

Beziehung und Bedeutung. Soziale und semantische Netzwerkanalyse religionshistorischer Korpora

Abstract für die Dhd-Tagung 2015 in Graz

Frederik Elwert, Ruhr-Universität Bochum

Simone Gerhards, Ruhr-Universität Bochum

Sven Sellmer, Ruhr-Universität Bochum

Mit der fortschreitenden Digitalisierung historischer Textsammlungen steht ein immer größerer Quellenfundus für computergestützte Analysen zur Verfügung. Auch wenn in den letzten Jahren neue Analysetechniken wie etwa das *topic modeling* zunehmend in den Geisteswissenschaften angewandt worden sind, ist das Potenzial der verfügbaren Daten für die Gewinnung neuer Erkenntnisse – um das Tagungsthema zu zitieren – längst noch nicht ausgeschöpft. Das BMBF-geförderte eHumanities-Projekt „Semantisch-soziale Netzwerkanalyse als Instrument zur Erforschung von Religionskontakten“ (SeNeReKo) ist mit der Anwendung und Weiterentwicklung von Methoden textbasierter Netzwerkanalyse befasst. Am Beispiel von religionshistorischen Korpora wie dem buddhistischen Pali-Kanon, dem altindischen Epos *Mahābhārata* und dem Thesaurus Linguae Aegyptiae, einer Sammlung altägyptischer Quellen, sollen die Erkenntnisse aus der dreijährigen Projektarbeit präsentiert werden.

Methoden der Netzwerkanalyse haben sich mittlerweile im Methodenkanon der *digital humanities* etabliert. Dafür lassen sich insbesondere zwei Gründe anführen: Zum einen entsprechen sie theoretischen Entwicklungen in den Geistes- und Sozialwissenschaften, die das Relationale gegenüber isolierten Entitäten betonen. Diese Tendenz findet sich etwa in der Geschichtswissenschaft (Verflechtungsgeschichte, Beziehungsgeschichte u.a., etwa Osterhammel 2001), in der Soziologie (Relational Sociology, Emirbayer 2007) oder in der Linguistik (Fuzzy Semantics, Rieger 1989). Zum anderen stellt die Graphentheorie ein formales Modell bereit, das relationale Systeme unterschiedlichster Provinienz repräsentieren und quantitativen Analysen zugänglich machen kann. Was die zentralen Netzwerkkomponenten, die Knoten und Kanten des jeweiligen Modells sind, unterscheidet sich dabei je nach Anwendungsfall. Netzwerkansätze in Geschichts- und Literaturwissenschaft beschäftigen sich zumeist mit (historischen oder literarischen) Akteuren und weisen damit eine gewisse Nähe zu *social network analysis* auf, wie sie sich im Kontext der Soziologie entwickelt hat (etwa Gramsch 2013). Ansätze in Linguistik und Computerlinguistik stellen dagegen eher auf sprachliche Strukturen ab, die als Netzwerke repräsentiert werden können (etwa Ferrer i Cancho, Solé, and Köhler 2004). Das SeNeReKo-Projekt hat zum Ziel, beide Perspektiven zu verbinden, um die Netzwerkanalyse für die Religionsforschung fruchtbar zu machen. Die interdisziplinäre Religionsforschung zeichnet sich dabei durch eine Pluralität sowohl der Gegenstände (nach Zeit, Ort und Genre) als auch der Zugänge (philologische, historische und soziologische) aus. Sie ist daher ein fruchtbares Feld, um neue methodische Ansätze zu erproben.

Der Vortrag stellt anhand ausgewählter, im SeNeReKo-Projekt bearbeiteter religionshistorischer Korpora zentrale Ergebnisse des Projekts vor. Im Zentrum der Methode stehen dabei Verfahren der textbasierten Netzwerk-Erzeugung, die Ausgangspunkt für unterschiedliche Analysestrategien sind.

Die Grundlage für die Untersuchung der altägyptischen Textzeugen bildet die digitale Datenbank des Thesaurus Linguae Aegyptiae (TLA) der Berlin-Brandenburgischen Akademie der Wissenschaften. Ihre Sammlung umfasst Textkorpora der unterschiedlichsten Gattungen, wie beispielsweise literarische Erzählungen, Jenseitsliteratur, königliche Dekrete oder magische Sprüche, und beinhaltet dabei mehr als 1.200.000 Textworte aus knapp 2.000 Jahren Geschichte.

