



Zeitschrift für Diskursforschung

Journal for Discourse Studies

Sonderausgabe zum zehnjährigen Jubiläum/Special Issue for the Tenth Anniversary

- **Teil A Grundlagen/**
Part A Fundamentals
- **Teil B Forschungsüberblicke und einzelne Ansätze/**
Part B Research Overviews and Specific Approaches
- **Teil C Exemplarische Anwendungen/**
Part C Exemplary Applications
- **Teil D Diagnosen – Diskussionen – Perspektiven/**
Part D Diagnoses – Discussions – Perspectives

Inhaltsverzeichnis

<i>Reiner Keller/Werner Schneider/Wolf Schünemann/Willy Viehöver</i> Editorial	124
---	-----

Teil A Grundlagen/Part A Fundamentals

<i>Christian Bär</i> Zwischen Diskurs und Ästhetik – Aspekte vor einem diskurslinguistischen Hintergrund	128
--	-----

<i>Dietrich Busse</i> Diskursanalyse als Wissensanalyse – aus der Perspektive einer linguistischen E pistemologie	137
---	-----

<i>Benno Herzog</i> Kritik in der Diskursforschung.....	148
--	-----

<i>Linda Maack/Inga Truschkat</i> Diskurs und Organisation – Theoretische Reflexionen eines rekursiven Verhältnisses	155
--	-----

<i>Boris Traue/Lena Schürmann/Lisa Pfahl</i> Beyond Negative Humanism. Subjectivation, the Lived Body and Human Rights	165
--	-----

<i>Ingo H. Warnke</i> Mythos Zuhören – Bemerkungen zur Diskursphänomenologie gerichteter Aufmerksamkeit	176
---	-----

Teil B Forschungsüberblicke und einzelne Ansätze/Part B Research Overviews and Specific Approaches

<i>Robert Aust/Arne Böker/Anne Mielke</i> Hochschule und Wissenschaft aus diskursanalytischer Perspektive: Eine Einladung....	184
--	-----

<i>Noah Bubenhofer</i> Corpus Linguistics in Discourse Analysis: No Bodies and no Practices?.....	195
--	-----

<i>Adele E. Clarke</i> Including Visual Discourse Materials in Situational Analysis Research.....	205
--	-----

Rainer Diaz-Bone

What difference does Foucault's discourse analysis make? Why discourse analysis needs to be based on the concepts of historical epistemology..... 217

Dinah K. Leschzyk

Digitalisierung – Daten – Deutungen. 15 Jahre Diskursforschung aus romanistisch-linguistischer Perspektive..... 227

Marcus Müller

Von methodischen Standards und kontroversen Diskursen: Zum Stand der digitalen Diskurslinguistik 235

Peter Stücheli-Herlach/Philipp Dreesen/Julia Krasselt

Öffentliche Diskurse modellieren und simulieren. Wege der transdisziplinären Diskurslinguistik 245

Ruth Wodak

(Kritische) Diskursforschung: Ansätze, Missverständnisse und neue Fragestellungen 257

Teil C Exemplarische Anwendungen/Part C Exemplary Applications

Maria Becker/Ekkehard Felder

Moralisierung zwischen den Zeilen: Auf den Spuren einer kommunikativen Praktik 266

Moritz Hillebrecht

Zwischen Sorgearbeit und Care. Spezialdiskursive Problematisierungen spätmoderner Sorgeordnungen..... 277

Anders Horsbøl

Discourse and the Environment: Complexity, Conflicts, and Crises 288

Heidrun Kämper

Sprachliche Umbrüche und Diskurs. Gedanken zu ihrer Analyse..... 296

Amira Malik

Sind Wissenschafts- und Meinungsfreiheit in Gefahr? Kämpfe um die Grenzen des Sagbaren als Gegenstand der Diskursforschung 307

Sabine Pfleger

Diskurs- und Identitätsforschung im Spannungsfeld von epistemologischer und methodologischer Lokalität und Translokalität 318

Dennis Puorideme

SKAD analysis of development intervention in contemporary African society:
reflecting on current and future challenges 326

Felix Schilk

Metapolitics as Programmatic Storytelling. New Right Discourses as a Challenge
for Discourse Research 339

Teil D Diagnosen – Diskussionen – Perspektiven/Part D Diagnoses – Discussions – Perspectives

Johannes Angermüller

Postfaktischer Diskurs und Wahrheit 349

Marlon Barbehön/Sybille Münch

The politics of narrative (research): A »success story« 358

Inka Bormann/Pavla Schäfer

Vertrauensdiskurse. Überlegungen zu Vertrauen als Interpretationskategorie
in der Diskursforschung 366

Simon Egbert

Das Potenzial der Diskursforschung für die Analyse digitalisierter Gesellschaften ... 376

Annika Harzmann

»Following Shadows« – zeittheoretische Überlegungen zur Zukunftskompetenz
der Diskursforschung 384

Sabine Heiss/Annette Knaut

(De-)Koloniale Diskursforschung – Kritik und Alternativen 393

Reiner Keller

Discourse and Violence 404

Angelika Pöferl

Diskurs – Alltag – Alltags(kosmo)politik. Anmerkungen zur Reflexivität
von Natur-, Selbst- und Sorgeverhältnissen 415

David Römer/Martin Wengeler

Back to the roots! Eine Verteidigungsrede der traditionellen themenbezogenen Dis-
kurslinguistik 426

