

Artikel aus:
Zeitschrift für digitale Geisteswissenschaften

Titel:
Die Vernetzung des Textes: Im Möglichkeitsraum digitaler Literaturanalyse

Autor/in:
Christine Ivanovic

Kontakt:
christine.ivanovic@univie.ac.at

Institution:
Abteilung für Vergleichende Literaturwissenschaft an der Universität Wien, Institut für Corpuslinguistik und Texttechnologien (ICLTT) der Österreichischen Akademie der Wissenschaften (ÖAW)


GND:
[115731776](#)

ORCID:

DOI des Artikels:
[10.17175/2016_010](https://doi.org/10.17175/2016_010)

Nachweis im OPAC der Herzog August Bibliothek:
[868714151](#)

Erstveröffentlichung:
10.02.2017

Lizenz:
Sofern nicht anders angegeben 

Medienlizenzen:
Medienrechte liegen bei den Autoren

Letzte Überprüfung aller Verweise:
10.02.2017

GND-Verschlagwortung:
[Literaturwissenschaft](#) | [Digital Humanities](#) | [Semantic Web](#) |

Zitierweise:
Christine Ivanovic: Die Vernetzung des Textes: Im Möglichkeitsraum digitaler Literaturanalyse. In: Zeitschrift für digitale Geisteswissenschaften. 2017. PDF Format ohne Paginierung. Als text/html abrufbar unter DOI: [10.17175/2016_010](https://doi.org/10.17175/2016_010).

Christine Ivanovic

Die Vernetzung des Textes: Im Möglichkeitsraum digitaler Literaturanalyse

Abstracts

Die Digitalisierung von Texten, die Erarbeitung komplexer digitaler Editionen und die Herstellung maschinell analysierbarer Korpora sind Hauptaufgaben gegenwärtiger digitaler Philologie. Diese baut vielfach auf Methoden und Erkenntnissen auf, wie sie in der Computerlinguistik seit Jahrzehnten erprobt werden. Sie müssen jedoch für die Anforderungen der literaturwissenschaftlichen Analyse beispielsweise in ihrer Granularität angepasst werden. Der Beitrag skizziert mögliche Vorgehensweisen bei der digitalen Verarbeitung großer Textmengen respektive der Entwicklung komplexer Annotationsverfahren. Sie verändern die Ausgangssituation philologischer Forschung noch weiter, wenn die Texte im Internet verfügbar sind. Daher werden in einem zweiten Schritt die daraus sich ergebenden neuen Möglichkeiten der Analyse literarischer Texte diskutiert. Neben der Entwicklung von Verfahren zur systematischen und nachprüfbareren Literaturanalyse großer Textkorpora geht es dabei um die Darstellung der Mehrdimensionalität literarischer Texte durch Datenverknüpfung im Semantic Web. Literarische Texte können dadurch auch anderen Disziplinen zur Evaluierung zugänglich gemacht werden.

The main tasks of contemporary digital philology are the digitization of texts, the development of complex digital editions, and the creation of text corpora that can be analyzed by machines. These activities build on methods and findings that have been explored in computer linguistics for decades. There is, however, a need to adapt those procedures to the requirements of literary analysis, such as in their granularity. This paper outlines possible approaches for the digital processing of large volumes of text and for the development of complex annotation processes. These methods change the very basis of literary analysis, especially if the texts are available on the internet. The next step, therefore, is to discuss the new possibilities for the analysis of literary text that arise from this situation. These possibilities go beyond the mere development of procedures for systematic and verifiable literary analysis of large text corpora. One new step, for example, would be the representation of multi-dimensional qualities of literary texts by data linkage in the Semantic Web, which could make literary texts available for other disciplines as well.

»Il faut confronter les idées vagues avec des images claires« – Jean-Luc Godard, 1967 («La Chinoise«)

»Alles, was Gegenstand wissenschaftlichen Denkens überhaupt sein kann, verfällt, sobald es zur Bildung einer Theorie reif ist, der axiomatischen Methode und damit mittelbar der Mathematik.« – David Hilbert, 1918

1. Vom Wörterzählen zur Datenvernetzung

Zwei Wege führen an den Punkt, wo sich die Computerphilologie heute befindet: Der eine geht auf eine Initiative des Jesuitenpaters Roberto Busa zurück, der zwischen 1940 und 1947 gemeinsam mit IBM Gründer Thomas J. Watson eine computergestützte Lemmatisierung der

Werke von Thomas von Aquin vornahm und mit dem *Index Thomisticus* das früheste Beispiel digitaler Textanalyse vorlegte (seit 2005 online verfügbar¹). Bis zu seinem Tod 2011 verfolgte Buso den Ansatz computergestützter Hermeneutik und Lexikographie. In Deutschland bemühte sich der Philosoph Max Bense um die Einbeziehung von Informationstheorie und Computertechnik in seine Analysen der Strukturen von Sprache.² In England versuchte der aus Südafrika eingewanderte Mathematiker und spätere Literaturnobelpreisträger J.M. Coetzee an die stilometrischen Analysen des deutschen Physikers Wilhelm Fucks anzuknüpfen, der in den fünfziger Jahren quantitative Methoden in Literaturwissenschaft und Linguistik eingeführt hatte.³ In seiner Dissertation von 1969 legte Coetzee erstmals stilometrische Analysen der englischen Texte von Samuel Beckett mit mathematischen Methoden vor.⁴

Der andere Weg kommt aus der langen Geschichte der Editionsphilologie. Die ersten digitalen Editionen entstanden Anfang der neunziger Jahre mit der kommerziellen Verbreitung der CD-ROM; erstmals war ein mobiler Datenträger gefunden worden, der genug Speichervolumen für ein ganzes Buch mit Anmerkungen (oder für 74 Minuten Musik) aufwies. Eines der frühesten Projekte war die Digitalisierung der Buchausgabe der Colli/Montinari-Edition der Werke Nietzsches, die der Verlag de Gruyter 1994–1997 auf CD-ROM herausbrachte und die inzwischen online zugänglich ist.⁵ Eines der jüngsten ist die ebenfalls online einsehbare 3D-Ausgabe der Ogham Stones durch das Dublin Institute of Advanced Studies; es handelt sich um ein irisches Nationaldenkmal und einmaliges Zeugnis einer frühen Stufe der irischen Sprache und ihres Alphabets, das als Inschrift auf Steinen überliefert ist.⁶

Digitale literaturwissenschaftliche Projekte sind mehrheitlich Editionen, die sich nun in den, wie oft betont wird, »schiefer unbegrenzten« Speicherraum des World Wide Web ausdehnen. Neben exklusiven, aufwendig erstellten kritischen digitalen Ausgaben repräsentativer Werke⁷ werden heute in großem Umfang Digitalisierungen ganzer Bibliotheken durchgeführt, um wertvolle Altbestände zu erhalten oder um große Korpora für die Öffentlichkeit zugänglich zu machen.⁸ Motivation, Organisation, technische Durchführung und Nutzungskonzepte können dabei stark variieren.⁹ Wichtige Korpora wurden und werden von nationalen Institutionen in Auftrag gegeben und verwaltet,¹⁰ daneben gibt es die digitale Texterfassung durch

¹ <http://www.corpusthomicum.org/it/index.age>.

² Bense 1962; Bense 1969.

³ Fucks 1955.

⁴ Coetzee 1969.

⁵ <http://www.nietzschesource.org/documentation/de/eKGWB.html>.

⁶ <https://ogham.celt.dias.ie/menu.php?lang=en>. Zur Entwicklungsgeschichte digitaler Editionen vgl. auch <http://www.digitale-edition.de>.

⁷ Als ein Beispiel sei hier die Klagenfurter Ausgabe der Werke Robert Musils genannt (Musil 2009). Vgl. dazu Salgado 2014.

⁸ Lauer 2013; Lauer 2014.

