

# Transbiblionome Daten in der Literaturwissenschaft. Texttechnologische Erschließung und digitale Visualisierung intertextueller Beziehungen digitaler Korpora

,  
wagner@lit-wiss.uni-siegen.de  
Beijing Institute of Technology, China, Volksrepublik

,  
mehler@em.uni-frankfurt.de  
Universität Frankfurt

,  
Hanno.Biber@oeaw.ac.at  
Österreichische Akademie der Wissenschaften

## Introduction

“digital humanities” boils down to using computers to do exactly the same silo-ed and intellectually buttoned down work that people did before. But it's always easier to get money for equipment (i.e. computers to make a million concordances) than it is to re-envision a field. People in this kind of digital humanities are very concerned with “preservation” in every sense of the word — preservation of the status quo, of themselves and their jobs, and of the methods and fields of the past.

(Lisa Nakamura)

Letztlich reicht es nicht aus, auf Seiten der Objektbasis unablässig neu digitales Material zu akkumulieren. Parallel dazu müsste auf Seiten der Forschung die Bereitschaft zum aktiven Einsatz technologisch und methodisch innovativer Verfahren gefördert werden. Die Digitalisierung allein ohne eine begleitende Theoriedebatte und ohne ein verfeinertes methodisches Rüstzeug betreiben zu wollen dürfte zu verkürzten Ergebnissen führen.

(Embach / Andrea 2008)

Die Sektion und ihr Leitkonzept, d. h. transbiblionome Daten, reagiert auf die zitierten Postulate, indem sie die Konferenz-Leitfrage: „Was sind Daten in den Geisteswissenschaften?“ an einem spezifischen Aspekt

literarischer Texte und ihrer Nutzung aufgreift, nämlich ihrer Intertextualität. Seit den Medienverbünden der literarischen Moderne funktioniert Literatur essentiell als Intertext, als Relais zwischen anderen Texten. Lektüre wird damit zu einer Suche „von Buch zu Buch“ nach „etwas, was zwischen den Inhalten aller einzelnen Bücher schwebt, was diese Inhalte in eins zu verknüpfen vermöchte“ (von Hofmannsthal 2000: 111). Im Unterschied zu *paratextuellen* Daten wie Textvarianten, Konkordanzen, Erläuterungen, Quellen etc. sind *intertextuelle* Daten relational, d. h. sie bezeichnen die (z. B. semantischen) *Beziehungen* eines literarischen Textes (Matrix-Text) zum Archiv aller vorgängigen und zeitgleichen Texte. Sie sind Daten (gegeben) in dem eingeschränkten Sinn, dass sie jeweils aus der Differenz zwischen zwei Texten emergieren. Diese *transbiblionomen*, die Ordnung des Buches und seiner Einheiten überschreitenden Daten bewirken zudem eine laufende Modifizierung der Daten des Matrix-Textes, indem sie dessen (lexikalischen, syntaktischen und semantischen) Einheiten je nach Kontextualisierung unterschiedliche Signifikanz bzw. Verknüpfungsdichte zuweisen. Hieraus ergibt sich für den textbezogenen Sektor der *Digital Humanities* die Herausforderung, das transbiblionome Medium des Computers nicht länger für rein biblionom formulierte Aufgaben zu verwenden, sondern dessen Potential für die Erschließung, Aufbereitung und kollaborative Nutzbarmachung transbiblionomer (intertextueller) Daten zu erproben.

In den Beiträgen dieser Sektion (1) wird eine literatur- und medientheoretische Definition transbiblionomer Daten gegeben, ihre Problematisierungsgeschichte knapp resümiert und ihr Stellenwert für Literaturforschung im digitalen Zeitalter illustriert; (2) ein erfolgreich etabliertes digitales Korpus ( fackel-online des Austrian Academy Corpus) vorgestellt, dessen Textbestand aufgrund seiner programmatischen (zitierenden, anspielenden, parodierenden) Referenz auf ein umfangreiches Korpus anderer Texte (die zeitgenössische Presseberichterstattung) besonders nachdrücklich die Öffnung auf transbiblionome Daten (die extrem große Zahl der zitierten Presstexte) nahelegt; (3) eine Texttechnologie namens *Wikidition* zur automatischen Generierung lexikalischer, syntaktischer, semantischer und textueller Links vorgeführt, die es Literaturforschern erstmals ermöglichen wird, intertextuelle Bezüge zwischen Matrix-Texten und großen Kontext-Korpora (etwa Jahressbänden einer Tageszeitung) umfassend zu erschließen, zu evaluieren, und zu visualisieren. Die Sektion reagiert damit theoretisch und praktisch auf die Leitfragen nach dem theoretischen und technologischen Status von Daten in den *Digital Humanities* sowie nach den Verfahren ihrer Modellierung und Visualisierung.

