

More Data – More Interpretation?

Zur interpretativen Diskursanalytik des Quantitativen

Lilli Alexa Braunisch

Beitrag zur Veranstaltung der Ad-hoc-Gruppe »Digitale Diskurstransitionen und algorithmische Macht/Wissen-Regime. Analysen zur digitalen Umordnung der Diskurse«

Einleitung: Digitale Diskurse und die Krise der Interpretativität

Die fortschreitende Digitalisierung gesellschaftlicher Kommunikation stellt die interpretative Diskursforschung vor grundlegende Herausforderungen. Zum einen ist sie mit einer bislang nicht gekannten Quantität an Daten konfrontiert, zum anderen mit neuen Datenformen, Analyseverfahren und epistemischen Versprechen, die insbesondere unter dem Label von *Big Data* verhandelt werden. Diese betreffen nicht nur die empirischen Gegenstände sozialwissenschaftlicher Forschung, sondern greifen tief in ihre methodologischen und erkenntnistheoretischen Grundlagen ein. Die empirische Sozialforschung befindet sich in diesem Zusammenhang seit einigen Jahren in einem tiefgreifenden Transformationsprozess. Digitale Kommunikationsinfrastrukturen, algorithmische Ordnungen und neue Formen der Datenerzeugung verändern, wie soziale Wirklichkeit erfasst, strukturiert und analysiert wird. Besonders deutlich zeigt sich dies in der Diskursforschung, die zunehmend mit digitalen Massendaten, umfangreichen Textkorpora und computerbasierten Analyseverfahren arbeitet.

Vor diesem Hintergrund rückt das Verhältnis von Quantifizierung und Interpretation in den Mittelpunkt diskursanalytischer Debatten. Während computergestützte Verfahren neue Möglichkeiten der Erschließung und Systematisierung großer Datenmengen eröffnen, werfen sie zugleich die Frage auf, wie diskursive Bedeutungsstrukturen unter Bedingungen quantitativer und algorithmischer Analyse rekonstruiert werden können. Dies gilt in besonderem Maße für die Analyse von neuen Machtpositionierungen oder vom (digitalen) Wandel von Wissensregimen, weil hier die Diskurselemente und deren Beziehung zueinander neue Dimensionen annehmen (etwa in der Gleichzeitigkeit realer und digitaler Raum-Zeit-Überschneidungen).

Damit berührt die Digitalisierung der Diskursforschung einen zentralen Kern interpretativer Sozialforschung, und zwar die Frage, wie die komplexe Bedeutungsebene diskursiver Praktiken kontextuell in einer für andere systematisch nachvollziehbaren Weise erschlossen werden kann – und ob bzw. wie sich dieser Anspruch unter den Bedingungen quantitativer Datenverarbeitung aufrechterhalten lässt.

Diese Entwicklungen stellen die interpretative Diskursforschung vor eine zentrale Frage: Wie lassen sich digitale (Massen-)Daten diskursanalytisch untersuchen, ohne den Anspruch interpretativer Sozialforschung aufzugeben? Anders formuliert: Können quantitative und algorithmische Verfahren in dis-

kursanalytische Forschungslogiken integriert werden, ohne die Interpretativität zu verlieren? Und folglich ohne dass Diskurse auf algorithmisierte Muster, Frequenzen oder Korrelationen reduziert werden?

Der vorliegende Beitrag argumentiert, dass digitale Diskurse nicht mit rein quantifizierenden Verfahren angemessen erfasst werden können. Vielmehr bedarf es einer interpretativen Diskursanalytik des Quantitativen, die computerbasierte Verfahren systematisch in einen hermeneutisch-reflexiven Forschungsprozess integriert. Ausgangspunkt ist die These, dass mehr Daten nicht weniger, sondern – im Gegenteil – *mehr* Interpretation erfordern. Diese Frage ist nicht nur methodisch, sondern grundlegend theoretisch gelagert; denn die digitalen Umordnungen von Diskursen lassen sich nicht mehr allein mit den Begriffen der foucaultschen Diskursordnung des 19. oder 20. Jahrhunderts fassen. Algorithmische Selektionsmechanismen, datenbasierte Sichtbarkeitsregime und automatisierte Klassifikationen verändern diskursive Macht- und Wissensverhältnisse auf eine Weise, die neue analytische Zugänge erfordert.

Im Folgenden werde ich auf die digitale Umordnung der Diskurse eingehen, bevor einige Potenziale und Grenzen computerbasierter Analyseverfahren diskutiert werden.

Digitale Umordnung von Diskursen

Die Digitalisierung der Gesellschaft bringt nicht nur eine enorme Zunahme an Daten mit sich, sondern auch eine qualitative Veränderung diskursiver Praktiken. Diskurse entstehen, zirkulieren und sedimentieren sich zunehmend in digitalen Infrastrukturen, die selbst nicht neutral sind, sondern spezifische Logiken der Sichtbarkeit, der Relevanz und der Anschlussfähigkeit implementieren.

