FREDERIK ELWERT

Welcher Text, warum Computer? Computergestützte Textanalyse im Kontext der Religionswissenschaft

Abstract

Der Beitrag diskutiert anhand der Forschungsbiografie des Autors die Rolle computationeller Textanalyse in der Religionswissenschaft als nicht bekenntnisgebundener Disziplin mit sozialwissenschaftlichen sowie historisch-philologischen Methodentraditionen. Daraus lassen sich drei zentrale Dimensionen von Text in der Forschung ableiten: Religiöse Literaturen, religionshistorische Quellen sowie sozialwissenschaftliche Dokumente. Die computergestützte Textanalyse hat dabei das Potenzial, gemeinsame Perspektiven auf das Medium Text zu stärken. Abschließend diskutiert der Beitrag die Rolle, die Forschende mit geistes- und sozialwissenschaftlichem Hintergrund in Bezug auf die computergestützte Textanalyse einnehmen können. Während ein Großteil der Methodenentwicklung im engeren Sinne in Fächern wie der Computerlinguistik oder Informatik stattfindet, bleiben Methodenanwendung, Methodenvermittlung und auch fachbezogener Methodenerforschung als lohnende Betätigungsfelder.

Based on the author's research biography, the article discusses the role of computational text analysis in religious studies as a non-denominational discipline with methodological traditions in the social sciences and historical philology. From this, three central dimensions of text can be derived: As religious literatures, historical sources, and social science documents. Computational textual analysis has the potential to strengthen shared perspectives on the medium of text. Finally, the article discusses the role that researchers from the humanities and social sciences can play in computational text analysis. While much of the development of methods in the narrower sense takes place in fields such as computational linguistics or computer science, the application of methods, the teaching of methods, and subject-related methodological research remain fruitful areas of activity.

Keywords: Computergestützte Textanalyse, Interdisziplinarität, Netzwerkanalyse, Philologie, Religionswissenschaft, topic modeling

computational text analysis, interdisciplinarity, network analysis, philology, religious studies, topic modeling

Osnabrücker Beiträge zur Sprachtheorie 104 (2025), 109-123.
Published online 20-05-25; DOI https://doi.org/10.17192/obst.2025.104.8808

1 Computergestützte Textanalyse per Zufall

Als ich angefragt wurde, für diesen Band einen Beitrag zu schreiben, war ich erst einmal sofort angetan von dem Vorhaben und dem etwas ungewöhnlichen Format, Forschungsbiografien mit allgemeineren Entwicklungen in verschiedenen Disziplinen zu verbinden. Im zweiten Moment habe ich dann einigen Respekt vor der Aufgabe bekommen. Kann ich überhaupt für „meine" Disziplin, die Religionswissenschaft, sprechen? Ist meine Forschungsbiografie, die mich zeitweise zu mehr Digital-Humanities-Tagungen als religionswissenschaftlichen geführt hat, überhaupt in irgendeiner Weise indikativ für allgemeinere Entwicklungen im Fach? Und andersherum, zwar habe ich neben unterschiedlichen anderen Methoden auch mit computergestützter Textanalyse gearbeitet, aber kann ich für mich in Anspruch nehmen, Forschender in diesem Feld zu sein? Im Folgenden möchte ich die Herausforderung annehmen und zum einen anhand meines Werdegangs einige Entwicklungen aus meiner subjektiven Perspektive nachzeichnen, und zum anderen daran einige Reflexionen über deren Bezug zu disziplinären Konstellationen anschließen. Ob dies fruchtbar ist, mögen die Lesenden beurteilen.

Die Religionswissenschaft ist nicht unbedingt in erster Linie eine Textwissenschaft, und ich habe mich nie als Text- geschweige denn Sprachwissenschaftler gesehen. Dass ich für mich die computationelle Textanalyse als Betätigungsfeld entdeckt habe, war daher im Prinzip ein Zufall. Oder vielmehr: Sie drängte sich auf, als ich eigentlich ein anderes Problem lösen wollte. Anfang der 2010er Jahre war immer noch die Frühzeit der Digital Humanities in Deutschland, aber das Feld nahm so langsam an Fahrt auf. 2011 veröffentlichte das BMBF eine Bekanntmachung zur Förderung von eHumanities-Projekten. Eine richtige Vorstellung davon, was mit computationellen Methoden möglich sein könnte, hatten wir alle nicht. Wir waren allerdings zu der Zeit an relationalen Gefügen in der Religionsgeschichte interessiert (Krech 2020). Nach meinem Studium hatte ich kurz in einem Projekt zu Politiknetzwerken gearbeitet, und dort war eine Methode entwickelt worden, um aus öffentlich verfügbaren Policy-Dokumenten verschiedener europäischer Akteure Informationen über Netzwerkbeziehungen zu extrahieren (Nagel 2008). Daher schlug ich als mögliches Projekt eine Adaption dieser Methode für religionshistorische Texte und unter Einbezug computergestützer Werkzeuge vor. Der informatische Anteil in dem Projekt sollte zunächst sein, ein Tool für die manuelle Annotation von Netzwerkbeziehungen in Texten zu entwickeln – weiter reichte unsere Fantasie in Bezug auf die mögliche Rolle des Computers in diesem Zusammenhang nicht. Das Trierer „Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften", wie das „Trier Center for Digital Humanities" damals hieß, übernahm nach ersten Sondierungen die Rolle des informatischen Partners

und brachte erstmals ins Spiel, dass man Entitäten in Texten ja auch automatisiert erkennen könne und dass eine Analyse syntaktischer Muster auch bei der Erkennung von Beziehungen im Text helfen könne. Daraus entwickelten sich dann die Konturen unseres Vorhabens.