## Transbiblionome Daten: Problemgeschichtlicher Abriss und aktuelle Herausforderungen

Benno Wagner; Beijing Institute of Technology

Moderne Literatur funktioniert essentiell intertextuell und intermedial. Mit dem Heraufkommen der neuen Konkurrenz-Medien (Photographie und Film, Telegraf und Telefon) und mit der Ablösung der Referenzinstitution Bibliothek durch globale Informationssysteme (Rayward 2008) sowie das alle Lebensbereiche durchdringende Wissen moderner Verwaltungen existiert Schrift nurmehr im Spannungsverhältnis zwischen Buchgebundenheit und einem zunehmend *transbiblionom* organisierten kulturellen Kontext. Der literarische Text gerät auf diese Weise einerseits zum „geometrischen Ort eines hors-texte“, zu einem „Kreuzungspunkt von Schichten, die Myriaden von Horizonten entspringen“ (Topia 1984: 103). Andererseits wird Literatur unter diesen Bedingungen zu einer besonderen Instanz des kulturellen Gedächtnisses. Sie lässt sich als komplexer „Spurenkörper“ (Pêcheux 1983: 55) beschreiben, oder – jedenfalls in ihren raffiniertesten und reflektiertesten Schreibweisen – auch als „hypermnemische Maschine“ (Derrida 1984: 147), als dynamischer Erinnerungsapparat, dessen virtuelles Verweispotential auf andere Texte die Ordnungsraster realer Wissensspeicher (Enzyklopädien, Bibliotheken, Archive) durchkreuzt. Bilden die letzteren Datensätze aus Relationen erster Ordnung, so sind die genuin intertextuellen Daten stets Relationen zweiter Ordnung (Relationen von Relationen). M.a.W.: sie lassen im Paradigma eines Textsyntagmas nicht nur lexikalische Einheiten erscheinen, sondern ganze Textfragmente, also wiederum Syntagmen.

So schrieb Derrida in der *Grammatologie* zunächst:

„Es geht [...] nicht darum, der Buchhülle noch nie dagewesene Schriften einzuverleiben, sondern endlich das zu lesen, was in den vorhandenen Bänden schon immer zwischen den Zeilen geschrieben stand. Mit dem Beginn einer zeilenlosen Schrift wird man auch die vergangene Schrift unter einem veränderten Organisationsprinzip lesen. [...] Was es heute zu denken gilt, kann in Form der Zeile oder des Buches nicht niedergeschrieben werden; ein derartiges Unterfangen käme dem Versuch gleich, die moderne Mathematik mit Hilfe einer Rechenschiebermaschine zu bewältigen.“ Stattdessen avisiert er, diesmal mit Leroi-Gourhan, „eine andere[n], bereits vorstellbare[n] Art der Speicherung [...], deren rasche Verfügbarkeit der des Buches überlegen sein wird: die große ‚Magnetothek‘ mit elektronischer Auswahl wird in naher Zukunft vorselektierte und sofort verfügbare Informationen liefern“ (Derrida 1967: 154-155).

Knapp zwei Jahrzehnte später hingegen, in einem Aufsatz aus dem Jahre 1984, konfrontiert uns Derrida mit einem

ganz anderen und scheinbar diametral entgegengesetzten Szenario. Im Bezug auf den *Ulysses* von Joyce heißt es nun:

„for there be no simple confusion between him [Joyce] and a sadistic demiurge, setting up a hypermnestic machine, there in advance, decades in advance, to compute you, control you, forbid you the slightest inaugural syllable because you can say nothing that is not programmed on this 1000th generation computer [...] beside which the current technology of our computers and micro-computerfied archives and translating machines remains a bricolage of a prehistoric child's toys“ (Derrida 1984: 147).

Hier hat sich offenbar das Komplexitätsgefälle zwischen Druckschrift und elektronischem Speicher verkehrt. Die lineare Schrift ist nicht länger Komplexitäts-Engpass, sondern sie fungiert selbst als Quelle einer überbordenden Komplexität. Sie ist nun der "Computer der 1000. Generation", im Vergleich zu dem die elektronischen Speichermedien als Problem erscheinen, als eine dem Gegenstand der Druckschrift unangemessene, prähistorische Spielerei.

Betrachtet man nun den Einsatz von Computern zu Zwecken der Literaturforschung seit den 1990er Jahren, so drängt sich der Eindruck auf, als habe jedes der beiden Zitate eine Arbeitsperspektive eröffnet, die von der jeweils anderen nichts zu wissen scheint. So haben Autoren wie George Landow und Jay Bolter, in einer eigentümlichen Einebnung des Unterschieds zwischen der syntagmatischen und der paradigmatischen Text-Dimension, den elektronischen Hypertext kurzerhand zu jenem Medium deklariert, mit dessen Hilfe sich das intertextuelle Verweispotential eines literarischen Textes restlos implementieren, der vieldimensionale literarische Text sich aus dem Zwang der Zeile befreien ließe. Hierzu konstatiert Moritz Baßler:

„Landows Parallelisierung von Hypertext mit jenem poststrukturalistischen Textbegriff, den Barthes in S/Z entwickelt, setzt sich über den elementaren Unterschied von syntagmatischer und paradigmatischer Textdimension großzügig hinweg. [...] Dabei handelt es sich jedoch um zwei vollkommen verschiedene Dinge, denn ein Hypertext mag so nonlinear sein wie er will – das betrifft doch immer nur die Sequenz, in seiner paradigmatischen Dimension dagegen unterscheidet er sich nicht vom normalen Text“ (Baßler 2005: 307-308).