Dabei lassen sich gegenwärtig zwei miteinander verschränkte Entwicklungen von Digitalisierungs- und Algorithmisierungsprozessen beobachten: erstens eine Welle der Quantifizierung von Daten, in deren Zuge soziale Zusammenhänge und Handlungen in Daten übersetzt werden, wie etwa durch Klicks, Likes, Verweildauern, Geodaten oder Interaktionsnetzwerke. Diese Formate produzieren eine enorme Menge an digitalen Daten, die häufig als unmittelbare Abbilder sozialer Realität gelten, obwohl sie selbst das Ergebnis spezifischer Mess- und Kategorisierungsprozesse sind. Zweitens lässt sich eine Digitalisierung qualitativer Daten beobachten. Texte, Bilder oder audiovisuelle Inhalte werden durch Text-Mining-Verfahren, Topic Models oder Large Language Models analysiert, klassifiziert und verdichtet. Qualitative Diskursdaten werden damit selbst zum Gegenstand quantifizierender Verfahren.

Für die Diskursforschung ergibt sich daraus eine doppelte Herausforderung: Einerseits eröffnen digitale Daten neue Möglichkeiten – etwa zur Analyse langfristiger diskursiver Entwicklungen oder zur Untersuchung umfangreicher Macht-/Wissensregime. Andererseits droht eine Reduktion diskursiver Phänomene auf formale Muster, wenn ihre tieferliegende Bedeutungsdimension aus dem Blick gerät.

Computergestützte Analyseverfahren: Potenziale und Grenzen

Computergestützte Analyseverfahren sind in der empirischen Sozialforschung keineswegs neu (u. a. Baur et al. 2020; Baur und Blasius 2022). In der quantitativen Forschung gehören statistische Auswertungen seit Langem zum Standard, und auch qualitative Forschung nutzt seit den 1990er-Jahren Software zur Verwaltung und Systematisierung von Daten (Kuckartz 2010; Kuckartz et al. 2007; Friese 2014). Programme wie MAXQDA, Atlas.ti oder TrAVis (Jost et al. 2013) unterstützen Kodierungsverfahren sowie Visualisierung und Vergleiche von Daten, ohne jedoch selbst zu interpretieren.

Neuere KI-gestützte Verfahren gehen darüber hinaus. Text-Mining-Tools zum Beispiel versprechen, große Textmengen automatisch zu strukturieren, semantische Regelmäßigkeiten zu erkennen und thematische Cluster zu bilden und damit einen Teil der Analyse übernehmen zu können. Gerade für die

Diskursforschung erscheint dies attraktiv, da sich damit auch umfangreiche Korpora analysieren lassen, die händisch nicht mehr bearbeitbar sind. Eine Erweiterung und „Verbesserung“ dieser Bearbeitungsmöglichkeiten für die qualitative Sozialforschung – und aufgrund ihrer häufigen Datenkomplexität vor allem für die Diskursforschung – versprechen folglich zunehmend KI-gestützte Analyseverfahren von großen Text- oder Videokorpora, wie zum Beispiel Text-Mining-Verfahren oder Large Language Models (LLMs).

Als Reaktion auf die neuen digitalen Datenbestände plädieren vor diesem Hintergrund einige Wissenschaftler:innen zunehmend für die adäquate Verwendung von entsprechenden Analyseverfahren (Goldberg 2015; McFarland et al. 2015; Mohr und Bogdanov 2013; Stulpe und Lemke 2015; Wagner-Pacifici et al. 2015; Wiedemann 2013, 2016; Wiedemann und Lemke 2015; Evans und Aceves 2016). Die Behauptung dabei sei, dass diese Instrumente nicht nur unstrukturierte Daten auswerten könnten, sondern auch die Möglichkeit hätten, etwa diskursive Muster oder Regelmäßigkeiten der Semantik zu erfassen und zu beschreiben. Sie versprechen damit, auch von großen Datenmengen inhaltliche Merkmale und Unterschiede ebenso wie – im Fall von Texten – die semantischen Strukturen offenlegen zu können (die andernfalls im manuellen Vorgehen untergehen würden oder mit einem immensen Aufwand verbunden wären).

