

Annäherungen an eine digitale Semiotik: Zwischen computergestützter Semiotik und Semiotik als Metadisziplin der digitalen Literaturwissenschaften*

Julia Nantke, Bergische Universität Wuppertal

Summary. From a literary studies point of view, the paper approaches the potential of digital semiotics from two angles: Based on the assumption of basic similarities between semiotic and computer-assisted methods (sign-theoretical approach, tendency towards model-building), it on the one hand examines the potentials and constrains of the computer-assisted methods for semiotic approaches to literary texts. Here, both convergences and challenges will be addressed in a sign-theoretical perspective. On the other hand, semiotics is regarded as a meta-discipline that is able to bridge the gap between digital and analogue methods and practices in literary studies.

Zusammenfassung. Der Beitrag nähert sich aus literaturwissenschaftlicher Perspektive in zweifacher Weise einer digitalen Semiotik an: Ausgehend von grundlegenden Gemeinsamkeiten zwischen semiotischen und computergestützten Methoden (zeichentheoretischer Ansatz, starke Tendenz zur Modellbildung) wird zum einen nach den Möglichkeiten und Grenzen der Nutzung computergestützter Methoden für semiotische Zugriffe auf literarische Texte gefragt. Hierbei werden sowohl Konvergenzen als auch Herausforderungen zeichentheoretisch beleuchtet. Zum anderen wird die Semiotik als Metadisziplin perspektiviert, die einen relevanten Beitrag zur Vermittlung zwischen digitalen und analogen Methoden und Praktiken der Literaturwissenschaften leisten kann.

1. Einleitung und Verortung der Untersuchungsperspektive

Jede geisteswissenschaftliche Methode modelliert den durch sie betrachteten Gegenstand, indem bestimmte Eigenschaften hervorgehoben, andere hingegen unberücksichtigt bleiben oder zumindest in den Hintergrund treten. Jeder geisteswissenschaftliche *turn* der vergangenen Jahrzehnte ermöglichte daher einerseits die Fokussierung zuvor vernachlässigter Aspekte der betrachteten Gegenstände, ging aber andererseits jeweils mit

neuen blinden Flecken in Bezug auf andere Bereiche einher. An diese Beobachtung anschließend perspektivierte Aleida Assmann in ihrem Beitrag auf der Jahrestagung der Deutschen Gesellschaft für Semiotik im September 2017 die Semiotik als Metadisziplin, welche eben jene Verschiebungen von methodologischen Rahmungen und die dabei jeweils entstehenden Leerstellen zu beobachten und zu beschreiben habe.¹ An diese Perspektivierung anschließend wird im vorliegenden Beitrag der Versuch unternommen, sich der Modellierung literarischer Texte im Rahmen der computergestützten Methoden der digitalen Literaturwissenschaften anzunähern. Dies gilt hinsichtlich der durch computergestützte Methoden eröffneten neuen Erkenntnismöglichkeiten, aber auch hinsichtlich ihrer Grenzen und blinden Flecken. Der Terminus der „computergestützten Methoden“ bezieht sich hierbei auf den Umstand, dass der Einsatz von Werkzeugen wie Computerprogrammen, Datenbanken und Algorithmen sowie die Verfahren, auf denen dieser Einsatz basiert (Statistik, Formalisierung, wahrscheinlichkeitstheoretische Ansätze), spezifische methodische Implikationen bergen, die sich von jenen der tradierten analogen Literaturwissenschaften – je nach Ausprägung auf der einen oder anderen Seite – mehr oder weniger stark unterscheiden.

Der folgende Beitrag nimmt das Potential einer digitalen Semiotik² in zweifacher Hinsicht in den Blick: Zum einen erscheint insbesondere die Semiotik als ein methodisches Feld, welches tendenziell in stärkerer Nähe zu den Methoden der digitalen Literaturwissenschaften anzusiedeln ist. Denn letztere setzen dezidiert auf der Ebene des Zeichens an und sind durch eine starke Tendenz zur formalisierenden Modellierung geprägt. Computergestützte Methoden könnten daher für eine semiotisch orientierte Analyse literarischer Texte neue Chancen bieten, um das von Rolf Parr und Jürgen Link für die Semiotik formulierte Ziel einer „möglichst stringenten strukturalen und strukturpragmatischen Modellbildung“ (Link und Parr 1990: 125) zu erreichen. Zum anderen werden im Folgenden im Sinne eines metadisziplinären Zugriffs die im Rahmen digitaler Methoden erfolgenden Modellierungen in semiotischer Perspektive hinsichtlich ihrer Relation zu den modellierten Gegenständen genauer betrachtet sowie die weiteren Möglichkeiten für eine Semiotik als Metadisziplin der Digitalen Literaturwissenschaften ausgelotet.

Bei den digitalen Literaturwissenschaften handelt es sich um ein noch verhältnismäßig junges Feld, das zudem aufgrund stetiger wechselwirkender technischer und methodischer Weiterentwicklungen permanenten Veränderungen unterliegt. Viele aktuelle Projekte zeichnen sich daher durch ihren Experimentcharakter aus und etliche Beiträge und Diskussionen zeugen vom laufenden Prozess der Selbstverortung im Bereich der digitalen Literaturwissenschaften (vgl. u.a. Gold 2012, Terras, Nyhan und Vanhoutte 2013, Klein und Gold 2016). Eine Betrachtung im Rahmen eines (relativ) kurzen Beitrags kann daher keinesfalls in aller Differenziertheit sämtlichen Bewegungen in diesem Feld gerecht werden. Denn im Gegensatz zu den auf Ganzheitlichkeit zielenden Darstellungen von Franco Moretti,

Matthew Jockers und anderen, die als Gründungsakte digitaler Literaturwissenschaften gelten können, zeigt sich in der Praxis, dass weniger von dem einen großen Modell zur digitalen Beschreibung des literarischen Feldes auszugehen ist. Vielmehr existiert eine Vielzahl von Ansätzen zur Modellierung von Zeichenbeziehungen.³ Diese beziehen sich historisch und gattungsmäßig auf unterschiedliche Bereiche der Literatur sowie auf verschiedene Aspekte literarischer Sprache und literarischen Sprechens.⁴ Um diese Vielgestaltigkeit zu betonen, ist im Folgenden konsequent von (digitalen) Literaturwissenschaften die Rede. Nicht einer Opposition zwischen der tradierten und der digitalen Literaturwissenschaft soll hier Vorschub geleistet werden, sondern es sollen verschiedene digitale Zugriffe auf ihre zeichentheoretischen Implikationen und ihr Verhältnis zu semiotischen Analysepraktiken befragt werden. Ziel dieses Beitrags ist es daher, orientiert an den Grundzügen digitaler Methoden, nach den Möglichkeiten und Grenzen für eine digitale Semiotik zu fragen und dabei so differenziert wie möglich die pluralen Entwicklungen der digitalen Literaturwissenschaften einzubeziehen.

Im Sinne der Orientierung an grundsätzlichen Vorgehensweisen der digitalen Literaturwissenschaften werden zwei methodische Ansätze heuristisch getrennt betrachtet, die sich in der Gegenüberstellung als quantitative bzw. qualitative Methoden beschreiben lassen, die wiederum jeweils vielfältige Modellierungspraktiken umfassen (vgl. zu dieser Unterscheidung auch Schöch 2017: 279). Diese Gegenüberstellung stellt eine Reduktion dahingehend dar, dass quantitative Methoden immer in einem gewissen Maße auch qualitative Zuschnitte (in Form von Korpusauswahl, Untersuchungsparametern, Vorannahmen usw.) bedingen und im Ergebnis auf qualitative Aussagen über literarische Korpora zielen. Umgekehrt können die sukzessive Vergrößerung des Untersuchungskorpus, die Akkumulation von Analyseergebnissen und die Automatisierung von (zunächst manuell ausgeführten) Annotationsprozessen ebenfalls als Ziele qualitativer Methoden gelten. Dennoch erweist sich die getrennte Betrachtung als heuristisch sinnvoll, da sie verschiedene Ausgangspunkte markiert, die grundlegende Auswirkungen auf die Korpusauswahl, den Zugriff auf das Material und die Arbeitsweisen haben. In den folgenden Ausführungen werden aber immer auch die jeweils in die andere Richtung weisenden Tendenzen sowie die Möglichkeiten zu Verschränkungen der beiden methodischen Richtungen berücksichtigt.

2. Quantitative Methoden: Modellierung von Textoberflächen

Die grundlegenden Gemeinsamkeiten verschiedener quantitativer computergestützter Methoden bestehen in der Zugrundelegung (verhältnismäßig) großer Textkorpora sowie der (möglichst) automatisierten Klassifikation von gemeinsamen Eigenschaften der Texte. Theoretisch begründet wurde dieser Ansatz am prominentesten von Franco Moretti (vgl. Moretti 2013). *Dis-*

tant Reading ist zum Schlagwort für die Beschreibung des methodischen Ansatzes geworden, auch wenn computergestützte Verfahren für Morettis in diesem Zusammenhang präsentierte Beispiele kaum eine Rolle spielen (vgl. Trilcke und Fischer 2016).

Vor allem zwei literaturgeschichtliche Argumente begründen dabei das innovative Potential der *Large Scale*-Perspektive: Zum einen ermöglicht die Analysierbarkeit unlesbar großer Textmengen eine Kanonerweiterung sowie die Wiederentdeckung vergessener Literatur, wobei die bisher als repräsentativ rezipierte, maßgeblich der westlichen Kultur entstammende Höhenkammliteratur eine relativierende Einbettung in ihr interkulturelles zeit-räumliches Gefüge erfährt (vgl. hierzu anknüpfend an Moretti z.B. Jannidis 2010: 131; Wilkens 2012). Zum anderen ermöglicht eine systematische Analyse großer Korpora einen alternativen Zugriff, der potentiell bisher unentdeckte Strukturen literarischer Diskurse erfassbar macht (vgl. hierzu z.B. Goldstone und Underwood 2014; Trilcke und Fischer 2016).

