Jan Christoph Meister

Geleitwort

Der vorliegende Band versammelt achtzehn Aufsätze, die – im ersten Abschnitt theoretische Aspekte behandelnd, im zweiten auf Grundsatzfragen der Methodenentwicklung fokussierend und im dritten anhand von Fallanalysen und Beispielen – um die Frage kreisen, welche Voraussetzungen eine theoretisch-methodologisch reflektierte Form der algorithmischen Textanalyse zu erfüllen hat. Das Buch präsentiert damit Ergebnisse der Forschungsarbeit eines interdisziplinären Forschungszentrums, das sich seit 2016 systematisch und im Modus eines stringent organisierten interdisziplinären Austauschs zwischen hermeneutisch-geisteswissenschaftlich orientierten Textdisziplinen, formalisierungsaffiner Computerlinguistik und datenmodellierender Informatik mit der Frage beschäftigt, wie praxeologische, methodologische und theoretisch-konzeptionelle Perspektiven in der Methodenentwicklung der Digital Humanities im Teilbereich der Textanalyse fruchtbar kombiniert werden können. Mit diesem ebenso ambitionierten wie klar fokussierten Ansatz hat CRETA, das Centrum für reflektierte Textanalyse an der Universität Stuttgart, mittlerweile national wie international ein herausragendes Profil gewonnen.

Die DH zählen heute, keine zehn Jahre nach der Gründung der Digital Humanities im deutschsprachigen Raum (DHd) im Jahr 2012, zum pluralistischen Methodenrepertoire der Geisteswissenschaften. Selbst von bekennenden Traditionalisten werden sie zunehmend als ein Methodenparadigma wahrgenommen, das den Geisteswissenschaften eine Option auf die Teilhabe am digitalen Zeitalter eröffnet, wenn nicht gar interessante methodische Impulse liefern könnte.

Was dabei allerdings in der Binnen- wie der Außenperspektive leicht in Vergessenheit gerät, ist, dass das neue Methodenparadigma seine gegenwärtige Kontur nicht erst Dank der strategischen Förderinitiativen der letzten zehn Jahre gewonnen hat, mit denen öffentliche und private Drittmittelgeber die Digital Humanities insbesondere im deutschsprachigen Raum entscheidend gefördert haben.¹

¹ Hervorzuheben sind hier im deutschen Kontext neben diversen föderalen und privaten Initiativen insbesondere die einschlägigen Förderlinien der Volkswagenstiftung, der Alexander-von-Humboldt-Stiftung und der Deutschen Forschungsgemeinschaft (DFG). Das Projekt CRETA verdankt seine Existenz der impulsgebenden eHumanities-Förderinitiative des Bundesministeriums für Bildung und Forschung (BMBF).

Jan Christoph Meister, Universität Hamburg

Der Erfolg der DH ist auch und nicht zuletzt der einer Forschungsidee, die im Verlauf von gut fünfzig Jahren Forschungsarbeit ihren Gegenstandsbereich wie ihre methodisch-theoretische Ausprägung einerseits sukzessive erweitert, dabei jedoch andererseits ihr interdisziplinäres Ethos bewahrt hat. In der Außenwahrnehmung der DH stand dabei oft die technologische Komponente – ‚der Computer‘ – im Vordergrund. Aber das ist eine probate Verdinglichung, die den Blick auf den konzeptionellen Kern und die historische Dimension eines philosophischen Anliegens verstellt, das bereits bei Raimundus Lullus und Gottfried Wilhelm Leibniz Konturen annimmt und auch in späteren Epochen immer wieder virulent wurde. Dieses Anliegen gilt der Frage, inwieweit nicht nur Natur-, sondern auch Geistesphänomene formal beschrieben, analysiert und rekonstruiert werden können.

