter« die homogenste Gruppe und sind eng am Ursprung positioniert. Auf dieser Basis scheint eine automatische Erkennung von »tugendhaften Töchtern« möglich.

Unter den blauen Punkten befinden sich in vielen Abbildungen auch rote Punkte – Figuren, die »Vater« beziehungsweise »Tochter« sind, aber in unseren Annotationen nicht als »zärtlich« beziehungsweise »tugendhaft« ausgewiesen sind. Dies sind nicht notwendigerweise Fehler, da der Status dieser Figuren bezüglich ihrer Eigenschaften zum Teil unterbestimmt ist. Wir konzentrieren uns im Folgenden auf das jeweils linke Diagramm, in dem die Figuren nach der ersten und zweiten Hauptkomponente dargestellt sind. Im Einzelnen handelt es sich bei den roten Punkten um Gertrud aus Gustav Freytags *Graf Waldemar* (1847) als potentiell »tugendhafte Tochter« sowie Präsident von Walter aus Schillers *Kabale und Liebe* (1784), den Baron aus Lessings *Die Juden* (1749) und Meister Anton aus Hebbels *Maria Magdalene* (1844) als potentielle »zärtliche Väter«.

<sup>139</sup> Vgl. dazu Jolliffe 2002, passim.

<sup>140</sup> Eine *interaktive Visualisierung* erlaubt die Exploration weiterer Hauptkomponenten.

Freytag, als einer der Hauptvertreter des Bürgerlichen Realismus, verhandelt in seinem Waldemar soziale Problemkonstellationen, die aus dem bürgerlichen Trauerspiel bekannt sind. Dabei setzt er auch auf ganz ähnliche Figurenkonstellationen. In diesem Drama ist es die Gärtnerstochter Gertrud, die dem durch Werteverfall charakterisierten Adel in Person des Grafen Waldemar den Spiegel vorhält, diesen läutert und zu einem »neue[n] Leben voll freier, gesunder Thätigkeit« verhilft.<sup>141</sup> Sie ist also mit Gründen als eine tugendhafte, »fromme [Gärtners-]Tochter« zu lesen.<sup>142</sup>

Bei den von der PCA als potentiell »zärtlich« erkannten »Vätern« sind ähnliche dramenhistorische Erklärungen der oben genannten Gruppierungen möglich. Hebbels Meister Anton gilt der Forschung zwar nicht als Vertreter der »zärtlichen Väter«, das Stück Maria Magdalene allerdings als das letzte bürgerliche Trauerspiel. Die durch Gattungskonventionen vorgeschriebenen Problemkonstellationen sind es auch in diesem Fall, die den »Vater« als standesstolz und von der Sorge um die Familienehre geprägt charakterisieren. Diese Attribute teilt er wiederum mit Schillers Präsident von Walter, der als herrschsüchtiger, gewissenloser Machtmensch prima vista alles andere als ein »zärtlicher Vater« zu sein scheint. Dass Präsident von Walter hier trotzdem in die Nähe der »zärtlichen Väter« gerückt wird, kann entweder daran liegen, dass andere ihm ähnliche »Väter« nicht in unseren Daten enthalten sind, oder dass hierfür eine Figureneigenschaft entscheidend ist, die durch unsere Operationalisierung nicht abgedeckt wird.

## 5.5 Diskussion

Die automatische Klassifizierung von Figurentypen muss – dafür gibt es eine ganze Reihe an Ursachen – sowohl mit Blick auf ihre theoretische Konzeption als auch die technische Umsetzung als komplexe multidimensionale Problemstellung aufgefasst werden. Eine Herausforderung bildet dabei das heterogene Untersuchungskorpus. Zwar konzentrieren wir uns mit Blick auf den Einflussbereich der Rollenfächer auf einen literarhistorischen Zeitraum von nur knapp über 100 Jahren, finden darin aber ein breites Spektrum an literarischen Strömungen, welches wiederum starken Einfluss auf poetologische Überlegungen der Autor\*innen ausübt.<sup>143</sup> Rollenvorstellungen ändern sich ebenso wie die an die Figurentypen geknüpfte Abstraktion vom Individuellen.<sup>144</sup> Das Korpus ist zudem weder auf eine spezifische Untergattung, etwa Komödien oder Tragödien, noch auf eine gezielt gewählte Gruppe von Autor\*innen beschränkt. Für die 43 Dramen im Korpus zeichnen sich 22 Autor\*innen verantwortlich. Zudem musste schnell deutlich werden, dass sich eine an literaturwissenschaftlicher Forschung orientierte Gruppierung von Figuren in Typen kaum durch die Fokussierung einzelner quantitativer Methoden, etwa mittels eines netzwerkanalytischen Zugangs alleine, abbilden lassen würde. Denn das war letztlich schon die Pointe, die Manfred Pfisters Ausführungen zur quantitativen Bestimmung von Haupt- und Nebenfiguren subsumierte.<sup>145</sup> Unser umfassender, in mehrere Schritte untergliederter Operationalisierungsansatz unterstreicht diese Beobachtung.

In Anbetracht dessen können die Klassifikationsergebnisse der »Intrigant\*innen« durchaus als Erfolg gewertet werden – gerade da die »Intrigant\*innen« einen recht heterogenen Figurentypus darstellen. So ist Ulfo, der Antagonist in Johann Elias Schlegels Trauerspiel Canut (1746), nicht nur Intrigant, sondern zugleich der nach Macht strebende Gegenspieler des dänischen Königs Canut. Durch eine typische Briefintrige nimmt er Canuts Schwester Estrithe für sich ein. Die Intrige ist jedoch weniger der Auslöser als ein Symptom des dramatischen Konflikts: Ulfo ist geleitet von einem nicht zu brechenden Verlangen nach Ruhm, das seinen Verrat an Canut motiviert. Auch Lessings Intrigantin Marwood (Miß Sara Sampson) vertraut unter anderem (neben Verkleidung, Gift, Dolch) auf den Brief als Mittel der Intrige, sogar gleich mehrfach. Ähnlich wie diejenige Ulfos scheitert Marwoods erste Intrige aber daran, »daß das Laster nur kurze Zeit Verwirrung in die Welt der Tugend bringen kann«, wie Karl Eibl ausführte.<sup>146</sup> Marwood weist aber auch signifikante Unterschiede zu Ulfo auf. Schon ihr Geschlecht unterscheidet die beiden Figuren. Und auch die Motivation ihres Handelns ist eine andere: Marwood versucht ihren ehemaligen Geliebten Mellefont – sie haben eine gemeinsame Tochter – zurückzugewinnen. Der eigentlich zentrale dramatische Familienkonflikt zwischen William Sampson und seiner Tochter Sara ist Marwoods Intrigen vorgelagert und bietet diesen überhaupt erst das Einfallstor. Während sowohl Ulfo als auch Marwood für ihre persönlichen Interessen eintreten, ist die eigene Motivation des Kammerherren Marinelli eher sekundär. Doch von der »Figur des Anstifters, der die bösen Triebe im Herrscher aktiviert«,<sup>147</sup> wird er selbst zur dynamischen Figur, die nicht mehr nur »ein kleines heilsames Verbrechen« einfädelt,<sup>148</sup> sondern für »eine fundamentale Destabilisierung geregelter Verhältnisse« sorgt.<sup>149</sup>

---

<sup>141</sup> Freytag 1920, S. 595.

<sup>142</sup> Freytag 1920, S. 572.

<sup>143</sup> Etwa die Unterscheidung von Auftritt und Szene.

<sup>144</sup> Vgl. dazu die Ausführungen von Alt zum Intriganten: Alt 2004, passim.