Den quantitativen Verfahren gemein – und somit wiederum als methodischer Kern aufzufassen – ist die Tatsache, dass diese zunächst ausschließlich an der Textoberfläche operieren, indem bestimmte Textmerkmale mit Hilfe von Algorithmen auf Signifikantenebene erhoben werden. Während bei der Konzentration auf spezifische Textelemente wie Figurenrede oder Plotstrukturen eher mittelgroße Korpora betrachtet und in diesen sämtliche Fälle erfasst werden, beziehen sich die zentralen Verfahren wie *Topic Modeling* und Stilometrie auf Worthäufigkeiten; in die Analyse einbezogen werden also jeweils die n häufigsten Wörter insgesamt oder bestimmter Wortarten (Nomen, Funktionswörter) in Relation zum Gesamtkorpus. Bei sehr großen Korpora, wie sie beispielsweise für das *Topic Modeling* als Verfahren unüberwachten *machine learnings* unabdingbar sind (vgl. hierzu die Erklärung in Jockers 2013: 123), erfolgen die Modellierungen zudem auf der Basis bedingter Wahrscheinlichkeiten (vgl. Blei 2012).⁵

Die den quantitativen Methoden zugrunde liegende Vorstellung vom Text als Zeichenformation entspricht grundsätzlich einem semiotischen Textverständnis. Die Interpretation der statistisch gewonnenen Ergebnisse wäre dabei mit Peirce als Abduktion zu beschreiben: Durch die Größe der Korpora und die systematische Erfassung der Signifikanten können z.B. im großen Stil Sprach- und teilweise Textstrukturen verglichen und auf diese Weise statistische Aussagen über relationale Parallelen und Differenzen in Bezug auf die beobachteten Einheiten getroffen werden, die allerdings eher „probeweise angenommene[n] (hypothetische[n]) Regel[n]“ (Nöth 2000: 68) entsprechen. Denn die Ergebnisse als „erklärungsbedürftige[] Resultat[e]“ (Nöth 2000: 68)⁶ sind in hohem Maße abhängig von der Korpusbildung, den vorgegebenen Parametern, deren Justierung aufgrund der Diversität des Gegenstands Literatur immer auch Teil der Validierung der verwendeten Programme ist,⁷ und der normalisierenden Angleichung der Werte. Hinzu kommt die kaum vollständig nachvollziehbare Funktionsweise der eingesetzten Algorithmen (vgl. Stalder 2016: 202).⁸ Bei wahrscheinlichkeitsbasierten Operationen vermehren sich die Unsicherheitsfaktoren aufgrund

der zufälligen Auswahl aus dem Zeichenrepertoire. Diese Perspektivierung quantitativer Methodik als abduktive Hypothesenbildung, die „from facts to hypothesis, that is, from the empirical world to an ideal world“ (Stjernfelt 2007: 333) führt, betont gleichermaßen deren innovatives Potential (vgl. Peirce 1960: 90 [§ 145]) und die Notwendigkeit der Überprüfung des aus der Abduktion deduzierten abstrakten Modells mittels induktiver Analysen.⁹

2.1 Abstraktionsgrade und Ebenendifferenzen

Der abduktive Charakter algorithmischer Untersuchungsmethoden bietet einen Anknüpfungspunkt für die Ergänzung digitaler Methoden durch analoge Praktiken der genauen Einzeltextanalyse zur Validierung der Ergebnisse. Dabei kommt allerdings eine Problematik zum Tragen, die durch die Differenz zwischen „data und meaning“ und die „Diskontinuität[,] die zwischen diesen beiden epistemischen Objekten besteht“ (Trilcke und Fischer 2016; vgl. auch Heuser und Le-Khac 2012: 46f.), bedingt ist, und mit der Umwandlung von Texten in Daten im Zuge der Anwendung statistischer Verfahren einhergeht. Diese Diskontinuität lässt sich semiotisch anhand des zeichentheoretischen Zugriffs auf die zu untersuchenden Korpora begründen. Die algorithmische Verarbeitung der Texte kommt einer starken Abstraktion gleich: Die analysierten Korpora unterliegen extremen Reduktionsprozessen durch Zergliederung, *Stopword*-Listen und die durch Algorithmen anhand bestimmter Parameter getroffene Auswahl der Signifikanten. Für die Auswertung werden Signifikanten-Kookkurrenzen in der Regel wiederum auf eine *Type-Token*-Relation reduziert (vgl. Schöch 2017: 280), die digitalen Erhebungen durch Normalisierungen bereinigt und in Form von Diagrammen, Netzwerken, Baumstrukturen und Graphen visualisiert.¹⁰ Diese Operationen resultieren in „weitgehend atomistischen Vorstellungen von den Konstituenten eines Textes“ (Bubenhofer und Scharloth 2015: 13), die eine große Distanz zur semiotischen Perspektivierung eines Textes als komplexes Zeichengewebe und dessen detaillierter Analyse markieren. Diese Distanz verstärkt sich nicht zuletzt durch die maßgeblich visuelle Repräsentation digitaler Analyseergebnisse, in deren Rahmen Relationen zwar einerseits in Form von Tabellen, Graphen oder Netzwerken verdeutlicht werden, in ihren jeweiligen bedeutungstragenden Implikationen andererseits aber unterspezifiziert bleiben.

Jene Abstraktionen gestalten sich in Bezug auf literarische Texte als besondere Herausforderung, denn mit Lotman stellt Literatur ein „sekundäres modellbildendes System“ (Lotman 1993: 39) dar, welches nicht primär Objekt-bezogen funktioniert, sondern von der poetischen Funktion der Sprache dominiert wird (vgl. Jakobson 1960: 356f.; Jakobson 1972: 415).¹¹ Die statistischen Verfahren der quantitativen Analyse operieren hingegen maßgeblich auf der primären Ebene der Lexik.¹² Die hierbei entstehende Kluft lässt sich anhand des Zugriffs auf verschiedene Zeichenebenen als methodologische beschreiben: Die aus der Computerlinguistik übernom-

mene Betrachtung von *Type-Token*-Relationen basiert auf Peirces Vorstellung von Legizeichen und Replik, für die Peirce synonym die Termini *Type* und *Token* einführte (vgl. Peirce 1933: 423 [§ 537]; Nöth 2000: 65), und gerade in dieser Übernahme durch die digitalen Literaturwissenschaften manifestiert sich die angesprochene Problematik: Statistische Verfahren zielen verstärkt auf grundsätzliche Muster. Besonderheiten und Abweichungen bleiben hingegen tendenziell unberücksichtigt, weil aus ihnen keine statistisch signifikanten Regelmäßigkeiten abgeleitet werden können (vgl. Gius und Jacke 2015, Kap. 3.1).¹³ Damit einher geht eine Reduktion auf das Zeichen als Replik eines Legizeichens, womit wiederum dessen Eigenschaft als „general type“ (Peirce 1932: 143 [§ 246]) in den Vordergrund rückt.

Während für computerlinguistische Anwendungen aufgrund ihres maßgeblichen Bezugs auf das primäre modellbildende System der lexikalischen Sprache ein Goldstandard angenommen wird, vor dessen Folie eine Modellierungspraxis *qua* Induktion¹⁴ als (nicht-)funktional qualifiziert werden kann, ist dies insbesondere in Bezug auf literarische Texte nicht in objektivierbarer Form möglich (vgl. Stjernfelt 2007: 335; Klimek und Müller 2015: 70f.).¹⁵ Literarisches Schreiben entfaltet sich maßgeblich in den Dimensionen von Sin- und Qualizeichen (vgl. Peirce 1932: 142 [§ 244 und 245]), indem sich Literatur gerade durch ihren von der Norm abweichenden Umgang mit Sprache und die Komplexität ihrer kommunikativen Funktion auszeichnet, die zu einer „entautomatisierten Beziehung zwischen Sender und Empfänger“ (Lotman 1981a: 109) führen.¹⁶ Auch wenn die Frage der Textsemantik im Rahmen der quantitativen Erhebung an dieser Stelle noch keine Rolle spielt, erweist sich der Umstand, dass *Tokens* aus ihrer ko- und kontextuellen Umgebung herausgelöst betrachtet bzw. auf *Types* reduziert werden und Einzeltextspezifika durch Normalisierung wegfallen, für die ergänzende Kombination von abduktiver Hypothesenbildung und induktiver Überprüfung *qua* Einzeltextanalyse als Herausforderung. Für zentrale literarische Charakteristika wie Perspektiv- und Ebenenwechsel, den Gebrauch von Tropen sowie die Selbstreferentialität und Polysemie literarischer Sprache gibt es keine Entsprechung in den Ergebnissen quantitativer Analysen.

Dies gilt im Prinzip auch für die ersten Ansätze seitens der digitalen Literaturwissenschaften, die Textoberfläche im Rahmen quantitativer Analysen mittels Sentimentanalysen oder Verfahren der Vektorensemantik wie *Word Embedding Models* in Richtung automatisierter qualitativer Auswertungen zu durchstoßen (vgl. zur Sentimentanalyse z.B. Jockers 2014; Jannidis u.a. 2016 und zur Vektorensemantik Mikolov u.a. 2013 sowie mit unmittelbar literaturwissenschaftlichem Bezug Schöch 2018). Insbesondere diese Ansätze verweisen zwar auf die zu Beginn angesprochene Instabilität der Trennung in quantitative und qualitative Methoden. Indem die mit jenen Ansätzen verbundenen Verfahren allerdings wiederum statistisch auf der Basis vorgegebener lexikalischer Wortbedeutungen bzw. umfangreicher Trainingskorpora arbeiten, besteht hier weiterhin die Aufgabe der Relationierung von primärer Sprach- und sekundärer Ebene literarischer Rede (vgl. Bubenho-

fer und Scharloth 2015: 13f.),¹⁷ wobei aufgrund der Komplexität der Modellierung im n-dimensionalen Vektorraum wiederum ein hoher Abstraktionsgrad herrscht.

Mit Blick auf die aktuellen Untersuchungspraktiken der digitalen quantitativen Methoden lassen sich viele Tendenzen der dargestellten Herausforderung zuschreiben. Viele Projekte knüpfen in der Korpusbildung eng an möglichst stabile, kanonbasiert generierte literaturwissenschaftliche Kategorien an: Eine bestimmte Gattung (Drama, Sonett, Historischer Roman) in einer spezifischen literaturwissenschaftlich begründeten und kulturhistorisch begrenzten¹⁸ Epoche¹⁹ sowie das Gesamtwerk eines Autors oder einer Autorin bilden die zentralen Orientierungspunkte bei der Textauswahl. Automatisierte Modellierungen von Textstrukturen werden also vor allem in Bezug auf inhaltlich und/oder strukturell stark formalisierte Textsorten, insbesondere (klassische) Dramen, Sonette und Kriminalgeschichten vorgenommen, die eine relationale Betrachtung aus der Distanz erleichtern.²⁰ Im Falle von Dramen und Krimis gilt dies auch aufgrund der Handlungslastigkeit der Textsorten. Eine Differenzierung von Erzählebenen bzw. Handlung und erzählerischer Reflexion entfällt hier weitgehend, gemeinsames Auftreten von Figurennamen entspricht meist einer Interaktion. Ebenso sind die Extraktion und Zuordnung von Figurenrede im Drama (relativ) einfach und daher teilweise automatisiert zu erreichen.

Aufgrund der Orientierung an stabilen Kategorien laufen die quantitativen Analysen allerdings leicht Gefahr, in einen Zirkelschluss zu geraten, wenn literaturwissenschaftliche Annahmen etwa über die Struktur bestimmter Dramen als Ausgangspunkt für eine zunächst zu validierende Modellierungspraxis dienen. Kann bei einer Bestätigung der literaturwissenschaftlichen Vorannahmen davon ausgegangen werden, dass das Modell funktioniert bzw. wäre eine Nicht-Bestätigung gleichzusetzen mit neuen Erkenntnissen hinsichtlich der Dramenstruktur (vgl. Jannidis 2010: 126)?