⁹ Das Spektrum reicht von der mit Deutschen Digitalen Bibliothek 18.098.542 Objekten (davon 6.129.980 mit Digitalisat) über das [Deutsche Textarchiv](#) mit 474.326 digitalisierten Seiten bis zum [zentralen Verzeichnis digitalisierter Drucke](#), das derzeit 1.339.130 Titel registriert, um nur einige Beispiele aus dem deutschsprachigen Raum zu nennen. Eine repräsentative Auswahl verzeichnet die [German Literature Collection](#), über die einschlägige Werkausgaben von Goethe, Schiller, Kafka und Brecht, sowie »Die deutsche Lyrik« (basierend auf Ausgaben aus Reclams Universalbibliothek) und »Deutsche Klassiker« (basierend auf der Bibliothek Deutscher Klassiker) abgerufen werden können. Für die wissenschaftliche Analyse bereits mit XML/TEI erschlossene Texte findet man als Digitale Bibliothek im [TextGrid Repository](#).

¹⁰ In Österreich wäre beispielweise das AAC zu nennen, das vom ICLTT an der ÖAW erstellt wurde <http://corpus1.aac.ac.at/fackel/>.

Freiwillige¹¹ oder durch Firmen. So hat Google einen immensen und teilweise frei zugänglichen Datenbestand mit Unterstützung zahlreicher Universitätsbibliotheken aufgebaut.¹² Datenerhalt und Archivierung großer Textmengen sowie leichtere Zugänglichkeit, Durchsuchbarkeit und bessere Auffindbarkeit von Textstellen sind die Vorzüge der Anwendung digitaler Methoden in diesem Bereich.

Obwohl schon vor mehr als einem halben Jahrhundert vielfältige Ansätze dazu gemacht wurden,¹³ hat sich die Literaturwissenschaft im Gegensatz zur statistischen Sprachanalyse, wie sie in der Computerlinguistik weit fortgeschritten ist, in der Anwendung von Informationstechnologie bei der Textanalyse, geschweige denn bei der Institutionalisierung einer ›Computerliteraturwissenschaft‹ bisher eher zurückhaltend gezeigt.¹⁴ Bei der digitalen Edition ist die Verankerung in einer traditionellen textphilologischen Teildisziplin unübersehbar. Hier kann die digitale Technologie nicht nur jene Prozesse unterstützen und vereinfachen, die zuvor in mühsamer Handarbeit durchgeführt werden mussten. Sie kann auch dazu beitragen, neue Formate für die Erfassung, Darstellung und Analyse der Mehrschichtigkeit textueller Strukturen zu erarbeiten. Dementsprechend präsent, gefördert und anerkannt sind die einschlägigen Projekte im Feld der Digital Humanities.¹⁵

Wie wäre weiter vorzugehen? Der Beitrag schlägt einen Weg vor, der die digitale Texterfassung mit den Methoden korpusbasierter Analyse enger zusammenführt. Dabei gehe ich vor allem auf die mit dem World Wide Web gegebenen Möglichkeiten ein. Mit dem Internet liegt ein potentiell Medium der Erkenntnisgewinnung vor: erstens ist es ein immenser, weithin zugänglicher Datenspeicher, und zweitens bietet es mit der Verknüpfbarkeit von Dokumenten die Möglichkeit, die Verweisstruktur von Texten auf neue Weise sichtbar zu machen. Es erlaubt mittels Datenvernetzung eine Darstellung der von Derrida fokussierten »differentiellen Verweisung von einer [Text]Spur auf die andere«¹⁶ im mehrdimensionalen

¹¹ <http://gutenberg.spiegel.de/>.

¹² <https://books.google.com/>. Vgl. auch die Textscans, die der [Hathi Trust](#) online gestellt hat.

¹³ Vgl. neben den Ansätzen bei den oben genannten Autoren auch noch [Kenny 1982](#). Auch er kommt wie [Busa](#) aus kirchlichem Kontext.

¹⁴ Die Entwicklung von Methoden digitaler Literaturanalyse und die Anzahl einschlägiger Publikationen hat sich seit Beginn dieses Jahrhunderts allerdings rapide entwickelt. Das im Wintersemester 2011/2012 an der Universität Göttingen entstandene [Wiki zur digitalen Textanalyse](#) beschreibt aus den schon zur Verfügung stehenden Tools zur digitalen Textanalyse eine Auswahl von acht. Die 2011 vom Cologne Center for eHumanities (CCEH) herausgegebene Broschüre [Digitale Geisteswissenschaften](#) skizziert zwar einschlägige Studiengänge an bereits 13 Studienorten, darunter aber noch kein Programm mit explizitem Schwerpunkt auf der Literatur(analyse). Lediglich die TU Darmstadt bot damals einen MA Studiengang »Linguistic and Literary Computing« (jedoch ohne speziellen Bezug zur Literaturanalyse) an. Am nächsten kommen diesem Interesse bisher die an der Universität Würzburg angebotenen BA und MA Studiengänge »Digital Humanities«.

¹⁵ Zu verweisen ist hier im Besonderen auf die speziell für die Geisteswissenschaften eingerichteten Forschungsumgebungen [TextGrid](#) (gefördert 2006–2105) sowie als deren Nachfolgerin die Digitale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften [DARIAH-EU](#) bzw. [DARIAH-DE](#).

¹⁶ Derridas im Gespräch formulierte Darstellung der Grundidee (literaturwissenschaftlicher) Dekonstruktion war Ausgangspunkt der im Folgenden entwickelten Überlegungen. Daher soll sie hier noch einmal im Zusammenhang zitiert werden: »Das, was ich also Text nenne, ist alles, ist praktisch alles. Es ist alles, das heißt, es gibt einen Text, sobald es eine Spur gibt, eine differentielle Verweisung von einer Spur auf die andere. Und diese Verweise bleiben nie stehen. Es gibt keine Grenzen der differentiellen Verweisung einer Spur auf die andere. Eine Spur ist weder eine Anwesenheit noch eine Abwesenheit. Folglich setzt dieser neue Begriff des Textes, der ohne Grenzen ist ich habe deshalb gesagt, auch als scherzhafte Bemerkung, es gäbe kein Außerhalb des Textes -, folglich setzt dieser neue Begriff des Textes voraus, daß man in keinem Moment etwas außerhalb des Bereiches der differentiellen Verweisungen fixieren kann, das ein Wirkliches, eine Anwesenheit oder eine Abwesenheit wäre, etwas, das nicht es selbst wäre, markiert durch die textuelle *différance*, durch den Text als *différance* mit einem »a.« (zit. nach [Engelmann 1988](#), S. 107f.) Zur Auseinandersetzung mit Derrida im Kontext der Digital Humanities vgl. u.a. [Baillot 2015](#).

Raum und bietet damit weitreichende Möglichkeiten für eine Literaturanalyse, die außer von der Literaturwissenschaft auch von anderen Disziplinen effektiv genutzt werden könnten.

2. Das World Wide Web als Möglichkeitsraum digitaler Literaturanalyse

Dem vor etwas mehr als 25 Jahren geschaffenen und stetig erweiterten digitalen Datenraum des World Wide Web (Netzes)¹⁷ werden unter anderem folgende Eigenschaften zugeschrieben:

- unbegrenzte Kapazität für langfristige Speicherung von Daten;
- unbegrenzte Verfügbarkeit der Daten in Raum und Zeit; im Prinzip ist es technisch möglich jederzeit von jedem Ort der Erde auf jedes Dokument zuzugreifen;
- freie Zugänglichkeit (open access) für eine große Menge von Daten, Tendenz steigend.

Das Netz stellt kein statisches Archiv dar, sondern einen virtuellen Speicherraum, der permanent neu organisiert wird. Die dabei vorgenommenen Operationen bewirken

- die anhaltende Evolution des World Wide Web durch beständige Erweiterung respektive Veränderung der schon vorhandenen Daten; jeder kann Daten eingeben oder, sofern diese zugänglich sind, bestehende Daten verändern;
- die Generierung separat evaluierbarer Metadaten;
- die Bildung der Netzstruktur durch Verknüpfung der Daten untereinander.

