TECHNISCHE UNIVERSITÄT DARMSTADT
DIGITAL HUMANITIES COOPERATION
UNIVERSITÄT KONSTANZ

Marcus Müller

Digitale Diskursanalyse

Pamphlet #5

Dezember 2017

Herausgegeben von Thomas Weitin

TECHNISCHE UNIVERSITÄT DARMSTADT
DIGITAL HUMANITIES COOPERATION
UNIVERSITÄT KONSTANZ

Marcus Müller

Digitale Diskursanalyse

Abstract

In der Diskursanalyse geht es um das Verhältnis von Sprache, Wissen und Gesellschaft. Die Rolle der Digitalen Linguistik in diesem Feld liegt darin, in Sprachkorpora, die mit Daten über Sprachverwendung und Sprecher*innen angereichert sind, Muster zu messen, kategorisieren und kontextualisieren, die als Spuren sozialer Interaktion gedeutet werden. Dabei kann man zwei Zugänge unterscheiden: Das forensische Programm begreift Sprache als Hinweis auf Phänomene, welche ideen-, kultur- und sozialgeschichtlich relevant sind; das explorative Programm nimmt die Prinzipien der Formierung von Sprache in sozialen Kontexten in den Blick. Das Pamphlet entwickelt die methodologischen Grundlagen der Digitalen Diskursanalyse entlang der Begriffe ‚Index', ‚Spur' und ‚Kontextualisierung'. Am Beispiel der Modalpartikelkonstruktion des Typs *doch eben, doch eigentlich* im Kontext der Bioethikdebatte stellt es die Forschungspraxis vor und plädiert schließlich für solche Forschungsmethoden, mit denen der textuelle Zusammenhang quantitativer Ergebnisse nicht aus dem Blick gerät.

Discourse Analysis deals with the interrelationship between language, knowledge, and society. The role of Digital Linguistics within this field lies in measuring, categorising, and contextualising patterns in language corpora enriched with data on language use and speakers. In Digital Discourse Analysis, we interpret those patterns as traces of social interaction. We distinguish two linguistic approaches to discourse: The forensic program conceives language as indication of phenomena that are relevant to conceptual, cultural, or social history. The explorative program examines the principles of language formation in social contexts. The pamphlet draws up the methodological foundations of Digital Discourse Analysis based on the terms 'index', 'trace', and 'contextualisation'. Taking the example of the modal particle construction of the type *doch eben, doch eigentlich* in the bioethics de-

bate, the research practice is presented. Finally, the pamphlet advocates research methods that help not to lose sight of the textual context of quantitative results.

Bibliografische Information der Deutschen Nationalbibliothek: Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie. Sie ist in der Zeitschriftendatenbank (ZDB) und im internationalen ISSN-Portal erfasst. Detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar. Alle Rechte, auch die des auszugsweisen Nachdrucks, der fotomechanischen Wiedergabe und der Übersetzung, vorbehalten. Dies betrifft auch die Vervielfältigung und Übertragung einzelner Textabschnitte, Zeichnungen oder Bilder durch alle Verfahren wie Speicherung und Übertragung auf Papier, Transparente, Filme, Bänder, Platten und andere Medien, soweit es nicht §§ 53 und 54 UrhG ausdrücklich gestatten.

©2017 Marcus Müller, mueller@linglit.tu-darmstadt.de

ISSN: 2628-4537

Hrsg. von Thomas Weitin

Marcus Müller

Digitale Diskursanalyse

I. Was ist Digitale Diskursanalyse?

Beginnen wir mit einer Beobachtung: Wenn Journalist*inn*en öffentliche Diskussionen moderieren, dann unterlaufen ihnen häufig Formulierungen, in denen zwei Modalpartikeln hintereinander verwendet werden, wie in den folgenden Beispielen:

frau späht vielen dank das ist natürlich ein wichtiger appell letztlich geht es ***natürlich auch*** *um risiken die bei der mutter entstehen*

ja uns geht es ***doch eigentlich*** *sehr gut*

gibt es ***denn auch*** *eine POSitive seite an diesem neuen gesetz für sie*

An diese Beobachtung kann man nun mehrere Fragen anschließen: Man kann zum Beispiel fragen, in welchen syntagmatischen Kontexten solche Doppelkonstruktionen vorkommen. Das wäre eine syntaktische Frage. Weiterführend interessant wäre, in welchen Kontexten die Konstruktion belegt ist und ob sie bei Journalist*inn*en und in Diskussionen häufiger vorkommt als in anderen Zusammenhängen. Das wäre eine variationslinguistische Frage. Wenn man nun noch weiterfragt, *warum* Journalist*inn*en in Diskussionen so sprechen und wie solche Formulierungen mit den situativen, kognitiven und gesellschaftlichen Bedingungen des journalistischen Sprachgebrauchs zusammenhängen, betreibt man linguistische Diskursanalyse.

