



DIGITAL  
HUMANITIES  
COOPERATION

**Evelyn Gius**

**Computationelle Textanalysen  
als fünfdimensionales Problem:**

**Ein Modell zur Beschreibung von Komplexität**

**LitLab**

**Pamphlet #8**

**November 2019**

Herausgegeben von Thomas Weitin

# Evelyn Gius

## Computationelle Textanalysen als fünfdimensionales Problem: Ein Modell zur Beschreibung von Komplexität

### Abstract

Ansätze der computationellen Literaturwissenschaft sind inzwischen fester Teil der Digital Humanities und basieren auf einer Reihe von etablierten und neuen Verfahren, die computationelle Textanalyse mit literaturwissenschaftlichen Forschungsinteresse ermöglichen. Entsprechend besteht ein großer Bedarf an der Beschreibung und Reflexion dieser Zugänge, sowohl innerhalb der Digital Humanities als auch mit Blick auf das Verhältnis von computationeller Literaturwissenschaft zur nicht-computationellen Literaturwissenschaft. Um eine solche Reflexion zu erleichtern, wird in diesem Beitrag ein Modell vorgestellt, das die Komplexität computationeller Textanalysen in Bezug auf die betrachteten Phänomene und Texte sowie auf die Erkenntnisse der Analyse fasst. Konkret werden fünf Dimensionen vorgeschlagen, anhand derer jede computationelle Textanalyse in der Literaturwissenschaft – und darüber hinaus – beschrieben werden kann: (1) die Zusammengesetztheit der analysierten Phänomene, (2) die Kontextualisierung der Phänomene, (3) die Heterogenität der betrachteten Texte, (4) der Analysemodus und (5) der Erkenntnisbeitrag der computationellen Analyse.

Bibliografische Information der Deutschen Nationalbibliothek: Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie. Sie ist in der Zeitschriftendatenbank (ZDB) und im internationalen ISSN-Portal erfasst. Detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar. Alle Rechte, auch die des auszugsweisen Nachdrucks, der fotomechanischen Wiedergabe und der Übersetzung, vorbehalten. Dies betrifft auch die Vervielfältigung und Übertragung einzelner Textabschnitte, Zeichnungen oder Bilder durch alle Verfahren wie Speicherung und Übertragung auf Papier, Transparente, Filme, Bänder, Platten und andere Medien, soweit es nicht §§ 53 und 54 UrhG ausdrücklich gestatten.

©2019 Evelyn Gius, [gius@linglit.tu-darmstadt.de](mailto:gius@linglit.tu-darmstadt.de)

ISSN: 2628-4537  
Herausgeber: Thomas Weitin

*Evelyn Gius*

## Computationelle Textanalysen als fünfdimensionales Problem: Ein Modell zur Beschreibung von Komplexität

### 1 Einleitung

Digital Humanities-Projekte sind bestimmt von einer Reihe von Aspekten, die sie zu einem komplexen Unterfangen machen. Diese Aspekte reichen von der Aushandlung von Vorannahmen, Methoden und ihrer Passung zum Gegenstand über die Gestaltung der konkreten interdisziplinären Zusammenarbeit, die fachlich, persönlich und oft auch karrierestrategisch eine große Herausforderung für die Beteiligten sein kann, bis hin zur Darstellung von Ergebnissen für eine oder mehrere Forschungscommunities. Neben Fragen der Projektplanung und -steuerung oder wissenschaftspolitischen, strategischen und wissenschaftskommunikativen Aspekten geht es auch um Fragen, die das eigentliche Forschungsgeschehen betreffen. Dieses wird aktuell verstärkt in Bezug auf seine Relevanz und Ausrichtung diskutiert: Eine harsche Kritik von Nan Z. Da (2019a) an den Verfahren der literaturwissenschaftlich ausgerichteten Digital Humanities, den Computational Literary Studies, initiierte eine mit dem etwas überzogenen Begriff "Digital Humanities War" bezeichnete Auseinandersetzung.<sup>1</sup> Diese Debatte wird u.a. als Auseinandersetzung zwischen angeblichen Strukturalisten und Poststrukturalisten dargestellt. Zumindest attestieren manche nach eigenem Verständnis nicht strukturalistische Literaturwissenschaftler\*innen der anderen Seite Strukturalismus und diagnostizieren eine Kluft zwischen dem Strukturalismus und den eigenen Zugängen (vgl. z.B. Dobson 2019 und Bode 2020). Verschärft wird die Auseinandersetzung dadurch, dass in ihr förderpolitische Aspekte zumindest als Hintergrund eine große Rolle spielen.<sup>2</sup>

Die Debatte ist insgesamt wichtig, allerdings wird sie z.T. sehr konfrontativ geführt und in diesen Fällen wird die Chance vergeben, wirklich auf die gut begründeten Argumente der jeweiligen Gegenseite einzugehen. Unter anderem wäre es aus Sicht der Digital Humanities *und* der Literaturwissenschaft erhellend, die diskutierten Verfahren oder gar Methodenlinien detaillierter zu beschreiben und ihre Bedeutung zu reflektieren, bevor man auf jene Aspekte eingeht, die kritisiert werden (müssen). Deshalb möchte ich hier ein Modell zur Betrachtung von computationellen Textanalyseansätzen vorschlagen,

---

<sup>1</sup>Vgl. dazu den Artikel "The Digital Humanities Debacle. Computational methods repeatedly come up short" von Da (2019b) und als „Digital Humanities War“ zusammengefasste Reaktionen auf <https://www.chronicle.com/article/The-Digital-Humanities-Debacle/245986>, und das „Special Forum on Responses to Nan Z.“ Da in *Cultural Analytics* auf <https://culturalanalytics.org/2019/09/special-forum-on-responses-to-nan-z-da/>.

<sup>2</sup>Nan Z. Da (2019b) könnte diesbezüglich so zusammengefasst werden, dass sie vorschlägt, keine Fördermittel mehr in die Computational Literary Studies zu investieren bzw. damit zu verschwenden.

welches vorerst unabhängig von derartigen Grundsatzdebatten einen detaillierten Blick auf ein Vorhaben ermöglicht. Das Modell dient einer groben Einschätzung der Komplexität von Forschungsansätzen, die sich Texten mit computationellen Analysen nähern, und wurde vor dem Hintergrund der (literaturwissenschaftlichen) Analyse von literarischen Texten entwickelt. Es sollte jedoch – ggf. mit leichten Anpassungen – für Textanalysen generell geeignet sein. Im Fokus der Komplexitätsbetrachtung steht das methodische Vorgehen in einem Projekt, wobei die vorgeschlagenen Dimensionen unabhängig von der Frage der dahinterstehenden Interpretationstheorie bzw. Schule sind. Damit ist das Modell für alle Textanalyseverfahren geeignet: für jene, die in einer strukturalistischen Tradition gesehen werden können, genauso wie für solche, die eher postmoderne Zugangsweisen umsetzen – oder andere Zugänge. Das Modell bildet fünf Dimensionen ab, die für jeden computationellen Textanalyseansatz relevant sind. Diese orientieren sich an den wesentlichen Aspekten von Forschungszugängen, die computergestützte Textanalyse betreiben, also an (i) den Phänomenen, denen das Interesse gilt, (ii) den Texten, die untersucht werden, und (iii) den Erkenntnissen, die dabei gewonnen werden.

Im weiteren Verlauf soll dieses Modell nun vorgestellt werden. Auf eine Einführung des übergeordneten Aspekts folgt die Erläuterung der zugehörigen Komplexitätsdimension(en). Diese werden anhand von Beispielen aus der eigenen aktuellen Forschung sowie drei weiteren literaturwissenschaftlichen Ansätzen, die bereits in dieser Reihe als Pamphlets erschienen sind, erläutert. Abschließend wird das Modell noch einmal zusammenfassend dargestellt und seine generellen Einsatzmöglichkeiten werden umrissen.