Prinzipielle Offenheit und unvergleichliche Kapazität des World Wide Web einerseits, die permanent weiter entwickelte Vielfalt und Variabilität der darauf bezogenen Operationen andererseits, legen es nahe zu überlegen, ob und inwieweit die Daten im World Wide Web einen neuartigen Möglichkeitsraum für die Literaturanalyse bilden, in dem sich die mehrdimensionale Vernetzung literarischer Texte anders als bisher aufzeigen, darstellen und erforschen lässt. Ansätze zur Visualisierung literaturwissenschaftlich relevanter Daten im Netz sind schon partiell erprobt worden und werden zunehmend weiter entwickelt.¹⁸ Eine Evaluierung der mit dem World Wide Web verbundenen Möglichkeiten wissenschaftlicher Analyse literarischer Texte hingegen steht noch aus. Im folgenden sollen bereits erprobte Verfahren wie auch absehbare Perspektiven einer durch die digitale Textverarbeitung unterstützten, netzbasierten Literaturwissenschaft skizziert werden, wobei der Schwerpunkt meiner Überlegungen auf der Verknüpfung von Daten liegt.

3. Texte im Netz

¹⁷ Berners-Lee / Fischetti 1999.

¹⁸ So gibt es beispielsweise im Bereich der Erforschung von Stadt(-literatur) schon diverse Darstellungen im Netz. Vgl. neben dem von Barbara Piatti u.a. erarbeiteten [Literaturatlas Europas](#), das von Todd Presner und anderen an der UCLA betriebene Projekt [Hypercities](#) oder eine typische Sammlung von stadtbezogener Literatur, hier zu [Edinburgh](#).

Ein – überraschend kleiner – Teil der im Netz mittels Informationstechnologie erzeugten und verwalteten Daten stammt aus Texten in natürlichen Sprachen. Literaturwissenschaftlich im engeren Sinne relevant sind

- Digitalisate, d.h. (literarische) Texte, die ursprünglich in anderen Formaten ediert wurden und weiter existieren, die nun zusätzlich im Netz verfügbar gemacht werden (DTA, Gutenberg, Google Books u.a.); nicht alle dieser Digitalisate sind schon maschinenlesbar, vielfach handelt es sich um Textscans;
- Digitale Literatur, d.h. (literarische) Texte, die ursprünglich in digitaler Form verfasst und ins Netz gestellt wurden; das Spektrum reicht von Weblogs über Netzliteratur(-magazine) bis zu hochgradig verlinkten Formen wie Hyperfiktio¹⁹;
- Texte und weitere Daten (Bilder, Sounddateien u.a.), die literaturwissenschaftlich relevante Informationen enthalten; traditionelle Hilfsmittel wie Wörterbücher, biographische Lexika, Bibliographien, Enzyklopädien und andere Nachschlagewerke finden sich heute zunehmend durch Internetdatenbanken wie Wikipedia ersetzt.

Die Anzahl der im Netz verfügbaren und literaturwissenschaftlich evaluierbaren Texte nimmt rapide zu. Texte, die bisher mangels Zugänglichkeit kaum wissenschaftliche Berücksichtigung gefunden haben, wie zum Beispiel nicht-kanonisierte Texte, Literaturen in ›kleinen‹ Sprachen, Texte in entlegenen, vergriffenen oder vergessenen Ausgaben etc., werden zunehmend greifbar und sollten in einschlägige Untersuchungen einbezogen werden. Damit entsteht ein enormer Druck auf die Philologien, die an der Textbasis veränderte Ausgangslage ihrer Forschungen zur Kenntnis zu nehmen und ihr methodisch zu entsprechen.

Untersuchungen, die allgemeine Aussagen beispielsweise zu einer Epoche, zu einer Gattung, oder zur Literatur einer Sprache machen wollen (als Standardbeispiel: »Der englische Roman des 19. Jahrhunderts«) müssen sich heute nicht mehr nur auf die schon bekannten respektive kanonisierten Texte beschränken, sondern können die bisherigen Erkenntnisse an großen Korpora überprüfen und weiter ausdifferenzieren. Trotz der zunehmenden Verfügbarkeit von Texten im Netz repräsentieren die vorhandenen Daten allerdings noch lange nicht das gesamte Spektrum der in der Literaturwissenschaft zu untersuchenden Literaturen. Restriktiert bleibt weiterhin der Zugang insbesondere zu urheberrechtlich geschützten Werken und damit zum größten Teil der Literatur des 20. und 21. Jahrhunderts. Digitale Literaturanalysen stützen sich deshalb bisher mehrheitlich auf vor 1900 erschienene Texte.

Gleichwohl werden Untersuchungen zur Gegenwartsliteratur die Webpräsenz zeitgenössischer Texte nicht ignorieren können. Dies betrifft Autoren und Autorinnen wie Elfriede Jelinek, die ihre Texte primär im Netz veröffentlichen,²⁰ ebenso wie mit digitalen Technologien entwickelte Textformate, welche sich im Druck gar nicht oder nicht mehr genreadäquat darstellen lassen wie Hyperfiktio oder Twitterature.²¹ Sie verlangen ebenso nach neuen Methoden der literaturwissenschaftlichen Evaluierung wie die digitalisierten Texte.

¹⁹ Vgl. dazu u.a. Herring et al. 2005; Suter 2000; Böhler 2001; Giacomuzzi 2012.

²⁰ <http://www.elfriedejelinek.com/>.

²¹ Aciman / Rensin 2009.

4. Netzstruktur und Textstruktur

Wenn digitalisierte, d.h. in maschinenlesbare Formate gebrachte, Texte ›ins Netz gestellt‹ werden, können sie dort mit anderen Daten verknüpft werden. Eben durch diese Verknüpfung wird die Netzstruktur des World Wide Web ermöglicht.

Hierbei erscheint bemerkenswert, dass der Aufbau des Netzes wie der Umgang mit dem Netz Parallelen zum Aufbau von und dem Umgang mit (literarischen) Texten aufweist, d.h. das Netz hat eine Reihe strukturell relevanter Eigenschaften, die sich mit denjenigen von in natürlichen Sprachen abgefassten Texten vergleichen lassen.

Ein verständiger Leser bezieht das Gelesene während der Lektüre unaufhörlich auf seinen Wissens- respektive Erfahrungshorizont inklusive früherer Lektüren; nur so kann er sich den Sinn des Textes erschließen.²² Während des Lesens ›vernetzt‹ er das Gelesene also unaufhörlich ›in seinem Kopf‹ mit einzelnen Elementen aus diesem Vorwissen: Er verknüpft den Text in einem Netz, dessen Konstellation durch jede neu aufgenommene Information wieder verändert wird.²³ Dieses Netz ist virtuell oder (im Sinne von Benedict Anderson) imaginär,²⁴ da es eine Menge von Informationen enthält, die vielen Menschen gemeinsam oder gleichzeitig zugänglich sind. Von jedem einzelnen Menschen wird aber nur ein für ihn spezifischer, einmaliger Teil davon erfasst. Ähnlich verhält es sich auch mit dem World Wide Web.

Gewöhnlich verstehen wir das World Wide Web als Akkumulation großer Mengen an digitalen Daten, die in unterschiedlicher Dichte miteinander verknüpft und weltweit zugänglich sind. Das Verhältnis der im World Wide Web verwalteten Daten zu dem, was wir als ›Welt‹ auffassen, ist mindestens zweifach strukturiert:

- semiotisch/symbolisch: im Netz werden Informationen, die sich auf etwas außerhalb davon Liegendes beziehen, in Daten übersetzt und verwaltet; der ›Entortung‹ der in Daten umgewandelten Informationen im geopolitischen Raum korrespondiert deren neue Verortung im global zugänglichen Webkontext (URI); die ›Entzifferung‹ der Informationen bedarf der Zuordnung zu Systemen, innerhalb derer der semiotische/symbolische Code Gültigkeit hat und Bedeutung entfalten kann;
- operationell: die Daten werden durch Operationen natürlicher oder künstlicher Intelligenz erzeugt und verwaltet; diese Operationen auszuführen ist technisch im Prinzip jedem Nutzer möglich.