Marcus Müller

Von methodischen Standards und kontroversen Diskursen: Zum Stand der digitalen Diskurslinguistik

Zusammenfassung: Der Beitrag erörtert den methodischen Erkenntnisstand der digitalen Diskurslinguistik. Dazu gibt er eine Skizze der Entwicklung digitaler Methoden in der linguistischen Diskursforschung, thematisiert Kernprobleme und stellt schließlich die DFG-Forschungsgruppe Kontroverse Diskurse vor, in der sich Expertinnen und Experten für unterschiedliche linguistische Ansätze der Diskursanalyse – von der auf Textlektüre basierenden Diskursforschung bis zu computerlinguistischen Verfahren – zusammengeschlossen haben, um eine integrative Methodologie auf Höhe des methodischen Forschungsstands zu entwickeln. Dabei wird Diskursgeschichte erstmals systematisch als Gruppenforschung entwickelt und betrieben.

Schlagwörter: Annotation, digitale Diskursforschung, Gruppenforschung, Korpuslinguistik, Methodologie

Abstract: The article discusses the methodological state of knowledge in digital discourse linguistics. For this purpose, it gives a sketch of the development of digital methods in linguistic discourse research, addresses core problems, and finally introduces the DFG research group Controversial Discourses, in which experts for different linguistic approaches to discourse analysis – from discourse research based on text reading to computational linguistic procedures – have joined forces to develop an integrative methodology at the height of the methodological state of research. In doing so, discourse history is for the first time systematically developed and conducted as group research.

Keywords: Annotation, digital discourse research, group research, corpus linguistics, methodology

1 Einleitung

In diesem Text möchte ich den Stand und die aktuellen methodischen Themen der digitalen linguistischen Diskursforschung darstellen, so wie sie sich mir zur Zeit der Abfassung dieses Beitrags darstellen. Ich tue das nicht im Sinne eines Überblicks, wie man ihn etwa in einem Handbuchartikel erwarten würde, sondern als positionsgewundene Bestandsaufnahme. Da es in der Diskursforschung – unabhängig von disziplinären, methodischen oder thematischen Spezifizierungen – immer um das Verhältnis von Zeichengebrauch und Kontexten (z. B. Situationen, Institutionen, Ideologien, sozialen Beziehungen) geht, muss man Diskurse interpretativ erschließen. Andererseits ist die Diskursanalyse im Gegensatz z. B. zur juristischen oder literaturwissenschaftlichen Hermeneutik nicht am einzelnen Text oder individuellen Intentionen interessiert, sondern am Systematischen,

Sozialen, Regelhaften des Zeichengebrauchs. Insofern liegt es nahe, sich mit mehreren – und eventuell möglichst vielen – Zeugnissen diskursiver Praktiken zu beschäftigen. So beginnen Foucaults Erörterungen einer Methodologie der Diskursanalyse in der *Archäologie des Wissens* (1972, S. 10 f.) mit Prinzipien und Verfahren der Korpusbildung, berühren Fragen der deskriptiven und schließenden Statistik und schließen interpretative Text- und Kontextanalyse mit ein.¹ Während sich Probleme der Korpus- und Kategorienbildung quer über alle disziplinären und methodischen Ausprägungen der Diskursforschung stellen, hatte sich bei der Frage nach dem Stellenwert des Messens und Zählens in der Diskursanalyse doch ein gewisses Schisma ergeben, im Sinne einer Fokussierung auf das Interpretieren einzelner Texte einerseits und auf das Messen partikularer Phänomene in möglichst vielen Texten andererseits. Klar ist, dass man sowohl kategoriale Bestimmungen braucht als auch Abschätzungen der Reichweite kategorial bestimmter Phänomene. Schließlich kann man nichts messen, was man nicht vorher segmental und kategorial bestimmt hat. Und ein Messergebnis, das nicht auf methodisch kontrollierte Weise kontextualisiert wird, ist (nicht nur) in der Diskursanalyse nicht hilfreich. In der DFG-Forschungsgruppe *Kontroverse Diskurse* arbeiten Diskursforscher:innen aus dem Feld der deutschsprachigen Linguistik daran, dieses Schisma nicht nur zu überwinden, sondern vor dem Hintergrund unterschiedlicher methodischer Schwerpunkte im Spannungsfeld zwischen Vermessen und Verstehen ein integratives Verfahrensmodell der Diskursgeschichte als kollaborative Gruppenforschung zu entwickeln, das auf größere digitale Datenbestände zurückgreift und aktuelle Verfahren der algorithmischen Texterschließung und Textklassifikation einbezieht, ohne triviale Ergebnisse zu produzieren und ohne auf das explikative Potenzial des spezialisierten Lesens von Texten und Kontexten zu verzichten. Darüber werde ich im letzten Teil dieses Beitrags berichten. Zuvor gebe ich eine fachhistorische Skizze der digitalen Diskurslinguistik.