Einige der Aspekte, die solche Tools wie Text-Mining-Verfahren leisten, scheinen dabei einen Anschluss an das Forschungsverständnis der qualitativen Sozialforschung zu ermöglichen.¹ Gleichzeitig sind diese Verfahren mit methodologischen Spannungen verbunden. Viele Ansätze beruhen etwa implizit auf strukturalistischen Sprachmodellen, in denen Wörter oder Wortformen als kleinste bedeutungstragende Einheiten gelten. Bedeutung erscheint hier als eine statistisch messbare Eigenschaft von Texten, nicht als kontextuell hergestellte soziale Relation. Zudem gibt es eine Reihe von extremen Positionen einiger Vertreter:innen der Big Data Sciences, die der Meinung sind, die Vielheit an Daten lasse die Daten für sich sprechen und bedürfe daher keinerlei Theorie oder Methodologie zu ihrer Analyse. Meist wird hierbei ein positivistischer Datenbegriff unterstellt, indem davon ausgegangen wird, dass man digitale Daten (vor allem Online-Daten) als soziale Gegebenheiten – also ohne Rückbezug auf ihren Kontext – interpretieren könne. Im Fall von Forschungen zu Social-Media-Plattformen vernachlässigt dieser Positivismus der Computational Social Sciences beispielsweise die sinnstiftende und handlungsweisende Rolle der Medien bei der Hervorbringung sozialer Rollen und Handlungsformen im Umgang mit dem Medium.

Hierbei zeigt sich eine epistemologische Verschiebung, die auch schon Kenneth Cukier und Viktor Mayer-Schönberger formulierten:

„Using great volumes of information [...] requires three profound changes in how we approach data. The first is to collect and use a lot of data rather than settle for small amounts or samples, as statisticians have done for well over a century. The second is to shed our preference for highly curated and pristine data and instead accept messiness [...]. Third, in many instances, we will need to give up our quest to discover the cause of things, in return for accepting correlations.“ (Cukier und Mayer-Schönberger 2013, S. 30 f.)

Zum einen wird folglich die Suche nach möglichst repräsentativen, kleinen Datensätzen zum Schließen auf die Allgemeinheit in Big Data durch das Ideal ersetzt, „alle möglichen und so viele Daten wie möglich zu sammeln“ (Cukier und Mayer-Schönberger 2013); zum anderen wird eine fehlende Präzision einzelner Datensätze durch die wesentlich größere Datenmenge aufgewogen und anstelle einer Systemati-

¹ Autor:innen wie Wiedemann (2015) und Lemke (2015) etwa beanspruchen dabei mit der Nutzung computerbasierter Analyseverfahren und -werkzeuge (für eine große Menge digitaler Daten) eine Ergänzung der hermeneutischen Interpretation für die empirische Sozialforschung.

sierung von Daten (und Aussagen) ein Datenchaos in Kauf genommen. Vernachlässigt oder gänzlich außer Acht gelassen werden dabei die Grundlagen der interpretativen Analytik und damit einhergehend die Unterschiede zwischen inhaltsanalytischen und interpretativ-hermeneutischen Verfahrensweisen.

Interpretative Analytik und Wissenssoziologische Diskursanalyse

Interpretative Diskursforschung setzt dem ein grundsätzlich anderes Verständnis von Erkenntnis entgegen. In der Tradition Foucaults (2005a [1961], 1977, 1981 [1969], 2023 [1974]) und der Wissenssoziologischen Diskursanalyse (WDA) (Keller 2001, 2005a, 2005b, 2007, 2011) werden Diskurse nicht als bloße Textmengen verstanden, sondern als symbolische Ordnungen, die Wissen strukturieren, Handeln ermöglichen und zugleich begrenzen. Diskurse sind demnach weder stabil noch homogen, sondern historisch kontingente und konflikthafte Sinnzusammenhänge. Subjekte sind nicht bloße Träger von Diskursen, sondern beteiligen sich aktiv an deren Reproduktion und Transformation, während Diskurse zugleich Subjektivierungsprozesse strukturieren.

Die WDA verbindet diese hier nur im Ansatz skizzierten diskurstheoretischen Annahmen mit einer methodisch kontrollierten interpretativen Analytik (Keller 2005b). Interpretation wird dabei nicht als ein subjektiver Akt verstanden, sondern als rekonstruktiver Prozess, der intersubjektiv nachvollziehbar sein muss. Zentrale Bedeutung kommt folglich der Reflexion der eigenen Forschungsentscheidungen zu.

Bereits Foucault legt die Basis für ein interpretatives Vorgehen,² um die Ausdifferenziertheit der Aussagezusammenhänge von Diskursen zu erfassen. Im Foucaultschen Diskursdenken – und später weiter ausgearbeitet durch Reiner Keller – wird berücksichtigt, dass Subjekte bzw. die Handelnden in ihren „Äußerungen“ diskursive „Aussagen“ deutend reproduzieren und schließlich auch verändern (können). Demnach gelten Diskurse als „umstrittene, vorübergehend kristallisierte symbolische Strukturen der Ordnung von Welt, die Handeln zugleich ermöglichen und beschränken. [...]“ (Keller 2005b, S. 190). Foucaults Diskursverständnis hilft dabei, ein laut Schütz in der empirischen Sozialforschung seit Langem bestehendes Erklärungsdefizit zwischen sozialer Ordnung und Handeln zu füllen. Das – und das später im Rahmen der Wissenssoziologischen Diskursanalyse ausgearbeitete interpretative Paradigma – ermöglichen eine feine, ausdifferenzierte Bestimmung jener Übergänge vom subjektiven zum objektiven Sinn und umgekehrt: Diskurse „subjektivieren“ Handelnde (Bosančić 2019; Pfahl et al. 2018; Knoblauch 2017, S. 183 ff.) und das kommunikative Handeln „objektiviert“ sich über mehrere Schritte (Knoblauch 2017).