Texte lassen sich desgleichen als weitreichend verknüpfte Akkumulationen von Informationen verstehen, die in natürlichen Sprachen repräsentiert sind und die als Daten erfasst werden können. Auch hier ist der ›Weltbezug‹ semiotisch/symbolisch; die jeweilige Entzifferung der Informationen aus den ihm vorliegenden Daten obliegt dem Leser, der dafür

²² Vgl. dazu sowohl Gadamers Hermeneutik (1960) wie die Dekonstruktion nach Derrida (1976, 1983).

²³ Diese Vorstellung formuliert sinngemäß erstmals 1919 T.S. Eliot in seinem Essay *Tradition and The Individual Talent* (Eliot 1932).

²⁴ Anderson 1983.

ebenfalls eine Reihe von Operationen ausführen muss, durch welche die Zeichen des Textes in Sinn übersetzt werden (vgl. **Abschnitt 7**).

Wesentlich in diesem Zusammenhang ist das für beide, Netz wie Text, konstitutive System der Verweisung bzw. der Verknüpfung: (Natürliche) Sprachen und darin formulierte Texte können nur dann Bedeutung generieren und kommunizieren, wenn sie auf Außersprachliches verweisen (d.h. sie repräsentieren dieses zeichenhaft), und wenn sie sich mit anderen sprachlichen Aussagen respektive Texten operationell verknüpfen lassen und in dem imaginären Netz identifizierbar werden. Texte sind autorisierte Knotenpunkte (nodes) im Netz sprachlicher Kommunikation. Sie generieren Bedeutung, indem sie sich sprachlich auf vereinbarte Zeichen- und Wertsysteme beziehen. Diese Bedeutung ist nie einsträngig; die Verknüpfungen müssen auf mehreren Ebenen des Textes vom Leser aktualisiert und weiter verfolgt werden (können).

Literaturwissenschaftliche Untersuchungen zielen auf Erkenntnis solcher Verknüpfungsstrukturen, die in literarischen Texten hochgradig elaboriert sind. D.h. es geht hier um den Nachweis der Mehrdimensionalität, die literarische Texte in ihren vielfältigen Verknüpfungen mit anderen Texten und Systemen herstellen und die sie als ein Qualitätsmerkmal auszeichnet.

Daraus ergibt sich die Frage, inwiefern die Verfügbarkeit digitalisierter Texte im Netz für den Nachweis, die Darstellung und die Analyse ihrer Mehrdimensionalität genutzt werden kann, d.h. inwiefern das World Wide Web nicht allein als Datenspeicher, sondern auch als Medium literaturwissenschaftlicher Erkenntnis fungieren kann.

5. Literaturwissenschaftliche Textanalyse

Um die Mehrdimensionalität von (literarischen) Texten aufzuzeigen, versuchen literaturwissenschaftliche Textanalysen

- Texte strukturell zu erfassen, d.h. ihre Eigenschaften und deren Relationen zueinander zu bestimmen, zu kategorisieren, und die Texte diesen Eigenschaften entsprechend zu klassifizieren;
- Texte hermeneutisch zu verstehen, d.h. den ›mehrfachen Schriftsinn‹, die Bedeutung von Inhalt und Aussage in Abhängigkeit von der gewählten sprachlichen Ausdrucksform, zu bestimmen;
- Texte kontextuell zu verorten, d.h. ihre Eigenschaften auf sprachlich und kulturell bedingte Ordnungen (Systeme) und deren historische Verläufe zu beziehen und einzuschätzen.

Die literaturwissenschaftliche Analyse besteht demnach (i) aus der Isolierung, Beschreibung und Klassifizierung ausgewählter Eigenschaften von Texten, und (ii) aus der differenzierenden Zuordnung dieser Merkmale zu bedeutungstragenden Systemen. Die Zuordnung (Verknüpfung) ist für das Verstehen eines Textes unabdinglich; sie kann nur dann

erfolgen, wenn die Bezugsgrößen (richtig) identifiziert werden. Die Mehrdimensionalität eines (literarischen) Textes – darin liegt ein Teil seines Reizes – impliziert aber durchaus auch multiple Möglichkeiten der Zuordnung. Diese mögliche Vielfalt muss in der Analyse mit erfasst werden. Sie zielt somit auf die komplexe Erfassung dessen, was ich im Folgenden das »Referenzprofil« eines Textes nennen möchte.²⁵

6. Literaturwissenschaftliche Analyse digitalisierter Texte

Literarische Texte mit Unterstützung von Informationstechnologie zu bearbeiten heißt, literaturwissenschaftliche Methoden teilweise Computerprogrammen zu übertragen in der Erwartung, das Erreichen fachspezifischer Arbeitsziele mit deren Unterstützung zu erleichtern oder zu verbessern. Dies betrifft insbesondere mehr oder weniger mechanische, systematisch durchzuführende Vorgänge wie zum Beispiel das Zählen und Sortieren festgelegter Einheiten, wo über die Anwendung von Programmen eine größere Vollständigkeit und geringere Fehlerrate erzielt werden können. Ein gutes Beispiel dafür sind die digitalen Editionen, die die Textaufzeichnung und -speicherung wie die Darstellungs- und Zugriffsmöglichkeiten gerade bei Werkausgaben erheblich verbessert haben.

Für die computergestützte literaturwissenschaftliche Textanalyse wird der Analysevorgang, den der wissenschaftliche Leser vollzieht, in einzelne Schritte zerlegt, von denen einige durch Computerprogramme ausgeführt werden können. Gegenwärtig geht es vor allem darum, Texte von Programmen lesen und nach bestimmten Wörtern oder Textstellen durchsuchen zu lassen. In einem zweiten Schritt geht es darum, Programme zu entwickeln, die die Verknüpfungen innerhalb eines Textes und zwischen Texten erkennen und annotieren können.

6.1 Vorbereitung maschinell durchgeführter Suchabfragen

Um Texte von Computerprogrammen durchsuchen zu lassen, müssen sie in maschinell analysierbare Textformate gebracht werden.

Zuerst werden die zu untersuchenden Texte mittels OCR (Optical Character Recognition) in ein maschinenlesbares Format gebracht. Dieses Verfahren ist seit Längerem gut etabliert, kann weitgehend automatisiert erfolgen und bedarf je nach Textvorlage manueller Nachbearbeitung in meist überschaubarem Ausmaß.

Anschließend werden Texteigenschaften in einer Weise ausgezeichnet (annotiert), die ihre Abfrage durch Computerprogramme ermöglicht. Auch diese Auszeichnung kann durch bereits erprobte Programme durchgeführt werden, bedarf aber ebenfalls der manuellen Ergänzung.

²⁵ Vgl. dazu auch Ramsay 2011.

Die Annotierung baut einerseits auf Verfahren und Erkenntnissen auf, wie sie zuerst in der computergestützten Edition Anwendung fanden (XML, TEI).²⁶ Des Weiteren kommen Verfahren aus der Computerlinguistik zur Anwendung (FSPar, Stanford Core NLP, CLAWS), wenn die digital bearbeiteten Texte zusätzlich mit Tags für automatisiert durchzuführende Abfragen versehen werden (z.B. Lemmatisierung, part of speech tags, TreeTagger, **semantic tags**).