Das Projekt „Semantisch-soziale Netzwerkanalyse als Instrument zur Erforschung von Religionskontakten" (SeNeReKo) wurde schließlich gefördert, und das Team bestand aus einer Ägyptologin und einem Indologen, die jeweils die von ihnen untersuchten Texte als Korpora einbrachten, einem Informatiker mit Erfahrung in der Anwendungsentwicklung und mir – mit der vorgesehenen Rolle als Projektkoordinator. Als „Mittler" zwischen den geisteswissenschaftlichen Kolleg*innen und dem Informatiker fiel es mir zunehmend zu, überhaupt erst einmal herauszuarbeiten, wie wir unsere inhaltlichen Interessen methodisch adäquat adressieren könnten. Ziel des Projekts war, nicht nur soziale Beziehungen (Strukturen) in Texten zu identifizieren, sondern auch die diesen Beziehungen zugeschriebenen Bedeutungen (Semantik). Eine Prämisse des Projekts war dabei, dass wir für die semantische Bestimmung der Beziehungen nicht einfach auf vordefinierte Kategorien setzen, sondern die Konnotationen induktiv aus dem Quellenmaterial selbst herausarbeiten wollten. Das erforderte, den Text nicht nur als Steinbruch für die Extraktion von Informationen über soziale Relationen zu verstehen, sondern zunehmend Fragen von Bedeutungsmustern im Text selbst in den Blick zu nehmen.

Niemand von uns hatte eine (computer-)linguistische Vorbildung. Meine geisteswissenschaftlichen Kolleg*innen hätten sich zwar durchaus als Philolog*innen bezeichnet, aber erst nach und nach wurde mir klar, dass die formale Linguistik hier noch einmal mit ganz anderen Konzepten operiert. Und auch unser informatischer Kollege hatte wenig Erfahrung im Umgang mit Text als Daten. Wir tasteten uns also langsam heran, wie sich unsere Idee von Text und Netzwerkanalyse realisieren ließe. Und gleichzeitig musste sich auch erst ein Modus der interdisziplinären Zusammenarbeit entwickeln. Für einige Aufgabenstellungen war es sehr hilfreich, einen Informatiker im Team zu haben, aber für die iterative Entwicklung und Verfeinerung unserer Analysemethoden waren die notwendigen Kommunikationsschleifen irgendwann zu langwierig.

Ich hatte zuvor schon ein wenig Erfahrung in der Programmierung mit R und Python gesammelt. Das meiste davon war bisher reines Hobby gewesen. Nur gelegentlich hatte ich diese – noch sehr basalen – Kompetenzen in einem akademischen Kontext eingesetzt, etwa als Hilfskraft in der Bremer Soziologie für die automatisierte Erstellung von Evaluationsberichten in einem Projekt zur gerade erst aufkommenden Praxis regelmäßiger Online-Lehrevaluationen. Mit meiner Arbeit in der Religionswissenschaft hatte dieses private Interesse aber so gut wie nichts zu tun – auch wenn allgemein meine Affinität für Technik sicherlich ein Grund war, warum man mir diese Koordinationsposition zutraute. Jetzt

war es irgendwann aber einfacher, unsere methodischen Überlegungen selbst in Code umzusetzen, als erst detailliert zu beschreiben, was uns vorschwebt, und dann auf die Umsetzung zu warten. Zunehmend wandelte sich so meine Rolle, und ich war nicht mehr nur Vermittler, sondern ein Stück weit ein zweiter Programmierer. So war eines der ersten Ergebnisse aus dem Projekt, dass ich den von Paranyushkin (2011) beschriebenen Algorithmus für die Erzeugung eines Textnetzwerks implementiert habe. Daraus ist dann später das Softwarepaket TCFnetworks hervorgegangen (Elwert 2015).

Diese Form von „Text als Netzwerk", mit der wir uns dann in erster Linie beschäftigten, hatte erst einmal nicht mehr viel mit der ursprünglichen Idee von sozialen Netzwerken und ihren semantischen Konnotationen zu tun. Sie erlaubte uns aber, uns allmählich an den algorithmischen Umgang mit Text anzunähern. Und über pragmatische Anforderungen wie die Implementierung von Wortartenfiltern, den Einbezug von syntaktischen Dependenzstrukturen für die Netzwerkgenerierung oder die Erkennung von *Named Entities* lernten wir auch zunehmend mehr über die zu der Zeit aktuellen Möglichkeiten der automatischen Sprachverarbeitung bzw. Computerlinguistik. Irgendwann stießen wir dann auf ein Paper, das Entity-Netzwerke in Texten mit einem angepassten Topic-Modeling-Verfahren kombinierte, um so Entity- und Relationsbeschreibungen aus Texten zu lernen (Chang/Boyd-Graber/Blei 2009). Es gelang uns, dieses Verfahren – dank einer verfügbaren R-Implementierung – auf das Sanskrit-Epos Mahābhārata anzuwenden (Elwert/Sellmer 2017, siehe Abbildung 1). Damit waren wir nicht nur unserer ursprünglichen Idee doch noch relativ nahe gekommen, sondern hatten zudem auch erste Erfahrungen mit Machine-Learning-Verfahren als analytische Werkzeuge jenseits der computerlinguistischen Vorverarbeitung gemacht.