## **2 Die Komplexität von Phänomenen**

In literaturwissenschaftlichen Textanalysen stehen normalerweise Phänomene im Sinne bestimmter Eigenschaften von Texten im Fokus. Während der Phänomenbegriff in der Literaturwissenschaft schwer zu fassen ist, und der Versuch, bestimmte Phänomene anhand von Konzepten zu fassen, Gegenstand ausgedehnter Theoriedebatten ist, bietet die im Bereich der computationellen Textanalyse nötige Operationalisierung eine geeignete Grundlage, um auch Phänomene unterschiedlicher Art in Bezug auf ihre Komplexität zu betrachten. Die Konzepte, mit denen man die erforschten Phänomene erfasst, müssen nämlich operationalisiert werden, um computationell bearbeitbar zu sein. Der Umgang und die Reflexion von Operationalisierung ist für die Entwicklung der Digital Humanities hin zu reflektiert(er)en Zugängen im letzten Jahrzehnt exemplarisch. Spätestens seit Franco Morettis Beitrag „‘Operationalizing’. Or, the Function of Measurement in Literary Theory“ vor einigen Jahren ist das Verfahren ein wichtiges Schlagwort, mit dem auch das Selbstverständnis des Felds gefasst werden kann. Denn

„operationalizing has certainly changed, and radicalized, our relationship to concepts“ (Moretti 2013, S. 119).

Die Operationalisierung der Phänomene wird im vorgeschlagenen Modell genutzt, um die zwei nachfolgend erläuterten Komplexitätsdimensionen der Zusammengesetztheit von Konzepten von Phänomenen und ihrer Kontextualisierung einzuschätzen. Damit steht wohlgerne nicht eine möglichst allgemeingültige Definition des entsprechenden Phänomens im Fokus, sondern die von den Forscher\*innen genutzte Beschreibung des Phänomens bzw. das, was man von ihrem Vorgehen über die Operationalisierung der Phänomene ablesen kann.

Die Phänomenbeschreibung bzw. Operationalisierung kann entsprechend für dasselbe Phänomen in unterschiedlichen Forschungsprojekten unterschiedlich ausfallen. Die Betrachtung der mit Phänomenen bzw. ihrer Operationalisierung zusammenhängenden Komplexität ist ein Vorgehen im Sinne von Alvarados Forderung: „As humanists, we should not accept the glib premise that the most easily operationalized ideas are the best ideas, but should instead engage in an overt and critical review of operationalization as a form of argument, even as we employ this form to test and explore a grand theory“ (Alvarado 2019).

## **2.1 Komplexitätsdimension 1: die Zusammengesetztheit von Phänomenen**

Die erste Dimension, für die man die Komplexität einer computationellen Textanalyse bestimmen kann, ist die Zusammengesetztheit des untersuchten Phänomens. Die Frage lautet: Wird das Phänomen als einfach, nicht weiter unterteilt, oder als aus mehreren Phänomenen zusammengesetzt betrachtet? In einem laufenden Forschungsprojekt zu Gender und Krankheit in literarischen Prosatexten<sup>3</sup> geht es etwa um das Phänomen Krankheit. Dieses Phänomen kann grundsätzlich unterschiedlich gefasst werden: Man könnte das Kranksein einer literarischen Figur zum Beispiel daran festmachen, ob diese ärztlich behandelt wird, man könnte aber ebenso eine Reihe von Phänomenen wie körperliche Reaktionen, Aussagen der Figur etc. nutzen, um Krankheit zu bestimmen. Ersteres wäre ein Beispiel für ein einfach gefasstes Phänomen, letzteres eines für ein zusammengesetztes Phänomen.

Betrachten wir die Zusammengesetztheit von Phänomenen bzw. ihrer Operationalisierung einmal beispielhaft in konkreten Forschungsprojekten. Die übergeordnete Fragestellung im eben genannten

---

<sup>3</sup>Vgl. dazu z.B. Gius et al. (2019), Adelmann et al. (2019) und <https://www.herma.uni-hamburg.de/subprojects.html>.

Projekt – der mögliche Zusammenhang zwischen Krankheit und Gender einer Figur – wird als zusammengesetztes Phänomen operationalisiert. Es werden nämlich Krankheitseffekte anhand von Figurencharakterisierungen betrachtet, wobei letztere anhand einer Reihe von Phänomenen bestimmt werden. In dem von Weitin & Herget (2016) beschriebenen Zugang geht es hingegen um ein einfacheres Phänomen. Es wird u.a. die Frage, ob es sogenannte „Falkentopics“ gibt, also Topics, die primär als Abbildung von Themen bzw. Zusammenfassungen des Plots spezifischer Texte gesehen werden können, anhand der Betrachtung von durch Topic Modeling ermittelten Topics umgesetzt. Das Topic Modeling-Verfahren selbst ist zwar komplex<sup>4</sup>, aber das Phänomen, das mit den so genannten Falkentopics erfasst wird, ist einfach. Es besteht aus einer Reihe von besonders häufig mehr oder weniger gemeinsam vorkommenden Wörtern.

Im Beitrag von Weitin (2018) geht es um die Frage der formalen Ausgestaltung von Novellen und ihre „durchschnittliche“ Form. Diese wird übersetzt in die Frage nach der Distinktivität einer Novelle innerhalb einer Reihe von Texten, wobei diese über Netzwerkmaße modelliert wird, die die Ergebnisse stylometrischer Analysen nutzen. Dadurch wird eine literaturhistorische Frage in vergleichsweise einfache Phänomene – die verhältnismäßige Häufigkeit von Wörtern im Kontext einer Gruppe von Texten – übersetzt.

Beim Zugang von Krautter et al. (2018) steht schließlich die Frage im Zentrum, wie eine Figurentypologisierung im Drama, insbesondere in Form der Unterscheidung zwischen Protagonisten und anderen Figuren, vorgenommen werden kann. Diese Typologisierung baut auf Aspekten wie Redeumfang, Themen und Interaktion der Figuren auf und modelliert Figurentypen entsprechend als zusammengesetztes Phänomen.

Man sieht bereits an diesen kurzen Beschreibungen, dass innerhalb eines Zugangs häufig mehrere Phänomene betrachtet werden, die zu einem übergeordneten Phänomen zusammengefasst werden. Dabei lassen sich jene Phänomene, aus denen ein komplexeres Phänomen zusammengesetzt ist, wiederum einzeln in Bezug auf ihre Komplexität betrachten. Wenn es etwa im Projekt Gender und Krankheit um die Analyse der Effekte von Krankheit auf Figuren geht, wird dort u.a. die bereits erwähnte Figurencharakterisierung als komplexes Phänomen abgebildet, das aus Figurenäußerungen und Figurenbeschreibung zusammengesetzt ist. Diese beiden Phänomene wiederum werden im Zugang als einfache Phänomene genutzt und nicht weiter differenziert nach Arten von Äußerungen oder Figurenhandlungen, Charakterzügen etc. Für die Einschätzung der Komplexität im Sinne der Zusammengesetztheit eines Phänomens

---

<sup>4</sup>Zur Komplexität von computationellen Verfahren bzw. den damit erzeugten Daten vgl. die Ausführungen im Abschnitt „Die Komplexität des Erkenntnisgewinns“ unten.

kann man deshalb neben der Zusammengesetztheit auch die Anzahl der analysierten Phänomene als Kriterium nutzen.

Auch wenn die Zusammensetzung von Phänomenen z.T. sehr abhängig vom konkreten Forschungszugang ist, gibt es doch Phänomene, die eine allgemeiner gültige bestimmte Zusammengesetztheit aufweisen. In den besprochenen Beispielen gilt das etwa für das Konzept von Topic im Topic Modeling. Auch für viele Zugänge relevante Phänomene wie Koreferenz sind meist etwas allgemeingültiger operationalisiert. Grundsätzlich gilt diese festgelegte Komplexität im Sinne der Zusammengesetztheit von Phänomenen insbesondere für Phänomene, die im Natural Language Processing (NLP) analysiert werden und für die entsprechend eine bereits computationell implementierte Operationalisierung vorliegt. Wenn Phänomene unter Rückgriff auf etablierte Konzepte in NLP-Verfahren operationalisiert werden, ist es deshalb meistens vergleichsweise einfach, die einzelnen Phänomene eines zusammengesetzten Phänomens zu identifizieren. Bei der Koreferenz sind diese Phänomene etwa Entitäten, Verweise usw., die relativ allgemein bestimmt sind.