- 1 »[...] the new history is confronted by a number of methodological problems, [...]. These include: the building-up of coherent and homogeneous corpora of documents (open or closed, exhausted or inexhaustible corpora), the establishment of a principle of choice (according to whether one wishes to treat the documentation exhaustively, or adopt a sampling method as in statistics, or try to determine in advance which are the most representative elements); the definition of the level of analysis and of the relevant elements (in the material studied, one may extract numerical indications; references – explicit or not – to events, institutions, practices; the words used, with their grammatical rules and the semantic fields that they indicate, or again the formal structure of the propositions and the types of connexion that unite them); the specification of a method of analysis (the quantitative treatment of data, the breaking-down of the material according to a number of assignable features whose correlations are then studied, interpretative decipherment, analysis of frequency and distribution); the delimitation of groups and sub-groups that articulate the material (regions, periods, unitary processes); the determination of relations that make it possible to characterize a group (these may be numerical or logical relations; functional, causal, or analogical relations; or it may be the relation of the ›signifier‹ (signifiant) to the ›signified‹ (signifié).« (Foucault 1972, S. 10 f.)

2 Digitale Sprachkorpora in der Diskursforschung

In der auf digitalen Sprachkorpora beruhenden Diskursforschung hat sich über die Jahrzehnte ein Standardverfahren etabliert, das Marchi/Taylor (2018, S. 6) folgendermaßen auf den Punkt bringen: »What we do when we analyse discourse using corpora ›is a qualitative analysis of quantifiable patterns««. Das Zitat stammt aus dem Band *Corpus Approaches to Discourse* (Taylor/Marchi 2018), der einen guten Überblick über die aktuelle Methodendiskussion der digitalen Diskursanalyse gibt. Er zeigt, dass viel über Deutung und Schlussverfahren im Kontext von digitalen Analysen nachgedacht wird. Gleichzeitig zeigt der Band aber eben auch, dass die Autorinnen und Autoren Deutung als Prozess-Schritt immer erst nach der Messung bestimmter Variablen ansetzen (Müller/Stegmeier 2022). Das handlungsleitende Motiv, Korpora und statistische Analyseverfahren in der Diskursanalyse einzusetzen, ist es, den subjektivistischen Anfechtungen einer rein auf Intuition und Erfahrung bauenden Hermeneutik zu begegnen. Bereits Michel Pêcheux in seinem bahnbrechenden Buch *Analyse automatique du discours* (1969) verwendet digitale Analysen sprachlicher Muster, um eine methodengeleitete semantische Analyse von Diskursen durchzuführen, die nicht von der positionsabhängigen Intuition des Analysierenden abhängt. Während es in Frankreich seit Pêcheux's Buch eine Kontinuität der digitalen Textanalyse gegeben hat, die sich immer auch in der Nähe zur poststrukturalistischen Wissens- und Machtanalyse gesehen hat (vgl. Scholz 2019), ist im englisch- und deutschsprachigen Raum eine explizite Verbindung von digitaler Textanalyse und Diskursanalyse Foucaultscher Prägung erst in den 1990er Jahren mit den Arbeiten Hardt-Mautners (1995) hergestellt worden. In den letzten Jahren haben digitale Analysemethoden (nicht nur) in der Diskursanalyse dann einen erheblichen Innovations-schub ausgelöst. Durch den Aufbau immer größerer digitaler Textkorpora und den Einsatz sprachstatistischer und algorithmischer Methoden (Baker 2020; Bubenhofer 2018) lassen sich Muster herausarbeiten und wird gleichzeitig die Theorie von Sprache als sozialer Praxis und die Kontextsensitivität von Sprachgebrauch durch Serialitätsanalyse empirisch unterfüttert (Müller 2015, 2017). Die Grundpraktiken der Analyse in diesen Arbeiten unterscheiden sich nicht von denen Pêcheux's, allerdings unterscheidet sich der Grad an Theoretisierung der Analyse: Während Vertreterinnen und Vertreter nicht-quantitativer Ansätze stark theoriegeleitet gearbeitet haben (z. B. Wengeler 2003; van Dijk 2009; Reisigl/Wodak 2016), kann man in vielen korpusgestützten Ansätzen eine bewusste Konzentration auf Methoden und einen Verzicht auf die Entfaltung theoretischer Konzepte erkennen (vgl. beispielhaft den oben erwähnten Band von Taylor/Marchi 2018). Die Reflexion dreht sich vor allem um die praktische Methodik der Korpuszusammenstellung, der Strukturierung und Vorverarbeitung sprachlicher Daten und der Weiterentwicklung und Evaluation statistischer Maße. Momentan wird die Anwendung heuristischer algorithmischer Verfahren wie Topic Modelling (Murakami et al. 2017), Word Embeddings (Bubenhofer/Calleri/Dreesen 2019) und künstlicher neuronaler Netze (Becker/Bender/Müller 2020) auf die Diskursanalyse diskutiert. Mit diesen Ansätzen verbindet sich das begründete Versprechen, durch distributionelle Analyse Sinnbeziehungen zwischen Wörtern in großen Textmengen herausarbeiten zu können, sie werfen aber auch Fragen auf, die z. B. die Zuverlässigkeit, methodische Transparenz, Reproduzierbarkeit und Interpretierbarkeit der Ergebnisse betreffen.