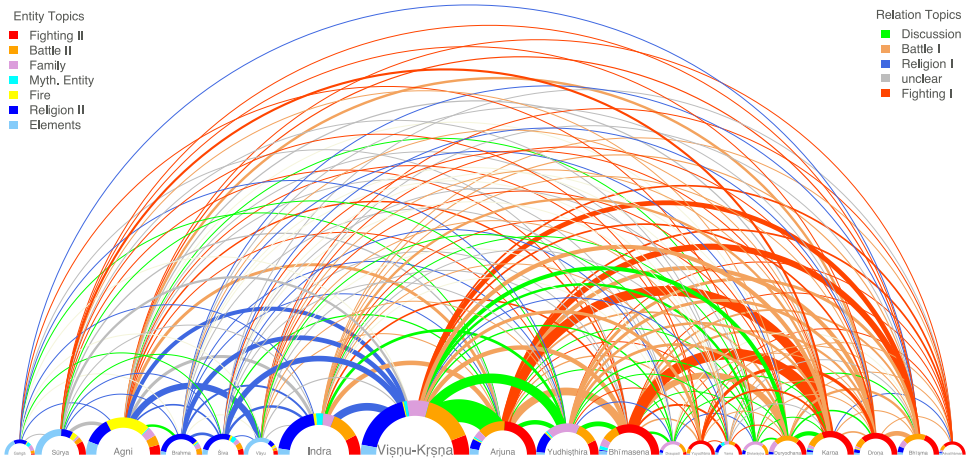

Abb. 1: Visualisierung zentraler Figuren und ihrer Beziehungen im Epos Mahābhārata auf Basis des nubbi-Modells

2 Ein Schritt zurück: Texte, Zahlen, Religionswissenschaft

Individuell waren diese Erfahrungen in den emergierenden Digital Humanities für uns sicherlich wegweisend, und strukturell haben sie auch zur zunehmenden Etablierung computergestützter Methodik am Centrum für Religionswissenschaftliche Studien, zu einem gewissen Maße vielleicht sogar an der Ruhr-Universität Bochum überhaupt, geführt. Aber inwieweit sind diese Entwicklungen charakteristisch für die Religionswissenschaft als Disziplin?

Die Religionswissenschaft ist ein relativ kleines Fach, dessen spezifisches Profil oft nicht sehr sichtbar und bekannt ist. Zudem ist die Abgrenzung zu anderen Disziplinen nicht immer einfach – sowohl aus der Außensicht, aber teilweise auch im Fach selbst. Was für viele Religionswissenschaftler*innen relativ klar ist, ist die Abgrenzung von den Theologien als bekenntnisgebundenen Wissenschaften.[1] Die Religionswissenschaft versteht sich im Gegensatz zu den Theologien als nicht an ein spezifisches Bekenntnis gebundenes, kulturwissenschaftliches Fach, das Religion als Teil menschlicher Kultur in Geschichte und Gegenwart empirisch erforscht. Während der Gegenstand Religion – und dieser nicht bekenntnisgebundene Blick auf diesen Gegenstand – der zentrale Fluchtpunkt des Faches ist, sind die Zugänge und Methoden sehr divers und häufig aus anderen Disziplinen übernommen. Im Kern kann man hier zwei Ansätze

1 Auch wenn es gelegentlich die Bezeichnung „Religionswissenschaft“ für Teilgebiete der interkulturellen Theologie gibt, was die Grenzziehung nicht einfacher macht.

unterscheiden: Einen sozialwissenschaftlichen Zugang, der stark von der empirischen Kulturwissenschaft und der qualitativen Sozialforschung beeinflusst ist, und einen historisch-philologischen Zugang, der sich gerade auch globalgeschichtlich mit historischen Texten und Zeugnissen religiöser Traditionen in verschiedenen Sprachen beschäftigt (Seiwert 2014).

Da sich diese vielfältigen methodischen Zugänge nur bis zu einem gewissen Grad in einem Studiengang vermitteln lassen, ist die konkrete methodische Praxis der Religionswissenschaftler*innen häufig auch von ihren weiteren Studienfächern abhängig. Wer sich tatsächlich intensiv mit der Geschichte des Buddhismus auseinandersetzen will, greift hierfür häufig auf die Sprachkenntnisse aus einem Indologie- oder Sinologiestudium zurück. Wer gegenwärtige religiöse Phänomene empirisch untersuchen will, nutzt dagegen Methodenkenntnisse etwa aus einem sozialwissenschaftlichen Fach. So war es auch bei mir: Neben Religionswissenschaft habe ich im Magister als zweites Hauptfach Soziologie studiert. Hierüber bin ich auch mit Grundzügen der quantitativen Sozialforschung und der Statistik in Kontakt gekommen – auch wenn wir immer etwas neidisch auf die noch einmal umfangreichere Statistikausbildung der Diplom-Soziolog*innen geschaut haben. Im Religionswissenschaftsstudium dagegen habe ich auch historische Seminare besucht und zumindest in Bruchstücken Sanskrit und Latein gelernt. Nicht genug, um damit sinnvoll arbeiten zu können, aber doch so viel, dass ich ein grundlegendes Verständnis für philologisches Arbeiten erworben habe.

Dieser Kontext ist noch einmal wichtig, um sich der Frage zu nähern, was überhaupt „Text" in der Religionswissenschaft ist und wo es überhaupt ein Potenzial für computergestützte Analysen gibt. Etwas vereinfachend kann man drei Aspekte von Text unterscheiden, die in der Religionswissenschaft analysiert werden: Zum einen religiöse Literaturen, bestehend aus den sogenannten heiligen Schriften und der kanonischen, aber auch nicht-kanonischen Überlieferung. Zentral ist hier, dass sie in der Regel als umfassendere Textsammlungen überliefert (und nicht selten von religiösen Institutionen selbst zugänglich gemacht) worden sind. In religionswissenschaftlicher Perspektive sind diese Texte eine wichtige Quelle zur Rekonstruktion einer Ideengeschichte religiöser Traditionen. Zum zweiten spielen in religionshistorischer Perspektive Texte eine Rolle, wie sie auch in der Geschichtswissenschaft relevant sind: als Zeugen bestimmter historischer Prozesse und Konstellationen. Texte in diesem Zusammenhang, etwa Urkunden oder Briefe, sind häufig weniger als kohärente Sammlungen relevant denn als punktuelle Manifestationen historischer Wirklichkeit. Und drittens sind Texte im Rahmen der sozialwissenschaftlichen Religionsforschung relevant, etwa als Transkripte von Interviews. In diesem Zusammenhang wird auch textbasierte digitale Kommunikation zunehmend in den Blick genommen,

die einen Einblick in den Austausch religiöser Gemeinschaften in bislang kaum erreichtem Umfang bietet.