Die Annotierung entspricht andererseits dem Ansatz klassischer und bis in jüngere Zeit immer noch weiter verfeinerter literaturwissenschaftlicher Kommentare zu Hauptwerken der Antike, des Mittelalters oder den Klassikern der Moderne (die jüdische Thora als Kommentar zum Alten Testament; Marsilius Ficino zu Platon; Boccaccio zu Dante...). Diese Kommentare verfahren ebenfalls lemmatisierend; sie liefern Annotationen zu einzelnen Worten oder Textstellen. Sie haben die Aufgabe, sprachliche und textuelle Besonderheiten sowie möglicherweise unverständliche Stellen hervorzuheben, zu beschreiben, zu erläutern, aufzuschlüsseln und durch Belege aus anderen Quellen zu erklären. Ging es zunächst darum, den ›mehrfachen Schriftsinn‹ eines hieratischen Textes zu (ent-)bergen, so dienen spätere Kommentierungen der Entzifferung komplexer, hybridisierter sprachlicher Strukturen²⁷ oder intertextueller Bezugnahmen²⁸, respektive der Bewahrung sprachlich-kulturellen Wissens, das zeitgenössischen Lesern oft nicht mehr zur Verfügung steht. Bei der digitalen Annotierung werden Erläuterungen in Form von Querverweisen eingefügt, die im Kontext des World Wide Web als verfolgbare Verknüpfungen (Links) fungieren können (vgl. **Abschnitt 7**).

Die Entwicklung von für die Literaturanalyse geeigneten Annotationsverfahren stellt gegenwärtig eine der zentralen Aufgaben in der Entwicklung von Methoden digitaler Literaturanalyse dar. Eines der wesentlichen Probleme dabei ist die Wahl der geeigneten Granularität. Die Granularität muss gröber sein, als die in der Korpuslinguistik angewandte (typische Einheiten wären Wort oder Satz); sie sollte feiner sein als die etablierten literaturwissenschaftlichen Kategorien (typische Einheiten wären Gattungen oder Schreibweisen). Mit der Quantität sollte auch die Distribution der gesuchten Größen innerhalb der Texte erfasst werden, indem man kleinere Texteinheiten (z.B. Absätze) bei der Auszeichnung mit berücksichtigt. So kann man zum Beispiel bei der Untersuchung des Auftretens von Personen in einer Netzwerkanalyse das Auftreten der Personen in Relation zum jeweiligen Raum registrieren und dabei gleichzeitig Texteinheiten (z.B. Absätze) notieren. So können personenbezogene Handlungsstränge innerhalb eines Textes abgegrenzt und verglichen werden.²⁹

6.2 Bearbeitung, Auswertung und Darstellung von Suchergebnissen

²⁶ Sahle 2013, *passim*.

²⁷ Zu Joyce vgl. Gifford / Seidman 1989.

²⁸ Zu Celan vgl. Lehmann / Ivanovic 2003.

²⁹ Ansätze dazu erproben wir gegenwärtig in einem auf die Analyse der Werke Ilse Aichingers bezogenen Projekt, vgl. dazu Ivanovic / Frank 2015.

Digitale Literaturanalysen, die auf der Auszeichnung von Texteigenschaften beruhen, entsprechen der unter **Abschnitt 5** genannten strukturellen Texterfassung. Das vereinheitlichte Verfahren lässt sich als automatisiertes Vorgehen auf eine unbegrenzt große Anzahl von Texten anwenden und erlaubt die gleichen überprüfbaren Abfragen in einem Textumfang, der weit über der menschlichen Lesekapazität liegt.

Wenn nicht nur ein einzelner Text, sondern wenn größere Textkorpora mit derselben Methode untersucht werden, d.h. wenn dieselbe Suchabfrage an vielen Texten durchgeführt wird, kommt man zu statistisch auswertbaren Ergebnissen, die mit den in der Statistik oder in der Informatik üblichen Mitteln dargestellt werden können: Graphen, Karten, Bäume, Netze, Blasen etc. visualisieren die erzielten Suchergebnisse. Ausführliche Beispiele dafür diskutiert Franco Moretti in seinen Büchern *Atlas des europäischen Romans* (1999) sowie *Graphs, Maps, Trees* (2005) und *Distant Reading* (2013). Solche Darstellungen gelten als attraktiv und anschaulich. In den Literaturwissenschaften sind sie bisher nur sehr begrenzt eingesetzt worden, finden aber zunehmend Akzeptanz. Dabei darf nicht übersehen werden, dass sie Ergebnisse von Suchanfragen zwar veranschaulichen, aber selbst noch der Analyse bedürfen.

7. Literaturanalyse im Semantic Web

Die digitale Literaturanalyse stützt sich in der Regel auf digitalisierte Texte oder Textkorpora, die von einzelnen Arbeitsgruppen oder Institutionen erarbeitet wurden. Diese Texte sind aber meist nicht frei zugänglich, und sie sind vor allem nicht mit externen Daten verknüpfbar. Digitalisierte, d.h. maschinenlesbare Texte, auf die im Netz zugegriffen werden kann, erlauben weitere Analysemöglichkeiten durch Datenvernetzung im Semantic Web. Diese Vorgehensweise geht auf Sir Tim Berners-Lee zurück, der 2001 vorgeschlagen hat, die Möglichkeiten des World Wide Web mit semantischen Annotationen zu ergänzen, um die in allen Dokument-Texten implizit enthaltenen (für menschliche Leser verständlichen) Informationen explizit (d.h. für Programme auswertbar) zu machen.³⁰

Das Semantic Web beruht darauf, dass bestimmte Informationen in digitalen Dokumenten mit einer URI identifizierbar gemacht werden. Über diese Identifikatoren kann in anderen Dokumenten auf sie verwiesen werden, und zwar in beliebiger Abfolge. Damit entsteht wiederum ein Netz (hier besser: ein Graph) von Verweisen, die ihrerseits wieder annotiert und abgefragt werden können.³¹ Der dafür zur Verfügung stehende Standard RDF³² ist ein nicht spezifisch für die Computerphilologie entwickeltes Auszeichnungsformat,³³ das in der Literaturanalyse aber vor allem deshalb mit Erfolg eingesetzt werden kann, weil es über die Erfassung der sprachlich-textuellen Merkmale hinaus die vollständige Erfassung der Struktur und der Kontexte eines Texts sein Referenzprofil ermöglicht. Die ebenfalls standardisierte Abfragesprache SPARQL erlaubt dann beliebige Suchen nach Mustern in mit RDF strukturierten Dokumenten.

³⁰ Berners-Lee et al. 2001.

³¹ Hitzler et al. 2008.

³² Vgl. den RDF Primer <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>.

³³ Curé / Blin 2015.

7.1 Erstellung eines vernetzten Texts

Die Zuordnung der evaluierten Textmerkmale zu textexternen Referenzgrößen wurde in **Abschnitt 5** als notwendige Voraussetzung für das Verstehen eines Textes – das Erfassen seiner Bedeutung³⁴ – genannt. Sie realisiert sich als Operation der Verknüpfung mit jeweils als zugehörig erachteten Bezugsgrößen, die der Leser vollziehen muss.

Für diese kontextualisierende Zuordnung mit Hilfe von Computerprogrammen bietet die Auszeichnung (literarischer) Texte im Netz die entscheidende Grundlage.

Texte, die im Netz zugänglich sind, können zwar zusätzlich mit Hyperlinks ausgestattet werden, die beim Lesen eine Weiterleitung zu anderen Webeinträgen ermöglichen. Hyperlinks sind für Suchabfragen jedoch ineffizient. Texte mit dem Standard RDF zu annotieren, ermöglicht hingegen eine effiziente Analyse auch riesiger Textmengen. Milliarden von Triples können innerhalb von Sekunden nach komplexen Mustern abgefragt werden. RDF erlaubt auch die Verknüpfung mit externen Datenbanken wie Wikipedia, Google Maps, Filmdatenbanken etc. Solche externen Datenbestände können, soweit sie im RDF Format vorliegen (wie dies etwa bei DBpedia der Fall ist), nicht nur aufgefunden, sondern wiederum in Abfragen verwendet werden. So ist es zum Beispiel umstandslos möglich, alle Regisseure und Schauspieler, die in einem Text erwähnt werden, herauszufinden und zu prüfen, ob ein Film zum fiktiven Zeitpunkt der Handlung in einem Text bereits öffentlich gezeigt worden ist oder nicht.