<sup>145</sup> Vgl. Pfister 2001, S. 226f. Eine ähnliche Erfahrung haben wir bereits bei der Klassifikation von Titelfiguren gemacht. Vgl. Krautter / Pagel 2019, passim.

<sup>146</sup> Eibl 1971, S. 144.

<sup>147</sup> Alt 2004, S. 7.

<sup>148</sup> Lessing 2000, S. 342.

<sup>149</sup> Alt 2004, S. 8.

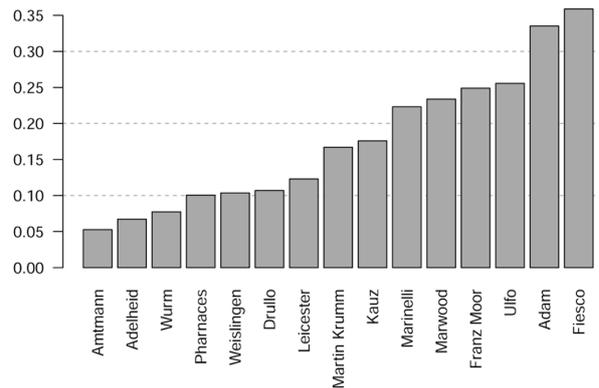


Abb. 6: Geäußerte Tokens der Intrigant\*innen; anhand der gesamten Figurenrede des jeweiligen Dramas normalisiert. [Krautter et al. 2020]

Die Herausforderungen im Umgang mit diesem wandelbaren Figurentypus lässt sich exemplarisch anhand der TOKENS erörtern, die die Figuren im Laufe der Dramen äußern. Während diesem feature für die Erkennung von Protagonisten und Titelfiguren plausiblerweise eine zentrale Funktion zukommt,<sup>150</sup> rangieren die TOKENS als Feature bei der Klassifizierung von »Intrigant\*innen« gerade noch im oberen Drittel der nach feature importance sortierten Features. **Abbildung 6** verdeutlicht, wie verschieden die Redeanteile der einzelnen Intrigant\*innen am jeweiligen Dramentext ausfallen. Gemessen in TOKENS nehmen Adam (Der zerbrochene Krug (1808)) und Fiesco (Die Verschwörung des Fiesco zu Genua (1783)) weit über 30 Prozent der Redezeit im Drama ein. Der Spielraum von Amtmann von Zeck (Die Jäger (1785)), Adelheid (Götz von Berlichingen (1773)) und Wurm (Kabale und Liebe (1784)) ist dagegen sehr viel eingeschränkter – ihre Redeanteile machen noch nicht einmal zehn Prozent der gesamten Figurenrede aus.

Trotz der schlaglichtartig skizzierten Unterschiede in der Figurenkonzeption gelingt es, mittels der in unserer Operationalisierung vorgeschlagenen Merkmalskombinatorik, »Intrigant\*innen« mit einem F1-Wert von 0,7 zu identifizieren. Die Erkennung »nicht-intrigant« Figuren in unserem Korpus ist sogar nahezu fehlerfrei (F1: 0,98).<sup>151</sup> Werden einzelne Features isoliert betrachtet, sind sie in unserem Zuschnitt der Untersuchungen kaum in der Lage, aussagekräftige Erkenntnisse beizusteuern.<sup>152</sup> Die Analysen lassen sich jedoch auf einem hohen Level kombinieren und sind so imstande, Figurengruppen zu differenzieren. Da sich hierbei aber ein Nexus aus einer ganzen Reihe an miteinander verflochtenen und in Abhängigkeitsverhältnissen stehenden Figureneigenschaften ausbildet, ist die Interpretation der Datenwerte keineswegs trivial.

Lässt sich diese Komplexität der Analysen durch ein *Heranzoomen* an die Dramentexte einfangen? Die in unseren Annotationen gewählten »tugendhaften Töchter« und »zärtlichen Väter« stammen in der Hauptsache aus vier Dramen, nämlich Die zärtlichen Schwestern (1747), Miß Sara Sampson (1755), Emilia Galotti (1772) sowie Kabale und Liebe (1784). Aus diesen Dramen lassen sich also vier Figurenpaare bestehend aus »Vater« und »Tochter« bilden, die abschließend näher betrachtet werden sollen. Nimmt man an, dass die Konfliktlinien der Vater-Tochter-Beziehung einerseits durch die Liebesbeziehung der Tochter und andererseits durch die Wahrung der familiären Ordnung durch den Vater abgesteckt werden, müsste sich dies in der Verwendung der Wortfelder *Liebe* und *Familie* widerspiegeln. **Abbildung 7** und **Abbildung 8** schlüsseln die Werte der beiden Wortfeldanalysen für die vier Figurenpaare auf. Tatsächlich lässt sich ein Muster wahrnehmen, das die formulierte Hypothese stützt. Während die »Töchter« eher in Liebessemantiken zu sprechen scheinen, bedienen die »Väter« eher das Familienwortfeld. Emilia, die sich häufiger im Familienwortfeld äußert als ihr Vater Odoardo, und Sir William, den im Vergleich zu Sara ein leicht erhöhter Wert für *Liebe* auszeichnet, sind die Ausnahmen. Diese Beobachtung, das veranschaulichen die Werte der beiden Abbildungen ebenfalls, haben aber einen eingeschränkten Geltungsbereich: Sie sind auf die Figurenpaare einzelner Dramen limitiert. Dramenübergreifende Analysen fallen deutlich heterogener aus. So bedient Sir William das Liebeswortfeld nicht nur in vergleichbarem Maße wie Sara, seine Figurenrede ist überdies stärker mit diesem Wortfeld assoziiert als diejenige Emilias oder Luises.

Während individuelle Features wie die vorgestellten Wortfeldanalysen durchaus in der Lage sind, Einzelphänomene mit spezifisch zugeschnittenen Hypothesen auszuführen, sind sie alleine kaum geeignet, um von diesem Einzelphänomen zu abstrahieren. Ihre Analyseresultate lassen sich also nur schwer extrapolieren.

<sup>150</sup> Vgl. Krautter / Pagel 2019, *passim*.

<sup>151</sup> Da die »nicht-intrigant« Figuren die deutliche Mehrheitsklasse darstellen, ist dieses Ergebnis allerdings weniger überraschend.

<sup>152</sup> Vgl. dazu auch die Klassifikationsergebnisse in Abschnitt 5.3 und die Verteilung der *Feature*-Werte in Abschnitt 5.2.

