

Digitale Philologie: Das Darmstädter Modell

Adler, Marc; Bartsch, Sabine; Becker, Maria et al.

(2020)

DOI (TUprints): <https://doi.org/10.25534/tuprints-00012476>

Lizenz:



CC-BY 4.0 International - Creative Commons, Namensnennung

Publikationstyp: Buch

Fachbereich: 02 Fachbereich Gesellschafts- und Geschichtswissenschaften

Quelle des Originals: <https://tuprints.ulb.tu-darmstadt.de/12476>

Digitale Philologie: Das Darmstädter Modell

Digital Philology |

Working Papers in Digital Philology – 01 | 2020

Herausgegeben von

Sabine Bartsch | Evelyn Gius | Marcus Müller | Andrea Rapp | Thomas Weitin

Marc Adler, Sabine Bartsch, Maria Becker, Michael Bender, Luise Borek, Cindy Brinkmann, Friedrich Michael Dimpel, Rotraut Fischer, Anastasia Glawion, Svenja A. Gülden, Canan Hastik, Philipp Hegel, Katharina Herget, Franziska Horn, Celia Krause, Marcus Müller, Alexandra Núñez, Andrea Rapp, Lisa Scharrer, Oliver Schmid, Jörn Stegmeier, Beate Thull, Thomas Weitin



TECHNISCHE
UNIVERSITÄT
DARMSTADT



Impressum

Postadresse:

Technische Universität Darmstadt

Institut für Sprach- und Literaturwissenschaft

Dolivostraße 15

64293 Darmstadt

Website: www.digital-philology.tu-darmstadt.de

Email: sprachli@linglit.tu-darmstadt.de



Zitierhinweis: Adler, Marc et al. (2020): Digitale Philologie: Das Darmstädter Modell. In: Digital Philology | Working Papers in Digital Philology 01 | 2020. Darmstadt: TUPrints.

Inhaltsverzeichnis

Vorwort der Herausgeber*innen zum Start der Reihe.....	ii
1. Einleitung	1
2. Ressourcen.....	3
3. Methodengnese	5
3.1 Zielsetzung.....	6
3.2 Philologische Methoden / Methoden in der Philologie.....	7
3.3 Disziplinäre Verankerung der Methoden.....	13
3.4 Anforderungen an die Digitalen Philologien	15
4. Annotation.....	16
4.1 Annotation als Auszeichnung, Anreicherung und Analysegrundlage	17
4.2 Annotation als kollaborative Interpretations- und Theoriearbeit	19
4.3 Annotation als Kontextualisierung.....	20
5. Thesauri, Taxonomien und Ontologien	21
6. Denkstile.....	26
7. Fazit	29
Literatur.....	31

Vorwort der Herausgeber*innen zum Start der Reihe

Mit der vorliegenden Publikation ist der Startschuss gefallen für die neue Open-Access-Schriftenreihe “Digital Philology | Working Papers in Digital Philology”, die von der Digitalen Philologie am Institut für Sprach- und Literaturwissenschaft an der Technischen Universität Darmstadt ins Leben gerufen wird. Sie soll ebenso wie eine weitere neue Reihe – “Digital Philology | Evolving Scholarship in Digital Philology” – die Publikationslandschaft der Digital Humanities in Deutschland bereichern und profilieren. Die folgende kurze Einführung zum Selbstverständnis der Herausgeber*innen und zur inhaltlich-fachlichen Ausrichtung beiden Reihen begleitet daher den ersten Band der Reihe ‘Working Papers in Digital Philology’.

Im Darmstädter Modell der Digital Humanities betrachten wir Digitalität als integrativen Bestandteil der jeweiligen fachlichen Identität; daher bezeichnen wir unsere Forschungstätigkeiten und auch unsere Studiengänge nicht mit dem übergreifenden Begriff bzw. umbrella term ‘Digital Humanities’, sondern spezifizieren das Feld als ‘Digitale Philologie’ bzw. ‘Linguistic and Literary Computing’. Die Beschäftigung mit Digitalität umfasst dabei sowohl die Aspekte der Materialität und Medialität – also die Befassung mit digitalen Objekten – als auch die Entwicklung und Anwendung digitaler Methoden und Verfahren. Mit seinen vier Professuren und einer weiteren Fachgebietsleitung mit dediziert digital-philologischer Ausrichtung hat das Darmstädter Institut für Sprach- und Literaturwissenschaft ein Alleinstellungsmerkmal – mindestens in Deutschland, aber auch weit darüber hinaus.

Die Herausgeber*innen stellen sich vor:

Sabine Bartsch ist Anglistin und Linguistin mit einem Schwerpunkt im Bereich Korpus- und Computerlinguistik und interessiert sich besonders für Kollokationsforschung, historische Register der Wissenschaftskommunikation, Multimodalität sowie Korpusdesign und -aufbau und Methoden korpusbasierter Analyse.

Evelyn Gius ist digitale Literaturwissenschaftlerin und interessiert sich besonders für Annotation, die narrative Struktur und die Segmentierung von literarischen Texten, die Automatisierung bzw. Automatisierbarkeit von Textanalyse sowie die Wechselwirkungen zwischen computationellen Verfahren und etablierten literaturwissenschaftlichen Methoden.

Marcus Müller ist Linguist und interessiert sich für Korpuslinguistik, digitale Diskursanalyse, Wissenschaftskommunikation, grammatische Variation sowie Sprache in der Kunstkommunikation.

Andrea Rapp ist germanistische Mediävistin und Computerphilologin mit Bibliothekserfahrung und interessiert sich besonders für die Digitalisierung und Erschließung mittelalterlicher Handschriften, die Analyse von Urkunden und Briefen, Varianz und Varietäten, Editionsphilologie, Lexikographie, Annotationsverfahren sowie Forschungs(daten)infrastrukturen.

Thomas Weitin ist digitaler Literaturwissenschaftler und interessiert sich für Modelle, die das Verhältnis des kanonischen Teils der Literaturgeschichte zum *great unread* sichtbar machen. Sein zweiter Forschungsschwerpunkt liegt in der kognitiven Rezeptionsanalyse.

International gibt es zahlreiche Publikationsorte, Zeitschriften und Reihen, die das Forschungsfeld der Digital Humanities bedienen. In Deutschland bzw. im deutschsprachigen Raum ist zum einen die ‘Zeitschrift für digitale Geisteswissenschaften ZfdG’ (www.zfdg.de) zu nennen, die der Forschungsverbund Marbach Weimar Wolfenbüttel in Zusammenarbeit mit dem Verband Digital Humanities im deutschsprachigen Raum (DHD) etabliert hat und die insbesondere bestandorientierte Forschungsbeiträge fördert. Zum anderen gibt es verschiedene Working-Papers-Reihen, unter denen die ‘DARIAH-DE Working Papers’ (de.dariah.eu/working-papers) beispielsweise für Beiträge zu digitaler Forschung oder digitalen Infrastrukturen in den Geisteswissenschaften offen sind. Die Max Weber Stiftung betreibt gemeinsam mit weiteren Institutionen mit dem deutschsprachigen Teil des Blogportals ‘hypotheses’ (de.hypotheses.org) einen nicht-kommerziellen Service für Wissenschaftsblogs. Wir meinen, dass in diesem Feld noch Raum ist für spezifischere, fachlich orientierte Open-Access-Reihen, die die vorhandenen Publikationsorte ergänzen: eine Reihe, die auf Digitale Philologie ausgerichtet ist sowie eine weitere Reihe, die hervorragende Abschlussarbeiten (in der Regel Master) aufnehmen soll. In beiden Reihen sind deutsch- und englischsprachige Publikationen willkommen. Mit dem Start dieser zwei neuen zusammenhängenden und dennoch klar distinkten digitalen Open-Access-Reihen soll also ein Publikations-Ökosystem geschaffen werden, das von Microblogging, individuellen und qualitätsgesicherten Blogs über die Publikation von Qualifikationsarbeiten und Working Papers, Diskussionsbeiträgen usw. hin zu etablierten Journals führt. Idealer Publikationsort für diese Schriftenreihen ist das institutionelle Repositorium der Technischen Universität Darmstadt mit dem Publikationsservice TUprints – er bietet die organisatorischen und technischen Rahmenbedingungen zur elektronischen Publikation wissenschaftlicher Dokumente im Sinne des offenen und freien Zugangs zu Wissenschaft, Lehre und Forschung.

Die ‘Working Papers in Digital Philology’ sollen ein niedrighschwelliges und offenes sowie qualitätsgesichertes Angebot für die Publikation von Working Papers, White Papers, Diskussionsimpulsen, Projektberichten und ähnlichen Formaten bieten. Interdisziplinäre und kollaborative

Publikationen sind ebenso willkommen wie solche von Personen und Gruppen außerhalb des engeren Darmstädter Digital-Philology-Kontexts.

Die Reihe ‘Evolving Scholarship in Digital Philology’ soll ein Ort für die Publikation studentischer Abschlussarbeiten sein. Die Qualitätssicherung erfolgt auf der Basis bestimmter Bewertungsvoraussetzungen (in der Regel Mindestnote 1,7) und erfordert zudem ein Peer Review von außerhalb des Darmstädter Herausgeberteams bzw. die Auszeichnung durch einschlägige Preise.

Band 1 der ‘Working Papers in Digital Philology’ startet mit dem Experiment einer kollaborativen Publikation. Beteiligt sind alle 23 zum damaligen Zeitpunkt (Sommer 2017) aktiven Mitarbeiter*innen und Doktorand*innen eines Digital Philology Retreats. Mit der seinerzeit erfolgten Neubesetzung zweier Professuren verstärkte sich die digitale Ausrichtung des Instituts entscheidend und das Team wurde größer und vielfältiger, so dass die Verständigung über Digitalität, die Ausgestaltung der ‘Digitalen Philologie’ im ‘Darmstädter Modell’ und die Erarbeitung einer gemeinsamen Publikation die Neuaufstellung des Instituts begleiten sollte. Mittlerweile ist die Gruppe um eine weitere Professur ergänzt und damit ist ein hervorragender Anlass für den Start der neuen Reihen gegeben.

Die Herausgeber*innen möchten sich bei allen bedanken, die das Vorhaben unterstützt haben und weiterhin unterstützen: zunächst bei ihren Teams, die die Digitale Philologie an der Technischen Universität Darmstadt zu einem so lebendigen und attraktiven Ort machen sowie bei der Universitäts- und Landesbibliothek, die insbesondere mit dem Team Digitales Publizieren Herausgeber*innen und Autor*innen vorbildlich betreut.

1. Einleitung

Wenn ‘Philologie’ die “Liebe zur Sprache” bezeichnet und ‘digital’ sich auf die Darstellung diskreter Einheiten in Binärcode bezieht, so scheint die Bezeichnung “Digitale Philologie” zunächst vor allem Gegensätze zusammenzubringen. Doch die Verbindungen zwischen diesen “Gegensätzen” sind vielfältig und bereits langdauernd: Die Anfänge der heute florierenden Digital Humanities lassen sich mit Roberto Busas Index Thomisticus markieren, dessen Planungen in die 1940er Jahre zurückreichen (Thaller 2017). Seit der Jahrtausendwende – seitdem zum einen zumindest in einigen Bereichen digital(isiert)es Kulturerbe in ausreichender Masse vorliegt und seitdem zum anderen Zugang zu Computerkapazität alltäglich geworden ist – haben die Digital Humanities Aufschwung genommen. Oft werden sie jedoch als eher übergreifender methodischer Zugang zur Erforschung des Kulturerbes verstanden und entwickeln sich so neben den Disziplinen der Humanities, während echte fachspezifische Transformationen selten sind. In der Computerlinguistik wurde zwar eine solche fachspezifische Transformation vollzogen, digitale Technologie mit Sprachwissenschaft verbunden und eine eigenständige Disziplin etabliert, sie ist jedoch meist sehr nahe an die Informatik heran und vom ursprünglichen sprachwissenschaftlichen Kern weggerückt.

In der Digitalen Philologie nach dem Darmstädter Modell sind die Methoden und Ansätze der Digital Humanities hingegen in die Fachgebiete integriert und treten nicht als neues Fach neben die geisteswissenschaftlichen Disziplinen. Seit 2016 sind am Institut für Sprach- und Literaturwissenschaft im Fachbereich 2 Gesellschafts- und Geschichtswissenschaften der Technischen Universität Darmstadt alle Fachgebiete der (germanistischen) Philologie in einer digitalen Ausrichtung in Forschung und Lehre vertreten: die Digitale Linguistik, die Digitale Literaturwissenschaft, die Computerphilologie und Mediävistik, ergänzt um das Fachgebiet Korpus- und Computerlinguistik mit einem Fokus in der englischen Philologie. Diese Integration der auf digitale Methoden und Gegenstände ausgerichteten wissenschaftlichen Forschungsarbeit in die Fächer ist ein Alleinstellungsmerkmal des Instituts, das dieses Modell seit 2002 konsequent entwickelt und in Forschung und Lehre ausgestaltet hat und weiter ausbaut. Mit dem seit 2006 eingeführten und akkreditierten Master-of-Arts-Studiengang ‘Linguistic and Literary Computing’ und dem im Wintersemester 2015/16 gestarteten Bachelor-of-Arts-Studiengang ‘Digital Philology’ hat es ein deutschlandweit einmaliges konsekutives Studienangebot mit der dedizierten Ausrichtung Digitale Philologie etabliert. Diese Konstellation erlaubt bereits institutsintern vielfältige Austauschprozesse zwischen den Fachgebieten und die Identifikation von Schnittstellen, Kooperations- bzw. Kollaborationsmöglichkeiten und Gemeinsamkeiten, aber auch von

Unterschieden im Hinblick auf Untersuchungsgegenstände, Methoden und Theorien. Gemeinsame Vorlesungen, Kolloquien, Workshops und Projekte fördern diesen Austausch. Im Mai 2017 wurden nach einem Jahr der Zusammenarbeit im Rahmen einer gemeinsamen Klausurtagung in Rieneck vorhandene und potenzielle Schnittstellen systematisch herausgearbeitet. Sie stellen die zentralen Punkte des folgenden Working Papers dar, das den Auftakt einer Open-Access-Publikationsreihe bildet, die den Austausch mit den verschiedenen Scientific Communities im Rahmen der Digitalen Philologie anregen soll.

Ausgangspunkt der Überlegungen sind die Ressourcen des digital-philologischen Forschens, also Untersuchungsgegenstände, Quellen bzw. Daten, die erforscht werden und mit denen gearbeitet wird. Dabei werden Merkmale fachspezifischer Ressourcen herausgearbeitet, beispielsweise in Repositorien, Katalogen, literaturwissenschaftlichen Editionen und linguistischen Korpora. Darüber hinaus wird zwischen analogen Ressourcen, digitalen Derivaten bzw. Surrogaten, Substituten sowie Repräsentationen und Modellen differenziert und die Implikationen der analogen oder digitalen Verfasstheit der Ressourcen für unterschiedliche Forschungsprozesse reflektiert und implementiert.