3 Verstehen vermessen – Vermessen verstehen: Annotation

Ein Beispiel für eine Studie, die algorithmische und interpretierende Verfahren miteinander vergleicht, ist der 2016 von Baker und Egbert herausgegebene Sammelband *Triangulating Methodological Approaches in Corpus-Linguistic Research*, in dem die zehn Beitragenden jeweils eine andere Methode anwenden, um die gleiche Forschungsfrage zu beantworten: Inwiefern unterscheidet sich der Sprachgebrauch in Online-Q+A-Forenantworten in vier verschiedenen Standardvarietäten des Englischen? Die Beiträge basieren auf der Analyse desselben Korpus aus Online-Foren, und die Autor:innen verwenden Methoden wie korpusbasierte Diskursanalyse, Wahrnehmungsforschung, Multidimensional Analysis, pragmatische Analyse und Schlüsselwortanalyse. Stegmeier et al. (2019) präsentieren einen mehrdimensionalen Ansatz, um Twitterdiskurse zu zwei politischen Themen von weltweiter Bedeutung zu untersuchen: Klimawandel und Netzneutralität. Dabei kombinieren sie Geolokationsanalysen, Netzwerkanalysen von Retweets und @-Mentions und Keyword-Analysen. Die Kombination dieser Methoden erlaubt es, den Grad der Transnationalisierung der Diskurse diskursstrukturell und inhaltlich miteinander zu vergleichen. Ancarno (2018) diskutiert einen auf digitalen Korpora beruhenden Methodenmix in interdisziplinären Diskursanalysen.

Diese Studien stellen verschiedene messende und verstehende Ansätze nebeneinander und arbeiten jeweils die methodischen Potenziale heraus. Ein Verfahren, in welchem das Messen und das Verstehen integriert werden, ist die Annotation – oder in den Sozialwissenschaften: Kodierung (Bender 2020; Bender/Müller 2020; Müller/Bartsch/Zinn 2021). Sie wird in vielen auf Korpuslinguistik basierenden Verfahren standardmäßig und automatisiert angewendet, wenn es um (mutmaßlich) unstrittige Kategorien wie Lemmata (Grundformen flektierter Wörter) und Wortarten geht, die im Forschungsdiskurs institutionalisiert und gleichsam naturalisiert wurden. Annotation ist aber auch in der qualitativen Diskursforschung ein Standardverfahren. Oft dienen Annotationsschemata und Code Books als interne Unterstützung zur Kategorienentwicklung in Projekten einzelner Forscherinnen und Forscher und werden in Publikationen auch nicht eigens thematisiert. Für eine methodisch integrative Diskursforschung ganz besonders interessant sind computerlinguistische Ansätze, die auf manueller Annotation komplexer sprachlicher Phänomene wie z. B. Argumenten (Habernal/Gurevych 2016) oder semantischen Satztypen (Becker et al. 2019) beruhen. Ziel ist jeweils die automatisierte Analyse der Phänomene auf großen Datenmengen. Auch wenn aus diesen Studien keine Tagger entstehen, die mit akzeptabler Zuverlässigkeit eingesetzt werden könnten, so lohnt es sich doch, sich mit diesen Ansätzen zu beschäftigen. Sie geben ein klares Prozessmodell der Annotation vor und helfen, über die einzelnen Schritte zu reflektieren, die anstehen, wenn qualitative Kategorisierung in einer Gruppe und mit dem Ziel einer Generalisierung des Kategorienschemas über die jeweils annotierten Daten hinweg durchgeführt werden soll: Tagsetmodellierung, unabhängige, mindestens doppelte Annotation eines Trainingsdatensatzes, Messung des Inter-Annotator-Agreements, Anwendung und Evaluation des automatischen Klassifikators auf einem Testdatensatz (Becker/Bender/Müller 2020, vgl. Müller/Stegmeier 2022).

4 Kollaboration

Während insbesondere die interpretativen Ansätze der Diskursforschung – nicht immer, aber doch meist – als Schreibtischprojekte Einzelner konzipiert waren, erfordern digitale Methoden in der Regel Teamarbeit – allein schon, weil das Sammeln und die Vorverbreitung größerer Datenmengen oft aufwendig ist und lange dauert. Spätestens wenn man verstehende und vermessende Ansätze integrieren will, ist Teamarbeit unerlässlich. Durch die Entstehung digitaler Plattformen haben sich Arbeitsformen etabliert, die auf der Kollaboration von Forscherinnen und Forschern auf allen Ebenen des Forschungsprozesses beruhen (Bender 2016). Solche kollaborativen Prozesse sind nicht auf digitale Umgebungen beschränkt, hängen jedoch in ihrer Organisationsform aufs Engste mit ihnen zusammen. So ermöglichen die genannten Plattformen (unter Maßgabe der jeweiligen spezifischen Schwerpunkte) Zusammenarbeit beim Edieren, Analysieren und auch beim Produzieren von Texten. Im Kontext der digitalen Diskursanalyse sind insbesondere kollaborative Annotationstools wie CATMA (Gius/Jacke 2015) oder INCEpTION (Eckart de Castilho et al. 2018) bedeutsam. Während CATMA für die digitale Literaturwissenschaft entwickelt wurde und besonders gut interpretative Verfahren auf Textebene unterstützt, ist INCEpTION für kollaborative Annotationen entwickelt worden und unterstützt neben der Tagsetentwicklung und Messung eines Inter-Annotator-Agreements auch den Einsatz von Machine-Learning-Algorithmen und die Integration von Wissensdatenbanken. INCEpTION ist in den letzten Jahren im Sinne eines »Human-in-the-loop«-Ansatzes weiterentwickelt worden (Klie et al. 2020). Damit ist gemeint, dass Benutzerinnen und Benutzer in verschiedenen Phasen von Annotationsprojekten durch automatische Verfahren der Klassifikation, Suche und Integration von Forschungsdaten unterstützt werden, ohne dabei die Kontrolle über deutungsrelevante Klassifikationsentscheidungen und die Validität entsprechender Messungen zu verlieren.