Dies ermöglicht eine Grundlage der Interpretation und des Deutens sozialen Handelns, wie es die Soziologie – auf Weber und Schütz zurückgehend – als „verstehend erklärende Wissenschaft“ ausmacht, die „soziales Handeln deutend [zu] verstehen und [dadurch ...] ursächlich erklären will“ (Weber 1980 [1921], S. 1; Knoblauch et al. 2018).

Die Rekonstruktion des komplexen Zusammenspiels vielfältiger Ereignisse für das Zustandekommen eines Diskurses sowie die Berücksichtigung auch nichtintendierter Wirkeffekte bedeuten, dass die Rekonstruktion und Deutung von diskursiven Macht- und Wissensregimen *nicht* durch das Aufbrechen einer einfachen Kausalkette von Ereignissen vollzogen werden kann. Anstelle singulärer Kausalmodelle bedarf es stattdessen der Analyse unterschiedlicher „Transformationstypen [...]“ (Foucault 1981 [1969], S. 19, 245). Foucault bezeichnet diesen Prozess auch als „kausale Demultiplikation“, die darin besteht, „das Ereignis den vielfältigen Prozessen entsprechend zu analysieren, die es konstituieren“ (Foucault 2005b, S. 29 f.).

² Auch wenn sich Foucault, wie er selbst bemerkt, nicht dazu äußert, wie er in seinen Analysen methodologisch und methodisch vorgeht.

In diesem Sinne formuliert Reiner Keller in der Wissenssoziologischen Diskursanalyse eine interpretative Analytik aus und schließt damit im Rahmen der WDA als Forschungsprogramm die Lücke, die Foucault in seinen Arbeiten offenlässt: eine systematisierte, nachvollziehbare Methodologie als elementarer Bestandteil der interpretativen Analytik von Diskursen. Durch die Zusammenführung des Foucaultschen Diskursansatzes mit dem sozialwissenschaftlichen interpretativen Paradigma bildet die WDA ein ergänzendes Lösungsangebot innerhalb der Hermeneutischen Wissenssoziologie (Hitzler et al. 1999). Sie entwickelt eine Methodologie des qualitativen Forschens, die zum einen die Position des Interpreten reflektiert, zum anderen Strategien der Dateninterpretation entwickelt – orientiert an Nachvollziehbarkeit und sozialer Objektivierung der Interpretationsschritte.

Für eine interpretative Diskursanalytik des Quantitativen

Vor diesem Hintergrund stellt sich gegenwärtig dringender denn je die Frage, wie quantitative und computerbasierte Verfahren in eine interpretative Diskursanalytik integriert werden können. Entscheidend ist, quantitative Analysen nicht als Ersatz, sondern als heuristische Zugänge zu begreifen. Sie können Hinweise auf diskursive Verdichtungen, Brüche oder Verschiebungen liefern, ersetzen jedoch nicht die interpretative Rekonstruktion von Bedeutung.

Methodologisch erfordert dies ein zirkuläres Forschungsdesign, in dem quantitative Exploration und qualitative Feinanalyse kontinuierlich aufeinander bezogen werden. Quantitative Ergebnisse dienen dabei nicht der Generalisierung im statistischen Sinne, sondern der Orientierung im Material und der gezielten Auswahl relevanter Textstellen. Interpretation beginnt dabei bereits bei der Korpusbildung, der Definition zentraler Begriffe und der Entscheidung für bestimmte technische Verfahren. Jede dieser Entscheidungen ist theoretisch eingebettet und damit selbst Teil des diskursanalytischen Prozesses.

Ein Beispiel: Der Innovationsdiskurs im Deutschen Bundestag

Diese Überlegungen lassen sich exemplarisch an einer wissenssoziologischen Diskursanalyse des Innovationsbegriffes im Deutschen Bundestag verdeutlichen.³ Auf Basis einer Vollerhebung aller Bundestagsdrucksachen aus dem Zeitraum von 1949 bis 2017 wurde darin der Innovationsbegriff in seiner diskursiven Genese und Dynamik untersucht. Das Forschungsinteresse lag auf der langzeitlichen Dynamik der Begriffsetablierung und seiner diskursiven Konstruktion, welche „Innovation“ zu einem zentralen Element politischer Planungs- und Entscheidungsprozesse im Deutschen Bundestag macht.