Eng verknüpft mit diesem Punkt sind die Methoden und ihre Genese. Ihre analoge und digitale Ausprägung und Entwicklung im jeweiligen Fachgebiet zur Erschließung und Analyse spezifischer Ressourcen und Bearbeitung bestimmter Fragestellungen werden mit Blick auf disziplinäre Erfordernisse und Möglichkeiten sowie auf interdisziplinäre Transfermöglichkeiten reflektiert.

Eine Schnittstelle zwischen den Fachgebieten sind die Prozesse und Produkte des Annotierens in unterschiedlichen Formen. Diesbezüglich wird diskutiert, inwiefern Annotation einerseits als analytische Methode und Erschließungs- und Anreicherungsverfahren, andererseits als Ausgangspunkt für die Theorieentwicklung (beispielsweise durch die Entwicklung taxonomischer Schemata) angesehen werden kann. Des Weiteren richtet sich die Diskussion darauf, inwiefern Annotationen einerseits als Ausgangsdaten für Untersuchungen, andererseits aber auch als (inter)textuelle Ebene betrachtet werden können, auf der Wissen ausgehandelt und dokumentiert wird. Gefragt wird auch danach, inwiefern diese Ebene dann selbst ein Produkt darstellen würde, das (mit)publiziert werden und inhaltliche Anreicherungen und Verlinkungen zu anderen Texten enthalten kann. Die Überlegungen in diesem Abschnitt nehmen ihren Ausgangspunkt in den Erfahrungen verschiedener konkreter Annotationsprojekte.

Anschließend werden Überlegungen zur Rolle von Taxonomien und Ontologien in der digitalen Forschungslandschaft angestellt – als Bindeglied zwischen Ressourcen und Methoden, aber auch als übergeordnete Strukturen.

Im abschließenden Kapitel werden Überlegungen zur Umwälzung von Denkstilen in den Philologien angestellt – im Hinblick auf die neue Umgangsweise mit Forschungsobjekten, auf die Entwicklung neuer Forschungsfragen, auf den Wandel in der Anwendung von Methoden und auf die Weiterentwicklung von Theorien bzw. des Denkens unter digitalen Bedingungen.

2. Ressourcen

Der Begriff der Ressource ist in den Geisteswissenschaften nicht einschlägig. Vor allem in der Ökonomie und Soziologie, aber auch in der Psychologie und Informatik ist er hingegen gebräuchlich. Hier bezeichnet er in der Regel Mittel, um bestimmte Handlungen und Vorgänge zu ermöglichen. Im digitalen Bereich wird der Ausdruck “Ressource” zum Teil sehr umfassend verstanden. Das W3C-Konsortium (www.w3.org) spricht bei allem, was mit dem Resource Description Framework (RDF) beschrieben werden kann, von Ressourcen und versteht sie als Instanz einer Klasse `rdfs:Resource`, von der es heißt: “This is the class of everything” (RDF Schema 1.1).

Allerdings verweist der Begriff auch auf die Quelle, die in seinem ursprünglich metaphorischen Sinn für die geisteswissenschaftliche Arbeit grundlegend ist. Sie bezeichnet dort ganz allgemein den Gegenstand, den insbesondere historisch orientierte Wissenschaftler*innen auf ihre Weisen betrachten und erschließen. Wenn im Folgenden von Ressource die Rede ist, soll dieser Bezug zum Material geisteswissenschaftlicher Tätigkeit bedacht werden.

Als Objekt der Geisteswissenschaften ist die Ressource oder Quelle zwar eine notwendige Bedingung ihrer Betrachtung und bestimmt die Möglichkeiten solcher Betrachtung selbst, aber sie ist auch für deren Selbstverständnis von entscheidender Bedeutung. Darin unterscheidet sich die Ressource in diesem Sinne von ökonomischen, sozialen und psychologischen Mitteln, die in ihrer Weise die geisteswissenschaftliche Arbeit beeinflussen. Diese Mittel sind nämlich nicht definitorisch entscheidend für die geisteswissenschaftliche Arbeit selbst. Auch die technischen Grenzen sind Bedingungen der Möglichkeit geisteswissenschaftlicher Arbeit und an sich nicht entscheidend für das disziplinäre Selbstverständnis. Aber in Hinblick auf die Nutzung digitaler Materialien und Methoden in den Geisteswissenschaften bedürfen diese Mittel einer besonderen Aufmerksamkeit. Im Gegensatz zu den ökonomischen, sozialen und psychologischen Ressourcen werden sie nicht nur bei dieser Arbeit benötigt und vorausgesetzt, sondern häufig auch selbst geschaffen. Die geisteswissenschaftliche Tätigkeit erzeugt Materialien und auch Methoden, die sie selbst im weiteren Verlauf nutzt. Die folgenden drei Thesen fassen verschiedene Aspekte und Probleme der insbesondere philologischen Ressource in der digitalen Welt zusammen.

Ressourcen sind nicht unmittelbar gegeben, sondern werden konstruiert. Sie drücken ein Erkenntnisinteresse und auch einen Denkstil im unten erläuterten Sinn aus (vgl. Kap. 6). Sie sind nicht nur ein Mittel zur Ausführung von Handlungen, sondern selbst das Ergebnis einer Handlung. In dieser Konstruktion sind sie aber in einer jeweils spezifischen Weise gegeben. Zum Beispiel sind raumzeitliche Objekte immer nur einmal gegeben, abstrahierte Objekte wie digitale Texte sind hingegen (verlustfrei) reproduzierbar und daher, wenn man an unikale kulturelle Objekte denkt, auch ohne größere Bedenken manipulierbar. In dieser Weise ist digitale Abstraktion ein Gewinn an Manipulierbarkeit durch Reduktion von Komplexität.

Ressource ist nicht gleich Ressource. Die Philologie spricht zunehmend von Ressourcen, die originär digital sind (originär digitale Ressourcen oder ‘born-digital objects’). Sie spricht aber unter anderem auch und vorrangig von Gegenständen, die nicht digital sind (‘analoge’ Ressourcen). Damit diese Gegenstände digital erforscht werden können, müssen andere digitale Gegenstände erzeugt werden, die nach bestimmten Kriterien als Repräsentationen der nicht-digitalen Gegenstände betrachtet werden (digitale Surrogate). In der digitalen Forschung entstehen aufgrund digitaler Ressourcen weitere digitale Ressourcen n-ten Grades wie Metadaten, Register, Konkordanzen, Taxonomien oder Korpora (digitale Derivate). Einige digitale Ressourcen stellen zum Beispiel als Annotationen primär Verbindungen zwischen digitalen Ressourcen beliebigen Grades her (digitale Bindeglieder). Diese Verknüpfung erfolgt traditionell meist diskursiv, digital jedoch konkret zwischen Ressourcen. Ergebnisse einer Untersuchung können dabei zur Ressource weiterer Untersuchungen werden. Ressource ist aber nicht gleich Ressource. Insbesondere sind digitale Surrogate nicht die analogen Ressourcen, sondern deren abstrahierte Modelle oder, etwa bei digitalen Faksimiles, “Abbilder” dieser Ressourcen. Fragen der Abstraktion, Modellierung und Repräsentation sind in der Regel Fragen wissenschaftlicher Adäquatheit in Hinblick auf Forschungsfragen und Erkenntnisinteressen.

Die Grenzen der Infrastrukturen, Werkzeuge und Dienste sind die Grenzen digitaler Forschung an Ressourcen. Vershoben werden können diese Grenzen durch die digitale Emanzipation der Geisteswissenschaftler*innen. Die Abwägungen von wissenschaftlichen und technischen Erfordernissen sind Ergebnisse von möglicherweise spannungsgeladenen Verhandlungen, oft zwischen Generalisierbarkeit und Spezifikation. Ressourcen sind Objekte, in denen sich geisteswissenschaftlicher und technischer Anspruch begegnen. Nicht nur neue digitale Ressourcen der philologischen Forschung werden in diesem Prozess selbst produziert, sondern auch die digitalen Methoden, Dienste und Werkzeuge, mit denen ältere Erkenntnisse geprüft, bestätigt oder in Frage gestellt sowie neue Erkenntnisse und Forschungsfragen generiert werden.

Alternativ zum Begriff der Ressource hätte von Daten gesprochen werden können. Aber auch dieser Begriff ist in den Geisteswissenschaften nicht einschlägig. Zum Teil wird er sogar als

problematisch empfunden. Dies mag an dem Verständnis liegen, dass Daten als “das Gegebene” das Wahre oder Faktische seien. Als solches wiederum wird unterstellt, dass in einer Form positivistischer Naivität nicht gesehen werde, dass auch Daten immer nur das sind, was als gegeben betrachtet wird und Menschen, auch in Gruppen, irren können.

Der Begriff des “Datums” erlaubt aber auch ein anderes Verständnis. Insofern “dare” den Vorgang des Gebens allgemein meint, kann Datum auch sein, was zu einem bestimmten Zeitpunkt vorliegt und weitergegeben werden kann. In diesem Sinne kann auch seine Bedeutung im Bereich digitaler Methoden verstanden werden. Daten sind dann das, was von einem Werkzeug übernommen und verarbeitet wird. Der Text beispielsweise, der verarbeitet wird, ist auch dann ein solches Datum, wenn er nicht das Wahre oder Faktische darstellt. Das Ergebnis einer solchen Verarbeitung, etwa eine Liste der im Text genannten Orte, kann ebenfalls Daten für weitere Schritte liefern. Diese Bedeutung des Ausdrucks “Datum” lag auch der früheren Erläuterung der DIN 44300 zugrunde. Sie definiert ihn als “Zeichen oder kontinuierliche Funktionen, die zum Zweck der Verarbeitung Informationen aufgrund bekannter oder unterstellter Abmachungen darstellen” (Deutsches Institut für Normung 1978: 3). Ähnlich klingt es auch in der aktuellen ISO/EC 2382:2015. Das Stichwort “data” definiert sie als “reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing”. Eine Erläuterung setzt hinzu: “Data can be processed by humans or by automatic means.”

Versteht man Daten in diesem engeren Sinn als Eingabe und Ausgabe digitaler Prozesse, so kann man demgegenüber von Ressourcen sprechen, wenn man das geisteswissenschaftliche Potential dieser Daten betonen möchte. Dies scheint der Bevorzugung der Rede von “digitalen Ressourcen” in Bibliotheken und Archiven zu entsprechen. In jedem Fall entspricht es dem Verständnis von “Quelle”, das in dem Ausdruck “Ressource” mitschwingen mag. Ressourcen und Daten können dasselbe bezeichnen, die begriffliche Unterscheidung kann aber eine Unterscheidung der Perspektiven widerspiegeln. Ressourcen werden dann im Sinne der ersten These erst zu Ressourcen durch ein geisteswissenschaftliches Erkenntnisinteresse und sind, so verstanden, auch von den Methoden zu unterscheiden, die zu ihrer Verarbeitung entwickelt und verwendet werden.

3. Methodengese

Ein zentrales Element in der Entwicklung einer digitalen Philologie ist die Einführung und Gese eines erweiterten Inventars an Methoden, die einerseits neu sind, andererseits aber auch in einer langen Tradition stehen. Die Reflexion über die Gese der Methoden speist sich nach unserer Wahrnehmung aus drei Perspektiven. Die erste Perspektive knüpft an die Geschichte

der Philologie an, die traditionell von der Entwicklung und Reflexion über Methoden der systematischen Modellierung von Sprache und sprachhistorischen Prozessen getrieben ist. Hieraus ergibt sich eine Reihe von Fragen, die im engen Zusammenhang mit der philologischen Tradition und ihren Gegenständen und Forschungsfragen sowie mit dem Spannungsfeld zwischen analogen und digitalen Gegenständen und Verfahren und deren Erfordernissen stehen.

Die zweite Perspektive betrifft die interdisziplinäre Interaktion zwischen den philologischen Teildisziplinen Linguistik und Literaturwissenschaft und die damit einhergehenden Fragen der methodischen Synergie zwischen den philologischen Fachgebieten einerseits und dem Spannungsfeld zwischen hermeneutischen und empirischen Forschungsansätzen andererseits.

Die dritte Perspektive betrifft schließlich die interdisziplinäre Interaktion zwischen den philologischen Teildisziplinen und weiteren Disziplinen, wie vor allem der Informatik, die einerseits Entwickler vieler für die digitale Philologie zentraler Techniken und Methoden ist, die andererseits aber nicht eins zu eins auf die digitale Philologie übertragbar sind.

Aus diesen Perspektiven und den sich hieraus ergebenden Spannungsfeldern entsteht ein reizvolles Forschungsfeld, in dem Methodengenerierung und -reflexion kritisch vergleichend stattfinden kann und muss, und in dem die digitale Philologie gleichzeitig gefordert ist, über die Entwicklung digitaler wie genuin philologischer Methoden zu reflektieren, die für ihre Untersuchungsgegenstände und philologischen Fragestellungen geeignet sind. Hier schließen sich auch Fragen nach den spezifischen Ressourcenerfordernissen einer gleichzeitig philologischen und digitalen Forschung an, die einerseits Theorien, Methoden und Algorithmen zur Modellierung philologischer Daten betreffen, andererseits aber auch die Eigenschaften der Forschungsgegenstände, also sprachlich verfasste Daten in Form natürlichsprachlicher Äußerungen aller Art, Quellenrepositorien, linguistischer Korpora und literarischer Texte in Form von Editionen umfassen, die im Zuge ihrer Digitalisierung Eigenschaften des Originals verlieren und neue hinzugewinnen. Dies betrifft z. B. die materialen und medialen Eigenschaften, die zwischen analogen und digitalen Fassungen stark differieren (Gehring / Rapp 2018).

3.1 Zielsetzung

Zielsetzung der Reflexion über die Genese von Methoden in der Digitalen Philologie ist neben einem besseren Verständnis für die philologischen Erfordernisse vor allem die Modellierung exemplarischer Workflows, die unterschiedliche Forschungsszenarien implementieren. In diesem Kontext ergeben sich Fragen, die aus den eingangs skizzierten Spannungsfeldern zwischen analogen und digitalen, hermeneutischen und empirischen Vorgehensweisen sowie in Reaktion auf die aus den Forschungsgegenständen und Fragestellungen der philologischen Teildiszipli-

nen Linguistik und Literaturwissenschaft erwachsen und denen sich die digitale Philologie stellen muss. Aus der Passung zwischen Forschungsfragen, Daten und Methoden ergibt sich eine Reihe weiterer Fragestellungen zu Technologien und Werkzeugen einerseits sowie den zu deren Nutzung erforderlichen Kompetenzen und deren Vermittlung an Wissenschaftler und Studierende.