Dieser Gefahr des Zirkelschlusses bzw. der Unentscheidbarkeit bezüglich der Validierung der Analyseergebnisse entzieht sich ein Stück weit die Stilometrie: Hier besteht die Möglichkeit, die erhobenen Merkmale mithilfe autorisierter Texte zu testen, und dabei andere Variablen wie Gattungszugehörigkeit zu kontrollieren.²¹ In Bezug auf diese Praxis wäre angesichts der beschriebenen Herausforderung der Relationierung quantitativer und analog-qualitativer Analyseergebnisse allerdings zu fragen, ob diese nicht auch umgekehrt besteht, inwieweit es also grundsätzlich plausibel ist, maßgeblich interpretativ und kanonbasiert erschlossene Kategorien wie Autor_innenstil, Gattung oder literarische Epochen als Grundlage für quantitative Analysen zu verwenden.²²

2.2 Semiotische Vermittlung zwischen Daten und Bedeutung

Gerade die Semiotik bzw. eine an ihren Methoden orientierte Literaturwissenschaft wäre besonders prädestiniert dafür, Ansätze für eine Überbrü-

ckung der Kluft zwischen Daten und Bedeutung sowie zwischen der klassisch-textuellen und der primär visuellen Repräsentation von Forschungsergebnissen in den digitalen Literaturwissenschaften zu liefern. Hierbei gilt es, jenseits einer Perspektivierung der Korpora quantitativer Analysen als „elektronische Zettelkästen“ (Bubenhofner und Scharloth 2015: 2), die über das rekurrente Auftreten spezifischer Signifikanten Hinweise auf das Emergieren bestimmter diskursiver Formationen in literarischen Texten liefern und umgekehrt Prozesse der textuellen Einschreibung in bestimmte Diskurse sichtbar machen, nach Möglichkeiten der Übersetzung zu suchen: Wie lassen sich Topics, verstanden als „a representation of ‚discourse‘ (language as it is used and as it participates in recognized social forms)“ (Rhody 2012: 29f.) in ein Verhältnis setzen zur „Tendenz literarischer Diskurse zur Multiplikation konnotativer Signifikate (d.h. zur Polysemie bzw. Polyisotopie)“ (Link und Parr 1990: 114)? Wie ist die Beziehung zwischen in Netzwerkanalysen modellierten Dramenstrukturen (vgl. Trilcke u.a. 2016) und den Untersuchungsergebnissen einer analogen Analyse zeichentheoretisch zu beschreiben? Und (wie) lässt sich die über eine Sentimentanalyse generierte Definition des Textendes semiotisch nachvollziehen? Diese und ähnliche Fragen umkreisen den von Lotman beschriebenen „Zustand ständiger Spannung und wechselseitiger Übersetzung (und zugleich nicht restloser gegenseitiger Übersetzbarkeit)“ (Lotman 1981b: 115) zwischen primärer und sekundärer Sprachebene.

Wenn eine semiotische Systematisierung dieser Zusammenhänge gelänge, wäre hier auch ein Ansatzpunkt für eine induktive Validierung abduktiver Hypothesen sowie für die Nutzung der Ergebnisse quantitativer Methoden für die Detailanalyse einzelner Texte gegeben.²³ Insbesondere die Ansätze zur automatisierten qualitativen Auswertung könnten Ausgangspunkte für zeichentheoretisch fundierte Übersetzungsversuche eröffnen, indem hier anhand von Wörterbüchern und umfangreichen Testkorpora historisch-kulturell spezifische Denotationen sichtbar gemacht werden, die sich beispielsweise in Beziehung zur Untersuchung „paradigmatisch expandierende[r] Zeichen“ (Link und Parr 1990: 116)²⁴ setzen ließen. Auf diese Weise ließen sich in der Kombination computergestützter und tradierter Analysepraktiken potentiell neue Einsichten in das Wechselverhältnis zwischen natürlicher Sprache und literarischem Sprechen gewinnen.²⁵

Umgekehrt kann eine semiotische Perspektive den Blick für Forschungsfragen schärfen, bei denen die beschriebenen Ebenendifferenzen weniger oder gar nicht ins Gewicht fallen. Dadurch wird der Abstraktionsgrad der quantitativ generierten Daten verringert und entsprechend die Kombination computergestützter und analoger Techniken der Textanalyse deutlich erleichtert.

So sind zum einen nicht alle sprachlichen Strukturen in literarischen Texten gleichermaßen von den Spezifika des sekundären modellbildenden Systems affiziert und eine Beschränkung quantitativer Analysen auf spezifische Charakteristika ermöglicht zudem einen zielgerichteten Nachvollzug der generierten Daten. Deshalb ist davon auszugehen, dass beispiels-

weise eine Erfassung der räumlichen Verhältnisse in fiktionalen Welten mithilfe der Erhebung von „Ortsmarkern“ sowie der „Anwendung von computerlinguistischen Methoden der Relationsextraktion“ (Barth und Viehhauser 2017) weniger mit den Herausforderungen der Ebenendiskrepanz konfrontiert ist, als ein auf die ganzheitliche Erfassung textueller Zusammenhänge gerichtetes *Topic Modeling*.

Zum anderen bestehen insbesondere in semiotischer Perspektive literaturwissenschaftliche Forschungsfragen, die sich dezidiert auf die Signifikantenebene beziehen. Dies betrifft etwa die poststrukturalistischen Vorstellungen vom Text als Schnittpunkt von Diskursen und vom „Papierautor“ (Barthes 2005: 47), der lediglich eine Funktion des Textes darstellt (vgl. Foucault 2000: 210) und dessen „Schreibweise“ (Barthes 2005: 45) sich als spezifische Kombinatorik bekannter Codes ausweist. Ausgangspunkt wären hierbei gerade nicht die instabilen Kategorien der stilistischen Einheit eines Autors/einer Autorin oder einer Gattung, sondern das auf der Ebene von Signifikantenkomplexen emergierende Netz sprachlicher Überschneidungen.²⁶

Ebenso zeigen erste Untersuchungen an Texten der klassischen Avantgarden, wie die maschinelle Signifikantenerhebung (in einer Erweiterung von reinen Types auf größere Textabschnitte) hilfreich bei der Erkennung strukturierender Muster sein kann, die im Zuge einer analogen Lektüre schwierig aufzufinden wären (vgl. Clement 2008). Indem sich die Strukturierung von stark konzeptuell organisierten Avantgardetexten und konkreter Poesie häufig maßgeblich anhand von Rekurrenzen auf Signifikantenebene vollzieht, können quantitative Analysen gerade hierfür Möglichkeiten der systematischen Erschließung von Textstrukturen als Montage-Muster bereitstellen. Das Potential des Computers zum *Close Reading* könnte auf diese Weise dazu beitragen, einer jenseits von inhaltlicher Geschlossenheit und kohärenter Sinnstiftung funktionierenden Bedeutungsbildung auf die Spur zu kommen (vgl. Weel 2015; Clement 2008). Ein ähnliches Vorgehen, welches auf der spezifischen Ausrichtung quantitativer Methoden auf die Entdeckung von Mustern auf der Sprachoberfläche basiert, könnte auch für weitere, maßgeblich auf Signifikantenebene analytisch zugängliche Rekurrenz-Phänomene wie beispielsweise Topoi neue Erkenntnisse erbringen. Quantitative Methoden können hierbei im Sinne einer induktiven Verifikation oder auch Falsifikation der anhand von Einzeltexten oder kleineren Korpora erarbeiteten Muster fungieren, die in analogen Studien lediglich deduktiv auf übergeordnete Systeme wie Werk, Gattung, Epoche usw. übertragen werden (vgl. Braun 2013: 87f.).

Diese letzten Überlegungen illustrieren abschließend nochmals die Vielfalt der Einsatzmöglichkeiten quantitativer Methoden und verweisen auf die Notwendigkeit, bei der Ermittlung von Konvergenzen mit analogen Analysepraktiken die um technische Faktoren vervielfältigten Untersuchungsparameter – methodischer Zugriff, eingesetzte Werkzeuge, veranschlagte Parameter, Operationen im Zuge der Präprozessierung, erhobene Merkmale, Untersuchungsgegenstände – und deren Relationen zu berücksichtigen.

3. Qualitative Methoden: Semantische Annotationen und semantische Netze

Annotationen in den Literaturwissenschaften stellen ein weites Feld dar, das von der Auszeichnung grundlegender Textstrukturen wie Überschriften, Kapitel, Zeilenumbrüche und *Named Entities* z.B. mittels XML bzw. TEI²⁷ über stärker kontextualisierende Informationen bis hin zu interpretativen Zuschreibungen reicht. Hierbei sind die Grenzen mindestens zwischen einer kontextualisierenden Annotation, die etwa in einer kritischen Edition der Klärung eines historischen Kontextes dient, und einer dezidiert interpretierenden Annotation, die eine Textstelle beispielsweise als ironisch gemeint deklariert, keinesfalls eindeutig zu ziehen.

Die folgende Betrachtung bezieht sich explizit auf solche Methoden, die Annotationen systematisch zur Modellierung spezifischen literaturwissenschaftlichen Wissens in Bezug auf literarische Texte einsetzen und die Texte entsprechend im Sinne eines durch ein Annotationsschema bzw. entsprechende *Guidelines* ausgedrückten heuristischen Fokus strukturieren (vgl. Piez 2010; Zirker und Bauer 2017: 147).

Im Gegensatz zu den bisher verhandelten quantitativen Methoden, die zunächst an der Textoberfläche operieren, steht für die semantischen Annotationen von vornherein die Bedeutung von Texten und Textteilen im Fokus (vgl. Meister 2012: 105). Das gilt ein Stück weit auch für die bereits verhandelten Ansätze zur Verknüpfung quantitativer und qualitativer Erhebungen. Indem hier teilweise ebenfalls Testkorpora händisch annotiert werden bzw. umgekehrt die im Folgenden dargestellten Ansätze teilweise das (Fern-) Ziel einer Automatisierung der Annotation verfolgen, werden erneut die fließenden Übergänge zwischen den Ansätzen sichtbar. Dennoch sei hier eine zentrale methodische Differenz im Zugriff auf die literarischen Texte hervorgehoben, welche aus der Verschiebung der Zielrichtung hervorgeht und entscheidend für die semiotische Perspektivierung ist.

Bisher ist maßgeblich die Differenz zwischen natürlicher Sprache, wie sie etwa im Lexikon erfasst ist, und Literatur als sekundärem modellbildenden System als Lücke in den dargestellten Operationalisierungs- und Abstraktionsprozessen im Rahmen der quantitativen Methoden hervorgetreten. In Bezug auf die qualitativen digitalen Methoden erfährt das dabei referenzierte Beziehungsgefüge von Sprach- und Sprecherebenen gleichermaßen eine Verschiebung und eine Erweiterung. Zum einen ermöglichen semantische Annotationen die strukturierende Anreicherung eines Textes bzw. Textkorpus mit formalisierten Informationen, die sich auf die Ebene des sekundären modellbildenden Systems beziehen. So können – hauptsächlich im Rahmen manueller Annotationen – z.B. narrative Strukturen auf verschiedenen Erzählebenen sowie Tropen und Polyisotopien in Bezug auf ihren spezifischen Status im jeweiligen Text systematisch verzeichnet, Handlung und erzählerischer Diskurs differenziert werden.