Phänomenbeschreibungen bzw. Operationalisierungen, die nicht auf Konzepten basieren, die in bereits computationellen Verfahren genutzt werden, sind hingegen meistens schwerer zu bestimmen. Hinzu kommt, dass die Operationalisierung eines Phänomens nicht immer expliziert ist. Trotzdem lässt sie sich meist bis zu einem gewissen Punkt aus den entsprechenden Publikationen zu einem Zugang rekonstruieren.

Grundsätzlich gilt, dass die Komplexität eines Vorhabens in Bezug auf die Zusammengesetztheit von Phänomenen alleine keine Aussagekraft in Bezug auf die Relevanz oder Qualität des Ansatzes hat. Als besonders komplex zusammengesetzt operationalisierte Phänomene sind an sich nicht besser oder schlechter als einfachere.

## **2.2 Komplexitätsdimension 2: die Kontextualisierung von Phänomenen**

Neben der Bestimmung der Zusammengesetztheit geht es bei Phänomenen auch um die Frage, welches Wissen gemäß der Operationalisierung zu ihrer Bestimmung herangezogen wird. Die Kernfrage ist entsprechend: Braucht man über das Textwissen hinausgehendes weiteres Wissen, um ein Phänomen zu identifizieren? Da das Komplexitätsmodell für die computationelle Textanalyse entwickelt wurde, ist es sinnvoll anzunehmen, dass der analysierte Text selbst immer eine Rolle bei der Identifikation von Phänomenen in diesem Text spielt. Entsprechend ist die Frage, ob – über die im Text vorhandene In-

formation hinaus<sup>5</sup> – auf weiteres Wissen zurückgegriffen wird, wie etwa auf spezielles Domänenwissen, zusätzliches (innerfiktionales oder außerfiktionales) Weltwissen u.ä. Dieses zusätzliche Wissen kann in Form von Lexika, Datenbanken, aber auch weiteren Texten etc. vorliegen.

In den diskutierten Beispielen werden die analysierten Phänomene hauptsächlich im Rückgriff auf die analysierten Texte erschlossen. Darüberhinausgehendes Wissen wird in den computationellen Analysen nur insofern genutzt, als es der Evaluation der Ergebnisse der computationellen Analyse dient. In den Zugängen zu Gender und Krankheit und zu den Figurentypen wird Wissen aus der literaturwissenschaftlichen Sekundärliteratur für die Erzeugung von (Evaluations-)Daten bzw. für die Evaluation der Ergebnisse genutzt. Figurendarstellungen im Textverlauf bzw. die Klassifizierung von Figurentypen werden mit diesem Fachwissen abgeglichen.

Dass alle hier diskutierten Zugänge auf die eine oder andere Weise ihre Ergebnisse zu literaturwissenschaftlichen Erkenntnissen in Bezug setzen, ist für die Beurteilung der Komplexitätsdimension Kontextualisierung irrelevant. Insbesondere die Beiträge zu den Falkentopics und die Netzwerkanalyse von Novellen nehmen zwar eine dezidiert literaturwissenschaftliche Perspektive ein, die über den reinen Textbezug hinausgeht und literaturhistorische bzw. methodische Überlegungen anstellt. Dies geschieht aber auf einer übergeordneten Reflexionsebene jenseits der computationellen Methode. Entsprechend gibt es auch keinen zwingenden Zusammenhang zwischen der Komplexität im Sinne der Kontextualisierung von Phänomenen und der literaturwissenschaftlichen Ausprägung bzw. Anschlussfähigkeit eines Zugangs.

In jedem Fall gilt auch hier: Die Einstufung der Komplexität gilt für den betrachteten Anwendungsfall. Andere Fälle haben ggf. andere Komplexitätsgrade für dieselben Phänomene und diese können außerdem nicht direkt in eine Bewertung der Qualität des Zugangs übersetzt werden. Ungeachtet dessen haben Verfechter\*innen bestimmter literaturwissenschaftlicher Methoden hier eindeutige Präferenzen zum Rückgriff auf zusätzliches Wissen und zur Art des zusätzlichen Wissens. So lassen sich etwa Ansätze in der poststrukturalistischen Tradition von solchen in der strukturalistischen Tradition u.a. insofern unterscheiden, als erstere wesentlich mehr Kontexte in die Phänomenbetrachtung mit einbeziehen.

In Abbildung 2.1 werden die beiden Dimensionen, die die Komplexität der untersuchten Phänomene bzw. ihrer Operationalisierung bestimmen, für die diskutierten Phänomene abgebildet. Dort

---

<sup>5</sup>Die Phänomenkomplexität könnte auch noch nach dem Umfang des für die Bestimmung benötigten intratextuellen Kontexts bzw. dem Textumfang des Phänomens selbst weiter differenziert werden (vgl. dazu auch meine Ausführungen in Gius 2016:11–15). Dies wird hier, ebenso wie andere weitere Differenzierungen, nicht vorgenommen, um, das Modell vorerst möglichst schlicht zu halten. Weitere Differenzierungen sollen bei Bedarf in der Anwendung des Modells vorgenommen werden.



sieht man u.a., dass im Projekt Gender und Krankheit ein vergleichsweise simpler Krankheitsbegriff verwendet wird (Krankheit wird über Wortfelder analysiert), während die Geschlechterrollenproblematik etwas komplexer verhandelt wird (sie wird über Figurencharakterisierung im Sinne eines komplexen Phänomens betrachtet). Auch sieht man, dass Figureninteraktion als Phänomen, das Teil der Figurentypen ist, entsprechend weniger komplex als diese ist.

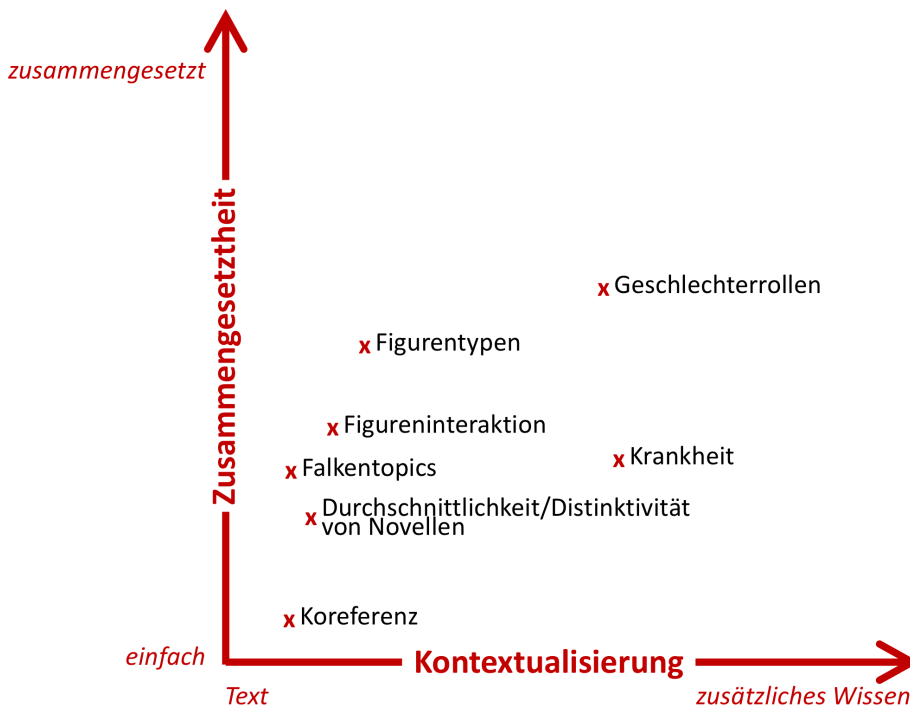


Abbildung 2.1: Komplexitätsdimensionen für Phänomene

### 3 Die Komplexität von Texten

Im Zuge von Digitalisierung und Big Data stehen Texte auf eine neue Art im Zentrum der Aufmerksamkeit. Die vermeintliche Verheißung der Digital Humanities ist, dass nun eine neue Ära angebrochen ist, in der wir massenhaft Texte analysieren können, die wir nicht gelesen haben (können). Tatsächlich scheint das Feld der computationellen Analyse literarischer Texte in Bezug auf die analysierte Textmenge aber zumindest zweigeteilt zu sein: Es gibt zwar durchaus Ansätze, die eine sehr große Anzahl von Texten analysieren. Allerdings handelt es sich bei diesen meist um eher informatikgetriebene Zugänge, in denen philologische Qualitätskriterien weder bei der Auswahl noch bei der Analyse eine wesentliche Rolle spielen. Ansätze wie der unter dem Schlagwort „Culturomics“ berühmt gewordene Zugang von Michel et al. (2011) bringen deshalb eher keine literaturwissenschaftlichen oder literaturhistorischen Erkenntnisse. Aus disziplinärer Sicht wesentlich interessanter sind tendenziell Zugänge, in denen die