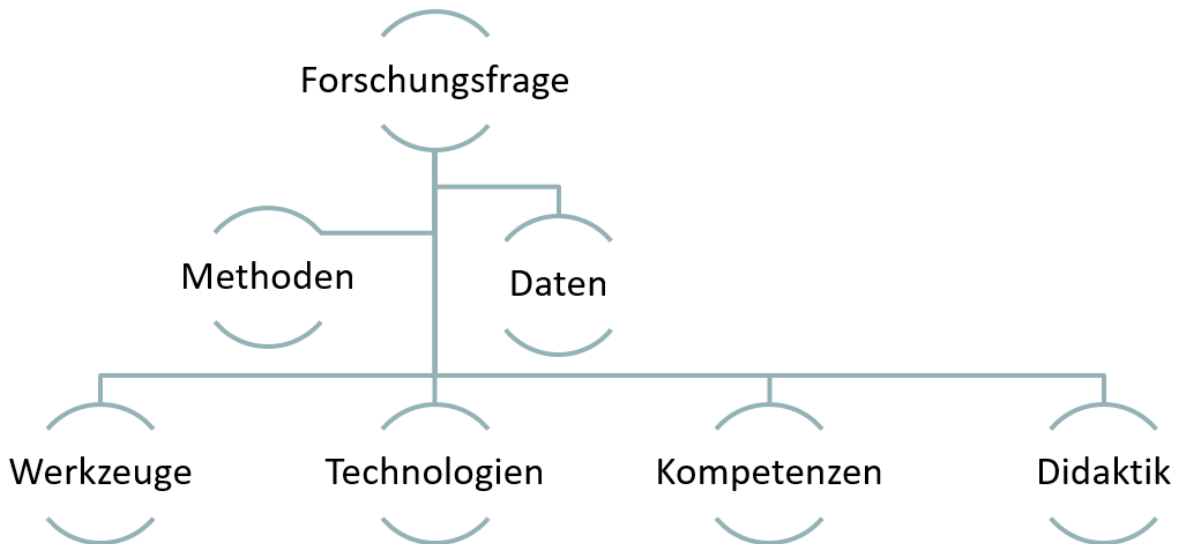


Abb. 1: Methodengenesen

3.2 Philologische Methoden / Methoden in der Philologie

Wie in jeder Disziplin orientieren sich die Methoden an den Gegenständen und deren inhärenten und zugeschriebenen Eigenschaften. Zentraler Gegenstand der Philologie ist die Sprache. Sekundär lassen sich hieraus sprachlich verfasste Artefakte, wie Texte, Diskurse etc. ableiten, die wiederum Eigenschaften aufweisen, die Effekte ihrer materialen und medialen Verfasstheit sowie ihrer kommunikativen Eigenschaften sind. Es sind die Artefakte, also Instanzen der Sprachproduktion, die mit unterschiedlichen philologischen Methoden untersucht werden. Diese Methoden dienen der Analyse auf den verschiedenen Ebenen der sprachlichen Organisation: Sprachlautung und Lautsystem werden traditionell in der philologischen Teildisziplin der Linguistik durch Phonetik und Phonologie untersucht. Hier wird die entlang einer zeitlichen Achse sich entfaltende Lautäußerung anhand der Beschreibung der Phone als distinktive Phöne sowie Eigenschaften wie Intonation und Prosodie untersucht. Die Beschreibung des Wortes als grundlegende bedeutungstragende Einheit findet auf der Ebene der morphologischen Ebene statt und widmet sich neben der Beschreibung der Flexion als ein Mittel der Repräsentation grammatischer und kombinatorischer Eigenschaften von Wörtern auch der Beschreibung ihrer lexikalischen Binnenstruktur anhand der Identifikation von Wortstämmen und deren Erweiterung durch Prozesse der Wortbildung wie Derivation und Komposition. Die Semantik widmet sich der Beschreibung der Bedeutungen und Bedeutungsbeziehungen innerhalb des

Lexikons einer Sprache. Die Syntax befasst sich mit der Beschreibung der Kombination von Wörtern zu größeren syntaktischen Einheiten wie Phrasen, Satzgliedern und schließlich dem Satz als größte syntaktischer Einheit. Zu diesen grundlegenden Beschreibungsebenen treten solche hinzu, die das Spektrum der Optionen der Anordnung der Satzglieder in den Sprachen der Welt sowie die in den Einzelsprachen verwendeten morphologischen Flexionsmuster beschreiben und so zu einer typologischen Klassifikation der Sprachen gelangen. Und schließlich finden sich solche Beschreibungsebenen, die Einheiten oberhalb der Satzgrenze als Texte und Diskurse im Sinne ihrer strukturellen und funktionalen Eigenschaften sowie bedeutungstragenden Einheiten beschreiben, die über den vorliegenden Text oder Diskurs hinausweisen (Pragmatik) und diesen in Beziehung zur Welt setzen. Diese grundlegenden Beschreibungsebenen natürlicher Sprache bilden die Basis der modernen Linguistik; sie greifen die Grundzüge der sprachlichen Organisation auf und sind konstitutiv für das Beschreibungsinventar der zu untersuchenden Einheiten und Gegenstände und für die darauf anwendbaren Methoden der Philologie.

Vor allem die empirische Beobachtung und systematische Beschreibung phonetischer, morphologischer und syntaktischer Regularitäten und deren regelhafte Veränderungen über die Zeit, wie sie vor allem die historische Sprachwissenschaft geprägt hat, bilden die Grundlage der modernen Philologie und ihrer hier skizzierten Analyseebenen. Es ist die Systematik der Beobachtung und Beschreibung und die zugrundeliegende Annahme, dass Regularitäten die Grundlage systematischer Veränderungen des Sprachsystems bilden müssen, die wiederum die Basis der modernen philologischen Forschung bildet, und die zu Beginn des 20. Jh. als Grundlage der Linguistik als eigenständige Disziplin in Abgrenzung von der Literaturwissenschaft aufgegriffen und ausdifferenziert wurde. So sind Linguistik und Literaturwissenschaft disjunkte Fächer einerseits, gleichzeitig aber bilden sie aufgrund ihres prinzipiell sprachlich verfassten Gegenstandes gemeinsam die Philologie. Und während seit dem Beginn des 20. Jh. vor allem die Unterschiede und Grenzen zwischen diesen beiden Fächern in den Vordergrund gerückt sind, hat die Hinwendung großer Teile der Philologie zum Digitalen die Gemeinsamkeiten wieder stärker in den Blick gerückt. Interessanterweise haben dabei gerade die Hinwendung zur Digitalität sowie die daraus sich ergebenden Methodensynergien diese Rückbesinnung auf das gemeinsame Fundament der Philologie begünstigt und befördert. Im Folgenden sollen deshalb Methoden und theoretische Vorannahmen der an der Philologie beteiligten Fächer die Folie sein, vor der exemplarisch das Methodeninventar der digitalen Philologie betrachtet wird.

Die Philologie definiert sich von jeher auch über ihre Methoden; in diesem Sinne ist die systematische Beobachtung von Mustern und Veränderungen von Sprachdaten konstitutiv für ihre

Entwicklung als eigenständige wissenschaftliche Disziplin. Das Methodeninventar der Philologie geht ursprünglich aus der empirischen Beobachtung sprachlicher Daten hervor, wie sie vor allem die historische, aber auch die synchronische Sprachwissenschaft kennzeichnet. Der empirische Impetus erklärt die Genese des Methodenspektrums. Grundlegend für Methoden in der Philologie ist die Untersuchung und Modellierung von an der Oberfläche der sprachlichen Instanz, vor allem der gedruckten und in der Erweiterung, der digitalisierten Form beobachtbaren Merkmale und deren Zusammenspiel auf unterschiedlichen Ebenen sprachlicher und textueller Organisation. Diese Methode der Betrachtung wird implementiert in Form von Wortformenlisten und Wortfrequenzlisten sowie von Konkordanzen, die in Form sogenannter Key-Word-in-Context-(KWIC)-Konkordanzen als wichtige Grundtechnologie aktueller Korpuswerkzeuge aus der empirisch arbeitenden Philologie nicht mehr wegzudenken sind. Wort- bzw. Vokabellisten sowie Konkordanzen sind ein gutes Beispiel für die gute Passung und Anpassung einer jahrhundertalten philologischen Technik der Textbetrachtung an die Erfordernisse digitaler Forschung. Bereits seit dem Mittelalter ist die Konkordanz ein wichtiges Instrument für die Exegese wichtiger philosophischer und religiöser Texte. Was ursprünglich und bis in die zweite Hälfte des 20. Jh. als Instrument zur Betrachtung ausgewählter kanonischer Texte wie die Bibel oder die Werke Shakespeares (Spevack 1973) aufgrund der aufwendigen Produktion nur wenigen Texten vorbehalten war, entwickelte sich mit der Einführung digitaler Computer zum grundlegenden Werkzeug zur Untersuchung digitaler Texte (Gärtner / Kühn 1998; Sinclair 1991). Deren Stärken bei der Suche und Sortierung über sprachliche Daten verhalfen der Konkordanz zu einer Renaissance und führten zu ihrer Transformation von einer spezifischen Anordnung, Darstellung und Indizierung ausgewählter Texte (Cameron et al. 1970) zu einem Werkzeug, das die programmatische Suche und Sortierung über jede Art von Texten in digitaler Form ermöglicht.

Die Annotation von Texten, von der in einem späteren Kapitel noch ausführlicher die Rede sein wird, ist ein weiteres Beispiel einer traditionellen philologischen Technik, die durch die Überführung in das digitale Medium einerseits verändert worden ist, andererseits aber auch viele ihrer ursprünglichen Eigenschaften und Bedeutung beibehalten hat. War Annotation seit jeher ein zentrales philologisches Werkzeug (bzw. ein Werkzeug des aneignenden Lesens), mit dem im einfachsten Fall im Text selbst oder in Marginalien dem sprachlichen Artefakt analytische und interpretatorische Information hinzugefügt und so der Text angereichert wurde, so ließ sich diese Technik nicht nur ideal in das digitale Medium überführen, sondern ist auch in Form automatischer und probabilistischer Klassifikationsinstrumente, wie Wortartentagger und syntaktischer Parser, zu teils sehr ausgereiften automatischen Annotationsverfahren weiterentwickelt worden.

Beide Methoden und Verfahren weisen einige Gemeinsamkeiten auf: beide sind lange in der Philologie verwurzelt und gehen zurück auf analoge Verfahren der prä-digitalen Zeit; beide erfüllen als Instrumente der Sichtung, Indizierung und Analyse sowie der Anreicherung philologischer Daten nach wie vor Funktionen, die sie schon vor Jahrhunderten erfüllt haben und beide haben sich in exzellenter Weise in das digitale Zeitalter überführen lassen, wodurch ihre angestammten Funktionen erhalten, aber auch erweitert und an die Erfordernisse digitaler Daten angepasst werden konnten. So gehört heutzutage Software, die textuelle Daten mit hoher Geschwindigkeit durchsuchen und dynamisch sortieren sowie als KWIC-Konkordanz darstellen kann, zur Basisausstattung des philologischen Werkzeugkastens. Das Verfahren der Erstellung einer Konkordanz ist von der Schaffung einer physischen Repräsentation einer Konkordanzdarstellung ausgewählter Text in Papierform in programmatische Implementierungen automatischer Suchalgorithmen und textueller Visualisierung übergegangen, der wiederum Philolog*innen in die Lage versetzt, das Verfahren selbst dynamisch auf immer neue Texte anzuwenden. Anwendung findet es sowohl in der Sprach- als auch in der Literaturwissenschaft. Ähnlich verhält es sich auch mit der Annotation: seit der Frühzeit der Philologie wird sie eingesetzt, um Texte und andere sprachlich verfasste Artefakte mit zusätzlicher Information anzureichern. Die Philologin oder der Philologe entwickelt, beschreibt und wendet ein Inventar an Kategorien auf den Text an und kann auf dieser Basis ihre Analysen vorantreiben. Doch im digitalen Zeitalter umfasst der Werkzeugkasten nicht mehr Stift und Papier, sondern einen Editor, das Inventar an Kategorien und seine Beschreibung werden durch die Software bereitgehalten und auf digitale Artefakte angewendet. Auch hier ist eine Automatisierung erfolgt, die einerseits auf Basis von Suchmustern das schnelle Durchsuchen von Texten nach einem Oberflächenmuster (Wortformen, Wortsequenzen, Buchstabenfolgen etc.) sowie die Zuweisung der Annotationskategorien auf die entsprechenden Formen ermöglicht. Darüber hinaus kommen bereits seit mehr als 20 Jahren zunehmend Annotationsverfahren zum Einsatz, die mittels statistischer und probabilistischer Verfahren von Philolog*innen identifizierte Oberflächenmuster anhand ihrer Verteilung automatisch identifizieren und Kategorien zuweisen. Gut erprobte Beispiele sind die Wortartenannotation (part of speech tagging) sowie die syntaktische Analyse (parsing). Auf Grundlage statistischer Modellierung kookkurrierender lexikalischer Einheiten werden Kollokationen beschrieben, die sich anhand der spezifischen probabilistischen Verteilung der miteinander auftretenden Wortformen im Sprachgebrauch in großen digitalen Korpora identifizieren lassen (Bartsch 2004; Bartsch / Evert 2014). In jüngerer Zeit ist es durch Verfahren des maschinellen Lernens möglich geworden, auch weniger gut beschriebene Phänomene auf der Basis expertenannotierter Daten automatisch zu lernen und diese mittels sogenannter

überwachter maschineller Lernverfahren Kategorien zuzuweisen oder sogar mittels nicht-überwachter Lernverfahren ohne Experteninput zu identifizieren und Kategorien zu bilden. Verfahren zur stilometrischen Klassifikation mit dem Ziel der Autorschaftserkennung und Genderidentifikation von Autoren bzw. Autorinnen sind Beispiele aus der Literaturwissenschaft, die sich u. a. aus linguistischen Verfahren zur Klassifikation von Registern und Genres und vergleichbare Untersuchungen der Verteilung linguistischer Merkmale in großen Korpora ableiten und die auf literarische Gegenstände angewandt werden. Das maschinelle Lernen und besonders das sogenannte Deep Learning verspricht in den kommenden Jahren hier weitere Fortschritte zur Verbesserung bei der Suche nach Merkmalen in großen Korpora.