Zum anderen implementieren die semantischen Annotationen und das hinter diesen stehende Kategorisierungsschema quasi eine tertiäre Text-

ebene, welche spezifischen Regeln und Restriktionen bezüglich der Repräsentation von Wissensstrukturen folgt und sich daher ebenfalls maßgeblich durch ihre modellbildenden Qualitäten auszeichnet.²⁸ Neben der Analyse konkreter Korpora stellt hierbei insbesondere die Taxonomie als aus den Annotationen abstrahiertes Modell ein zentrales Ergebnis semantischer Annotationsverfahren dar, welches als formalisiertes Methodengerüst zur Nachnutzung in weiteren Analysen angelegt ist.

3.1 Formalisierung literaturwissenschaftlicher Perspektiven

Diese beiden Aspekte der Verortung und Funktionsweise semantischer Annotationen haben verschiedene Auswirkungen auf die Frage nach den Potentialen der Wechselwirkung von digitalen Methoden und semiotischen Ansätzen.

Die Systematik semantischer Annotationen ermöglicht eine Modellierung von Signifikant-Signifikat-Beziehungen auf Einzelzeichen- oder Zeichenformationsebene sowie von strukturellen Eigenschaften literarischer Texte. Jene Modellierung ist zwar formalisiert und damit stringent, kann dabei aber in Abhängigkeit von der Anlage des jeweiligen *Frameworks* ebenfalls flexibel in Bezug auf überlappende oder widerstreitende Bedeutungszuschreibungen sein.²⁹ Die Modellierungsprozesse zeichnen sich zudem durch Rückkopplungsschleifen zwischen Annotation und Modellbildung aus, wodurch in der Modellierung Abweichungen und Einzeltextspezifika erfasst werden können (vgl. Schöch 2013: 5; Gius und Jacke 2017: 241f.).

Indem im Rahmen semantischer Annotationen formalisierte Kategorien jeweils aus einer an tradierten literaturwissenschaftlichen Maßgaben orientierten Analysepraxis heraus entwickelt werden, ist diese Form der digitalen Modellierung zunächst in deutlich geringerem Maße von jenen Übersetzungsschwierigkeiten zwischen analogen und digitalen Literaturwissenschaften geprägt, als dies für die statistischen Methoden der quantitativen Analyse dargestellt wurde.³⁰ Die entwickelten *Guidelines* bilden hier eine Referenz für die vergleichende Betrachtung digitaler Modellierung mit literaturwissenschaftlichen Methoden der Texterschließung.

Dabei knüpfen digitale Annotationskonzepte insbesondere an solche literaturwissenschaftlichen Methoden an, die bereits eine Tendenz zur formalisierenden Erfassung von Zeichenbeziehungen und damit eine starke Anbindung interpretierender Aussagen an ein relativ stabiles Gerüst analytischer Kategorien aufweisen.³¹ Dies verdeutlichen verschiedene Ansätze zur Modellierung narratologischer Strukturen.³² Parallel existieren aber auch *Frameworks*, die in stärkerer Unabhängigkeit von spezifischen methodischen Vorgaben Annotationskategorien entwerfen.³³ Die resultierende Taxonomie stellt entsprechend eine formalisierte Version des interpretativen Prozesses bzw. der dabei adressierten Textebenen dar.

Semantische Annotationen können also gewissermaßen als formalisierende Überführung tradierter literaturwissenschaftlicher Methoden in maschi-

nenlesbare und -auswertbare Form angesehen werden, die – ähnlich dem *Close Reading* mittels Werkzeugen zur quantitativen Analyse – in der Systematisierung neue Perspektiven auf die Gegenstände erlaubt, dabei aber gleichermaßen die Möglichkeiten zur Modellierung jeweils unterschiedlicher Perspektiven, also von Polysemie offenhält. Dabei trägt die für die Funktionalität der Modellierung unabdingbare Notwendigkeit zur Systematisierung dazu bei, Unschärfen und Inkonsistenzen in analogen Modellen und methodischen Konzepten aufzudecken (vgl. hierzu Gius und Jacke 2017: 243f.).³⁴ In diesem Sinne reduzieren semantische Annotationen den Abstraktionsgrad analoger Modellbildung, indem Kategorien einerseits konkret an eine Vielzahl jeweils relevanter Textstellen angebunden werden, andererseits aber die enge Perspektivierung auf den Einzeltext überwunden wird. Dabei spielt zwar die jeweilige Sichtweise des oder der Modellierenden eine Rolle, dieser Faktor kann jedoch durch die kollaborative Ausrichtung der Projekte im Vergleich mit analogen Praktiken demokratisierend wirken, indem Entscheidungen in einer Fachgemeinschaft ausgehandelt und Interpretationsvielfalt im Sinne von „agreed disagreement“ (Gius und Jacke 2017: 235, 238) systematisch erfasst werden können.

Dies entspricht grundsätzlich der von Link und Parr vorgegebenen semiotischen Zielsetzung einer strukturalen und strukturpragmatischen Modellbildung.³⁵ Als Elemente einer digitalen Semiotik könnten semantische Annotationen daher beispielsweise die systematische Relationierung verschiedener Textebenen unterstützen oder eine Kategorisierung inner- und intertextueller Zeichenrelationen ermöglichen. Ebenso könnte der Versuch unternommen werden, mehrgliedrige Zeichenmodelle wie jenes von Peirce mithilfe semantischer Annotationen auf konkrete Texte zu übertragen und hierbei potentiell auch deren Unschärfen systematisch zu beschreiben.

3.2 Semiotische Vermittlung zwischen Modellierung und Modell(en)

Die bisherigen Ausführungen verweisen auf die Interaktion von Einflussfaktoren, welchen der Modellierungsprozess im Rahmen semantischer Annotationen unterliegt: Je nachdem anhand welches literaturwissenschaftlichen Methodengerüsts bzw. auf der Basis welcher literarischen Texte und theoretischen Vorannahmen die Annotationskategorien entworfen werden, schreiben sich Perspektiven in die Modelle ein, die sich auch auf die im Rahmen einer Nachnutzung analysierten Texte auswirken.³⁶

In Bezug auf die durch semantische Annotation generierten Taxonomien als generalisierte Modelle besteht die Möglichkeit „disagreement“ abzubilden daher nur noch innerhalb der durch die *Guidelines* und Taxonomien vorgegebenen Strukturen. Etwas, das in deren Rahmen nicht beschreibbar bzw. nicht in den Annotationskategorien vorgesehen ist, kann nicht als interpretative Perspektive modelliert und entsprechend nicht aus den generierten Datenbanken abgefragt werden. Einheitlichkeit und Eindeutigkeit in der Modellierung gehen also potentiell mit einer Konzentration auf spezifi-

sche, in den *Guidelines* erfasste Textmuster einher, auch wenn hierbei Sonderfälle stärker berücksichtigt werden können. Die Notwendigkeit zur größtmöglichen Eindeutigkeit wird aber auch dadurch befördert, dass semantische Annotationen letztendlich ebenfalls auf Möglichkeiten der Automatisierung der Annotationsprozesse zielen. Insbesondere Mehrfachkodierungen durch interpretative Diskrepanzen sowie die Komplexität der modellierten Strukturen stellen allerdings weiterhin eine Herausforderung für die Automatisierung dar (vgl. Gius und Jacke 2017: 240f.).³⁷

Angesichts dieser Auswirkungen semantischer Annotationen auf die Wahrnehmung von Textstrukturen wäre es wiederum die Aufgabe einer Semiotik als Metadisziplin die komplexen Wechselbeziehungen von modellierten Gegenständen und dem Modell als eigener Zeichenebene vergleichend zu beschreiben und dabei herauszuarbeiten, welche Strukturen jeweils durch ein Modell zum Ausdruck gebracht werden und welche blinden Flecken hierbei möglicherweise entstehen.

Dies gilt bereits für die bisher fokussierten, relativ spezifischen Annotationsmodelle: Zum einen könnte eine semiotische Analyse der Relationen von Signifikant-Signifikat-Beziehungen auf der tertiären Ebene der Annotation beleuchten. Dies wird umso notwendiger, wenn die Dynamik semiotischer Systeme auf die Ebene der Annotations-Sprache übergreift. So können etwa verschiedene Projekte an *Guidelines* zur Auszeichnung narratologischer Kategorien arbeiten. Die Formalisierungen der Konzepte und damit deren modellierende Wirkung auf die annotierten Gegenstände müssen einander aber nicht entsprechen, selbst wenn möglicherweise dieselben Termini verwendet werden.³⁸ Zum anderen ließen sich verschiedene methodische Zugriffe anhand ihres Niederschlags in den Taxonomien und *Guidelines* der Modelle erfassen und auf diese Weise eine in formalisierte Modelle überführte Methoden- bzw. Perspektivenvielfalt systematisch für potentielle Nachnutzer_innen aufbereiten.³⁹

Grundsätzlich sind die „theoretischen und technologischen Entsprechungen“ semantischer Annotation aber „in der Idee des *Semantic Web* und den *Linked Open Data* (LOD) zu sehen“, das heißt in der wissenschaftsübergreifenden Kommunikation sowie der „Erschließung und Vernetzung von Weltwissen“ (Rapp 2017: 254; vgl. auch Langmead u.a. 2016). In dieser wiederum auf Large Scale zielenden Perspektive ist eine Orientierung an gemeinsamen Kategorien unabdingbar, wie sie durch das Ontologien-Konzept des *Semantic Web* repräsentiert wird. Deshalb werden insbesondere von großen Forschungsinstitutionen Anstrengungen zur Etablierung von Ontologien unternommen. Hierbei verstärkt sich die beschriebene Tendenz zur impliziten Implementierung von Perspektiven und entsprechend die Notwendigkeit einer Modell-Kritik auf Metaebene noch, indem die kontrollierten Vokabulare und Verknüpfungsmöglichkeiten großer Ontologien wie FRBR, Dublin Core oder CIDOC CRM die Wahrnehmung von Literatur bzw. kulturwissenschaftlichen Gegenständen entscheidend prägen (vgl. Rehbein 2017: 174; Eide 2014).⁴⁰ Bedingt durch die Vorgabe möglichst breiter Einsetzbarkeit neigen jene Ontologien zudem dazu, textspezifische