5 In Richtung einer kollaborativen Diskursgeschichte: Kontroverse Diskurse

Um also die Potenziale der verstehenden und der vermessenden Diskursforschung im Feld der deutschsprachigen Linguistik zusammenzuführen, hat sich auf Initiative von Martin Wengeler die DFG-Forschungsgruppe *Kontroverse Diskurse. Sprachgeschichte als Zeitgeschichte seit 1990* zusammengefunden (Wengeler 2022).² Es geht dabei darum, die Sprachgeschichte seit der deutschen Wiedervereinigung als eine Geschichte der kontroversen sprachlichen Bearbeitung öffentlicher Themen zu rekonstruieren. Das geschieht anhand von vier semantischen Grundfiguren, die als diskursrelevant identifiziert wurden: Partizipation & Egalität, Mensch & Technologie, Individuum & Gesellschaft, Freiheit & Sicherheit.

2 Martin Wengeler ist auch der Sprecher der Forschungsgruppe. Weiterhin gehören ihr an: Noah Bubenhofer (Zürich), Nina Janich (Darmstadt), Jörg Kilian (Kiel), Kristin Kuck (Magdeburg), Marcus Müller (Darmstadt), Juliane Schröter (Genf), Constanze Spieß (Marburg). Angaben zur Intention der Gruppe sowie zu den Teilprojekten finden sich unter www.kontroverse-diskurse.net

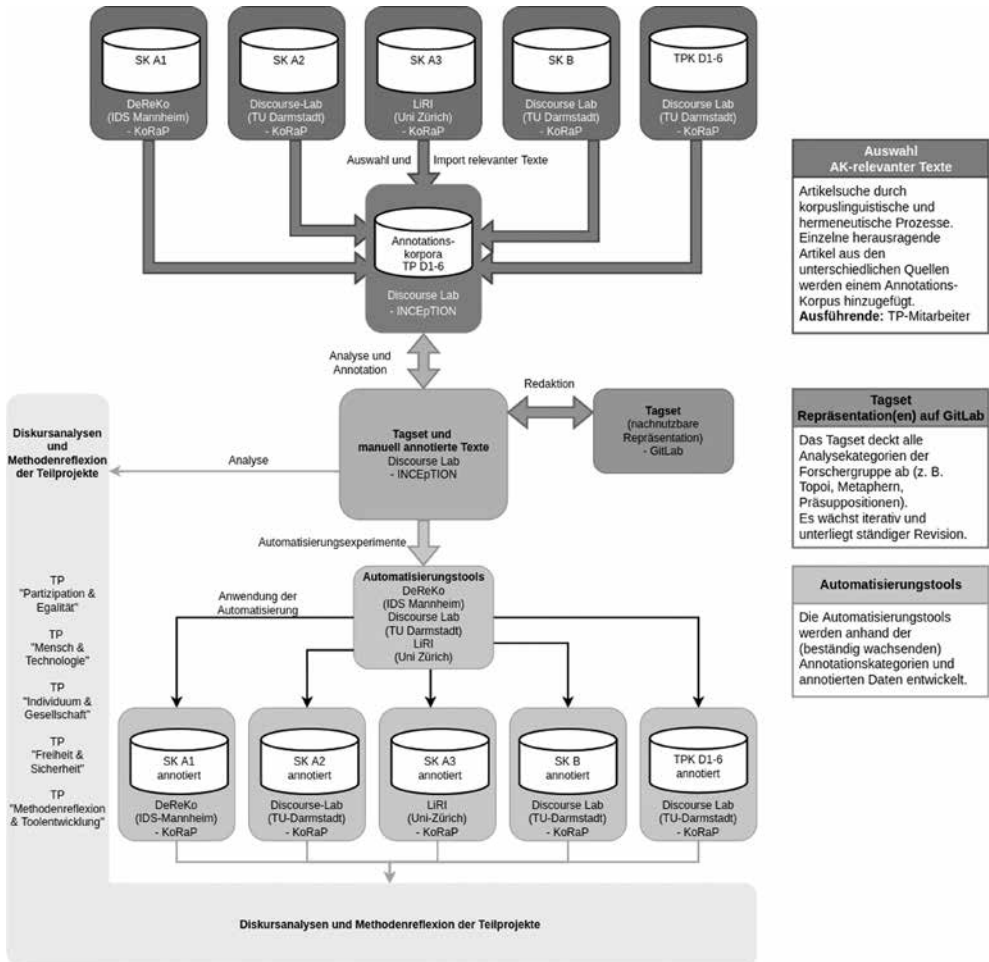
Diese sind jeweils einem Teilprojekt zugeordnet. Dabei wird Diskursgeschichtsschreibung zum ersten Mal in dieser Größenordnung als echte Gruppenforschung organisiert. Damit ist u. a. gemeint, dass nicht nur eine gemeinsame Korpusinfrastruktur entsteht und genutzt wird, sondern auch, dass die Gruppe gemeinsam ein Annotationsschema entwickelt und damit Analysen einzelner Teilprojekte für die gesamte Gruppe fruchtbar machen kann. Die Arbeit baut auf einem Prozessmodell auf, in dem algorithmische und interpretative Verfahren möglichst systematisch und transparent ineinandergreifen. Daher wird in einem methodologischen Teilprojekt die gemeinsame Annotationsarbeit der Gruppe und die Entwicklung eines gemeinsamen Annotationsschemas begleitet und reflektiert. Außerdem werden dort Automatisierungsexperimente zu den qualitativen Annotationen der inhaltlichen Teilprojekte durchgeführt mit dem Ziel, möglichst zuverlässige Tagger zu bauen, die ein dicht und hochwertig diskurssemantisch und pragmatisch annotiertes Gesamtkorpus ermöglichen. In Bezug auf den methodologischen Erkenntnisgewinn ist das Ziel der Forschungsgruppe die Evaluierung, Reflexion und inkrementelle Optimierung der kollaborativen Methoden. Es soll ein Workflow für digitale Analysen interpretativer Kategorien entwickelt werden, der den Qualitätsmaßstäben und dem Theorierahmen der Diskursanalyse genügt. Dabei stellen sich eine Reihe im Fach noch ungeklärter Verfahrensfragen, die einerseits die Qualitätsmaßstäbe pragmatischer Annotation betreffen, z. B. das Verhältnis von Segmentierung, Kategorisierung und Interpretationstiefe beim Annotieren, und andererseits den doppelten Operationalisierungs-Schritt von der Forschungsfrage zum Annotationsschema zur Mustererkennung.