Sehr viel erfolgreicher gestaltet sich computergestützte Literaturforschung immer dann, wenn sie die transbiblionome Dimension der Intertextualität von vornherein aus ihrem Gegenstandsbereich ausschließt. Dies geschieht zumeist stillschweigend, bisweilen aber auch mit programmatischem Nachdruck, wenn sich etwa die in Deutschland etablierte Computerphilologie explizit auf die Befassung mit „traditionellen philologischen Gegenständen“ und damit das Potential des Computers auf die Optimierung biblionomer Funktionen sowie auf die Herstellung dezentraler Forschungskollaborationen

beschränkt (Meister 2005: 326). Als ein auf dieser biblionomen Basis erfolgreich etabliertes Projekt stellen wir die digitale Edition der *Fackel* des Austrian Academy Corpus vor. Das digitale *Fackel*-Korpus kann zugleich als Testfall für die hier postulierte transbiblionome Erweiterung dienen, weil sein Textbestand aufgrund der programmatischen (zitierenden, anspielenden, parodierenden) satirischen und polemischen Referenz auf umfangreiche Korpora anderer Texte (die zeitgenössische Presseberichterstattung, die zeitgenössische Literatur) besonders nachdrücklich die Öffnung auf intertextuelle Daten (die extrem große Zahl der zitierten Presstexte) nahelegt.

Eine auf intertextuelle Verlinkung orientierte *transbiblionome Literaturforschung* wird also auf biblionome Digitalisierungen von Datensätzen erster Ordnung funktional aufsetzen. Ihre methodische und technische Entwicklung hätte sich vor dem skizzierten Erfahrungshintergrund an drei Leitlinien zu orientieren:

(1) *Zielsetzung*: Transbiblionome Digital Humanities zielt auf die computergestützte Erschließung und Darstellung der intertextuellen und intermediären Dimension literarischer Texte jenseits einer Beschränkung auf einen biblionomen Forschungshorizont und der Fixierung auf das Visualisierungspotential von Hypertext. Als technische Grundlage hierfür hätte eine auf die Generierung und Verwaltung intertextueller Daten zielende digitale Arbeitsumgebung zu dienen, die zugleich die dezentrale Kollaboration von Experten(gruppen) und eine nutzerspezifisch differenzierte Aufbereitung der Forschungsergebnisse ermöglicht. In dieser Sektion stellen wir mit *Wikidition* eine Software vor, welche die ersten beiden Aspekte dieses Anforderungsprofils texttechnologisch umsetzt.

(2) *Theorie*: Unter den genannten Bedingungen kann und muss sich das zugrundeliegende Intertextualitäts-Modell der methodischen Alternative entziehen, mittels derer die printorientierte Methodendiskussion (insbesondere der 70er und 80er Jahre) sich um eine „Zähmung“ (Lachmann 1984: 137) des transbiblionomen Potentials ‚moderner‘ Intertextualität bemüht hatte: Hier eine ‚geschlossene‘, durch unterstellte Verknüpfungsabsichten des Autors oder faktische Verknüpfungsoperationen des Lesers begrenzte, dort eine unmittelbar auf das System der langue bezogene ‚offene‘ und daher, zumal unter Bedingungen einer printfixierten Forschung, forschungspraktisch niemals einholbare Intertextualität. Baßlers Entwurf eines ‚archivimmanenten Strukturalismus‘ (Baßler 2005), der den intertextuellen Raum (das ‚Paragrammaire‘ nach J. Kristeva) eines Bezugstextes auf eine historische Positivität von Kontext-Dokumenten bezieht, die er ‚Archiv‘ nennt, kann hier als fundierte und konstruktive Alternative dienen, deren Begrifflichkeit sich unmittelbar auf die Konzepte und Leistungen einer digitalen Texttechnologie beziehen lässt. Präzisierungen und Erweiterungen der Theorie werden dort anzustreben sein, wo das ‚Archiv‘ nicht nur den biblionomen Raum, sondern zugleich den der Textualität überschreitet, indem es sich multimedial konstituiert.

(3) *Methode*: Intertextuelle Computerphilologie dieser transbiblionomen Art zielt nicht auf die *Implementierung* literarischer Intertextualität: ihre ‚Befreiung‘ aus dem ‚Gefängnis‘ der Druckzeile und ‚vollständige Entfaltung‘ im barrierefreien digitalen Schreibraum, sondern auf ihre *Supplementierung*: auf die forschungstechnische Unterstützung der selbstverständlich stets selektiven und (projekt- und methodenspezifisch) perspektivischen Erschließung des intertextuellen Potentials eines je gegebenen literarischen Texts. Bei der Entwicklung einer zweckmäßigen Arbeitsumgebung hätte die Kooperation zwischen Philologie, Medienwissenschaft, Texttechnologie und Informatik einer Logik *pragmatischer Schnittstellenbildung* zu folgen. Statt entweder digitale Lösungsmöglichkeiten mit philologischen Problemstellungen zu überfordern, oder umgekehrt von vornherein die philologischen Problemstellungen an das Leistungsvermögen digitaler tools anzupassen, wären für jede Teilaufgabe die Schnittstellen zwischen humaner Intelligenz und künstlicher Intelligenz präzise zu definieren, um die Leistungsvermögen von Menschen und Rechnern möglichst effizient miteinander zu verschalten.