Eigenschaften und Ambiguitäten zu nivellieren. Diese Perspektiven übertragen sich entsprechend auf Projekte, die Ontologien (nach)nutzen oder mit deren Hilfe modellierte Wissensbestände mit eigenen Projekten vernetzen. Dies bringt bereits für die Vernetzung von (historischen) Personen oder kulturhistorischen Fakten Herausforderungen mit sich (vgl. Shore 2014), welche in Bezug auf ambige literarische Konstruktionen wie in Romanen neben fiktionalen Charakteren auftretende historische Personen sowie metaleptisch oder autofiktional in den eigenen Roman eingeschriebene Autornamen umso problematischer erscheinen. Diese literarischen Konstrukte bewegen sich konstitutiv in einem Spannungsfeld widerstreitender Zuschreibungen, wobei eine Einordnung in übergeordnete Kategorien jeweils eine Reduktion bedeuten würde. Ähnliches gilt für abstrakte und vieldiskutierte Kategorien der Literaturwissenschaft wie „Text“ oder „Werk“. Eine vereinheitlichende Bedeutungsfestschreibung in einer Ontologie kann der Komplexität des bestehenden literaturwissenschaftlichen Diskurses nicht gerecht werden. So entspricht beispielsweise die hierarchische Modellierung von Werk, Expression, Manifestation und Exemplar in FRBR einem emphatischen Werkverständnis, bei dem das Werk sich als „abstract entity“ und „distinct intellectual or artistic creation“ (FRBR 2009: 17) definiert, welche man „am ehesten [...] vielleicht im Kopf ihres Schöpfers verorten“ (Wiesenmüller 2008: 350) kann. Die modellierte Hierarchie ist offensichtlich an der Printkultur orientiert, da sich nur hier eindeutig zwischen Expression, Manifestation und Exemplar unterscheiden lässt. Zudem impliziert sie weitere Zuschreibungen wie etwa jene, dass die Aufführung eines Werks als Expression desselben Werks zu gelten hat, die Überführung in einen Film hingegen ein eigenes Werk markiert (vgl. FRBR 2008: 17–20). Diese Vorstellungen werden mit Sicherheit nicht in allen Teilen der literaturwissenschaftlichen Forschungsgemeinschaft geteilt. In diesem Sinne erscheinen Forderungen an die digitalen Geisteswissenschaften nach der Etablierung einer „tragfähigen Ontologie“ (Loescher 2017) nicht nur unrealisierbar, sondern ein solcher Ansatz der Vereinheitlichung von Perspektiven konterkariert die Funktionsweise des geisteswissenschaftlichen und insbesondere des literaturwissenschaftlichen Forschungsdiskurses, der maßgeblich auf dem produktiven Potential widerstreitender methodischer Zugänge, Forschungsmeinungen und Interpretationen basiert.

Insbesondere im Hinblick auf die Komplexität und das diskursstrukturierende Potential dieser großen Ontologien gilt es daher von Seiten einer Semiotik als Metadisziplin die Wechselbeziehungen von modellierten Gegenständen, dem Modell als eigener Zeichenebene mit einer spezifischen Grammatik und eigenen Signifikant-Signifikat-Beziehungen sowie den das Modell beeinflussenden Theorien, Methoden und/oder abstrakten Vorannahmen systematisch darzustellen. Eine derartige Modell-Kritik in systematischer Form könnte entscheidend dazu beitragen, eine kritische Sichtung der Möglichkeiten und Restriktionen eines Modells bei den Nachnutzenden zu erleichtern und auf diese Weise Vereinheitlichungen zu verhindern, die zulasten der verhandelten Gegenstände sowie eines dynamischen

schen Forschungsdiskurses gehen. Stattdessen könnte eine semiotische Systematisierung der tertiären Ebene der Annotationen und Verknüpfungen dabei helfen, spezifizierende Brücken zwischen verschiedenen Modellierungen zu etablieren sowie das Bewusstsein für die durch das Modell vermittelte Perspektive und deren Reichweite bei den Rezipient_innen der resultierenden Visualisierungen und Datenbanken zu schärfen.

4. Abschließende Überlegungen zu einer semiotischen Rekontextualisierung computergestützter Methoden

Im Rahmen der vorangegangenen Überlegungen konnten verschiedene Verbindungslinien und Anknüpfungspunkte zwischen digitalen Methoden zur Analyse literarischer Texte und semiotischen Perspektiven aufgezeigt werden, die in deren gemeinsamer Orientierung am Zeichencharakter von Sprache und dem Ansatz der Systematisierung von Zeichenbeziehungen begründet sind. In den konkreten Vorgehensweisen wurden allerdings ebenfalls Differenzen sichtbar, die sich aus epistemischen Verschiebungen in Abhängigkeit von den Veränderungen im Rahmen der computergestützten Zugriffsweisen ergeben (vgl. Berry 2012: 4; Ramsay und Rockwell 2012). Die Rekontextualisierung literarischer Einzeltexte in einem breiteren literarischen Feld sowie die Detektion und Systematisierung übergreifender Muster, welche digitale Methoden ermöglichen, können einerseits neue oder zumindest erweiterte Perspektiven für eine semiotische Analyse eröffnen. Andererseits geht jedoch insbesondere die Dekontextualisierung der Zeichen im Rahmen der maschinellen Verarbeitung der Texte mit Deformationen der literarischen Zeichenkomplexe einher, die zunächst analysierend erfasst und in ihrer Relation zu den Funktionsweisen der Zeichenkomplexe als kontinuierliche Einheiten beschrieben werden müssen.

Die dargestellten computergestützten Methoden bedingen Reduktionen und Formalisierungen literarischer Texte auf dem Weg in ein für Computerprogramme lesbares Format, welche die Interpretation der Ergebnisse vom dabei jeweils implementierten Abstraktionsniveau abhängig machen. Die Differenzierung in quantitative und qualitative Methoden hat hierbei unterschiedliche Herausforderungen sichtbar gemacht. Gemeinsam ist den computergestützten Methoden aber die ausgeprägte Tendenz zur Modellbildung, die sich nicht zuletzt in der Visualisierung der Ergebnisse in Diagrammen, Netzen und/oder als Graphen niederschlägt.

Entsprechend dieser Heuristik ist von Seiten der Forschung der Versuch unternommen worden, die Modellierungspraxis der digitalen Geisteswissenschaften mit Bezug auf Peirces Zeichenkonzept zu beschreiben (vgl. Ciula und Eide 2017). Die vorangegangenen Ausführungen haben allerdings gezeigt, dass sich die Übersetzung zwischen Gegenstand und Modell zumindest für die Literaturwissenschaften aufgrund der Reduktion, Expansion und Deformation von Textebenen als deutlich zu komplex gestaltet, um die resultierenden Modelle in ihrem Verhältnis zu den literarischen

Gegenständen mit Peirces Vorstellung des *Icons* zu erfassen. Denn die textuellen Merkmale und Relationen, die in Visualisierungen und Datenbanken abgebildet werden, sind zu großen Teilen die Produkte komplexer Verarbeitungsverfahren im Spannungsfeld zwischen menschlichen und maschinellen Impulsen und Interventionen.⁴¹ Diese gehen mit veränderten epistemischen Techniken einher, die neben dem wissenschaftlichen Schreiben als grundlegender Diskursform auch Praktiken des Programmierens von Werkzeugen und des Aufbaus von technischen Infrastrukturen etablieren (vgl. Ramsay und Rockwell 2012). Visualisierungen als maßgebliche Formate der Ergebnisaufbereitung stellen hierbei zudem wiederum gänzlich eigene Zeichensysteme dar,⁴² die nur bedingt von den durch sie repräsentierten Gegenständen abhängig sind und als „Denkzeuge und Erkenntnis-mittel“ (Krämer 2014: 14) jeweils spezifische heuristische Funktionen erfüllen.⁴³ Denn es ist durchaus möglich, sowohl gleichförmig strukturierte Daten in völlig unterschiedlichen Visualisierungen zu repräsentieren als auch aus vollkommen verschiedenen analytischen Praktiken resultierende Daten in analoger Form visuell aufzubereiten.

Lediglich das Ergebnis mit dem Ausgangsobjekt zu vergleichen – wie dies anhand des Apfel-Beispiels von Ciula und Eide nahegelegt wird (vgl. Ciula und Eide 2017 [Abschnitt 3.4])⁴⁴ – reicht aus all diesen Gründen nicht aus, weil mindestens für den Fall literarischer Texte nicht ohne Weiteres von einer „analogy between the relations of the parts of each“ (Peirce 1932: 158 [§ 279]) auszugehen ist. Eine solche Sichtweise verkürzt die komplexen heuristischen Praktiken digitaler Analysen um jene Schritte, die erklären wie hier Wissen und Bedeutung generiert werden, auf die Frage, was in den einzelnen Modellen dargestellt ist. Diese Verkürzung wird zum einen der Vielfalt computergestützter methodischer Zugriffe nicht gerecht. Zum anderen resultieren daraus die hinsichtlich der quantitativen Verfahren beschriebenen Schwierigkeiten der Relationierung von digitalen Analyseergebnissen mit tradierten literaturwissenschaftlichen Konzepten sowie die Problematik der potentiellen Vernachlässigung der verschiedenen Möglichkeiten zur Strukturierung literarischer Texte durch die tertiäre Ebene semantischer Annotation (siehe oben, Abschnitt 3) und Vernetzung. Deshalb sollten in eine vergleichende Betrachtung die einzelnen Verfahrensschritte der Modellierung und die dabei wirksamen Übertragungs- und Rekodierungsprozesse von Zeichenkomplexen systematisch einbezogen werden, woraus eine deutlich komplexere Relation von Gegenstand und Modell folgt.⁴⁵ Diese wäre ausgehend vom Gegenstand als Kette von Übersetzungsprozessen zu beschreiben, deren einzelne Teile jeweils in ihren transformierenden Auswirkungen zu erfassen wären. Eine semiotische Strukturierung dieser Beziehungen erscheint nicht nur wünschenswert, sondern aufgrund der starken Formalisierung digitaler Methoden auch durchaus in systematischer Form möglich.

Voraussetzung für eine auf diese Weise gedachte Semiotik als Metadisziplin zur Reflexion digitaler Modellierungspraktiken wäre zum einen, dass die jeweiligen Vorgehensweisen der Modellierung durch die Akteure in aller

Detailliertheit ausgewiesen werden. Diese Forderung beschreibt allerdings ein genuines Interesse der digitalen Literaturwissenschaftler_innen selbst, denn auch sie sind für das Nachvollziehen der Analysepraktiken anderer, das eine Vergleichbarkeit mit den eigenen Ergebnissen garantiert, auf die genaue Kenntnis der Werkzeuge, Formalisierungsschritte und der den Algorithmen vorgegebenen Parameter sowie auf nachvollziehbare Visualisierungen angewiesen.⁴⁶

Zum anderen bedürfte es einer detaillierten Verständigung zwischen den Akteuren, die maßgeblich mit praktischen Modellierungen befasst sind, und jenen, die sich einer theoretischen Aufarbeitung im Sinne der bei der Modellierung hergestellten Modelle und deren Relation zu den modellierten Texten sowie zu anderen Modellen widmen. Damit die Semiotik ihrer von Aleida Assmann postulierten Rolle als Metadisziplin in Bezug auf die digitalen Literaturwissenschaften gerecht werden kann, muss sie initial an einem Dialog interessiert sein, was allerdings Kenntnisse der technologischen und mathematischen Vorgehensweisen sowie der spezifischen Terminologie der digitalen Literaturwissenschaften voraussetzt. Im Rahmen eines solchen Dialogs kann einer digitalen Semiotik Vorschub geleistet werden, die wiederum zu einer Rekontextualisierung digitaler Methoden im Feld literaturwissenschaftlicher Konzepte beiträgt.