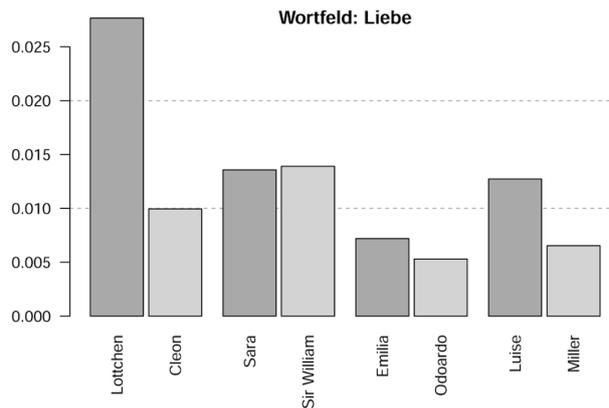


Abb. 7: Verwendung des Wortfelds ›Liebe‹, normalisiert anhand der Länge der Figurenrede. [Krautter et al. 2020]

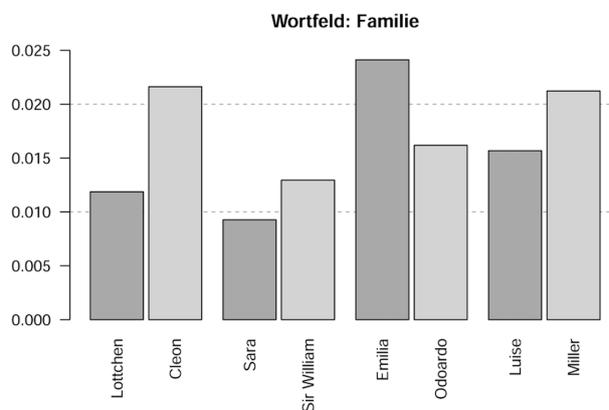


Abb. 8: Verwendung des Wortfelds ›Familie‹, normalisiert anhand der Länge der Figurenrede. [Krautter et al. 2020]

## 6. Fazit und Ausblick

In diesem Beitrag haben wir eine literaturwissenschaftlich motivierte Operationalisierung von Figurentypen vorgenommen und vorgestellt. Dabei gehen wir von einem weiten Operationalisierungsbegriff aus: Sowohl die Entwicklung von Annotationsrichtlinien und -schemata ist eine Art der Operationalisierung als auch die Entwicklung von Programmen zur automatischen Erkennung, z. B. mit maschinellen Lernverfahren. Die Arten der Operationalisierung unterscheiden sich hinsichtlich derjenigen, die sie umsetzen: Annotationsrichtlinien werden von Menschen ausgeführt, Programme von Computern. Dies wiederum zieht nach sich, dass unterschiedlich *atomare* Indikatoren für die Erkennung verwendet werden können.

Unsere Annotationsmethodik unterscheidet sich von anderen Annotationsprojekten,<sup>153</sup> insofern wir nicht nur auf Primärtexten und Annotationsrichtlinien basierende Annotation anstreben. Stattdessen verfolgen wir eine Bottom-up- / Top-down-Doppelstrategie, bei der wir uns vor allem auf existierendes Wissen aus der Literaturgeschichte stützen – Morettis frühem Distant-Reading-Verständnis folgend. Als Problem hat sich dabei herauskristallisiert, dass Literaturgeschichten nur sporadisch einzelne Dramen in ihren analytischen Fokus nehmen. Stattdessen werden größere Entwicklungslinien diskutiert, die durch (kanonische) Beispiele *belegt* werden. Jeder auf diese Weise gewonnene Blick leidet unter den Selektionsmechanismen, die den Literaturgeschichten ohnehin zugrunde liegen. Die Literaturgeschichten haben wir deshalb von Beginn an um Dramenbeziehungsweise Literaturlexika sowie Handbucheinträge zu einzelnen Autor\*innen und Dramen ergänzt. Ein detaillierter Eigenschaftenkatalog, wie er in Abschnitt 3.2 für Sara Sampson abgebildet ist, stellt dennoch eher die Ausnahme in unserer Datengrundlage dar. Nicht selten sind Figuren nur durch wenige Eigenschaften verschlagwortet, das gilt insbesondere für weniger kanonische Dramen und weniger zentrale (Neben-)Figuren der Stücke. Auch die Wahl der annotierten Figuren wird somit durch die Sekundärliteratur präfiguriert. Dadurch ergeben sich Folgeprobleme für die Bottom-up-Annotation: Um eine kohärente

<sup>153</sup> Im *Shared Task SANTA* werden z. B. Erzählebenen in narrativen Texten annotiert. Vgl. Gius et al. 2019, passim. Als Eingabe sollen die Annotator\*innen dabei nur den Text und die Annotationsrichtlinien erhalten und bei der Annotation gerade kein Wissen verwenden, das nicht in den Richtlinien steht.

Annotation der Figuren zu ermöglichen, müssen die herausgebildeten Attribute auf einem eher abstrakten Niveau verbleiben. Sie orientieren sich deshalb auch an den Konventionen der Rollenfächer. Nimmt man diese Einschränkungen als methodische Herausforderung an, können die verwendeten Attribute gleichwohl nutzbar gemacht werden.

Das Annotationsverfahren hat zudem zur Folge, dass der erzeugte Datensatz eher klein ist. Die Notwendigkeit, aus einer verhältnismäßig kleinen Menge (*Small Data*) an Beobachtungen Generalisierungen abzuleiten, ist im Bereich der digitalen Geisteswissenschaften eine weit verbreitete Herausforderung: Aus wenigen Beispielen sollen möglichst generalisierbare Zusammenhänge gelernt werden. In dieses Szenario können quantitative Ansätze, die für *Big-Data*-Szenarien entwickelt wurden, nicht eins zu eins übertragen werden. Unsere oben beschriebene Strategie, um mit dieser Herausforderung umzugehen, ist mehrschrittig: Zunächst verwenden wir interpretierbare Teilbegriffe anstelle einer sogenannten *End-to-End*-Architektur. Dies führt bereits bei der Feature-Extraktion zu einer Abstraktion von der Textebene und damit zu einer Generalisierung. Hier lässt sich auch von Entwicklungen in der Computerlinguistik und maschinellen Lernverfahren profitieren, etwa Operationalisierungen von Teilbegriffen mit großen Datenmengen (wobei dann die Passung der Operationalisierung auf den Gegenstand zu validieren ist). Der zweite Teil der Strategie besteht in der zielgerichteten Zusammenstellung von Datensätzen und insbesondere der Auswahl von Negativbeispielen. Unser Ansatz besteht hier in der Auswahl möglichst sicherer negativer Beispiele: Durch den Ausschluss möglicherweise überlappender Figurengruppen (siehe Abschnitt 5.1) wurde sichergestellt, dass Figuren, die Positivbeispiele sein könnten, aber entweder in der konsultierten Forschung oder in unserem Zugriff (noch) nicht als solche dokumentiert sind, nicht als Negativbeispiele verwendet werden. Der dritte Teil der Strategie besteht in der Inspizierung und Visualisierung der gelernten Zusammenhänge, etwa durch Feature Importance oder die Principle Component Analysis. Irreführende Generalisierungen können auf diese Weise frühzeitig sichtbar gemacht und vermieden werden.

Wie bereits der Titel des Beitrags verdeutlicht, gehen wir von einer starken Prämisse aus, nämlich dass »ein Vater [...] doch immer ein Vater« ist.<sup>154</sup> Weniger bildlich gesprochen heißt dies, dass das meta-dramatische, dem Diener Waitwell von Lessing in den Mund gelegte Diktum, eine Entsprechung in den dramatischen Texten hat: Es gibt so etwas wie auf Textbasis identifizierbare Figurentypen, die auch über die besetzungspragmatische Dimension der Theaterpraxis des 18. Jahrhunderts hinausgehen. Das Ergebnis unserer Analysen stützt diese, von der Literaturwissenschaft natürlich als Konsenswissen bekannte Annahme dahingehend, dass sich »Intrigant\*innen« recht gut automatisch erkennen lassen und dass »zärtliche Väter« und »tugendhafte Töchter« jeweils eine vergleichsweise homogene Teilmenge der »Väter« beziehungsweise »Töchter« bilden. Dabei wurden insbesondere drei Vaterfiguren identifiziert, deren genaue Bewertung noch aussteht. Auf den ersten Blick scheint es zumindest bei Meister Anton aus Hebbels *Maria Magdalene* gute Gründe zu geben, ihn in die Nähe der »zärtlichen Väter« zu rücken. Für die anderen Figuren müsste eine eingehende literaturwissenschaftliche Überprüfung klären, inwieweit sie als »zärtliche Väter« beziehungsweise »tugendhafte Töchter« zu charakterisieren sind. Denn lassen sich solche Befunde hermeneutisch rekonstruieren, wäre ein schönes Beispiel für den Einsatz computationaler Methoden gegeben, der jene Ähnlichkeiten literarischer Textelemente identifizieren kann, die der literaturwissenschaftlichen Forschung noch nicht bekannt sind. Doch auch unabhängig von dieser Rekonstruktion ergeben sich literaturwissenschaftlich relevante Anschlussfragen: Denn selbst, wenn Figuren nicht dem gleichen Figurentypus entsprechen, können strukturelle, thematische oder stilistische Analogien zwischen ihnen intendiert sein und für eine Interpretation herangezogen werden.