Ausgehend von diesen Überlegungen stellen wir, mit den Bezugshorizonten Franz Kafka und Karl Kraus, eine texttechnologische Anwendung (*Wikidition*) für die Entwicklung eines *Literary Memory Information System* (LiMeS) vor. Dh. einer literaturwissenschaftlichen Forschungsumgebung, die *literarische Texte* nicht einfach als Gegenstände, sondern *als Medien des kulturellen Gedächtnisses* behandelt, indem sie ihr intertextuelles Verweispotential erschließbar, darstellbar und für unterschiedliche Verwertungszusammenhänge nutzbar macht. Die texttechnologischen Entwicklungen der letzten Jahre bieten u. E. eine tragfähige Basis für die Konzeption und Implementierung einer digitalen Arbeitsumgebung für eine solche intertextuell und transbiblionom orientierte Literaturforschung.

## "Die grellsten Erfindungen sind Zitate". Corpusbasierte Erkennung und Analyse von Zitaten und intertextuellen Referenzen in literarischen Texten

Hanno Biber; Österreichische Akademie der Wissenschaften

„Die unwahrscheinlichsten Taten, die hier gemeldet werden, sind wirklich geschehen; ich habe gemalt, was sie nur taten. Die unwahrscheinlichsten Gespräche, die hier geführt werden, sind wörtlich gesprochen worden; die grellsten Erfindungen sind Zitate.“ (Kraus 1926: VII) Im Vorwort der Buchausgabe des aus den die Worte und Taten, die Presseberichte und Propaganda, die Phrasen und Tonfälle des Ersten Weltkrieges aus- und aufrufenden,

satirisch in Szene gesetzten Dokumenten bestehenden monumentalen Dramas "Die letzten Tage der Menschheit" steht diese Aussage von Karl Kraus zu seiner satirischen Methode des Zitats. Sie hebt hervor, wie der Autor das gestaltet hat, was er für seine Leser, die noch hören und lesen konnten, was und wie es gesagt wurde, wie auch für zukünftige Leser hörbar und sichtbar gemacht hat, in den aus Zitaten bestehenden Szenen seines Dramas ebenso wie in den Texten seiner Zeitschrift *Die Fackel*.

Die *AAC-Fackel*, die digitale Ausgabe der von Karl Kraus vom 1. April 1899 bis Februar 1936 in Wien herausgegebenen Zeitschrift *Die Fackel* wurde unter Anwendung computerphilologischer Methoden im Rahmen des *AAC-Austrian Academy Corpus* an der Österreichischen Akademie der Wissenschaften erstellt. Das literarische Werk des wohl bedeutendsten Satirikers deutscher Sprache und darin die 1899 bis 1936 in Wien herausgegebene Zeitschrift *Die Fackel* des 1874 in Jihlava in Böhmen geborenen Karl Kraus, der in Wien gelebt, in vielen Städten Europas seine Texte vorgetragen hat und weit darüber hinaus rezipiert wurde und wird, ist als überaus bedeutender Beitrag der deutschsprachigen Literatur zur Weltliteratur zu betrachten. Die Überlieferung und Verfügbarkeit der satirischen und polemischen Texte im digitalen Medium der *AAC-Fackel*, der Texte der *Fackel* mit ihrer thematischen Vielfalt, sprachlichen Komplexität und historischen Relevanz, muss für eine von computergestützten Verfahren der digitalen Literaturwissenschaft bestimmten wissenschaftlichen Beschäftigung, die sich der zeitgemäßen Erforschung des literarischen Werkes nach neuesten wissenschaftlichen Standards verpflichtet, als eine überaus wichtige Aufgabe betrachtet werden. In der digitalen Edition der *AAC-Fackel* wird der gesamte Text dieser Zeitschrift nicht nur einer an Literatur, Sprache und Geschichte interessierten Leserschaft zugänglich gemacht, sondern auch die einzigartige sprachliche und literarische Qualität dieser Texte unter Nutzung computerphilologischer Methoden und texttechnologischer Instrumente durch verschiedene Suchmöglichkeiten und Register in einer neuen Weise erschlossen.

Die Methode der Satire von Karl Kraus ist das Zitat. Seine Texte folgen mithin einer transbiblionomen Logik des laufenden Verweises auf andere Texte. Aus diesem Grunde kann unter Nutzung texttechnologischer Möglichkeiten, wie sie in der im anschließenden Beitrag beschriebenen *Wikidition* implementiert werden, das digital verfügbare Textcorpus der *Fackel* zur Analyse der in ihm zitierten Texte genutzt werden. Karl Kraus findet zumeist in den zeitgenössischen Texten der Presse und der Publizistik die Prätexte für seine Texte, indem er das, was er verarbeitet, einfach zitiert, oder einschöpft in seine Texte und satirisch kommentiert. In seinen satirischen Texten muss er nichts mehr erfinden, sondern vielfach einfach nur zitieren, in Anführungszeichen setzen, den Text einbetten, ihn graphisch so anordnen, dass seine satirische Qualität sichtbar werden kann, dem Leser sichtbar gemacht werden kann. Die in den tausenden Zitaten