Aufgrund des langen Zeitraums mit entsprechenden Drucksachendokumenten umfasste der analysierte Gesamtkorpus 123.082 Textdokumente. Da der Korpus mit diesem Umfang nicht mehr händisch zu bearbeiten war, wurde ein computergestütztes Vorgehen gewählt, wobei es jedoch konsequent ein interpretativer Forschungsprozess blieb. Über Webscraping wurden die Dokumente erhoben und in einer relationalen Datenbank aggregiert. Anschließend wurde ein Subkorpus aller Dokumente extrahiert, die den Wortstamm *nnovat*⁴ enthielten (N = 6.245).

Der Analyseprozess erfolgte in drei iterativen Schritten:

- (1) Systematisierung der Metadaten zur zeitlichen und institutionellen Einordnung,
- (2) quantitative Erhebung semantisch-lexikalischer Merkmale,
- (3) interpretative Feinanalyse zentraler Textsequenzen.

³ Die hier skizzierte wissenssoziologische Diskursanalyse führte ich im Rahmen meiner Dissertation mit dem Titel „Innovation und Politik. Zur diskursiven Konstruktion von Innovation im Deutschen Bundestag“ durch (Braunisch 2026).

⁴ Die lemmatisierte Form *nnovat* wurde für die Wortzählung aller lexikalischen Vorkommnisse von „Innovation“ im Datenkorpus verwendet.

Die Datenerhebung und -analyse erfolgte grundlegend in Form von drei Periodisierungsschritten: erstens in Form der Verschlagwortung von Metadaten (Urheber, Datum, Texttyp, Drucksachennummer). Zweitens erfolgte eine Periodisierung semantisch-lexikalischer Merkmale der *nnovat*-Vorkommnisse. Zur quantitativen Erhebung gehörten etwa Wortzählungen von Innovations-Vorkommnissen nach Wahlperiode, Texttyp und Urheber sowie die Erfassung des Wortfeldes über die Zeit. Auf diese Weise ließ sich rekonstruieren, welche Instanzen (z. B. Bundesregierung, Ausschüsse) Innovationssemantiken zu welchem Zeitraum, in welcher Häufigkeit und in welchem Kontext thematisierten (oder nicht thematisierten). Hierbei wurden sieben Perioden der unterschiedlichen Verwendung von Innovationssemantiken identifiziert. Ausgehend von diesen sieben Perioden wurden, drittens, diskursive Fokusphasen identifiziert, die jeweils unterschiedliche Zeiträume des Diskurses kennzeichnen.

Besonders wichtig war hierbei die ständige Rückkopplung zwischen quantitativer Exploration und interpretativer Tiefenanalyse. Die quantitativen Auswertungen dienten zum einen dazu, richtungsweisende Tendenzen für die Auswahl relevanter Textstellen zur Feinanalyse zu identifizieren; zum anderen ermöglichte die interpretative Feinanalyse einen gezielten Rückbezug auf die quantitativen Befunde, indem Deutungsangebote für die Interpretation typisierbarer diskursiver Formationen herausgearbeitet wurden. So wurden die quantitativen Schritte der lexikalisch-semantischen Eigenschaften und die interpretative Feinanalyse am Text immer wieder wechselseitig aufeinander bezogen.

Auf diese Weise wurden drei diskursive Fokusphasen des Innovationsdiskurses rekonstruiert, die nicht allein auf Basis von Worthäufigkeitsanalysen erarbeitet werden konnten.

Während das Wort „Innovation“ erstmals im Jahr 1959 auftritt, beginnt sein diskursives Aufkommen um 1967 zur Zeit der Regierung Kurt G. Kiesingers. Der sich zu dieser Zeit – um 1967 – konstituierende Innovationsdiskurs *bezeichnet* nicht nur den Gegenstand der Innovation, vielmehr *konstituiert* er diesen in den parlamentarischen Prozessen des Deutschen Bundestages aktiv mit. Während Innovation dabei seit 1965 zunächst zu einem *inhaltlich relevanten* Konzept der politischen Agenda wird, wandelt sich der Gebrauch der Innovationssemantik seit 1972 zu einem zunehmend *programmatischen* und ab 2000 zu einem *reflexiven* Element politischer Entscheidungsprozesse des Deutschen Bundestages (siehe Tabelle 1).

Tabelle 1: Übersicht der Fokusphasen des Innovationsdiskurses 1949–2017

	Diskursive Fokusphase	Merkmal der Innovation
I. 1965–1969	Aufkommen von Innovation	Relevantwerden der Innovation
II. 1974–1983	Aufholphase	Programmatische Innovation
III. 2005–2009	Innovation globalen Wachstums	Reflexive Innovation

Der Blick über den Gesamtverlauf semantisch-lexikalischer Merkmale seit 1949 in Form quantitativer Erhebungsschritte zeigt dabei grundlegende Entwicklungen des Innovationsvorkommens auf. Die diskursiven Fokusphasen (Tabelle 1) zeigen darüber hinaus, wie diese Entwicklungen konkret vonstattengehen und wie die Dynamik im diskursiven Gefüge aussieht.