Der Blick auf „Text" ist also sehr divers, und dementsprechend unterscheiden sich auch die Potenziale computergestützter Verfahren der Textanalyse und ihre Anknüpfungspunkte an die Methodenentwicklung in anderen Disziplinen, von den Computational Literary Studies über die Digital History bis hin zu den Computational Social Sciences.

3 Text als Forum und Grenze

Der Zugang zu Text ist in der Religionswissenschaft in erster Linie ein instrumenteller: Texte sind in vielen Fällen wichtige Quellen (oder auch die einzig verfügbaren), aber das Interesse der Religionswissenschaft gilt in erster Linie nicht dem Text an sich. Auch in meiner eigenen Forschung beschäftige ich mich zwar intensiv mit computationeller Textanalyse, aber je nach Anwendungsfall eben auch mit Netzwerkanalyse oder Knowledge Graphen. Zugleich ist Text als Medium aber auch nicht beliebig, die Entscheidung, religionswissenschaftliche Forschung auf Text als Daten zu stützen, hat konkrete Konsequenzen.

Zunächst einmal zeigt sich in der Religionswissenschaft mit ihren heterogenen Methodentraditionen ein Potenzial von computationeller Textanalyse, das auch über das Fach hinaus relevant ist, nämlich geisteswissenschaftlich-historische und sozialwissenschaftliche Trennlinien überbrücken zu können. In Bezug auf etablierte Zugänge zur Textanalyse gibt es in den Forschungstraditionen begründet kaum Berührungspunkte, etwa zwischen philologischer Hermeneutik und den Verfahren der qualitativen Sozialforschung. Mit der Hinwendung zur computationellen Textanalyse, die für beide Bereiche neu ist, zeigen sich Gemeinsamkeiten und mögliche Schnittstellen viel klarer.[2] So haben wir Methoden, die im SeNeReKo-Projekt auf religionshistorische Texte angewendet wurden, später auch auf religiöse Online-Kommunikation übertragen (Elwert/Tabti/Pfahler 2020, siehe Abbildung 2). Und auch in der Lehre lassen sich hier interdisziplinäre Verbindungen herstellen. In Bochum habe ich gemeinsam mit Kolleg*innen aus der Germanistik, Philosophie und Soziologie etwa eine Summer School zur digitalen Analyse großer Textkorpora durchgeführt, die für Studierende aller Fachbereiche geöffnet war. Es war vielversprechend, zu sehen, wie die geteilten Methoden hier einen intensiven Austausch zwischen ansonsten weitgehend getrennten Fachkulturen ermöglicht haben.

2 Innerhalb der Sozialwissenschaften lässt sich auf einer anderen Ebene ein vergleichbares Potenzial ausmachen, nämlich in Bezug auf die Überbrückung der Kluft zwischen qualitativen und quantitativen Ansätzen der Sozialforschung.

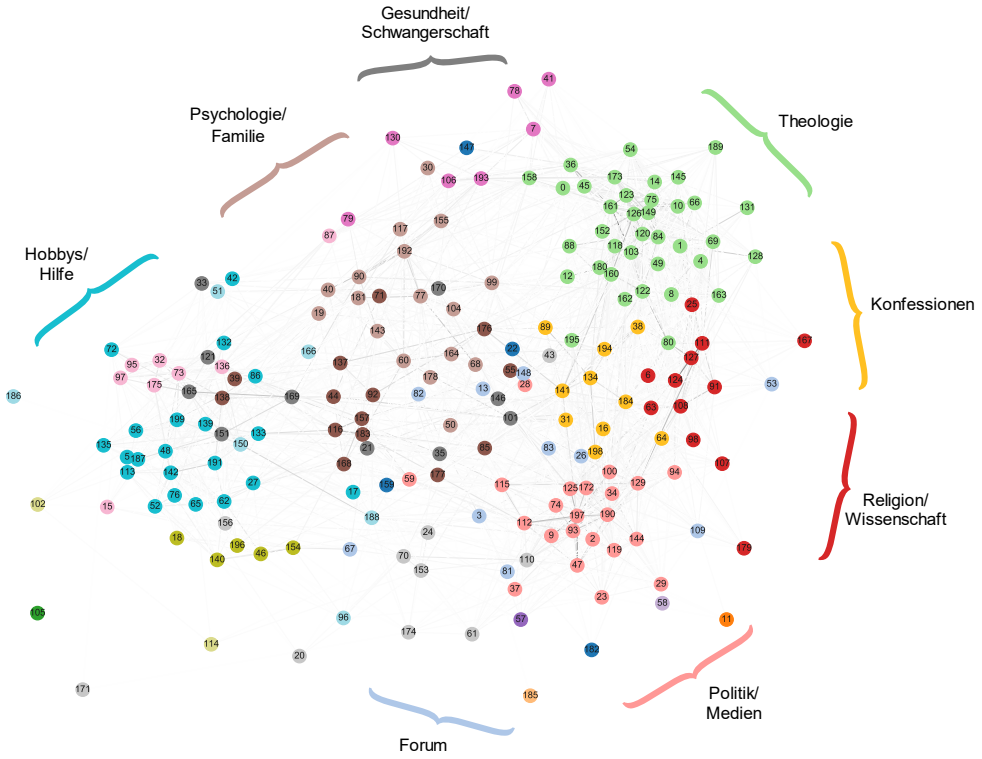

Abb. 2: Topics eines LDA-Modells auf Grundlage eines großen christlichen Online-Forums, geclustert