So sind Verfahren der Mustersuche, Sortierung, Indizierung und Darstellung in Form von Listen und Konkordanzen also ebenso wie die Annotation traditionelle und auch bereits prä-digital gut etablierte Methoden der Philologie, die durch die Transformation in das digitale Zeitalter bereichert und erweitert werden konnten. Andere Verfahren, wie die beispielhaft genannten Verfahren zur probabilistisch basierten Wortartenannotation oder zur Klassifikation mittels maschineller Lernverfahren, beruhen zum Teil ebenfalls auf lange existierenden Hypothesen, wie etwa Halliday (1978) probabilistische Definition linguistischer Register, die die Untersuchung der Verteilung linguistischer Merkmale in großen Korpora fordert, bevor dies in der heute möglichen Art und Weise denkbar war.

Digitale Werkzeuge unterstützen nicht nur die Anwendung einer bestimmten Methode, sie inspirieren gleichzeitig auch zu bestimmten Vorgehensweisen. Ein Werkzeug, das auf die manuelle Annotation mit nutzerdefinierten Kategorien ausgelegt ist, führt zu anderen Ergebnissen als ein Werkzeug, das zur algorithmischen Aggregation bestimmter Textelemente dient und diese Aggregationen statistisch auswertet.

Das Interesse am Einsatz digitaler Werkzeuge in geisteswissenschaftlichen Forschungsprojekten setzt keine speziellen Kenntnisse im digitalen Bereich voraus. Allerdings stößt dieses reine Interesse schnell an Grenzen: auf der einen Seite ist die Wahl des richtigen Werkzeuges kaum noch möglich, sobald die wahren Ausmaße des Möglichen erfasst sind. Und auf der anderen Seite kann der bzw. die (noch) nicht digitale Geisteswissenschaftler bzw. -wissenschaftlerin schnell in Schwierigkeiten geraten, wenn er oder sie diese Werkzeuge anwendet und z. B. mit absoluten Frequenzen argumentiert, wo nur relative Frequenzen aufschlussreich sind.

Aus den zahllosen Tools für den eher vagen Bereich der “Digital Humanities” genau das herauszupicken, was diese Anforderungen am besten erfüllt, ist jedoch eine zeitaufwendige Aufgabe. Portale wie vormals DiRT ((Digital Research Tools) Directory, CenterNet Initiative, 2011-2017; Cordell / Dombrowski 2011), das in die aktuelle Version von TAPoR3 (Text Analysis Por-

tal for Research: tapor.ca) integriert wurde, geben zwar einen sehr hilfreichen Überblick darüber, welche digitalen Werkzeuge zur Verfügung stehen. Doch trotz der hilfreichen Kategorisierungen und Beschreibungen fällt es häufig schwer, genau das richtige Werkzeug bzw. Werkzeugbündel für das eigene Projekt herauszufiltern und in einem geeigneten Workflow bzw. einer geeigneten Pipeline zu der Forschungsfrage entsprechend zu organisieren. Dies führt dazu, dass ein großer Teil der Laufzeit eines Projekts damit verbracht werden kann, Werkzeuge daraufhin zu testen, ob sie sich für das eigene Projekt eignen.

Daher wäre ein Desiderat eine ständige Meta-Studie zum Einsatz von digitalen Werkzeugen in geisteswissenschaftlichen Forschungsprojekten, die zeigt, welche Fragestellungen, Theorien, Modelle, Methoden und Ergebnisse mit welchen digitalen Werkzeugen bearbeitet wurden. Die Grundlage für die Analyse sind Veröffentlichungen, also Aufsätze, Sammelbandbeiträge usw. Die digitalen Werkzeuge werden dabei nach Arbeitsphasen getrennt kategorisiert, sodass unterschieden wird zwischen Werkzeugen für Preprocessing, Analyse, Korpusverwaltung, Ergebnisdarstellung (inkl. Visualisierung). Die Kategorisierungen werden dabei auf bestehende Taxonomien wie z. B. TaDiRAH (tadirah.dariah.eu) abgestimmt bzw. es werden entsprechende Matching-Tabellen erstellt.

Die Kategorisierungen sollten als Erweiterung zu TAPoR3 und anderen ähnlichen Portalen online zugänglich gemacht werden und würden es erlauben, jedes Projektelement – also Fragestellung, Theorien, Modelle usw. – als Zugang für die Suche nach digitalen Werkzeugen zu nutzen. Einer der großen Vorteile eines solchen Portals wäre es, dass die gefundenen Werkzeuge nicht nur aufgrund eines Handbuchs beurteilt werden können, sondern aufgrund eines Aufsatzes oder Sammelbandbeitrags.

Eine mögliche Erweiterung eines solchen Portals bestünde darin, traditionelle Arbeiten auf die gleiche Weise zu kategorisieren, um so graduell Unterschiede und Gemeinsamkeiten zwischen traditionell und digital ausgerichteten Projekten herauszuarbeiten. Dies würde einerseits einen Blick auf die historische Genese von Methoden und Theorien und eine fortlaufende Bestandsaufnahme der Praktiken der Digital Humanities ermöglichen, andererseits aber auch eine vergleichende Betrachtung solcher Praktiken über Disziplinengrenzen hinweg ermöglichen. Dies wäre vor allem im Hinblick auf den Austausch zwischen den philologischen Teildisziplinen Linguistik und Literaturwissenschaft sowie den interdisziplinären Austausch mit benachbarten geistes- und sozialwissenschaftlichen Disziplinen und mit der Informatik wünschenswert, da oftmals übersehen wird, dass Verfahren und Praktiken in anderen Forschungsbereichen existieren, die sich in der interdisziplinären Zusammenarbeit und über Disziplinengrenzen hinweg nutzen lassen.

3.3 Disziplinäre Verankerung der Methoden

Dass die Philologien schon von ihrer historischen Genese her eine hohe Methodensensibilität mitbringen und von jeher methodenbewusst und methodenstark sind, zeigt sich an der Entwicklung und Etablierung von Verfahren zur vergleichenden Untersuchung sprachlicher Merkmale ebenso wie an der Entwicklung von Verfahren und Methoden zur Sortierung, Indizierung und Visualisierung sprachlicher Daten. Das wohl bekannteste Beispiel hierfür ist die KWIC-Konkordanz, die eine lange Genese aus dem analogen in das digitale Zeitalter hinter sich gebracht hat und noch immer in vielfältigen Kontexten der Untersuchung linguistischer Daten selbstverständlich gebraucht wird (s. o.). Doch gerade die Genese des Prozesses der Konkordanzerstellung ist auch ein gutes Beispiel dafür, wie Prioritäten und Forschungsfragen sich verschieben, wenn man vom Analogen ins Digitale wechselt. Lag der Fokus in gedruckten Konkordanzen auf der Indizierung und damit Zugänglichkeit von Belegstellen in Texten, wie der Bibel oder den Werken William Shakespeares, so sind digitale Konkordanzprogramme dynamische Werkzeuge, die schnell große und flexibel zusammenstellbare Korpora immer wieder neu und rasch durchsuchen, sortieren und filtern können und deren Einsatz mehr und mehr zu einer Vorstufe für Prozesse der Sortierung und Filterung sowie der statistischen Auswertung zum Beispiel zur Untersuchung der relativen Kontextverteilung lexikalischer Ausdrücke als Grundlage der Kollokationsforschung weiterentwickelt wurden.

Auch die statistische Modellierung der Verteilung von Merkmalen in Texten, die im Analogen Grundlage der Beschreibung typologischer und historischer Unterschiede zwischen Sprachen und Sprachstufen ist und auf vielfältige Art und Weise genutzt wurde, um Verteilungsmuster von Merkmalen in Sprachdaten zu beschreiben, wie zum Beispiel in Untersuchungen zu Autorschaftsfragen der Werke William Shakespeares oder Kaedings Voruntersuchungen eines elf Millionen Wortformen umfassenden Korpus des Deutschen, das Grundlage für die Stenographie bildete (Kaeding 1898). Im Zeitalter der modernen digitalen Computertechnologie und der digitalen Verfahren in den Philologien werden Verfahren zur Klassifikation und zum Clustering eingesetzt, um inhaltliche Bezüge und Ähnlichkeiten innerhalb großer Textsammlungen durch statistische Verfahren aus der Informatik herauszuarbeiten. Einige davon, wie beispielsweise das Topic Modeling (Blei / Lafferty 2009) genannte maschinelle Verfahren zur Klassifikation sogenannter Topics, also semantisch ähnlicher Wortgruppen in Textkorpora, sind aus der Informatik bereitwillig in die Philologien übernommen worden und erfreuen sich aufgrund ihrer relativen Einfachheit und Robustheit in der Anwendung auf große Korpora großer Beliebtheit. Jedoch zeigt sich auf den zweiten Blick auch eine der Herausforderungen, denen sich die Digital Humanities stellen müssen – und zwar die Frage der Interpretation der Ergebnisse als Funktion des Verständnisses der zugrundeliegenden Verfahren. Es ist nämlich nur indirekt der Fall, dass

sich zwischen plausibel anmutenden Gruppierungen von Wörtern in den Topics auch direkte semantische Beziehungen ausmachen lassen. Die semantische Beziehung resultiert in diesem Fall aus der Gruppierung von Wörtern ähnlicher statistischer Verteilung, nicht aber aus einer im linguistischen bzw. literaturwissenschaftlichen Sinne notwendigerweise belegbaren semantischen Relation. Das Verfahren an sich ist hierfür nicht zu kritisieren, da es nicht vorgibt, etwas zu tun, was es nicht tut. Blei & Lafferty (2009) und andere erläutern detailliert, was das Verfahren anbietet und was es nicht anbietet und es hat in vielen Bereichen, auch in der Philologie interessante Anwendungsszenarien gefunden, in denen es hilft, große Korpora zu erschließen. Für die Interpretation und Rückbindung der Ergebnisse an die Texte sind auf jeden Fall kontextsensitive Methoden wie eine Konkordanzanalyse gewichtiger Topicwörter mit heranzuziehen. Ein bleibendes Problem ist die Reproduzierbarkeit der Modelle und ihrer Daten. Die Interpretation, Reproduzierbarkeit und Rückbindung der Ergebnisse an die Texte zeigt aber auch Grenzen und vor allem Herausforderungen bei der Interpretation der resultierenden Daten auf. Es zeigt sich hier, dass einerseits Verfahren aus der Informatik sich sehr gut in die Philologien und andere textbasierte Wissenschaften überführen lassen, dass aber die Interpretation der Ergebnisse auch an die disziplinären Erwartungen und Gepflogenheiten rückgebunden und mit der notwendigen kritischen Distanz betrachtet werden muss.

Diese Beispiele mögen genügen, um aufzuzeigen, dass jede Disziplin Herrin ihrer Gegenstände und interpretatorischen Prozesse bleiben muss und dass Methoden und Algorithmen aus benachbarten Disziplinen zwar übernommen, dann aber mindestens aus der disziplinären Perspektive kritisch betrachtet und ggf. auch angepasst werden müssen. Eine reine (unreflektierte) Übernahme von Werkzeugen und Algorithmen aus der Informatik würde dazu führen, dass die Prozesse und damit die Ergebnisse immer nur den Anforderungen der “Spenderdisziplin” genügen. Die Digital Humanities sind also gefordert, die Bereitschaft und die notwendigen Kompetenzen zur kritischen Evaluation adoptierter Verfahren zu entwickeln, Werkzeuge ggf. für die eigenen Forschungsfragen zu adaptieren, aber vor allem auch eigene Algorithmen und Werkzeuge zu entwickeln. Täten sie dies nicht, würden sie immer nur die Sicht auf die Daten erlangen, die ein Werkzeug oder ein Algorithmus vorgibt. Auch und vor allem in diesem Sinne haben sich die Darmstädter Digital Humanities die fachnahe und fachimmanente Implementierung digitaler Verfahren und Methoden und die kritische und kompetente Auseinandersetzung mit eigenen und adaptierten Algorithmen und Methoden aus benachbarten Disziplinen in ihr Pflichtenheft geschrieben und zum grundlegenden und programmatischen Credo gemacht.

3.4 Anforderungen an die Digitalen Philologien

Um diesen selbstdefinierten Anforderungen gerecht zu werden, müssen die Digital Humanities, hier im Sinne der Digitalen Philologien, sich ihrer angestammten Anforderungen und Kompetenzen bewusst sein, aber sich auch der Entwicklung neuer Kompetenzprofile öffnen. Nur so sind Wissenschaftler*innen in den Digitalen Philologien in der Lage, ihre eigenen Verfahren und Methoden aus dem Kern ihrer fachlichen Identität heraus weiterzuentwickeln, interdisziplinär zu kooperieren, von benachbarten Disziplinen zu lernen und Ideengeber für andere Disziplinen zu sein und dabei gleichzeitig die eigenen Gegenstände konsequent weiterzuentwickeln.

Zur Erreichung dieser Ziele müssen die Daten und Workflows klar definiert, Forschungsfragen operationalisiert und Algorithmen und Methoden in geeigneter Weise implementiert werden. Die hierzu notwendigen Kompetenzen müssen in den Digital Philologies entwickelt und an Studierende und junge Forschende weitergegeben werden, so dass die Implementierungskompetenz in den geisteswissenschaftlichen Fächern und aus diesen heraus und nicht als permanente Anforderung an benachbarte Disziplinen, wie vor allem die Informatik, formuliert wird. Hier müssen mehrere Anforderungen erfüllt werden.

Erstens müssen die Daten in einer algorithmisch zugänglichen Form erfasst, gespeichert und archiviert werden. Zweitens müssen existierende und entlehnte Algorithmen auf ihre Tauglichkeit für die Verarbeitung der fachlichen, in unserem Fall konkret der philologischen Daten überprüft und ggf. angepasst oder aus der Disziplin heraus entwickelt werden. Methoden und Werkzeuge, die diese Algorithmen implementieren, müssen auf die Anforderungen der Forschungsfragen und den Einsatz in den Disziplinen zugeschnitten werden. Dies betrifft beispielsweise den Umgang mit entsprechenden Datentypen und -mengen, aber auch Fragen der Passung für philologische Forschungsfragen. Drittens muss die Ergebnissicherung und -interpretation den disziplinären Anforderungen entsprechen, die nicht permanent die Besonderheit der digitalen Herangehensweise an sich in den Vordergrund rücken, sondern den Erkenntnisgewinn durch die digitalen Verfahren und die Satisfaktionsfähigkeit innerhalb der Basisdisziplinen befördern, sichern und aufzeigen. Nur so tragen die Digital Philologies ihre Erkenntnisse und Verfahren in die Mitte ihrer Ausgangsdisziplinen. In diesem Sinne verfolgen die Darmstädter Digital Humanities konsequent die Zielsetzung der Entwicklung und Verortung digitaler Verfahren in der Mitte der Fächer.

