

Fotis Jannidis, Gerhard Lauer, Andrea Rapp (Göttingen)  
**Hohe Romane und blaue Bibliotheken. Zum Forschungsprogramm einer computergestützten Buch- und Narratologiegeschichte des Romans in Deutschland (1500–1900)**

Im Jahr 1742 wurde Johann Friedrich Bachstrom in der litauischen Stadt Nieswicz erwürgt. So wenig erfreulich endete das Leben eines Mediziners, Theologen, Pädagogen und Romanschriftstellers,<sup>1</sup> dessen 1736/37 anonym in Breslau in zwei Teilen erschienener utopischer Roman *Bey zwey hundert Jahr lang unbekante, nunmehr aber entdeckte vortreffliche Land der Inquiraner* heute nicht einmal einer Handvoll Spezialisten bekannt ist. In ihm erzählt Bachstrom auf der Grundlage französischer und englischer Vorlagen und seiner eigenen Reisebegegnungen in der Türkei von den Abenteuern verschiedener religiöser Dissidenten. Verurteilt zur Galeere, erleiden sie Schiffbruch, machen eine Zeit der Robinsonade durch, um schließlich wiedervereint am Rande eines nicht genauer benannten nordafrikanischen Gebirges einen neuen Staat der unbedingten Religionstoleranz zu errichten. Der ungewöhnliche Roman Bachstroms verknüpft Robinsonade mit Religionskritik und -utopie, in der Absicht, »daß weder die Gemeinschaft der Bürger durch die Religion noch die Gemeinschaft der Religion durch die bürgerliche Gesellschaft im geringsten nicht gekräncket, gehindert und gestöret werde.«<sup>2</sup> Ein solcher Roman war im frühen 18. Jahrhundert so heterodox wie die Meinungen seines Autors, dass Früchte gegen Skorbut helfen würden, Matrosen Schwimmen lernen sollten oder Mädchen das Medizinstudium erlaubt werden dürfe, und zugleich so konventionell, benutzt er doch fast alle damals gängigen Erzählmuster von der menippeischen Satire Lukians über Montesquieus *Lettres persanes* bis zu Defoes *Robinson Crusoe*. In den heutigen Literatur- und Gattungsgeschichten zum Roman kommt er nicht mehr vor.<sup>3</sup>

Romane wie dieser des unglücklichen Johann Friedrich Bachstrom sind aus unserem kulturellen Gedächtnis getilgt, aus den Kanones der Literaturgeschichte und fachwissenschaftlichen Gegenstände verschwunden. Die Geschichte des Romans in Deutschland scheint sich in wenigen Sprüngen von Grimmelshausen über Schnabel und Gellert hin zu Goethe erzählen zu lassen. Die meisten der tatsächlich geschriebenen und gelesenen Romane kommen in dieser Literaturgeschichte nicht vor. Weder von Bachstrom noch von den galanten Romanen ist die Rede,<sup>4</sup> die mal Übersetzungen sind, mal sich als solche ausgeben, noch von den Ritter- und Feenromanen, die auf blauem, dafür billigem Papier gedruckt und verbreitet wurden, noch weniger von den vielen Büchern, die bei einem erfundenen Drucker namens Pierre Marteau angeblich über mehr als hundert

---

<sup>1</sup> Herbert Schönebaum: *Art. Johann Friedrich Bachstrom*. In: *Allgemeine Deutsche Biographie*. Hg. v. der Historischen Kommission bei der Bayrischen Akademie der Wissenschaften. Leipzig: Duncker & Humblot 55 (1911), S. 664–667 und 56, S. 395 (Korrektur) und Hermann Ullrich: *Art. Johann Friedrich Bachstrom*. In: *Neue Deutsche Biographie*. Hg. v. der Historischen Kommission bei der Bayrischen Akademie der Wissenschaften. Berlin: Duncker & Humblot 1 (1953), S. 503.

<sup>2</sup> A.B.C. [Johann Friedrich Bachstrom]: *Das Bey zwey hundert Jahr [...]*. Frankfurt u. Leipzig [wohl: Breslau: Korn], 1737, 2. Teil, S. 137.

<sup>3</sup> Dafür wenigstens unter den Rubriken Gelehrtengeschichte und Orientalismuskforschung, vgl. Hermann Ullrich: *Johann Friedrich Bachstrom. Ein Gelehrtenleben aus der ersten Hälfte des 18. Jahrhunderts*. In: *Euphorion* 16 (1909), S. 28–58 und 320–349 und Christoph Boehinger: *Pietistische Identität zwischen persönlicher Frömmigkeit und Gruppenprozessen*. In: *Hallesche Beiträge zur Orientwissenschaft* 24 (1997), S. 33–44.

<sup>4</sup> Einen Einblick gewähren Thomas Borgstedt u. Andreas Solbach (Hg.): *Der galante Diskurs. Kommunikative Ideale und Epochenwelle*. Dresden: Thelem 2001 (= Arbeiten zur neueren deutschen Literatur; 6).

Jahre hinweg in Köln gedruckten wurden<sup>5</sup> oder von Aetna-Romanen, die man nicht unter der Rubrik *fabulae romanensis*, sondern unter den medizinischen Schriften eingereiht hat. Die Geschichte des Romans in den deutschsprachigen Ländern ist, bis auf einige stark befahrene Küstenstrassen, ein weitgehend unbekannter Kontinent.

Das ist in anderen nationalen Forschungskontexten anders. In Spanien, Frankreich oder England gehört die Geschichte der Entstehung des Romans geradezu zu den nationalen Gedächtnisorten.<sup>6</sup> Dabei geht es nicht nur um die bibliographische Nachweise immer weiterer, bislang unbekannter Titel des versunkenen Kontinents der Drucke,<sup>7</sup> sondern auch um die Geschichte des Buches und die des Erzählens. Wer hat, wann und wo zum ersten Mal die Form der erlebten Rede eingesetzt, wer die episodische Reihung zu psychologischer Figurenzeichnung verdichtet? Wo verknüpfen sich populäre Novel-lenstoffe mit hochkulturellen Erzähltechniken, wo Pornographie mit philosophischen Ansprüchen des Erzählers? Das sind Fragen, die zu beantworten voraussetzt, Buch- und Kulturgeschichte des Romans mit der Geschichte des Erzählens zu verknüpfen. Nur ist ihre methodische Integration alles andere als einfach, und deshalb ist es der Forschung bislang überhaupt nur punktuell gelungen, beides zu integrieren. Das hängt nicht zuletzt damit zusammen, dass eine Geschichte des Erzählens ein statistisch hinreichendes Korpus braucht, um zu validen Aussagen über die Entwicklung der Gattung über einen längeren historischen Zeitraum kommen zu können. Für den deutschsprachigen Raum liegen dafür noch nicht einmal hinreichende bibliographische Daten vor.<sup>8</sup> Wer daher eine Kulturgeschichte des Romans im deutschsprachigen Raum angehen will und zugleich eine Geschichte des Erzählens schreiben möchte, der muss drei Dinge tun: erstens Rekanonisieren, also die Verengung auf die nur hochkulturelle Romanliteratur zugunsten eines weiten Begriffs des Romans öffnen auf das, was tatsächlich gedruckt und gelesen wurde, zweitens Bibliographieren und das auch in den Bibliotheken vor Ort und er muss drittens Digitalisieren. Denn nur mit Hilfe des Computers lässt sich ein hinreichend großes Korpus über einen langen Zeitraum hinweg untersuchen und damit etablierte literaturhistorische Methoden um serielle, computerbasierte Verfahren so ergänzen, dass eine Geschichte des Erzählens überhaupt erst geschrieben werden kann.

Das alles als Forschungsprogramm anzugehen, erhält seinen besonderen Nachdruck durch den digitalen Kampf um das kulturelle Gedächtnis. Die Transformation des kulturellen Erbes vom Medium des Buches in den digitalen Speicher hat seit dem Einstieg der großen Suchmaschinen-Firmen eine ganz neue Dramatik erreicht als dies akademi-

<sup>5</sup> Vgl. das Projekt *Pierre Marteau's Publishing House* unter <<http://www.pierre-marteau.com>> (5.1.2006) und die exzeptionelle Arbeit dazu von Olaf Simons: *Marteaus Europa oder Der Roman, bevor er Literatur wurde. Eine Untersuchung des deutschen und englischen Buchangebots der Jahre 1710 bis 1720*. Amsterdam u. Atlanta: Rodopi 2001 (= Internationale Forschungen zur allgemeinen und vergleichenden Literaturwissenschaft; 52).