Der Entwicklung von Verfahren zur (automatisierten) Erkennung und Auszeichnung der entsprechenden Textstellen gilt daher gegenwärtig unser vorrangiges Interesse. Ziel wäre es, diese Verfahren so weit zu entwickeln, dass die Verweisstruktur eines Textes, also dessen Referenzprofil, möglichst automatisch und möglichst vollständig erfasst werden kann. Dies betrifft Referenzen auf Personen, Orte, Zeitangaben, Ereignisse, Sachverhalte etc., aber auch Bild- oder Textreferenzen, Zitate, intertextuelle Bezugnahmen, Anspielungen, Redewendungen u.v.m. Letztlich geht es um die Verknüpfung des Textes mit Informationen, wie sie bereits in klassischen Textkommentaren beigebracht wurden, also um die Erfassung struktureller Merkmale, um Wort- und Sachverweise, intertextuelle Bezüge etc., nicht aber um Abbildung des subjektiven Assoziationsraums, in den jeder private Leser *seine* Lektüre eines Textes notwendigerweise einbetten wird.

7.2 Literaturanalyse im Semantic Web

Die mit RDF ausgezeichneten Dokumente können mit der Abfragesprache SPARQL untersucht werden, wobei sich über die Verbindung zu anderen Datenbanken durch die spezifische Strukturierung (Verknüpfbarkeit) der Suchkriterien erweiterte Suchmöglichkeiten ergeben. So erlaubt beispielsweise die Erfassung sämtlicher in einem Korpus genannter

³⁴ Vgl. zu dieser Problematik u.a. Jannidis et al. 2003, hier S. 3–30 und insbesondere die Einleitung von Gerhard Lauer zum Abschnitt »IV. Historische Aspekte literarischer Bedeutung«, wo er auf die Praxis der »Bedeutungszuweisung« einght, S. 559–565.

Ortsnamen durch RDF eine Verknüpfung mit Gazetteers (wie OpenStreetMap), um die geographischen Koordinaten zu bestimmen und daraus Karten z.B. mit Google Maps oder Google Fusion Tables zu zeichnen, über die sich die Frequenz und Distribution der in den erfassten Texten auftretenden Ortsnamen relationieren lassen. Der DARIAH-DE [GeoBrowser](#) erlaubt darüber hinaus das Zeichnen von Karten, die Veränderungen in der Zeit anzeigen.

Da die Texte im Netz evaluierbar sind, werden die ermittelten Referenzierungen dann aber nicht nur vom Einzeltext her lesbar. Es lässt sich auch in umgekehrter Richtung evaluieren, welche Texte sich auf einen bestimmten Referenzpunkt (z.B. ein bestimmtes Zitat oder einen Ort etc.) beziehen.³⁵

7.3 Digitale Literaturanalyse durch andere Disziplinen

Diese Methoden der Literaturanalyse könnten sich schließlich weit über ästhetisch-literaturwissenschaftliche Fragestellungen hinaus als nützlich erweisen. Denn die so erfassten literarischen Texte sind auf diese Weise der Grundanlage des World Wide Web entsprechend für jeden überall zugänglich und werden so auch für andere Disziplinen evaluierbar. Dies kann neben den historischen Wissenschaften auch für alle weiteren Forschungsrichtungen von Interesse sein, die sich mit Fragen der menschlichen Kultur respektive mit der Entwicklungsgeschichte des Menschen befassen, also auch für die Lebenswissenschaften sowie alle Fachrichtungen, die Kognitionsprozesse untersuchen. Wie Menschen beispielsweise Raum(-strukturen) erfahren, erfassen, beschreiben, repräsentieren und kommunizieren, ist eine für alle Raumwissenschaften wichtige Frage.³⁶ Sie könnte an (literarischen) Texten mit großem Gewinn untersucht werden.³⁷

Obwohl in literarischen Texten eine Fülle an (nicht nur kulturellem) Wissen enthalten und greifbar ist, wird Literatur in dieser Hinsicht bisher nur partiell evaluiert. Durch die Indizierung literarischer Texte im Semantic Web können auch andere Disziplinen deren Daten systematisch für ihre Erkenntnisbemühungen nutzen.

8. Perspektiven

8.1 Abschied vom biblischen Zeitalter

Es zeichnet sich ab, dass die hier in Betracht gezogenen Methoden literarische Texte zu analysieren eher darauf ausgerichtet sind, große Mengen von Texten zu bearbeiten als einzelne Werke zu analysieren. Das mit den Computerprogrammen gegebene Potential, Textmengen in einem Ausmaß zu untersuchen, das kein einzelner Leser bzw. Leserin je überblicken könnte, ist eine attraktive Herausforderung für die heutige Literaturwissenschaft. Der Positionswechsel

³⁵ Vgl. dazu auch Barbaresi 2016.

³⁶ Günzel 2009.

³⁷ Entsprechend multidisziplinär ausgelegt sind einschlägige Initiativen wie DARIAH-Eu.

weg von der hermeneutischen Auslegung des ›einen‹ Buchs zum »Distant Reading« hunderter von Texten, wie dies Franco Moretti experimentell pointiert hat,³⁸ ermöglicht neuartige Fragestellungen, die keineswegs als disziplinfremd oder -fremd eingestuft werden können. Im Gegenteil, letztlich hat uns Morettis *Atlas* ›den‹ europäischen Roman auf neue Weise nahe gebracht.³⁹

Wir hoffen, dass sich bei einer solchen Betrachtungsweise ›das System‹ Literatur, und das würde heißen die Funktionen und das Funktionieren literarischer Texte im Gesamtgefüge kultureller Gemeinschaften, in seinen Konturen deutlicher erkennen lässt. Literarische Texte sind weder Unikate noch Einzelgänger. Mehrfach reproduziert erscheinen sie an vielen Orten gleichzeitig. Sie treten im Gefüge literarischer Traditionen und aktueller Strömungen auf und bezeugen historische Ereignisse und Entwicklungen. Sie partizipieren an den Diskursen der Gesellschaft und sie sind Teil ökonomischer Beziehungen, die Entstehung und Rezeption ebenso wie Thema und Argumentation der Texte bestimmen. Diese Art der Verankerung oder Vernetzung von Texten in größeren systemischen Zusammenhängen sichtbar zu machen, kann mit den hier skizzierten Methoden in einer Weise geleistet werden, die die Evaluierung großer Textgruppen wie einzelner Texte ermöglicht und erfordert.

8.2 Die Vernetzung des Texts

Gerade jene monumentalen Einzeltexte, denen in der Kulturgeschichte größte Bedeutung zuerkannt wird, scheinen intensiv systemisch vernetzt; dies gilt für die Bibel ebenso wie für Homer, für Dante wie für die Märchen aus Tausendundeiner Nacht, für Shakespeare wie für Joyce. Evaluiert man das Referenzprofil respektive den Verknüpfungsgrad eines einzelnen Textes, lässt sich dessen Verhältnis zum ›unendlichen‹ Text der »Weltliteratur« wie auch sein referentielles Verhältnis zur »Welt« genauer bestimmen. Dies könnte nahelegen, dass sich aus der Ermittlung des jeweiligen Vernetzungsgrads neue Kriterien zur literarischen Wertung respektive zur Bestimmung der ästhetischen Qualität von Texten erarbeiten lassen.

