

Es ist an der Zeit, dass auch in der Zeitschrift «Italienisch» die Diskussion über Nutzen und Grenzen der digitalen Medien und des Internet geführt wird. Längst werden Studiengänge in der Disziplin der «Digital Humanities» angeboten und digitale Medien haben in Forschung und Lehre an den Universitäten auch in den Geisteswissenschaften Einzug gehalten. In diesem Kontext versteht sich der folgende Beitrag als Anregung an Vertreterinnen und Vertreter des Fachs an Schulen und Hochschulen, weitere Aufsätze zu dem Thema vorzulegen und die hier vorgebrachten Thesen zu vertiefen und zu diskutieren. (Herausgeber und Redaktion)

THOMAS KREFELD / STEPHAN LÜCKE / ISABEL VON EHRlich

Digitalianistica. Die italienische Philologie unterwegs in die *digital humanities*

Anna log (immer). Die Gitta L lügt weniger.

Emergenz eines Paradigmas

Die Schlagworte, mit denen die großen wissenschaftsgeschichtlichen Strömungen belegt werden, sind in der Regel ambivalent; sie liefern zwar bequeme Label, bezahlen ihre schnelle und flächendeckende Verbreitung jedoch mit erheblichen inhaltlichen Unschärfen. Der heuristisch zu verstehende Ausdruck *digital humanities* (= DH) ist ebenso wie seine synonymen Konkurrenten *humanities computing*, *eHumanities* (z.B. <http://www.cceh.uni-koeln.de/>), gegen die er sich zunehmend durchsetzt, in dieser Hinsicht paradigmatisch: Er bezeichnet einerseits alle Optionen, die sich für die Geistes- und Sozialwissenschaften aus den so genannten Neuen Medien¹ ergeben, also neue Formen der sprachlichen Kommunikation, neue Formen der Textgenese und Rezeption, neue Wege der Datenerhebung und -analyse, neue Formen der Forschungskooperation und andererseits die daraus entstehende Notwendigkeit neuer theoretischer Modellierungen. Der Ausdruck ist übrigens erst aufgekommen, nachdem sich ganze Forschungsrichtungen, wie zum Beispiel die Korpuslinguistik, längst konstituiert hatten. Die ganz selbstverständliche Adaption und Integration digitaler Techniken hatte eine «Ausdifferenzierung» (Jannidis 1999) der etablierten Fächer zur Folge, etwa in Gestalt der Computerphilologie und Computerlinguistik, ohne dass während dieses progressiven Übergangs zunächst der Eindruck eines wirklichen Paradigmenwechsels entstanden wäre. Die Begriffsgeschichte von DH und damit einhergehend ein nicht zuletzt auch terminologisch greifbarer Paradigmenwechsel hingegen ist relativ jung: 2001 brachte John Unsworth, der Gründungsdirektor des Institute for Advan-

ced Technology in the Humanities der Universität Virginia die Termini erstmals in dieser Verbindung als Titelvorschlag für den inzwischen kanonischen Blackwell *Companion to Digital Humanities* auf (Schreibman/Siemens/Unsworth 2004).² Spätestens seit 2006, als sich eine jährliche Konferenzserie, die ebenfalls bereits seit den 80er Jahren Wissenschaftler aus IT- und Geisteswissenschaften zusammenbrachte, schlicht in «Digital Humanities» umbenannte, wurde der Ausdruck zum Standard. «Digital Humanities», zuletzt in Lausanne (DH 2014), ist inzwischen der weltweit größte und wichtigste internationale Kongress in diesem Bereich (vgl. zur Geschichte der DH Kirschenbaum 2012 und Hockey 2004). Auch in Italien haben sich inzwischen zahlreiche universitäre Institute, Exzellenzinitiativen («centri d'ecellenza») und Vereinigungen zur Koordination der neuen Initiativen gebildet: die älteren von ihnen noch unter Bezeichnungen wie *linguistica computazionale* oder *Informatica umanistica*; die umgekehrte Gewichtung dieses zuletzt genannten Ausdrucks verbirgt übrigens die eigentliche Herausforderung der *humanities* durch Zuweisung an die Informatik (vgl. z.B. in Orlandi 2007 und 2012). In italianistischen Publikationen der letzten Jahre hingegen hat sich ebenfalls meist die Bezeichnung DH durchgesetzt (Vgl. Fiormonte 2012).³

In der sich nunmehr abzeichnenden Wahrnehmung eines Paradigmenwechsels spiegelt sich deutlich die totale Durchsetzung der sozialen Medien und des so genannten Web 2.0 in allen Sphären unserer Lebenswelt seit ca. 2003 wider; markant sind die Starts von Myspace 2003, Facebook 2004, Twitter 2006, Googlemaps 2005 u.a. (vgl. Jones 2014, S. 4). Auch die Institutionalisierung der DH ist mittlerweile unübersehbar. Ganz abgesehen von den nationalen und internationalen Fachverbänden (<<http://www.dig-hum.de/ueber-dhd>> und <<http://www.umanisticadigitale.it/>>) wurden in Deutschland ebenso wie in Italien mehrere einschlägige Lehrstühle und Studiengänge eingerichtet (http://dig-hum.de/sites/dig-hum.de/files/cceh_broschuereweb.pdf). Ganz sicher handelt es sich aber bei den DH nicht um ein «interdisziplinär ausgerichtetes Fach», wie es im Eintrag der deutschsprachigen Wikipedia heißt (http://de.wikipedia.org/wiki/Digital_Humanities). Zwar ist die interdisziplinäre Fundierung im Blick auf die Verschränkung einer spezifischen Disziplin mit der Informationstechnologie konstitutiv, aber von einem «Fach» bzw. einer «Disziplin» im eigentlichen Sinn kann man wegen der Unvergleichbarkeit der Erkenntnisinteressen nicht sprechen. Vielmehr handelt es sich um eine Methodologie bzw. einen methodologischen Rahmen (ein «framework»), der unterschiedlichen Disziplinen gemein ist. Mithin schafft die Anwendung mehr oder weniger identischer Technologien und Tools durchaus vielfältige Parallelen und Verknüpfungen zwischen diesen heterogenen Disziplinen, wie es sie im Übrigen auch in früheren Paradigmen, etwa in der Hochzeit des Strukturalismus oder der Semiotik, schon gab. Aus den neuen (digitalen) Verknüpfun-

gen ergeben sich nun naturgemäß neue Perspektiven und die Möglichkeit neuer interdisziplinärer Erkenntnis. Es ist zudem keineswegs nebensächlich, dass auch die IT-Wissenschaften (abgesehen davon, dass sich neue Betätigungsfelder ergeben) profitieren, sich verändern und sich in neue Richtungen weiterentwickeln.⁴

Mit der sich etablierenden Bezeichnung ›Digital Humanities‹ geht offensichtlich auch eine nachhaltige (Selbst-)Legitimierung digital gestützter oder gar komplett digital und somit zumindest methodologisch quantitativ⁵ konzipierter philologischer Forschung einher.⁶ Dies ist symptomatisch für die endgültige Überwindung des durch die Moderne geprägten, und vielerorts noch immer geniezentrierten Selbstverständnisses der Geisteswissenschaften.⁷ Die gegenwärtigen Ansätze sind oft von dem Versuch gekennzeichnet, einem neuen ›weiteren‹ Verständnis der Gegenstände ›Literatur‹ und ›Sprache‹ gerecht zu werden.⁸ Man kann daher von einem neuen Selbstverständnis oder gar einem Aufbruch zu einer neuen epistemologischen Verfassung (etwa einer Ära ›Post-Gutenberg‹) der Disziplinen sprechen, in der die Möglichkeiten der Forschung zwar nicht komplett erneuert, aber auf jeden Fall integriert, entgrenzt und nicht zuletzt neu gedacht werden (müssen).⁹ Eine solche fachgeschichtliche Wende lässt sich aber naturgemäß in aller Klarheit erst retrospektiv beschreiben.¹⁰