Zur befördernden Innovationsdynamik gehört dazu unter anderem, erstens, ein eng miteinander verwobenes Verhältnis zwischen der Europäischen Union und Deutschland als Mitgliedsstaat. Der Einfluss der EU auf den Innovationsdiskurs im Kontext einer wachsenden Wissensökonomie wird zu einem zentralen Motor für die Innovationsetablierung in Deutschland. Zweitens spielt der globale Wettbewerb zwischen den großen Industrienationen wie den USA, Japan, Russland und europäischen Ländern eine wichtige Rolle für die Herausbildung von Innovation. Drittens bedingen gesellschaftspolitische Herausforderungen der jeweiligen gesellschaftlichen Situation die Förderung und Forderung von Innovation,

wie beispielsweise der Klimadiskurs und daraus folgende Innovationsforderungen im Bereich erneuerbare Energien zeigen.

Als grundlegender Wandel im Diskurs lässt sich dabei insbesondere ein techno-industrielles Innovationsverständnis hin zu einem wissensbasierten Innovationsverständnis feststellen, welches Prinzipien einer wissensbasierten Ökonomie fortführt.

Die Periodisierungsschritte der Analyse des Diskurses folgten dabei nicht einem linearen Forschungsprozess; sie sind vielmehr ein Ergebnis des steten Rückbezuges zwischen den verschiedenen Periodisierungen, zwischen quantitativen Analyseschritten und Feinanalyse. Die Bezugnahme auf die Wissenssoziologische Diskursanalyse (Keller 2005) und das theoretische Fundament des Kommunikativen Konstruktivismus (Knoblauch 2017; Keller et al. 2013) ermöglicht hierbei den analytischen Blick von der Ebene der Drucksachen als Träger diskursiver Äußerungen bis auf die Ebene konkreter (institutioneller wie auch personeller) Akteure – wie die Bundesregierung oder Abgeordnete des Deutschen Bundestages – zu richten. So wurde im Rahmen der hier nur im Ansatz skizzierten Forschung nicht nur sichtbar, *dass* bestimmte Begrifflichkeiten derart wirkmächtigen Eingang in die Steuerungsmechanismen politischen Handelns finden; auch und vor allem lässt sich in einem so verknüpften Verfahren aufzeigen, *wie* diese Etablierungsprozesse vonstattengehen.

Das Beispiel verdeutlicht, dass ein stetes Hinterfragen und Vermitteln zwischen quantitativen (Zwischen-)Ergebnissen und interpretativen Deutungen für die Diskursforschung unumgänglich sind. Gerade weil große digitale Korpora zwar strukturelle Muster und Häufigkeiten etwa in der Verschiebung oder Verdichtung der Innovationssemantik sichtbar machen, deren diskursive Bedeutung und Kontextualisierung jedoch nicht aus den Zahlen selbst hervorgehen, bedarf es einer systematischen Verknüpfung quantitativer und interpretativer Verfahren.

Darüber hinaus stellen computerbasierte Verfahren keine neutralen Werkzeuge dar. Sie strukturieren den Forschungsprozess, setzen implizite Annahmen über Sprache und Bedeutung voraus und beeinflussen, welche Phänomene sichtbar werden. Dies zeigte zum Beispiel die Wortzählung in der vorab genannten Diskursanalyse des Innovationsbegriffes auf Basis von unterschiedlichen Wortdefinitionen: Da die Ergebnisse der Wortzählungen im Analyseprozess je nach Wortdefinition variierten, wurde die Bestimmung der Worteinheit, d. h. der Einheit dessen, was als Wort gelten soll, im Rahmen der SQLite-Datenbank von mir als Forscherin selbst vorgenommen. Da dies in vorprogrammierten Softwares weder transparent noch möglich ist, müssen die blinden Flecken über Forschungsimplicationen umso mehr transparent gemacht und reflektiert werden.

Eine interpretative Diskursanalytik des Quantitativen muss diese Implikationen explizit berücksichtigen und reflektieren. Besonders problematisch sind dabei sogenannte „Black-Box-Verfahren“, deren Funktionsweise für Forschende nicht transparent ist. Ohne Wissen über algorithmische Entscheidungen lassen sich Ergebnisse weder interpretieren noch kritisch einordnen. Dies hat auch Konsequenzen für die Ausbildung: Der kompetente Umgang mit digitalen Verfahren erfordert nicht nur technische, sondern auch theoretische und methodologische Kompetenzen.