Anmerkungen

- * Ich danke John Bateman, Maria Hinzmann sowie den Herausgebern und dem/der anonymen Reviewer_in für ihre kritische Lektüre und hilfreiche Anmerkungen zu diesem Beitrag.
- 1 Vortrag im Rahmen der Podiumsdiskussion mit Aleida Assmann und Peter Fröhlicher zum Thema *Grenzenloser Pluralismus der Methoden. Die hermeneutische Funktion der Semiotik*, 12.9.2017.
- 2 Semiotik wird in diesem Beitrag verstanden als die „Wissenschaft von den Zeichen“ bzw. Zeichensystemen und die „Theorie von den Zeichenprozessen“ (Nöth 2000: Xlf.). Es geht also nicht um die digitale Realisierung eines spezifischen semiotischen Konzepts (etwa von Peirce oder Eco), sondern vielmehr wird danach gefragt, inwieweit computergestützte Methoden grundsätzlich für zeichenbasierte Analysezugriffe auf literarische Texte fruchtbar gemacht werden (können). Dies schließt den Bezug auf verschiedene Ansätze im Rahmen der Semiotik ein.
- 3 Je nach Schwerpunktsetzung differenziert Eide (2014, Abschnitt 15) heuristisch in „modeling for understanding“ mit dem Hauptziel des „learning new things through the modeling activity“ und „modeling for production“, wo „a version of the modeled entity is the main goal of the work“. Unter letztere Kategorie fällt auch die Herstellung des Modells im Sinne einer übergeordneten Einheit bzw. Praxis zur Repräsentation von Daten selbst. Die hier diskutierten Beispiele bewegen sich im Spannungsfeld zwischen diesen Polen.
- 4 Vgl. in diesem Zusammenhang auch die von Klein und Gold beschriebene „realization that the challenges currently associated with the digital humanities involve a

- shift from congregating in the big tent to practicing DH at a field-specific level, where DH work confronts disciplinary habits of mind“ (Vorwort zu Klein und Gold 2016).
- 5 De facto werden also nicht sämtliche Signifikanten für die Modellierung verwendet. Neben der Beschränkung durch die zufällige Auswahl im Rahmen der Modellierung werden die Texte bereits im Rahmen der Präprozessierung um Flexionen sowie meist um Artikel, Konjunktionen und Hilfsverben reduziert, um verwertbare Ergebnisse zu erzeugen.
 - 6 Vgl. hierzu auch Nöth (2000: 68): „Eine Abduktion schließt von einem erklärungsbedürftigen Resultat auf eine bisher unbekannte und nur probeweise angenommene (hypothetische) Regel, um einen Fall zu erklären. [...] Die Abduktion führt von Wahrnehmungsurteilen zunächst nur vorläufig zu einer allgemeinen Regel.“ Vgl. hierzu ausführlicher Peirce (1932: 496f. [§ 776]).
 - 7 Vgl. Eder (2017) sowie die Anmerkungen in Jannidis (2014: 182f. und 188) bezüglich der Merkmalsvarianz und der Unerklärlichkeit des guten Funktionierens von Burrows' Delta zur stilometrischen Autorschaftsattribuion.
 - 8 Bubenhöfer und Scharloth (2015: 15) sehen dieses Problem allerdings nicht als ein kategorisches an: „Statt auf sog. Black-box-Data-Mining-Algorithmen wie etwa Support Vector Machines zu setzen, sollte die sozial und kulturwissenschaftlich interessierte maschinelle Textanalyse White-box-Algorithmen benutzen, um die Konstruktivität des Algorithmus transparent und nachvollziehbar zu machen“.
 - 9 „Deduction works in this ideal world – the only sphere where necessity is possible – and traces certain ideal consequences in the model so proposed. Induction takes these consequences and returns to the world of facts, seeking out further empirical data which may (or may not) corroborate these new consequences of the ideal model“ (Stjernfelt 2007: 333). „Aus dieser abduktiven Regel müssen weitere allgemeine Sätze deduziert werden, und diese müssen danach noch induktiv überprüft werden. Je mehr Beobachtungen einen induktiven Schluß auf die abduktiv postulierte Regel erlauben, desto mehr gewinnt diese an Erklärungspotential“ (Nöth 2000: 68).
 - 10 Zu den Interdependenzen von Operationalisierungsfaktoren und Visualisierung vgl. Eder (2017: 51f.).
 - 11 Im Prinzip besteht diese Herausforderung nicht nur für literarische Texte. Poetizität stellt mit Jakobson ein graduelles Phänomen dar, welches für literarische Texte konstitutiv, aber in anderen Kommunikationssituationen ebenfalls wirksam ist. Vgl. hierzu auch Hjelmslevs Modell konnotativer Semiotik (Hjelmslev 1969: 114–119).
 - 12 In diesem Sinne lässt sich auch die „grundlegende Verschiebung“ fassen, die Trilcke und Fischer (2016) in Bezug auf den Terminus „machine reading“ feststellen: „Denn was für ein Verständnis von der Praxis des Lesens wird hier aktualisiert? Haben wir es in einem solchen Fall nicht mit einem ebenso posthumanistischen wie posthumanen Begriffsverständnis zu tun? Ein Lesen frei von Bedeutung, frei von Techniken wie der Interpretation, von Techniken der Hermeneutik – frei auch von einem menschlichen Subjekt, das liest?“ Vgl. in diesem Zusammenhang bezogen auf die Kategorie „Autorstil“ auch Jannidis (2014: 191).
 - 13 Allerdings besteht auch im Rahmen bestimmter quantitativer Verfahren die Möglichkeit – etwa durch das Plotten von Ausreißern – Abweichungen von Mustern sichtbar zu machen.

- 14 Zu Peirces Differenzierung in quantitative und qualitative Induktion vgl. Stjernfelt (2007: 335–338).
- 15 Bubenhofer und Scharloth (2015: 12) sehen dieses Problem grundsätzlich für geistes und sozialwissenschaftliche Anwendungen. Stubbs zweifelt hingegen die Möglichkeiten automatisierter Induktion auch beim Bezug auf die lexikalische Ebene an, weil „even corpus-based dictionaries, which seem the ultimate example of linguistic description, are in fact prescriptive, since they do not fully take into account variation in usage“ (Stubbs 2007: 236).
- 16 Vgl. den gesamten Text auch allgemein zur Dynamik semiotischer Systeme und insbesondere zu jenem der „poetischen Sprache“ sowie zu letzterem Punkt Link und Parr (1990: 114–117).
- 17 Jockers (2015, Fußnote 1), beschreibt die aus diesem Umstand resultierenden Probleme der Sentimentanalyse, postuliert jedoch gleichermaßen deren generelles Funktionieren für die automatische Plotanalyse.
- 18 Die Begrenzung auf kulturell und zeitlich homogene Räume ist auch durch die für die Automatisierung notwendige sprachliche Übereinstimmung der Texte begründet. Des Weiteren stellt aktuell noch die beschränkte (freie) Verfügbarkeit von maschinenlesbaren Texten einen Faktor dar, der sich auf die Korpusbildung auswirkt (der aber durch die Urheberrechtsnovelle 2018 an Gewicht verlieren dürfte). Grundsätzlich wären jedoch Vergleiche von Analysen verschiedensprachlicher Korpora möglich, wenn Varianzen von Parametern und Werkzeugen als Einflussfaktoren auf die Ergebnisse auszuschließen sind.
- 19 Im Rahmen des Projekts *dlina* erfolgen allerdings aktuell Modellierungen auf der Basis von Dramenkorpora, die längere Zeiträume abdecken (vgl. Trilcke u.a. 2016).
- 20 Vgl. in diesem Zusammenhang auch die Hypothese in Jannidis u.a. (2016) bezüglich der „correlation between publication date and automatic classification quality, i.e. novels published before the period of Realism are more easily classifiable in terms of having a happy ending than realistic novels. A possible explanation is that many novels of that earlier period are more schematically structured“ (Hervorhebung J. N.).
- 21 Vgl. dazu die Beschreibung in Jannidis (2014: 182 und 189). Es wird immer wieder betont, dass stilometrische Analysen auch jenseits der Autorschaftsattribuion etwa für die Erfassung von Gattungs- oder Epochenspezifika eingesetzt werden können. Grundsätzlich ändert sich dadurch allerdings nichts an der Abhängigkeit von etablierten literaturwissenschaftlichen Kategorien und den Schwierigkeiten der Relationierung von Daten und Bedeutung.
- 22 Zur Problematik der Vermittlung zwischen literaturwissenschaftlichen Genres und stilometrisch erschließbaren linguistischen Text Types vgl. Viehhauser (2015).
- 23 Heuser und LeKhac beschreiben für ihr Projekt ausführlich das systematische Ineinandergreifen von digitaler Daten- und analoger Textanalyse: „Turning to the novels helps us interpret the data in terms meaningful to literary history, while turning to the data helps us see literary history in new ways“ (Heuser und LeKhac 2012: 31).
- 24 Vgl. hierzu auch die Analyse des Wechselverhältnisses von Denotationsebene und Konnotationsebenen in Goethes Gedicht *Geständnis* in Link und Parr (1990: 120).

- 25 Grundsätzlich problematisch erscheinen allerdings Zeichen, „die bloß Elemente einer übergeordneten Polyisotopie sind“. Deren Elemente werden von einer „übergreifenden Struktur generiert“ und entsprechend „läßt sich der literarische Diskurs als sekundäres semiotisches System keineswegs aus ‚kleinen Einheiten‘ – etwa kombinatorisch – generieren“ (Link und Parr 1990: 116f.).
- 26 Einen vielversprechenden Ansatz hierzu bietet das noch in Entwicklung befindliche Projekt *Wikidition* mit Bezug auf theoretische Konzepte von Hjelmslev, Derrida und Baßler (vgl. Wagner, Mehler und Biber 2016).
- 27 *Text Encoding Initiative*, vgl. <http://www.tei-c.org>. Auch die quantitativen Methoden basieren allerdings auf grundlegenden, häufig automatisierten Annotationen wie dem Part of Speech Tagging oder der Zuordnung von *Named Entities*. Vgl. hierzu auch das „scalar model that extends between the ideal-types of performative versus hermeneutic markup“ in Meister (2012: 115).
- 28 Rapp (2017: 257) sieht in Annotationen einen „eigenen Textsortentyp“. Diese Vorstellung rückt den Eigenwert der Annotationen als Interpretationen in den Vordergrund. Indem bei den hier betrachteten Annotationen die Textstrukturierung mittels Kategorisierung in unmittelbarer Wechselwirkung mit dem annotierten Text funktioniert, betont ein Verständnis als eigene Ebene desselben deren modellierende Wirkung. Beide Perspektiven sind aber in Bezug auf die hier diskutierten Methoden als einander ergänzende zu betrachten.
- 29 Dies bedingt *Stand-off-Markup*, welches sich im Vergleich zum *Inline-Markup* nicht deterministisch in Bezug auf Überlappungen und Mehrfachauszeichnungen verhält. Vgl. hierzu auch Meister (2012: 117–120).
- 30 Schöch (2013: 5) postuliert dies grundsätzlich für den Umgang mit „smart digital data“ im Gegensatz zu „big data“.
- 31 Zirker und Bauer (2017: 213) betonen hingegen zurecht, dass „[f]rom a post-structuralist point of view, annotation is a vain attempt at delimiting the free play of signifiers or exerting a hegemonial influence on the reader, and is therefore to be rejected“.
- 32 Im Projekt *HeureCLÉA* orientiert sich die Modellierung explizit an den Genette’schen Kategorien zur Markierung und Analyse zeitlicher Strukturen in Erzähltexten (vgl. zu dem Projekt u.a. Gius und Jacke 2015; Gius und Jacke 2017). Auch das noch in einem frühen Entwicklungsstadium befindliche Annotationsprojekt *SANTA* verfolgt die Zielsetzung der Entwicklung von *Guidelines* zur Annotation narratologischer Kategorien.
- 33 So ist beispielsweise das Ziel des Annotations-Frameworks *TEASys* insbesondere die Hierarchisierung und Kategorisierung von Erklärungen und Bedeutungszuschreibungen im Spannungsfeld zwischen Information und Interpretation (vgl. Zirker und Bauer 2017: 224f.). Ein Modell zur Beschreibung narrativer Strukturen jenseits eines konkreten Bezugs auf methodische Konzepte stellt zum Beispiel das *OntoMedia Model* dar (vgl. Jewell, Lawrence u.a. 2005).
- 34 In ähnlicher Weise führt Jannidis (2014: 190f.) folgend das Verfahren computergestützter stilometrischer Untersuchungen zur Reflexion der Unterbestimmtheit des literaturwissenschaftlichen Stilbegriffs.
- 35 Vgl. hierzu auch die von Meister (2012: 107f.) hervorgehobene Vorreiterschaft Barthes’ in Bezug auf computergestützte Analysepraktiken.