Die Disposition der Fächer den skizzierten Paradigmenwechsel zu vollziehen ist offenkundig (noch) unterschiedlich stark entwickelt. Auch wenn es kaum möglich ist, sich einen schnellen Überblick zu verschaffen, scheinen die textorientierten Disziplinen bisher besonders empfänglich zu sein, und zwar vor allem dann, wenn es gilt, Audio- oder Bildmaterialien einzubinden und somit Hypertexte zu generieren. Ein unangefochten akzeptierter Einsatz digitaler Techniken, der hier schon fast keiner Erwähnung mehr bedarf, findet sich außerdem in der Editionsphilologie mit unzähligen Projekten digitaler Auszeichnung (meist mit TEI – Text Encoding Initiative) der Werke kanonischer Autoren.¹¹ Im Bereich der Sprachwissenschaften im Allgemeinen, und der Erforschung gesprochener Sprache im Besonderen (sei sie dialekt- oder standardnah) darf man jedoch sagen, dass die DH gewissermaßen zur ›Entphilologisierung‹ dieses Fachs beitragen. Einen punktuellen, aber vielleicht nicht ganz zufälligen Eindruck gibt das LMU Center for Digital Humanities (IT-Gruppe Geisteswissenschaften; <http://www.itg.uni-muenchen.de/index.html>); dort wurden oder werden in Zusammenarbeit mit den Fächern Ägyptologie, Anglistik, Finno-Ugristik, Hebraistik, Kunstgeschichte, Musikwissenschaft, Neuere und Neueste Geschichte, Romanische Philologie, Slavistik, Theaterwissenschaft und Europäische Volkskunde 23 Projekte entwickelt, von denen immerhin 18 um sprachliche Daten zentriert sind (darunter sind übrigens allein vier romanistisch-italianistische Unternehmungen). Auffällig ist im

Umfeld der beteiligten Fakultäten die schwache Präsenz der Geschichtswissenschaften, unter denen z.B. die Archäologie vollkommen fehlt.

Es bietet sich vor dem skizzierten Horizont ein prototypisches Verständnis der DH an, die idealerweise den folgenden, allgemeinen Ansprüchen genügen sollten.

(1) Die empirische Grundlage der Forschung besteht in Daten (vgl. Schöch 2013), d.h. in digital kodierten und strukturierten oder mindestens strukturierbaren Einheiten, unabhängig davon, ob es sich um originär erhobene, sekundär digitalisierte Daten oder um eine Kombination beider Typen handelt. Dabei werden in Abhängigkeit vom Forschungsgegenstand möglichst umfangreiche Datenbestände angestrebt.

Die Methode ist also immer quantitativ und weitgehend induktiv. Eine ausschließlich deduktive Arbeit ist in der Perspektive der DH nicht sinnvoll.

(2) Die Forschungskommunikation erfolgt unter den medialen Bedingungen des Internets. Dadurch ergibt sich einerseits ganz selbstverständlich die Möglichkeit, bei der Publikation unterschiedliche Medien (Schrift, Ton, Bild, Film, u.a.) oder unterschiedliche Gattungen (Text, Kommentar, kritischer Apparat, Wörterbuch, Grammatik) hypertextuell zu verflechten; andererseits können die als Forscher und/oder Probanden beteiligten Personen kontinuierlich miteinander kommunizieren und kooperieren.

(3) Isolierte Individualforschung ist nicht mehr sinnvoll. Um den notwendigen Schritt von «big data» zu «smart big data» oder umgekehrt, von «smart data» zu «smart big data» zu bewältigen, ist *crowdsourcing* ein unbedingt notwendiger Vorgang.¹² Anne Burdick ersehnt – vielleicht etwas utopisch – einen «state of ubiquitous scholarship», of evermore interconnected, publicly engaged, participant citizens» (Burdick 2012, S. 30). Jedenfalls kann man davon ausgehen, dass das «read-only»-Ethos der alten Geisteswissenschaften von einem «read/write/rewrite»-Ethos abgelöst werden wird. (Vgl. Porsdam 2013).

Vor allem aber ist das Spezialwissen von IT-Wissenschaftlern effizient mit dem Wissen des Geisteswissenschaftlers zu fusionieren. Das ist bislang nur in sehr wenigen Fällen in ein und derselben Person gegeben; überdies ist gerade das Ideal des autonomen, so-

zusagen enzyklopädischen, wenn nicht genialischen Individualforschers mittlerweile verblasst.

(4) Das Ziel der Forschung besteht vorrangig nicht mehr darin, das resultierende Wissen in abgeschlossenen medialen Gegenständen zu materialisieren und dauerhaft niederzulegen; vielmehr entstehen Forschungsplattformen, die grundsätzlich und auch in längerer Perspektive modifiziert und erweitert werden können, obwohl die Garantie einer dauerhaften Verfügbarkeit technisch noch schwer umzusetzen ist (vgl. hierzu die wichtige CLARIN-D-Forschungsinfrastruktur <<http://de.carin.eu/de/home.html>>). DH-Projekte lassen sich leicht so einrichten, dass sekundär Publikationen auf dinglichen Trägermedien erzeugt werden, wie etwa die Druckoption des in vielerlei Hinsicht exemplarischen *Tesoro della Lingua Italiana delle Origini* (<http://tlio.ovi.cnr.it/TLIO/>) zeigt. Aber eine ausschließlich auf Produkte wie Bücher, CDs oder DVDs zielende Forschung ist nicht mehr sinnvoll.

Definitiv grundlegend ist das aus fachwissenschaftlicher Sicht zunächst unscheinbare Kriterium(1), denn es liefert die Voraussetzungen für (2), (3) und (4). Es ist daher notwendig, auf diese unabdingbare Basis ausführlicher einzugehen.

Das Digitale in den Digital Humanities

Das «Digital» im Kontext der Humanities weist auf einen Umstand, der möglicherweise nicht allen, die sich der einschlägigen Ressourcen, Konzepte und Methoden bedienen, in vollem Umfang und letzter Konsequenz bewusst ist: Sämtliche Daten liegen entweder bereits in Gestalt von binären kodierten Daten vor oder müssen vor dem Einsatz entsprechender Methoden in diese Gestalt überführt werden (sog. «Digitalisierung»). De facto stellt sich das Datenmaterial als eine endliche Reihe bestehend aus den Ziffern 0 und 1 dar. Computer können – und daran wird sich auch in der Zukunft niemals etwas ändern – Daten ausschließlich in diesem Format verarbeiten. Dieses zunächst banal anmutende Faktum bedingt eine ganze Reihe von Konsequenzen, deren Kenntnis grundlegend für den souveränen und fruchtbringenden Umgang mit digitalen Daten ist. Ganz wesentlich ist dabei, dass Daten im Zuge einer digitalen Abbildung grundsätzlich in der einen oder anderen Weise «kodiert» werden müssen (vgl. Renear 2004). Dies beginnt bereits auf der Ebene der elementarsten Bestandteile eines Textes: den einzelnen Buchstaben. Durch die Beschränkung auf das duale Zahlensystem müssen sämtlichen Buchstaben –

und natürlich auch den übrigen Zeichen, aus denen Texte üblicherweise bestehen, wie Zahlen, Satzzeichen und sogar dem Spatium – Zahlenwerte zugewiesen werden.