Fazit und Ausblick

Digitale Massendaten stellen die interpretative Diskursforschung vor erhebliche Herausforderungen, eröffnen aber zugleich neue analytische Möglichkeiten und Potenziale. Entscheidend ist, dass diese Möglichkeiten nicht in einem datenpositivistischen Verständnis von Erkenntnis verspielt werden, sondern stets als Ergänzung zu interpretativen Verfahren eingesetzt werden.

Vor diesem Hintergrund ist – unter Berücksichtigung der Interpretativität sowie der theoretischen und methodologischen Grundlagen – eine Schnittmenge zwischen den Werkzeugen und Forschungs-

verständnissen zu suchen und anschlussfähig zu machen. Die interpretative Analytik bietet hierzu einen konstruktiven Ansatz.

Für den Umgang mit digitalen computerbasierten Verfahrensweisen der interpretativen Diskursforschung des Quantitativen erachte ich dabei folgende Aspekte als notwendig:

1. Computergestützte Verfahren müssen mit dem interpretativen Analyseprozess der Forschenden kombiniert werden, etwa im iterativen Zirkulieren der Erkenntnisgewinnung. Dies ermöglicht ein reflektiertes Zweifeln und Hinterfragen der im computerbasierten Prozess erworbenen Kenntnisse über Diskursfragmente, Machtverhältnisse oder Deutungsmuster (wie es die Hermeneutik grundlegend für den Interpretationsprozess im Sinne des Verstehens des Verstehens vorschlägt).
2. Empirisch qualitative Forschung erfordert eine Form der Interpretation, die Verstehen und Erklären nicht voneinander trennt. Dies gilt umso stärker, je mehr digitale Daten und Datenverfahren eingesetzt werden. Die Interpretation beginnt bei der Auswahl der Daten, betrifft die Datenerhebungsmethode und die -auswertung und endet bei der Ergebnisdarlegung.
3. Keine Maschine, weder KI-basiert noch anderweitig programmiert, kann digitale Diskursdaten von selbst analysieren oder interpretieren: Das Verstehen und deutende Erklären komplexer, vielfältiger Diskursereignisse zu unterschiedlichen Zeitpunkten kann keine KI leisten. Dies ist insbesondere nicht mit einer Black Box möglich, die einprogrammierte Implikationen zur Erhebung, Bearbeitung oder Auswertung enthält, ohne dass diese bekannt sind. Als Konsequenz erfordert dies die Vermittlung von Wissen darüber, was die Software genau tut, beispielsweise als Bestandteil eines Curriculums zur Vermittlung solcher Tools.

Grundlegend muss stets die Position der Forschenden reflektiert und die Nachvollziehbarkeit des Forschungsprozesses gewährleistet werden.

Mehr Daten bedeuten nicht automatisch mehr Erkenntnis. Sie erfordern vielmehr *mehr* Theorie, *mehr* Methodologie und *mehr* Interpretation. Eine interpretative Diskursanalytik des Quantitativen kann dazu beitragen, digitale Diskurse nicht zu messen, sondern diese in ihrer sozialen Bedeutungsdimension zu rekonstruieren – und damit auch neue Formen diskursiver Macht- und Wissensverhältnisse der diskursiven Konstruktion von Wirklichkeit sichtbar zu machen.

Literatur

- Baur, Nina, und Jörg Blasius. 2022. Methoden der empirischen Sozialforschung. In *Methoden der empirischen Sozialforschung*, Hrsg. N. Baur und J. Blasius, 1–32. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Baur, Nina, Graeff, Peter, Braunisch, Lilli, und M. Schweia. 2020. The Quality of Big Data: Development, Problems, and Possibilities of Use of Process-Generated Data in the Digital Age. *Historical Social Research*, 45 (2020) 3:209–243. <https://doi.org/10.12759/hsr.45.2020.3.209-243>
- Bosančić, Sasa. 2019. Die Forschungsperspektive der Interpretativen Subjektivierungsanalyse. In *Subjekt und Subjektivierung*, Hrsg. Alexander Geimer, Steffen Amling, und Sasa Bosančić, 43–64. Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-658-22313-7_3
- Cukier, Kenneth, und Victor Mayer-Schoenberger. 2013. The Rise of Big Data: How It's Changing the Way We Think About the World. *Foreign Affairs*, Vol. 92, No. 3:28–40.
- Evans, James A., und Pedro Aceves. 2016. Machine Translation: Mining Text for Social Theory. *Annual Review of Sociology*, 42:21–50.
- Foucault, Michel. 1977. *Überwachen und Strafen. Die Geburt des Gefängnisses*. Frankfurt am Main: Suhrkamp.
- Foucault, Michel. 1981 [1969]. *Archäologie des Wissens*. Frankfurt am Main: Suhrkamp.
- Foucault, Michel. 2005a [1961]. *Analytik der Macht*. Frankfurt am Main: Suhrkamp.