Voraussetzung hierfür ist einerseits die sachliche und instrumentelle Ausstattung, die dies ermöglicht. In diesem Sinne verfügt das Darmstädter Institut über eine Ausstattung mit digitalen

Laboren, die das Discourse Lab und das LitLab, Rechnerpools für die Lehre sowie eine entsprechende Arbeitsplatzausstattung der Wissenschaftler*innen beinhaltet. Hinzu kommt eine aufgeschlossene institutionelle Umgebung, die die Bibliothek, das Rechenzentrum sowie die Einstellung der Universität umfasst. Weitere Voraussetzung ist die Ausbildung von Kompetenzen der Wissenschaftler*innen sowie vor allem des wissenschaftlichen Nachwuchses. Diese sich stetig weiterentwickelnden Kompetenzen verlangen eine fundierte digitalen und philologische (bzw. geisteswissenschaftliche) Aus- und Weiterbildung ebenso wie eine Einarbeitung in Verfahren und Methoden benachbarter und kooperierender Disziplinen, wie etwa der Geschichtswissenschaften, der Sozialwissenschaften und vor allem der Informatik. Dies ist elementare Voraussetzung für die kritische Beurteilung und avancierte Weiterentwicklung der eigenen Theorien und Methoden, so dass Geisteswissenschaftler*innen in die Lage versetzt werden, nicht nur Anwender*innen von Werkzeugen zu sein, sondern Entwickler*innen ihrer eigenen Methoden und Verfahren mindestens im fachlichen, häufig aber auch im technologischen Sinne sind. In diesem Sinne bildet das Institut für Sprach- und Literaturwissenschaft seit den frühen 2000er Jahren Studierende in digitalen Studiengängen, wie dem Master of Arts Linguistic and Literary Computing (seit 2006), dem Bachelor of Arts Digital Philology (seit 2015) und dem interdisziplinären Master of Arts Data and Discourse Studies (seit 2019), in philologischen Studienprogrammen auf Grundlage digitaler Verfahren und Methoden aus. Und schließlich hat das Institut seine inhaltliche Ausrichtung und personelle Ausstattung seit den frühen 2000ern konsequent auf sein digitales Profil ausgerichtet, so dass derzeit fünf der sieben Professuren sowie eine nicht-professorale Fachgebietsleitung und ein großes Team junger Forschender die digitale Forschung und Lehre betreiben und weiterentwickeln.

Dieser Abschnitt hat ein Schlaglicht auf die Methodengenesen in den Philologie(n) als typischerweise methodenstarke Wissenschaft geworfen. Er hat einerseits aufgezeigt, in welcher Tradition die Darmstädter Digital Humanities stehen und trägt andererseits zu Klärung bei, was notwendig ist, um in einen sinnvollen und kritischen interdisziplinären Austausch mit benachbarten Disziplinen zu treten. Dabei wurden nicht nur fachliche, sondern auch institutionelle Bedingungen in den Blick genommen.

4. Annotation

Annotation kann als Kulturtechnik angesehen werden, die sich – insbesondere vor dem Hintergrund der Digitalisierung – im Wandel befindet. Sie gewinnt dabei als wissenschaftliche Praktik unter digitalen Bedingungen generell an Bedeutung und spielt in der Digitalen Philologie an der Technischen Universität Darmstadt insofern eine besondere Rolle, als hier u. a. im Rahmen

des BMBF-Verbundprojekts DARIAH schwerpunktmäßig über fachwissenschaftliche Annotationen gearbeitet wird – auch im Sinne der Reflexion dieses Themas (vgl. Lordick et al. 2016) – und Annotation in vielen Projekten ein zentraler Aspekt ist. Das Spektrum an annotierten Gegenständen erweitert sich (beispielsweise um Materialität physischer Objekte, nicht-schriftliche Aspekte usw.), die Formen und Funktionen von Annotationen werden ebenfalls vielfältiger. Sie können der individuellen Erschließungsarbeit dienen, aber auch Teil von kollaborativen Verfahren sein und als Ergebnis veröffentlicht werden.

Annotation ist also einerseits als Prozess bzw. Methode und andererseits als Produkt in mehrerlei Hinsicht zentraler Bestandteil der digitalen Forschung – gerade in den philologischen Fachdisziplinen: In Form von Auszeichnungen und Anreicherungen von Korpora bildet Annotation die Grundlage für analytische Zugänge zu Forschungsobjekten, trägt als kollaborative Interpretations-, Operationalisierungs- und Wissensaushandlungspraktik zur Weiterentwicklung von Theorien und Modellen bei und stellt eine zusätzliche textuell-diskursive Ebene dar, auf der Untersuchungsgegenstände kontextualisiert und intertextuell verknüpft werden können. In diesen drei Funktionen werden digitale Annotationswerkzeuge und -verfahren als wichtige Bedarfsaspekte und als Mehrwert-Faktoren des digital gestützten Arbeitens beschrieben (vgl. Bender 2016a); dabei werden konkrete Annotations-Projekte unterschiedlicher Zielsetzung als Ausgangspunkt der Überlegungen und zur Veranschaulichung herangezogen.

4.1 Annotation als Auszeichnung, Anreicherung und Analysegrundlage

Grundlage zahlreicher Projekte der Digitalen Philologie in Darmstadt sind strukturelle Annotationen in einer Auszeichnungssprache wie XML (eXtensible Markup-Language) nach den Standards der TEI (Text Encoding Initiative) sowie die Vergabe von Metadaten, hauptsächlich in digitalen Korpora, Sammlungen, Editionen und Wörterbüchern. Diese Verfahren unterstützen die hermeneutische Analyse, dienen der standardisierten Erschließung und Archivierung sowie der Nachnutzbarkeit, können aber gleichzeitig zur (automatisierten) Analyse genutzt werden. Sowohl die Auszeichnung z. B. lexikalischer und grammatischer Einheiten als auch die Kennzeichnung bestimmter semantischer bzw. inhaltlicher oder nach bestimmten Kriterien interpretierter Phänomene – in unterschiedlicher Tiefe und Komplexität – können die Basis automatisierter Verarbeitungsprozesse bilden. In den Darmstädter Projekten ist dies ein zentraler Aspekt.

So werden im Projekt “Digitale Zugänge zu fachwissenschaftlichen Heuristiken im akademischen Diskurs (DIGIF)” sprachliche Muster und ihr Vorkommen in wissenschaftlichen Arbeiten verschiedener Disziplinen annotiert und analysiert, die als Spuren heuristischer Praktiken – im

Sinne von komplexen Sprachhandlungen mit epistemischer, komplexitätsreduzierender Funktion – angesehen werden können. Ziel ist es, ein besseres Verständnis der verschiedenen Fachkulturen zu erlangen. Neben einer linguistischen Basisannotation (Tokenisierung, Lemmatisierung, Part-of-Speech-Tagging) werden auf Satzebene Indikatoren für heuristische Sprachhandlungen getaggt – z. B. Relevanzmarkierungen im Sinne von Sachbeschreibungen und Diskursreferenzierungen, Zielformulierungen, Definitionen, Begründungen und Argumentationen – und hinsichtlich ihrer fachspezifischen Verteilung und musterhaften Kombination ausgewertet.

Als weiteres Beispiel kann das 2016 abgeschlossene Projekt “ePoetics. Korpuserschließung und Visualisierung deutschsprachiger Poetiken (1770-1960) für den *Algorithmic criticism*” angeführt werden, in dem Komponenten des theoretischen Metaphernbegriffs sowie die damit verbundenen Referenzstrukturen annotiert wurden, um den Diskurs um den Begriff und seine Entwicklung untersuchen zu können (vgl. Alscher / Bender 2016 und ePoetics im DTA: www.deutschestextarchiv.de/search/metadata?corpus=epoetics, siehe auch Abschnitt 5.3).

Die Detektion und das Clustern von Metaphern stand im Mittelpunkt der Projekte “Arabischer Frühling” (vgl. Núñez et al. 2017) und “Natur & Staat (1903-1911)” (vgl. Núñez et al. 2015). Die Projekte zielen auf die digitale Annotation und schließlich (automatische) Detektion von usuellen Metaphern und ihren spezifischen grammatischen Konstruktionsmustern sowie von kreativen Metaphern und Begriffsbildungen. Dadurch lassen sich in Medienkorpora textsortenspezifisch z. B. in Zeitungskommentaren aktuelle Wahrnehmungsweisen, akteurspezifische Wissenshorizonte und Vermittlungsstrategien eruieren (“Arabischer Frühling”) sowie Annotationsguidelines und -schemata entwickeln, die dem internen Sprachusus der (populär-)wissenschaftlichen, sozialdarwinistischen Untersuchungstexte um 1900 im “Natur & Staat”-Korpus gerecht werden.

Als weiteres Beispiel sei das Projektvorhaben “Ein narratologisches Annotationssystem für Märe und Novelle: Konzeption und Korpusauszeichnung” genannt, in dem literarische Texte auf zehn verschiedenen narratologischen Ebenen wie Fokalisierung, Figurenrede oder Figurenbezug ausgezeichnet werden sollen (vgl. Dimpel 2017).

Diese Projekte verfolgen die Zielsetzung, ein einheitlich ausgezeichnetes Korpus zu erstellen, das quantitative Auswertungen möglich macht und das sich auf konkrete oder multiple Fragestellungen hin auf verschiedenen Abstraktionsebenen untersuchen lässt.

4.2 Annotation als kollaborative Interpretations- und Theoriearbeit

Das Ziel der standardisierten Auszeichnung und der quantitativen Analyse steht jedoch nicht gleichermaßen in allen Annotationsprojekten im Zentrum: So setzen z. B. Meister et al. im Kontext der Hamburger Projekte wie *heureCLÉA* (vgl. Bögel et al. 2015) bei der Annotation von Zeitphänomenen gerade auf den Umstand, dass verschiedene Bearbeiter*innen Zeitphänomene unterschiedlich einstufen können: Zwar werden dort offensichtliche Fehler, die auf einem “falschen Verständnis der fraglichen Textstelle oder der Kategoriedefinition” beruhen, korrigiert (vgl. Gius & Jacke 2016: 169). Wird jedoch als Grund für widersprüchliche Annotationen textuelle Polyvalenz identifiziert, werden “die widersprüchlichen Annotationen als legitimiert verstanden” (ebd.). Eine solche Widersprüchlichkeit von mehreren Annotator*innen ist beim kollaborativen Annotieren dann zu vermeiden, wenn ein auswertbares Korpus erstellt werden soll. Hier wird ein Inter-Annotator-Agreement angestrebt; polyvalente Textpassagen müssen dann als solche ausgezeichnet werden, unsichere Aussagen ebenfalls.

Bei der Gestaltung der Prozesse und Qualitätsmaßstäbe des kollaborativen Annotierens müssen projektspezifische Erfordernisse berücksichtigt werden. In den Projekten können Phasen der unabhängigen Doppelannotation mit Phasen der gemeinsamen Annotationsarbeit alternieren. Die Doppelannotation kann aber auch der Kontrolle der Konsistenz sowohl der konkreten Auszeichnung als auch der Annotationsrichtlinien und damit der Kategorienbildung dienen. Die Kategorienbildung wird im Prozess der Annotation evaluiert und weiterentwickelt. In der Gruppe bilden sich Maßstäbe dafür aus. Ebenso, wie man sich bei Editionsprojekten auf Kollationsrichtlinien verständigen muss, ist es bei Annotationsprojekten nicht mit der Erstellung des Tagsets getan: Zentral ist die Erstellung und Weiterentwicklung von Annotations-Schemata und vor allem der entsprechenden Guidelines, die oft eine recht ausführliche Beschreibung der Konzepte und Fallsammlungen zu bestimmten Phänomenen mitführen.

In den beschriebenen Projekten tritt bei der Entscheidung, welche Modelle und Kategorien der Annotation zugrunde gelegt werden, zu Tage, dass fachlich geläufige Modelle und Kategorien nicht immer trennscharf sind. Meist muss zuerst eine exaktere Modellbildung, die digital operationalisierbar ist, vorgenommen werden. Im Zuge dessen werden Schemata bzw. Taxonomien entwickelt, die Datenmodellierungen darstellen (vgl. McCarty 2004: 254–271) und zugleich Forschungsfragen und grundlegende Konzepte im Hinblick auf den Forschungsgegenstand repräsentieren. Dabei zeigen sich schon vor der informationstechnischen Auswertung – mit dem digitalen, algorithmischen Paradigma als Folie – Erkenntnismöglichkeiten und -grenzen im kollaborativ-diskursiven Annotationsprozess. Hermeneutische und algorithmische Verfahren können in Wechselwirkung treten. Annotator*innen gewinnen dabei nicht nur Erkenntnisse über den Gegenstand selbst, es werden auch die zugrundeliegenden Kategorien überprüft, präzisiert

und weiterentwickelt, so dass eine Rückkopplung der digitalen Annotation an den konventionellen Fachdiskurs erfolgen kann. Während des Annotationsworkflows und der Weiterentwicklung der Annotationsguidelines wird außerdem der Annotationsprozess selbst zum Objekt der Reflexion (Autoreflexivität etwa in methodischer Hinsicht).

Voraussetzung für die beschriebenen kollaborativen und diskursiven Aushandlungsprozesse im Zuge der Annotation und Guideline-Entwicklung ist eine virtuelle Forschungsumgebung, die die Organisation und Umsetzung dieser Verfahren unterstützt, die also die kollaborative Erstellung von Schemata erlaubt und idealerweise auch ein Nutzerrollenmanagement sowie Speicher-, Versionierungs- und Freigabe-Optionen auch auf der Annotationsebene bietet (vgl. Bender 2016a).

Im oben genannten Narratologie-Projekt, das verschiedene narratologische Ebenen annotiert, tritt bei der Entscheidung, welche Modelle und Kategorien der Annotation zugrunde gelegt werden, zu Tage, dass fachlich geläufige Modelle und Kategorien nicht immer trennscharf sind; gerade narratologische Kategorien sind oft – recht kontraintuitiv! – derart unscharf gefasst, dass bei einer narratologischen Annotation meist zuerst eine exaktere Modellbildung, die digital operationalisierbar ist, vorgenommen werden muss. Gerade hier ist die Weiterentwicklung von Annotationsguidelines, die oft eine recht ausführliche Beschreibung der Konzepte und gar Fall-sammlungen zu bestimmten Phänomenen mitführen, wichtig. Beim Annotationsvorgang sind stets Entscheidungen nötig, während man bei einer konventionellen Textanalyse nicht jeden einzelnen textuellen Sachverhalt in Rubriken einordnen muss. Diese Entscheidungen können zwar auch als subjektive Setzungen betrachtet werden, allerdings lässt sich die Subjektivitätsproblematik dadurch erheblich begrenzen, dass Setzungen möglichst in die Annotationsguidelines delegiert werden. Auch in annotationsfernen Wissenschaftsbereichen sind Setzungen nötig – wie etwa expliziert man ein Wort? Konzeptualisiert man ‘Fokalisierung’ nach Genette oder nach Bal? Nicht nur solche unausweichlichen Setzungen können in den Annotationsguidelines vorgenommen werden, vielmehr können hier auch Grenzfälle beschrieben werden, unter welchen Bedingungen ein Phänomen nicht mehr als zur Kategorie A, sondern zu B rubriziert wird.