Fundamental für das Funktionieren eines solchen Systems ist die Etablierung und Einhaltung von Kodierungsstandards. Im Bereich der Zeichenkodierung am bekanntesten und weitesten verbreitet sind der «American Standard Code for Information Interchange» (ASCII) und der sog. «Unicode». Parallel dazu existieren bis heute konkurrierende und von den genannten teils abweichende Kodierungen, z.B. die ISO-8859-Familie. In den weiteren Kontext der Zeichenkodierungsthematik, aber nicht mit dieser im engeren Sinn zu verwechseln, gehören auch die verschiedenen Verfahren zur – möglichst ökonomischen – Abbildung der dualen Zahlenwerte auf Frames zu je acht Bit, die als «Bytes» bezeichnet werden und in allen gängigen Betriebssystemen (Windows, Unix inklusive Mac OS X) als rechnerische Basiseinheit fungieren. Das bekannteste dieser Verfahren ist utf-8, daneben existiert eine Vielzahl weiterer, von denen durchaus einige, wie z.B. utf-16 oder utf-32, immer noch Verwendung finden.

Aus Sicht eines Literatur- oder Sprachwissenschaftlers ist die Kenntnis dieser Konzepte und Verfahren insofern unverzichtbar, als diese sich unmittelbar auf das Funktionieren von Suchen nach Zeichenmustern oder, auf höherer Ebene, nach Inhalten sowie – vielleicht weniger wichtig – auf Sortierreihenfolgen auswirken. Im schlimmsten Fall kann die Nichtbeachtung dieser Thematik dazu führen, dass mühsam zusammengetragene Textkorpora in inhomogener Zeichenkodierung vorliegen und somit praktisch unbrauchbar sind. Voraussetzung für eine sprach- und literaturwissenschaftliche Analyse unter Einsatz von DH-Methoden ist demnach das Vorliegen eines – aus einem oder mehreren Texten bestehenden – Textkorpus in homogener Zeichenkodierung. Das Konzept der Kodierung ist nicht auf die Ebene der einzelnen Zeichen beschränkt, sondern setzt sich gleichsam nach oben in die Bereiche komplexerer inhaltlicher Kategorien fort. Dabei ist es durchaus unterschiedlich, welche und wie solche Kategorien «höherer» Art in den elektronischen Texten kodiert werden. So müssen beispielsweise Überschriften als solche gekennzeichnet/kodiert werden. Im traditionellen Buchsatz erfolgte dies mit den Mitteln der Typographie, also etwa durch das Setzen in Majuskeln, durch Fettdruck oder auch durch Einrückungen. Textverarbeitungs- oder sog. Desktoppublishingprogramme waren und sind bemüht, diese traditionellen Konzepte in die digitale Welt zu übertragen und den Nutzern die Anmutung des auf Papier gedruckten Textes zu vermitteln (WYSIWYG – ‚What you see is what you get‘). In der Realität freilich sind auch diese Programme dazu gezwungen, die entsprechenden Informationen in Gestalt binärer Zahlen zu kodieren. Anders als auf der Ebene der Zeichenkodierung hatten sich diesbe-

zöglich jedoch lange keine Standards durchgesetzt oder vielfach, noch problematischer, war die entsprechende Kodierung nicht dokumentiert und blieb also ein Geheimnis der Softwareproduzenten. Mit dem Aufkommen der sog. Extensible Markup Language (XML) hat sich dies nun ein wenig geändert und entsprechende Kodierungen sind nunmehr transparent und nachvollziehbar. Als Beispiel mag das weit verbreitete Programm Microsoft Word, hier in der Version 2010, dienen. Die beiden folgenden Abbildungen zeigen den Anfang des *Secretum* von Francesco Petrarca («De secreto conflictu curarum mearum»). Abbildung 1 zeigt den Anfang des Textes in der vom Buchdruck her gewohnten Weise, so wie das Programm Word ihn auf dem Bildschirm präsentiert:¹³

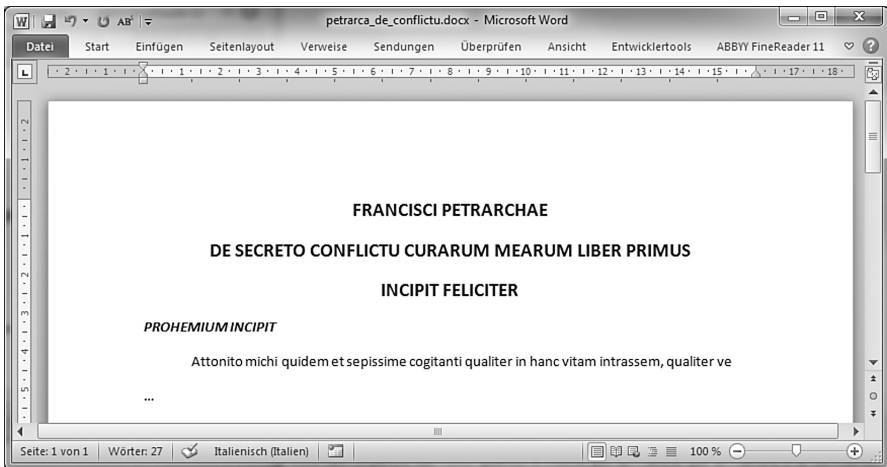


Abbildung 1: Microsoft Word 2010, traditionelle Kodierung mittels Typographie (WYSIWYG)

Aufgrund der typographischen Auszeichnung ist der Leser dazu in der Lage, die einzelnen Textteile bestimmten Kategorien wie «Überschrift erster Ordnung», «Überschrift zweiter Ordnung», «eigentlicher Text» etc. zuzuordnen. Die computergerechte Kodierung sieht notgedrungen anders aus. Zusätzlich zum «eigentlichen» Text müssen sämtliche Informationen wie etwa die Schriftgestalt, Zeilenausrichtung und -abstände, Einrückungen etc. unter Verwendung von binären Zahlen kodiert werden.¹⁴ Im Prinzip, d.h. sofern der Urheber der Datei keine entsprechenden Label (z.B. Formatvorlagen «Überschrift 1» etc.) verwendet hat, bleiben dabei semantische Kategorien weitgehend unberücksichtigt. Sieht man sich die oben abgebildete Datei mit einem als «Editor» bezeichneten Bearbeitungsprogramm an, ergibt sich folgendes Bild:¹⁵

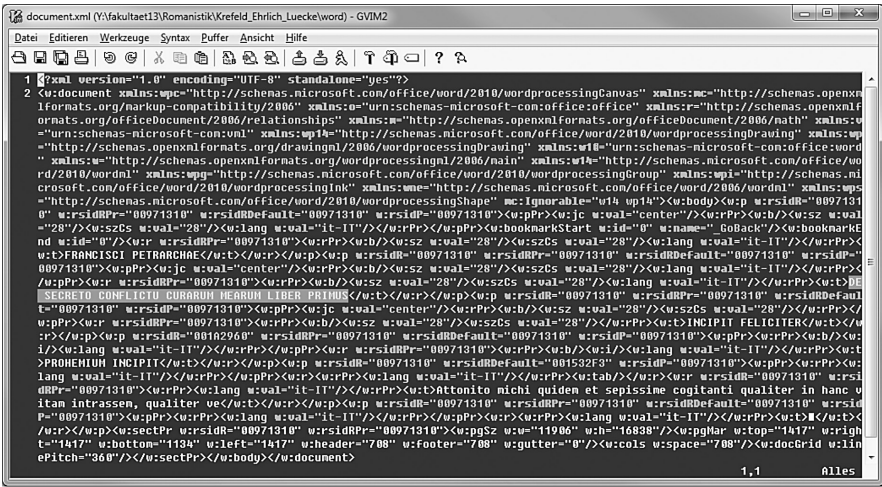


Abbildung 2: Microsoft Word-Dokument im XML-Format in der Ansicht des Editors gVIM

