

Erkenntnisse aus der digitalen Auswertung der „Listen ziviler Dienstränge“ aus dem späten Zarenreich

Benedikt Tondera

1. Einführung

2012 stellten Stephen Ramsay und Geoffrey Rockwell die These auf, dass Softwareanwendungen als Theoriemodelle (*theoretical frameworks*) für geisteswissenschaftliche Forschung betrachtet werden können. Gerade Visualisierungstools funktionierten wie „Teleskope für das Gehirn“ (Ramsay/Rockwell 2012: 79). Ähnlich wie theoretische Literatur trügen sie dazu bei, neue Sinnhorizonte zu kulturellen Artefakten zu erschließen. Das Problem bestünde darin, die Funktionsweise der Software und damit die programmiersprachlichen Grundannahmen, die in die jeweiligen Tools einfließen, offenzulegen und zu verstehen. Wie kritisch dieser Aspekt ist, lässt sich am Beispiel von *deep learning*-Technologien in den *digital humanities* (DH) nachvollziehen. Bei Verfahren wie dem *topic modeling*, bei dem durch maschinelles Lernen signifikante Worthäufungen (*topics*) aus Textkorpora erzeugt werden, stellt sich die Frage, ob und wie die Ergebnisse genutzt werden können, wenn die zugrundeliegenden Rechenschritte nicht rekonstruierbar sind und die Wirkungsweise der komplexen Algorithmen nur abstrahiert nachvollzogen werden kann (Hodel et al. 2022). Darüber hinaus muss auch das sogenannte *preprocessing* – also die Vorverarbeitung von Textdaten, um sie maschinenlesbar zu machen (vgl. ebd.: 188 f.) – kritisch reflektiert werden: Welche Implikationen hat es, wenn strukturierter Text für den Computer einerseits seiner grammatischen, orthographischen und semantischen Struktur entledigt und in „Wortbeutel“ (*bag of words*) verpackt wird, andererseits ein großer Teil des Textes als nicht bedeutungsrelevante Partikel oder Füllwörter schon vor dem Analyseprozess ausgesondert wird?

Im Gegensatz zu den komplexen methodischen und erkenntnistheoretischen Fragen, die der Einsatz von *machine learning* für geisteswissenschaftliche Zwecke aufwirft, scheinen prosopografische Tools wie das in Stanford entwickelte *Palladio*¹ und das *nodegoat*² in dieser Hinsicht unproblematischer: Statt mit unübersichtlichen Textmassen wird hier mit tabellarisch vorsortiertem Datenmaterial gearbeitet, das eindeutigen Kategorien zugeordnet ist (in der Regel Personen-, Raum-, Objekt- und Zeitdaten). Daraus generieren die genannten Programme Karten- und Netzwerk-Visualisierungen. Selbst wenn die Algorithmen der Netzwerkgraphen den meisten AnwenderInnen ähn-

1 <http://hdlab.stanford.edu/palladio/> (26.10.2023).

2 Vgl. Bree/Kessels (2013) und *nodegoat*: a web-based data management, network analysis & visualisation environment, <https://nodegoat.net> from LAB1100, <https://lab1100.com> (26.10.2023).

lich unverständlich bleiben dürften wie beim *topic modeling*, so arbeiten diese immerhin deterministisch und erzeugen intuitiv verständliche Ergebnisse. Anders gesagt: Derselbe Input führt zu demselben Output, und die erzeugten Cluster bilden die Nähe oder Distanz der untersuchten Objekte zueinander ab. Während die Arbeit mit Tools wie *nodegoat* und *Palladio* daher auf den ersten Blick transparenter und verständlicher erscheint als KI-gestützte Verfahren wie das *topic modeling*, lautet eine weiter unten ausgeführte These meines Beitrags, dass sie anders gelagerte methodische Schwierigkeiten aufwerfen. Diese lassen sich zwei Bereichen zuordnen, erstens der Objektorientierung und zweitens einer Tendenz zum Positivismus beim Umgang mit Quellenmaterial. Mit „Schwierigkeiten“ sind in diesem Fall nicht Nachteile oder Unzulänglichkeiten gemeint, sondern Sachverhalte, die meines Erachtens bei dem Umgang mit den von diesen Tools erzeugten Visualisierungen berücksichtigt werden sollten.

Darüber hinaus wird in diesem Beitrag der Versuch vorgestellt, mithilfe von *nodegoat* Teile eines theoretischen Konzepts aus den Geschichtswissenschaften digital umzusetzen. Es handelt sich um den Ansatz der „imperialen Biographien“, der von Tim Buchen und Malte Rolf in den 2010er-Jahren entwickelt wurde (Rolf 2014; Buchen/Rolf 2015). Zusammengefasst geht es dabei darum, die kollektivbiographische Spezifik der Lebensläufe imperialer Eliten im 19. Jahrhundert zu erfassen. Der Grundgedanke ist, dass hohe Beamte, Wissenschaftler und Politiker (in den meisten Fällen tatsächlich nur Männer) an der gesellschaftlichen Spitze besondere Lebensmuster in Bezug auf die berufliche Mobilität, die Ausbildung und die kollektiven Selbstbilder aufwiesen, die spezifisch für den Kontext von Großreichen waren. Ein zentraler Aspekt ist die räumliche Dimension der betrachteten Lebensläufe, also die Frage, wie imperiale Eliten die politisch, kulturell und geographisch häufig sehr disparaten Karrierestationen gedanklich miteinander verbanden und auf diese Weise zur Entstehung einer kollektiven Vorstellung davon beitrugen, was die Großreiche im Inneren zusammenhielt. Das Konzept selbst ist in den vergangenen Jahren insbesondere in Hinblick auf die zu hinterfragende Spezifik der Elitenbiographien in Imperien im Unterschied zu Nationalstaaten kritisch diskutiert worden; darüber hinaus existieren ähnliche Konzepte, die wie die „imperialen Biographien“ Anstöße aus der „neuen Biographik“ (Raulff 2002) aufgenommen haben und sich mit dem Einfluss von Räumen, Mobilität und kollektiven Diskursen auf die Identitätsbildung auseinandersetzen (Rolf/Tondera 2021; Aust/Schenk 2021).

Anstatt diese theoretischen Erörterungen zu vertiefen, soll in diesem Beitrag untersucht werden, inwiefern sich abstrakte Konzepte wie „Ämterrotation“, „berufliche Mobilität“ oder „Peripherie“ mithilfe von *nodegoat* anhand eines prosopografischen Samples in den digitalen Raum transferieren lassen. Insbesondere ist dabei von Interesse, welche Reibungsverluste, aber auch welche neuen Einsichten bei der Digitalisierung geisteswissenschaftlicher Konzepte entstehen. Das hier vorgestellte digitale Fallbeispiel befindet sich noch in einer recht frühen Bearbeitungsphase, insofern handelt es sich bei den folgenden Ausführungen um einen klassischen Werkstattbericht.