Autor\*innen auf literaturwissenschaftliche Expertise sowie auf teilweise auch profunde Textkenntnis zurückgreifen und in denen entsprechend weniger umfangreiche Textbestände betrachtet werden.<sup>6</sup>

### **3.1 Komplexitätsdimension 3: Textheterogenität**

Die Frage, ob es sich um (vermeintliche) Big Data handelt oder nicht, ist deshalb aus Sicht der computationellen Textanalyse hauptsächlich insofern interessant, als damit die Frage zusammenhängt, ob man die Texte, die man analysiert, kennt oder nicht. In Bezug auf die Komplexität der genutzten Texte relevanter ist hingegen die umfassendere Frage: Wie viele (wie) verschiedene Texte werden analysiert? Die Heterogenität von Texten ist von der Anzahl der Texte selbst, aber auch von der Anzahl von verschiedenen Texteigenschaften bestimmt, die für die Fragestellung relevant sind bzw. sein könnten. Dazu gehören im Fall literarischer Texte typischerweise Eigenschaften wie Gattung, Genre, Epoche, Autorgender, Erscheinungsort etc.

In Bezug auf die Komplexität eines Vorhabens geht es damit auch um die Frage, wie viele Texte und Texteneigenschaften eine Rolle spielen, wobei die Skala von einem Text bis zu sehr vielen, sehr heterogenen Texten reicht.<sup>7</sup>

Eine vergleichsweise hohe Textheterogenität liegt im Projekt Gender und Krankheit vor. Das Korpus besteht aus über 2.000 deutschsprachigen Texten einer Gattung, aber verschiedener Genres, die von diversen Autor\*innen aus verschiedenen Epochen zwischen Naturalismus und Moderne stammen. Bei den Falkentopics und der Netzwerkanalyse werden 86 deutschsprachige Texte aus dem *Deutschen Novellenschatz* genutzt. Es handelt sich damit um ein Korpus aus einem Genre (Novellen) von insgesamt 82 Autor\*innen, das mehrere Epochen zwischen Klassik und Realismus umfasst. Dabei kann den Texten aufgrund ihrer Zusammenstellung in Sammelbänden eine gewisse Homogenität unterstellt werden. Entsprechend ist die Textbasis von mittlerer Heterogenität. Für die Figurentypologisierung wurde ein Korpus von 98 deutschsprachigen Dramen aus mehreren Epochen von der Aufklärung bis zur Moderne und von diversen Autor\*innen genutzt, wobei daraus vier Datensätze mit jeweils 39 bis 42 Dramen erstellt wurden, die überwiegend einzeln oder im Vergleich zu einem weiteren betrachtet werden. Damit ist die Textbasis ebenfalls von mittlerer Textheterogenität.

Da Textheterogenität Heterogenität in verschiedenen Hinsichten umfassen kann, kann die Komplexität dieser Dimension auf vielfältige Weise gesteigert werden. Deshalb kann die Komplexitätsdimension der

---

<sup>6</sup>Vgl. dazu auch Herget & Weitin (2016, S.4).

<sup>7</sup>Dies ist genau genommen keine einzelne Skala, weil es zwei Gegensatzpaare gibt, die abgebildet werden: Anzahl (von Texten) und Heterogenität (von Texteneigenschaften). Diese Eigenschaften werden aber zu einer Dimension zusammengefasst, da sie die Komplexität von Texten vergleichbar steigern. Vergleichbare Vereinfachungen wurden auch für die anderen Dimensionen vorgenommen.

Textheterogenität für einen Zugang am besten im Verhältnis zu anderen Zugängen bestimmt werden. So sind mittlere Korpora im Kontext von Einzeltextanalysen verhältnismäßig variant und im Kontext von Analysen, die tausende von Texten benutzen, eher weniger. Außerdem sind für die Betrachtung einzelner Textphänomene ggf. schon verhältnismäßig kleine Textmengen als sehr variant einzuschätzen, wenn sie eine große Diversität in Bezug auf das betrachtete Phänomen aufweisen.

Auch bei der Textheterogenität gibt es aber keine mit der Komplexität einhergehende Bewertung der Qualität eines Ansatzes, da alle Ausprägungen potentiell literaturwissenschaftlich interessant sind.

Abbildung 3.2 bildet die Textheterogenität der diskutierten Verfahren im Verhältnis zueinander ab.

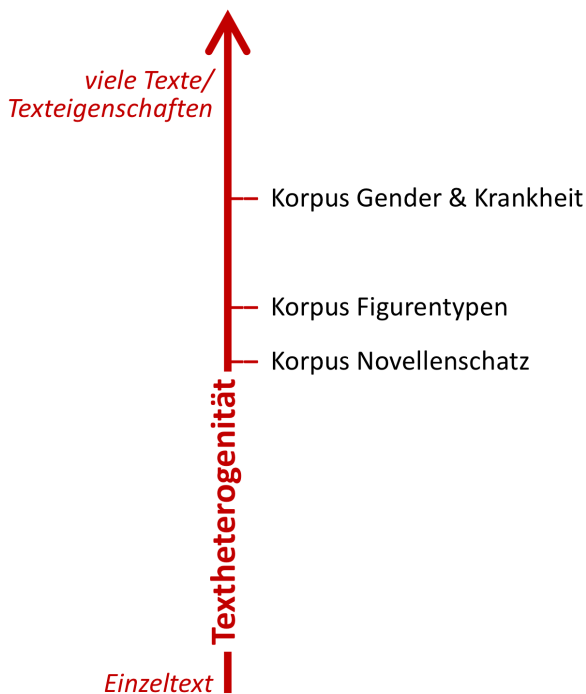


Abbildung 3.2: Komplexitätsdimension für Text(e)

## 4 Die Komplexität des Erkenntnisgewinns

Neben der Betrachtung der genutzten Phänomene und Texte geht es bei der Beurteilung der Komplexität von computationellen Textanalysen auch um die übergeordnete Frage, wie der Computer eingesetzt wird, um Erkenntnisse zu generieren. Mit „Erkenntnis“ sind hier Ergebnisse eines Textanalyseprozesses gemeint. Zu diesen gehört eine Zuweisung wenig elaborierter Phänomene an kleine Textsegmente ebenso wie Befunde über literaturhistorische Zusammenhänge.

Wenn man den Umgang mit Erkenntnissen in computationellen Textanalysen in Bezug auf ihre Komplexität betrachtet, sollte man zwei Komplexitätsdimensionen in den Blick nehmen: Zum einen geht es um die Frage des Analysezugangs – wer analysiert? – zum anderen um die Frage der Integration der Ergebnisse der computationellen Analyse in den Erkenntnisprozess – wie tragen die Ergebnisse zu Erkenntnissen bei? Da beide Dimensionen elaborierte computationelle Verfahren beinhalten können, spielt in ihnen z.T. auch künstliche Intelligenz eine Rolle. Dies bedeutet zum einen, dass die Komplexität dieser beiden Dimensionen besonders schwer zu bestimmen ist. Zum anderen handelt es sich um die Aspekte, die zudem oft in der Diskussion der Relevanz computationeller Textanalyse in den Geisteswissenschaften besonders erhitzt diskutiert werden.