Für eine sprach- oder literaturwissenschaftliche Analyse dürften die meisten der hier kodierten Informationen in aller Regel irrelevant sein. Es erscheint daher vernünftig, Kodierungssysteme zu verwenden, die gezielt und ausschließlich Informationen enthalten, die im Sinne des jeweiligen Erkenntnisinteresses von Bedeutung sind, wobei die entsprechende Auswahl von eben jenem abhängt und daher nicht allgemeingültig definiert werden kann. In diesem Zusammenhang empfiehlt sich die Verwendung der eben erwähnten Editoren und die Abspeicherung der Daten in reinen Textdateien, die keine anderen Informationen enthalten, als den blanken Text, aus dem sie bestehen. Eine Möglichkeit, den soeben präsentierten Petrarca-Text im Sinne dieses Konzepts einzurichten, könnte so aussehen:

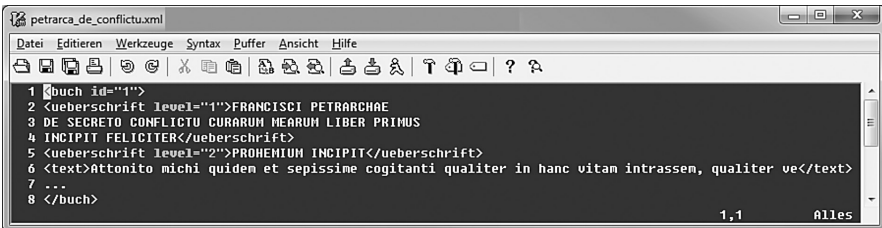


Abbildung 3: Individuelle Informationskodierung (Pseudo-XML)

Das hier angewandte Verfahren der Informationskodierung unter Verwendung sog. Tags, also Metadaten zwischen spitzen Klammern, deren Bezug jeweils

durch Anfangs- (<quodlibet>) und End-Tag (</quodlibet>) eindeutig angegeben ist und die in nahezu beliebiger Tiefe geschachtelt werden können, ist der XML entlehnt. Von echtem XML kann man in diesem Fall jedoch nicht sprechen, da hierfür die Einhaltung komplexer und klar definierter Syntaxregeln erforderlich ist. Da dergleichen recht aufwendig ist, ist die Definition und Anwendung solcher individueller, an XML angelehnter Informationskodierung pragmatisch und so lange legitim, wie es nicht um Datenaustausch mit Dritten, Publikation oder gar Langzeitarchivierung geht.

Selbstverständlich sind auf diese Weise auch literarische bzw. literaturwissenschaftliche – ebenso wie sprachwissenschaftliche – Kategorien kodierbar. So lassen sich beispielsweise Gedichtpassagen innerhalb eines Prosatextes mit entsprechenden <Tags> (<gedicht>...</gedicht> oder <lyrik>...</lyrik>) umfassen. Das skizzierte System ist im Prinzip beliebig erweiterbar und in sich robust, Voraussetzung ist lediglich die logische Konsistenz des gewählten oder entworfenen Tag-Repertoires sowie eine gewisse Sorgfalt bei der entsprechenden Auszeichnung der Korpustexte (Achten auf Orthographie der Tags, strukturelle Integrität etc.).¹⁶

Dergestalt etikettierte und strukturierte Daten können also im Prinzip beliebige und beliebig viele Informationen in kodierter Form enthalten. Ein weiterer Vorteil besteht darin, dass entsprechend aufbereitete Texte nach wie vor als Fließtext vorliegen und in gewohnter Weise lesbar sind – sämtliche <Tags>, also Zeichenfolgen zwischen spitzen Klammern, können, bei Verwendung geeigneter Programme und bei Bedarf, problemlos aus- und eingeblendet werden.

Im Hinblick auf rechnergestützte Analysen quantitativer Art allerdings bevorzugen die Autoren dieses Artikels das relationale Datenformat, in dem sämtliche Daten in Tabellen organisiert und abgespeichert werden. Zur Illustration bilden wir hier das oben gegebene Textbeispiel in relationaler Gestalt ab:

id_token	buch	ueberschrift_level	token	wortart	lyrik
1	1	1	FRANCISCI	Eigennamen	0
2	1	1	PETRARCHAE	Eigennamen	0
3	1	1	DE	Praeposition	0
4	1	1	SECRETO	Adjektiv	0
5	1	1	CONFLICTU	Nomen	0
6	1	1	CURARUM	Nomen	0
7	1	1	MEARUM	Pronomen	0
8	1	1	LIBER	Nomen	0
9	1	1	PRIMUS	Numeral	0
10	1	1	INCIPIIT	Verb	0
11	1	1	FELICITER	Adjektiv	0
12	1	2	PROHEMIUM	Nomen	0
13	1	2	INCIPIIT	Verb	0

Abbildung 4: Relationale Datenstruktur

Dies ist die Datenstruktur, die in relationalen Datenbanken wie etwa MySQL Verwendung findet. Weitere Kategorien von Metadaten, z. B. textkritischer Art (Varianten) oder metrischer Art (Versform), usw. können ebenso wie im Fall XML-artiger Strukturierung jederzeit nachträglich hinzugefügt werden. In diesem Fall erfolgt das durch Erweiterung der Tabelle um zusätzliche, in ihrer Anzahl theoretisch unbegrenzte, Kolumnen, in der Fachsprache der Datenbanken auch als «Attribute» bezeichnet. Datenbankmanagementsysteme (DBMS) wie MySQL sind problemlos dazu in der Lage, den ursprünglichen Fließtext – im Bedarfsfall auch unter Einbindung beliebig vieler Attribute – zu «re-synthetisieren» oder in XML-(artiger)-Gestalt zu exportieren. Folgende Abbildung zeigt die in Abbildung 4 präsentierten Daten als Fließtext ohne und mit Einbindung ausgewählter Metadaten:

Fließtext
FRANCISCI PETRARCHAE DE SECRETO CONFLICTU CURARUM MEARUM LIBER PRIMUS INCIPIIT FELICITER PROHEMIUM INCIPIIT
FRANCISCI [1 1 Lyrik=0 Wortart=Eigennamen] PETRARCHAE [1 1 Lyrik=0 Wortart=Eigennamen] DE [1 1 Lyrik=0 Wortart=Praeposition] SECRETO [1 1 Lyrik=0 Wortart=Adjektiv] CONFLICTU [1 1 Lyrik=0 Wortart=Nomen] CURARUM [1 1 Lyrik=0 Wortart=Nomen] MEARUM [1 1 Lyrik=0 Wortart=Pronomen] LIBER [1 1 Lyrik=0 Wortart=Nomen] PRIMUS [1 1 Lyrik=0 Wortart=Numeral] INCIPIIT [1 1 Lyrik=0 Wortart=Verb] FELICITER [1 1 Lyrik=0 Wortart=Adjektiv] PROHEMIUM [1 2 Lyrik=0 Wortart=Nomen] INCIPIIT [1 2 Lyrik=0 Wortart=Verb]