In der Diskursanalyse generell wird das Verhältnis von Sprache, Wissen und Gesellschaft erforscht. Sie ist ein florierendes Forschungsfeld in den Geistes- und Sozialwissenschaften und hat in den verschiedenen Fächern viele unterschiedliche Ausprägungen erfahren. Digitale Diskursanalyse kann man dabei solche Ausprägungen nennen, die sich nicht nur auf digitale oder digitalisierte Forschungsdaten stützen, sondern die sich auch in ihren Analyseverfahren auf die genuine Beschaffenheit digitaler Daten stützen.[1] Das kann etwa bedeuten, dass Daten ausgezählt werden und ihre Häufigkeit ins Verhältnis zu

[1] Unter einem *Datum* verstehe ich ein als Zeichen gedeutetes Phänomen, das im Zuge eines Forschungsprozesses aus einer gegebenen Komplexion extrahiert und rekontextualisiert wird. Vgl. zum Datenbegriff und seiner Differenzierung Michael Bender: Forschungsumgebungen in den Digital Humanities: Nutzerbedarf, Wissenstransfer, Textualität. Reihe: Sprache und Wissen (SuW) 22. Berlin, Boston: de Gruyter 2016, S. 27–37.

anderen Daten gesetzt werden. Typischerweise bedeutet es, dass Daten auf der Grundlage von Kategorisierungsmodellen vorklassifiziert werden, zum Beispiel indem man Buchstabenketten der Kategorie ‚Wort' zuordnet (Tokenisierung) oder Wörter der Kategorie ‚Wortart' (Part-of-speech-Tagging). Das Auszählen von Daten ist natürlich keine Diskursanalyse, sondern nur ein Zwischenschritt. Entsprechend müssen wir zu Beginn klären, welche Systematik in der Verbindung zwischen Sprachdatum und Diskurs angenommen werden kann und welche methodischen Möglichkeiten und Restriktionen sich dabei ergeben.

II. Datum und Diskurs: Sprache als Spur sozialer Interaktion

Die Textbelege, die ich oben gegeben habe, sind bereits *linguistisch gedeutete* Daten, weil ihre Darbringungsform nahelegt, dass es sich um in Wörter segmentierte kommunikative Äußerungen handelt. Allerdings scheint es so zu sein, dass die Deutung dieser Sprachdaten noch nicht allzu weit getrieben wurde, z. B. wurde in den Belegen auf Groß- und Kleinschreibung gemäß der deutschen Rechtschreibung und auf Satzzeichen verzichtet. Dagegen wurde im letzten Beleg eine ganze Silbe großgeschrieben. Das weist die Belege für linguistisch geschulte Leser*innen als Transkripte mündlich produzierter Sprache aus, die Großschreibung indiziert eine Betonung, die über die erwartbare Sprachmelodie hinausgeht. In der Abweichung von der Standardgraphie liegt also eine weitere Deutungsschicht der zugrunde gelegten Sprachdaten. Im Kontext der Belege habe ich einige (wenn auch nicht besonders viele) Informationen zu den Belegen gegeben. Die Sprachdaten erscheinen so skizzenhaft als ein Aspekt der gesellschaftlichen Ereignisse, in deren Rahmen sie produziert wurden.

Diskursanalyse hat Sprache *als gesellschaftliches Phänomen* zum Gegenstand. Damit sprechen wir einen überaus weiten Phänomenbereich an, der z. B. die Variation postnominaler Genitivattribute genauso umfasst wie Machteffekte von Sprache, Praktiken des Schmeichelns oder der Verleumdung, hate speech und die sprachliche Perspektivierung von Wissen. Man könnte auch sagen: Möglicher Gegenstand der Diskursanalyse ist alles, was Teil der sozialen Wirklichkeit und damit *erfahrbar* ist und in Verbindung mit sprachlichen Mustern gebracht werden kann. Die Erfahrung sprachlich-sozialer Realität allein ist aber noch keine Diskursanalyse, selbst wenn sie sich methodisch kontrolliert vollzieht und sich den Transparenzbedingungen wissenschaftlicher Forschung stellt. Diskursanalytiker*innen wollen Sprachgebrauch nicht *verstehen*, in dem Sinne, dass sie an den kommunikativen Intentionen von Sprecher*inne*n und Schreiber*inne*n interessiert wären. Sie wollen Sprachgebrauch im Kontext vielmehr *erklären*. Was die Journalist*inn*en tatsächlich *gemeint* haben, als sie die Äußerungen mit den dop-