Aus dem indischen Bereich beschäftigt sich das Projekt zum einen mit der als *Pāli-Kanon* bekannten Sammlung altbuddhistischer Texte (ca. 4.-1. Jh. v. u. Z.). Aus diesem wurden die aus religionswissenschaftlicher Sicht interessantesten Teile (im Umfang von ca. 1,7 Mio. Wörtern) ausgewählt. Zwar lagen diese Daten schon in digitalisierter Form vor, mussten jedoch für die durchgeführten Analysen in umfangreicher Weise aufgearbeitet werden. Zum anderen wird auch das in Sanskrit verfasste indische Nationalepos *Mahābhārata* (ca. 3. Jh. v. – 3. Jh. n. u. Z.) untersucht. Dieser umfangreiche Text (ca. 1 Mio. Wörter) ist ein dankbares Studienobjekt, da er eine Art Gründungsdokument des Hinduismus darstellt und daher verschiedene religiöse Strömungen seiner Entstehungszeit widerspiegelt. Die computergestützte Analyse des *Mahābhārata* wird dadurch erleichtert, dass ein großer Teil semantisch und syntaktisch annotiert vorliegt.

Um nun von den Textdaten zu Erkenntnissen zu gelangen, werden auf Basis linguistischer Annotationen alle relevanten Untersuchungseinheiten (Entitäten und Lemmata) und ihre relationale Struktur als Graph dargestellt und so für die Netzwerkanalyse erschlossen. Relationen von Einheiten werden dabei durch Kookkurrenzen definiert, wobei die Kontextfenster je nach Untersuchungsziel unterschiedlich weit gefasst werden können. Dies erlaubt dann unterschiedliche Analysen. Für die Bedeutungsanalyse einzelner Worte kann das Netzwerk auf ein zu erforschendes Lemma und alle mit diesem in Verbindung stehenden Begriffe reduziert werden. Auf diese Weise können sowohl kookkurrente Worte mit einer hohen Zentralität als auch mögliche Cluster engverwandter Begriffe aufgedeckt werden. Dieses Verfahren ermöglicht es, relevante Informationen aus einem umfangreichen Textkorpus (1) zu ermitteln und sie (2) durch verschiedene Visualisierungstechniken schneller und leichter interpretierbar zu machen. Für eine soziale Netzwerkanalyse kann für das Korpus (oder ausgewählte Teilkorpora) ein Netzwerk nur der sozialen Akteure erstellt werden. Ein besonderer Ansatz ergibt sich aus der Kombination beider Ansätze: In den jeweiligen Netzwerken können nicht nur alle mit einem bestimmten Lemma in Verbindung stehenden Begriffe als solche abgebildet, sondern auch die in diesem Kontext agierenden sozialen Akteure dargestellt werden. Darüber hinaus können über Kookkurrenzen einer bestimmten Entität nicht nur alle verbundenen Akteure aufgedeckt werden, sondern zudem gezeigt werden, mit welchen semantischen Begriffen diese verbunden sind. Anhand von konkreten Beispielen sollen die Verbindung aus semantischer und sozialer Netzwerkanalyse vorgestellt und ihre Möglichkeiten für die Fächer Religionswissenschaft, Ägyptologie und Indologie diskutiert werden.

Literatur

Emirbayer, Mustafa. 2007. 'Manifesto for a Relational Sociology.' *American Journal of Sociology* 103 (2): 281–317.

- Ferrer i Cancho, Ramon, Ricard V. Solé, and Reinhard Köhler. 2004. 'Patterns in Syntactic Dependency Networks'. *Physical Review E* 69 (5): 051915.
doi:10.1103/PhysRevE.69.051915.
- Gramsch, Robert. 2013. *Das Reich als Netzwerk der Fürsten: politische Strukturen unter dem Doppelkönigtum Friedrichs II. und Heinrichs (VII.) 1225-1235*.
- Osterhammel, Jürgen. 2001. *Geschichtswissenschaft jenseits des Nationalstaats: Studien zu Beziehungsgeschichte und Zivilisationsvergleich*. Kritische Studien zur Geschichtswissenschaft ; 147. Göttingen: Vandenhoeck & Ruprecht.
- Rieger, Burghard B. 1989. *Unscharfe Semantik: Die Empirische Analyse, Quantitative Beschreibung, Formale Repräsentation Und Prozedurale Modellierung Vager Wortbedeutungen in Texten*. Frankfurt am Main u. a.: Lang.