

Relating the Unread

Modellierungen der Literaturgeschichte

Brottrager, Judith

judith.brottrager@tu-darmstadt.de
TU Darmstadt, Germany

Von Beginn an war die Inklusion einer größeren Textmenge und damit auch nicht-kanonisierter Werke eines der Hauptargumente für digitale Ansätze in der Literaturwissenschaft: Durch Distant Reading, könne, so Moretti, eine alternative Literaturgeschichte erfasst werden, die auch das sogenannte „Great Unread“ mit einschließt (Moretti 2013: 48f). Forschungsprojekte der letzten Jahre haben jedoch gezeigt, dass der Mehrwert der digitalen Literaturwissenschaft nicht nur in der rein quantitativen Erweiterung der Untersuchungsgegenstände, sondern auch in der tiefgreifenden Kontextualisierung der untersuchten Texte in einer, wie es Bode nennt, „data-rich literary history“ (Bode 2018: 37–57) liegt. Diese kontextreichen Ansätze, wie beispielsweise Analysen zum Verhältnis von Prestige und Popularität (Underwood/Sellers 2016; Algee-Hewitt et al. 2016; Porter 2018; Underwood 2019: 68–110), zeigen, wie quantitative Methoden trotz nötiger Formalisierungen und Abstraktionen die Komplexität des literarischen Systems (cf. Bode 2017: 91) beschreiben können.

Mein Dissertationsprojekt folgt dieser Tradition, indem kanonisierte und nicht-kanonisierte literarische Werke auf unterschiedlichen Ebenen mit Rückgriffen auf literaturhistorische Daten miteinander in Beziehung gesetzt werden. Die für die Untersuchungen erstellten Korpora und Datensätze umfassen etwa 1.200 englisch- und deutschsprachige Texte von 1688 bis 1914, wodurch diachrone und synchrone Vergleiche von Kanonisierungsprozessen und -mustern über Sprachgrenzen hinweg ermöglicht werden. Durch diese Datenvielfalt sollen einerseits etablierte literaturhistorische Narrative untersucht und quantitativ überprüft werden und andererseits literaturwissenschaftliche Kategorien wie Kanonisierung und Wertung so operationalisiert werden, dass sie mit computationellen Methoden zur Bestimmung von Textähnlichkeiten gewinnbringend kombiniert und diese Ähnlichkeit schließlich als Netzwerkmodelle dargestellt werden können.

Korpusaufbau

Der Korpusaufbau folgt einem systematisch angepassten Ansatz von Algee-Hewitt und McGurl, der darauf abzielt, von einem vorgefundenen zu einem maßgeschneiderten Korpus zu gelangen, indem Bestenlisten, Bestsellerlisten und von Expert*innen kuratierte Literaturlisten kombiniert werden, um ein repräsentatives Korpus für die englischsprachige Literatur des 20. Jahrhunderts zu erstellen (Algee-Hewitt/McGurl 2015). Durch die Kombination dieser Listen decken Algee-Hewitt und McGurl drei Ebenen der literarischen Produktion ab: den normativ-exklusiven Kanon, populäre Texte und von Expert*innen für Postkoloniale und Feministische Literaturwissenschaft vorgeschlagene Werke. Für die Umsetzung für die Zeitspanne von 1688-1914 wurde dieser Ansatz systematisch adaptiert, indem narrative Literaturgeschichten, Anthologien und (spezialisierte) Sekundärtexte, die diese Ebenen abdecken, identifiziert und als bibliografische Quellen für die Korpuserstellung genutzt wurden. Der Workflow umfasst Web-

scraping, X-Technologien/Transformationen und Retro-Digitalisierungen.

Metadatenätze

Analog zur Korpuserstellung wurden Daten zur Kontextualisierung der jeweiligen Korpus-texte, aber auch der gesamten literarischen Produktion gesammelt. Durch diese Daten können die Korpora mit den von Algee-Hewitt et al. als „the published“¹ und „the archive“² (Algee-Hewitt et al. 2016: 2) bezeichneten Ebenen der Literaturgeschichte verglichen werden, wodurch wiederum eine Einordnung der Untersuchungsergebnisse und eine Einschätzung der unvermeidlichen Selektionsprozesse bei der Korpuserstellung möglich sind (cf. Bode 2017: 85). Die Daten reichen von Publikationslisten und Leihbibliothekskatalogen bis hin zu Rezensionen und Daten zu Zweitaufgaben; der Aufbau der Datensätze orientiert sich am Beispiel der von Garside zur Verfügung gestellten Datenbank *British Fiction 1800-1829* (Garside 2011). In Anlehnung an erfolgreiche Metadatenanalysen (z.B. Jockers 2013: 35–62) sollen diese Daten zur Überprüfung von Hypothesen zum Zusammenhang zwischen dem Aufschwung des Romans als Gattung und gegenderten Kanonisierungsprozessen verwendet werden (cf. Watt 1957; Raven 1987; Raven/Forster 2000; Tuchman/Fortin 2012).