wiedergegebenen, satirisch bearbeiteten Sprachstücke der *Fackel* sind vom Autor montiert und typographisch aufbereitet, was in vielen Fällen, besonders in den Glossen der 'Fackel' durch einfache Spationierung von bestimmten Textpassagen erfolgen kann. Die Textsorte der Glosse ist beispielhaft für das satirische Verfahren von Karl Kraus, wo es auf die besondere Behandlung eines zitierten, aus der Presse stammenden Prätextes ankommt. Kurt Krolop schreibt dazu in einem Aufsatz über Karel Apek und Karl Kraus über die wichtigste Funktionsweise der Satire von Karl Kraus: „Das auf solche Weise schöpferische Vernichtungsarbeit leistende Zitat, in dem die Spationierung den Gestus sichtbar macht, reduziert wolkigen Bombast auf den zugrunde liegenden reinen Unsinn, vorgetäuschte Fülle auf die tatsächliche Leere, falsches Pathos auf das echte Blech, die hohle Phrase auf die dicke Lüge. Der gestischen Entlarvung dient auch die für die Fackel so charakteristische Verschränkung von klassischem Sprachgut und Jargon, welche etwa die wahre Haltung eines liberalen Journalisten zur Anschauung bringen soll, wenn dieser Posas berühmte Aufforderung zitiert und das so bewirkt „Sire, bittsie, was liegt Ihnen schon dran, geben Sie Gedankenfreiheit!“ (890,276)“ (Krolop 1992: 305-306.) Der Text „Blendwerk der Hölle“ (F 366,30), in dem der Satiriker auch am Ende seines Textes über die besondere Funktionsweise seiner Satire reflektiert, ist beispielhaft für dieses dann in der satirischen Tragödie "Die letzten Tage der Menschheit" dramatisch gestalteten Verfahrens. Diese Glosse ist wie viele andere auch Teil einer Reihe von oft thematisch in Zusammenhang stehenden Glossen in den Heften der Zeitschrift, einer für diese besondere Form der Satire von Karl Kraus typischen Textsorte und thematisiert in besonderer Weise die Arbeit des Satirikers mit der „überevollen Realität“ (F 366,32), wie sie sich ihm sprachlich deutlich und in seinen eigenen Texten durch Zitierung in seiner Zeitschrift deutlich gemacht in der Presse seiner Zeit darbietet. Karl Kraus demonstriert und reflektiert seine Methode der Satire, die nicht Satiren verfassen muss, sondern Texte aus der Zeitung bloß noch anordnen muss, damit sie zu Satiren werden. Diese besondere Methode der Satire ist die Methode des Zitats. Der Schriftgrößenwechsel zur Kenntlichmachung des Zitats und die Hervorhebung einzelner sprachlicher Äußerungen durch Sperrung sind der zeitgenössischen typographischen Konvention entlehnte Zeichen, die hier zum kritischen Zweck der Hervorhebung im satirischen Text eingesetzt werden.

Im Zentrum des satirischen Verfahrens steht das Zitat. Die zitierten Prätexte der satirischen Texte, die in den Glossen, im Drama und in den anderen Texten von Karl Kraus aufgerufen und zitiert werden, können mit Hilfe von texttechnologischen und corpusbasierten Methoden der digitalen Literaturwissenschaft ebenso wie andere intertextuelle Verbindungen zwischen den Texten aufgefunden und systematisch erforscht werden. Wie erste Versuche eines solchen Verfahrens aussehen könnten, wäre ein wichtiger Beitrag zur Erforschung der komplexen

Textreferenzen der *Fackel*. Als ein wichtiges Instrument dazu kann neben anderen zur Verfügung stehenden Ressourcen die in der Österreichischen Akademie der Wissenschaften erstellte Textsammlung des AAC-Austrian Academy Corpus dienen. Das AAC ist ein umfangreiches und komplex strukturiertes Textcorpus zur deutschen Sprache und Literatur im Untersuchungszeitraum von 1848 bis 1989 und Basis für die Entwicklung texttechnologischer Anwendungen im Bereich der Corpusforschung sowie philologischer und textwissenschaftliche Forschungen, die am Institut für Corpuslinguistik und Texttechnologie der Österreichischen Akademie der Wissenschaften mit diesen Textressourcen neben anderen Textcorpora in den Bereichen der Corpusforschung im Kontext digitaler Literaturwissenschaft und philologischer Grundlagenforschung durchgeführt werden. Die AAC-*Fackel* kann neben der digitalen Edition der literarischen Zeitschrift *Der Brenner* in diesem Kontext als exemplarischer Anwendungsfall einer aus der Corpusforschung und computergestützten Textwissenschaft entstandenen digitalen Musteredition eines literaturgeschichtlich bedeutenden Textes betrachtet werden. Ihr Zustandekommen und die dafür notwendigen Bedingungen resultieren aus einer sich mit Sprache und Fragen des besonderen Sprachgebrauchs auf empirischer Textbasis in einem bestimmten, mit besonderen Eigenschaften ausgestatteten Textcorpus befassenden Forschungsrichtung, wie sie im ICLTT und seinen Vorgängerunternehmungen seit vielen Jahren betrieben wird. Die computergestützte Auffindung von Zitaten in Texten und der Erforschung des Gebrauchs von Zitaten in Texten, die der Satiriker Karl Kraus in seinen Texten literarisch nutzt, wird durch große Textcorpora erst systematisch möglich und soll als exemplarischer Anwendungsfall einer derart texttechnologisch ausgeübten Literaturwissenschaft verstanden und gezeigt werden.