<sup>6</sup> Einen Überblick über den Stand der Forschungsdiskussion bei English Showalter (Hg.): *Transformation du genre romanesque au XVIII<sup>e</sup> siècle*. In: *Eighteenth-Century Fiction* 13, 2–3 (2001).

<sup>7</sup> Besonders eindrucksvoll ist die Ausdehnung des zunächst auf das 18. Jahrhundert beschränkten *English Short Title Catalogue* (ESTC) auf die gesamte englischsprachige Druckwelt seit dem 15. Jahrhundert, so dass auch portugiesische Drucke, die im Bereich des englischen Empire gedruckt wurden, etwa in Indien, verzeichnet werden oder auch deutsche Drucke in Nordamerika, vgl. <<http://www.bl.uk/collections/early/holdingenglish.html>> (5.1.2006).

<sup>8</sup> Bibliographien wie etwa Ernst Weber u. Christine Mithal: *Deutsche Originalromane zwischen 1680 und 1780*. Berlin: Erich Schmidt 1983 oder auch die *Personalbibliographien zu den Drucken des Barock* von Gerhard Dünnhaupt (2. verb. Aufl. Stuttgart: Anton Hiersemann 1990–1993) erfassen nur einen Bruchteil der tatsächlich im deutschsprachigen Raum zirkulierenden Romanliteratur.

sche Projekte wie *American Memory* oder *English Short Title Catalogue* tun konnten.<sup>9</sup> Die kulturpolitischen Debatten um die universelle Online-Bibliothek und die mit ihr verbundenen Probleme, wer darüber entscheidet, welche Inhalte wie und unter welchen Bedingungen zugänglich gemacht werden, der nicht nur kulturpolitische Streit um die so genannte »Googleization« unseres kulturellen Wissens, hat längst die Politik auf den Plan gerufen.<sup>10</sup> Es geht dabei um nicht weniger als die Definitionsmacht über die Kultur, wie es der Direktor der Pariser *Bibliothèque nationale de France* Jean Noël Jeanneney prominent formuliert hat.<sup>11</sup>

Vor diesem Hintergrund ist das Forschungsprogramm einer computergestützten Buch- und Narratologiegeschichte des Romans in Deutschland (1500–1900) *Charikleia*<sup>12</sup> zu sehen. *Charikleia* ist selbst Teil eines größeren Langfristvorhabens mehrerer deutscher Universitäten und außeruniversitärer Forschungseinrichtungen, dem Projekt *DeutschDiachronDigital* (DDD).<sup>13</sup> Ziel des DDD-Vorhabens ist die Erstellung eines für die deutsche Sprach- und Literaturwissenschaft zentralen Korpus von den althochdeutschen und altsächsischen Anfängen bis zum älteren Neuhochdeutsch um 1900. Dieses Korpus wird wissenschaftlichen Ansprüchen auf Textgenauigkeit bzw. Diplomtizität und Historizität entsprechen müssen, außerdem allen Anforderungen an statistische Validität und große Erschließungstiefe der Textauszeichnungen. Gewährleistet sein muss dabei die Kompatibilität mit den international gängigen Standards (vor allem XML, TEI), wie sie andere Projekte bereits etabliert haben: das *Helsinki Corpus of English Texts* mit mehreren Millionen Wortformen seit dem Altenglischen,<sup>14</sup> erweitert durch das *Penn Parsed Corpus of Early Modern English*<sup>15</sup> oder das Textarchiv *Medieval Nordic Text Archiv* (*Menota*),<sup>16</sup> das *Hellenic National Corpus*,<sup>17</sup> das *Corpus de Referencia del Español Actual* (CREA),<sup>18</sup> die Korpora des Instituts für deutsche Sprache in Mannheim<sup>19</sup> oder das

---

<sup>9</sup> Besonders die in diesem Umfang einmalige retrospektive Digitalisierungspraxis von *Google Print* (<<http://www.print.google.com>> [5.1.2006]) und *Google Scholar* (<<http://www.scholar.google.com>> [5.1.2006]) bzw. *Microsoft MSN Book Search* (vorläufig noch unter <<http://www.archive.org>> [5.1.2006]) und *Yahoo (Open Content Alliance)* (<<http://www.opencontentalliance.org>> [5.1.2006]) stellen trotz der oft ungeklärten Rechtelage jetzt schon mehrere Millionen Aufsätze und Bücher, im wachsenden Maße auch andere Medien digital zur Verfügung. Vgl. das Interview von Wade Rush mit David Mandelbrot (Yahoo) *Digitize This. Yahoo hopes to trump Google with its Open Content Alliance publishing venture*. In: *Technology Review* (<[http://www.technologyreview.com/articles/05/10/wo/wo\\_102005roush.asp](http://www.technologyreview.com/articles/05/10/wo/wo_102005roush.asp)> vom 20.10.2005 [5.1.2006]).

<sup>10</sup> Etwa *i2010 Digital Library*-Projekt der Europäischen Union (<[http://europa.eu.int/information\\_society/activities/digital\\_libraries/index\\_en.htm](http://europa.eu.int/information_society/activities/digital_libraries/index_en.htm)> [5.1.2006]) oder die *European Library* (<<http://www.theeuropeanlibrary.org>> [5.1.2006]), in Deutschland das *Digital Library-forum* (<<http://www.dl-forum.de>> [5.1.2006]), *Vascoda* (<<http://www.vascoda.de>> [5.1.2006]) und *ZVDD* (<<http://www.zvdd.de>> [5.1.2006]) als Grundsteine einer nationalen digitalen Bibliothek, um nur ein paar der Initiativen und Projekte zu nennen.

<sup>11</sup> Jean Noël Jeanneney: *Quand Google défie l'Europe. Plaidoyer pour un sursaut*. Paris: Mille et une nuits 2005.

<sup>12</sup> *Charikleia. Deutsches Romanarchiv* <<http://www.charikleia.uni-goettingen.de>> (5.1.2006).

<sup>13</sup> *DeutschDiachronDigital*, <<http://www.deutschdiachrondigital.de>> (5.1.2006); gemeinsam mit dem Vorhaben *Deutsches Textarchiv* DTA <[www.bbaw.de/forschung/dta](http://www.bbaw.de/forschung/dta)> (5.1.2006) haben beide Initiativen den Zusammenschluss *Deutsches Diachrones Textarchiv* DDTA gebildet.

<sup>14</sup> <<http://khnt.hit.uib.no/icame/manuals/HC/INDEX.htm>> (5.1.2006).

<sup>15</sup> <<http://www.ling.upenn.edu/mideng>> (5.1.2006).

<sup>16</sup> <<http://gandalf.aksis.uib.no/menota>> (5.1.2006).

<sup>17</sup> <[http://www.ilsp.gr/hnc\\_eng.html](http://www.ilsp.gr/hnc_eng.html)> (5.1.2006).

<sup>18</sup> <<http://corpus.rae.es/creanet.html>> (5.1.2006).

<sup>19</sup> <<http://www.ids-mannheim.de/kt/corpora.html>> (5.1.2006).

*Digitale Wörterbuch der Deutschen Sprache* (DWDS).<sup>20</sup> DDD wird als digitales Korpus ein Forschungsinstrument von einer Art sein müssen, wie es keines der Massendigitalisierungen der großen Information Retrieval-Firmen sein kann. Seine Logik der Annotierung und seine Erschließungstiefe werden allein wissenschaftlichen Vorgaben folgen und im geglückten Fall selbst wiederum Modellcharakter für weitere wissenschaftliche Korpora haben.

Für *Charikleia* gelten die gleichen Anforderungen, andere kommen hinzu. Denn die Literaturwissenschaft hat noch nicht wie etwa die Korpuslinguistik etablierte Auszeichnungen entwickelt. Wie man narratologisch, buch- und kulturgeschichtlich Texte auszeichnet, wie ein historisch differenziertes Tagset zur Annotierung erzählerischer Texte aussieht, ist eine Forschungslücke der Literaturwissenschaft. Valide literaturwissenschaftliche Forschungsinstrumente für das digitale Zeitalter zu entwickeln, wird daher eine der vordringlichen Aufgaben des Forschungsprogramms *Charikleia* sein müssen. Sein Gegenstand sind die blauen Bibliotheken und die hohen Romane, sind Bachstrom und Goethe. Was heißt das im Detail, digitale Annotierungsverfahren und die Kulturgeschichte des Romanlesens zusammenzuführen?