pelten Modalpartikeln getan haben, ist diskursanalytisch gesehen bedeutungslos. Interessant ist es dagegen zu fragen, welche Bedingungen gegeben sind, wenn diese Konstruktionen auftreten.

Das kann zu zwei Richtungen im Forschungsprozess führen: Erstens kann man untersuchen, was (wahrscheinlich - plausiblerweise - typischerweise - tendenziell) in der sozialen Wirklichkeit der Fall ist, wenn Sprache in einer bestimmten Weise von einer bestimmten Person in einer bestimmten Situation gebraucht wird. Dieses Vorgehen kann man mit dem eines Detektivs vergleichen, der die Spuren am Tatort sichert, in einen Zusammenhang bringt und kontextualisiert, um einen Mord aufzuklären. Wir können es das *forensische* Programm der Diskursanalyse nennen. In unserem Beispiel nähmen wir den Befund mit den Modalpartikeln lediglich zum Anlass, um die epistemischen, interaktionalen oder sozialen Situationen in denen sich Journalist*inn*en im Moment der Äußerung befinden, zu erforschen. Zweitens ist es ein Programm insbesondere der linguistischen Diskursanalyse zu erklären, welche *Formen* Sprache in gegebenen sozialen Situationen typischerweise annimmt, wenn Sprache in einer bestimmten Weise von einer bestimmten Person in einer bestimmten Situation gebraucht wird. Diese auf die sprachlichen Ausdrucksmittel gerichtete Form können wir die *explorative* Richtung der Diskursanalyse nennen.[2] Hier ginge es also darum, ein Inventar der Doppelkonstruktionen, ihr Kontextualisierungspotenzial und ihre syntagmatischen Einbettungsmuster relativ zu den typischen Situationen ihres Gebrauchs zu bestimmen. Die oben eingeführte Forschungsskizze werde ich unten als Beispiel für das explorative Vorgehen weiterführen.[3]

Die linguistische Sicht der Diskursanalyse bringt es in jedem Fall mit sich, dass wir einerseits von ‚Sprache' in dem oben angedeuteten ganz weiten Verständnis ausgehen müssen. Andererseits kann eine methodisch kontrollierte Sprachanalyse damit wenig anfangen. Wir brauchen also zwei Begriffe von Sprache, einen möglichst engen und einen möglichst weiten, die systematisch aufeinander bezogen werden. Der enge Begriff von Sprache S_1 zielt auf den Ausgangsbereich der Analyse; er beschreibt die Menge an im Diskurs konfigurierten materialen Sprachdaten (also Schriftzeichen und Sprachlaute, ggf. multimodale Zeichen wie Bilder, Mimik, Gestik, Proxemik) sowie die Muster ihrer Konfiguration. Der weite Begriff von ‚Sprache' S_2 umreißt den Zielbereich der Analyse; er umfasst neben den aus

[2] Diese Unterscheidung wird in Bezug auf Grammatik eingeführt in Marcus Müller *Diskursgrammatik*. In: Ingo H. Warnke (Hg.): Handbuch Diskurs (Handbücher Sprachwissen) Berlin, Boston: De Gruyter (im Druck).

[3] Ein Beispiel für die forensische Richtung findet sich in Marcus Müller & Jörn Stegmeier: *Investigating risk, uncertainty and normativity within the framework of Digital Discourse Analysis. The example of future technologies in climate change discourse.* In: Anna Olofsson & Jens O. Zinn (eds.): Researching risk and uncertainty – methodologies, methods and research strategies. Basingstroke: Palgrave (im Druck). Darin untersuchen wir mit korpuslinguistischen Methoden, wie Risiken und Unsicherheiten im deutschen und englischen Mediendiskurs thematisiert werden. Man lernt dabei nichts Neues über das Deutsche oder das Englische, wohl aber über die diskursive Formierung von Risikokonzepten in beiden Ländern. Die linguistische Methode bringt also einen eher soziologisch interessanten Befund hervor.