4.3 Annotation als Kontextualisierung

Annotation lässt sich in mehrerlei Hinsicht als Kontextualisierung begreifen.

Erstens stellen Annotationen, die als Grundlage für Analysen erstellt wurden, eine metatextuelle Ebene dar, die aus einer bestimmten Perspektive im Zusammenhang mit einem bestimmten Forschungskontext erstellt wurde, aber auch Anknüpfungspunkte für Untersuchungen aus anderen Blickwinkeln bieten kann.

Zweitens können Annotationsschemata und die damit verknüpften Guidelines als Explikationen einer Forschungsperspektive mit dem entsprechenden Kontext angesehen werden. Einerseits werden dabei abstrakte Konzepte und Aspekte von Forschungsfragen operationalisiert, andererseits können Richtlinien anhand von konkreten Fallbeispielen diskursiv ausgehandelt werden.

Drittens gehen die Inhalte von Annotationen in einigen Projekten über das reine Tagging im Sinne einer Zuweisung von Kategorien eines Tagsets bzw. Annotationsschemas hinaus und umfassen beispielsweise Kommentierungen und Verknüpfungen mit Primär- und Sekundärquellen oder andere intertextuelle Verlinkungen. So wurden im Projekt ePoetics einerseits quantitativ auswertbare Einordnungen in bestimmte taxonomische Kategorien annotiert, die Textebenen und Referenzstrukturen der Poetik sowie die Explikation des sprach- und literaturtheoretischen Begriffs der Metapher umfassen, andererseits aber auch Angaben zur zitierten Primärliteratur mit einem Link zum Primärtext im TextGrid-Repositorium (vgl. Alscher / Bender 2016).

Dieses Beispiel zeigt zumindest ansatzweise, dass Annotationen auch Ergebnisse wissenschaftlichen Arbeitens im Sinne von Mikropublikationen darstellen können, z. B. als eine Art Stellenkommentar (vgl. Bradley 2012, Lordick et al. 2016). Es besteht also ein Zusammenhang zwischen Annotation, intertextuellen Verknüpfungen und hypertextueller bzw. -medialer Textproduktion. Außerdem ist Annotation im Bezug zu digitalen Repositorien, Editionen und Korpora zu sehen, an die auf der Annotationsebene angeknüpft werden kann (vgl. Bender 2016b). Hier wird erneut das Potenzial virtueller Forschungsumgebungen sichtbar, die nicht nur im Sinne einer kollaborativen Arbeitsumgebung bzw. eines digitalen Labors, sondern auch als Repositorium einen Raum mit entsprechenden technischen, infrastrukturellen und inhaltlichen Ressourcen bieten (vgl. Bender 2016a).

5. Thesauri, Taxonomien und Ontologien

Ein verbindendes Element philologischer Forschungsinteressen besteht in der Textualität ihrer Ressourcen. Textuell ist dabei nicht nur das Ausgangsmaterial, auch das Vorgehen zu seiner digitalen Erschließung und seiner Analyse erfolgt in der Regel textbasiert. In der Praxis können Forschungsdaten von unterschiedlichen Wissenschaftler*innen entsprechend ihrer individuellen und heterogenen Forschungsinteressen herangezogen werden. Mit dem TEI/XML-Standard hat sich in der digitalen Philologie eine Auszeichnungssprache etabliert, die einen übergreifenden Zugang auf erschlossene Texte ermöglicht. In dem Bereich der Metadaten ist die TEI-Auszeichnung stärker standardisiert als bei den Elementen, die innerhalb eines Textes zu seiner Annotation herangezogen werden. Diese Diskrepanz ist eine bewusst herbeigeführte, die es einerseits erlaubt, einen disziplinenübergreifenden philologischen Standard zu implementieren,

diesen andererseits jedoch so flexibel einsetzen zu können, dass wohlgeformte, aber maßgeschneiderte Lösungen entstehen.

Die philologischen Ressourcen – und zwar auch ihre digitalen Repräsentanten – sind damit in erster Linie menschenlesbar. Zwar bieten Metadaten und Auszeichnungssprache wie z. B. TEI zusätzliche (textuelle) Ebenen, auf denen implizit Informationen hinzukommen, diese Informationen lassen sich jedoch nicht wie bei RDF automatisch, sondern ausschließlich durch menschliche Rezipient*innen semantisch erfassen.

Die multiperspektivischen Forschungsinteressen an philologische und qualitativ angereicherte Daten sowie deren massives Anwachsen erfordern jedoch Zugänge, die formalisiert und maschinenlesbar sind. Beispielhaft ist der Webservice XTriples zu nennen, der dazu verwendet werden kann, aus verschiedensten Ressourcen im XML-Format RDF-Triples zu gewinnen. Die Explizierung des implizit bereits Vorhandenen führt schließlich einzelne Ressourcen zu einem komplexen Wissensnetz zusammen, was das Adressieren weiterer Forschungsfragen ermöglicht und somit das Potential ihrer Nachnutzung steigert. Auf diese Art wird es folglich möglich, individuelle Forschungsfragen an Datenbestände zu stellen, die für einen anderen Zweck aufbereitet wurden.

Im World Wide Web haben sich taxonomische Strukturen, insbesondere Thesauri und Ontologien, als das Mittel der Wahl zur Semantifizierung und Vernetzung von Datenbeständen etabliert. Taxonomien sind als übergeordnete Ordnungsstrukturen zu verstehen, die der Klassifikation und Konzeptualisierung dienen. Mit der Entwicklung von Thesauri und Ontologien wird ein Vokabular als gemeinsame Basis für die nachhaltige Erschließung und den Zugang zu Ressourcen zur Verfügung gestellt. Zudem lassen sich mit diesen Taxonomien nicht nur systematisch unterschiedliche Ressourcen, sondern auch Methoden bündeln.

Taxonomien zu erstellen und zu verwenden, bedeutet, sich mit dem eigenen Forschen intensiv auseinanderzusetzen. Das Verwenden konzeptueller Strukturierung in Bezug auf geisteswissenschaftliche Ressourcen ist bislang insbesondere im archivarischen, musealen und bibliothekarischen Bereich besonders ausgeprägt. Dabei erfährt gerade die forschungsgetriebene Datenmodellierung durch die Konzeptualisierung eine zusätzliche Aufwertung. Die konzeptualisierten Einheiten emanzipieren sich aus ihren spezifischen Domänen und werden disziplin- und sprachunabhängig adressier- und auffindbar. Sie werden zudem Teil eines Wissensnetzes, das für die integrierten Objekte neue Dimensionen der Kontextualisierung ermöglicht. Durch explizite Relationen zwischen einzelnen Entitäten lassen sich diesen automatisch weitere Informationen beifügen, die zu einem Erkenntnisgewinn beitragen können. Auch geht mit diesem Verfahren eine basale Disambiguierung einher – etwa wenn die polyseme Einheit ‘Jaguar’ als

Tier konzeptualisiert und damit spezifiziert ist. Die Koppelung von Forschungsdaten an Taxonomien hat somit das Potential, verschiedenartige Ressourcen derart miteinander zu verbinden, dass eine transparente, interoperable und nachhaltige Form des Wissensaustauschs und Erkenntnisgewinns möglich wird.

Mit der Anbindung an Taxonomien gehen auch zusätzliche – automatische – Annotationen einher, die komplexe Suchanfragen ermöglichen und darüber hinaus ein wichtiges Werkzeug zur Interoperabilität darstellen. Taxonomien sind ein wesentlicher Faktor der Grundlage einer nachhaltigen digitalen Forschungslandschaft. Um ihr Potential auszuschöpfen, sind darüber hinaus Schnittstellen und Werkzeuge wichtig, die die Nachnutzung befördern und einen interoperablen Zugriff ermöglichen. Zur Veranschaulichung wird im Folgenden dargelegt, wie Daten im Bereich der Digitalen Philologie grundlegend strukturiert werden können. Anschließend wird die *Taxonomy of Digital Research Activities in the Humanities* (TaDiRAH, s. tadi-rah.dariah.eu/) exemplarisch herangezogen, um die Rolle von Taxonomien für die Vernetzung von Forschungsdaten zu veranschaulichen.

Der TEI-Standard ist die führende Annotationsrichtlinie für die Digitale Philologie. Während mit dem Fortschreiten der technologischen Möglichkeiten zunehmend deduktive Methoden und Verfahren zum Erkenntnisgewinn eingesetzt werden, wird in den digitalen Geisteswissenschaften weiterhin eine nativ induktive Forschungsperspektive vermittelt. Geprägt von einem hermeneutischen Forschungsparadigma wird die Ressource induktiv strukturiert und formalisiert. Ein induktiver Ansatz beginnt mit der spezifischen Beobachtung und Suche nach Mustern oder Regelmäßigkeiten, um dann über Generalisierung allgemeine Theorien zu entwickeln und Schlussfolgerungen zu ziehen. Durch die aktive Gestaltung relevanter Daten bottom-up, ausgehend von Details, werden sinnvolle Konzepte und Beziehungen entwickelt, um eine möglichst genaue Beschreibung der Realität zu erhalten. Das induktiv erzeugte beschreibende Datenmodell dient dann als Ausgangslage für ein domänenspezifisches konzeptuelles Metamodell als Basis für die Überprüfung von Forschungsfragen.

An dieser Stelle ist es sinnvoll, auf Institutionen zu schauen, die bereits eine lange Tradition der Verwendung von Taxonomien, Metadatensätzen und der Anwendung standardisierter Regeln zur Beschreibung von Ressourcen haben. Hier wird im Gegensatz dazu meist ein top-down Ansatz verfolgt, angetrieben von der Suche nach einem kohärenten Metadatensatz, der als Standard die Integrität des domänenspezifischen Datenmodells gewährleistet. In diesem Zusammenhang funktionieren Ontologien, die als formale Konzeptmodelle bekannt sind, als Schlüsseltechnologie für die Interoperabilität und Integrität von Ressourcen und unterstützen einen verbesserten Zugang, fördern das Verständnis und ermöglichen die Analyse und Nachnutzung der Daten. Das bietet große Flexibilität in Bezug auf die Verwendung von Ressourcen

und ermöglicht die Erweiterbarkeit und die Kombination von verschiedenen Metadatenstandards.

Insbesondere das Conceptual Reference Model (CRM) ist eine sogenannte Top-Level-Ontologie für die Dokumentation und den Austausch von Daten im Bereich des Kulturerbes. Als ISO-Norm (21127:2014) gewinnt das CRM über die Archäologie hinaus zunehmend an Popularität und entwickelt sich zu einer disziplinübergreifenden lingua franca für virtuelle Forschungsumgebungen. Als Konzeptmodell stellt das CRM ein standardisiertes Vokabular zur Verfügung und ermöglicht gleichzeitig die Spezifizierung einer Domäne u. a. über die Konnektivität mit anderen Metadatenstandards, wie z. B. TEI.

Einen communitygetriebenen und anwendungsorientierten Ansatz verfolgt die o. g. Taxonomy of Digital Research Activities in the Humanities (TaDiRAH). Sie dient der Klassifikation von Digital-Humanities-Verfahren etwa in Publikationen, Werkzeugen oder Projekten und reichert diese mit Metainformationen an, die Vergleichbarkeit herstellen und Impulse zur Methodenreflexion liefern können. Die strukturierte Auszeichnung gemäß den darin angelegten Kategorien schafft nicht nur eine Schnittstelle für andere – auch disziplinenübergreifende – Anwendungen, sie leistet auch einen Beitrag zur transparenten Darstellung digitaler geisteswissenschaftlicher Forschung. TaDiRAH fungiert somit sowohl als Bindeglied zwischen Methoden und Ressourcen als auch als Indikator einer sich verändernden geisteswissenschaftlichen Forschungspraxis. Aufbauend auf vergleichsweise generischen Vorarbeiten zur Methodologie wie die Scholarly Primitives (Unsworth 2000) und die Methodological Commons (McCarty / Short 2002) lassen sich mit TaDiRAH alle Phasen des Forschens klassifizieren.

Der Prozess des Konzeptualisierens, Abwägens, Benennens und Präzisierens – insbesondere im kritischen Austausch mit Fachkolleg*innen – stellt an sich schon einen Mehrwert dar, weil er den Blick auf den eigenen Forschungsgegenstand schärft und neue Perspektiven schafft. Der hier skizzierte praxisorientierte Ansatz, der Anspruch auf eine relativ weitgefaste Anwendbarkeit erhebt, unterliegt dabei den Bedürfnissen einer sich verändernden Forschungskultur und kann daher niemals finalisiert werden. Die Taxonomie muss auf methodische Veränderungen in einem kollaborativen, reflexiven und vor allem iterativen Prozess reagieren, um weiterhin praxisorientiert anwendbar zu sein (vgl. Borek et al. 2016).

Ein Einsatzgebiet der Taxonomie, in dem sich aktuelle Entwicklungen der Digital Humanities ablesen lassen, ist die europäische DH-Course Registry von CLARIN und DARIAH-EU, die mithilfe von TaDiRAH systematisch Studienangebote aus dem DH-Bereich erfasst (registries.clarin-dariah.eu/courses/). Das Beispiel zeigt auch, dass Forschungsergebnisse – oder ganze Disziplinen – zudem nach außen nachvollziehbar werden, wenn Wissenschaftler*innen ihr Vorgehen an bestehende Taxonomien koppeln.

Wie groß der Nutzen ist, den strukturierte Daten erbringen können, ist stark davon abhängig, wie Daten organisiert und miteinander assoziiert sind. Während offene Standards längst existieren, um eine terminologische Interoperabilität zu gewährleisten, ist ihre philologische Verwendung unterschiedlich stark ausgeprägt. Es fehlt ein Zugang, der digitalen Geisteswissenschaftler*innen die Anbindung ihrer digitalen Forschungsdaten an übergeordnete Taxonomien erleichtert. Gleichzeitig bedarf es technischer Infrastrukturen, die mit Hilfe entsprechender Taxonomien disperse Datenbestände integrieren und als Linked (Open) Data zugänglich machen.