Vergleicht man jedoch Qualität und Intensität der Verknüpfungen, ergibt sich eher eine ganze Reihe von Relationen als eine Werteskala. Sie können beispielsweise im Hinblick auf die Mechanismen der Textrezeption ebenfalls statistisch untersucht werden: einem hohen Grad an Verknüpfung dürfte ein hohes Maß an Anschlussmöglichkeiten für viele Leser entsprechen; ein geringer Grad an Verknüpfung hingegen kann große Eigenständigkeit der literarischen Struktur signalisieren; sehr spezielle Verknüpfungen können hermetische Tendenzen anzeigen; abstrakt-allgemeine Merkmale könnten eine weiträumige Rezipierbarkeit ermöglichen etc. Es sind auch bisher nicht angestellte Überlegungen, die sich hier als weiterführende Forschungsfragen abzeichnen. Für deren Evaluierung können schließlich auch Metadaten wie die Registrierung von Abfragehäufigkeit einzelner Texte oder einzelner Verknüpfungen genutzt werden, sofern man deren Erhebung ins Programm implementiert hat.

³⁸ Moretti 2013.

³⁹ Moretti 1999.

8.3 Alte und neue Aufgaben

Der skizzierte Ansatz digitaler Literaturanalyse durch Datenvernetzung mündet in eine Art relationierende Literaturanalyse, in der die Eigenschaften einzelner Texte im Verhältnis zu Textgruppen evaluiert und auf die Systeme bezogen werden, in die sie produktions- wie rezeptionsästhetisch eingebettet sind. Nach wie vor wird ein Schwerpunkt der Analyse auf der genauen Untersuchung von Textstrukturen und deren Erfassung durch literaturwissenschaftliche Kategorien liegen.

Eine der wesentlichen Ideen dieser Methode ist es, das in und mit literarischen Texten kommunizierte Wissen schärfer hervortreten zu lassen und systematischer als bisher zu bergen. Dies betrifft einerseits spezifisch literaturwissenschaftlich relevantes Wissen, wie etwa die Wanderung ästhetischer Konzepte und Ausdrucksformen, und andererseits die Möglichkeit außerliterarische Sachzusammenhänge abzufragen, die in Texten abgebildet oder thematisiert werden wie zum Beispiel die Ermittlung von Daten zu Kaufpreisrelationen, wie sie für die Wirtschaftsgeschichte von Belang sind, oder Daten zur Erforschung des Klimawandels und Beschreibungen medizinisch relevanter Krankheitsbilder. Die Daten der »vernetzten«, d.h. für automatisierte Abfragen vorbereiteten Texte im World Wide Web können und sollten also auch von anderen Disziplinen genutzt und damit das Potential von Literatur in weiterem Umfang als bisher erfasst und wissenschaftlichen Erkenntnisbemühungen verfügbar gemacht werden.

9. Conclusio

Für die Entwicklung digital gestützter Literaturanalysen scheint es sinnvoll, an bereits bewährte Methoden und Ansätze der Computerlinguistik respektive der Korpuslinguistik anzuknüpfen. Wie viele andere Bereiche der Computerwissenschaften wird auch die digitale Literaturwissenschaft in Zukunft in hohem Maße auf Erkenntnisse und Methoden des Natural Language Processing angewiesen sein.

Vorerst erscheint die Entwicklung komplexer Annotationsverfahren respektive zugehöriger Abfragen als ein begehbarer Weg, um systematische und nachprüfbar Literaturanalysen anhand großer Textkorpora zu ermöglichen. Diese Analysen können insbesondere bei großräumig angelegten Untersuchungen unter anderem dazu beitragen, basale Einschätzungen der Disziplin zum Beispiel in Bezug auf Formierung und Ausbreitung bestimmter literarischer Genres, Strömungen oder Gruppierungen, oder aber im Hinblick auf die Untersuchung von literarischem und Wissenstransfer zu überprüfen und zu differenzieren. Andererseits zeichnet sich ab, dass diese Verfahren auch im Bereich von Einzeltextanalysen fruchtbar genutzt werden können, wenn es darum geht, die Spezifika eines Werks vergleichend zu evaluieren.

Wenn Textkorpora im World Wide Web gespeichert werden, so fördert dies nicht allein deren allgemeine ubiquitäre Zugänglichkeit. Eine Integration der Texte ins Semantic Web eröffnet darüber hinaus weitere Optionen: sie ermöglicht die Darstellung der

Mehrdimensionalität literarischer Texte durch Datenverknüpfung und fungiert damit sowohl als Speichermedium wie als Medium wissenschaftlicher Erkenntnisgewinnung. Eine solche Textdarstellung wird das Spektrum literaturwissenschaftlicher Erkenntnismöglichkeiten zweifellos erweitern. Nicht zuletzt aber ermöglicht die Verknüpfung von literarischen Texten mit dem Semantic Web auch anderen an der Entwicklung des Menschen und den von ihm geschaffenen Systemen interessierten Disziplinen Zugang zu dem in literarischen Texten enthaltenen Wissen, das von diesen bisher kaum evaluiert werden konnte.⁴⁰

⁴⁰ Dieser Beitrag entstand im Zusammenhang eines Forschungsprojekts zur digitalen Literaturanalyse, das ich gegenwärtig mit einer Gruppe von Studierenden und in Verbindung mit Andrew U. Frank (TU Wien) am ICLTT der ÖAW durchführe.