## Automatische intra- und intertextuelle Vernetzung literarischer Texte mit Hilfe von Wikidition

Alexander Mehler; Universität Frankfurt

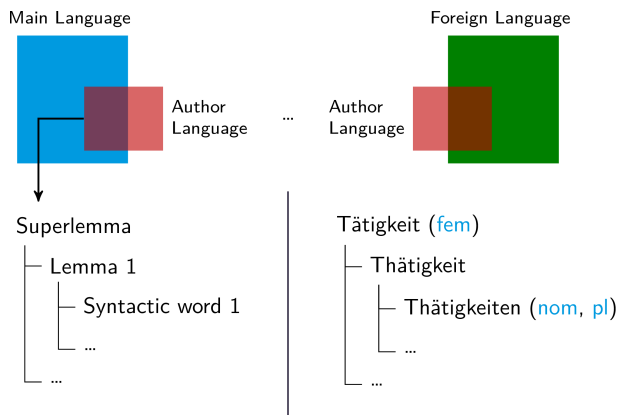
Der Beitrag führt Wikidition als neuartiges *Literary Memory Information System* ein, das die in den beiden vorigen Beiträgen theoretisch und im Hinblick auf ein existierenden digitales Korpus formulierten transbiblionomen Postulate einlösen soll. Wikidition zielt auf die automatische Generierung von online Editionen von Korpora natürlichsprachlicher Texte. Dazu zählen insbesondere literarische Texte, welche als Matrixtexte im Kontext ihrer Echotexte untersucht werden (siehe Abbildung 1). Wikidition kombiniert eine Vielzahl von Text-Mining-Verfahren für die automatische Vernetzung ( *Linkification*) von Wort-,

Satz- und Texteinheiten unter den Bedingungen ihrer syntagmatischen *Verwendung* einerseits und ihres paradigmatischen *Gebrauchs* andererseits. Darüber hinaus beinhaltet Wikidition eine Lexikonisierungskomponente zur Ermittlung jenes Teillexikons einer Sprache, welches der Erzeugung der jeweiligen Inputtexte zugrunde liegt. Auf diese Weise lassen sich die Texte des Eingabekorpus nicht nur mehrschichtig (auf Wort-, Satz- oder Textebene) hypertextuell traversieren. Vielmehr können die Leser einer Wikidition das jeweilige Autorenvokabular bis hinab zur Ebene einzelner syntaktischer Wörter analysieren (siehe Abbildung 2). Während die Linkifizierungskomponente an Prinzipien von Wikipedia bzw. WikiSource anknüpft, ist es das Wiktionary-Projekt, an dem sich die Lexikonisierungskomponente orientiert. Wikidition verwendet zahlreiche computerlinguistische Methoden zur Gewährleistung von Interoperabilität, Offenheit und Erweiterbarkeit der resultierenden Editionen nach dem Wiki-Prinzip. Auf diese Weise wird die Dissemination computerbasierter Methoden über die *Digital Humanities* hinaus unterstützt, indem es Geisteswissenschaftlern ermöglicht wird, eigene explorative Analysen durchzuführen, und zwar ohne Implementierungsaufwand.

	{Echo Text;  i = 1}	{Echo Text;  i > 1}	{Echo Text;  i >> 1}
{Matrix Text;  i = 1}	1	2	3
{Matrix Text;  i > 1}	4	5	6
{Matrix Text;  i >> 1}	7	8	9

**Abb. 1:** Transbiblionome Szenarien, auf welche Wikidition zielt. Beispiele: (1) Kafkas Bericht für eine Akademie (als Matrixtext) versus Hauffs Der Affe als Mensch (als Echotext); (2) Kafkas Bericht für eine Akademie versus alle „Affentexte“ (R. Borgards) seit Ende des 18. Jahrhunderts (Hauff, Hoffmann, Flaubert, ...); (3) Kafkas Beim Bau der Chinesischen Mauer versus Prager Tagblatt vom 08.1914–03.1917; (4) Kafkas Werk versus ein Einzeltext aus dem Werk Nietzsches; (5) eine Werkauswahl Kafkas versus eine Werkauswahl Nietzsches; (6) Kafkas Werk versus ein Zeitungskorpus (z. B. basierend auf dem Prager Tagblatt); (7) sämtliche Werke einer Reihe von Schriftstellern versus ein Einzeltext (z. B. Goethes Faust); (8) sämtliche Werke einer Reihe von Schriftstellern versus ein Korpus von Faust-Texten; (9) sämtliche Werke einer Reihe deutschsprachiger Schriftsteller versus sämtliche Werke einer Reihe englischsprachiger Schriftsteller. Anstelle von Matrix- und Echotexten (Wagner 2015) kann alternativ von Hyper- und Hypotexten (Genette 1997) gesprochen werden. In ersterem Fall liegt eine rezeptionsorientierte Sichtweise zugrunde.