### 1. *Deutsch Diachron Digital. Ein historisches Referenzkorpus für das Deutsche*

In Deutschland haben sich verschiedene Fachwissenschaften und Akademien, Kompetenzzentren und Digitalisierungszentren von Universitätsbibliotheken zusammengefunden, um die bislang verstreuten Unternehmungen zu einem Projekt eines historischen Referenzkorpus zusammenzuführen. Wie bei vergleichbaren Forschungsprojekten weltweit stehen auch hier grundsätzliche Entscheidungen vorne an, die die erheblichen Kosten eines solchen Unternehmens rechtfertigen müssen. Durch die retrospektive Digitalisierung in den letzten Jahren haben sich die Anforderungen an historische Korpora herauskristallisiert. Dazu gehören: die weitgehend fehlerfreie Volltextfassung, die komplexe Erhebung von Metadaten, die Aufnahme mehrerer, im Idealfall aller relevanten Textfassungen sowie die Alignierung von Volltext und digitalen Faksimiles<sup>21</sup>. Diplomazität ist eine selbstverständlich gewordene Forderung. Sprach-, literatur- und kulturwissenschaftliche, aber auch informationstechnologische Auswertungen und Erschließungsroutinen zählen zu den Anforderungen, die an wissenschaftliche Korpora gestellt werden müssen. Weiterreichende multidirektionale Verknüpfungen mit Kontextwissen, etwa historischen Wörterbüchern, sind möglich und bieten einen erheblichen Mehrwert. Etabliert sind auch flache Annotationen, die bei modernen Sprachstufen automatisiert an einzelnen Wortformen (Tokens) hängen. Der öffentliche Zugang und die virtuelle Zusammenführung getrennter, oft auch nur noch kaum zugänglicher Ressourcen im Internet erschließen in einer solchen Zusammenstellung bereits neue Blickwinkel, neue Zusammenhänge und ermöglichen innovative Fragestellungen. Entscheidend für den Wert, den ein digitales Korpus für die Forschung besitzt, ist jedoch seine Erschließung, die Kontextualisierung der Materialien. An diese Standards knüpft das Pro-

<sup>20</sup> <<http://www.dwds.de>> (5.1.2006).

<sup>21</sup> Vgl. zum Themenkreis Digitale Faksimiles als Forschungsinstrument Stephan Füssel: *Vom Steindruck zum Internet. Faksimiles im Medienumbruch*. In: Sven Hanschek: *Die Struktur medialer Revolutionen. Festschrift für Georg Jäger*. Frankfurt u. a.: Lang 2000, S. 134–142; Andrea Rapp: *Digitale Faksimiles*. In: *Aus dem Antiquariat* 5 (2004), S. 361–366; vgl. zum wissenschaftlichen elektronischen Publizieren auch Carlos J. Alonso u. a.: *Crises and Opportunities. The Futures of Scholarly Publishing. ACLS Occasional Paper*, No. 57; insbesondere den Beitrag von John M. Unsworth S. 37–44 <<http://www.acls.org/op57.pdf>> (5.1.2006).

jekt eines digitalen historischen Referenzkorpus an, weicht aber auch an einigen Stellen begründet davon ab.

Zunächst zur Auswahl der Texte. Eine digitale Bibliothek ist dann ein Korpus, wenn sie Texte nach definierten philologischen Kriterien auswählt und einheitlich mit weiteren Angaben versieht. Das Referenzkorpus des Deutschen soll von den ältesten überlieferten Sprachstufen bis zum Neuhochdeutschen reichen. Während für die älteren Sprachstufen des Deutschen Vollständigkeit (Althochdeutsch) bzw. eine breite Repräsentativität (Mittelhochdeutsch) gefordert werden muss, können für das Neuhochdeutsche nur mehr exemplarische Teilkorpora erfasst werden. Diese Teilkorpora sind textsortenbestimmt und bilden diese Textsorte in einer historisch repräsentativen Auswahl ab. Gewählt wurden drei Bereiche, die möglichst systematisch verschiedene Felder der Sprachverwendung abdecken: Zeitungstexte (öffentliche Sprache), Briefe (private Sprache), Romane (künstlerische Sprache). Die Möglichkeiten einer dynamischen Erweiterung der Korpora sind von Anfang an im Gesamtplan der Korpusarchitektur, vor allem im Hinblick auf die technologischen Grundlagen, mit berücksichtigt. Wenn dabei »deutsch« klein geschrieben wird, dann auch deshalb, weil die lateinischen Interlineartexte zum Althochdeutschen oder auch Textzeugen in Jiddisch zu dem Korpus hinzugezählt werden. Erarbeitet wird das Korpus in getrennten Sprachstufenprojekten für das Althochdeutsche, das Mittelhochdeutsche, das Niederdeutsche, das frühe Neuhochdeutsche und das Neuhochdeutsche. Übergreifende Architekturprojekte gewährleisten die Einheitlichkeit der Erfassung der Metadaten und die Standards der Annotierung.

### *1.1. Struktur des Gesamtkorpus DDD und seine Architektur<sup>22</sup>*

Bislang gibt es kaum umfangreiche Korpora, an die so flexible Anforderungen gestellt werden, wie es das DDD-Konzept vorsieht. Die Summe der Anforderungen macht die Entwicklung einer ganz neuen Korpusarchitektur notwendig.

Ausgangspunkt ist das Original. Jede auch noch so exakte (analoge oder digitale) Reproduktion des Originals beinhaltet nicht nur ein gewisses Maß an »Verfremdung«, sondern mehr noch, sie beinhaltet auch immer eine Interpretation. Dies hat zwei Konsequenzen für die Konzeption des Korpus: 1. größtmögliche Originalnähe, 2. eine klare Trennung von »Basisdaten« und Metadaten. Weitere Forschungen mit anderen Schwerpunkten können das möglichst originalnahe Digitalisat und die diesem eindeutig zugeordneten, aber auch technisch klar getrennten erschließenden Metadaten nach Bedarf nachnutzen. Interoperabilität ist nur dann dauerhaft gewährleistet, wenn flexible Anforderungen im Hinblick auf die Datenmenge, die unterschiedlichen Annotationstiefen, die Multimodalität, die Multilingualität, die Anbindung von externen Ressourcen sowie effiziente Suchmöglichkeiten auf allen Ebenen sichergestellt sind. Das macht für das digitale Referenzkorpus die Entwicklung einer eigenen technologischen Korpusarchitektur notwendig. Im Unterschied zu herkömmlichen Textkorpora erfordern multimodale Korpora eine so genannte Stand-off-Architektur, in der jeder Datentyp (einschließlich der verschiedenen Metadaten) in einer eigenen unabhängigen Datei gespeichert ist. Die einzelnen Dateien beziehen sich alle auf eine Referenzdatei und können auf diese Weise

---

<sup>22</sup> Vgl. zum Folgenden Anke Lüdeling, Thorwald Poschenrieder und Lukas Faulstich: *Deutsch Diachron Digital – Ein diachrones Korpus des Deutschen*. In: *Jahrbuch für Computerphilologie* 6 (2004), S. 119–136 und <<http://computerphilologie.uni-muenchen.de/jg04/luedeling/ddd.html>> (5.1.2006).

exakt aligniert werden. Hierfür werden auch Modelle und Konzepte aus den Naturwissenschaften, insbesondere aus der Molekularbiologie bzw. der Bioinformatik auf ihre Übertragbarkeit überprüft. Für jede Ebene wird ein spezifisches, standardisiertes, für alle Gruppen verbindliches Tagset entwickelt.<sup>23</sup> Diese Tagsets werden entweder von bereits bestehenden Standards, vor allem dem Auszeichnungssystem der *Text Encoding Initiative* (TEI), übernommen,<sup>24</sup> teilweise aber auch neu entwickelt. Ein zu lösendes Problem und damit ein Forschungsauftrag bleibt dabei jedoch die hierarchisch verschachtelte Referenzierung von Annotationen.