Operationalisierungen

Die gesammelten Daten werden neben diesen Metadatenanalysen auch für die Operationalisierungen der literaturwissenschaftlichen Konzepte der Kanonisierung und Wertung eingesetzt. Aufbauend auf die theoretischen Grundlagen von Heydebrand und Winko werden Kanonisierung und Wertung als Scores implementiert, die ausdrücken, wie hoch die Wahrscheinlichkeit ist, dass ein bestimmter Text sehr kanonisiert ist beziehungsweise zur Entstehungszeit sehr gut rezipiert wurde (Heydebrand/Winko 1996). Ein besonderes Augenmerk liegt hierbei auf der Differenzierung der Konzepte und der Einbindung der Rezeptionsebene über Marker für Publikumsinteresse (wie Einträge in Leihbibliothekskatalogen und Zweitaufgaben innerhalb einer Generation) und sprachliche Werturteile, die durch Sentiment Analysen vergleichbar werden.

Ausblick

Unter Verwendung der generierten Scores soll schließlich untersucht werden, ob Kanonisierung und Wertung mit textin-trinsischen Merkmalen in Verbindung gebracht werden können. Stilometrische Berechnungen, Topic Modeling und Word Embeddings sowie wortartenbasierte Ansätze sollen dabei als allein-stehende Analysen der Textähnlichkeiten durchgeführt werden. Als zusätzliche Analysen- und Visualisierungsmethode dienen Netzwerkmodelle, die anhand der Ergebnisse der Textähnlichkeitsberechnungen erstellt werden, zur Exploration von Ähnlichkeitsstrukturen. Besonders auf dieser Ebene soll der Bezug zur literaturwissenschaftlichen Forschung durch die Identifikation von dichten stilistischen Ähnlichkeitsgruppen und durch aus den Modellen abgeleitete Einzeltextanalysen hergestellt werden.

Fußnoten

1. i.e. die Menge aller veröffentlichten Texte
2. i.e. die Menge der in Bibliotheken und Archiven gesicherten Texte

Bibliographie

Algee-Hewitt, Mark / Allison, Sarah / Gemma, Marissa / Heuser, Ryan / Moretti, Franco / Walser, Hannah (2016): "Canon/Archive. Large-scale dynamics in the literary field", *Pamphlets of the Stanford Literary Lab* 11. <https://litlab.stanford.edu/LiteraryLabPamphlet11.pdf> [letzter Zugriff 15. Juli 2021].

Algee-Hewitt, Mark / McGurl, Mark (2015): "Between canon and corpus: Six perspectives on 20th-century novels", *Pamphlets of the Stanford Literary Lab* 8. <https://litlab.stanford.edu/LiteraryLabPamphlet8.pdf> [letzter Zugriff 15. Juli 2021].

Bode, Katherine (2017): "The equivalence of "close" and "distant" reading; or, toward a new object for data-rich literary history", in: *Modern Language Quarterly* 78, 77–106.

Bode, Katherine (2018): *A world of fiction: Digital collections and the future of literary history*. Ann Arbor: University of Michigan Press.

Garside, Peter (2011): *British fiction 1800-1829. A database of production, circulation & reception*. <http://www.british-fiction.cf.ac.uk/> [letzter Zugriff 15. Juli 2021].

Heydebrand, Renate von / Winko, Simone (1996): *Einführung in die Wertung von Literatur: Systematik - Geschichte - Legitimation*. Paderborn: Schöningh

Jockers, Matthew Lee (2013): *Macroanalysis: Digital methods and literary history*. Urbana: University of Illinois Press.

Moretti, Franco (2013): *Distant reading*. London / New York: Verso.

Porter, J.D. (2018): "Popularity/Prestige", *Pamphlets of the Stanford Literary Lab* 17. <https://litlab.stanford.edu/LiteraryLabPamphlet17.pdf> [letzter Zugriff 15. Juli 2021].

Raven, James (1987): *British fiction: 1750 – 1770. A chronological check-list of prose fiction printed in Britain and Ireland*. Newark: University of Delaware Press.

Raven, James / Forster, Antonia (eds.) (2000): *The English novel 1770 - 1829: A bibliographical survey of prose fiction published in the British Isles*. Oxford: Oxford University Press.

Tuchman, Gaye / Fortin, Nina E. (2012): *Edging women out: Victorian novelists, publishers and social change*. London: Routledge.

Underwood, Ted (2019): *Distant horizons: digital evidence and literary change*. Chicago: The University of Chicago Press.

Underwood, Ted / Sellers, Jordan (2016): "The Longue Durée of Literary Prestige", in: *Modern Language Quarterly* 77, 321–344.

Watt, Ian (1957): *The rise of the novel*. Berkeley: University of California Press.