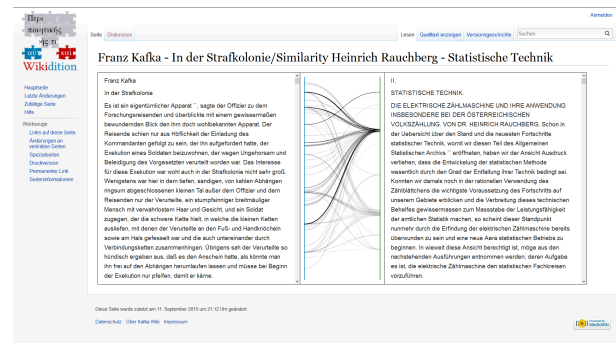


**Abb. 2:** Linke Bildseite: schematischer Ausschnitt (rot) des autorenpezifischen Vokabulars im Schnittmengenbereich der zugrundeliegenden Matrixsprache (blau). Beispielhafte Fragestellung: Welches Deutsch verwendet Franz Kafka in seinem Werk und worin weicht sein Deutsch von der jeweiligen Referenzsprache ab? Da Autoren mehrsprachige Texte verfassen können, ist diese Betrachtungsweise auf mehrere Bezugssprachen auszudehnen (rechte Bildseite). Hierzu unterscheidet Wikidition drei Ebenen der lexikalischen Resolution: die Ebene der Superlemmata, der Lemmata und der syntaktischen Wörter (Wortformen plus grammatische Merkmale) (untere Bildhälfte).

Die Basis für die Vernetzung von Texteinheiten bildet das Konzept der Intertextualität einerseits bzw. der intratextuellen Kohärenz andererseits. Hierzu werden je Sprachebene (Wort, Satz, Paragraph, Text) zwei Arten von Relationen unterschieden: syntagmatische Relationen, welche auf Kontiguitätsrelationen von Texteinheiten beruhen, sowie paradigmatische Relationen, welche auf Relationen der Similarität bzw. Substituierbarkeit von Texteinheiten basieren. Dieses Modell orientiert sich an Hjelmslevs (1969) Konzept der sprachlichen Relation, um eine linguistische Basis für die automatisch zu ermittelnden Vernetzungsrelationen zu gewinnen. So werden beispielsweise alle Sätze des Inputkorpus danach paarweise analysiert ob sie (1) syntaktisch ähnlich konstruiert sind (syntagmatische Ähnlichkeit) oder (2) lexikalisch ähnlich aufgebaut sind (paradigmatische Ähnlichkeit). Ausgehend von einem zweidimensionalen Koordinatensystem, in dem syntagmatische und paradigmatische Ähnlichkeit die Dimensionen aufspannen, kann auf dieser Grundlage für alle texttechnologischen Satzähnlichkeitsmaße theoretisch bestimmt werden, wo sie innerhalb dieses Bezugssystems angesiedelt sind. Mit Hilfe von Wikidition soll auf diese Weise zugleich eine theoretische und praktische Basis der Text- bzw. Satzähnlichkeitsmessung gelegt werden, die es Usern erlaubt, den Ergebnissen der von Wikidition vorgelegten Ähnlichkeitsmessungen explorativ nachzuspüren. Gleichzeitig wird so eine Basis für die Analyse von Satzähnlichkeitsphänomenen im Hinblick auf

Zitate gelegt, wie sie insbesondere für das Werk von Karl Kraus maßgeblich sind.

Analog zu diesem Ansatz wird auf Wortebene zwischen der (syntagmatischen) Tendenz von Wörtern zur Kookkurrenz einerseits und der (paradigmatischen) Tendenz zum Vorkommen in ähnlichen Kontexten andererseits unterschieden. Mittels texttechnologischer Erweiterungen der MediaWiki-Software werden zudem sämtliche dieser Verweisbeziehungen interaktiv visualisiert, so dass Nutzer beispielsweise in Echotexten jenen Satzkandidaten nachspüren können, welche dem Autor des jeweiligen Matrixtexts als Vorlage gedient haben könnten (siehe Abbildung 3).



**Abb. 3:** Franz Kafkas *In der Strafkolonie* (als Matrixtext) im Verhältnis zu Heinrich Rauchbergs *Statistische Technik* (als Echotext). In dem bipartiten Graphen (Bildmitte) verbinden Kanten solche Sätze, die als Text-Reuse-Kandidaten ermittelt wurden. Nutzer können per Fish-Eye-View den bipartiten Graphen traversieren und Verweisbeziehungen einzelner Sätze, welche in dem links- bzw. rechtsseitig dargestellten Text entsprechend hervorgehoben werden, analysieren.