Dabei wird von den Kritiker\*innen oft nicht erkannt, dass die informierte Auseinandersetzung über diese Aspekte ein zentraler Bestandteil der Geisteswissenschaften sein sollte. Man kann computationelle Textanalyse mit Luhmann als Beobachtungen zweiter Ordnung verstehen, da es sich um Beobachtungen von Beobachtungen handelt, die analog zum Lesen sind.<sup>8</sup> In Analogie zur Entwicklung des professionellen Lesens haben die Literaturwissenschaften – bzw. die Geisteswissenschaften generell – deshalb die wichtige Aufgabe, Verfahren der ‚computationellen Erkenntniserzeugung‘ zu beschreiben und im geisteswissenschaftlichen Sinne (weiter) zu entwickeln.

#### **4.1 Komplexitätsdimension 4: Analysemodus**

In der Komplexitätsdimension des Analysemodus geht es darum, wer die Erkenntnisse produziert. Wie bereits erläutert, sind Erkenntnisse alle denkbaren Analyseergebnisse. Deshalb geht es um die Frage, wer die (operationalisierten) Phänomene in der Textbasis identifiziert. Die Hauptfrage ist: Wird die Textbasis durch Menschen oder durch Computer erschlossen? Dabei wird für alle Zugänge als gegeben vorausgesetzt, dass der Computer in beiden Fällen genutzt wird. Schließlich handelt es sich um ein Modell für computationelle Zugänge. Außerdem werden im Normalfall die Ergebnisse des Analyseprozesses zumeist von Forscher\*innen, also von Menschen, weiter analysiert. Interessant ist beim Analysemodus hauptsächlich, wie die Daten für den in einem Zugang erzielten Erkenntnisgewinn erzeugt werden, ob durch menschliche Analyse oder durch den Einsatz des Computers. Lesen an sich ist natürlich ein hochkomplexer Vorgang. In einem computationellen Textanalysezugang wird er aber meist ohne weitere Reflexion als gegeben vorausgesetzt. Hingegen steigt die Komplexität mit dem verstärkten Einsatz des Computers, weil stark computerbasierte Ansätze insofern komplexer sind, als sie

---

<sup>8</sup>Diese Analogie soll wohlgerne nicht implizieren, dass der Computer liest oder gar denkt. Die computationellen Analysen, die durchgeführt werden, können aber auch ohne eine solche irreführende Anthropomorphisierung des Computers als beobachtende Verfahren gesehen werden.

schlechter für Menschen verständlich (z.B. im Fall von Deep Learning) oder sehr aufwändig in ihrer Implementierung sind, wenn sie ein Phänomen detailliert modellieren (z.B. bei regelbasierten Systemen). In computationellen Zugängen ist es meistens so, dass das menschliche Lesen und Texterfassen typischerweise in Annotationen von Textstellen oder zumindest in eine Ergänzung der Texte um Metainformationen resultiert, während beim maschinellen Erschließen hauptsächlich Textmining betrieben wird, das wiederum zusätzliche Daten erzeugt. Beide Analysemodi können weiter differenziert werden, z.B. nach der Interpretationstheorie (etwa in text-, leser- oder autororientierte Zugänge) bzw. dem angewendeten maschinellen Verfahren (etwa in regelbasierte und Lernverfahren).

In konkreten Forschungsprojekten kommen fast immer beide Modi vor. So werden im Projekt Gender und Krankheit manuelle Annotationen von Textpassagen und halb-automatische Verfahren zur Wortfeldgenerierung für die weitere Verarbeitung oder die Methodenentwicklung mit automatischen Verfahren zur Figurenerkennung, Segmentierung und Sentimentanalyse kombiniert. Da die Zwischenschritte in der Analyse zumeist manuell überprüft und teilweise ergänzt werden, handelt es sich hier um ein Verfahren zwischen Lesen und automatischem Erschließen und damit von geringerer Komplexität. Bei der Erkennung der Falkentopics wird mit MALLET ein Topic Modeling-Ansatz gewählt. Dieser wird durch die manuelle Überarbeitung der Stopwörter und durch Experimente mit der Parametrisierung zwar erheblich gestaltet, der Fokus liegt aber auf dem automatischen Erschließen, womit die Komplexität des Zugangs vergleichsweise höher ist.

Etwas anders sieht es bei den Netzwerkanalysen des *Novellenschatzes* aus. Dort werden bestehende Verfahren zur Netzwerkgenerierung genutzt, deren Auswahl und Parametrisierung literaturwissenschaftlich motiviert wird. Die Anwendung ist aber primär eine des maschinellen Erschließens.

Bei der Figurentypologisierung kommt ebenfalls eine Mischung von manuellen und automatischen Zugängen zum Einsatz. Der Fokus liegt allerdings auf dem automatischen Erschließen, da eine Klassifikation der Figurentypen als Lernverfahren implementiert wird, das die Tokenzahl von Figurenrede, sowie Daten aus dem Topic Modeling und Netzwerkmaße als Features nutzt. Damit ist der Ansatz eher komplex.

Die hier beschriebenen Aktivitäten legen nahe, dass Arbeitsschritte wie das Preprocessing, das Identifizieren von Features und das Manipulieren von Parametern Aspekte sind, die man bei der Betrachtung der Komplexitätsdimension Analysemodus als zusätzliche Kriterien einführen kann. Schließlich handelt es sich dabei um z.T. sehr aufwändige manuelle oder computationelle Datenaufbereitung, die nicht nur meist sehr zeitintensiv ist, sondern auch erhebliche Auswirkungen auf das Analyseergebnis haben kann.

## 4.2 Komplexitätsdimension 5: der Erkenntnisbeitrag computationeller Analysen

In der literaturwissenschaftlichen Praxis ist die in Bezug auf den Erkenntnisgewinn komplexeste Aufgabe der Textanalyse jene, die Textbasis insgesamt im Hinblick auf die gewählte Fragestellung zu interpretieren. Textinterpretation in diesem literaturwissenschaftlichen Sinne ist jedoch nicht im Fokus computationeller Zugänge zu literarischen Texten, die Vorstellung vom Computer als literaturwissenschaftlichem Textinterpreten erscheint bislang wenig plausibel. Trotzdem lohnt es sich, computationelle Interpretation als ein Extrem der Dimension des Erkenntnisbeitrags zu denken. In Anlehnung an die literaturwissenschaftliche Praxis kann man die Komplexitätsdimension des Erkenntnisbeitrags computationeller Analysen als von der Analyse des Textes für ein erstes Textverständnis bis hin zur Interpretation der Textbasis als Ganzes ausgedehnt sehen.<sup>9</sup> Alternativ kann auch in Anlehnung an Peirce (1935) die insbesondere in den Sozialwissenschaften verbreitete Kategorisierung von Forschungslogiken in Deduktion, Induktion und Abduktion als Skala für die Beschreibung des Erkenntnisbeitrags computationeller Analysen genutzt werden.<sup>10</sup>

Unabhängig von der Frage, welche Systematik man für die Tätigkeiten verwendet, die mit Textverstehen befasst sind, ist die zentrale Frage in der letzten Komplexitätsdimension: Wie weit geht der Erkenntnisbeitrag der computationellen Methoden? Der mögliche Erkenntnisbeitrag reicht von einfacher Textanalyse bis hin zur Interpretation bzw. von Deduktion bis hin zu Induktion oder Abduktion. Grob kann man die Komplexitätsstufen wie folgt beschreiben: Werden bereits formalisierte Analysekatégorien und -verfahren genutzt, um Texte zu analysieren, handelt es sich um eine deduktive Analyse. Dient die Analyse der Entwicklung neuer Analysekatégorien oder der Offenlegung bestimmter Zusammenhänge, so ist es ein eher induktives Verfahren. Wenn es schließlich darum geht, Hypothesen über größere, neu entdeckte Zusammenhänge in den Texten zu entwickeln, handelt es sich um Abduktion bzw. Interpretation.<sup>11</sup> Dabei betreffen die ersten beiden Aktivitäten, Deduktion und Induktion bzw. Textanalyse im engeren Sinne,<sup>12</sup> eher den Sinn nach Frege (1982), während der abduktive bzw. interpretative Modus sich mit der Fregeschen Textbedeutung beschäftigt. Im Zentrum der Frage nach dem Erkenntnisbeitrag steht in jedem Fall die Beurteilung der Neuheit der computationell erzeugten Er-

---

<sup>9</sup>Vgl. dazu Winko (2003).