## Danksagungen

Die in diesem Artikel vorgestellten Arbeiten wurden freundlicherweise durch die VolkswagenStiftung und ihre Förderung des Projektes *Quantitative Drama Analytics* (QuaDramA) ermöglicht, wofür wir danken. Wir danken außerdem Nora Ketschik und Sandra Schell für die kritische Durchsicht des Artikels sowie Nathalie Schürmann und Martin Glasbrenner für die Annotationsarbeit.

---

<sup>154</sup> Lessing 2003, S. 475.

## Primärliteratur im Text

Die vollständige Liste der für die Analyse verwendeten Dramen lässt sich dem [German Drama Corpus](#) entnehmen.

- Gustav Freytag: Graf Waldemar. Schauspiel in fünf Akten (1847). In: Gustav Freytag: Gesammelte Werke: neue wohlfeile Ausgabe. Serie 1, 8 Bände. Leipzig u. a. 1920. Erste Serie. Bd. 6 (1920): Dramatische Werke: Lustspiele, Trauerspiele, Schauspiele, S. 507–595. [\[Nachweis im GBV\]](#)
- Christian Fürchtegott Gellert: Die zärtlichen Schwestern (1747). In: Christian Fürchtegott Gellert: Werke. Hg. von Gottfried Honnefelder. Frankfurt/Main 1979. Bd. 1 (1979): Einleitung, Fabeln und Erzählungen, Geistliche Oden und Lieder, Moralische Gedichte, Moralische Charaktere, Lustspiele, S. 375–443. [\[Nachweis im GBV\]](#)
- Johann Wolfgang von Goethe: Götz von Berlichingen (1773). In: Johann Wolfgang Goethe: Sämtliche Werke. Briefe, Tagebücher und Gespräche. Hg. von Dieter Borchmeyer u. a. Frankfurt/Main 1985–2013. 1. Abteilung: Sämtliche Werke. Bd. 4 (1985): Dramen 1765–1775, S. 279–389. [\[Nachweis im GBV\]](#)
- Johann Christoph Gottsched: Sterbender Cato (1732). In: Johann Christoph Gottsched: Ausgewählte Werke. Hg. von Joachim Birke / Phillip Marshall Mitchell. 12 Bände. Berlin u. a. 1968–1995. Bd. 2 (1970): Sämtliche Dramen, S. 1–114. [\[Nachweis im GBV\]](#)
- Friedrich Hebbel (1963a): Agnes Bernauer (1851). In: Friedrich Hebbel: Werke. 5 Bände. Hg. von Gerhard Fricke / Werner Keller / Karl Pörnbacher. München 1963–1967. Bd. 1 (1963): S. 679–765. [\[Nachweis im GBV\]](#)
- Friedrich Hebbel (1963b): Maria Magdalena (1844). In: Friedrich Hebbel: Werke. 5 Bände. Hg. von Gerhard Fricke / Werner Keller / Karl Pörnbacher. München 1963–1967. Bd. 1 (1963): S. 301–382. [\[Nachweis im GBV\]](#)
- August Wilhelm Iffland: Die Jäger. Ein ländliches Sittengemälde in 5 Aufzügen (1785). Hg. von Jürg Mathes. Stuttgart 1976. [\[Nachweis im GBV\]](#)
- Heinrich von Kleist: Der zerbrochene Krug (1808). In: Heinrich von Kleist: Sämtliche Werke. Brandenburger Ausgabe. Basel u. a. 1988–2010. Bd. I.3 (1995): Der zerbrochene Krug, S. 7–149. [\[Nachweis im GBV\]](#)
- Friedrich Maximilian Klinger: Die neue Arria. Ein Schauspiel (1776). In: Friedrich Maximilian Klingers dramatische Jugendwerke. Hg. von Hans Berendt / Kurt Wolff. 3 Bände. Leipzig 1913. Bd. 2 (1913): Die neue Arria, Simsons Grisaldo, Sturm und Drang, Pyrrhus, S. 7–124. [\[Nachweis im GBV\]](#)
- Friedrich Maximilian Klinger: Sturm und Drang (1777). Friedrich Maximilian Klingers dramatische Jugendwerke. In drei Bänden. Hg. von Hans Berendt / Kurt Wolff. Leipzig 1912–1913. Bd. 2 (1913): Die neue Arria, Simsons Grisaldo, Sturm und Drang, Pyrrhus, S. 261–355. [\[Nachweis im GBV\]](#)
- Gotthold Ephraim Lessing: Emilia Galotti. Ein Trauerspiel in fünf Aufzügen (1772). In: Gotthold Ephraim Lessing: Werke und Briefe in zwölf Bänden. Hg. von Wilfried Barner / Klaus Bohnen. München 1985–2003. Bd. 7 (2000): Emilia Galotti, Anmerkungen über das Epigramm, Zur Geschichte und Litteratur I/II, S. 291–371. [\[Nachweis im GBV\]](#)
- Gotthold Ephraim Lessing: Miss Sara Sampson. Ein bürgerliches Trauerspiel, in fünf Aufzügen (1755). In: Gotthold Ephraim Lessing: Werke und Briefe in zwölf Bänden. Hg. von Wilfried Barner / Conrad Wiedemann. München 1985–2003. Bd. 3 (2003): Werke 1754–1757. Rettungen, Miss Sara Sampson, Briefwechsel über das Trauerspiel, S. 431–526. [\[Nachweis im GBV\]](#)
- Gotthold Ephraim Lessing: Die Juden (1754). In: Gotthold Ephraim Lessing: Werke. Hg. von Herbert Georg Göpfert. 8 Bände. München 1970–1978. Bd. 1 (1989): Werke 1743–1750, S. 447–488. [\[Nachweis im GBV\]](#)
- Gotthold Ephraim Lessing: Minna von Barnhelm, oder das Soldatenglück (1767) In: Gotthold Ephraim Lessing: Werke. Hg. von Herbert Georg Göpfert. 8 Bände. München 1970–1978. Bd. 6 (1985): Werke 1767–1769, S. 9–110. [\[Nachweis im GBV\]](#)
- Gotthold Ephraim Lessing: Hamburgische Dramaturgie (1767–1769). In: Gotthold Ephraim Lessing: Werke. Hg. von Herbert Georg Göpfert. 8 Bände. München 1970–1978. Bd. 6 (1985): Werke 1767–1769, S. 181–694. [\[Nachweis im GBV\]](#)
- Johann Nestroy: Das Mädl aus der Vorstadt (1767). In: Johann Nestroy: Werke. Hg. von Oskar Maurus Fontana. München 1962. [\[Nachweis im GBV\]](#)
- Johann Gottlob Benjamin Pfeil: Lucie Woodvil (1756). Wien 1767. [\[Nachweis im GBV\]](#)
- Friedrich Schiller: Die Räuber (1781). In: Schillers Werke. Nationalausgabe. Hg. von Julian Petersen / Hermann Schneider / Herbert Stubenrauch. Weimar 1943 ff. Bd. 3 (1953): Die Räuber, S. 1–135. [\[Nachweis im GBV\]](#)
- Friedrich Schiller: Die Verschwörung des Fiesco zu Genua (1783). In: Schillers Werke. Nationalausgabe. Hg. von Norbert Oellers / Siegfried Seidel / Edith Nahler / Horst Nahler. Weimar 1943 ff. Bd. 4 (1983): Die Verschwörung des Fiesco zu Genua, S. 5–121. [\[Nachweis im GBV\]](#)
- Friedrich Schiller: Kabale und Liebe (1784). In: Schillers Werke. Nationalausgabe. Hg. von Norbert Oellers / Herbert Kraft / Claudia Pilling / Gert Vonhoff. Weimar 1943 ff. Bd. 5 (2000): Kabale und Liebe, Semele, Der versöhnte Menschenfeind, Körners Vormittag, S. 5–193. [\[Nachweis im GBV\]](#)
- Friedrich Schiller: Maria Stuart (1800). In: Schillers Werke. Nationalausgabe. Hg. von Norbert Oellers / Nikolas Immer. Weimar 1943 ff. Bd. 9.1 (2010): Maria Stuart, S. 5–180. [\[Nachweis im GBV\]](#)
- Johann Elias Schlegel: Canut (1746). In: Johann Elias Schlegel: Werke. Hg. von Johann Heinrich Schlegel. Frankfurt/Main 1971 [1771]. Bd. 1: Werke. Erster Theil, S. 209–282. [\[Nachweis im GBV\]](#)