- 36 Dies stellt kein Spezifikum digitaler Methoden dar, sondern gilt beispielsweise auch für analoge Anwendungen von Genettes, maßgeblich auf der Basis eines literarischen Texts entworfener narratologischer Taxonomie.
- 37 Einen Ansatz zur Komplexitätsreduktion beschreiben Bögel u.a. (2016).
- 38 Eide (2014, Abschnitt 32f.) exemplifiziert dies anhand der semantischen Unterschiede, die sich in verschiedenen Modellen mit der Kategorie ‚Person‘ verbinden.
- 39 Vgl. hierzu auch den Vergleich verschiedener Schemata zur Modellierung von Personennetzwerken in Abhängigkeit von den jeweiligen historischen und geografischen Schwerpunkten der Projekte in Langmead u.a. (2016: 31f.) sowie deren daran anschließenden Vorschlag eines „network of networks‘ to help scholars consider not only the messy data they deal with every day, but also the messy structures they create to corral that data into pens“ (Langmead u.a. 2016: 33).
- 40 Eide (2014) diskutiert zudem, inwieweit auch der Textauszeichnungsstandard TEI bereits als Ontologie angesehen werden kann.
- 41 Bubenhofer und Scharloth (2015: 15) beschreiben die erzeugten Modelle als „Interpretationen von Interpretationen“. Vgl. außerdem Fraas und Pentzold (2015: 114) bezogen auf das Feld der Sozialwissenschaften.
- 42 Krämer (2014: 27) geht sogar davon aus, dass es sich bei den von ihr verhandelten Diagrammen nicht nur um „eine Art von Zeichensystemen [handelt], vielmehr sind sie Medien. [...] Situiert zwischen zwei Seiten stiften Medien eine Verbindung bzw. realisieren einen Austausch zwischen heterogenen Welten/Feldern/Systemen, der ohne Medien in dieser Weise nicht möglich wäre.“
- 43 Vgl. zur Differenzierung in „narrative[r] Diagrammatik“, die sich auf die „Tiefenstruktur eines Textes sowie damit verbundene kognitive Bewegungen seitens des Rezipienten“ bezieht, und „Diagramme“ als „Oberflächenphänomen[e]“ Krämer (2014: 13) mit Bezug auf Peirce.
- 44 Als Beispiel dient hier die Frucht selbst, auf die sich eine morphologische Abbildung, eine botanische Visualisierung sowie der Apfel als Zeichen für die Sünde beziehen.
- 45 Stjernfelt (2007: 59) formuliert für komplexe Beziehungen zwischen Objekt und Zeichen in ähnlicher Weise: „[in these cases; J. N.] similarity is a sign to be interpreted; the precise character of the property space in question and its topology and transformations is a question to be investigated“. Es ließe sich diskutieren, inwieweit diese Fälle – wie Stjernfelt nahelegt – noch in die Kategorie ikonischer Zeichen fallen oder ob hierfür ein anderes Konzept der Gegenstand-Modell-Relation notwendig wird.
- 46 Vgl. z.B. bezogen auf Kriterien der Korpusbildung Klimek und Müller (2015: 73).

Literatur

Sämtliche digitalen Quellen wurden zuletzt am 19.12.2017 eingesehen.

Barth, Florian und Gabriel Viehhauser (2017). Digitale Modellierung literarischen Raums. *Konferenzabstracts DHd 2017 Bern: Digitale Nachhaltigkeit*, 128–132.

- Barthes, Roland (2005). Vom Werk zum Text. In: Stephan Kammer und Roger Lüdeke (eds.). *Texte zur Theorie des Textes*. Stuttgart: Reclam, 40–47.
- Berry, David M. (2012). Introduction: Understanding the Digital Humanities. In: David M. Berry (ed.). *Understanding Digital Humanities*. Basingstoke u.a.: Palgrave Macmillan, 1–20.
- Blei, David M. (2012). Probabilistic Topic Models. *Communications of the ACM* 55, 4, 77–84, DOI: <https://doi.org/10.1145/2133806.2133826>.
- Bögel, Thomas, Evelyn Gius, Janina Jacke und Jannik Strötgen (2016). From Order to Order Switch. Mediating between Complexity and Reproducibility in the Context of Automated Literary Annotation. In: *Digital Humanities 2016: Conference Abstracts*. Kraków: Jagiellonian University & Pedagogical University, 379–82.
- Braun, Manuel (2013). Kodieren, Annotieren, Theoretisieren. Zur Wiederannäherung von Literatur und Sprachwissenschaft über Korpora. *LiLi Zeitschrift für Literatur und Linguistik* 43, 4, 83–90.
- Bubenhöfer, Noah und Joachim Scharloth (2015). Maschinelle Textanalyse im Zeichen von Big Data und Data-driven Turn – Überblick und Desiderate. *Zeitschrift für germanistische Linguistik* 43, 1, 1–26.
- Ciula, Arianna und Øyvind Eide (2017). Modelling in digital humanities: Signs in context. *Digital Scholarship in the Humanities* 32, 1, DOI: <https://doi.org/10.1093/llc/fqw045>.
- Clement, Tanya E. (2008). ‚A thing not beginning and not ending‘: using digital tools to distand-read Gertrude Stein’s *The Making of Americans*. *Literary and Linguistic Computing* 23, 3, DOI: <https://doi.org/10.1093/llc/fqn020>.
- Eder, Maciej (2017). Visualization in stylometry: Cluster analysis using networks. *Digital Scholarship in the Humanities* 32, 1, 50–64, DOI: <https://doi.org/10.1093/llc/fqv061>.
- Eide, Øyvind (2014). Ontologies, Data, and TEI. *Journal of the Text Encoding Initiative* 8, DOI: <https://doi.org/10.4000/jtei.1191>.
- Fraas, Claudia Fraas und Christian Pentzold (2015). Big Data vs. Slow Understanding? Voraussetzungen und Vorgehen computergestützter Analyse transmedialer multimodaler Diskurse. *Zeitschrift für germanistische Linguistik* 43, 1, 112–133.
- International Federation of Library Associations and Institutions (2009). Functional Requirements for Bibliographical Records (FRBR), Final Report September 1997, as amended and corrected through February 2009, URL: https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf.
- Foucault, Michel (2000). Was ist ein Autor? In: Fotis Jannidis u.a. (eds.). *Texte zur Theorie der Autorschaft*. Stuttgart: Reclam, 198–229.
- Gius, Evelyn und Janina Jacke (2015). Informatik und Hermeneutik. Zum Mehrwert interdisziplinärer Textanalyse. In: Constanze Baum und Thomas Stäcker (eds.). *Grenzen und Möglichkeiten der Digital Humanities*. (Sonderband der *Zeitschrift für digitale Geisteswissenschaften* 1). DOI: https://doi.org/10.17175/sb001_006.
- Gius, Evelyn und Janina Jacke (2017). The Hermeneutic Profit of Annotation: On Preventing and Fostering Disagreement in Literary Analysis. *International Journal of Humanities and Arts Computing* 11, 2, 233–254.
- Gold, Matthew K. (ed.) (2012). *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, URL: <http://dhdebates.gc.cuny.edu/debates/1>.

- Goldstone, Andrew und Ted Underwood (2014). The Quiet Transformations of Literary Studies: What Thirteen Thousand Scholars Could Tell Us. *New Literary History* 45, 359–384; hier zitiert nach dem Preprint, URL: <https://www.ideals.illinois.edu/bitstream/handle/2142/49323/QuietTransformations.pdf?sequence=2>.
- Heuser, Ryan und Long Le-Khac (2012). A Quantitative Literary History of 2,958 Nineteenth-Century British Novels: The Semantic Cohort Method. *Pamphlets of the Stanford Literary Lab* 4, 46, URL: <http://litlab.stanford.edu/LiteraryLabPamphlet4.pdf>.
- Hjelmslev, Louis (1969). *Prolegomena to a Theory of Language*. Madison: University of Wisconsin Press.
- Jakobson, Roman (1960). Closing Statement: Linguistics and Poetics. In: Thomas A. Sebeok (ed.). *Style in Language*. Cambridge, MA: MIT Press, 350–377.
- Jakobson, Roman (1972). Was ist Poesie? In: Wolf-Dieter Stempel (ed.). *Texte der russischen Formalisten II*. München: Fink, 392–417.
- Jannidis, Fotis (2010). Methoden der computergestützten Textanalyse. In: Ansgar Nünning und Vera Nünning (eds.). *Methoden der literatur- und kulturwissenschaftlichen Textanalyse*. Stuttgart und Weimar: Metzler, 109–132.
- Jannidis, Fotis (2014). Dem Autor ganz nah. Autorstil in Stilistik und Stilometrie. In: Matthias Schaffrick und Marcus Willand (eds.). *Theorien und Praktiken der Autorschaft*. Berlin und Boston: De Gruyter, 169–195.
- Jannidis, Fotis, Isabella Regner, Albin Zehe, Martin Becher, Lena Hettinger und Andreas Hotho (2016). *Analyzing Features for the Detection of Happy Endings in German Novels*. arXiv:1611.09028v1 [cs.IR], URL: <https://arxiv.org/abs/1611.09028>.
- Jannidis, Fotis, Hubertus Kohle und Malte Rehbein (eds.) (2017). *Digital Humanities: Eine Einführung*. Stuttgart: Metzler.
- Jewell, Michael O., Faith K. Lawrence, Mischa M. Tuffield, Adam Prugel-Bennett, David E. Millard, Mark S. Nixon, M. C. Schraefel und Nigel R. Shadbolt (2005). *OntoMedia: An Ontology for the Representation of Heterogeneous Media*. URL: <https://eprints.soton.ac.uk/id/eprint/261024>.
- Jockers, Matthew L. (2014). A Novel Method for Detecting Plot. Blogpost vom 5.6.2014, URL: <http://www.matthewjockers.net/2014/06/05/a-novel-method-for-detecting-plot/>.
- Jockers, Matthew L. (2015). The Rest of the Story. Blogpost vom 25.2.2015, URL: <http://www.matthewjockers.net/2015/02/25/the-rest-of-the-story/>.
- Klein, Lauren F. und Matthew K. Gold (eds.) (2016). *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, URL: <http://dhdebates.gc.cuny.edu/debates/2>.
- Klimek, Sonja und Ralph Müller (2015). Vergleich als Methode? Zur Empirisierung eines philologischen Verfahrens im Zeitalter der Digital Humanities. *Journal for Literary Theory* 9, 1, 53–78.
- Krämer, Sybille (2014). Zur Grammatik der Diagrammatik. Eine Annäherung an die Grundlagen des Diagrammgebrauches. *LiLi Zeitschrift für Literaturwissenschaft und Linguistik* 176, 44, 11–30.
- Langmead, Alison, Jessica M. Otis, Christopher N. Warren, Scott B. Weingart und Lisa D. Zilinski (2016). Towards Interoperable Network Ontologies for the Digital Humanities. *International Journal of Humanities and Arts Computing* 10, 1, 22–35, DOI: <https://doi.org/10.3366/ijhac.2016.0157>.