Das gemeinsame Korpus der Forschungsgruppe besteht einerseits aus Artikeln deutschsprachiger nationaler Tages- und Wochenzeitungen und andererseits aus den bereits eingeführten Plenarprotokollen des deutschen Bundestags (vgl. Müller 2023). Es war der Forschungsgruppe wichtig, nicht einfach auf die Daten zurückzugreifen, die zur Verfügung stehen, sondern die tatsächlich projektrelevanten Daten auch in die Analyse einzubeziehen. Deshalb wurden Lizenzverträge mit der BILD-Zeitung und der FAZ abgeschlossen, die ansonsten aus nutzungsrechtlichen Gründen eher nicht beforscht werden. Neben BILD-Zeitung und FAZ bilden außerdem die Zeitungen, die am Institut für Deutsche Sprache in Mannheim (IDS) lizenziert sind (Der Spiegel, Die Zeit, SZ, taz), die Neue Zürcher Zeitung, die als maschinenlesbares Korpus an der Universität Zürich gehostet wird, und die von Discourse Lab aufbereiteten Plenarprotokolle des Deutschen Bundestags (Müller 2022a) die Datengrundlage, aus denen das Stammkorpus der Forschungsgruppe zusammengestellt wird. Außerdem baut jedes Teilprojekt spezifische Korpora auf. Die Korpora werden in der am IDS entwickelten Umgebung KorAP verwaltet (vgl. Müller 2023).

Die Forscher:innen in den unterschiedlichen Teilprojekten legen den untersuchten Teildiskursen und ihren jeweiligen Expertisen gemäß methodische Schwerpunkte bei der Diskursanalyse, z. B. Topos- (Wengeler 2003) oder Metaphernanalyse (Spieß 2012). Mit den daraus erwachsenden interpretativen Kategorien werden die Teilprojektkorpora annotiert und in ein Gesamt-Tagset integriert. Dazu müssen alle Revisionsschritte und Neufassungen nicht nur innerhalb eines der Teilprojekte am Teilprojekt-Tagset durchgeführt werden, sondern auch immer für das gesamte Projekt am Gesamt-Tagset.

Dies erfordert einen ständigen Austausch der Mitarbeiter:innen und eine klare Versionierungs-Strategie. Der Austausch wird über eine Gruppe unserer Forschungsplattform Discourse Lab (Müller 2022b) organisiert. Abb. 1 zeigt die Korpusinfrastruktur der Forschungsgruppe.

Abb. 1 Forschungsinfrastruktur der DFG-Forschungsgruppe Kontroverse Diskurse (aus Müller/ Stegmeier 2022)



Im Teilprojekt *Methodologie* werden auf der Basis der annotierten Daten Automatisierungsexperimente mit Recurrent Neural Networks (Becker/Bender/Müller 2020), aber auch einfacheren Recommender-Systemen (Bender 2023) und mit Hilfe linguistischer Modellierung und stringsbasierter Suche (Müller/Bartsch/Zinn 2021) durchgeführt. Insofern dabei reliable Tagger für die diskurssemantischen und pragmatischen Kategorien entste-

hen, wird das Gesamtkorpus mit diesen Annotationen angereichert. Diese werden über KorAP durchsuchbar gemacht, so dass die Verteilung und Kombinatorik der qualitativ erhobenen Kategorien im Gesamtkorpus gemessen werden kann. Damit soll erreicht werden, dass erstens die Analysen der Teilprojekte möglichst eng verschränkt werden können und zweitens die Methoden der qualitativen, verstehenden Diskursgeschichte auf große Datenbestände möglichst ohne Qualitätsverlust angewendet werden können.