## Bibliographische Angaben

- Peter-André Alt: Tragödie der Aufklärung. Eine Einführung. Tübingen u. a. 1994. [\[Nachweis im GBV\]](#)
- Peter-André Alt: Schiller. Leben – Werk – Zeit. 2 Bände. München 2000. Bd. 1. [\[Nachweis im GBV\]](#)
- Peter-André Alt: Dramaturgie des Störfalls. Zur Typologie des Intriganten im Trauerspiel des 18. Jahrhunderts. In: Internationales Archiv für Sozialgeschichte der deutschen Literatur 29 (2004), S. 1–28. [\[Nachweis im GBV\]](#)
- Bernhard Asmuth: Einführung in die Dramenanalyse. 8., aktualisierte und erweiterte Auflage. Stuttgart 2016. [\[Nachweis im GBV\]](#)
- Wilfried Barner: Lessing als Dramatiker. In: Handbuch des deutschen Dramas. Hg. von Walter Hinck. Düsseldorf 1980, S. 106–119. [\[Nachweis im GBV\]](#)
- Wilfried Barner / Gunter E. Grimm / Helmuth Kiesel / Martin Kramer: Lessing. Epoche – Werk – Wirkung. 6. Auflage. München 1998. [\[Nachweis im GBV\]](#)
- Wilfried Barner: »Zu viel Thränen« – »nur Keime von Thränen«. Über Miß Sara Sampson und Emilia Galotti beim zeitgenössischen Publikum. In: Wilfried Barner: »Laut denken mit einem Freunde«. Lessing-Studien. Hg. von Kai Bremer. Göttingen 2017, S. 125–144. [\[Nachweis im GBV\]](#)
- Christian Begemann: Einleitung. In: Realismus. Epoche – Autoren – Werke. Hg. von Christian Begemann. Darmstadt 2007, S. 7–10. [\[Nachweis im GBV\]](#)
- David Blei / Andrew Yan-Tak Ng / Michael Irwin Jordan: Latent Dirichlet Allocation. In: Journal of Machine Learning Research 3 (2002/2003), H. 3, S. 993–1022. PDF. [\[online\]](#) [\[Nachweis im GBV\]](#)
- Peter Josef Brenner: Neue deutsche Literaturgeschichte. Vom »Ackermann« zu Günter Grass. 3., überarbeitete und erweiterte Auflage. Berlin u. a. 2011. [\[Nachweis im GBV\]](#)