Der Beitrag exemplifiziert Wikidition anhand ausgewählter Texte von Franz Kafka einerseits und von Karl Kraus andererseits. Er ergänzt die beiden literaturwissenschaftlichen Vorträge der Sektion, indem er das texttechnologische Instrumentarium einer transbiblionomen Literaturwissenschaft entwirft und experimentell erprobt. Insbesondere die Verweisstruktur der resultierenden Wikiditionen wird auf der Grundlage einer größeren Zahl von Netzwerkcharakteristika quantitativ evaluiert. Diese Evaluation zeigt, dass Wikiditionen jeweils auf der Basis einer sehr kleinen Zahl von Verweisen, die theoretisch möglich sind, hochgradig clusternde Graphen erzeugen, die es ihren Lesern erlauben, dasselbe Textkorpus in jedwede Richtung und ausgehend von jedwedem (Wort-, Satz- oder Text-)Kontext zu traversieren. Auf diese Weise wird eine Brücke zwischen Literaturwissenschaft und Informatik gespannt, und zwar am Beispiel der Analyse und Synthese intertextueller Beziehungen in digitalen Korpora literarischer Texte.

## Fußnoten

1. Kurznotation für Die Fackel Nr. 366, S. 30
2. Hier sprechen wir auch von syntagmatischer Kontiguität der betroffenen Sätze in dem Sinne, dass ihre Konstituenten füreinander austauschbar sind, ohne paradigmatisch oder inhaltlich verwandt sein zu müssen
3. Hier sprechen wir von paradigmatischer Similarität der betroffenen Sätze in dem Sinne, dass ihre Konstituenten füreinander austauschbar sind, ohne dass die Sätze syntaktisch ähnlich zu sein brauchen. Im Extremfall zweier identischer Sätze gehen syntagmatische und paradigmatische Ähnlichkeit zusammen.

## Bibliographie

**AAC – Austrian Academy Corpus:** *AAC-Fackel*. Online Version: 'Die Fackel. Herausgeber: Karl Kraus, Wien 1899–1936.' AAC Digital Edition No. 1. <http://www.aac.ac.at/fackel> [letzter Zugriff 15. Oktober 2015].

**AAC-Austrian Academy Corpus und Brenner-Archiv:** *Brenner Online*. Online Version: 'Der Brenner. Herausgeber: Ludwig Ficker, Innsbruck 1910–1954.' AAC Digital Edition No.2. <http://www.aac.ac.at/brenner> [letzter Zugriff 15. Oktober 2015].

**Baßler, Moritz** (2005): *Die kulturpoetische Funktion und das Archiv*. Eine literaturwissenschaftliche Text-Kontext-Theorie. Tübingen: Francke.

**Borgards, Roland** (2012): „Der Affe als Mensch und der Europäer als Ureinwohner“, in: Wellbery, David E. (ed.): *Kultur-Schreiben als romantisches Projekt*. Romantische Ethnographie im Spannungsfeld zwischen Imagination und Wissenschaft. Tübingen: Königshausen & Neumann 17-42.

**Derrida, Jacques** (1967): *Grammatologie*. Frankfurt a.M.: Suhrkamp.

**Derrida, Jacques** (1984): "Two words for Joyce", in: Attridge, Derek / Ferrer, Daniel (eds.): *Poststructuralist Joyce*. Essays from the French. Cambridge: CUP 145-159.

**Embach, Michael / Rapp, Andrea** (05.08.2008): „Die intelligentere Expansion der Gutenberg-Galaxis“, in: *Frankfurter Allgemeine Zeitung*.

**Genette, Gérard** (1993): *Palimpseste*. Die Literatur auf zweiter Stufe. Frankfurt am Main: Suhrkamp.

**Hjelmslev, Louis** (1969): *Prolegomena to a Theory of Language*. Madison: University of Wisconsin Press.

**Kraus, Karl** (1926): *Die letzten Tage der Menschheit*. Wien: Verlag Die Fackel.

**Krolop, Kurt** (1992): "Solche Erfolge erreichen nur deutsche Molche", in: Krolop, Kurt: *Sprachsatire als Zeitsatire bei Karl Kraus*. Neun Studien. Berlin: Akademieverlag 305-306.

**Meister, Jan Christoph** (2005): „Projekt Computerphilologie. Über Geschichte, Verfahren und Theorie rechnergestützter Literaturwissenschaft“, in: Segeberg, Harro / Winko, Simone (eds.): *Literarität und*

*Digitalität*. Zur Zukunft der Literatur. München: Fink 315-341.

**Pêcheux, Michel** (1983): „Über die Rolle des Gedächtnisses als interdiskursives Material“, in: Geier, Manfred / Woetzel, Harold (eds.): *Das Subjekt des Diskurses*. Beiträge zur sprachlichen Bildung von Subjektivität. Berlin: Argument-Verlag 50-58.

**Rayward, Warden Boyd** (2008): *European modernism and the information society*. Informing the present, understanding the past. Aldershot: Ashgate.

**Topia, André** (1984): "The Matrix and the Echo: Intertextuality in Ulysses", in: Attridge, Derek / Ferrer, Daniel (eds.): *Post-Structuralist Joyce*. Essays From the French. Cambridge: Cambridge UP 103-125.

**von Hofmannsthal, Hugo** (2000): *Der Brief des Lord Chandos*. Schriften zur Literatur, Kultur und Geschichte, hg. v. Mathias Mayer. Stuttgart: Reclam.