Dass so ein Unternehmen irgendwann darauf hinauslaufen müsste, nicht nur das Denken und Schlussfolgern an sich, sondern auch die kulturellen Produkte dieses Denkens und Folgerns formal zu behandeln, hat pikanterweise ausgerechnet der Satiriker Jonathan Swift 1726 als einer der ersten prognostiziert (Swift 1986, Teil III, Kapitel 7). Sein Held *Gulliver* wird Zeuge, wie ein Professor der *Akademie von Lagado* Studenten an einer Maschine kurbelnd nach dem Zufallsprinzip Wortkombinationen generieren lässt. Dabei kommt natürlich viel Unsinn heraus, aber manchmal fügen sich die Worte auch zu sinnvollen Folgen und Sätzen und die werden dann manuell selektiert, um sie wiederum zu Texten zusammenzufügen. Das Ganze ist wie gesagt ein Akademieprojekt: Als Langzeitvorhaben angelegt zielt es deshalb, wie Swift seinen *Gulliver* berichten lässt, darauf, die Universalbibliothek aller nur möglichen Bücher überhaupt zu generieren.

Wer sich um die statistischen und mathematischen Feinheiten aktueller Verfahren wie *topic modeling* oder *word2vec* nicht weiter bekümmern will, kann Swifts Persiflage der Leibnizschen Kombinatorik natürlich umstandslos als ein Menetekel lesen, das bereits die Kapitulation des geisteswissenschaftlichen Forschungsethos vor der ‚anderen‘, dem Geistigen inkompatiblen Kultur mathematisch-naturwissenschaftlicher Disziplinen voraussagt. Aber das ist selbst unter dieser problematischen Voraussetzung schlicht zu kurz gedacht: denn der Topos von den ‚zwei Kulturen‘ des Wissens war schon 1959, als C. P. Snow ihn formulierte, genau das – ein Topos, ein rhetorisches Stilmittel. Als Snow das Bild in seiner berühmten *Rede Lecture* verwendete, um auf die Problematik zweier sich verfestigender, voneinander zunehmend abgeschotteter ‚Kulturen‘ des Wissens hinzuweisen, tat er das in polemischer Absicht. Bewirken wollte der Bildungspolitiker Snow hingegen die Wiederaufnahme des intellektuellen Diskurses zwischen den *scientists* und den *men of letters*. In der Pflicht sah er dabei in erster Linie die Vertreter der *liberal arts* und die Kulturschaffenden, die Snow einer bornierten, wissenschaftsfeindlichen Gesprächsverweigerung bezichtigte (Snow [1959] 2001).

Dieser kleine Exkurs zeigt: Um ein Forschungsparadigma in seinem Anspruch zu verstehen, sollten wir es nicht nur pragmatisch und methodologisch-theoretisch, sondern auch wissenschafts- und ideengeschichtlich kontextualisieren. Wenn sich heute die Digital Humanities darum bemühen, die der Domäne der Geisteswissenschaften zugerechneten Phänomene der menschlichen Kultur, des Sinnverstehens, der Konstituierung, Kommunikation, Tradierung und Analyse von Bedeutung bis hin zur kritischen Reflexion der Verstehensprozesse selbst auf dem Wege einer formalen Modellierung und in Zusammenarbeit mit Informatikern zu erforschen, dann tragen sie insofern seit einem halben Jahrhundert zur Verwirklichung der im Laufe der Moderne immer wieder eingeforderten domänenübergreifenden Gesprächskultur bei. Und umgekehrt setzt nun die Frage, die im Fokus des vorliegenden Bandes steht – nämlich wie man den grundsätzlich interdisziplinären Forschungsansatz der DH im spezifischen Bezug auf den engeren Anwendungsbereich der algorithmischen Textanalyse modular organisieren, schrittweise ausbauen, systematisch validieren und zyklisch optimieren kann –, an genau der Problemkonstellation an, die für die Entwicklung der digitalen Geisteswissenschaften insgesamt die frühesten Impulse geliefert hat. Diese Konstellation lässt sich vielleicht am besten so umreißen: Wie lassen sich das komplexe Phänomen Text, das semiotische System der natürlichen Sprache und die Methode der formalen Modellierung und Analyse von Objekten und Prozessen mit rechnergestützten Verfahren so zueinander in Beziehung setzen, dass daraus beispielhaft eine neue epistemische Praxis für die Geisteswissenschaften resultiert?