<sup>10</sup>Vgl. zu den verschiedenen Forschungslogiken auch die Arbeit im Projekt hermA, die diese im Kontext von Annotationen untersucht (u.a. Gaidys et al. 2017 und [www.herma.uni-hamburg.de](http://www.herma.uni-hamburg.de)).

<sup>11</sup>Vgl. dazu auch Eco (1987): „[D]er Text ist ein Objekt, das die Interpretation im Verlauf ihrer zirkulären Anstrengungen um die eigene Schlüssigkeit bildet auf der Basis dessen, was sie als ihr Resultat erschafft. Ich schäme mich nicht, daß ich auf diese Weise den alten und immer noch gültigen hermeneutischen Zirkel definiere. Die Logik der Interpretation ist die Peircesche Logik der ‚Abduktion‘“ (S. 45).

<sup>12</sup>„Textanalyse“ ist literaturwissenschaftlich mehrdeutig, da der Begriff sowohl eine Textanalyse meint, die das Textverständnis im Fokus hat und der anschließenden Interpretation als Voraussetzung dient, als auch den Prozess der Analyse und Interpretation insgesamt, vgl. dazu Winko (2003).

gebnisse, also der innovativen Erkenntnis, die diese – vor der evtl. weiteren Analyse und Interpretation – beinhalten.

Bei der Auseinandersetzung mit der Komplexitätsdimension des Erkenntnisbeitrags ist zu beachten, dass in einer typischen literaturwissenschaftlichen Textanalyse meist alle Modi vorliegen und fließend ineinander übergehen. Interessant ist für die Komplexitätseinschätzung nun wie gesagt, welche Modi davon computationell unterstützt werden sollen. Dabei sollte nachvollziehbar sein, inwiefern der Modus literaturwissenschaftlichen Prinzipien genügt. Entsprechend gelten etwa Deep Learning-Verfahren, die durchaus Analogien zu induktiven – wenn nicht sogar abduktiven – Verfahren aufweisen, im Komplexitätsmodell nicht als abduktiv bzw. interpretativ, solange sie nicht nach literaturwissenschaftlichen Kriterien entwickelt wurden bzw. nach diesen beurteilbar sind. Grundsätzlich kann man aber davon ausgehen, dass regelbasierte Verfahren weniger komplex sind als Maschinenlernverfahren, wobei hier wiederum die überwachten Verfahren weniger komplex als die unüberwachten sind.<sup>13</sup> Eine Betrachtung der generellen Abhängigkeit zwischen computationellen Verfahren und der Komplexität ihres Erkenntnisbeitrags steht allerdings noch aus.

Aufgrund der meist gemischt vorliegenden Modi ist es in dieser Dimension besonders wichtig, bei der Betrachtung eines Zugangs entsprechend zu differenzieren. Im Fall des Projekts zu Gender und Krankheit wird etwa deduktiv die Veränderung der Figurenkonstellation anhand der Figurennennungen analysiert werden. Ein induktives Verfahren liegt vor, wenn Genderkategorien durch Clustering von Figurenrede herausgearbeitet werden (die dann wieder als neues Verfahren deduktiv in der Analyse genutzt werden). Und schließlich liegt ein abduktiver Zugang vor, wenn durch eine Gesamtbetrachtung ein neues Element entdeckt wird, das Figurenkrankheit beeinflusst.

Insgesamt wird, im genannten Anwendungsbeispiel eher deduktiv vorgegangen. Es werden hauptsächlich einfache Extraktionsverfahren genutzt, die allerdings z.T. auf die Entwicklung eines computationellen Verfahrens abzielen, das Figurenanalyse automatisiert und entsprechend induktiver bzw. textanalytisch anspruchsvoller arbeitet.

Im Beitrag von Weitin & Herget (2016) geht es darum, Topic Modeling auf seine Eignung für die literaturwissenschaftliche Textbeschreibung hin zu prüfen. Die dahinterstehende Idee, dass Topics ggf. als Beschreibung einzelner Texte geeignet sind, ist eine literaturwissenschaftliche Sicht auf die Anwendbarkeit des Verfahrens. Entsprechend ist der rein computationell mögliche Erkenntnisgewinn hier höher

---

<sup>13</sup>Für eine Beschreibung der Nutzung computationeller Verfahren in der computationellen Textanalyse vgl. Underwood (2015), der sieben Typen vorschlägt.

einzuschätzen. Im Kontext der Falkentopics kann Topic Modeling mindestens als induktives Verfahren betrachtet werden, das zwischen Textanalyse und Textinterpretation angesiedelt ist.

Ähnlich verhält es sich bei den Netzwerkanalysen, in denen ebenso die literaturwissenschaftliche Eignung der computationallyen Verfahren geprüft wird. Da die Verfahren als zur Analyse der Novellen geeignet herausgestellt werden, liegt hier ebenso ein automatisches Verfahren vor, dessen Erkenntnisgewinn vergleichsweise hoch ist.

Die Figurentypologien nutzen wiederum eine Reihe weniger komplexer (deduktiver) Verfahren zur Extraktion von Features, die sie dann in eine Klassifikation überführen, was ein stärker erkenntnisgenerierender Vorgang ist, der entsprechend als induktiv bezeichnet werden kann. Abbildung 4.3 bildet die diskutierte Erkenntnisgenerierung der Projekte beispielhaft auf die entsprechende Komplexitätsdimension ab.

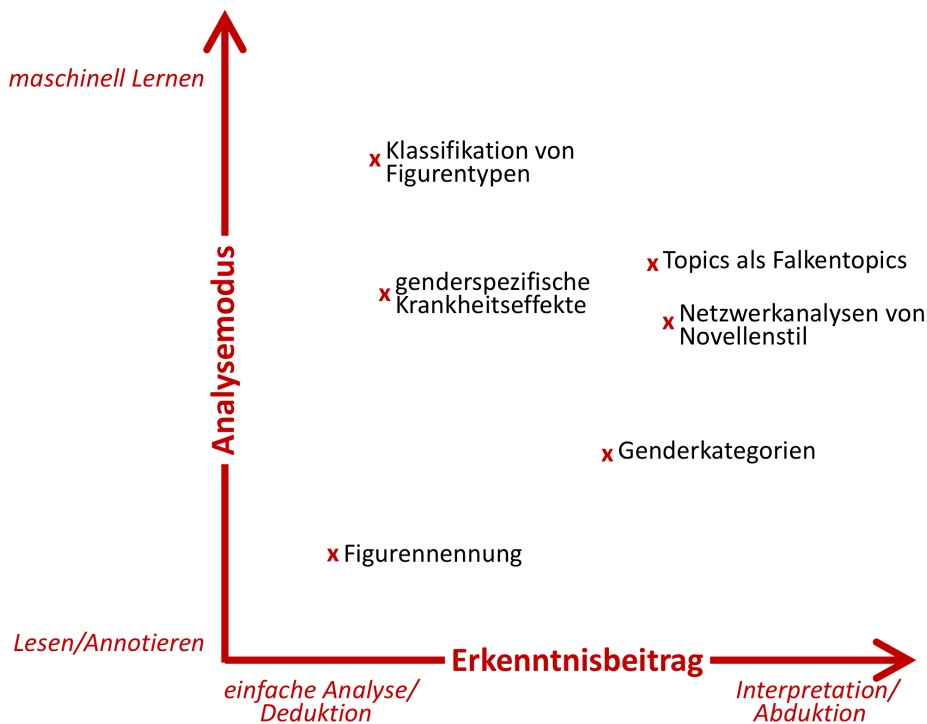


Abbildung 4.3: Komplexitätsdimension für Erkenntnis

## 5 Das Modell zur Komplexität computationeller Textanalyse

Das vorgestellte Modell ist als allgemeines Instrumentarium angelegt, das zur Betrachtung computationeller Textanalysen genutzt werden kann. Eine Spezifizierung oder auch Anpassung der Komplexitätsaspekte soll je nach Anwendungssituation erfolgen. Die Möglichkeiten der Anwendung wurden



bereits beispielhaft skizziert. Abbildung 5.4 zeigt die Komplexität der diskutierten Projekte in allen fünf Dimensionen, wobei die diskutierten Aspekte zu einer Gesamtbeurteilung der einzelnen Dimension für jedes Projekt zusammengefasst wurden.

