

Rechtsgeschichte Legal History

www.rg.mpg.de

<http://www.rg-rechtsgeschichte.de/rg24>

Rg **24** 2016 335 – 336

Manfred Thaller

Was sind (keine) methodischen Implikationen der Digital Humanities?

Dieser Beitrag steht unter einer
Creative Commons cc-by-nc-nd 3.0



Manfred Thaller

Was sind (keine) methodischen Implikationen der Digital Humanities?

Der Begriff der Digital Humanities ist derzeit ebenso modisch wie unscharf. Er wird verwendet, um die Nutzung von Twitter oder von Blogs durch GeisteswissenschaftlerInnen zu beschreiben; dient der Diskussion der Implikationen der Kommunikation mit Studierenden via Facebook; deckt die Möglichkeiten digitaler Bibliotheken im Allgemeinen und derer von Manuskriptbeständen im Besonderen ab; betont die Möglichkeiten dynamischer Visualisierungen geographischer Verteilungen im Rahmen eines *spatial turn* der Geisteswissenschaften; schließt eine selbstbewusste Community ein, die sich der Frage widmet, wie man Texte mit XML, am besten auf Grund des aus den Disziplinen entstandenen Standards der Text Encoding Initiative, zur maschinellen Bearbeitung aufbereitet; dient einer nicht minder selbstbewussten Community, die sich dem statistischen Nachweis der Autorschaftszuschreibung strittiger Texte widmet; umfasst die Versuche, Raumerlebnisse in nicht mehr existierenden Gebäuden mit den Mitteln der Virtual Reality nachzuzeichnen, ebenso wie die akustische Simulation der Rahmenbedingungen historisch bedeutsamer Reden und ist mit der Entwicklung von OCR-Verfahren für Texte in den Keilschriftsprachen noch lange nicht zu Ende.

Die Beschäftigung von GeisteswissenschaftlerInnen mit Medien wie Twitter oder Facebook scheint mir der systematischen Diskussion nicht zu bedürfen. Es mag ein in persönlicher Arroganz begründeter Fehler sein: Die Tatsache, dass GeisteswissenschaftlerInnen nicht wirklich dümmer sind als der Rest der Bevölkerung, sich der aktuellen Kulturtechniken also ebenso bedienen wie dieser Rest, überrascht mich persönlich nicht.

Aus anderem Grunde möchte ich auch darauf verzichten, näher darauf einzugehen, wie weit die Geisteswissenschaften genuine Neuentwicklungen der Informatik benötigen. Nicht, weil mir das fern läge; dies ist mein eigentliches Forschungsgebiet. Der Umgang mit widersprüchlichen, inkonsistenten und unvollständigen Daten, eine Kernkompetenz der Geisteswissenschaften, könnte ebenso wichtige Impulse für die nächsten Generationen von Informatikanwendungen geben, wie die Ingenieurwissenschaften dies für die bisherigen getan

haben. Dies zu erläutern, würde aber ein weiteres Ausholen erfordern, als es der zur Verfügung stehende Raum erlaubt.

Zwischen beiden Extremen gibt es zahlreiche Ansätze, bei denen nicht immer klar ist, ob es sich um neutrale Werkzeuge oder Techniken handelt, die nur im Rahmen einer bestimmten Methode sinnvoll sind. Auch wenn man mit Paul Feyerabend die Existenz verbindlicher Methoden verneint, kann man immer noch Werkzeuge der Informationstechnologie nutzen, um relevante Texte zu finden: Die Digital Humanities insgesamt finden also in der ganzen Breite der möglichen methodischen Haltungen der Geisteswissenschaften ihre Berechtigung. Wir verzichten deshalb darauf, von methodischen *Positionen* zu sprechen, und beschränken uns auf die Betrachtung methodischer *Implikationen*. Zwei Beispiele dafür:

Beispiel 1: Die als Optical Character Recognition (OCR) bekannte Wandlung gedruckter Texte zu Daten, die mit informationstechnischen Mitteln bearbeitet werden können, ist für die Geisteswissenschaften insgesamt von erheblichem potentiellen Wert. Trotzdem verlief ihre Rezeption seit den frühen achtziger Jahren eher zögernd. Dies lag daran, dass auch eine zunächst respektabel wirkende Richtigkeit der Texterkennung von 99 % bedeutet, dass nahezu in jeder Zeile ein Zeichen falsch erkannt wird. Dies erlaube, so die Argumentation vor allem der frühen neunziger Jahre, *nicht*, sich darauf zu verlassen, dass ein nicht gefundenes Wort tatsächlich nicht im Text enthalten sei, eine Beweisführung per argumentum ex silentio sei also unmöglich.