### 1.2. Digitalisierung

Die Texte des DDD sollen durchgehend als elektronische Volltexte und weitgehend mit ergänzenden Faksimiles verfügbar gemacht werden. Ziel der Faksimilierung ist es, die haptische Qualität verschrifteter Sprache sinnfällig zu machen. Daher sollten möglichst in einem ersten Schritt hochqualitative Farbscans hergestellt werden, deren größtmögliche Qualität und Originalnähe ferner hilft, Verfremdungseffekte zu vermeiden.<sup>25</sup> Diese digitalen Abbilder können auch als Vorlagen für die Volltexterfassung dienen und damit evtl. empfindliche Originale schonen helfen. Als Format wird ein standardisiertes TIFF-Format mit Metainformationen im Header jedes Images gewählt (bibliographische, organisatorische und technische Metadaten). Diese Metainformationen sind notwendig zur dauerhaften Identifizierung jedes Images und damit auch eine wichtige Voraussetzung für die Langfristarchivierung.

Alle weiteren sprach- und literaturgeschichtlichen Auswertungen lassen sich nur mit digitalen Volltexten angehen. Verlangt wird eine weitestgehend fehlerfreie Volltextversion, sonst sind zuverlässige sprachwissenschaftliche Analysen nicht zu realisieren. Eine OCR-Erfassung scheidet aus verschiedenen Gründen aus: Vor allem die Heterogenität der Vorlagen bereitet der automatischen Erkennung Schwierigkeiten. Dazu kommen teilweise schlechte Druckqualität, wechselnde Papierqualitäten sowie vor allem häufig wechselnde Schriftarten auch innerhalb eines Werkes (verschiedene Fraktur- und Antiquaschriften) sowie ein komplexes Layout, das weit heterogener ist als heutige Fließtexte. Das Verfahren der Wahl ist daher das *double keying*-Verfahren, bei dem der Text zweimal unabhängig voneinander eingetippt, anschließend automatisch verglichen und auf dieser Basis korrigiert wird. Ergebnis ist ein Text, der zu mindestens 99,997% korrekt ist, d. h. maximal 3 Fehler auf 100.000 Zeichen enthält. Während für die Texteingabe insbesondere historischer Texte die muttersprachliche Interferenz eine Fehlerquelle darstellt und daher Nicht-Muttersprachler für die Texterfassung ideal sind, müssen für den Fehlerabgleich Muttersprachler, zum Teil auch Fachwissenschaftler eingesetzt werden.

### 1.3. Annotierung bzw. inhaltlich-strukturelles Markup

Zu den Grundprinzipien für den Einsatz aller Technologien im Rahmen des Gesamtprojektes gehört, dass konsequent auf ›offene Standards‹ gesetzt wird, das heißt auf stan-

---

<sup>23</sup> Das bedeutet nicht, dass alle Teilgruppen alle Annotierungsebenen beachten müssen, sondern dass sie den für die jeweilige Annotierungsebene verbindlichen Standard verwenden müssen.

<sup>24</sup> Zur TEI vgl. die Website <<http://www.tei-c.org/>> sowie Fotis Jannidis: *Wider das Altern elektronischer Texte. Philologische Textauszeichnung mit TEI*. In: *editio* 11 (1997), S. 152–177.

<sup>25</sup> Durch den Medienwechsel vom Buch zum digitalen Medium ist ferner ausreichend kritische Distanz gegeben, die z. B. bei gedruckten Faksimiles eher verwischt wird.

dardisierte Methoden und Konzepte, die herstellerübergreifend bzw. plattformunabhängig formuliert und angewendet werden. Das gilt für die Bilddateien der Faksimiles, die als TIFF-Dateien gespeichert werden sollen. Auch sämtliche Basis- und Metadaten werden mit XML<sup>26</sup> nach den Richtlinien der *Text Encoding Initiative* (TEI) kodiert. Sind noch keine Standards etabliert, wie etwa für ein buchgeschichtliches und narratologisches Tagging, wird eine neue DTD entwickelt, die möglichst in die Richtlinien der TEI aufgenommen werden soll. Die Rückkopplung mit TEI gehört zu den notwendigen institutionellen Vernetzungen des Forschungsvorhabens. Alle Texte sollen ein einfaches strukturelles Markup erhalten, das Texteinheiten wie Buch, Brief, Kapitel, Seiten, Absätze usw. festhält. Hinzu kommt eine durchgehende Lemmatisierung der Texte, die eine einfache einheitliche Suche über die durch den Sprachwandel sehr heterogenen Textbestände ermöglichen wird. Für linguistische Analysen wird ein Kernbereich mit einem sehr aufwändigen morphosyntaktischen Tagging versehen werden.

#### 1.4. Publikation

Für alle Materialien gilt das Prinzip des freien Zugangs im Sinne der *Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen* vom 22.10.2003.<sup>27</sup> So genannte offene Schnittstellen ermöglichen auch eine dezentrale Vernetzung von und mit externen Ressourcen. Erst die wissenschaftliche Erschließung und damit die Kontextualisierung der Materialien führen dazu, dass solche Ressourcen auch tatsächlich breite Akzeptanz in der *scientific community* finden. Die technologische Offenheit der Materialien ermöglicht es ferner, multidirektionale Verweise zu anderen Ressourcen zu generieren und damit vernetzte Wissensstrukturen zu schaffen. Die Vorteile solcher Erschließungsstrategien liegen auf der Hand: Werden zum Beispiel die Romane auch mit digitalen Wörterbüchern wie dem *Digitalen Grimm*<sup>28</sup> vernetzt, eröffnen sich ganz neue Nutzungs- und Auswertungsmöglichkeiten. Um nur zwei Beispiele zu nennen: Durch die Verbindung von historischem Text und Wörterbuch sind die Materialien für den akademischen, aber auch schulischen Unterricht bestens präpariert. Aus der anderen Perspektive werden Wörterbuchartikel durch die direkte Verknüpfung mit den Primärquellen transparent bzw. lassen sich durch neue Belege dynamisch ergänzen.

#### 2. Charikleia. Deutsches Roman-Archiv

Die Textsorte Roman repräsentiert die Leitgattung unter den literarisch-fiktionalen Textsorten der Neuzeit. An ihr lassen sich zentrale sprach- und literaturgeschichtliche, erzähltechnische, lese- und buchgeschichtliche Wandlungen verfolgen. Diese umfassende Zielsetzung bestimmt damit nicht nur die interdisziplinäre Herangehensweise des Forschungsprogramms, sondern auch die Überlegungen zur Konzeption des Korpus, das zugrunde gelegt wird, wie auch die technologischen Aspekte des Vorhabens.

Eine Erforschung der Entwicklungsgeschichte des Romans von den Anfängen bis in die Neuzeit kann nicht losgelöst von dem Medium geschehen, in dem diese Textsorte sich durchgesetzt hat, dem Buch. Romangeschichte und Buchgeschichte gehören zusammen. Daher kann sich das Vorhaben nicht auf die reine Textfassung beschränken, sondern muss auch die Originale, das heißt in der Regel die Erstdrucke, aber gebe-

---

<sup>26</sup> *Extensible Markup Language*, vgl. <<http://www.w3.org/TR/REC-xml>> (5.1.2006).

<sup>27</sup> <[http://www.mpg.de/pdf/openaccess/BerlinDeclaration\\_dt.pdf](http://www.mpg.de/pdf/openaccess/BerlinDeclaration_dt.pdf)> (5.1.2006).

<sup>28</sup> <<http://www.dwb.uni-trier.de>> (5.1.2006).

nenfalls auch spätere Auflagen als Faksimiles abbilden, als optisches Faksimile wie als Metadaten. Die Verbindung von digitalem Faksimile (Image) und digitalem Volltext mit buchgeschichtlichen Metadaten versucht, die Texte in einem Format zugänglich zu machen, das für viele gegenwärtige, auch viele zukünftige Forschungsprogramme und -ansätze brauchbar ist.