Gleichzeitig stellt Text als Gegenstand aber auch eine Grenze dar. Hier gibt es durchaus ernstzunehmende Vorbehalte von Religionswissenschaftler*innen. In der Fachgeschichte hat sich die Religionswissenschaft erst in jüngerer Zeit von den stark textfokussierten Forschungspraktiken emanzipiert. Neuere Ansätze wie Material Religion (Hazard 2013) oder Lived Religion (Knibbe 2020) werden hier explizit als Gegenentwürfe positioniert. Mit dem Fokus auf Text, so die Kritik, gehe eben auch eine Betonung bestimmter Teilaspekte von Religion (nämlich der kognitiv-glaubensförmigen) sowie eine Bevorzugung einzelner religiöser Traditionen (nämlich der sogenannten Schriftreligionen und hier in erster Linie des Christentums) einher. Eine Kritik an computationeller Textanalyse ist also nicht allein als Verharren im Vertrauten zu verstehen, sondern geht mit bestimmten und berechtigten methodologischen Anfragen einher. Und in der Tat sehe ich hier auch eine Gefahr in den Digital Humanities, dass zugunsten des Machbaren weiterführende Theorieentwicklungen ignoriert werden.

In diesem Zusammenhang lohnt sich auch ein Blick in ein Forschungsfeld, das sich innerhalb der Religionswissenschaft zunehmend etabliert hat: die Forschung zu „digital religion“, also religiöser Kommunikation und Praxis in digital

vermittelten Medien. Hier sieht man, dass sich auch unsere Gegenstände selbst und mit ihnen der Stellenwert von Text verändern: Während textbasierte Kommunikation für die frühe Internetkultur zentral war, kommen mit neueren sozialen Medien auch zunehmend Bild und Video als Genres hinzu. Eine Religionsforschung, die diesen Teil des Materials schlicht ignoriert, droht hier zu kurz zu greifen. Zugleich lassen sich unsere computationellen Methoden nicht so ohne weiteres auf Bilder übertragen. Zwar macht auch die computationelle Bildanalyse große Fortschritte, ist aber in den meisten Fällen auf vortrainierte Modelle angewiesen, deren Biases gerade im Bereich von Religion bei der Anwendung zu Forschungszwecken große Vorsicht erfordern (Berg/Valaskivi 2023).

Bei der Einführung von Methoden computationeller Textanalyse in der Religionswissenschaft ist daher auch eine Sensibilität für die Grenzen der Methoden und ihre Passung für konkrete Forschungsfragen, aber auch theoretische Perspektiven zentral, um mit den Fachkolleg*innen auf Augenhöhe sprechen zu können. Und an diesem Umgang zeigt sich auch die Positionierung von methodenbezogener Forschung im Verhältnis zur „Herkunftsdisziplin". Forschende, die sich mit digitalen Methoden befassen, haben häufig die Tendenz, dies entweder als neues, ganz eigenes Feld zu verstehen, oder aber als quasi evolutionäre Weiterentwicklung. Im ersten Fall gibt es kaum noch die Notwendigkeit eines Austauschs, und im zweiten besteht die Gefahr eines Überlegenheitsgestus auf der einen und einer entsprechend defensiven Abwehrhaltung auf der anderen Seite. In meiner Funktion als methodisch ausgerichteter Forscher an einem religionswissenschaftlichen Institut hat mich daher lange die Frage umgetrieben, wie ich meine Rolle in Bezug auf die Digital Humanities als Forschungsfeld einerseits und auf meine geistes- und sozialwissenschaftlichen Kolleg*innen andererseits definiere.

4 Methodenentwicklung, Methodenvermittlung, Methodenerforschung

In der damaligen Bekanntmachung des BMBF war die Erwartung an die Rollenverteilung in der Methodenentwicklung relativ klar:

> Anhand fachspezifischer Forschungsfragen sollen geisteswissenschaftliche Fachgebiete und qualitative Sozialforscher neue Forschungsmethoden insbesondere unter Nutzung von Informatikmethoden (z. B. Textmining, Text Reuse, Mapping, Visualisierung, Komponenten zur Analyse von Bewertungen u. a.) in ihren Fachgebieten entwickeln. Die Bekanntmachung fokussiert dabei nicht auf spezifische Disziplinen. Erwartet wird eine intensive interdisziplinäre Zusammenarbeit zwischen Geistes-, Kultur- oder Sozialwissenschaftlern einerseits und Informatikern andererseits bei der Lösung einer spezifischen Forschungsfrage (Bundesministerium für Bildung und Forschung 2011).