- Percy Williams Bridgman: *The Logic of Modern Physics*. New York, NY 1927. [\[Nachweis im GBV\]](#)
- Hans Richard Brittnacher: Sündenbock und Opferlamm. Soziologischer Realismus in Hebbels Agnes Bernauer. In: *Hebbel-Jahrbuch* 51 (1996), S. 77–99. [\[Nachweis im GBV\]](#)
- Hasok Chang: Operationalism. In: *The Stanford Encyclopedia of Philosophy*. Hg. von Edward N. Zalta. Stanford CA 2019. Artikel vom 17.09.2019. [\[online\]](#)
- Nitesh Vijay Chawla / Kevin W. Bowyer / Lawrence O'Higgins Hall / William Philip Kegelmeier: SMOTE: Synthetic Minority Over-sampling Technique. In: *Journal of Artificial Intelligence Research* 16 (2002), S. 321–357. DOI: [10.1613/jair.953](#) [\[Nachweis im GBV\]](#)
- Anke Detken / Anja Schonlau: Das Rollenfach – Definition, Theorie, Geschichte. In: *Rollenfach und Drama*. Hg. von Anke Detken / Anja Schonlau. Tübingen 2014, S. 7–30. [\[Nachweis im GBV\]](#)
- Bernhard Diebold: *Das Rollenfach im deutschen Theaterbetrieb des 18. Jahrhunderts*. Nachdruck. Nendeln 1978 [1913]. [\[Nachweis im GBV\]](#)
- Hans Doerry: *Das Rollenfach im deutschen Theaterbetrieb des 19. Jahrhunderts*. Berlin 1926. [\[Nachweis im GBV\]](#)
- Manfred Durzak: »Außer der Bernauerin ist Niemand naß geworden.« Hebbels problematischer Beitrag zur Geschichte des Bürgerlichen Trauerspiels in Agnes Bernauer. In: *Hebbel-Jahrbuch* 56 (2001), S. 83–102. [\[Nachweis im GBV\]](#)
- Jens Eder / Fotis Jannidis / Ralf Schneider: Characters in Fictional Worlds. An Introduction. In: *Characters in Fictional Worlds. Understanding Imaginary Beings in Literature, Film, and Other Media*. Hg. von Jens Eder / Fotis Jannidis / Ralf Schneider. (Konferenz, Bielefeld, 28.–03.02.2007) Berlin u. a. 2010, S. 3–64. (= Revisionen: Grundbegriffe der Literaturtheorie, 3) [\[Nachweis im GBV\]](#)
- Karl Eibl: *Gotthold Ephraim Lessing: Miss Sara Sampson. Ein bürgerliches Trauerspiel*. Frankfurt/Main 1971. [\[Nachweis im GBV\]](#)
- Monika Fick: *Lessing-Handbuch. Leben–Werk–Wirkung*. 4., aktualisierte und erweiterte Auflage. Stuttgart 2016. [\[Nachweis im GBV\]](#)
- Frank Fischer / Ingo Börner / Mathias Göbel / Angelika Hecht / Christopher Kittel / Carsten Milling / Peer Trilcke: Programmable Corpora – Die digitale Literaturwissenschaft zwischen Forschung und Infrastruktur am Beispiel von DraCor. In: *DHD 2019 Digital Humanities: multimedial & multimodal*. Konferenzabstracts. Hg. von Patrick Sahle. (DHD: 6, Frankfurt am Main u. a., 25.–29.03.2019) Frankfurt/Main 2019, S. 194–197. DOI: [10.5281/zenodo.2596095](#)
- Frank Fischer / Peer Trilcke / Christopher Kittel / Carsten Milling / Daniil Skorinkin: To Catch a Protagonist: Quantitative Dominance Relations in German-Language Drama. In: *Digital Humanities 2018: Puentes-Bridges. Book of Abstracts*. Hg. von Jonathan Girón Palau / Isabel Galina Russell. (DH 2018, Mexico City, 26.–29.06.2018) Mexico City 2018, S. 193–201. PDF. [\[online\]](#)
- Erika Fischer-Lichte: *Geschichte des Dramas. Epochen der Identität auf dem Theater von der Antike bis zur Gegenwart*. 2 Bände. Tübingen u. a. 1990. Bd. 1 (1999): *Von der Antike bis zur deutschen Klassik*. [\[Nachweis im GBV\]](#)
- David Fishelov: Types of Character, Characteristics of Types. In: *Style* 24 (1990), H. 3, S. 422–439. [\[Nachweis im GBV\]](#)
- Edward Morgan Forster: *Aspects of the Novel*. New York, NY 1927. [\[Nachweis im GBV\]](#)
- Evelyn Gius / Nils Reiter / Marcus Willand: A Shared Task for the Digital Humanities. In: *Journal of Cultural Analytics* 2 (2019), H. 1. Artikel vom 01.12.2019. DOI: [10.18419/opus-10884](#)
- Gunter E. Grimm: Miß Sara Sampson. In: *Kindlers Literatur-Lexikon*. Hg. von Heinz Ludwig Arnold. 18 Bände. 3., völlig neu bearbeitete Auflage. Stuttgart u. a. 2009 Bd. 10: Leo–Mar, S. 57–58. [\[Nachweis im GBV\]](#)
- Edward P. Harris: Lessing und das Rollenfachsystem. Überlegungen zur praktischen Charakterologie im 18. Jahrhundert. In: *Schauspielkunst im 18. Jahrhundert. Grundlagen, Praxis, Autoren*. Hg. von Wolfgang F. Bender. Stuttgart 1992, S. 221–235. [\[Nachweis im GBV\]](#)
- Brita Hempel: Sara, Emilia, Luise: drei tugendhafte Töchter. Das empfindsame Patriarchat im bürgerlichen Trauerspiel bei Lessing und Schiller. Heidelberg 2006. [\[Nachweis im GBV\]](#)
- Walter Hinck: *Das deutsche Lustspiel des 17. und 18. Jahrhunderts und die italienische Komödie. Commedia dell'Arte und Théâtre Italien*. Stuttgart 1965. [\[Nachweis im GBV\]](#)
- Baruch Hochman: *Character in literature*. Ithaca, NY u. a. 1985. [\[Nachweis im GBV\]](#)
- Fotis Jannidis: Character. In: *Handbook of Narratology*. Hg. von Peter Hühn et al. 2 Bände. 2. edition, fully revised and expanded. Berlin u. a. 2014. Bd. 1: S. 30–45. [\[Nachweis im GBV\]](#)
- Fotis Jannidis: Typologie3. In: *Reallexikon der deutschen Literaturwissenschaft. Neubearbeitung des Reallexikons der deutschen Literaturgeschichte*. Hg. von Klaus Weimar / Harald Fricke / Jan-Dirk Müller. 3 Bände. Berlin u. a. 2007, Bd. 3: P–Z, S. 712–713. [\[Nachweis im GBV\]](#)
- Fotis Jannidis / Isabella Reger / Markus Krug / Lukas Weimer / Luisa Macharowsky / Frank Puppe: Comparison of Methods for the Identification of Main Characters in German Novels. In: *Digital Humanities 2016. Conference Abstracts*. Hg. von Maciej Eder / Jan Rybicki. (DH 2016, Krakau, 11.–16.07.2016) Krakau 2016, S. 578–582. PDF. [\[online\]](#)
- Ian T. Jolliffe: *Principal Component Analysis*. 2. Auflage. New York, NY 2002. [\[Nachweis im GBV\]](#)
- Gerhard Kaiser: Krise der Familie. Eine Perspektive auf Lessings »Emilia Galotti« und Schillers »Kabale und Liebe«. In: *Recherches Germanique* 14 (1984), S. 7–22. [\[Nachweis im GBV\]](#)
- Folger Karsdorp / Mike Kestemont / Christof Schöch / Antal van den Bosch: The Love Equation: Computational Modeling of Romantic Relationships in French Classical Drama. In: *6th Workshop on Computational Models of Narrative*. Hg. von Mark A. Finlayson / Ben Miller / Antonio Lieto / Remi Ronfard. (CMN'15: 6, 26–28.05.2015, Atlanta). Saarbrücken/Wadern 2015. DOI: [10.4230/OASICS.CMN.2015.98](#)
- Alexander Košenina: Ifflands und Schillers dramatischer Start von Mannheims Bühnenrampe. In: *Mannheimer Anfänge. Beiträge zu den Gründungsjahren des Nationaltheaters Mannheim 1777–1820*. Hg. von Thomas Wortmann. Göttingen 2017, S. 135–150. [\[Nachweis im GBV\]](#)
- Benjamin Krautter / Janis Pagel: Klassifikation von Titelfiguren in deutschsprachigen Dramen und Evaluation am Beispiel von Lessings »Emilia Galotti«. In: *DHD 2019 Digital Humanities: multimedial & multimodal*. Konferenzabstracts. Hg. von Patrick Sahle. (DHD: 6, Frankfurt/Main u. a., 25.–29.03.2019) Frankfurt/Main 2019, S. 160–164. DOI: [10.5281/zenodo.2596095](#)
- Benjamin Krautter / Janis Pagel / Nils Reiter / Marcus Willand: Titelhelden und Protagonisten – Interpretierbare Figurenklassifikation in deutschsprachigen Dramen. In: *Litlab Pamphlet* 7 (2018), S. 1–56. PDF. [\[online\]](#)
- Benjamin Krautter / Janis Pagel / Nils Reiter / Marcus Willand: Properties of Dramatic Characters: Automatic Detection of Gender, Age, and Social Status. In: *Plotting Poetry and Poetics: Computational Stylistics for the Close and Distant Reading of Poetic Texts* (provisional title). Hg. von Anne-Sophie Bories / Peter Plecháč / Pablo Ruiz Fabo. Berlin u. a. 2021. (im Erscheinen)
- Nicolette Kretz: Bausteine des Dramas (Figur, Handlung, Dialog). In: *Handbuch Drama. Theorie, Analyse, Geschichte*. Hg. von Peter W. Marx. Stuttgart u. a. 2012, S. 105–121. [\[Nachweis im GBV\]](#)
- Gotthold Ephraim Lessing: Briefe von und an Lessing, 1770–1776. In: *Gotthold Ephraim Lessing: Werke und Briefe in zwölf Bänden*. München 1985–2003. Hg. von Wilfried Barner / Helmuth Kiesel. Bd. 11,2 (1988): Briefe von und an Lessing, 1770–1776. [\[Nachweis im GBV\]](#)
- Otto Mann: *Geschichte des deutschen Dramas*. Stuttgart 1960. [\[Nachweis im GBV\]](#)
- Peter von Matt: *Ästhetik der Hinterlist. Zur Theorie und Praxis der Intrige in der Literatur*. München 2002. [\[Nachweis im GBV\]](#)
- Sybille Maurer-Schmooch: *Deutsches Theater im 18. Jahrhundert*. Tübingen 1982. [\[Nachweis im GBV\]](#)
- Urs Hans Mehlin: *Die Fachsprache des Theaters. Eine Untersuchung der Terminologie von Bühnentechnik, Schauspielkunst und Theaterorganisation*. Düsseldorf 1969. [\[Nachweis im GBV\]](#)