Bibliographische Angaben

- Alexander Aciman / Emmet Rensin: *Twitterature: The World's Greatest Books Retold Through Twitter*. New York 2009. [[Nachweis im GBV](#)]
- Benedict Anderson: *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London 1983. [[Nachweis im GBV](#)]
- Anne Baillot: *Reconstruire ce qui manque ou le déconstruire ? Approches numériques des sources historiques. Entre sources données et réseaux (Programme MORE-HIST)*. Mar 2015, Toulouse, France 2015. [[online](#)]
- Adrien Barbaresi / Hanno Biber: *Extraction and Visualization of Toponyms in Diachronic Text Corpora Digital Humanities Kraków*, 11–16 July 2016, <http://dh2016.adho.org/abstracts/317> [[online](#)]
- Max Bense: *Theorie der Texte. Eine Einführung in neuere Auffassungen und Methoden*. Köln 1962. [[Nachweis im GBV](#)]
- Max Bense: *Einführung in die informationstheoretische Ästhetik. Grundlegung und Anwendung in der Texttheorie*. Reinbek 1969. [[Nachweis im GBV](#)]
- Tim Berners-Lee / Mark Fischetti: *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its inventor*. San Francisco 1999. [[Nachweis im GBV](#)]
- Tim Berners-Lee / James Hendler / Ora Lassila: *The semantic web*. Scientific America 2001. [[online](#)]
- John Maxwell Coetzee: *The English Fiction of Samuel Beckett. An Essay in Stylistic Analysis*. Austin 1969. [[Nachweis im GBV](#)]
- Olivier Curé / Guillaume Blin: *RDF Database Systems: Triples Storage and SPARQL Query Processing*. Amsterdam 2015. [[Nachweis im GBV](#)]
- Jacques Derrida: *Die Schrift und die Differenz*. [L'écriture et la différence. Paris 1967]. Frankfurt/Main 1976. [[Nachweis im GBV](#)]
- Jacques Derrida: *Grammatologie* [De la grammatologie. Paris 1967]. Frankfurt/Main 1983. [[Nachweis im GBV](#)]
- Thomas Sterns Eliot: *Selected Essays, 1917–1932*. London, New York 1932. [[Nachweis im GBV](#)]
- Peter Engelmann: *Jacques Derridas Randgänge der Philosophie*. Wien 1988. [[Nachweis im GBV](#)]
- Wilhelm Fucks: *Mathematische Analyse von Sprachelementen, Sprachstil und Sprachen*. [Veröffentlichungen der Arbeitsgemeinschaft für Forschung des Landes Nordrhein-Westfalen, H. 34a]. Opladen 1955. [[Nachweis im GBV](#)]
- Hans-Georg Gadamer: *Wahrheit und Methode. Grundzüge einer philosophischen Hermeneutik*. Tübingen 1960. [[Nachweis im GBV](#)]
- Renate Giacomuzzi: *Deutschsprachige Literaturmagazine im Internet. Ein Handbuch*. Innsbruck 2012. [[Nachweis im GBV](#)]
- Don Gifford / Robert J. Seidman: *Ulysses Annotated. Notes for James Joyce's Ulysses*. Los Angeles, London 1989. [[Nachweis im GBV](#)]
- Stephan Günzel: *Raumwissenschaften*. Frankfurt/Main 2009. [[Nachweis im GBV](#)]
- Susan C. Herring / Lois Ann Scheidt / Elijah Wright / Sabrina Bonus: *Weblogs as a bridging genre*. In: *Information, Technology & People* 18 (2005), H. 2, S. 142–171. [[online](#)]
- David Hilbert: *Axiomatisches Denken*. *Mathematische Annalen* Bd. 78, S. 405–415 (1918). Wieder abgedruckt in: David Hilbert: *Gesammelte Abhandlungen. Band III: Analysis; Grundlagen der Mathematik; Physik; Verschiedenes; Lebensgeschichte*. Heidelberg 1970, S. 146–155. [[Nachweis im GBV](#)]
- Pascal Hitzler / Markus Krötzsch / Sebastian Rudolph / York Sure: *Semantic Web. Grundlagen*. Berlin 2008. [[Nachweis im GBV](#)]
- Fotis Jannidis / Gerhard Lauer / Matías Martínez / Simone Winko: *Regeln der Bedeutung. Zur Theorie der Bedeutung literarischer Texte*. Berlin 2003. [[Nachweis im GBV](#)]
- Anthony Kenny: *The Computation of Style*. Oxford, New York 1982. [[Nachweis im GBV](#)]
- Christine Ivanovic / Andrew U. Frank: *Viennavigator: Digitale Formalisierung literarischer Topographien am Beispiel des Gesamtwerks von Ilse Aichinger*. In: Norbert Bachleitner / Christine Ivanovic (Hg.): *Nach Wien! Sehnsucht, Distanzierung, Suche. Literarische Darstellungen Wiens aus komparatistischer Perspektive*. Frankfurt/Main u.a. 2015, S. 289–322. [[Nachweis im GBV](#)]
- Gerhard Lauer: *Die Vermessung der Kultur. Geisteswissenschaften als digital Humanities*. In: *Big Data. Das neue Versprechen der Allwissenheit*. Hg. von Heinrich Geiselberger. Berlin 2013, S. 99–116. [[Nachweis im GBV](#)]
- Gerhard Lauer: *Die Weltbibliothek und ihre Korpora*. In: *Zeitschrift für Bibliothekswesen und Bibliographie* 61 (2014) H. 4–5, S. 251–253. [[online](#)]
- Jürgen Lehmann / Christine Ivanovic: *Kommentar zu Paul Celans »Die Niemandrose«*. Heidelberg 2003. [[Nachweis im GBV](#)]
- Franco Moretti: *Atlas of the European Novel 1800–1900*. New York 1999. [[Nachweis im GBV](#)]

Franco Moretti: *Graphs, Maps, Trees: Abstract Models for a Literary History*. New York 2005. [\[Nachweis im GBV\]](#)

Franco Moretti: *Distant Reading*. New York 2013. [\[Nachweis im GBV\]](#)

Robert Musil: *Klagenfurter Ausgabe. Kommentierte digitale Edition sämtlicher Werke, Briefe und nachgelassener Schriften. Mit Transkriptionen und Faksimiles aller Handschriften*. Klagenfurt 2009. [\[Nachweis im GBV\]](#)

Stephen Ramsay: *Reading Machines. Toward an Algorithmic Criticism*. University of Illinois Press 2011. [\[Nachweis im GBV\]](#)

Patrick Sahle: *Digitale Editionsformen. Teil 3: Textbegriffe und Recodierung. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Schriften des Instituts für Dokumentologie und Editorik*. Norderstedt 2013. [\[online\]](#)

Robert Musil in der *Klagenfurter Ausgabe*: Bedingungen und Möglichkeiten einer digitalen Edition. Hg. von Massimo Salgaro. München 2014. [\[Nachweis im GBV\]](#)

Beat Suter: *Hyperfiktion und interaktive Narration: Im frühen Entwicklungsstadium zu einem Genre*. Zürich 2000. [\[Nachweis im GBV\]](#)

Beat Suter / Michael Böhler: *Hyperfiction. Hyperliterarisches Lesebuch: Internet und Literatur*. Frankfurt/Main 2001. [\[Nachweis im GBV\]](#)

Weiterführende Literatur

Steven Bird / Edward Loper / Ewan Klein: *Natural Language Processing with Python*. Sebastopol 2009. [\[Nachweis im GBV\]](#)

Alexander Clark / Chris Fox / Shalom Lappin: *The Handbook of Computational Linguistics and Natural Language Processing*. Chichester 2010. [\[Nachweis im GBV\]](#)

James H. Coombs / Allen H. Renear / Steven J. DeRose: Markup systems and the future of scholarly text processing. In: *Communications of the ACM* 30 (1987), H. 11, S. 933–947. [\[online\]](#)

David Hoover: The end of the irrelevant text: Electronic texts, linguistics, and literary theory. In: *DHQ* 2 (2007). [\[online\]](#)

Fotis Jannidis / Gerhard Lauer / Andrea Rapp: Alte Romane und neue Bibliotheken. Zum Projekt eines digitalen historischen Referenzkorpus des Deutschen. In: *Die innovative Bibliothek*. Hg. von Erland Kolding Nielsen / Klaus Gerhard Saur / Klaus Ceynowa. München 2005, S. 139–150. [\[Nachweis im GBV\]](#)

Fotis Jannidis / Gerhard Lauer / Andrea Rapp: Hohe Romane und blaue Bibliotheken. Zum Forschungsprogramm einer computergestützten Buch- und Narratologiegeschichte des Romans in Deutschland (1500–1900). In: *Literatur und Literaturwissenschaft auf dem Weg zu den neuen Medien. Eine Standortbestimmung*. Zürich 2007, S. 29–42. [\[Nachweis im GBV\]](#)

Fotis Jannidis: Methoden der computergestützten Textanalyse. In: *Methoden der literatur- und kulturwissenschaftlichen Textanalyse*. Hg. von Vera Nünning. Stuttgart 2010, S. 109–132. [\[Nachweis im GBV\]](#)

Fotis Jannidis / Gerhard Lauer: Burrows's delta and its use in german literary history. [\[Nachweis im GBV\]](#) In: *Distant Readings. Topologies of German Culture in the Long Nineteenth Century*. New York 2014, S. 29–54. [\[Nachweis im GBV\]](#)

Gerhard Lauer: Introduction: Empirical Methods in Literary Studies. In: *Journal of Literary Theory* 9 (2015), H.1, S. 1–3. [\[Nachweis im GBV\]](#)

Christopher D. Manning / Hinrich Schütze: *Foundations of Statistical Natural Language Processing*. Cambridge, MA 2001. [\[Nachweis im GBV\]](#)

Thomas Rommel: *Literary studies*. In: *Companion to Digital Humanities*. Hg. von R. Siemens / J. Unsworth / S. Schreibman. Oxford 2004. [\[Nachweis im GBV\]](#)

Desmond Schmidt: The inadequacy of embedded markup for cultural heritage texts. In: *Literary and Linguistic Computing* 25 (2010), S. 337–356. [\[Nachweis im GBV\]](#)

Jan Van Looy / Jan Baetens: *Close Reading New Media. Analyzing Electronic Literature*. Leuven 2003. [\[Nachweis im GBV\]](#)