Die von Jenn Riley bereits 2010 erstellte Visualisierung bestehender Metadatenstandards (Riley 2010) gewichtet diese nach Verwendung und clustert sie nach Funktion, Zweck, Domäne und Institution. Zwar ist diese Darstellung nicht primär auf die Digitale Philologie ausgerichtet – hierfür müsste sie im besten Fall neu gedacht werden – bietet aber dennoch eine Orientierung zur Wahl eines passenden Meta-Datenmodells. Taxonomien wie TaDiRAH und die Aufstellung zu Metadaten von Riley zu erstellen und zu verwenden, sind zentrale Schritte, um einen nachhaltigen Umgang mit Forschungsgegenständen zu schaffen. Die Visualisierung der Metadatenstandards von Riley bietet interdisziplinär z. B. für Forscher*innen der Musik-, Bild- wie auch Textwissenschaften einen Überblick. Sie ist jedoch nicht spezifisch auf ihre Bedarfe ausgerichtet und bietet auch keine Verknüpfung zu Beispielprojekten, in denen diese abgebildeten Standards Verwendung finden. Beides sind Desiderate, deren Behebung die Verwendung von Taxonomien in den digitalen Geisteswissenschaften begünstigen können. Sobald eine kritische Masse an Forschungsdaten, Werkzeugen, Methoden und resultierenden Publikationen standardisiert miteinander verknüpft vorliegt, entsteht ein Wissensnetz, das die Aktivitäten der Community transparent abbildet und auf methodisches Vorgehen ausgewertet werden kann. Entsprechend ergibt sich aus dem Einsatz von Taxonomien auch die in Kapitel 3 geforderte transparente Nachvollziehbarkeit methodischer Vorgehensweisen, die ähnlichen Forschungsprojekten zur Orientierung dienen kann und diese Prozesse erstmals so visualisiert, dass sie miteinander vergleichbar werden.

Das Schaffen von niedrigschwelligen Angeboten zur Implementierung von Linked (Open) Data und die Veröffentlichung der eigenen Forschungsdaten als solche sind jedoch nur die ersten Schritte, um auch in den Philologien ein qualitatives Wissensnetz mitzugestalten. Während diese Schritte essentiell sind, um den bestehenden Bedarf anzuzeigen, können einzelne Forschende und Forschungsprojekte die für Linked (Open) Data erforderliche langfristige Perspektive nicht aus eigener Kraft gewährleisten. Damit resultierende Forschungsdaten als Linked (Open) Data vollumfänglich genutzt werden können, müssen übergeordnete infrastrukturelle

Lösungen etabliert werden. Exemplarisch sei hier auf ein Projekt der Universitäts- und Landesbibliothek (ULB) Darmstadt verwiesen: Im Zuge der Digitalisierung der zehnbändigen “Flora Graeca Sibthorpiana” hat eine Anreicherung mit Linked Open Data stattgefunden, die eine einladende und interaktive Präsentationsmöglichkeit erlaubt (Stille 2017). Auf diese Art können Nutzer*innen direkt mit dem Werk interagieren und zu weiterführenden Informationen gelangen. Die ULB generiert so eine Sichtbarkeit und effiziente Nutzbarkeit ihres Angebots und installiert gleichzeitig ein Pilotprojekt zur Konservierung von Linked (Open) Data.

Taxonomien sind ein solides Hilfsmittel, um Daten einheitlich zu organisieren und nachhaltig zugänglich zu machen. Insbesondere Ontologien bieten Beschreibungsmöglichkeiten, die natürliche Denkprozesse nachbilden und ein gemeinsames Verständnis von Konzepten einer Domäne fördern. Implementiert in einer universellen Beschreibungssprache wie RDF/XML wird so nicht nur eine Maschinenlesbarkeit, sondern auch Technologieunabhängigkeit erzielt, die eine disziplinübergreifende Nachnutzung der generierten Daten ermöglicht. Bei der Konzeptualisierung einer Domäne und der Entwicklung einer Ontologie handelt es sich um mehrfach iterative Prozesse, die einen gemeinsamen Konsens aller Teilhabenden einfordern und mit wachsendem Umfang nicht selten problembehaftet sind. Folglich tragen diese Prozesse auch dazu bei, Forschungsgegenstände und assoziierte Methoden immer wieder neu zu hinterfragen und ihr Verständnis im Zuge von Aushandlungsprozessen weiter zu schärfen.

Das Einbinden von Taxonomien in die digital arbeitenden Geisteswissenschaften wird nicht nur ihren Forschungsergebnissen zu mehr Sichtbarkeit verhelfen und sie in ein übergreifendes Wissensnetz integrieren – in ihm liegt auch eine Chance, den Forschungsprozess erstmals mit abzubilden und so einen Beitrag zu leisten für eine transparente und nachhaltige Publikationspraxis.

6. Denkstile

Ludwig Fleck begreift einen Denkstil als “gerichtetes Wahrnehmen, mit entsprechendem gedanklichen und sachlichen Verarbeiten des Wahrgenommenen” (Schäfer / Schnelle 1980: 130). Als richtungsweisend für einen Denkstil definiert er zudem die spezielle Fähigkeit, nicht nur “Sinn, Gestalt und geschlossene Einheit” wahrzunehmen, sondern gerade auch “der Gestalt Widersprechendes” zu sehen. (ebd.: 121). Ein Wandel des Denkstils vollziehe sich in enger Verbindung mit einem Wandel in der Wahrnehmung, der wiederum zu einer neuen Sichtweise führe. Fleck weist seine Leser außerdem darauf hin, dass dieses neue Sehen in der Wissenschaft erst gelernt werden müsse (vgl. ebd.: 41).

In den Wissenschaften, die mit digitalen Methoden und Werkzeugen für die Beantwortung statistisch relevanter Fragen arbeiten, sind Denkstile einerseits geprägt durch passive Wahrnehmung von Trends und Ausreißern in den generierten Daten. Zu neuen Erkenntnissen führt gerade die Detektion von Phänomenen, die der Gestalt der beobachteten Objekte widersprechen. Andererseits wird die Denkweise im Digitalen verstärkt bestimmt durch das aktive Visualisieren – eine Dimension, die auch bei Fleck schon anklingt, wenn er von ‘Sinn-Bildern’ spricht und ausführt: “Wir haben Ideogramme vor uns, d. h. graphische Darstellung gewisser Ideen, gewissen Sinnes, einer Art des Begreifens. Der Sinn ist in ihnen dargestellt wie eine Eigenschaft des Abgebildeten.” (ebd.: 183. 186)

Der Denkstil innerhalb eines Wissenschaftsbereichs wird durch die Kommunikation beeinflusst und wirkt umgekehrt auf diese zurück, etwa indem in einer langen Serie von Übersetzungen die ursprünglichen Messdaten schrittweise in symbolische Darstellungen umgewandelt werden – “Tabellen, Bilder, Kurven, die (...) ihrerseits wieder auf andere Darstellungen verweisen.” (Heintz / Huber 2001: 12) Diese Funktion leitet wiederum über zur Operationalisierung im Digitalen, also zu der Frage, wie Hypothesen durch bestimmte Instrumente, Erhebungsmethoden und die Auswahl von Messgrößen überprüft werden können. Um zu ermitteln, in welcher Form sich ein für Digitale Philologien charakteristischer Denkstil äußern kann, erscheint es zunächst lohnend, einen Blick auf die grundlegenden Differenzen zwischen der traditionellen und der digitalen Arbeitsweise zu werfen, denn nur so werden die jeweils zugrundeliegenden Denkmotive evident.

In den Literaturwissenschaften, die hier beispielhaft herangezogen werden, zeichnet sich die Arbeit grob gesprochen durch das Lesen, das Verstehen und das Analysieren von Literatur aus. Im menschlichen Lesen äußert sich ein “verstehendes qualitativ-hermeneutisches ‘Lesen’”, das sich vom maschinellen Lesen, von einem algorithmusgesteuerten “sammelnd-suchende(n) quantitative(n) ‘Auslesen’” unterscheidet (vgl. Weitin 2015: 1). Dem Vorgang des menschlichen Lesens entspricht auf digitaler Seite das ‘Ein-Lesen’ der Informationen in die Maschine oder auch das ‘Aus-Lesen’ formaler Merkmale von Texten mit Hilfe von statistischen Verfahren. Das menschliche ‘Verstehen’ eines Textes kann wiederum dem ‘Erkennen’ eines Computers gegenübergestellt werden. Wo der Mensch den Inhalt eines Textes versteht, erkennt der Rechner (z. B. annotierten Text oder die Anweisung in einer Programmiersprache). Der durch den Menschen vorgenommenen ‘Analyse’ von Literatur entspricht schließlich die rechnergestützte ‘Suche’. In diesem Punkt treffen sich beide Arbeitsweisen grosso modo, denn genau wie der Mensch sucht die Maschine nach Mustern, die ihr bereits bekannt sind. Werden qualitativ-hermeneutische und quantitativ-statistische Methoden verbunden, so integriert die Arbeitsweise menschliches und maschinelles Lesen und Analysieren gleichermaßen (vgl. ebd.: 9).

Im letzten Abschnitt wurde deutlich, dass wir es in der analogen und der digital betriebenen Philologie mit unterschiedlichen Arbeitsgängen zu tun haben. Auch die zugrundeliegenden Denkmodelle differieren. Der Denkstil in den Wissenschaften, die mit qualitativ-hermeneutischen Verfahren befasst sind, ist in aller Regel stark verifikationsorientiert, das heißt, es geht darum, eigene Thesen, die an die behandelte(n) Quelle(n) gerichtet sind, in einer Untersuchung zu bestätigen. Ein Negativergebnis hinterlässt dann eher den fragwürdigen Eindruck, dass die Studie schon bei der Wahl des Themas nur ungenügend durchdacht worden sei. Durch digitale Methoden scheint diese traditionelle Denkweise aufgebrochen zu werden, denn quantitative Forschungsansätze, die streng an Operationalisierung gebunden sind, machen ein prinzipielles Umdenken notwendig. Endet die Analyse mit einem negativen Ergebnis, so muss auch dieses akzeptiert werden, d. h. auch eine falsifizierte Theorie ist als forschungsrelevant bzw. wertvoll für die Forschung anzusehen. Dasselbe gilt für die Arbeitsweise in digitalen Fächern, die im Gegensatz zur traditionellen Arbeitsweise geprägt ist durch kollaborative Laborforschung mit flachen Hierarchien, und in denen grundsätzlich nicht zwischen Haupt- und Hilfsdisziplinen unterschieden wird.

Traditionelle Denkstile werden auch umgewälzt, wenn es um die Wahrnehmung der Ressourcen geht, an denen geforscht wird. Auch hier spielt die Operationalisierung wieder eine entscheidende Rolle, wenn man diese als Übersetzung einer Fragestellung an ein kulturelles Objekt in elementare Merkmale, Formen und Strukturen begreift, die am Ende gezählt und in Visualisierungen sichtbar gemacht werden können (vgl. ebd.: 11). Für die quantitative Analyse muss ein Text zunächst in Struktureinheiten aufgeteilt werden und diese müssen dann in einem zuvor festgelegten Prozedere geordnet, also modelliert werden, bevor sie wiederum rechnerisch gezählt werden können. Damit ähnelt die digitale Methodik einer rein strukturalistischen Tätigkeit, welche die Sicht auf den Forschungsgegenstand im Vergleich zu den qualitativ-hermeneutischen Verfahren entscheidend verändert. Das untersuchte Objekt – sei es ein ganzer Text, ein Satz, ein Wort, ein Buchstabe oder ein Abbild derselben – ist dann austauschbar und fungiert im Forschungsprozess nur noch als eine Art ‘Container’, der bei der Modellierung in einzelne Bestandteile zerlegt wird. Durch die Beliebigkeit der ins Digitale übertragenen Ressource geht mit der ihr inhärenten natürlichen Qualität zum einen etwas Entscheidendes verloren. Zum anderen bringt die herausgearbeitete formale Struktur aber auch etwas zum Vorschein, “das im natürlichen Objekt unsichtbar oder, wenn man lieber will, unverständlich bliebe” (vgl. Barthes 1966: 191). Die Wahrnehmung des Forschungsobjekts im Digitalen sollte also nochmals überdacht und neu definiert werden.

Wissenschaftliche Denkstile werden nicht zuletzt auch durch die verwendeten Werkzeuge beeinflusst. Dieser Frage ist beispielsweise Noah Bubenhofer nachgegangen und kommt in Anlehnung an Ludwik Fleck zu dem Schluss: “Mit der Wahl des Tools (Fleck spricht vom wissenschaftlichen Gerät) ordnen Sie sich einem wissenschaftlichen Denkstil unter und reproduzieren ihn wieder.” (vgl. Bubenhofer 2015). Unter die in den Digitalen Philologien gebräuchlichen Werkzeuge fallen Programmiersprachen, Algorithmen, Metadaten und Annotationsschemata, Visualisierungsformen und statistische Modelle. Wenn beispielsweise das Programmieren von Software zu einer neuen Kulturtechnik geworden ist und die Programmiersprache selbst Ausdruck kultureller Praxis, dann wirkt sich diese auch wieder auf die wissenschaftliche Praxis aus und formt die dahinterliegenden Denkstile. Unterschiedliche Annotationsformen und der Einsatz von Taxonomien und Ontologien bei der Aufbereitung der Daten haben einen Einfluss auf das ‘Denkgebäude’, in dem sich das wissenschaftliche Kollektiv (man könnte im Digitalen genauso von der Community sprechen) bewegt.