5 Fazit und Ausblick

Dieser positionsgebundene Blick auf die methodischen Themen und Entwicklungen der digitalen Diskurslinguistik soll die Chancen und Probleme algorithmischer und statistischer Verfahren bei der Erforschung von Diskursen aufzeigen. Eines hat sich nicht geändert: Seit den 1960er Jahren dreht sich die Diskussion darum, wie man die möglichst genaue und umfassende Freilegung der Materialität der Diskurse, die letztlich naturwissenschaftliche Verfahren und gedankliche Prozeduren erfordern, mit der Analyse der diskursiven Sinnerzeugung und der Rekontextualisierung ihrer soziosemiotischen Ermöglichungsformationen vermitteln kann, welche wiederum ohne Interpretation und verstehenden Nachvollzug nicht zu haben ist. Während in den letzten zwei Jahrzehnten diese Vorhaben zwar nicht voneinander getrennt, aber doch mit recht deutlich verteilten Rollen und Schwerpunktsetzungen angegangen worden sind, unternimmt es die Forschungsgruppe *Kontroverse Diskurse*, diese beiden Kernaufgaben der Diskursanalyse in einem integrativen Ansatz der Gruppenforschung zusammenzuführen. Die Forschungsgruppe will in diesem Sinne ein Labor kollaborativer Methoden einer verstehenden Diskursgeschichte sein. Diesen Ansatz habe ich im letzten und ausführlichsten Teil meines Beitrags dargestellt. Damit verbunden ist das Ziel, auf der Basis der produzierten Forschungsdaten methodische Innovationen der verstehenden digitalen Diskursforschung voranzutreiben und zu evaluieren. Dieser Beitrag mag gezeigt haben, dass die technischen und methodologischen Voraussetzungen für dieses Vorhaben zum jetzigen Zeitpunkt gegeben sind. Bei allem methodologischen Optimismus allerdings wird man auch einräumen müssen, dass die epistemologische Lücke zwischen den vermessenen und den verstandenen Diskursen wohl umspielt, überbrückt und eingeht, wohl nie aber ganz geschlossen werden kann.

Literatur

- Ancarno, C. (2018): Interdisciplinary approaches in corpus linguistics and cads. In: Taylor, C./Marchi, A. (Hrsg.): *Corpus Approaches to Discourse: A Critical Review*. London und New York: Routledge, S. 130–156.
- Baker, P. (2020): *Corpus-assisted discourse analysis*. In Hart, C. (Hrsg.): *Researching Discourse: A student guide*. London und New York: Routledge, S. 124–142.
- Baker, P./Egbert, J. (Hrsg.) (2016): *Triangulating Methodological Approaches in Corpus-Linguistic Research*. London und New York: Routledge.

- Becker, M./Staniek, M./Nastase, V./Palmer, A./Frank, A. (2019): Classifying semantic clause types with recurrent neural networks: Analysis of attention, context and genre characteristics. In: *TAL Journal (Traitement Automatique des Langues/Natural Language Processing)* 59(2), 15–48.
- Becker, M./Bender, M./Müller, M. (2020): Classifying heuristic textual practices in academic discourse. A deep learning approach to pragmatics. In: *International Journal of Corpus Linguistics* 25(4), S. 426–460.
- Bender, M. (2016): *Forschungsumgebungen in den Digital Humanities: Nutzerbedarf, Wissenstransfer, Textualität*. Berlin und Boston: de Gruyter.
- Bender, M. (2020): Annotation als Methode der digitalen Diskurslinguistik. In: Gredel, E. (Hrsg.): *Diskurse digital. Theorien – Methoden – Fallstudien*. 2, S. 1–35.
- Bender, M. (2023): Pragmalinguistische Annotation und maschinelles Lernen. In: Bülow, L./Marx, K./Meier-Vieracker, S./Mroczynski, R. (Hrsg.): *Digitale Pragmatik*. Stuttgart: Metzler, S. 267–286.
- Bender, M./Müller, M. (2020): Heuristische Textpraktiken. Eine kollaborative Annotationsstudie zum akademischen Diskurs. In: *Zeitschrift für Germanistische Linguistik (ZGL)* 48(2), S. 1–46.
- Bubenhof, N. (2018): Diskurslinguistik und Korpora. In: Warnke, I.H. (Hrsg.): *Handbuch Diskurs, Sprachwissen*. Berlin und New York: de Gruyter, S. 208–241.
- Bubenhof, N./Calleri, S./Dreesen, P. (2019): Politisierung in rechtspopulistischen Medien: Wortschatzanalyse und Word Embeddings. In: *OBST. Osnabrücker Beiträge zur Sprachtheorie* Nr. 95, S. 211–242.
- Devlin, J./Chang, M.-W./Lee, K./Toutanova, K. (2018): BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805v2 [cs.CL].
- Eckart de Castilho, R./Klie, J.-C./Kumar, N./Boullosa, B./Gurevych, I. (2018): INCEpTION – Corpus-based Data Science from Scratch. In: *Digital Infrastructures for Research (DI4R)*, 9-11 October 2018. Online unter https://public.ukp.informatik.tu-darmstadt.de/UKP_Webpage/publications/2018/2018_DI4R_INCEpTION-abstract.pdf (Abruf 25.11.2022).
- Foucault, M. (1972): *The Archaeology of Knowledge and The Discourse on Language*. New York: Pantheon Books.
- Gius, E./Jacke, J. (2015): Informatik und Hermeneutik. Zum Mehrwert interdisziplinärer Textanalyse. In: *Sonderband der Zeitschrift für digitale Geisteswissenschaften* 1, 2015.
- Habernal, I./Gurevych, I. (2016). Which argument is more convincing? Analyzing and predicting convincingness of Web arguments using bidirectional LSTM. In: Erk, K./Smith, N. (Hrsg.): *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Berlin: Association for Computational Linguistics, S.1589–1599.
- Hardt-Mautner, G. (1995): ›Only Connect.‹ *Critical Discourse Analysis and Corpus Linguistics*. UCREL Technical Paper 6. University of Lancaster, <http://ucrel.lancs.ac.uk/papers/techpaper/vol6.pdf> (Abruf 21.07.2022).
- Klie, J.-C./Bugert, M./Boullosa, B./Castilho, R. E. de/Gurevych, I. (2018): The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation. In: Zhao, D. (Hrsg.): *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, S. 5–9.
- Marchi, A./Taylor, C. (2018): Introduction: partiality and reflexivity. In: Taylor, C./Marchi, A. (Hrsg.): *Corpus Approaches to Discourse: A Critical Review*. London und New York: Routledge, S. 1–15.
- Müller, M. (2015): Sprachliches Rollenverhalten: Korpuspragmatische Studien zu divergenten Kontextualisierungen in Mündlichkeit und Schriftlichkeit. Berlin und Boston: de Gruyter.
- Müller, M. (2017): *Digitale Diskursanalyse*. LitLab Pamphlet #5, https://www.digitalhumanities-cooperation.de/wp-content/uploads/2019/06/p05_mueller_de-1.pdf (Abruf 25.11.2022).
- Müller, M. (2022a): Die Plenarprotokolle des Deutschen Bundestags auf Discourse Lab. In: *Korpora Deutsch als Fremdsprache (KorDaF)* 2 (1), 123–127.
- Müller, M. (2022b): Discourse Lab – eine Forschungsplattform für die digitale Diskursanalyse. In: Bender, M./Jacob, K. (Hrsg.): *Digitales Forschen. Daten – Werkzeuge – Methoden*. Mitteilungen des Deutschen Germanistenverbands 69(2), S. 152–159.