Die Methoden für das neue Gebiet der eHumanities als *enhanced humanities* sollen also aus der Informatik übernommen werden, während die jeweilige Fragestellung aus den Geisteswissenschaften oder den qualitativ arbeitenden Sozialwissenschaften kommt. In unseren Projekten hatten wir jedoch oft ganz andere Konstellationen der Interdisziplinarität. Im SeNeReKo-Projekt waren wir schon auf religionswissenschaftlicher Seite ein Indologe (zunächst Sven Wortmann, später Sven Sellmer), eine Ägyptologin (zunächst Beate Hofmann, später Simone Gerhards) sowie ich als Sozialwissenschaftler. Unsere Zugänge zu Text waren auch komplett unterschiedlich: Während die Indologie auf umfassende, gut überlieferte und digitalisierte Texttraditionen zurückgreifen kann – wir haben vor allem mit dem buddhistischen Pali-Kanon und dem Sanskrit-Epos Mahābhārata gearbeitet –, ist die Textüberlieferung in der Ägyptologie viel stärker fragmentarisch. Vor allem aber waren es auch epistemologische Fragen, die wir diskutiert haben, um die Angemessenheit verschiedener Verfahren zu eruieren. Wie verhalten sich etwa unsere zeitgenössischen Analysekategorien zu den emischen Kategorisierungen historischer Gesellschaften? Wie lässt sich aus einem dogmatischen oder literarischen Text auf dahinterliegende religiöse Abgrenzungsprozesse schließen? Der Beitrag von Jürgen Knauth als Informatiker im Team war entsprechend sehr hilfreich bei der praktischen Entwicklung unserer Forschungsinfrastruktur, etwa wenn es um die linguistische Annotation indischer Texte ging. Auf der Ebene der *Methodologie* selbst hingegen war die Auseinandersetzung mit den Möglichkeiten und Implikationen digitaler Methoden innerhalb des religionswissenschaftlichen Projektteams viel zentraler.

Dabei ist es im Rahmen der Digital Humanities durchaus eine berechtigte Frage, ob eine echte Methodenentwicklung im Sinne der Einführung neuer oder zumindest substanziellen Weiterentwicklung bestehender Methoden durch *digital humanists* mit geisteswissenschaftlichem Hintergrund geleistet werden kann. In unserem Fall habe ich einfache Verfahren zur Extraktion von Netzwerken aus Texten selbst implementiert, aber dies auch mit einer Portion Dilettantismus. Trotzdem war diese intensive Auseinandersetzung mit den Daten selbst, mit den Anforderungen an eine algorithmische Verarbeitung und mit den methodischen Grundlagen der jeweiligen Ansätze unverzichtbar für eine echte Auseinandersetzung mit den Möglichkeiten und Implikationen computergestützter Forschungsmethoden. Kein Gespräch mit einem Informatiker während einer Projektbesprechung hätte das ersetzen können.

Gleichzeitig gibt es natürlich Methoden, die eine derart vertiefte Kenntnis über grundlegende Fragen von Informatik und Statistik erfordern, dass hier kaum substanzielle Beiträge von Forschenden erwartet werden können, die nicht sehr tief in diesen Forschungstraditionen verankert sind. Ein mittlerweile schon historisches Beispiel ist die Entwicklung von LDA als richtungsweisendes Verfahren für das Topic Modeling (Blei/Ng/Jordan 2003). Und die Einführung

neuerer Ansätze wie BERT (Devlin et al. 2019) hat diese Entwicklung noch einmal zementiert.

Im Rahmen eines späteren Projekts, relNet, haben wir zusammen mit Informatiker*innen von der TU Dortmund an der Erforschung religiöser Online-Kommunikation gearbeitet. Die Arbeitsgruppe um Katharina Morik beschäftigt sich aus der Perspektive der Informatik intensiv mit Anwendungen künstlicher Intelligenz für die Sprachverarbeitung. Dies gab uns die Gelegenheit, tatsächlich die Methodenentwicklung innerhalb der Informatik mit religionswissenschaftlichen Fragestellungen zu verbinden. Lukas Pfahler entwickelte hierfür etwa einen alternativen Topic-Modeling-Ansatz, der dynamisch eine nachträgliche Anpassung der Topic-Anzahl sowie eine interaktive Verschmelzung oder Auftrennung gefundener Topics ermöglichte (Pfahler et al. 2018). Letztlich hat sich aber auch hier gezeigt, dass ein gutes Verständnis der jeweiligen Methode und eine iterative, schnelle Anpassung der Analyse- und Visualisierungsmöglichkeiten, die wir auch auf eigene Faust umsetzen konnten, für die inhaltliche Arbeit ertragreicher war.

Daraus ergibt sich für mich eine relevante Reflexion in Bezug auf die eigene Rolle als jemand, der mit einem sozial- und geisteswissenschaftlichen Hintergrund im Bereich der computergestützten Textanalyse aktiv ist. Sollten wir unsere Aufgabe in der Methodenentwicklung sehen? Oder geht es in erster Linie um die Anwendung bestehender Tools, die für die eigene, inhaltliche Forschung eingesetzt werden?

Zunächst einmal spricht meines Erachtens gar nichts gegen die Anwendung. In Fächern wie der Informatik oder der Computerlinguistik herrscht in Bezug auf computationelle Textanalyse ein großer Innovationsdruck hinsichtlich der Methoden: Die neue Methode ist gerade das publikationswürdige Ergebnis. In den Digital Humanities kann man gut ähnliche Tendenzen beobachten. Neue, innovative Methoden können die Forschung auch durchaus voranbringen, aber dies ist auch kein Automatismus. Um einen Vergleich mit der Soziologie zu bemühen: Es gibt dort zwar durchaus Statistiker*innen, die neue Modelle entwickeln. Es lassen sich aber nach wie vor relevante Forschungen durchführen, die mit seit Jahrzehnten bewährten Methoden wie der linearen Regression arbeiten, wenn sie über eine Fragestellung und geeignete Daten verfügen. Dieser Aspekt scheint mit in den Digital Humanities derzeit etwas zu kurz zu kommen: Ja, wir kennen LDA seit mittlerweile zwanzig Jahren, aber das bedeutet nicht, dass damit nicht nach wie vor erkenntnisreiche Forschung produziert werden kann. Und vielleicht sind dann Digital-Humanities-Konferenzen und -Zeitschriften nicht der richtige Ort dafür. Umso wichtiger erscheint es mir, in den inhaltlichen Bezugswissenschaften einen Fuß in der Tür zu behalten, um dort auch inhaltlich relevante Forschung zu publizieren, die auf der Grundlage computationeller Methoden entstanden ist. Die Gewichtungen und Schreibstile unterscheiden sich