- Wolfgang Mertens: Zum Marinelli-Typus vor Lessing. In: Ethik und Ästhetik. Werke und Werte in der Literatur vom 18. bis zum 20. Jahrhundert. Festschrift für Wolfgang Wittkowski zum 70. Geburtstag. Hg. von Richard B. Fisher. Frankfurt/Main u. a. 1995, S. 69–79. [[Nachweis im GBV](#)]
- Peter Michelsen: Zur Entstehung des bürgerlichen Trauerspiels. Einige geistes- und literaturgeschichtliche Vorüberlegungen zu einer Interpretation der »Miß Sara Sampson«. In: Literaturwissenschaft und Geistesgeschichte. Festschrift für Richard Brinkmann. Tübingen 1981, S. 83–98. [[Nachweis im GBV](#)]
- Cornelia Mönch: Abschrecken oder Mitleiden. Das deutsche bürgerliche Trauerspiel im 18. Jahrhundert. Versuch einer Typologie. Tübingen 1993. [[Nachweis im GBV](#)]
- Franco Moretti (2000a): The Slaughterhouse of Literature. In: *Modern Language Quarterly* 61 (2000), H. 1, S. 207–227. [[Nachweis im GBV](#)]
- Franco Moretti (2000b): Conjectures on World Literature. In: *New Left Review* 1 (2000), H. 1, S. 54–68. [[online](#)] [[Nachweis im GBV](#)]
- Franco Moretti: Network Theory, Plot Analysis. In: *Pamphlets of the Stanford Literary Lab* 2 (2011), S. 1–12. PDF. [[online](#)]
- Franco Moretti: »Operationalizing«: or, the Function of Measurement in Modern Literary Theory. In: *Pamphlets of the Stanford Literary Lab* 6 (2013), S. 1–13. PDF. [[online](#)]
- Eric T. Nalisnick / Henry S. Baird: Character-to-Character Sentiment Analysis in Shakespeare's Plays. In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. Hg. von Hinrich Schuetze / Pascale Fung / Massimo Poesio. 2 Bände. (ACL 2013, Sofia, 04.–09.08.2013) Red Hook, NY 2013. Bd. 2: Short Papers, S. 479–483. PDF. [[online](#)] [[Nachweis im GBV](#)]
- Janis Pagel / Nils Reiter: GerDraCor-Coref: A Coreference Corpus for Dramatic Texts in German. In: *Proceedings of the 12th Language Resources and Evaluation Conference*. Hg. von Nicoletta Calzolari / Frédéric Béchet / Philippe Blache / Khalid Choukri / Christopher Cieri / Thierry Declerck / Sara Goggi / Hitoshi Isahara / Bente Maegaard / Joseph Mariani / Hélène Mazo / Asuncion Moreno / Jan Odijk / Stelios Piperidis. (LREC: 12, Marseille, 11.–16.05.2020) Paris 2020, S. 55–64. PDF. [[online](#)]
- Manfred Pfister: Das Drama: Theorie und Analyse. 11., erweiterte und bibliographisch aktualisierte Nachdruck der durchgesehenen und ergänzten Auflage. München 2001 [1988]. [[Nachweis im GBV](#)]
- Axel Pichler / Nils Reiter: Reflektierte Textanalyse. In: *Reflektierte algorithmische Textanalyse*. Hg. von Nils Reiter / Axel Pichler / Jonas Kuhn. Berlin u. a. 2020, S. 43–60. [[Nachweis im GBV](#)]
- Lothar Pikulik: »Bürgerliches Trauerspiel« und Empfindsamkeit. Köln u. a. 1981. [[Nachweis im GBV](#)]
- Elke Platz-Waury: Drama und Theater. Eine Einführung. 5., vollständig überarbeitete und erweiterte Auflage. Tübingen 1999. [[Nachweis im GBV](#)]
- Vladimir Propp: Morphologie des Märchens. Hg. von Karl Eimermacher. München 1972. [[Nachweis im GBV](#)]
- Peter Pütz: Die Zeit im Drama. Zur Technik dramatischer Spannung. Göttingen 1970. [[Nachweis im GBV](#)]
- Nils Reiter / Benjamin Krautter / Janis Pagel / Marcus Willand: Detecting Protagonists in German Plays around 1800 as a Classification Task. In: *Abstracts of EADH: Data in the Digital Humanities*. (EADH 2018, Galway, 07.–09.12.2018) Stuttgart 2018, S. 1–5. DOI: [10.18419/opus-10162](https://doi.org/10.18419/opus-10162)
- Nils Reiter / Marcus Willand: Poetologischer Anspruch und dramatische Wirklichkeit: Indirekte Operationalisierung in der digitalen Dramenanalyse. Shakespeares natürliche Figuren im deutschen Drama des 18. Jahrhunderts. In: *Quantitative Ansätze in den Literatur- und Geisteswissenschaften. Systematische und historische Perspektiven*. Hg. von Toni Bernhart / Marcus Willand / Sandra Richter / Andrea Albrecht. Berlin u. a. 2018, S. 45–76. [[Nachweis im GBV](#)]
- Robert Remus / Uwe Quasthoff / Gerhard Heyer: SentiWS – a Publicly Available German-language Resource for Sentiment Analysis. In: *Proceedings of the 7th International Language Resources and Evaluation Conference*. Hg. von Nicoletta Calzolari / Khalid Choukri / Bente Maegaard / Joseph Mariani / Jan Odijk / Stelios Piperidis / Mike Rosner / Daniel Tapias. (LREC'10, Valetta/Malta, 17.–23.05.2010) Paris 2010. PDF. [[online](#)] [[Nachweis im GBV](#)]
- Paul Rilla: Lessing und sein Zeitalter. 2., unveränderte Auflage. München 1977. [[Nachweis im GBV](#)]
- Shlomith Rimmon-Kenan: Narrative Fiction. Contemporary Poetics. 2. Auflage. London u. a. 2002. [[Nachweis im GBV](#)]
- Anja Schonlau: Rollenfach und Geschlecht im Drama des 18. Jahrhunderts. In: *Der Deutschunterricht* 62 (2010), H. 6, S. 82–87. [[Nachweis im GBV](#)]
- Anja Schonlau: Emotionen im Dramentext. Eine methodische Grundlegung mit exemplarischer Analyse zu Neid und Intrige 1750–1800. Berlin u. a. 2017. [[Nachweis im GBV](#)]
- Christof Schöch: Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama. In: *Digital Humanities Quarterly* 11 (2017), H. 2. [[online](#)]
- Martina Schönenborn: Tugend und Autonomie. Die literarische Modellierung der Tochterfigur im Trauerspiel des 18. Jahrhunderts. Göttingen 2004. [[Nachweis im GBV](#)]
- Jörg Schönert: Zu Geltungsansprüchen und Durchsetzbarkeit von Standards für die Interpretation literarischer Texte. In: *Literatur interpretieren. Interdisziplinäre Beiträge zur Theorie und Praxis*. Hg. von Jan Borkowski / Stefan Descher / Felicitas Ferder / Philipp David Heine. (Tagung, Göttingen, 07.–08.03.2013) Münster 2015, S. 317–337. [[Nachweis im GBV](#)]
- Horst Steinmetz: Die Komödie der Aufklärung. 3., durchgesehene und bearbeitete Auflage. Stuttgart 1978. [[Nachweis im GBV](#)]
- Ingo Steinwart / Andreas Christmann: Support Vector Machines. New York, NY 2008. [[Nachweis im GBV](#)]
- Inge Stephan: »So ist die Tugend ein Gespenst«. Frauenbild und Tugendbegriff im bürgerlichen Trauerspiel bei Lessing und Schiller. In: *Lessing Yearbook* 17 (1985), S. 1–20. [[Nachweis im GBV](#)]
- Claudia Stockinger: Miß Sara Sampson. Ein bürgerliches Trauerspiel, in fünf Aufzügen. In: *Dramenlexikon des 18. Jahrhunderts*. Hg. von Heide Hollmer / Albert Meier. München 2001, S. 205–207. [[Nachweis im GBV](#)]
- Ludwig Storch: Charakter-Rollen. In: *Allgemeines Theater-Lexikon oder Encyclopädie alles Wissenserwerthen für Bühnenkünstler, Dilettanten und Theaterfreunde*. Hg. von Robert Blum / Karl Herloßsohn / Herrmann Marggraff. 7 Bände. Altenburg / Leipzig 1839–1842. Bd. 2 (1839): Boulanger bis Devise, S. 122. [[Nachweis im GBV](#)]
- Rhea Sukthanker / Soujanya Poria / Erik Cambria / Ramkumar Thirunavukarasu: Anaphora and Coreference Resolution: A Review. In: *Information Fusion* 59 (2020), S. 139–162. PDF. [[online](#)] [[Nachweis im GBV](#)]
- Gisbert Ter-Nedden: Der fremde Lessing. Eine Revision des dramatischen Werks. Hg. von Robert Vellusig. Göttingen 2016. [[Nachweis im GBV](#)]
- Gisbert Ter-Nedden: Lessings Trauerspiele. Der Ursprung des modernen Dramas aus dem Geist der Kritik. Stuttgart 1986. [[Nachweis im GBV](#)]
- Friedrich Vollhardt: Gotthold Ephraim Lessing. Epoche und Werk. Göttingen 2018. [[Nachweis im GBV](#)]
- Stanley Wasserman / Katherine Faust: Social Network Analysis. Cambridge u. a. 1994. [[Nachweis im GBV](#)]
- Thomas Weitin / Thomas Gilli / Nico Kunkel: Auslegen und Ausrechnen. Zum Verhältnis hermeneutischer und quantitativer Verfahren in den Literaturwissenschaften. In: *Vom Experiment zur Neuorientierung: Forschungswege der Germanistik im 21. Jahrhundert*. Hg. von Ana Maria Bernardo / Fernanda Mota Alves / Ana Margarida Abrantes. Berlin 2017, S. 161–177. [[Nachweis im GBV](#)]
- Marcus Willand / Nils Reiter: Geschlecht und Gattung: Digitale Analysen von Kleists »Familie Schroffenstein«. In: *Kleist-Jahrbuch* (2017), S. 177–195. [[Nachweis im GBV](#)]
- Susanne Winter: Von der Maske zur Rolle, vom Magnifico zum Familienvater. Die Fächerrezeption der Commedia dell'arte am Beispiel des Pantalone. In: *Rollenfach und Drama*. Hg. von Anke Detken / Anja Schonlau. Tübingen 2014, S. 33–47. [[Nachweis im GBV](#)]
- Wolfgang Wittkowski: Väter und Töchter im bürgerlichen Drama. In: *Über deutsche Dichtungen*. Hg. von Wolfgang Wittkowski. 8 Bände. Frankfurt/Main u. a. 2002–2015. Bd. 7 (2013): Hausväter im Drama Lessings und des Sturm und Drangs, S. 17–47. [[Nachweis im GBV](#)]
- Marcus Willand / Benjamin Krautter / Janis Pagel / Nils Reiter: Passive Präsenz tragischer Hauptfiguren im Drama. In: *DHd 2020. Spielräume: Digital Humanities zwischen Modellierung und Interpretation*. Konferenzabstracts. Hg. von Christof Schöch. (DHd: 7, Paderborn, 02.–06.03.2020) Paderborn 2020, S. 177–181. DOI: [10.5281/zenodo.3666690](https://zenodo.org/record/3666690)