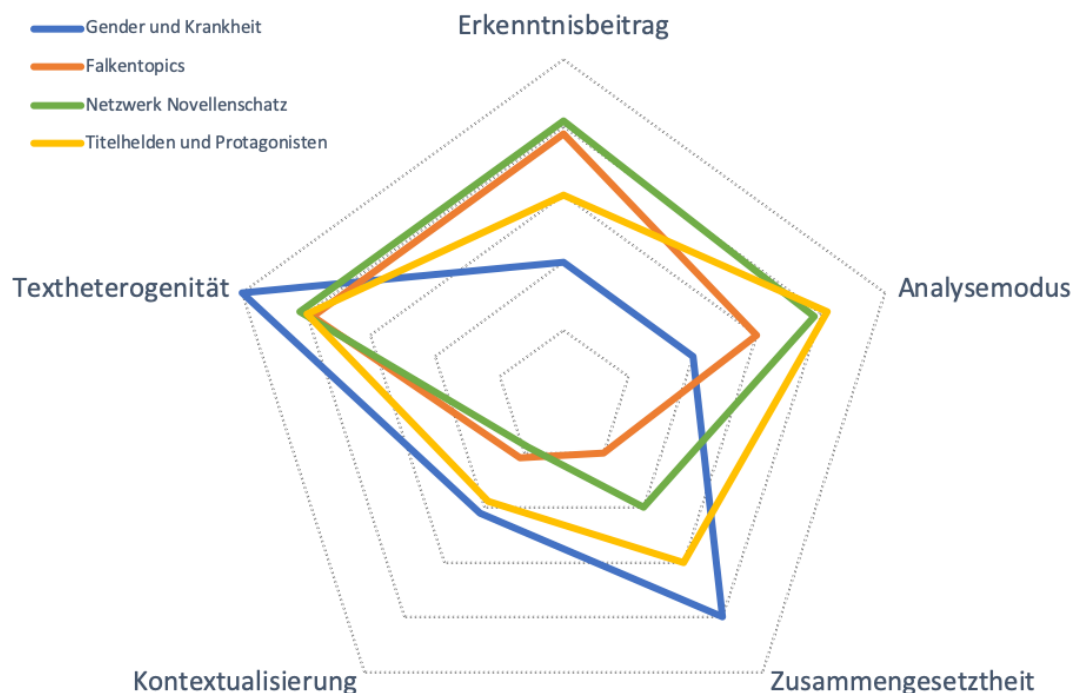


Abbildung 5.4: Komplexität der diskutierten Zugänge in allen fünf Dimensionen.

An dieser Stelle sei noch einmal auf die Vereinfachungen verwiesen, die bei der Konzipierung und beispielhaften Anwendung des Modells vorgenommen wurden. So wurde Kontextwissen nicht weiter differenziert, die Anzahl von Texten und Texteigenschaften unter Heterogenität subsumiert, keine Abgrenzung von literaturwissenschaftlicher Analyse und Interpretation vorgenommen und deren Zusammenhang mit der Forschungslogik von Peirce und maschinellen Verfahren nicht detailliert betrachtet. An einigen Stellen der Modellbeschreibung bzw. -anwendung wurden bereits mögliche weitere Kriterien herausgestellt, die in das Modell einfließen können, wie etwa Parametermanipulation und Preprocessing in der vierten Komplexitätsdimension, dem Analysemodus.

Grundsätzlich soll das Modell als ein erster Entwurf verstanden werden, der sich nun in seiner Anwendung bewähren muss, wobei sich dabei auch mögliche Erweiterungen zeigen sollen.

Das vorgeschlagene Modell wird nachfolgend noch einmal komprimiert dargestellt. Wichtig ist, dass es sowohl für die Darstellung von Teilaktivitäten als auch für die Darstellung eines gesamten Forschungsprojektes genutzt werden kann (vgl. dazu Abbildungen 2.1, 3.2, 4.3 bzw. Abbildung 5.4). Außerdem ist es möglich, nur ausgewählte der fünf Dimensionen zu betrachten.

- **Komplexitätsdimension 1: die Zusammengesetztheit von Phänomenen**

Frage: *Wird das Phänomen als einfach, nicht weiter unterteilt, oder als aus mehreren Phänomenen zusammengesetzt betrachtet?*

Komplexität: von einfach bis hin zu vielfach zusammengesetzten Phänomenen

- **Komplexitätsdimension 2: die Kontextualisierung von Phänomenen**

Frage: *Braucht man über das Textwissen hinausgehendes weiteres Wissen, um ein Phänomen zu identifizieren?*

Komplexität: von Textwissen bis hin zu verschiedenen Arten von umfangreichem weiterem Wissen

- **Komplexitätsdimension 3: Textheterogenität**

Frage: *Wie viele (wie) verschiedene Texte werden analysiert?*

Komplexität: von einem Text mit homogenen Eigenschaften bis hin zu vielen, in sich und zueinander heterogenen Texten

- **Komplexitätsdimension 4: Analysemodus**

Frage: *Wird die Textbasis durch Menschen oder durch Computer erschlossen?*

Komplexität: von von Menschen annotiert bis hin zu von Maschinen durch Lernen analysiert

- **Komplexitätsdimension 5: der Erkenntnisbeitrag computationeller Analysen**

Frage: *Wie weit geht der Erkenntnisbeitrag der computationellen Methode?*

Komplexität: von der Anwendung simpler Regeln auf einzelne Textelemente bis zur Interpretation der gesamten Textbasis

## 6 Zur Nutzung des Modells

Wie bereits dargelegt, betrifft die Bestimmung der Komplexität eines computationellen Textanalysezugangs primär die normativen Setzungen durch die Forscher\*innen in den fünf Dimensionen. Ausschlaggebend ist weniger, wie Phänomene, Texte und Erkenntnis an sich modelliert werden sollten, sondern vielmehr, wie sie tatsächlich umgesetzt werden. Daran sollte sich auch eine Kritik eines Zugangs orientieren, die produktiv ist. Wenn man nämlich z.B. einen Zugang, der ein Phänomen stark vereinfacht modelliert, in allen fünf Dimensionen anschaut, ist das erkenntnisproduktiver, als wenn

man sich auf die Kritik der unterkomplexen Phänomendarstellung beschränkt und dadurch die weiteren Aspekte ignoriert. Insbesondere ist es wenig erhellend, wenn man ein Projekt, das aus bestimmten Gründen eine sehr simple Operationalisierung der Phänomene nutzt, hauptsächlich in Bezug auf diese Operationalisierung diskutiert. Oder wenn fehlende literaturhistorische Erkenntnis angemahnt wird, wenn auf Basis nur teilweise literaturwissenschaftlich motiviert ausgewählter Texte an der Erkennung bestimmter Phänomene gearbeitet wird. Denn, auch wenn in manchen Fällen eine wenig komplexe Umsetzung in einer Dimension unterkomplex im Sinne von nicht gegenstandsadäquat ist, so ist es doch häufig so, dass Schwierigkeiten von Ansätzen nicht hauptsächlich darin bestehen.

Wenn man die Zusammenhänge zwischen den einzelnen Dimensionen im Komplexitätsmodell berücksichtigt, sieht man, dass der Vorwurf von Unterkomplexität in vielen Fällen nur bedingt gültig ist. Denn die Steigerung der Komplexität in einer Dimension führt praktisch immer zu erhöhter Gesamtkomplexität. Wenn ich etwa als überzeugte Poststrukturalistin für die Phänomenerkennung nicht nur auf die untersuchten Texte, sondern auch auf weiteres Kontextwissen zurückgreife und mein Korpus außerdem maschinell erschließen will, wird letzteres entsprechend anspruchsvoller. Dasselbe gilt, wenn ich zahlreiche, sehr heterogene Texte in Bezug auf ein Phänomen untersuchen will und dieses entsprechend mit Blick auf ihre Heterogenität operationalisieren muss. Oder wenn ich Interpretation oder zumindest abduktive Vorgänge computationell umsetzen will, dann wird auch die einfachste Operationalisierung in wenigen, gleichförmigen Texten zur großen Herausforderung.