### 2.1. Korpusbildung

Anders als in der angelsächsischen Forschungstradition, in der Arbeiten zur *rise of the novel* ein nicht zuletzt auch kulturpolitisch zentrales Forschungsfeld ausmachen, fehlt in Deutschland eine damit vergleichbare Forschung. Hier dominiert – mit Ausnahmen<sup>29</sup> – die Konzentration auf wenige, hochkanonische Romane vor allem der Goethe-Zeit. Eine annähernd vollständige bibliographische Erfassung der kulturhistorisch relevanten Romanliteratur<sup>30</sup> einschließlich der Übersetzungen und Bearbeitungen fehlt,<sup>31</sup> eine nach internationalen Standards annotierte Digitalisierung liegt nur mit wenigen Ausnahmen vor.<sup>32</sup> Das Arbeitsprogramm des Teilprojekts *Charikleia* wird diese Lücken schließen und damit beginnen, etwa 350 Romane zwischen 1500 und 1900 als Images und Volltext zu digitalisieren und zu annotieren. Die Zahl von 350 Romanen ist nicht eben hoch angesichts der ungleich höheren Zahlen der gedruckten und publizierten Romane nicht erst im 19. Jahrhundert. Daher ist die möglichst vollständige bibliographische Ermittlung der tatsächlich seit 1500 im deutschsprachigen Raum zirkulierenden Romanliteratur für das Forschungsprogramm ein zentrales Vorhaben. Erst sie erlaubt eine Einschätzung, welche Romane überhaupt gelesen wurden, was in welchen Sprachen in Deutschland gedruckt und vertrieben wurde, was wie oft Neuauflagen erlebte und was, obgleich angekündigt, nie erschienen ist. Man sieht dann schnell, dass der literaturhistorische Blick auf wenige kanonische Texte beschränkt ist, aber weder ein Roman wie der Bachstroms, noch jene »Perlen« noch die »neuesten und besten Romane aller Nati-

<sup>29</sup> Simons: *Marteaus Europa* und sein Projekt *Pierre Marteau Publishing House* unter <<http://www.pierre-marteau.com>> (5.1.2006).

<sup>30</sup> Vgl. Anm. 8.

<sup>31</sup> Übersetzungen sind seit dem 18. Jahrhundert bibliographisch meist nach Landessprachen getrennt erfasst, etwa in Heynes Bandkatalog der Staats- und Universitätsbibliothek Göttingen. Neuere Bibliographien, die das Ausmaß der tatsächlichen oder nur fingierten Übersetzungen erkennen lassen, sind etwa für das Englische: Wilhelm Graeber u. Geneviève Roche: *Englische Literatur des 17. und 18. Jahrhunderts in französischer Übersetzung und deutscher Weiterübersetzung. Eine kommentierte Bibliographie*. Hg. v. Jürgen Stackelberg. Tübingen: Niemeyer 1988 oder aus dem Französischen: *Bibliographie deutscher Übersetzungen aus dem Französischen 1700–1948*. Bearb. von Hans Fromm. 6 Bde. Baden-Baden: Verl. für Kunst und Wissenschaft 1950–1953, oder aus dem Italienischen: Frank-Rutger Hausmann u. Volker Kapp (Hg.): *Bibliographie der deutschen Übersetzungen aus dem Italienischen von den Anfängen bis zur Gegenwart*. Tübingen: Niemeyer 1992–2004 und Alberto Martino: *Die italienische Literatur im deutschen Sprachraum. Ergänzungen und Berichtigungen zu Frank-Rutger Hausmanns Bibliographie*. Amsterdam u.a.: Rodopi 1994 (= Chloe; 17), oder für die antike Literatur: Franz-Josef Worstbrock: *Deutsche Antikenrezeption 1450–1550*. Mit einer Bibliographie der Übersetzer. Boppard: Boldt 1976 um nur ein paar Beispiele zu nennen.

<sup>32</sup> Von mehreren kanonisierten Romanen liegen elektronische Texte vor, die etablierten Studienausgaben in der Textgestalt folgen; vgl. Mathias Bertram: *Deutsche Literatur von Luther bis Tucholsky. Großbibliothek*. Digitale Bibliothek, Bd. 125. Berlin: Directmedia Publishing 2005. Allerdings sind diese Texte als kommerzielle Produkte mit ihrer Bindung an eine bestimmte Software nicht langfristig verfügbar. Ein anderes Beispiel bietet die *Römische Octavia* von Herzog Anton Ulrich, die in einer per *double keying* erstellten Volltextfassung vorliegt, die jedoch vor allem zum Zweck der Satzerstellung erfasst und nicht systematisch ausgezeichnet wurde.

onen«, die man im 19. Jahrhundert wert befand, zu drucken, berücksichtigt sind.<sup>33</sup> Der Aufwand für diese bibliographische Arbeit ist nicht gerade gering, was noch dadurch erschwert wird, dass »Romane« eine unscharfe, weil historisch variierende und poetologisch nur gering abgestützte bibliographische Klassifikation darstellt. Oft sind »Roman« bzw. »fabulae romanensis« nur eine Sammelbezeichnung für die vielen Bücher, die den Bibliothekaren vergangener Jahrhunderte in keine Klassifikation zu passen schienen. Auch lassen sie sich von Novellen und »kleinen Romanen«<sup>34</sup> nicht so absetzen, wie wir das heute zu tun gewohnt sind. Vieles lässt sich daher nur durch Autopsie und Querweise unterschiedlicher Kataloge erfassen. Ausgesprochen hilfreich für die Rekonstruktion der versunkenen Lesewelten sind dabei auch historische Messkataloge und Auktionskataloge.<sup>35</sup> In den Auktionskatalogen etwa sind jene Romane aufgelistet, die man anders als andere Gattungen gerne bei Bibliotheksaufösungen verkauft hat. Die für das 18. Jahrhundert so repräsentative Bibliothek Johann Wilhelm Ludwig Gleims beispielsweise wurde nach seinem Tod 1803 selbstverständlich zusammen belassen. Verkauft wurde ebenso selbstverständlich seine Sammlung der Romane, wie der Auktionskatalog im Detail ausweist.<sup>36</sup> Nicht weniger hilfreich sind auch die Kataloge adliger Ritterroman-Sammlungen wie etwa in Kassel. Erst in der Kombination verschiedener bibliographischer Instrumente lässt sich rekonstruieren, wie die Kulturgeschichte des Lesens und des Buchdrucks der Romane ausgesehen hat und erst auf dieser Grundlage kann dann der Anspruch auf die Erstellung eines kulturgeschichtlichen Archivs der literarischen Leitgattung der Neuzeit überhaupt erhoben und entsprechend valide Forschungsmöglichkeiten eröffnet werden. Um daher den Anspruch des Projektes an Repräsentativität der Auswahl zu stützen, wird daher aus vorhandenen Bibliographien, der Forschungsliteratur, vor allem aber aus Messkatalogen, historischen Buchverzeichnissen und durch Autopsie verschiedener historischer Bibliotheken die bislang umfassendste Bibliographie der Romanliteratur erarbeitet. Sie erschließt die Originalromane ebenso wie Übersetzungen, Bearbeitungen und fremdsprachige Romane, die im Alten Reich und im Deutschland des 19. Jahrhunderts zirkulierten. Nicht die Originalromane, sondern die Übersetzungen, Bearbeitungen und fremdsprachigen Romane sind es gewesen, die besonders im 17. und 18. Jahrhundert den größeren Teil der kulturgeschichtlich relevanten Romanliteratur ausgemacht, aber nicht den Weg in den literaturhistorischen Kanon gefunden haben. Die Bibliographie soll selbständig publiziert und im Internet für Forschungszwecke zur Verfügung gestellt werden.

Ausgangspunkt der retrospektiven Digitalisierung ist zunächst ein Kanon von knapp 150 Romantexten. Die Ermittlung dieses Kanons folgt den etablierten romangeschichtlichen Darstellungen. Auch wenn man über den einen oder anderen Text streiten mag,

---

<sup>33</sup> Perlen. *Journal für gute deutsche Original-Novellen nebst den besten Übersetzungen der berühmtesten Schriftsteller des Auslandes*. Berlin: Sacco 1846–1865; *Belletristisches Lese-Cabinet der neuesten und besten Romane aller Nationen in sorgfältigen Übersetzungen*. Wien, Leipzig u. a.: Hartleben 1850–1853.

<sup>34</sup> Friedrich Schulz: *Kleine Romane*. 5 Bde. Leipzig: Göschen 1788–90.

<sup>35</sup> Dabei ist zu bedenken, dass rund ein Viertel der Ankündigungen in den Buchmesskatalogen des 18. Jahrhunderts nie gedruckt worden sind; vgl. Erich Schön: *Gibt es Literatur ohne Leser*. In: Simone Winko, Fotis Jannidis u. Gerhard Lauer (Hg.): *Literatur, Literaturbegriff, literarische Kommunikation*. Berlin: de Gruyter 2007 (in Vorb.).