dann aber entsprechend der Zielgruppe. Allzu technische Details hinsichtlich der genutzten Methode schrecken in einem geisteswissenschaftlichen Kontext eher ab und sind für eine grundsätzliche Nachvollziehbarkeit des gewählten Ansatzes vielleicht auch gar nicht erforderlich. Und auch wenn ein überzeugendes Narrativ hinsichtlich des Beitrags zur Forschung vielleicht mehr Gewicht bekommt, ist nicht mein Eindruck, dass quantitative Ansätze zur Evidenzerzeugung grundsätzlich abgelehnt werden. Historisch arbeitenden Kolleg*innen ist es ja nicht fremd, etwa in Aussagen zu Veränderungen über die Zeit, auch zu quantifizieren. Die Erkenntnisse aus computationellen Analysen müssen aber zum Fachdiskurs anschlussfähig gemacht werden – und wecken dann im Idealfall vielleicht sogar ein wenig die Neugier auf die Methode.

Das führt zu einem zweiten Aspekt, der zunehmend meine Praxis bestimmt: der Methodenvermittlung auf verschiedenen Ebenen. Denn in der Religionswissenschaft kommen computationelle Methoden nach wie vor nur am Rande vor. Die Gründe dafür sind dabei nicht nur in einem methodischen Konservatismus und einer – zum Teil auch legitimen – Skepsis gegenüber dem Mehrwert computergestützter Analysen zu suchen, sondern pragmatisch auch in einem Mangel an entsprechenden Lerngelegenheiten. Es zeigt sich hier ein typisches Henne-Ei-Problem: Da die Lehrenden in religionswissenschaftlichen Studiengängen in der Regel nicht mit computationellen Methoden vertraut sind, finden diese keinen Eingang in die Curricula, und folglich ist die nächste Generation ebenfalls nicht in der Lage, diese Methoden anzuwenden oder zu vermitteln. Und anders als bei sozialwissenschaftlichen oder philologischen Methoden sind entsprechende Fächerkombinationen, bei denen vertiefte Methodenkenntnisse im Zweitfach gewonnen werden, auch eher selten. Dabei gibt es durchaus ein Interesse an diesen Methoden. In neueren Auflagen einschlägiger Handbücher wie des *Handbook of Research Methods in the Study of Religion* werden nun Kapitel zu „Computational text analysis“ aufgenommen (Elwert 2021). Bei Studierenden, aber auch bei Forschenden verschiedener Karrierestufen gibt es hier einen Bedarf, mehr über die methodischen Möglichkeiten zu lernen.

Vielleicht noch einmal besonders ausgeprägt ist dies im Forschungsfeld der „digital religion“. Hier drängt sich die Erkenntnis geradezu auf, dass die verfügbaren Inhalte weit über das hinausgehen, was man mit etablierten, qualitativen Methoden erfassen könnte. Hier muss also entweder das Material stark eingegrenzt werden, um für eine qualitative Analyse handhabbar zu sein, oder es müssen Methoden hinzugezogen werden, die mit entsprechend großen Datenmengen umgehen können. Dies schließt sich dabei auch nicht aus: Häufig besteht zunächst einmal ein Interesse, Daten – etwa von Social-Media-Plattformen – systematisch zu speichern. Im zweiten Schritt können dann computationelle Analysemethoden herangezogen werden, um überhaupt einmal in der Fülle des Materials Teile zu identifizieren, die für die eigene Forschung von gesteigertem

Interesse sind. Und bei diesen ist dann eine vertiefte Analyse mittels qualitativer Methoden sowohl machbar als auch fruchtbar.

Neben der eigenen Anwendung und Vermittlung computationeller Methoden sehe ich aber noch einen dritten Bereich, in dem computationell Forschende gerade mit geistes- und sozialwissenschaftlichem Hintergrund gefragt sind. Wenn auch die Methodenentwicklung im engeren Sinne überwiegend in anderen Disziplinen stattfindet, so gibt es doch einen großen Bedarf, die etwa in der Informatik entwickelten Methoden selbst systematisch zu erforschen und auf dieser Grundlage ihre Eignung für geistes- und sozialwissenschaftliche Fragestellungen zu beurteilen. In Bezug auf Topic Modeling hat das etwa Melanie Althage (2022) vorgemacht. In meinem Arbeitsbereich beschäftigt sich etwa Lina Rodenhausen mit der Frage, mit welchen computationellen Methoden sich die metaphorische und wörtliche Verwendung eines Wortes unterscheiden lassen. Auch hier steht nicht die Entwicklung neuer Methoden im Vordergrund, aber schon für eine passgenaue Auswahl aus den verfügbaren Werkzeugen braucht es einiges an Experimenten und ein Verständnis dafür, wie diese Methoden auf bestimmte Textphänomene reagieren. Gerade bei neueren Modellen, etwa von LDA hin zu BERT und dann zu GPT, ergibt sich ein wesentlicher Teil der hinzugewonnenen Mächtigkeit aus dem Pre-Training. Ungeachtet der pragmatischen Nützlichkeit dieser Modelle sind hier inhärente Biases bei der Anwendung in der Forschung zu beachten, die gerade im Fall von Religion eine Rolle spielen (Abid/Farooqi/Zou 2021).

In diesem Dreieck aus eigener Forschung *mit* computergestützter Textanalyse, der Forschung *über* die Methoden selbst, und der Vermittlung der Methoden an Fachwissenschaftler*innen verschiedener Karrierestufen sehe ich ein fruchtbares Aufgabenfeld für computergestützte Forschung innerhalb der Religionswissenschaft. Und mit dieser Orientierung an den Bedarfen der Fachwissenschaft kann dann durchaus auch eine gewisse Entschleunigung gegenüber dem Innovationsdruck in der methodenorientierten Forschung einhergehen.

