

Annotation und Distant Reading: Probleme, Synergien, Perspektiven

,
angelika.zirker@uni-tuebingen.de
Eberhard Karls Universität Tübingen, Deutschland

,
m.bauer@uni-tuebingen.de
Eberhard Karls Universität Tübingen, Deutschland

In unserem Vortrag möchten wir Methoden des *close* und *distant reading* miteinander in Beziehung setzen und Probleme sowie Synergien im Rahmen der Digital Humanities diskutieren. Unser Ziel ist dabei, mögliche Perspektiven im Zusammenspiel der beiden Herangehensweisen aufzuzeigen. Ausgangspunkt ist ein Projekt, das bei der Tagung der Digital Humanities im deutschsprachigen Raum 2015 in Graz vorgestellt wurde und das sich mit der erklärenden Annotation von literarischen Texten im Kontext der Digital Humanities befasst.

Unsere Ausgangsfrage ist, inwiefern Methoden des *distant reading* bei Annotationen hilfreich sein können und wo ihre (derzeitigen) Grenzen liegen bzw. in welchen Fällen sie sogar hinderlich oder nicht zielführend sind. Franco Moretti (in seiner Essaysammlung zum *Distant Reading*) wie auch Fotis Jannidis haben aufzeigen können, inwieweit qualitative Methoden Aufschluss über bestimmte Entwicklungen und Trends in der Literaturgeschichte oder in der Geschichte von Gattungen geben können; ebenso ermöglichen quantitative Methoden z. B. die Aufschlüsselung von Charakterkonstellationen. Dabei werden jedoch auch Probleme deutlich, etwa wenn für den identischen Charakter unterschiedliche Namen und Referenzen gebraucht werden: an diesen Punkten versagen automatisierte Verfahren häufig. Eine weitere Schwierigkeit ergibt sich daraus, dass bei der Ermittlung von Worthäufigkeiten die Semantik unberücksichtigt bleibt: eine rein quantitative Analyse etwa des Wortes „bank“ in einem englischen Text kann ggf. nicht zwischen den verschiedenen Bedeutungen des Wortes unterscheiden und übersieht somit Ambiguitäten ebenso wie unterschiedlichen Funktionen von Wörtern in der Syntax und Grammatik eines Satzes oder Textes. Dieses Problem ergibt sich beispielsweise bei x-ray, das Verlinkungen auf Wikipedia-Einträge anbietet, dabei aber häufig Ambiguitäten nicht erkennt bzw. lediglich einen Link auf die Disambiguierung von Wikipedia selbst liefert (der dann wiederum dem Leser, dem ja eigentlich geholfen werden soll, die Interpretation des Begriffs überlässt, die Ambiguität also erkennt, aber nicht auflöst). Ebenso ergibt

sich ein Problem im Verhältnis von Quantität und Qualität: wenn ein Wort in einem Text nicht häufig genannt wird, heißt das nicht zwangsläufig, dass es nicht wichtig ist und für die Gesamtbedeutung des Textes relevant ist. Vor diesem Hintergrund stellt sich somit auch die Frage, wie Entscheidungen über Bedeutungen im Verhältnis zu distant reading-Methoden getroffen werden können: gibt es hierzu systematische Ansätze? Und (wie) können im Hinblick auf diese Probleme Annotationen von quantitativen Verfahren profitieren?

Im zweiten Teil des Vortrags werden Synergieeffekte von Methoden der Annotation und des *distant reading* vorgestellt. Tools wie etwa der *google Ngram Viewer* erlauben die sehr schnelle Durchsicht von großen Daten- und Textmengen, die manuell nicht zu leisten ist. Sie verschafft dem Annotierenden – einen Überblick, der die erklärende Annotation von Texten erleichtert und etwa auch Querverweise und interne Verlinkungen erleichtert. Der Leser profitiert vom Zusammenspiel der Methoden, denn die individuelle Annotation, ausgerichtet an dem Bedarf sowohl von individuellen Nutzern wie auch von social communities und Lernern, kann die automatisierten und quantitativen Verfahren anreichern.

Aus diesen Synergieeffekten ergibt sich der Anschluss an Perspektiven zum Verhältnis von (erklärender) Annotation und quantitativen Methoden des *distant reading*. Es ist denkbar, dass Datenbanken es künftig ermöglichen, Vorgänge des erklärenden Annotierens zu automatisieren, etwa in Fällen von Ambiguität, deren Semantik erkannt und die entsprechend aufgelöst wird. In diesem Zusammenhang bedarf er der Ergänzung unserer literaturwissenschaftlichen Perspektive durch technische Expertise. Unser Vortrag stellt somit auch Fragen, die insbesondere im Zuge der Tagung diskutiert werden können. Dabei sollen auch Ideen diskutiert werden, welche Möglichkeiten es geben könnte, dass die manuelle Markierung dessen, was erklärend annotiert (also nicht im Sinne von markup) werden soll, entfällt und Erklärungskontexte definiert werden, innerhalb derer eine datenbankgestützte – und damit „automatische“ – Annotation erfolgen kann.