<sup>36</sup> *Verzeichniß eines Theils der Kupferstich- und Bücher-Sammlung Johann Wilhelm Ludewig Gleim's : welcher den 22sten October d.J. u. folgende Tage, in d. gewöhnlichen Nachmittagsstunden, im sogenannten großen Judenhause öffentl. versteigert werden soll; nebst einigen Anhängen*. Repr. d. Orig.-Ausg. 1804. Mit einem Nachw. von Reinhard Selz. München u. a.: Saur 1987.

so zeigen die literaturhistorischen Darstellungen seit dem 19. Jahrhundert eine hinreichend große Konstanz, um diesen Kanon ermitteln zu können. Über die Bedeutung des »Werther« für die deutsche Romangeschichte muss man nicht lange diskutieren. Diese Romane als Kernkorpus in einem Schritt zu digitalisieren, ist aber auch deshalb unumgänglich, um dem Projekt den Anschluss an die etablierte Romanforschung zu sichern. Von diesem Korpus ausgehend wird dann in einem zweiten Schritt auf der Grundlage der Bibliographie ein kulturgeschichtlich relevantes Korpus von etwa 200 deutschsprachigen Romanen seit 1500 zusammengestellt und digitalisiert. Hierbei spielen auch die Kriterien der Auflagenhöhe, der verschiedenen Untertextsorten und der sozialen, regionalen und konfessionellen Verbreitung eine Rolle.

Diesem kulturgeschichtlichen Anspruch entsprechend soll das Romankorpus auch nicht aus deutschen Originalromanen, sondern aus Romanen in deutscher Sprache bestehen. Wirkungsmächtige Übersetzungen und Bearbeitungen werden einbezogen. Es wird eine daran anschließende Aufgabe sein, zusammen mit den jeweiligen Fremdsprachenphilologien, auch die fremdsprachigen Texte, die in den deutschsprachigen Ländern gelesen wurden, zu erfassen. Das kann aus letztlich pragmatischen Gründen erst ein dritter Schritt sein, denn die schiere Masse der in den deutschsprachigen Territorien distribuierten fremdsprachigen Romane vor allem vor 1730 erzwingt diese Entscheidung ebenso wie die spezifische romanistische, anglistische und andere Kompetenz, die für die Annotierung der fremdsprachigen Romane erforderlich ist. Gleichwohl sind gerade diese Romane für die Lesegeschichte des Romans in Deutschland von zentraler kulturhistorischer Bedeutung. Daher soll und muss das Projekt *Charikleia* in späteren Stufen in möglichst internationaler Kooperation um Digitalisate fremdsprachiger Texte erweitert werden. Alignierungen mit bereits digitalisierten deutschsprachigen Bearbeitungen und Adaptionen wären dann auch möglich und würden die Filiationen des europäischen Romanmarktes auf ganz neue Weise sichtbar machen.

Der Digitalisierung zugrunde gelegt wird im Regelfall der Erstdruck. Das Prinzip ist im Einzelfall immer wieder zu prüfen, da nicht für jeden Text gilt, dass der Erstdruck der tatsächlich gelesene Text war. So wird man etwa bei den Romanen aus der zweiten Hälfte des 19. Jahrhunderts, wo der Erstdruck zumeist in einer Zeitschrift erfolgte, dem aber dann sofort die Buchausgabe folgte, wohl in der Regel auf die Buchausgabe zurückgreifen, da diese häufig noch Verbesserungen des Autors enthält, die Texte in dieser Form langfristig wirksam blieben und die Ausrichtung unseres Korpus auf das Medium Buch nicht systematisch durchbrochen wird. Umgekehrt sind nicht wenige Romane auch in Roman-Journalen veröffentlicht worden, die eine größere Leserschaft erreicht haben als etwa Einzel- oder Werk-Ausgaben. Nicht ganz unproblematisch ist dieses Prinzip in den Fällen, in denen mit dem Autormanuskript ein besserer Text vorliegt – »besser« aus der Sicht der Autorenphilologie. Hier wird abzuwägen sein zwischen dem Anspruch des Korpus, die öffentliche Verwendung der deutschen Sprache zu dokumentieren, und dem literaturwissenschaftlichen Anspruch, einen möglichst autornahen Text zu konstituieren. Das Erstdruckprinzip hat einen aufwändigen Erstellungsprozess zur Folge, weil die Quellentexte aus verstreuten Bibliotheken für die Digitalisierungszentren ausgeliehen werden müssen. Editorisches Prinzip ist aber auch hier, möglichst genau die Romantexte in ihrem historischen Umfeld aufzusuchen und diese kulturhistorischen Rahmendaten ebenfalls als medienspezifisches Markup auszuzeichnen.

## 2.2. Medienspezifisches Markup

Wie bereits erwähnt, wird das Korpus auf mehreren Ebenen annotiert werden. Zum einen wird das Romankorpus wie das gesamte Korpus der digitalen Bibliothek die basalen Strukturinformationen, zum Beispiel Kapitel, Seite, Absatz, Inhaltsverzeichnis, auszeichnen. Außerdem werden auch alle Romane lemmatisiert, nur ein Teil aber wird morphosyntaktisch annotiert, um etwa Vergleiche zwischen literarischer und nichtliterarischer Sprache zu ermöglichen. Hinzu kommen zwei korpuspezifische Auszeichnungsperspektiven:

1. buchgeschichtliche Informationen, 2. erzähltheoretische Aspekte.

Die Materialität literarischer Texte, ja von Texten überhaupt, ist in vielen Disziplinen lange Zeit vernachlässigt worden. Sie gingen von einem idealen Text aus, dessen materiale Grundlage für die Analyse als belanglos erachtet wurde. Das gilt z. B. für die Linguistik, aber auch für die Literaturwissenschaft. Andere Formen historischer Wissensverwaltung waren sich dagegen der Relevanz dieses Wissens immer bewusst. So gibt es im Antiquariatshandel ausgesprochen detaillierte Beschreibungen von Büchern und auch die Bibliotheken haben in ihrer Katalogisierung materiale Aspekte des Buchs mit erfasst. Hinzu kommen Unterschiede in den nationalen Philologien. Die angelsächsische Editionswissenschaft hat, nicht zuletzt aufgrund der Besonderheiten der Shakespeare-Überlieferung, ein sehr differenziertes Beschreibungsinventar für die Erfassung von buchhistorischen Besonderheiten entwickelt.<sup>37</sup> Wenn auch in der deutschsprachigen Literaturwissenschaft unter dem Stichwort »Materialität der Kommunikation«<sup>38</sup> diesem Bereich inzwischen eine größere Aufmerksamkeit geschenkt wird, so fehlt es hier doch noch an integrativen systematisch-historischen Studien. Ohne Zweifel lassen sich die unterschiedlichen Perspektiven der diversen Disziplinen auf die Materialität der Kommunikation auch nur mit Schwierigkeiten aufeinander abbilden. Die Editionsphilologie ist vor allem daran interessiert, diese Informationen auszuwerten, um einen besonders guten Text zu erstellen. Bibliotheken beschreiben die Aspekte, die unter archivarischen Gesichtspunkten relevant sind. Ein kulturhistorisches Verständnis dieses Aspekts wird sich jedoch gerade für die (schwachen) Zeichen interessieren, die in ständig veränderlichen Konstellationen gebildet werden. So ist es, um ein willkürliches Beispiel zu nennen, sicherlich eine Besonderheit der letzten Jahrzehnte, dass kleine, nicht sonderlich großzügig gedruckte, gelb eingeschlagene Bände mit klassischer Literatur und damit mit Schule und Mühen verbunden werden.