Oder zumindest *wurde* dieses Argument gerne gebracht: Die DFG referiert es zwar noch in den gültigen Empfehlungen für die Digitalisierung, fördert derzeit aber ein Projekt zur Schaffung technischer Strukturen, die ermöglichen sollen, alle gedruckten Texte zwischen 1500 und 1800 mittels verbesserter OCR-Verfahren umzuwandeln – mit Überlegungen, wie diese Techniken auf die Zeit der Inkunabeln und auf das 19. Jahrhundert ausgedehnt werden können.

Nicht deshalb, weil die Technik entscheidende Fortschritte gemacht habe, sondern wegen einer

methodischen Entwicklung innerhalb vor allem der Philologien. Das intensive Studium einer kleinen Gruppe kanonischer Texte erlaubt Aussagen über die Literatur insgesamt, so die bisherige Annahme. Wenn wirklich große Mengen von Texten zur Verfügung stehen: Wäre es nicht mindestens genauso sinnvoll, aus statistischen Aussagen über einen großen Teil der existierenden Texte einer bestimmten Epoche / Stilrichtung – also über hunderttausende oder Millionen Druckseiten – Erkenntnisse zu gewinnen, mit deren Hilfe danach auch die einzelnen Texte besser eingeordnet werden können? Was zur Antwort auf das oben angeführte Argument führt: Es mag ja sein, dass menschliche LeserInnen 500 Seiten so sorgfältig durchgehen können, dass ihnen kein Auftreten einer Wortform entgeht, sie sind aber definitiv *nicht* in der Lage, eine Million Seiten durchzusehen. Eine fast beliebig große Fehlerrate bei deren maschineller Durchsuchung verbessert das Ergebnis also immer noch ungemein.

Die Verfügbarkeit großer Textmengen erfordert kein bewusstes methodisches Umdenken. Sie *impliziert* aber eine Erweiterung des methodischen Arsenal. Und führt zur Frage, ob, aristotelisch gesehen, geisteswissenschaftliche Kanones Essenz der Geisteswissenschaften oder Akzidenz der bisherigen Arbeitsbedingungen sind.

Beispiel 2: Eine der prominentesten Studien mindestens der französischen Geschichtswissenschaft in der zweiten Hälfte des vorigen Jahrhunderts war unstreitig Emmanuel Le Roy Laduries 1975 erschienenes *Montaillou, village occitan de 1294 à 1324*. In der gedruckten Diskussion tauchten bald Stimmen auf, der Verfasser habe das Bild verfälscht, da die vom Inquisitor, der die ausgewerteten Protokolle im Lateinischen niederschrieb, befragten Personen okzitanisch gesprochen hätten – und die Zitate im französischen Text seien dann durch die Ladurie'sche Übersetzung

nochmals verfälscht worden. Im Smalltalk von den Kritikern dann gerne auf den Punkt gebracht, Laduries Argumentation sei deshalb so konsistent, weil er durch Auswahl und Übersetzung der Zitate diese Konsistenz künstlich generiert habe.

Selbstverständlich hat Ladurie nach allen handwerklichen Regeln gearbeitet – die Zitate aus den archivischen Quellen sind präzise und zahlreich. Wenn Kritiker eines der bekanntesten Historiker seiner Zeit einen so schwerwiegenden Vorwurf gegen eines seiner am weitesten verbreiteten Werke erheben, warum hat trotzdem nie jemand den Versuch unternommen, an Hand der ausgewiesenen Quellen nachzuprüfen, ob die Übersetzung mindestens des Lateinischen ins Französische und die Wahl der Zitate den Vorwurf rechtfertigen? Natürlich ist die Antwort offensichtlich: Der Aufwand, systematisch große Mengen nicht publizierter Quellen, die nur vor Ort im Archiv zugänglich sind, nachzuprüfen, ist schlicht prohibitiv. Wenn selbst die Anmerkungen eines der prominentesten Historiker Frankreichs de facto nicht nachprüfbar sind, wozu dienen dann Quellennachweise archivischen Materials in im Eigenverlag publizierten Dissertationen? Der schamanistischen Befriedung der Stammesgötter der historischen Zunft?

Wenn freilich der Inhalt der Archive systematisch digitalisiert würde, was heute schon in Angriff genommen werden könnte, *wenn* die handschriftlichen Texte mit jenen Ansätzen zur OCR für handschriftliches Material, die jetzt im Labor so weit gediehen sind, dass ihr Einsatz auf Produktionsniveau in etwa zehn Jahren wahrscheinlich ist, darauf angewendet werden ... änderte sich methodisch nichts. Bloß die fundamentalsten Grundlagen, auf Basis derer über historische Erkenntnisse diskutiert werden kann, als Implikation neuer Formen der Verfügbarkeit der Quellen.