- Link, Jürgen und Rolf Parr (1990). Semiotische Diskursanalyse. In: Klaus-Michael Bogdal (ed.). *Neue Literaturtheorien vom Dekonstruktivismus bis zur historischen Diskursanalyse. Eine Einführung*. Opladen: Westdeutscher Verlag, 107–130.
- Loescher, Jens (2017). Garagenbastler der Geisteswissenschaften. *Der Tagesspiegel* 23 098, 27.4.2017, 19.
- Lotman, Juri M. (1981a). Das dynamische Modell eines semiotischen Systems. In: Juri M. Lotman. *Kunst als Sprache. Untersuchungen zum Zeichencharakter von Literatur und Kunst*. Leipzig: Reclam, 89–110.
- Lotman, Juri M. (1981b). Über das Wechselverhältnis von ‚primär‘ und ‚sekundär‘ in kommunikativen modellbildenden Systemen. In: Juri M. Lotman. *Kunst als Sprache. Untersuchungen zum Zeichencharakter von Literatur und Kunst*. Leipzig: Reclam, 111–115.
- Lotman, Juri M. (1993). *Die Struktur literarischer Texte*. 4. Auflage. München: Fink.
- Meister, Jan-Christoph (2012). Crowd Sourcing ‚True Meaning‘: A Collaborative Approach to Textual Interpretation. In: Marilyn Deegan und Willard McCarty (eds.). *Collaborative Research in the Digital Humanities: A Volume in Honour of Harold Short, on the Occasion of his 65th Birthday and his Retirement*. Farnham: Ashgate, 105–122.
- Mikolov, Tomas, Kai Chen, Greg Corrado und Jeffrey Dean (2013). *Efficient estimation of word representations in vector space*. arXiv:1301.3781v3 [cs.CL], URL: <https://arxiv.org/abs/1301.3781>.
- Moretti, Franco (2013). *Distant Reading*. London: Verso.
- Nöth, Winfried (2000). *Handbuch der Semiotik*. 2. vollständig neu bearbeitete und erweiterte Auflage. Stuttgart und Weimar: Metzler.
- Peirce, Charles S. (1932). *Collected Papers*. Hg. v. Charles Hartshorne und Paul Weiss, Bd. II: *Elements of Logic*. Cambridge: Harvard University Press.
- Peirce, Charles S. (1933). *Collected Papers*. Hg. v. Charles Hartshorne und Paul Weiss, Bd. IV: *The Simplest Mathematics*. Cambridge: Harvard University Press.
- Peirce, Charles S. (1960). *Collected Papers*. Hg. v. Charles Hartshorne und Paul Weiss, Bd. V: *Pragmatism and Pragmaticism*. Cambridge: Belknap Press of Harvard University Press.
- Piez, Wendell (2010). Towards Hermeneutic Markup: an Architectural Outline. *Digital Humanities 2010 Conference Abstracts*, King's College London, 202–205.
- Ramsay, Stephen und Geoffrey Rockwell (2012). Developing Things: Notes toward an Epistemology of Building in the Digital Humanities. In: Gold 2012, URL: <http://dhdebates.gc.cuny.edu/debates/text/11>.
- Rapp, Andrea (2017). Manuelle und automatische Annotation. In: Jannidis, Kohle und Rehbein 2017, 253–267.
- Rehbein, Malte (2017). Ontologien. In: Jannidis, Kohle und Rehbein 2017, 162–176.
- Reiter, Niels und Marcus Willand (2017). Analyse der Figurenrede (im Rahmen des Panels Aktuelle Herausforderungen der Digitalen Dramenanalyse). *Konferenzabstracts DHD 2017 Bern: Digitale Nachhaltigkeit*, 48f.
- Rhody, Lisa M. (2012). Topic Modeling and Figurative Language. *JDH: Journal of Digital Humanities* 2, 1, 19–35, URL: <http://journalofdigitalhumanities.org/2-1/topic-modeling-and-figurative-language-by-lisa-m-rhody/>.

- Schöch, Christof (2013). Big? Smart? Clean? Messy? Data in the Humanities. *JHD: Journal of Digital Humanities* 2, 3, 2–13, URL: <http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/>.
- Schöch, Christoph (2017). Quantitative Analyse. In: Jannidis, Kohle und Rehbein 2017, 279–298.
- Schöch, Christoph (2018). Quantitative Semantik. Word Embedding Models für literaturwissenschaftliche Fragestellungen. Paper-Präsentation im Rahmen des DFG-Symposiums Digitale Literaturwissenschaft, Villa Vigoni 9.–13.10.2017 (erscheint 2018).
- Shore, Dan (2014). On Categories and Relations in Networks: or, Most Abstract Blog Post Title Ever? Blogpost vom 21.8.2014, URL: <http://6dfb.tumblr.com/post/95378633456/on-categories-of-relations-in-networks-or-most>.
- Stalder, Felix (2016). *Kultur der Digitalität*. Frankfurt a.M.: Suhrkamp.
- Stjernfelt, Frederik (2007). *Diagrammatology. An Investigation on the Borderlines of Phenomenology, Ontology, and Semiotics*. Dordrecht: Springer.
- Stubbs, Michael (2007). Inferring Meaning: Text, Technology and Questions of Induction. In: Alexander Mehler und Reinhard Köhler (eds.). *Aspects of Automatic Text Analysis*. Berlin u.a.: Springer, 233–253.
- Terras, Melissa, Julianne Nyhan und Edward Vanhoutte (2013). *Defining Digital Humanities: A Reader*. Farnham: Ashgate.
- Trilcke, Peer, Frank Fischer, Mathias Göbel und Dario Kampkaspar (2016). Theatre Plays as ‚Small Worlds‘? Network Data on the History and Typology of German Drama, 1730–1930. In: *Digital Humanities 2016: Conference Abstracts*. Kraków: Jagiellonian University & Pedagogical University, 385–387.
- Trilcke, Peer und Frank Fischer (2016). Fernlesen mit Foucault? Überlegungen zur Praxis des distant reading und zur Operationalisierung von Foucaults Diskursanalyse. *Le foucauldien* 2, 1, DOI: <https://doi.org/10.16995/lefou.15>.
- Viehhauser, Gabriel (2015). Historische Stilometrie? Methodische Vorschläge für eine Annäherung textanalytischer Zugänge an die mediävistische Textualitätsdebatte. In: Constanze Baum und Thomas Stäcker (eds.). *Grenzen und Möglichkeiten der Digital Humanities*. (Sonderband der *Zeitschrift für digitale Geisteswissenschaften* 1). DOI: <https://doi.org/10.17175/sb01>.
- Wagner, Benno, Alexander Mehler und Hanno Biber (2016). Transbiblionome Daten in der Literaturwissenschaft. Texttechnologische Erschließung und digitale Visualisierung intertextueller Beziehungen digitaler Korpora. In: *Konferenzabstracts DHd 2016 Modellierung, Vernetzung, Visualisierung. Die Digital Humanities als fächerübergreifendes Forschungsparadigma*, Universität Leipzig, 87–94.
- Weel, Adriaan van der (2015). Feeding our reading machines: From the typographic page to the docuverse. In: Brent Nelson und Richard Cunningham (eds.). *Digital Studies / Le champ numérique 6: Beyond Accessibility: Textual Studies in the Twenty-First Century*, URL: https://www.digitalstudies.org/ojs/index.php/digital_studies/article/view/352/458.
- Wiesenmüller, Heidrun (2008). Zehn Jahre ‚Functional Requirements for Bibliographic Records‘: Vision, Theorie und praktische Anwendung. *Bibliothek, Forschung und Praxis* 32, 3, 348–359.

- Wilkins, Matthew (2012). Canons, Close Reading, and the Evolution of Method. In: Matthew K. Gold (ed.). *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, URL: <http://dhdebates.gc.cuny.edu/debates/text/17>.
- Zirker, Angelika und Matthias Bauer (2017). Explanatory Annotation in the Context of the Digital Humanities: Introduction. *International Journal of Humanities and Arts Computing* 11, 2, 145–152.

Erwähnte Projekte und Datenmodelle

- CIDOC CRM: CIDOC Conceptual Reference Model, entwickelt von der Documentation Standards Working Group des internationalen Museumverbands ICOM, URL: <http://www.cidoc-crm.org/>
- dlina: digitale literaturwissenschaftliche Netzwerkanalyse, URL: <https://www.uni-potsdam.de/lit-19-jhd/digitale-literaturwissenschaft/dlina.html>
- Dublin Core: entwickelt von der Dublin Core Metadata Initiative, URL: <http://dublincore.org/>
- FRBR: Functional Requirements for Bibliographical Records, entwickelt von der International Federation of Library Associations and Institutions, URL: <http://www.ifla.org/VII/s13/frbr/>
- HeureCLÉA Collaborative Literature Exploration and Annotation, URL: <http://heureclea.de>
- OntoMedia Model: Jewell, Lawrence u.a. 2005
- TEASys: Tübingen Explanatory Annotations System, URL: http://www.annotation.es.uni-tuebingen.de/?page_id=200
- TEI: Text Encoding Initiative, URL: <http://www.tei-c.org>

Dr. Julia Nantke
Bergische Universität Wuppertal
Graduiertenkolleg Dokument – Text – Edition
Gaußstraße 20
D-42119 Wuppertal
E-Mail: nantke@uni-wuppertal.de