Abbildung 5: Relational strukturierte Daten in Fließtextgestalt

Zwar ist auch die Konvertierung von XML-Strukturen in ein relationales Datenformat grundsätzlich möglich, nach unseren bisherigen Erfahrungen erscheint der umgekehrte Weg jedoch technisch einfacher realisierbar. Ein großer Vorteil speziell relationaler Datenbanken besteht in der vergleichsweise hohen Geschwindigkeit, in der quantifizierende Berechnungen vorgenommen werden können. Es liegt in der Natur eines Computers, dass sich damit in erster Linie quantitative Analysen durchführen lassen: Die Geräte sind dazu in der Lage, in großer Geschwindigkeit große Mengen an Text quantifizierend zu durchleuchten. Es lässt sich auf diese Weise eine ganze Reihe von «Kennzahlen» ermitteln und zueinander in Beziehung setzen. So ist es, um bei oben gewähltem Beispiel zu bleiben, möglich, die Anzahl von Gedichten und deren Verteilung innerhalb eines Prosatextes festzustellen. Es ließe sich die durchschnittliche Länge der Gedichte berechnen, und generell können alle Werte auf bestimmte textexterne Parameter wie etwa «Textgattung», «Autor», «Entstehungsepoche», «Entstehungsort» etc. bezogen und zueinander in Beziehung gesetzt werden. Die sich ergebenden Zusammenhänge lassen sich über ihre gemeinsame Geo- und Chronoreferenzierung visualisieren. Dabei ist der Computer natürlich stets nur ausführendes Organ. Relevanz und Aussagekraft der mit diesen Methoden gewonnenen Ergebnisse können nur und müssen vom Menschen bewertet werden.

Aus methodischer Perspektive ist in diesem Zusammenhang z.B. ein heuristisches Vorgehen denkbar, das auf ein Wechselspiel zwischen induktiver Ermittlung spezifischer Kennzahl-Relationen (neben den bereits genannten z.B. Type-Token-Relation, Z-Score, usw.; der Phantasie sind hier kaum Grenzen gesetzt: Möglich wären beispielsweise auch Parameter wie die durchschnittliche Wortlänge oder das Verhältnis von Vokalen zu Konsonanten) für bekannte Textparameter und deduktivem Rückschluss auf korrelierende Kategorien in bis dato nicht analysierten Texten oder Textpassagen abzielt. Ein solcher Ansatz ist im Hinblick auf Fragen nach Urheberschaft oder Entstehungszeit eines Textes oder auch seine diesbezügliche Homogenität nahelegend. Der Italiener Franco Moretti, einer der Pioniere quantitativer Forschung in den Literaturwissenschaften, schlägt in seinem Aufsatz «Trees» (Moretti 1997) etwa vor, auf diese Weise (als morphologischen Vergleich), die historische Entwicklung der sog. «erlebten Rede», des «free indirect style»¹⁷ in den großen europäischen narrativen Texten von 1800–2000 aufzuzeigen. In dem kurzen Absatz geht er unter anderen auf die sozial-politischen Implikationen dieses im 19. Jahrhundert aufkommenden revolutionären narratologischen Verfahrens ein, das weder direkte Rede noch indirekte Rede ist und den Leser im Unklaren darüber lässt, ob es sich um subjektive Figurenrede oder (vermeintlich) objektive Erzählerrede handelt:

«Emotions plus distance: it is truly a peculiar mix, [...]. Placed as it is halfway between social doxa and the individual voice, free indirect style is a good indicator of their changing balance of forces, of which the tree offers a schematic visualization.» (Moretti 2005, S. 81–85)

Franco Moretti erläutert sein methodisches Vorgehen in dem Text leider nicht näher – möglicherweise handelt es sich noch um analog gewonnene quantitative Ergebnisse –, aber es kann davon ausgegangen werden, dass das erstmals 1887 in der Zeitschrift für romanische Philologie beschriebene Verfahren der erlebten Rede, als Mischform von direkter und indirekter Rede, von Erzähler- und Figurenperspektive,¹⁸ anhand linguistischer Annotation in einer relationalen Tabelle gut digital erfasst werden kann: Formal lässt sich die erlebte Rede mit Hilfe einer Kombination von positiven und negativen Merkmalen suchen, die jedes für sich genommen unspezifisch für diese Erzähltechnik sind: kein redeenleitendes Sprechaktsverb, dritte Person, kein Präsens, lokal- und temporaldeiktischer Bezug zum hic-et-nunc der Figur (und nicht des Erzählers), Gesprächswörter und Modalpartikel, die eigentlich für die direkte Rede charakteristisch sind. Ein willkürliches Beispiel für einen Satz in erlebter Rede

könnte sein: «Morgen war Ferragosto», oder aber ein italienisches Beispiel aus Niccolò Ammanitis *L'ultimo capodanno della umanità*:

«Cristiano Carucci aveva in testa tre possibilità per sfangare quella maledettissima notte. [...] In programma quella sera [hic-et-nunc] c'era [Vergangenheitstempus; 3. Person] la megaspinnellata di capodanno [subjektive Figurenperspektive] e il concerto degli Animal Death.» (Ammaniti 1996, S. 11)

Da die aufgezeigten sprachlichen Kategorien in annotierten digitalen Texten rasch aufgespürt werden können, lassen sich auch umfangreiche Korpora nach Vorkommnissen der erlebten Rede durchforsten, so dass historische Entwicklungen, Abhängigkeiten hervortreten und sich die Differenzierung von Gattungsmerkmalen aufzeigen lässt.¹⁹

Im Sinne des diesem Aufsatz vorangestellten Mottos ist schließlich noch hervorzuheben, dass sämtliche auf diese Weise gewonnenen Erkenntnisse stets bis zur darunterliegenden Datenbasis zurückverfolgbar und somit verifizierbar sind, wodurch ihnen selbst wie auch allen darauf fußenden Schlüssen und Thesen ein besonderes Maß an Autorität und Relevanz verliehen wird.

Es sei darauf hingewiesen, dass es im Internet eine Reihe von Programmen und Online-Tools gibt, die mehr oder weniger komfortable Lösungen für quantitative Analysen von Texten und ganzen Textkorpora anbieten und deren Stärke nicht selten in unterschiedlichen und teils aufwendigen Visualisierungsoptionen in Form von Diagrammen besteht. Eine – sicherlich nicht vollständige – Liste solcher Dienste und Programme findet sich im Anhang.

Forschungshorizont

Da quantitative Analyse immer nur dann möglich ist, wenn die zugrunde gelegten Einheiten hinreichend definiert sind, und da die grundlegende Struktur von Sprache sehr viel klarer auf positive Art zu definieren ist als diejenige von «Text» oder gar «Literatur», scheint es im Bereich der Textwissenschaften sinnvoll, aus einer gemeinsamen Perspektive zu operieren. Das bedeutet ganz konkret, dass der Versuch unternommen werden muss, literarische Phänomene über sprachliche und somit «rechenbare» Merkmale zu beschreiben und womöglich in Bezug zu anderen, demographischen, politischen, ökonomischen etc. Daten zu setzen. Im Blick auf den philologischen Kernbereich eröffnet sich auf digitaler Grundlage zweifellos ein neu gewonnenes, oder: wiedergewonnenes, Feld der Zusammenarbeit von Literaturwissenschaft und Linguistik. Ist die Linguistik seit jeher daran gewohnt, Fragestellungen quantitativer Art zu

entwickeln, gelingt das der Literaturwissenschaft meist über den Umweg des (genuin sprachwissenschaftlichen) Strukturalismus.²⁰

Die Initiative ›DH‹ ist indes aber kein ›Rückschritt‹ mit aktualisierten Tools, sondern kann und muss die theoretischen Vorarbeiten des Strukturalismus ebenso wie die des Poststrukturalismus konstruktiv zur Gewinnung neuer Erkenntnisse – auch und vor allem außerhalb des philologischen Kontexts und bezüglich des spezifischen Einzelfalls – nutzbar machen.