Der Vortrag bewegt sich an der Schnittstelle von Automatisierung und individuellen hermeneutischen Akten und damit entlang des Problems, wie im Markup eines Textes Entscheidungen getroffen werden können, welche Aspekte in einem Text relevant sind und die dem individuellen Text gerecht werden können. Wir möchten verschiedene Fallstudien aus englischsprachigen literarischen Texten vorstellen, etwa anhand von automatisierten Annotationssystemen wie x-ray, die oben geschilderte Probleme exemplarisch aufzeigen, die aber zugleich auch Synergieeffekte deutlich machen. Letztlich sollten bei dem Verhältnis von qualitativen Methoden des *close reading* und quantitativen Herangehensweisen die Nutzerfreundlichkeit sowie die Individualisierung im Vordergrund stehen. Diese Individualisierung ist dabei zweifach: zum einen bezieht sie sich auf Informationen aus Datenbanken, zum anderen auf den Text. Texte, die annotiert werden sollen, müssen mit Datenbanken

korrelieren, und die erklärenden Annotationen müssen offenlegen, was im Text gemeint ist sowie was der potentielle Leser erfahren und wissen möchte. Eine solche Anreicherung von Texten ist von quantitativen Methoden bislang nicht zu leisten; umgekehrt sind qualitative Methoden momentan dadurch eingeschränkt, dass sie sich manueller Verfahren bedienen müssen. Das Zusammenspiel von Quantität und Qualität, von *close* und *distant reading*, erklärender Annotation und computergestützter Textanalyse öffnet neue Perspektiven im Bereich der Digital Humanities und kann auch einen Beitrag zu fächerübergreifenden Paradigmen leisten.

Moretti, Franco (2013): *Distant Reading*. New York: Verso.

Stroud, Matthew D. (2006): "The Closest Reading: Creating Annotated Online Editions", in: Bass, Laura R. / Greer, Margaret R. (eds.): *Approaches to Teaching Early Modern Spanish Drama*. New York: The MLA of America 214-129.

Bibliographie

Battestin, Martin C. (1981): „A Rationale of Literary Annotation: The Example of Fielding's Novels“, in: *Studies in Bibliography* 34: 1-22.

Bauer, Matthias / Angelika Zirker (2015): „Whipping Boys Explained: Literary Annotation and Digital Humanities“, in: Siemens, Ray / Price, Kenneth M. (eds.): *Literary Studies in the Digital Age: An Evolving Anthology* <http://dlsanthology.commons.mla.org/under-review-matthias-bauer-and-angelika-zirker-whipping-boys-explained-literary-annotation-and-digital-humanities/> [letzter Zugriff 09. Januar 2016].

Cummings, James (2013): „The Text Encoding Initiative and the Study of Literature“, in: Siemens, Ray / Schreibman, Susan (eds.): *A Companion to Digital Literary Studies*. Oxford: Blackwell 451-76.

Drucker, Johanna (2012): „Humanistic Theory and Digital Scholarship“, in: Gold, Matthew K. (ed.): *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press 85-95.

Jannidis, Fotis (2010): „Methoden der computergestützten Textanalyse“, in: Nünning, Vera (ed.): *Methoden der literatur- und kulturwissenschaftlichen Textanalyse*. Stuttgart: Metzler 109-132.

Jannidis, Fotis / Flanders, Julia (2015): „Knowledge Organization and Data Modeling in the Humanities. A whitepaper“ <http://www.wwp.northeastern.edu/outreach/conference/kodm2012/> [letzter Zugriff 09. Januar 2016].

Jannidis, Fotis / Krug, Markus / Toepfer, Martin / Puppe, Frank / Reger, Isabella / Weimer, Lukas (2015): „Automatische Erkennung von Figuren in deutschsprachigen Romanen“. Abstract für die DHd 2015 in Graz.

McCarty, Willard (2012): „Collaborative Research in the Digital Humanities“, in: Deegan, Marilyn / McCarty, Willard (eds.): *Collaborative Research in the Digital Humanities*. Farnham: Ashgate 1-10.

Meister, Jan-Christoph (2012): „Crowd Sourcing 'True Meaning': A Collaborative Approach to Textual Interpretation“, in: Deegan, Marilyn / McCarty, Willard (eds.): *Collaborative Research in the Digital Humanities*. Farnham: Ashgate 105-122.