Diese Frage begleitet uns seit den Anfängen der disziplinären Formierung unseres Forschungsfeldes. *Nomen est omen*: Schon 1964 gründete der Germanist und Mediävist Roy Wisbey an der University of Cambridge das erste *Literary and Linguistic Computing Centre* und 1970 veranstaltete er dort zusammen mit Michael Farringdon auch das erste internationale Symposium „The Computer in Literary and Linguistic Research“. Aus dem Veranstaltungstitel leitete sich dann der Name für den 1973 am King's College London gegründeten ersten Fachverband ab: *Association for Literary and Linguistic Computing* (ALLC). Einen programmatischen Fokus auf die Trias von Literatur, Linguistik und Computing signalisierte in gleicher Weise die von 1973 bis 1988 jährlich ausgerichtete *ALLC Conference* und ebenso die erste einschlägige Fachzeitschrift, das *Journal of Literary and Linguistic Computing*. Erst als in den 1990er Jahren das Spektrum digitalisierter Medienformate breiter wurde, setzte sich allmählich die Bezeichnung *Humanities Computing* durch. Die heute international geläufige Bezeichnung *Digital Humanities* kam hingegen erst nach der Jahrtausendwende auf; wir verdanken sie, wie sich John Unsworth erinnert hat, einer Entscheidung, die einen gelungenen Kompromiss zwischen den Interessen von Verlagsmarketing und wissenschaftlicher Profilbildung darstellt:

The real origin of that term [digital humanities] was in conversation with Andrew McNeillie, the original acquiring editor for the Blackwell *Companion to Digital Humanities*. We started talking with him about that book project in 2001, in April, and by the end of November we'd lined up contributors and were discussing the title, for the contract. Ray [Siemens] wanted "A Companion to Humanities Computing" as that was the term commonly used at that point; the editorial and marketing folks at Blackwell wanted "Companion to Digitized Humanities." I suggested "Companion to Digital Humanities" to shift the emphasis away from simple digitization. (Kirschenbaum 2010)

Wir dürfen die Nuancierung der wechselnden Bezeichnungen *Literary and Linguistic Computing* über *Humanities Computing* bis hin zu *Digital Humanities* also nicht überinterpretieren; Namensgebungen können auch kontingenten Faktoren geschuldet sein. Aber die Abfolge der Bezeichnungen macht immerhin deutlich, dass neben der schrittweisen Erweiterung des Skopus vom engeren Gegenstandsbereich der Textwissenschaften hin zur Gesamtdomäne der Geisteswissenschaften sich zugleich das Bewusstsein für den methodologischen Effekt einer Kombination von verstehensorientiert-interpretierenden mit formal-modellierenden Verfahren geschärft hat. Statt um eine ‚simple digitization‘ von Objekten und Verfahren der traditionellen Geisteswissenschaften mit den Techniken der digitalen Datenerfassung und Datenmanipulation ging und geht es in den DH dabei um eine methodologische Erweiterung und Bereicherung der Geisteswissenschaften.