Es wurde vielerorts polemisch die Befürchtung geäußert, dass die natürliche Kontingenz literarischer/künstlerischer Phänomene auf Grund von Durchschnittsberechnungen verlustig ginge. Tatsächlich birgt die Abstraktion die Gefahr, das Besondere durch das Allgemeine zu überdecken. Es ist doch gerade das Besondere, der überraschende Einzelfall das, was Literatur immer wieder aufs Neue ausmacht. Mit diesem Paradoxon gilt es sich unbedingt auseinanderzusetzen!

Sicherlich können errechnete Standards die Interpretation des «besonderen» Texts nicht ersetzen. Den Einzelfall müssen Verfahren des ›close-reading‹ beschreiben.

Dennoch: Wenn Moretti zugespitzt formuliert: «Reading ›more‹ seems hardly to be the solution» (Moretti 2000), hat er nicht ganz unrecht. Denn nur über eine große (qualifizierte) Menge an Daten können wir die Abweichungen vom (tatsächlichen) Standard einer Literatur – sei es einer Epoche, eines Autors oder eines geographischen Zusammenhangs – überhaupt erst erkennen und eben den Einzeltext finden, den andernfalls vielleicht gar niemand gelesen hätte.

Das Vorgehen hat sich in Zeiten von DH also nicht in jeder Hinsicht verschlankt. Im Gegenteil entstehen nun zwei Schritte:

(1) «Es muss [...]abstrahiert werden von den kontingenten Elementen, um die stabilen, dauerhaften Elemente zu ermitteln» (Jannidis 1999).

Daraus können wichtige Schlüsse insbesondere für die Literatur- und Sprachgeschichte gezogen werden.

(2) Vor dem Hintergrund von (1) erhalten die notwendigen Einzelfallanalysen und Interpretationen eine ebenso transparente wie solide historische Kontextualisierung.

Es ist also keineswegs zu befürchten, dass den Literatur- und Sprachwissenschaften durch die neuen Technologien ein Aufgabenbereich abhanden kommt, sondern im Gegenteil zeichnet sich eine bedeutsame Erweiterung der möglichen Erkenntnisziele und formulierbaren Fragen ab. Fragen an einen immer wieder neu zu definierenden Gegenstand.

Hinzu kommt die Chance einer wesentlich breiteren gesellschaftlichen Wahrnehmung geisteswissenschaftlicher Argumentation und Erkenntnisinteressen durch die digitale Infrastruktur und Methoden wie *crowdsourcing* sowie die neue interdisziplinäre Kooperation genuin unterschiedlicher Fächer wie der IT-Wissenschaften und der Geisteswissenschaften. Mindestens ebenso wichtig wie die bloße Bereitstellung und Entwicklung der Tools ist die tatsächliche und fundierte Wissensvermittlung in alle Richtungen zwischen den Fakultäten: so kann einerseits dem Philologen ein kritischer Umgang mit den Medien und Tools erst ermöglicht werden, der neben dem Bewusstsein für die Möglichkeiten auch die Grenzen und Gefahren kennt, und andererseits werden dem IT-Spezialisten philologische Entwicklungen, spezielle Fragestellungen und Erkenntnisziele, auch analoger (!) Art, nicht vorenthalten. Nur aus einem möglichst exhaustiven gegenseitigen Verständnis kann wirklich effiziente Kollaboration entstehen. Helle Porsdam, Altphilologin und Professorin für Amerikanistik an der Universität Kopenhagen hat sich jüngst in einem Papier für «a better balance between the ›how‹ and the ›what‹ of DH» ausgesprochen. Ihrem Plädoyer, als Antwort auf das Manifest von Burdick et al. 2012²¹ «for a qualitative turn of sorts away from an interest in gaining and making accessible more information only, to an interest in also making sense of and understanding that information» darf man sich wohl anschließen. Die humanities sind zweifellos aufgefordert, der kritischen Interpretation der gewaltigen Informationsmasse den Weg zu weisen. In diesem Sinne wollen wir abschließend den folgenden, von Domenico Fiormonte lancierten Appell weitergeben:

«in questo delicato momento di passaggio di formato della nostra memoria culturale, gli umanisti [sono] chiamati a dare un contributo non in veste di semplici fornitori o consumatori di contenuti, ma come creatori e architetti delle nuove forme della conoscenza.»²²

Abstract. Oggigiorno il passaggio delle scienze umanistiche alle *digital humanities* è inevitabile. Che piaccia o no, sono ormai pochi coloro si ostinano a non (volere) riconoscere la formazione di un nuovo orizzonte di ricerca. In linea di massima esso è caratterizzato da quattro fattori: (1) La base empirica è costituita da dati digitali, la cui struttura sia possibilmente già stata annotata a priori; i metodi sono quantitativi e in larga misura induttivi. (2) La comunicazione scientifica si svolge in base alle condizioni dettate da internet, combinando e incrociando media diversi (testo, audio, immagine, film ecc.) e generi diversi (corpus, lessico, apparato critico, grammatica ecc.). (3) Il lavoro (progettuale)

diventa maggiormente collaborativo, fino a mettere fortemente in dubbio l'ideale del ricercatore e del genio individuale. (4) La registrazione dei risultati sotto forma esclusivamente materiale e circoscritta (libri, cd, dvd) non è più adatta alla comunicazione scientifica mediata da internet; va quindi sostituita con piattaforme di ricerca che possono essere costantemente modificate e arricchite sia dal punto di visto empirico che metodologico.

Anmerkungen

1 Hier wird ein pragmatischer Medienbegriff zugrunde gelegt, der alle Verfahren und Mittel die durch neue, digitale Technologien ermöglicht werden impliziert. Sabine Heuser definiert «Neue Medien» in diesem Sinne als einen «multifunktionalen Verbund, der digitale und herkömmliche Medien mittels Computer verbindet. Der digitale Code ist universell und kann alle analogen Zeichensysteme verarbeiten, deshalb ist er nicht nur intermedial sondern transmedial» (Heuser 2008, S. 468).

2 Der «Companion to DH» sollte anfangs «A Companion to Humanities Computing» oder «Companion to Digitized Humanities» heißen, wurde dann aber aus ideologischen Gründen unter dem bekannten Titel publiziert, «to shift the emphasis away from simple digitization» (vgl. Kirschenbaum 2012, S. 3). Brett Bobley, Chief Information Officer des National Endowment for the Humanities (<http://www.neh.gov/>) in den USA, das unter seiner Führung 2006 die Digital Humanities Initiative (<http://www.dhinitiative.org/>) gegründet hatte, erklärt in diesem Zusammenhang bezeichnenderweise: «[...] <digital humanities> implied a form of humanism.» (Ebd., S. 4).

3 Domenico Fiormonte (Università Roma Tre) setzt sich in mehreren Artikeln detailliert mit nicht anglo-amerikanischer Institutionalisierung der DH auseinander, vgl. insbesondere auch Numerico/Fiormonte/Tomasi (2010) und den von Fiormonte gemeinsam mit Paolo Sordi betriebenen Blog: <http://infolet.it/>.

4 Dies ist z. B. bei der fachwissenschaftlichen Operationalisierung des sogenannten *crowdsourcing* der Fall; aus dieser facettenreichen wissenschaftlichen Herausforderung ergeben sich auch für die Informatik zahlreiche Impulse.