- Müller, M. (2023): Korpora für die Diskursanalyse. Ressourcen und Lösungen im Discourse Lab. In: Deppermann, A./Fandrych, C./Kupietz, M./Schmidt, T. (Hrsg.): Korpora in der germanistischen Sprachwissenschaft. Mündlich, schriftlich, multimedial. Berlin und Boston: de Gruyter, S. 161–180.
- Müller, M./Bartsch, S./Zinn, J. O. (2021): Communicating the unknown. An interdisciplinary annotation study of uncertainty in the coronavirus pandemic. In: *International Journal of Corpus Linguistics*. 26(4), S. 498–531.
- Müller, M./Stegmeier, J. (2022): Zur Methodologie der kollaborativen Diskursgeschichte. In: *Aptum, Zeitschrift für Sprachkritik und Sprachkultur*, S. 280–296.
- Murakami A./Thompson P./Hunston S./Vajn, D. (2017): ›What is this corpus about?‹ Using topic modelling to explore a specialised corpus. In: *Corpora* 12(2), S. 243–277.
- Pêcheux, M. (1969): *Analyse automatique du discours*. Paris: Dunod.
- Reisigl, M./R. Wodak (2016): *The Discourse-Historical Approach*. In: Wodak, R./Meyer, M. (Hrsg.): *Methods of Critical Discourse Studies*. Los Angeles: SAGE, S. 23–61.
- Scholz, R. (2019): Lexicometry: A Quantifying heuristic for social scientist in discourse studies. In: Scholz, R. (Hrsg.): *Quantifying approaches to discourse for social scientists*. Basingstoke: Palgrave, S. 123–153.
- Spieß, C. (2012): Metaphern als sprachliche Strategien – Zur sprachlichen Manifestation von Konflikthaftigkeit im Stammzelldiskurs. In: Spieß, C. (Hrsg.): *Sprachstrategien und Kommunikationsbarrieren. Zur Rolle und Funktion von Sprache in bioethischen Diskursen*. Bremen: Hempen S. 175–198.
- Stegmeier, J./Schünemann, W.J./Müller, M./Becker, M./Steiger, S./Stier, S. (2019): Multi-method discourse analysis of twitter communication. A comparison of two global political issues. In: Scholz, R. (Hrsg.): *Quantifying Approaches to Discourse for Social Scientists*. Basingstoke: Palgrave, S. 285–314.
- Taylor C./Marchi A. (Hrsg.) (2018): *Corpus Approaches to Discourse: A Critical Review*, London und New York: Routledge.
- van Dijk, T. A. (2009): *Society and discourse. How social contexts influence text and talk*. Cambridge: Cambridge University Press.
- Wengeler, M. (2003): *Topos und Diskurs. Begründung einer argumentationsanalytischen Methode und ihre Anwendung auf den Migrationsdiskurs (1960-1985)*. Tübingen: Niemeyer.
- Wengeler, M. (2022): Kontroverse Diskurse. Sprachgeschichte als Zeitgeschichte seit 1990. Ein Forschungsprojekt. In: *Aptum, Zeitschrift für Sprachkritik und Sprachkultur*, S. 201–206.

Anschrift:

Prof. Dr. Marcus Müller
 Germanistik – Digitale Linguistik
 Institut für Sprach- und Literaturwissenschaft
 Technische Universität Darmstadt
 Residenzschloss 1
 64283 Darmstadt
 marcus.mueller@tu-darmstadt.de