## Abbildungslegenden und -nachweise

Abb. 1: Tab. 1: Brandes' Auflistung der Rollenfächer. [Harris 1992, S. 231]

Abb. 2: Tab. 2: Eigenschaftenkatalog von Sara Sampson. [Krautter et al. 2020]

Abb. 3: Tab. 3: Beispiele für die Bottom-up-Annotation. [Krautter et al. 2020]

Abb. 4: Tab. 4: Übersicht der ausgewählten »tugendhaften Töchter«. [Krautter et al. 2020]

Abb. 5: Tab. 5: Übersicht der ausgewählten »zärtlichen Väter«. [Krautter et al. 2020]

Abb. 6: Tab. 6: Übersicht der ausgewählten »Intrigant\*innen«. [Krautter et al. 2020]

Abb. 7: Hierarchische Operationalisierung von Figurentypen in Dramen. [Krautter et al. 2020]

Abb. 8: Tab. 7: Die häufigsten zehn Verben in Figurenrede (links) beziehungsweise Bühnenanweisungen (rechts), die als Approximation an Handlung eingesetzt werden. [Krautter et al. 2020]

Abb. 9: Tab. 8: Auflistung aller im Folgenden verwendeten Features, gruppiert nach den Teilbegriffen wie sie in Abbildung 1 verwendet werden. [Krautter et al. 2020]

Abb. 10: Tab. 9: Übersicht über den Datensatz und seine Verwendung in den Experimenten. »SVM« bezeichnet das Klassifikationsexperiment (Abschnitt 5.3), »PCA« das Experiment zur Bildung der Teilmengen (Abschnitt 5.4). Die Spalte »Exkludierte Charakterisierungen« beschreibt, welche Eigenschaften eine Figur nicht haben darf, damit sie als Negativbeispiel verwendet werden kann. [Krautter et al. 2020]

Abb. 11: Verteilung der Feature-Werte für die drei Figurentypen »tugendhafte Tochter« (t. T., orange), »zärtlicher Vater« (z. V., rot) und »Intrigant\*in« (I., blau). Alle Feature-Werte werden hier numerisch repräsentiert. [Krautter et al. 2020]

Abb. 12: Tab. 10: Ergebnisse der automatischen Klassifikation. Gezeigt ist die positive Klasse, als Negativbeispiele dienen jeweils spezifisch ausgewählte Figurengruppen (siehe Abschnitt 5.1). [Krautter et al. 2020]

Abb. 13: Klassifikationsergebnisse für die Erkennung von Intrigant\*innen für einzelne Gruppen von Features. Die Einteilung folgt den in Abschnitt 4.1 diskutierten Dimensionen. [Krautter et al. 2020]

Abb. 14: Visualisierung der Principle Component Analysis für den Figurentyp »tugendhafte Tochter«. Gezeigt sind die X-Y-Diagramme für jeweils zwei Hauptkomponenten (PC1–3). [Krautter et al. 2020]

Abb. 15: Visualisierung der Principle Component Analysis für den Figurentyp »zärtlicher Vater«. Gezeigt sind die X-Y-Diagramme für jeweils zwei Hauptkomponenten (PC1–3). [Krautter et al. 2020]

Abb. 16: Geäußerte Tokens der Intrigant\*innen; anhand der gesamten Figurenrede des jeweiligen Dramas normalisiert. [Krautter et al. 2020]

Abb. 17: Verwendung des Wortfelds »Liebe«, normalisiert anhand der Länge der Figurenrede. [Krautter et al. 2020]

Abb. 18: Verwendung des Wortfelds »Familie«, normalisiert anhand der Länge der Figurenrede. [Krautter et al. 2020]