5 Christof Schöch weist zurecht darauf hin, dass «language, texts, paintings, and music are semiotic systems that have dimensions beyond the physically measurable, dimensions which depend on semantics and pragmatics, that is on meaning in context. For this latter reason particularly, speaking of <data> [also dem per definitionem <Quantifizierbaren>] in the humanities is problematic and has been challenged.» Er verweist auf die Möglichkeit eine semantisch präzisere und weniger problematische Terminologie zu verwenden wie dies von Joanna Drucker vorgeschlagen wird, «[who] prefers to speak of <capa> instead of data, literally <that which has been captured or gathered>, underlining the idea that even the very act of capturing data in the first place is oriented by certain goals, done with specific instruments, and driven by a specific attention to a small part of what could have been captured given different goals and instruments. In other words, capturing data is not passively accepting what is given, but actively constructing what one is interested in.» (Schöch 2013, S. 3).

6 Die Methoden und Erkenntnisziele der Literaturwissenschaften galten ja seit dem späten 17. Jahrhundert bis mindestens in die 60er Jahre des 20. Jahrhunderts (produktiver Höhepunkt des Strukturalismus als interdisziplinäre Forschungsrichtung) als gerade nicht

quantitativ formulierbar, oder formalisierbar und ob der Subjektivität ihres Gegenstandes, um eine je individuelle, textnahe («close reading!») Lektüre von Artefakten bemüht. Derrida spricht von Literatur als «eine[r] Institution, die alle referentialisierenden Bedeutungs-zuweisungen als trügerische Referenzen unterläuft» (Derrida, *Acts*, zit. nach Jannidis 2009, S. 4).

7 Vgl. für eine kurze Begriffsgeschichte z. B. Simonis (2008). Um ein rezentes italienisches Beispiel für den «Geniekult» zu nennen, sei Andrea Cortelassa (2011) auf der Suche nach einer zeitgemäßen Positions- und Aufgabenbestimmung des Literaturwissenschaftlers in dem – im Grunde sehr gelungenen – Sammelband *dove siamo. nuove posizioni della critica* zitiert. Er beruft sich auf keinen geringeren als Pier Paolo Pasolini, wenn er den Intellektuellen im Allgemeinen und den Literaturwissenschaftler im Besonderen stilisiert als «colui che sa, ma non ha le prove» [Hervorhebung im Original] (S. 18).

8 Eine intensive Diskussion um Forschungsgegenstand und Erkenntnisinteresse der Philologischen Fächer und im speziellen um die definitorischen Grenzen von Text und Literatur begann bereits in den 70er Jahren; die Publikationen zum Thema sind seit der digitalen Wende fast unüberschaubar: es sei hier, stellvertretend, für den deutschen Raum, auf den 2009 erschienene Sammelband *re/visionen. Grenzen der Literatur*, hrsg. von Simone Winko, Fotis Jannidis und Gerhard Lauer verwiesen. Für den italienischen Sprachraum vgl. z. B. Matteo di Gesù (2011).

9 Vgl. auch Porsdam (2012): «The more widespread the use of digital technology becomes, and the more humanities scholars start working through as opposed to merely with computers, the more important it becomes for us to look at the ways in which digital technologies transform knowledge production and scholarly research.» (§7).

10 Vgl. z. B. das einleitende Plädoyer von Jannidis/Lauer/Winko (2009) für eine radikale Historisierung und damit einen pragmatischen Literaturbegriff (S. 3–40).

11 Dazu muss bemerkt werden, dass ohne diese großen und teils sehr arbeits- und kostenintensiven Projekte um digitale Editionen, welche wissenschaftlich mit Metadaten aufbereiteten Text zur Verfügung stellen, kaum weitere qualitative digitale Forschung in den Textwissenschaften möglich ist. Vgl. dazu Schöch 2013: «Only smart big data enables intelligent quantitative methods» (S. 11). Vgl. z. B. das Dante Dartmouth Project: <http://dantelab.dartmouth.edu/> und <http://dante.dartmouth.edu/>

12 Vgl. Schöch 2013. Schöch argumentiert sehr aufschlussreich wie und warum *crowd-sourcing* und *automatic (linguistic) annotation* Hand-in-Hand gehen müssen und werden. (Ebd., S. 10)

13 Quelle: http://www.liberliber.it/mediateca/libri/p/petrarca/de_secreto_conflictu_curarum_mearum/pdf/de_sec_p.pdf (22.09.14).67

14 In Fiormonte 2008, S. 77 (Esempio 3) findet sich ein Beispiel für die Syntax einer XML-TEI-Kodierung eines Autographs mit unterschiedlichen, zu verschiedenen Zeitpunkten wieder verworfenen Varianten von Valerio Magrelli.

15 Eine Microsoft Word-Docx-Datei ist eigentlich ein ZIP-Archiv, in dem eine ganze Reihe von Einzeldokumenten im XML-Format mit unterschiedlicher Funktion zusammengefasst ist. Das hier abgebildete XML-Dokument «document.xml» enthält den eigentlichen Text der Docx-Datei.

16 Unterstützung bei der Arbeit mit XML-Dateien bzw. XML-artigen Daten bieten einschlägige Programme wie z. B. Oxygen. Diese verhindern versehentliches Eingeben nicht vorgesehener Tags oder Syntaxverstöße wie etwa das Vergessen korrespondierender End-Tags.

17 Italienisch: stile indiretto libero

18 Hinter dem Erzähler wird unweigerlich der Autor vermutet, was insbesondere bei der Rezeption von Romanen mit personaler Erzählhaltung zu Mehrdeutigkeiten und Rezeptionsschwierigkeiten führen kann; das historisch vielleicht bekannteste Beispiel dafür ist der Skandal um *Madame Bovary*.

19 Hier ist selbstverständlich auch mit Unschärfen oder gar mit der Unmöglichkeit der Umsetzung zu rechnen: Der Versuch jedoch bedarf im Sinne der heuristischen Methode keiner Rechtfertigung.

20 Literarizität und Poetizität, also die Eigenschaften, die bei sprachlichen Äußerungen überhaupt erst das Literarische ausmachen, sind aber in der neueren literaturtheoretischen Diskussion längst keine rein ästhetischen, nur ‚gefühlten‘ Merkmale mehr. So findet man im Reallexikon der deutschen Literaturwissenschaft unter dem Lemma ‚Literarizität‘ die Bedeutung «(1) eine besondere Art der Sprachverwendung [...], die Syntax bzw. Grammatik (Kohäsion), Semantik (Referenz) und Pragmatik (situativer Bezug) betrifft.» Diese Kategorien kann die Linguistik längst ‚rechnen‘!

21 Anne Burdick, *Digital Humanities*, Cambridge, Mass. u.a.: MIT Press 2012.

22 Domenico Fiormonte, «Pragmatica Digitale. Paratesti, Microtesti e ‚Metatesti‘ nel Web,» in: Dardano, Maurizio/Frenguelli, Gianluca/De Roberto, Elisa, *Testi Brevi. Atti del Convegno internazionale di studi* (Università Roma Tre, 8–10 giugno 2006), Roma: Aracne 2008.

Bibliographie

Alfano, Giancarlo et. al. (2011): *dove siamo. nuove posizioni della critica letteraria*.

Palermo: duepunti edizioni.

Ammaniti, Niccolò (1996): «L'ultimo capodanno dell'umanità», in: ders., *Fango*, Milano: Mondadori, S. 7–136.

Burdick, Anne (2012): *Digital Humanities*. Cambridge, Mass. u.a.: MIT Press.

Cortelassa, Andrea (2011): «Intellettuale, Anni Zero,» in: Alfano et. al. (2011), S. 15–35.

Dante Dartmouth Project: <http://dantelab.dartmouth.edu/> und <http://dante.dartmouth.edu/> (Zugriff am 21.9.2014).

Fotis Jannidis (1999): «Was ist Computerphilologie?», in: *Jahrbuch für Computerphilologie* 1 . URL: <http://computerphilologie.uni-muenchen.de/jahrbuch/jb1/jannidis-1.html> (Zugriff am 21.9.2014).