Aufgabe des Tagsets zur buchgeschichtlichen Erschließung ist die differenzierte Beschreibung der Materialität des Buches, allerdings nicht erschöpfend, sondern in einer pragmatisch handhabbaren und kulturhistorisch relevanten Auswahl. Zu jedem Roman wird ein standardisierter Header aufgebaut, der neben den üblichen bibliographischen Angaben beispielsweise Informationen zur Größe des Buches und zur Qualität des Papiers und zur Art des Druckbildes enthält und nach diesen Kriterien – verknüpft mit al-

---

<sup>37</sup> Vgl. die Standardwerke: Fredson Bowers: *Principles of Bibliographical Description*. Princeton: Princeton University Press 1949 sowie Philip Gaskell: *A New Introduction to Bibliography*. Oxford: Oxford University Press 1972. In deutscher Sprache gibt es kaum Vergleichbares außer Martin Boghardt: *Analytische Druckforschung. Ein methodischer Beitrag zu Buchkunde und Textkritik*. Hamburg: Hauswedell 1977. Die Einzelforschung insbesondere zur Frühzeit des Buchdrucks ist auch in Deutschland sehr ausgeprägt, aber es fehlt noch an systematischen Überblicken vor allem zum 19. Jahrhundert.

<sup>38</sup> Vgl. zu der Formulierung Hans U. Gumbrecht u. K. Ludwig Pfeiffer (Hg.): *Materialität der Kommunikation*. Frankfurt: Suhrkamp 1988 (= Suhrkamp-Taschenbuch Wissenschaft; 750).

len übrigen Ebenen der Annotierung – recherchierbar ist. *Charikleia* wird sich hierbei auf die buchanalytischen Forschungen stützen, die vor allem in den angelsächsischen Ländern eine lange Tradition haben, diese aber durch weitere historische Arbeiten ergänzen.<sup>39</sup> Zurzeit arbeitet eine Arbeitsgruppe der TEI daran, ein Tagset zu entwickeln, das es ermöglicht, Texte entsprechend auszuzeichnen.<sup>40</sup>

Schon aus ganz praktischen Gründen wird man nicht 350 Romane in der Intensität und Genauigkeit beschreiben können, wie das etwa mit einem Foliodruck von Shakespeares Dramen geschieht. Ein erster Schritt wird daher sein, festzulegen, welche buchhistorischen Informationen voraussehbar von allgemeinerem Interesse sind und in einem überschaubaren Zeitrahmen erhoben werden können. Gedacht ist zurzeit an Informationen, die in einem erweiterten Header notiert werden würden: Buchgröße, Drucktype, Illustrationen, Seitenlayout. Hinzu kommen seitenspezifische Informationen zu illuminierten Buchstaben, Kopfzeilen usw.

Das Tagset zur Erfassung von erzähltheoretischen Aspekten kann sich zwar auf eine umfangreiche narratologische Forschung stützen,<sup>41</sup> wird aber wohl der erste Versuch sein, mehrere Kategorien systematisch und mit technischen Mitteln auf ein größeres historisches Korpus anzuwenden, um damit zur Verwissenschaftlichung der Literaturgeschichtsschreibung beizutragen.<sup>42</sup> Auch hier ist das Hauptproblem die Auswahl geeigneter Kategorien. Einerseits sollen die Kategorien sich genau genug spezifizieren lassen, um intersubjektiv verbindlich, schnell einsichtig und leicht erlernbar zu sein, da die konkrete Textauszeichnung ja von einer größeren Arbeitsgruppe durchgeführt werden soll und nicht zuletzt auch weil die Benutzer der Edition das Markup möglichst schnell verstehen und plausibel finden sollen. Andererseits sollen die entsprechend kodierten Phänomene auch von wissenschaftlichem Interesse sein. Unproblematisch sind wohl Informationen zu Zeitwechsel, direkter und indirekter Rede, Erzählerwechsel. Weniger einfach, wenn auch besonders interessant, wäre die Auszeichnung von erlebter Rede.

Um für diese beiden Aspekte den richtigen Mittelweg zwischen Handhabbarkeit und Komplexität zu finden, haben die Herausgeber ein überschaubares Probekorpus konzipiert, das zur Zeit erstellt wird und an dem diese Probleme durchgespielt werden.<sup>43</sup> Sobald das Korpus erstellt ist und einige Erfahrungen im Umgang mit den Auszeichnungsaspekten gewonnen wurden, sollen zwei Arbeitsgespräche mit Buchhistorikern und Narratologen dazu beitragen, die Ermittlung dieses Mittelwegs möglichst früh auch mit der Gemeinschaft der interessierten Forscher abzustimmen.

<sup>39</sup> Für das Projekt wird außerdem im Rahmen eines Promotionsprojekts in Darmstadt untersucht, inwieweit diese auf die angelsächsische Welt bezogenen Konzepte für die Situation in Deutschland im 18. und 19. Jahrhundert übertragbar sind.

<sup>40</sup> Vgl. <<http://www.tei-c.org/Activities/PB/>> (8.2.2005).

<sup>41</sup> Insbesondere Genettes Entwurf stellt so etwas wie eine *lingua franca* der Narratologie dar; vgl. Gérard Genette: *Die Erzählung*. München: Fink 1994 (= UTB für Wissenschaft).

<sup>42</sup> Anknüpfen kann man dabei natürlich an die Arbeiten zur computergestützten Stilistik und Inhaltsanalyse; vgl. dazu und zum Stand der Computerphilologie insgesamt jetzt Susan Schreibman, Ray Siemens u. John Unsworth (Hg.): *A Companion to Digital Humanities*. Malden: Blackwell 2004.

<sup>43</sup> Das Probekorpus besteht aus Erzählungen, entstanden zwischen 1700 und 1900. Die Texte liegen inzwischen als Faksimile und als digitalisierte Volltexte vor. In diese Volltexte werden die basalen strukturellen Informationen einschließlich der Verlinkung von Text und Bild sowie einige darüber hinausgehende publikationshistorische Aspekte eintragen.

### *2.3. Perspektiven für Forschung, Lehre und die allgemeine Öffentlichkeit*

Mit dem Korpus, das im Rahmen des Projekts DDD erstellt werden soll, können für das Deutsche erstmals sprachwissenschaftliche Analysen auf allen sprachsystematischen Ebenen auf repräsentativer Basis in diachroner Langzeitperspektive von den Anfängen der schriftlichen Überlieferung bis zum Jahr 1900 bzw. bei Anbindung an moderne Korpora<sup>44</sup> bis in die Gegenwart durchgeführt werden.

Literaturwissenschaftliche Untersuchungen erhalten eine breite Basis und die Möglichkeit, standardisierbare Abfragen zu entwickeln. Das Romankorpus ermöglicht empirisch-systematische Vergleiche von Romanen untereinander und mit anderen Textsorten. Durch die ausführlichen bibliographisch-buchgeschichtlichen Header-Annotationen entstehen soziolinguistisch auswertbare und mit sprach- und literatursystematischen Phänomenen verknüpfbare Metadaten. Das umfassende Konzept des DDD-Gesamtkorpus erlaubt ferner, dass die Textsorten im Verbund für Untersuchungen zur Verfügung stehen, etwa auch über die Grenze zwischen faktualen und fiktionalen Textsorten hinweg. So wird die Entwicklung von Textsorten erstmals auf breiter repräsentativer und verlässlicher Materialbasis – hier exemplarisch anhand der modernen Leitgattung ›Roman‹ – diachron nachvollziehbar. Für die lese- und buchgeschichtliche Forschung und damit für eine neue Form der Literaturgeschichtsschreibung in Deutschland eröffnet das Projekt ein singuläres Untersuchungsfeld. Mit ihm würden computergestützte Formen vom Rand der Literaturwissenschaft in das Zentrum der Disziplin rücken und damit zur Verwissenschaftlichung des Faches beitragen.

Die Narratologie gewinnt durch das Projekt nicht nur eine einzigartige Materialbasis. Die Erschließung des Romankorpus durch narratologische Annotationen setzt schon während des Projekts die Entwicklung standardisierbarer, erzählanalytischer Auszeichnungssets voraus und wirkt so begriffsklärend auf die Narratologie zurück. Gelingt das Forschungsvorhaben, dann wird eine neue digitale Bibliothek hohe Romane und blaue Bibliotheken ins digitale Gedächtnis des 21. Jahrhunderts heben, aber auch das Fach verändern, das dieses kulturelle Gedächtnis zu verwalten beansprucht, die Literaturwissenschaft.

---

<sup>44</sup> Z. B. an die Korpora des Instituts für Deutsche Sprache Mannheim <<http://www.ids-mannheim.de/kt/corpora.html>> (5.1.2006) oder an das DWDS-Korpus (Digitales Wörterbuch der deutschen Gegenwartssprache) <<http://www.dwds.de>> (5.1.2006).