Literatur

Abid, Abubakar/Farooqi, Maheen/Zou, James (2021) Persistent Anti-Muslim Bias in Large Language Models. In: *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, AIES* '21. New York, NY, USA: Association for Computing Machinery, 298–306. DOI: https://doi.org/10.1145/3461702.3462624

Althage, Melanie (2022) Potenziale und Grenzen der Topic-Modellierung mit Latent Dirichlet Allocation für die Digital History. In: Döring, Karoline Dominika/Haas, Stefan/König, Mareike/Wettlaufer, Jörg (Hrsg.) *Digital History*.

Berlin/Boston: De Gruyter, 255–278. DOI: https://doi.org/10.1515/9783110757101-014

Berg, Anton/Valaskivi, Katja (2023) Representational silence and racial biases in commercial image recognition services in the context of religion. In: *Handbook of Critical Studies of Artificial Intelligence*. Cheltenham: Edward Elgar Publishing, 607–618. DOI: https://doi.org/10.4337/9781803928562.00062

Blei, David M./Ng, Andrew Y./Jordan, Michael I. (2003) Latent Dirichlet Allocation. In: *Journal of Machine Learning Research* 3, 993–1022. DOI: https://dl.acm.org/doi/10.5555/944919.944937

Bundesministerium für Bildung und Forschung (2011) Bekanntmachung des Bundesministeriums für Bildung und Forschung von Richtlinien zur Förderung von Forschungs- und Entwicklungsvorhaben aus dem Bereich der eHumanities. 10. Mai 2011. https://www.bmbf.de/foerderungen/bekanntmachung-643.html [zuletzt aufgerufen am 26.09.2024].

Chang, Jonathan/Boyd-Graber, Jordan/Blei, David M. (2009) Connections between the lines: augmenting social networks with text. In: *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 169–178. DOI: https://doi.org/10.1145/1557019.1557044

Devlin, Jacob/Chang, Ming-Wei/Lee, Kenton/Toutanova, Kristina (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Burstein, Jill/Doran, Christy/Solorio, Thamar (Hrsg.) *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* 1. (Long and Short Papers). Minneapolis. Minnesota: Association for Computational Linguistics, 4171–4186. DOI: https://doi.org/10.18653/v1/N19-1423

Elwert, Frederik (2015) TCFnetworks. Python. https://github.com/SeNeReKo/TCFnetworks

Elwert, Frederik (2021) Computational Text Analysis. In: Engler, Steven/Stausberg, Michael (Hrsg.) *Handbook of Research Methods in the Study of Religion*. 2. Auflage. Abingdon/New York: Routledge, 164–179.

Elwert, Frederik/Sellmer, Sven (2017) Modelling Structure and Content. Socio-Semantic Network Analysis of the Mahābhārata. In: *Leonardo* 50(5), 501. DOI: https://doi.org/10.1162/LEON_a_01277

Elwert, Frederik/Tabti, Samira/Pfahler, Lukas (2020) Me, myself and the other. Interreligious and intrareligious relations in neo-conservative online forums. In: *Religion* 50(3), 414–436. DOI: https://doi.org/10.1080/0048721X.2020.1754603

Hazard, Sonia (2013) The Material Turn in the Study of Religion. In: *Religion and Society: Advances in Research* 4, 58–78. DOI: https://doi.org/10.3167/arrs.2013.040104

Knibbe, Kim Esther (2020) Theorizing lived religion: introduction. In: *Journal of contemporary religion* 35(2), 157.

Krech, Volkhard (2020) Relational religion: manifesto for a synthesis in the study of religion. In: *Religion* 50(1), 97–105. DOI: https://doi.org/10.1080/00487 21X.2019.1686847

Nagel, Alexander-Kenneth (2008) Analysing change in international politics: a semiotic method of structural connotation. 70. *TranState working papers*. University of Bremen, Collaborative Research Center 597: Transformations of the State. http://hdl.handle.net/10419/24983 [zuletzt aufgerufen am 26.09.2024].

Paranyushkin, Dmitry (2011) Identifying the Pathways for Meaning Circulation using Text Network Analysis. Nodus Labs. http://noduslabs.com/research/pathways-meaning-circulation-text-network-analysis/ [zuletzt aufgerufen am 26.09.2024].

Pfahler, Lukas/Elwert, Frederik/Tabti, Samira/Morik, Katharina/Krech, Volkhard (2018) What do you do with 5 million posts? Versuche zum distant reading religiöser Online-Foren. In: Vogeler, Georg (Hrsg.) *Kritik der digitalen Vernunft. Abstracts zur Jahrestagung des Verbandes Digital Humanities im deutschsprachigen Raum*. Köln: Universität zu Köln, 335–338.

Rodenhausen, Lina (2024) „Mutter, Vater, Kind". Ressourcenarme automatische Metaphernverarbeitung für religionswissenschaftliche Fragestellungen. In: *Book of Abstracts. – DHd2024*. Passau. DOI: https://doi.org/10.5281/zenodo.10698475

Seiwert, Hubert (2014) Religionswissenschaft zwischen Sozialwissenschaften, Geschichtswissenschaft und Kognitionsforschung. In: Franke, Edith/Maske, Verena (Hrsg.) *Religionswissenschaft zwischen Sozialwissenschaften, Geschichtswissenschaft und Kognitionsforschung*. Ein Autoren-Workshop mit Hubert Seiwert. Marburg: Philipps-Universität Marburg, 15–31.