Dazu ist nun allerdings ein Brückenschlag erforderlich, denn insbesondere der hermeneutisch orientierte geisteswissenschaftliche Forschungsansatz ist seiner Natur nach zunächst phänomenologisch motiviert: Die traditionellen, idealtypisch im 19. Jahrhundert entstandenen Geisteswissenschaften fragen nach der Bedeutung und lebensweltlichen Relevanz symbolischer Artefakte für den Menschen; sie begnügen sich nicht mit der abstrakten und interessefreien analytischen Bestimmung von Phänomenen als Objekten an sich. Erst im 20. Jahrhundert kommen mit der strukturalen Linguistik und der formalistischen Erzählforschung die ersten Disziplinen auf, für die nicht mehr diese phänomenologisch-historische Perspektive auf die Wirkung eines Zeichens auf und für den Menschen und die Kultur im Vordergrund steht, sondern die analytische auf das immanente, dynamische Funktionsgefüge des Zeichenkomplexes und dessen Prozesslogik. Diese zweite Fragehaltung setzt allerdings mindestens zweierlei voraus: erstens die Bereitschaft, im Interesse des Erkenntnisfortschritts vorübergehend von der Problematik historischer Kontingenz und der Kontextualisierung des Beobachtungsfeldes zu abstrahieren, es also gewissermaßen bewusst ‚stillzustellen‘; zweitens die Bereitschaft, das lebensweltlich als ein auf uns wirkendes Ganzes erfahrene – die Sprache, die Erzählung, das Kunstwerk, die Gesellschaft etc. – als ein aus diskreten Elementen zusammengesetztes zu konzeptualisieren. Denn eben dies sind die beiden notwendigen Voraussetzungen einer digitalen Konzeptuali-

sierung: Abstraktion von der je konkreten Kontextgebundenheit, Erfassung und Modellierung der Objektdomäne in Form diskreter Beobachtungsinstanzen.

Schon Willard McCarty hat in seiner 2005 vorgelegten Monographie *Humanities Computing* allerdings darauf aufmerksam gemacht, dass das Projekt der Digital Humanities sich nicht auf ein Pipelinemodell reduzieren lässt, in dem geisteswissenschaftliche ‚Probleme‘ isoliert, formal beschrieben, dann an die Informatik weitergereicht, dort operationalisiert und mit teils generischen, teils spezialisierten Tools bearbeitet und als ‚gelöst‘ an die Geisteswissenschaften rückgemeldet werden, um zum nächsten ‚Problem‘ voranzuschreiten. Am Ende seiner Überlegungen stellt McCarty *A preliminary Agenda*. Sie umfasst zehn Punkte, von denen heute, fünfzehn Jahre später, die letzten neun weniger dringlich und aktuell erscheinen mögen – nicht aber der erste Punkt ‚Analysis‘. Dazu schreibt McCarty:

The first item is to understand how computing affects analysis in the humanities beyond simply fetching, counting and formatting data – how, in other words, it affects analysis itself rather than its scope, speed or convenience. [...] I have placed this question first on the agenda because otherwise we elide the profound difference between artifact and data and so are in no position to understand the significance and reception of the results we obtain from these data. We need to know what that difference is, and how an analysis based on it might differ from what we have been doing all along. (McCarty 2005, S. 206)

Wie aber können wir diese Fragestellung in der konkreten interdisziplinären Forschungsarbeit nicht nur abstrakt bewusst halten, sondern sie konzeptionell so in den Forschungsprozess einfließen lassen, dass daraus mehr wird als ein fruchtloses philosophisches *ceterum censeo*?

CRETA hat dieses Problem mit einem Ansatz zur reflektierten Erarbeitung textanalytischer DH-Workflows gelöst, der nach meiner Einschätzung vorbildhaft ist. Dazu trägt erstens die ‚Werkstatt-Idee‘ bei, die als organisatorisches Prinzip einen explorativen, bottom-up Austausch zwischen den beteiligten Disziplinen ermöglicht. Zweitens berücksichtigt der Ansatz von CRETA die geisteswissenschaftliche Anforderung, nach der Abstraktion vom Kontext, die für den formal-modellierenden Zugang Voraussetzung ist, auch wieder zum Kontext zurückzukehren – und dann womöglich in einen neuen Zyklus einzutreten, der ein komplexeres Gegenstandsmodell zur Folge hat. Die interdisziplinäre Forschung ist damit konzeptionell als ein iterativer Prozess organisiert², der wechselseitig befruchtend sein kann, statt eine modulare „Entkoppelung der disziplinären Fel-