## Literaturverzeichnis

- A.B.C. [Johann Friedrich Bachstrom]: *Das Bey zwey hundert Jahr [...]*. 2. Teil. Frankfurt u. Leipzig [wohl: Breslau: Korn] 1737.
- Alonso, Carlos J. u. a.: *Crises and Opportunities. The Futures of Scholarly Publishing*. ACLS Occasional Paper, No. 57 <<http://www.acls.org/op57.pdf>> (5.1.2006)
- Belletristisches Lese-Cabinet der neuesten und besten Romane aller Nationen in sorgfältigen Übersetzungen*. Wien, Leipzig u. a.: Hartleben 1850–1853.
- Bertram, Mathias: *Deutsche Literatur von Luther bis Tucholsky. Großbibliothek. Digitale Bibliothek*. Bd. 125. Berlin: Directmedia Publishing 2005.
- Bibliographie deutscher Übersetzungen aus dem Französischen 1700–1948*. Bearb. von Hans Fromm. 6 Bde. Baden-Baden: Verl. für Kunst und Wissenschaft 1950–1953.
- Bochinger, Christoph: *Pietistische Identität zwischen persönlicher Frömmigkeit und Gruppenprozessen*. In: *Hallesche Beiträge zur Orientwissenschaft* 24 (1997), S. 33–44.
- Boghardt, Martin: *Analytische Druckforschung. Ein methodischer Beitrag zu Buchkunde und Textkritik*. Hamburg: Hauswedell 1977.
- Borgstedt, Thomas u. Solbach, Andreas (Hg.): *Der galante Diskurs. Kommunikationsideal und Epochenschwelle*. Dresden: Thelem 2001 (= Arbeiten zur neueren deutschen Literatur; 6).
- Bowers, Fredson: *Principles of Bibliographical Description*. Princeton: Princeton University Press 1949.
- Dünnhaupt, Gerhard: *Personalbibliographien zu den Drucken des Barock*. 2. verb. Aufl. Stuttgart: Anton Hiersemann 1990–1993.
- Füssel, Stephan: *Vom Steindruck zum Internet. Faksimiles im Medienumbruch*. In: Hanuschek, Sven (Hg.): *Die Struktur medialer Revolutionen. Festschrift für Georg Jäger*. Frankfurt u. a.: Lang 2000, S. 134–142.
- Gaskell, Philip: *A New Introduction to Bibliography*. Oxford: Oxford University Press 1972.
- Genette, Gérard: *Die Erzählung*. München: Fink 1994 (= UTB für Wissenschaft).
- Graeber, Wilhelm u. Roche, Geneviève: *Englische Literatur des 17. und 18. Jahrhunderts in französischer Übersetzung und deutscher Weiterübersetzung. Eine kommentierte Bibliographie*. Hg. v. Jürgen Stackelberg. Tübingen: Niemeyer 1988.
- Gumbrecht, Hans U. u. Pfeiffer, K. Ludwig (Hg.): *Materialität der Kommunikation*. Frankfurt: Suhrkamp 1988 (= Suhrkamp-Taschenbuch Wissenschaft; 750).
- Hausmann, Frank-Rutger u. Kapp, Volker (Hg.): *Bibliographie der deutschen Übersetzungen aus dem Italienischen von den Anfängen bis zur Gegenwart*. Tübingen: Niemeyer 1992–2004.
- Jannidis, Fotis: *Wider das Altern elektronischer Texte. Philologische Textauszeichnung mit TEI*. In: *editio* 11 (1997), S. 152–177.
- Jeanneney, Jean Noël: *Quand Google défie l'Europe. Plaidoyer pour un sursaut*. Paris: Mille et une nuits 2005.
- Lüdeling, Anke; Poschenrieder, Thorwald u. Faulstich, Lukas: *Deutsch Diachron Digital – Ein diachrones Korpus des Deutschen*. In: *Jahrbuch für Computerphilologie* 6 (2004), S. 119–136 <<http://computerphilologie.uni-muenchen.de/jg04/luedeling/ddd.html>> (5.1.2006).

- Martino, Alberto: *Die italienische Literatur im deutschen Sprachraum. Ergänzungen und Berichtigungen zu Frank-Rutger Hausmanns Bibliographie*. Amsterdam u. a.: Rodopi 1994 (= Chloe; 17)
- Perlen. *Journal für gute deutsche Original-Novellen nebst den besten Übersetzungen der berühmtesten Schriftsteller des Auslandes*. Berlin: Sacco 1846–1865.
- Rapp, Andrea: *Digitale Faksimiles*. In: *Aus dem Antiquariat* 5 (2004), S. 361–366.
- Rush, Wade: *Digitize This. Yahoo hopes to trump Google with its Open Content Alliance publishing venture*. In: *Technology Review* (<[http://www.technologyreview.com/articles/05/10/wo/wo\\_102005roush.asp](http://www.technologyreview.com/articles/05/10/wo/wo_102005roush.asp)> vom 20.10.2005 [5.1.2006]).
- Schön, Erich: *Gibt es Literatur ohne Leser*. In: Winko, Simone; Jannidis, Fotis u. Lauer, Gerhard (Hg.): *Literatur, Literaturbegriff, literarische Kommunikation*. Berlin: de Gruyter 2007 (in Vorb.).
- Schönebaum, Herbert: *Art. Johann Friedrich Bachstrom*. In: *Allgemeine Deutsche Biographie*. Hg. v. der Historischen Kommission bei der Bayerischen Akademie der Wissenschaften. Leipzig: Duncker & Humblot 55 (1911), S. 664–667 und 56, S. 395 (Korrektur).
- Schreibman, Susan; Siemens, Ray u. Unsworth, John (Hg.): *A Companion to Digital Humanities*. Malden: Blackwell 2004.
- Schulz, Friedrich: *Kleine Romane*. 5 Bde. Leipzig: Göschen 1788–90.
- Showalter, English (Hg.): *Transformation du genre romanesque au XVIII<sup>e</sup> siècle*. In: *Eighteenth-Century Fiction* 13, 2–3 (2001).
- Simons, Olaf: *Marteaus Europa oder Der Roman, bevor er Literatur wurde. Eine Untersuchung des deutschen und englischen Buchangebots der Jahre 1710 bis 1720*. Amsterdam u. Atlanta: Rodopi 2001 (= Internationale Forschungen zur allgemeinen und vergleichenden Literaturwissenschaft; 52).
- Ullrich, Hermann: *Art. Johann Friedrich Bachstrom*. In: *Neue Deutsche Biographie*. Hg. v. der Historischen Kommission bei der Bayerischen Akademie der Wissenschaften. Berlin: Duncker & Humblot 1 (1953), S. 503.
- Ders.: *Johann Friedrich Bachstrom. Ein Gelehrtenleben aus der ersten Hälfte des 18. Jahrhunderts*. In: *Euphorion* 16 (1909), S. 28–58 und 320–349.
- Verzeichniß eines Theils der Kupferstich- und Bücher-Sammlung Johann Wilhelm Ludewig Gleim's: welcher den 22sten October d.J. u. folgende Tage, in d. gewöhnlichen Nachmittagsstunden, im sogenannten großen Juden-Hause öffentl. versteigert werden soll; nebst einigen Anhängen*. Repr. d. Orig.-Ausg. 1804. Mit einem Nachw. von Reinhard Selz. München u. a.: Saur 1987.
- Weber, Ernst u. Mithal, Christine: *Deutsche Originalromane zwischen 1680 und 1780*. Berlin: Erich Schmidt 1983.
- Worstbrock, Franz-Josef: *Deutsche Antikenrezeption 1450–1550*. Mit einer Bibliographie der Übersetzer. Boppard: Boldt 1976.

**Empfohlene Zitierweise:**

Jannidis, Fotis, Lauer, Gerhard und Rapp, Andrea: Hohe Romane und blaue Bibliotheken. Zum Forschungsprogramm einer computergestützten Buch- und Narratologiegeschichte des Romans in Deutschland (1650-1900). <[http://www.germanistik.ch/publikation.php?id=Hohe\\_Romane\\_und\\_blaue\\_Bibliotheken](http://www.germanistik.ch/publikation.php?id=Hohe_Romane_und_blaue_Bibliotheken)>

**germanistik.ch**  
Verlag für Literatur- und Kulturwissenschaft