7. Fazit

Im vorliegenden Working Paper konnten wir beispielhaft zeigen, wie die verschiedenen, oben erwähnten Bereiche interagieren und die Denkstile in den digitalen Philologien prägen und beeinflussen können. Unsere Ausführungen konnten deutlich machen, dass sich ein Wandel des Denkstils grundsätzlich in enger Verbindung mit einem Wandel in der Wahrnehmung vollzieht. Die Ressource war der Ausgangspunkt unserer Überlegungen, aus denen deutlich geworden ist, wie sich der Zugang zu den Quellen, aber auch ihre Wahrnehmung im Zuge der Digitalisierung gewandelt hat. Zu den hauptsächlichen Erkenntnissen gehört, dass der Terminus ‘Ressource’ im digitalen Bereich viel umfassender verstanden werden kann als in den ‘traditionellen’ Geisteswissenschaften. Ressourcen lassen sich in drei wesentliche ‘Aggregatzustände’ einteilen: analoge Ressourcen, digitale Surrogate der analogen Quellen und sogenannte ‘born-digital objects’. Ergebnis ihrer Bearbeitung im Analogen wie im Digitalen sind Derivate, die sich beliebig vervielfachen lassen. Wesentliche Unterscheidungsmerkmale digitaler Surrogate gegenüber den analogen Ressourcen sind ihre Abstrahierung, ihre Modellierung, die daraus resultierende Reduktion von Komplexität und die ihnen eigentümliche Präsentation im digitalen Medium. Im Zuge ihrer digitalen Verarbeitung verlieren diese einige der Eigenschaften der Originalfassungen, gewinnen aber zugleich auch neue hinzu. Verknüpfungen zwischen Ressourcen im Sinne digital umgesetzter Intertextualität bzw. Hypertextualität können durch Annotationen hergestellt werden. Annotationsverfahren dienen der standardisierten Erschließung und Archivierung sowie der tiefenhermeneutischen wie auch der automatischen Analyse. Im Rahmen von

Annotationsprozessen findet außerdem die diskursive Aushandlung von Wissen auf einer zusätzlichen textuellen Ebene statt – in Form von Tags, deren Bedeutung und Differenzierung in Form von Annotationsschemata und -richtlinien entwickelt wird, aber auch in Form von Kommentierungen, Verlinkungen und Quellenangaben als Annotationsinhalte. Annotation kann also als Kulturtechnik und wissenschaftliche kollaborative Operationalisierungs-, Interpretations- und Wissensaushandlungspraktik als Grundlage für analytische Zugänge zu Forschungsgegenständen angesehen werden. Im Zuge von Annotationsverfahren und der Weiterentwicklung von Schemata und Guidelines wird darüber hinaus der Annotationsprozess selbst zum Objekt der Reflexion – auch im Hinblick auf Denkstile.

Daran anknüpfend haben wir festgestellt, dass die Entwicklung von Methoden eng an die Erfordernisse von Ressourcen gekoppelt ist. Die Modellierung von Arbeitsabläufen anhand der digitalen Ressourcen kann durch die Methodengense in wesentlichen Punkten Unterstützung erfahren, beispielsweise beim Annotieren. Untrennbar mit dem Annotieren verbunden sind wiederum übergreifende Ordnungsstrukturen (Taxonomien), die einen systematischen Umgang mit Ressourcen erlauben und zugleich als Basis für die angewandte wissenschaftliche Methode gelten können. Die Konzeption von Taxonomien schärft den Blick des bzw. der forschenden Philologen bzw. Philologin für seine bzw. ihre Textinhalte und verändert zugleich die Sicht auf den betrachteten Forschungsgegenstand.

Literatur

Letzter Zugriff für alle genannten Internetreferenzen: 20.05.2020.

Alscher, Stefan; Bender, Michael (2016): Auf der Suche nach dem “goldnen Baum”. Digitale Annotation als Erkenntnisprozess und ‘tertium comparationis’ – am Beispiel der Auszeichnung des Metaphernbegriffs in Poetiken. In: Zeitschrift für digitale Geisteswissenschaften. DOI: 10.17175/2016_004.

Barthes, Roland (1966): Die strukturalistische Tätigkeit. In: Kursbuch 5, 190-196.

Bartsch, Sabine (2004): Structural and Functional Properties of Collocations in English. Tübingen: Verlag Gunther Narr.

Bartsch, Sabine; Evert, Stefan (2014): Towards a Firthian notion of collocation. In: Vernetzungsstrategien, Zugriffsstrukturen und automatisch ermittelte Angaben in Internetwörterbüchern. 2. Arbeitsbericht des wissenschaftlichen Netzwerks “Internetlexikografie” / Network Strategies, Access Structures and Automatic Extraction of Lexicographical Information. 2nd Work Report of the Academic Network “Internet Lexicography”, hrsg. von Andrea Abel und Lothar Lemnitzer. Mannheim: Institut für Deutsche Sprache (OPAL – Online publizierte Arbeiten zur Linguistik 2/2014: ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/2402/file/Abel_Lemnitzer_Vernetzungsstrategien_2014.pdf).

Bender, Michael (2016a): Forschungsumgebungen in den Digital Humanities: Nutzerbedarf, Wissenstransfer, Textualität (Sprache und Wissen 22). Berlin, Boston: de Gruyter.

Bender, Michael (2016b): Digitale Methoden und Kulturtechniken. In: Handbuch soziale Praktiken und digitale Alltagswelten, hrsg. von Heidrun Friese, Gala Rebane, Marcus Nolden, Miriam Schreiter. Wiesbaden: Springer. DOI: 10.1007/978-3-658-08460-8_48-1.

Blei, David M.; Lafferty, John D. (2009): Topic Models. In: Text Mining: Classification, Clustering, and Applications, eds. Ashok N. Srivastava and Mehran Sahami (Chapman & Hall / CRC Data Mining and Knowledge Discovery Series). Boca Raton / London, New York: Taylor and Francis Group. www.cs.columbia.edu/~blei/papers/BleiLafferty2009.pdf.

Borek, Luise; Dombrowski, Quinn; Perkins, Jody; Schöch, Christof (2016): TaDiRAH – A Case Study in Pragmatic Classification. In: Digital Humanities Quarterly (DHQ) 10(1), hrsg. von James O’Sullivan, Mary Galvin und Diane Jakacki. www.digitalhumanities.org/dhq/vol/10/1/000235/000235.html.

Bögel, Thomas; Gertz, Michael; Gius, Evelyn; Jacke, Janina; Meister, Jan Christoph; Petris, Marco; Strötgen, Jannik (2015): Gleiche Textdaten, unterschiedliche Erkenntnisziele? Zum Potential vermeintlich widersprüchlicher Zugänge zu Textanalyse. In: DHd 2015: Digital Humanities im deutschsprachigen Raum. Graz. gams.uni-graz.at/o:dhd2015.abstracts-gesamt.

Bradley, John (2012): Towards a richer sense of digital annotation: Moving beyond a “Media” orientation of the annotation of digital objects. Digital Humanities Quarterly 6(2). www.digitalhumanities.org/dhq/vol/6/2/000121/000121.html.

Bubenhofner, Noah (2015): Coding Cultures: Über den Zusammenhang von Programmiersprachen und Denkstilen, In: Sprechtafel. Linguistische Notizen. www.bubenhofner.com/sprechtafel/2015/08/08/coding-cultures-ueber-den-zusammenhang-von-programmiersprachen-und-denkstilen/.

Cameron, Angus; Frank, Roberta; Leyerle, John (eds.) (1970): Computers and Old English Concordances. Toronto / Buffalo: University of Toronto Press.

Cordell, Ryan; Dombrowski, Quinn. 2011. Chicago Colloquium on Digital Humanities and Computer Science: poster, "DHCommons and Bamboo DiRT: integrating digital humanities directories": www.flickr.com/photos/quinnanya/6368652141/sizes/1

Deutsches Institut für Normung (Hg.) (1978): DIN 44300. (Informationsverarbeitung 1) Berlin.

Dimpel, Friedrich Michael (2017): Novellenschätze narratologisch auszeichnen und analysieren am Beispiel Victor von Scheffels ‘Hugideo’ und der sozialen Netzwerkanalyse. In: LiLi 47, 87–108 [zugleich Band zur Tagung “Scalable Reading. Paul Heyses Deutscher Novellenschatz zwischen Einzeltext und Makroanalyse”, 4.–6. Juni 2015 in Konstanz, veranstaltet von Thomas Weitin].

Do Dinh, Erik-Lân; Gerloff, Malte; Núñez, Alexandra (2016): Metaphern digital – Auf dem Weg von der Annotation zur automatischen Detektion. Leipzig: Konferenzabstracts DHd 2016. www.dhd2016.de/abstracts/posters-011.html.

Gärtner, Kurt; Kühn, Peter (1998): Indices und Konkordanzen zu historischen Texten des Deutschen. Bestandsaufnahme, Typen, Herstellungsprobleme, Benutzungsmöglichkeiten. In: Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung. 2. Auflage. 1. Halbband. Hrsg. von Werner Besch u. a. 2. Auflage. Berlin/New York: de Gruyter, 715-742.

Gehring, Petra; Rapp, Andrea (2018): Vordigitale und digitale Buchseite – der Text als Raum. In: Technik – Macht – Raum. Das topologische Manifest im Kontext interdisziplinärer Studien,

hrsg. von Andreas Brenneis, Oliver Honer, Sina Keesser, Annette Ripper und Silke Vetter-Schultze (Technikzukünfte, Wissenschaft und Gesellschaft / Futures of Technology, Science and Society). Wiesbaden: Springer, 281-292.

Gius, Evelyn; Jacke, Janina (2015): Informatik und Hermeneutik. Zum Mehrwert interdisziplinärer Textanalyse. In: Grenzen und Möglichkeiten der Digital Humanities, hrsg. von Constanze Baum und Thomas Stäcker (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 1) DOI: 10.17175/sb001_006.

Gius, Evelyn; Jacke, Janina (2016): Kollaboratives Annotieren literarischer Texte. In: DHd 2016 Modellierung, Vernetzung, Visualisierung: die Digital Humanities als fächerübergreifendes Forschungsparadigma: Konferenzabstracts, Leipzig: nisaba verlag, 167–170. www.dhd2016.de/abstracts/vortr%C3%A4ge-039.html.

Halliday, M.A.K. (1978): Language as social semiotic: The Social Interpretation of Language and Meaning. London: Edward Arnold.

Halliday, M.A.K. (1988): On the language of physical science. In: Registers of written English: situational factors and linguistic features, ed. Mohsen Ghadessy. London: Pinter Publishers, 162-178.

Heintz, Bettina; Huber, Jörg (2001): Der verführerische Blick: Formen und Folgen wissenschaftlicher Visualisierungsstrategien. In: Mit dem Auge denken, hrsg. von Bettina Heintz und Jörg Huber. Zürich: Ed. Voldemeer, 9-40.

ISO/IEC JTC 1, Information technology Committee. ISO/IEC 2382:2015(en): Information technology – Vocabulary. www.iso.org/standard/63598.html.

Kaeding, Friedrich Wilhelm (Hrsg.) (1898): Häufigkeitswörterbuch der deutschen Sprache: Festgestellt durch einen Arbeitsausschuss der deutschen Stenographiesysteme *Selbstverlag des Herausgebers; E.S. Mittler & Sohn*, 1898

Lordick, Harald; Becker, Rainer; Bender, Michael; Borek, Luise; Hastik, Canan; Kollatz, Thomas; Mache, Beata; Rapp, Andrea; Reiche, Ruth; Walkowski, Niels-Oliver (2016): Digitale Annotationen in der geisteswissenschaftlichen Praxis. In: Bibliothek. Forschung und Praxis 40/2, 186-199.

McCarty, Willard (2004): Modelling. A study in words and meanings. In: A companion to digital humanities, hrsg. von Susan Schreibman, Ray Siemens und John Unsworth, Oxford: Blackwell, 254–271. www.digitalhumanities.org/companion/.

McCarty, Willard; Short, Harold (2002): Mapping the Field. Bericht des EADH-Meetings in Pisa. eadh.org/mapping-field.

Núñez, Alexandra; Gerloff, Malte; Do Dinh, Erik-Lân; Rapp, Andrea; Gehring, Petra; Gurevych, Iryna (2017): A “Wind of Change” – Shaping Public Opinion of the "Arab Spring" Using Metaphors. In: Digital Humanities 2017. Conference Abstracts. McGill University & Université de Montréal, Montréal, Canada. August 8–11. dh2017.adho.org/abstracts/041/041.pdf.

Núñez, Alexandra; Gerloff, Malte; Do Dinh, Erik-Lân; Rapp, Andrea; Gehring, Petra; Gurevych, Iryna (2015): Natur und Staat: Automatische Metapherndetektion im biopolitischen Diskurs des 20. Jahrhunderts. Berlin: Poster presentation at the DH Summit 2015, Berlin. download.hrz.tu-darmstadt.de/media/FB20/Dekanat/Publikationen/KDSL/dh-summit-2015-nus_final.pdf.

Núñez, Alexandra (2014): Wenn das ‘Embodiment’ politisch wird: Das Image-Schema PATH und seine Realisierung im Mediendiskurs zum „Arabischen Frühling“. In: Cognitive Explorations into Metaphor and Metonymy, hrsg. von Frank Polzenhagen, Sonja Kleinke, Zoltán Kövecses und Stefanie Vogelbacher. Frankfurt a. M.: Peter Lang, 149-164.

Ore, Christian Emil; Doerr, Martin; Le Bœuf, Patrick; Stead, Stephen (Hgg.) (2017): Definition of the CIDOC Conceptual Reference Model. Version 6.2.2 (September 2017). www.cidoc-crm.org/Version/version-6.2.2.

RDF Schema 1.1. (2014). www.w3.org/TR/rdf-schema/.

Riley, Jenn (2010): Seeing Standards: A Visualization of the Metadata Universe. jenn-riley.com/metadatamap/.

Schäfer, Lothar; Schnelle, Thomas (Hgg.) (1980): Ludwik Fleck: Entstehung und Entwicklung einer wissenschaftlichen Tatsache. Einführung in die Lehre vom Denkstil und Denkkollektiv (= suhrkamp taschenbuch wissenschaft. Nr. 312). Frankfurt a. M.: Suhrkamp.

Schrade, Torsten (o. J.): XTriples. A generic webservice to extract RDF statements from XML resources. Version 1.3.0. Akademie der Wissenschaften und der Literatur, Mainz. xtriples.spatialhumanities.de/index.html.

Sinclair, John McHardy (1991): Corpus, Concordance, Collocation. Describing English Language Series. Oxford et al.: Oxford University Press.

Spevack, Marvin (1973): The Harvard concordance to Shakespeare. Cambridge, Massachusetts: Belknap Press of Harvard University Press.

Stille, Wolfgang (2017): Flora Graeca Digitalis: Altbestand trifft Linked Open Data und Mobiles Computing. In: Vom Sinn der Bibliotheken: Festschrift für Hans-Georg Nolte-Fischer, hrsg. von Irmgard Siebert und Dietmar Haubfleisch. Harrassowitz, Wiesbaden.

TAPoR3 (Text Analysis Portal for Research 3.0). tapor.ca

Thaller, Manfred (2017): Geschichte der Digital Humanities. In: Digital Humanities. Eine Einführung, hrsg. von Fotis Jannidis, Hubertus Kohle und Malte Rehbein. Stuttgart: Metzler, 3-12.

Weitin, Thomas (2015): Thinking slowly. Literatur lesen unter dem Eindruck von Big Data, LitLingLab Pamphlet #1. digitalhumanitiescenter.de/pamphlets/kl3-01_weitin-thinking-slowly.pdf.