Wichtig ist: Wenn man fordert, dass bestimmte Aspekte an einem Ansatz in einem positiven Sinne komplexer umgesetzt werden, sollte man diese Dynamik mitbedenken. Außerdem sollte man bedenken: Je komplexer ein Zugang im Sinne des vorgestellten Modells, desto zahlreicher die möglichen Fehler – und desto anspruchsvoller sowohl die Kritik als auch die Umsetzung des Zugangs.

Insbesondere dieser pragmatische Aspekt von Komplexität macht das Modell über seine Funktion als Heuristik für die strukturierte Kritik von einem Zugang oder mehreren Zugängen hinaus zu einem Planungstool. Es bietet sich nämlich ebenso als Instrument für den Entwurf eines Zugangs an und sollte in allen Phasen computationeller Textanalyse genutzt werden – vom Design des Forschungszugangs zu Beginn der Forschungsarbeit über die wiederholte Bestandsaufnahme oder Nachjustierung im Projektverlauf bis hin zur Einordnung der erzielten Ergebnisse am Ende und der Reflektion des gesamten Prozesses.

Neben jeweils projektspezifischen Fragen, die man damit beleuchten kann, können die Ausprägung und Kombination der fünf Dimensionen für folgende Punkte genutzt werden:

- Abschätzung von Innovation und Risiko: Was besonders ausgeprägte Komplexität aufweist, bedarf besonders gründlicher Betrachtung. Darin steckt das größte Innovationspotential – und das größte Risiko. Idealerweise wird die Abschätzung sowohl für einzelne Aufgaben als auch für das Gesamtvorhaben vorgenommen, um zusätzlich eine Einschätzung der Bandbreite der Komplexitäten zu erhalten.
- Ressourcenplanung: Je weniger ausgeprägt die einzelnen Dimensionen sind, umso schneller kann das Projekt bearbeitet werden. Umgekehrt bedeutet eine tendenziell hohe Komplexität einen größeren Bearbeitungsaufwand. Kleinere Projekte (für kleinere Beiträge, Projektarbeiten etc.) sollten im Komplexitätsmodell tendenziell innen angesiedelt sein bzw. sich auf innen angesiedelte Aspekte fokussieren. Projekte mit zeitlich und personell umfangreichen Rahmen sollten hingegen außen angesiedelt sein.
- Nachjustierung: Wenn sich Projekte im Verlauf als komplexer herausstellen, kann man durch Vereinfachung einer Dimension die Gesamtkomplexität und damit die benötigten Ressourcen erheblich reduzieren. Umgekehrt könnte diese bei freien Kapazitäten auch gesteigert werden. Beides kann gezielt dort eingesetzt werden, wo die Ressourcen fehlen bzw. frei sind. Dies gilt in Projekten von Einzelforscher\*innen genauso wie in Teamprojekten oder Verbundforschungsvorhaben.

## 7 Literaturverzeichnis

- Adelmann, Benedikt, Melanie Andresen, Anke Begerow, Lina Franken, Evelyn Gius, und Michael Vauth. 2019. „Evaluation of a Semantic Field-Based Approach to Identifying Text Sections about Specific Topics“. *DH2019 Book of Abstracts*. Utrecht. <https://dev.clariah.nl/files/dh2019/boa/0895.html>.
- Alvarado, Rafael C. 2019. „Digital Humanities and the Great Project: Why We Should Operationalize Everything—and Study Those Who Are Doing So Now“. In *Debates in the Digital Humanities 2019*, herausgegeben von Matthew K. Gold und Lauren F. Klein. University of Minnesota Press. <https://doi.org/10.5749/j.ctvg251hk>.
- Bode, Katherine. 2020. „Why you can’t model away bias“. Preprint: *Modern Language Quarterly* 80 (3). [https://katherinebode.files.wordpress.com/2019/08/mlq2019\\_preprintbode\\_why.pdf](https://katherinebode.files.wordpress.com/2019/08/mlq2019_preprintbode_why.pdf).
- Da, Nan Z. 2019a. „The Computational Case against Computational Literary Studies“. *Critical Inquiry* 45 (3): 601–39. <https://doi.org/10.1086/702594>.
- Da, Nan Z. 2019b. „The Digital Humanities Debacle“. *The Chronicle of Higher Education*, 27. März 2019. <https://www.chronicle.com/article/The-Digital-Humanities-Debacle/245986>.
- Eco, Umberto. 1987. *Streit der Interpretationen*. Herausgegeben von Rolf Eichler. Konstanzer Bibliothek 8. Konstanz: Universitäts-Verlag.
- Frege, Gottlob. 1892. „Über Sinn und Bedeutung“. *Zeitschrift für Philosophie und philosophische Kritik*, 100 (1), 25–50..
- Gaidys, Uta, Evelyn Gius, Margarete Jarchow, Gertraud Koch, Wolfgang Menzel, Dominik Orth, und Heike Zinsmeister. 2017. „Project description – hermA: Automated modelling of hermeneutic processes“. *Hamburger Journal für Kulturanthropologie*. <https://journals.sub.uni-hamburg.de/hjk/article/view/1213>.
- Gius, Evelyn. 2016. „Narration and Escalation. An Empirical Study of Conflict Narratives“. *Diegesis* 5 (1): 4–25.
- Gius, Evelyn, Katharina Krüger, und Carla Sökefeld. 2019. „Korpuserstellung als literaturwissenschaftliche Aufgabe“. *DHd 2019 Digital Humanities: multimedial & multimodal Konferenzabstracts*, Frankfurt & Mainz: 164–166.
- Krautter, Benjamin, Janis Pagel, Nils Reiter, und Marcus Willand. 2018. „Titelhelden und Protagonisten - Interpretierbare Figurenklassifikation in deutschsprachigen Dramen“. Herausgegeben von Thomas Weitin. *LitLab Pamphlets*, Nr. 7. [https://www.digitalhumanitiescooperation.de/wp-content/uploads/2018/12/p07\\_krautter\\_et\\_al.pdf](https://www.digitalhumanitiescooperation.de/wp-content/uploads/2018/12/p07_krautter_et_al.pdf).

- Michel, Jean-Baptiste, Yuan K. Shen, Aviva P. Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, J. P. Pickett, u. a. 2011. „Quantitative Analysis of Culture Using Millions of Digitized Books“. *Science* 331 (6014): 176–82. <https://doi.org/10.1126/science.1199644>.
- Moretti, Franco. 2013. „‘Operationalizing’. Or, the Function of Measurement in Literary Theory“. *New Left Review*, Nr. 84 (Dezember): 103–19.
- Peirce, Charles S. 1935. *Collected Papers of Charles Sanders Peirce, Volumes V and VI: Pragmatism and Pragmaticism and Scientific Metaphysics*. Herausgegeben von Charles Hartshorne und Paul Weiss. Cambridge, Mass: Belknap Press of Harvard Univ. Press.
- verschiedene Autor\*innen. „Special Forum on Responses to Nan Z. Da“. 2019. *Journal of Cultural Analytics*. 17. September 2019. <https://culturalanalytics.org/2019/09/special-forum-on-responses-to-nan-z-da/>.
- Weitin, Thomas. 2018. „Average and Distinction. The Deutsche Novellenschatz Between Literary History and Corpus Analysis“. Herausgegeben von Thomas Weitin. *LitLab Pamphlets*, Nr. 6. <http://bit.ly/2CBXNo1>.
- Weitin, Thomas, und Katharina Herget. 2016. „Falkentopics“. Herausgegeben von Thomas Weitin. *LitLab Pamphlets*, Nr. 4. <https://www.digitalhumanitiescooperation.de/pamphlete/pamphlet-4-falkentopics/>.
- Winko, Simone. 2003. „Textanalyse“. In *Reallexikon der deutschen Literaturwissenschaft: Neubearbeitung des Reallexikons der deutschen Literaturgeschichte*, herausgegeben von Harald Fricke, Klaus Grubmüller, Jan-Dirk Müller, und Klaus Weimar, 3., neubearb. Aufl. Berlin: De Gruyter: 597–601.