- Foucault, Michel. 2005b. *Dits et Écrits. Schriften* Bd. 4, Hrsg. Defert, Daniel, Francois Ewald, Hans-Dieter Gondek, und Hermann Kocyba. Frankfurt am Main: Suhrkamp.
- Foucault, Michel. 2023 [1974]. *Die Ordnung der Dinge. Eine Archäologie der Humanwissenschaften*. Frankfurt am Main: Suhrkamp.
- Friese, Susanne. 2014. *Qualitative data analysis with ATLAS.ti*, London.
- Goldberg, Amir. 2015. In defense of forensic social science. *Big Data & Society*, July-December, 1–3.
- Hitzler, Ronald, Jo Reichertz und Norbert Schröer, Hrsg. 1999. *Hermeneutische Wissenssoziologie*. Konstanz: UVK.
- Jost, Christofer, Daniel Klug, Axel Schmidt, Armin Reautschnig, und Klaus Neumann-Braun. 2013. *Computergestützte Analyse von audiovisuellen Medienprodukten*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Keller, Reiner. 2001. Wissenssoziologische Diskursanalyse. In *Handbuch Sozialwissenschaftliche Diskursforschung, Band 1: Theorien und Methoden*, Hrsg. R. Keller, A. Hirsland, Schneider und W. Viehöver, 113–145. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Keller, Reiner. 2005a. *Wissenssoziologische Diskursanalyse. Grundlegung eines Forschungsprogramms*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Keller, Reiner. 2005b. Wissenssoziologische Diskursanalyse als interpretative Analytik. In *Die diskursive Konstruktion von Wirklichkeit: zum Verhältnis von Wissenssoziologie und Diskursforschung*, Hrsg. Reiner Keller, Alexander Hirsland, Werner Schneider und Willy Viehöver, 49–75. Konstanz: UVK.
- Keller, Reiner. 2007. *Diskursforschung*, Wiesbaden: VS Verlag für Sozialwissenschaften.
- Keller, Reiner. 2011. *Wissenssoziologische Diskursanalyse. Grundlegung eines Forschungsprogramms*, 3. Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Knoblauch, Hubert. 2017. *Die kommunikative Konstruktion der Wirklichkeit*. Wiesbaden: Springer Fachmedien.
- Knoblauch, Hubert, Baur, Nina, Traue, Boris, und Leila Akremi. 2018. Was heißt „interpretativ forschen“? In *Handbuch Interpretativ forschen*, Hrsg. Leila Akremi, Nina Baur, Hubert Knoblauch, und Boris Traue, 9–35. Weinheim, München: Beltz Juventa.
- Kuckartz, Udo. 2010. *Einführung in die computergestützte Analyse qualitativer Daten*. 3. Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kuckartz, Udo, Grunenberg, Heiko, und Thorsten Dresing. 2007. *Qualitative Datenanalyse: computergestützt. Methodische Hintergründe und Beispiele aus der Forschungspraxis*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- McFarland, Daniel, Kewin Lewis, und Amir Goldberg. 2015. Sociology in the era of big data: The ascent of forensic social science. *American Sociologist*, DOI 10.1007/s12108-015-9291-8
- Mohr, John W., und Petko Bogdanov. 2013. Introduction – Topic models: What they are and why they matter. *Poetics* 41(6):545–569.
- Pfahl, Lisa, Schürmann, Lena, und Boris Traue. 2018. Subjektivierungsanalyse. In *Interpretativ forschen. Ein Handbuch für die Sozialwissenschaften*, Hrsg. Leila Akremi, Nina Baur, Hubert Knoblauch, und Boris Traue, 858–886. Weinheim: Beltz Juventa.
- Stulpe, Alexander, und Matthias Lemke. 2015. Blended Reading. In *Text Mining in den Sozialwissenschaften. Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse*, Hrsg. Matthias Lemke, und Gregor Wiedemann, 17–61. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Wagner-Pacifici, Robin, John W. Mohr, und Ronald L. Breiger. 2015. Ontologies, methodologies, and new uses of Big Data in the social and cultural sciences. *Big Data & Society*, December, 1–11.
- Wiedemann, Gregor. 2013. Opening up to Big Data: Computer-Assisted Analysis of Textual Data in Social Sciences. *Forum Qualitative Sozialforschung* 14, Art. 13.
- Wiedemann, Gregor. 2016. *Text Mining for Qualitative Analysis in the Social Sciences: A Study on Democratic Discourse in Germany*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Wiedemann, Gregor, und Matthias Lemke. 2015. Text Mining für die Analyse qualitativer Daten. In *Text Mining in den Sozialwissenschaften. Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse*, Hrsg. Matthias Lemke und Gregor Wiedemann, 397–419. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Weber, Max. 1921 [1980]. *Wirtschaft und Gesellschaft*. Tübingen: Mohr.