Fiormonte, Domenico (2008): «Pragmatica digitale. Paratesti, microtesti e ‚metatesti‘ nel web,» in: Dardano, Maurizio/Frenguelli, Gianluca/De Roberto, Elisa (Hrsg.), *Testi Brevi. Atti del Convegno internazionale di studi* (Università «Roma Tre», 8–10 giugno 2006). Roma: Aracne.

Fiormonte, Domenico (2012): «Towards A Cultural Critique of Digital Humanities», in: Manfred Thaler (Hrsg.) *Controversies Around the Digital Humanities (Historical Social Research/Historische Sozialforschung; Special Issue)*, S. 59–76.

di Gesù, Matteo (2011): «L'affidabilità di un marchio garantito: Letteratura italiana. Since 1870 (se non prima)», in: Alfano et. al. (2011), S. 61–75.

Gold, Matthew K. (Hrsg.) (2012): *Debates in the Digital Humanities*. Minnesota: University of Minnesota Press. Open-Access Edition: <http://dhdebates.gc.cuny.edu/> (Zugriff am 21.9.2014).

Hockey, Susan (2004): «The History of Humanities Computing», in: Schreibman/Siemens/Unsworth (2004), Kap. 1.

- Heuser, Susanne (2008): «Neue Medien,» in: *Metzler Lexikon Literatur und Kulturtheorie*, hrsg. von Ansgar Nünning, 4. Auflage, Stuttgart: Metzler, S. 468–469.
- Jones, Steven E. (2014): *The Emergence of the Digital Humanities*. New York/London: Routledge.
- Kirschenbaum, Matthew (2012): «What is Digital Humanities and What's It Doing in English Departments», Einleitung in: Gold, Matthew K. (2012).
- Lazzari, Marco/Betella, Alberto/Bianchi, Alessandro et. al. (2014): *Informatica umanistica*. New York: McGraw-Hill.
- LMU Center for Digital Humanities: <http://www.itg.uni-muenchen.de/index.html> (Zugriff am 21.9.2014)
- Moretti, Franco (1999): *Atlante del romanzo europeo*. Torino: Einaudi 1997; (dt. Ausgabe: *Atlas des europäischen Romans*, Köln: Dumont.)
- Moretti, Franco (2000): «Conjectures on World Literature», in: *New Left Review* 1, <http://newleftreview.org/II/1/franco-moretti-conjectures-on-world-literature> (Zugriff am 22.9.2014).
- Moretti, Franco (2007): *Graphs, Maps, Trees: Abstract Models for a Literary History* (dt. Ausgabe: gemeinsam mit Florian Kessler, *Kurven, Karten, Stammbäume: Abstrakte Modelle für die Literaturgeschichte*, Frankfurt: Suhrkamp 2009.)
- Moretti, Franco (2013): *Distant Reading*. London/New York: Verso.
- Numerico, Teresa/Fiormonte, Domenico/Tomasi, Francesca (2010): *L'umanista digitale*. Bologna: il Mulino.
- Orlandi, Tito (2007): «Un ultimo bilancio dell'informatica umanistica», in: *Atti del Convegno E-laborare il sapere nell'era digitale*, Montevarchi 22–23 novembre 2007, URL: <http://rmcisadu.let.uniroma1.it/~orlandi/pubbli/informatica/montevarchi.pdf> (Zugriff am 21.9.2014).
- Orlandi, Tito (2012): «Per una storia dell'informatica umanistica», in: Ciotti, Fabio/Crupi, Gianfranco (Hrsg.), *Dall'Informatica umanistica alle culture digitali. Atti del convegno di studi* (Roma, 27–28 ottobre 2011) in memoria di Giuseppe Gigliozzi, Roma, La Sapienza: Digilab, S. 49–102.
- Porsdam, Helle (2013): «Digital Humanities: On Finding the Proper Balance between Qualitative and Quantitative Ways of Doing Research in the Humanities», in: *Digital Humanities Quarterly* 7/3. <http://www.digitalhumanities.org/dhq/vol/7/3/000167/000167.html> (Zugriff am 21.9.2014).
- Renear, Allen H. (2004): «Text Encoding», in: Schreibman/Siemens/Unsworth (Hrsg.), Kap. 17.
- Schreibman, Susan/Siemens, Ray/Unsworth, John (2004) (Hrsg.): *A Companion to Digital Humanities*, Oxford: Blackwell. Open-Access Edition: <http://www.digitalhumanities.org/companion/> (Zugriff am 21.9.2014).
- Schöch, Christof (2013): «Big? Smart? Clean? Messy? Data in the Humanities», in: *Journal of the Digital Humanities* 2.3, 2013, 2–13. URL: <http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/> (Zugriff am 21.9.2014)
- Simonis, Linda (2008): «Geistesgeschichte», in: *Metzler Lexikon Literatur und Kulturtheorie*, hrsg. von Ansgar Nünning, 4. Auflage, Stuttgart: Metzler, S. 242–244.
- Winko, Simone/Jannidis, Fotis/Lauer, Gerhard (2009): *rel/visionen. Grenzen der Literatur*. Berlin/New York: De Gruyter. http://de.wikipedia.org/wiki/Digital_Humanities (Zugriff am 21.9.2014).

Fachverbände, Forschungszentren und Forschungsplattformen

- Associazione per l'informatica umanistica e la cultura digitale (AIUCD):
<http://www.umanisticadigitale.it/> (Zugriff am 21.9.2014).
- Alliance of Digital Humanities Organizations (ADHO): <http://adho.org/> (Zugriff am 21.9.2014).
- CLARIN-D Common Language Resources and Technology Infrastructure,
 URL: <http://de.clarin.eu/de/home.html> (Zugriff am 21.9.2014)
- Cologne Center for eHumanities: <http://www.cceh.uni-koeln.de/>(Zugriff am 21.9.2014).
- Digital Humanities im deutschsprachigen Raum (DHd): <http://www.dig-hum.de/>
 (Zugriff am 21.9.2014).
- Digital Humanities Initiative: <http://www.dhinitiative.org/> (Zugriff am 21.9.2014).
- National Endowment for the Humanities (NEH): <http://www.neh.gov/> (Zugriff am 21.9.2014). Tesoro della Lingua Italiana delle Origini, URL: <http://tlio.ovi.cnr.it/TLIO/> (Zugriff am 21.9.2014).

Tools

a) Editoren

- vim: <http://www.vim.org>
- jEdit: <http://www.jedit.org/>
- notepad++: <http://notepad-plus-plus.org/>
- Oxygen: <http://www.oxygenxml.com/>; XML-Editor (kostenpflichtig)

b) Textanalyse

- catma - computer aided textual markup & analysis: <http://www.catma.de/>
- Google ngram viewer: <https://books.google.com/ngrams>
- juxta commons: <http://juxtacommons.org/> Kollationierungstool, Editionsphilologie
- mallet: <http://mallet.cs.umass.edu/topics.php> 'topic-modelling', entdeckt thematische Strukturen
- MySQL: <http://www.mysql.de/>- R: <http://www.r-project.org/> 'Statistical Computing'
- TreeTagger: <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/> Part-of-Speech (POS) Tagging
- tropes: <http://www.semantic-knowledge.com/tropes.htm>, (kostenpflichtig)
- txm: <http://textometrie.ens-lyon.fr/spip.php?article67&lang=en> Textométrie, POS-Tagging- Unix-Tools: <http://www.cygwin.com/>
- Voyant-Tools: <http://voyant-tools.org/> Visualisierungen, Clouds, tabellarisierte Wortfrequenzen, Kontextanalyse