² Siehe hierzu Abb. 1, „Prototypischer Arbeitsablauf“ in Pichler und Reiter 2020, S. 44.

der“ (Kuhn 2020b, S. 19) zu erzwingen.³ Drittens lernen die an dem CRETA-Projekt beteiligten Disziplinen voneinander nicht nur in Bezug auf die Konzepte, Modelle und Verfahren, die bei der Erarbeitung, Validierung und Anwendung von Prozessen und Werkzeugen der algorithmischen Textanalyse im Vordergrund stehen. Sie begegnen darüber hinaus auch einer fremden disziplinären Forschungspraxis, reflektieren diese und erhalten so Gelegenheit, etablierte Routinen in die eigene Praxis zu übernehmen, wo dies sinnvoll erscheint. Ein Beispiel für den letzten Punkt und damit zugleich für die programmatische Dimension von CRETA diskutiert Kuhn (2020a) ab Seite 84; hier wird wegweisend die Möglichkeit einer Integration des in der Informatik etablierten *rapid probing of analysis models* mit einem hermeneutischen Ansatz erörtert.

Die Beiträge des vorliegenden Bandes dokumentieren so einerseits auf beeindruckende Weise, wie und mit welchen Ergebnissen die Kooperationspartner in CRETA die praxeologische Aufgabe, „technische Methoden und Arbeitsablaufpraktiken zur Textanalyse im Forschungsbereich der DH zu entwickeln“ (Pichler und Reiter 2020, S. 43), angegangen sind. Aber der Band als ganzer ist zugleich mehr als die Summe seiner Teile: Er demonstriert auch das ambitionierte methodologische Niveau, auf dem sich die Digital Humanities als interdisziplinäres Forschungsprojekt im Schnittfeld von Geisteswissenschaften und Informatik mittlerweile bewegen können. Die Voraussetzung dafür signalisiert CRETA, das „Centrum für reflektierte Textanalyse“, mit der Wahl des Adjektivs in seinem Namen. Es sei hiermit doppelt unterstrichen.

Literatur

- Kirschenbaum, Matthew G. (2010). „What Is Digital Humanities and What’s It Doing in English Departments?“ In: *ADE Bulletin* 150. URL: <https://mkirschenbaum.files.wordpress.com/2011/03/ade-final.pdf> (besucht am 1. Juni 2020).
- Kuhn, Jonas (2020a). „Computational Text Analysis within the Humanities“. In: *Reflektierte Algorithmische Textanalyse*. Hrsg. von Nils Reiter, Axel Pichler und Jonas Kuhn. Berlin: De Gruyter, S. 63–106.
- Kuhn, Jonas (2020b). „Einleitung“. In: *Reflektierte Algorithmische Textanalyse*. Hrsg. von Nils Reiter, Axel Pichler und Jonas Kuhn. Berlin: De Gruyter, S. 9–40.
- McCarty, Willard (2005). *Humanities Computing*. Houndmills/New York: Palgrave Macmillan.

³ Das schließt nach der CRETA-Philosophie jedoch umgekehrt eine Wahl des Domänenexperten/Entwickler-Modells nicht aus, wo diese pragmatisch begründet werden kann. Siehe Kuhn 2020b, S. 23, Fußnote 13

Pichler, Axel und Nils Reiter (2020). „Reflektierte Textanalyse“. In: *Reflektierte Algorithmische Textanalyse*. Hrsg. von Nils Reiter, Axel Pichler und Jonas Kuhn. Berlin: De Gruyter, S. 43–59.

Snow, Charles Percy [1959] (2001). *The Two Cultures*. London: Cambridge University Press.

Swift, Jonathan (1986). *Gulliver's Travels*. Wiederauflage 2008. Oxford: Oxford University Press.