der Systemlinguistik bekannten Erscheinungen zusätzlich alle Aspekte sozialen Miteinanders, in die sprachliche Zeichen involviert sind (z. B. auch Interaktionsmuster, Sozialstrukturen, Machtgefüge) sowie alle Aspekte möglicher Gedanken, Einstellungen und Gefühle über die Welt, die in sozialen Gruppen geteilt werden können.[4] Für Diskursanalytiker*innen ist S_2 interessant, S_1 aber gegeben. Die Aufgabe besteht nun darin, mittels offen gelegter und reproduzierbarer Verfahren Tendenzen und Konjunkturen der Ausdruckskonfiguration in gegebenen Diskursen aus S_1 zu ermitteln, mittels derer dann auf diskursspezifische Verhältnisse in S_2 zu schließen und aus den so gewonnen Hypothesen über S_2 Aussagen über S_1 zu gewinnen, die dann wiederum überprüfbar sind. Es gilt nämlich (nicht nur) für die Korpuslinguistik das Diktum von Griess:

> ... there are no meanings, no functions, no concepts in corpora – corpora are (usually text) files and all you can get out of such files is distributional (or quantitative / statistical) information.[5]

Die Frage ist nun, wie man den Zusammenhang von S_1 mit S_2 so modellieren kann, dass er die methodologische Grundlage für linguistische Diskursforschung ergibt. Dazu lohnt ein Blick auf die semiotische Dimension von Sprache. Die grundlegende Beziehung zwischen zwei als Zeichen wahrnehmbaren Phänomenen ist nach Peirce die indexikalische:

> An Index is a sign which refers to the Object that it denotes by virtue of being really affected by that Object.[6]

Diese bekannte Bestimmung Peirces markiert den Blickwinkel, aus dem heraus die linguistische Diskursanalyse die Sprache betrachtet. Dass der linguistische Blick vom indexikalischen Aspekt sprachlicher Zeichen gelenkt wird, ist keineswegs selbstverständlich. Schließlich wird in jedem Einführungskurs in die Linguistik gelehrt, bei sprachlichen Zeichen handele es sich um symbolische, welche ihre Zeichenhaftigkeit Konventionen verdanken. Oft wird dabei aber unterschlagen, dass symbolische Zeichen, sofern sie natürliche sind, weder vom Himmel fallen noch in großen Sprachkonferenzen festgelegt werden, sondern immer dann entstehen, wenn Sprecher*innen Wörter in *ähnlichen* Situationen zu *ähnlichen* Aussagezwecken verwenden. Die Wiedererkennbarkeit und damit die Symbolizität sprachlicher Zeichen

[4] Die folgenden Gedanken in Kap. 2 und 3 sind eine Kurzfassung der Ausführungen in Marcus Müller: Vom Wort zur Gesellschaft: Kontexte in Korpora. Ein Beitrag zur Methodologie der Korpuspragmatik. In: Ekkehard Felder, Marcus Müller & Friedemann Vogel (Hgg.): Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analysen 2012, S. 33–82, hier S. 34 ff., und in Marcus Müller: *Sprachliches Rollenverhalten: Korpuspragmatische Studien zu divergenten Kontextualisierungen in Mündlichkeit und Schriftlichkeit.* Berlin, Boston: De Gruyter (Sprache und Wissen) 2015, S. 48–99.

[5] Stefan Th. Gries: What is Corpus Linguistics?. Language and Linguistics Compass, 3, 2009, S. 1225–1241. doi:10.1111/j.1749-818X.2009.00149.x, hier S. 1226.

[6] Charles S Peirce: Collected Papers. Cambridge/Massachusetts: Harvard University Press 1960, 2.247.

entsteht dadurch, dass im Sprachgedächtnis der Leser*innen oder Hörer*innen ein Zeichen von der typischen Situation seines Gebrauchs „really affected" ist. Es geht um die Tatsache, dass jedes sprachliche Zeichen symbolische Bedeutungen in den sozialen Routinen seiner Verwendung ausprägt. Diese sind aber an – mehr oder weniger – spezifische Kontexte gebunden. In diesem Sinne sind sprachliche Zeichen (S_1) als *Spuren sozialer Situationstypen* zu begreifen.

Der Terminus ‚Spur' ist in die jüngere semiotische Debatte vor allem von Sybille Krämer eingebracht worden.[7] In der jüngeren Debatte hat er besondere Aufmerksamkeit erfahren, da er drei viel diskutierte Dimensionen von Kommunikation miteinander verbindet:[8]

a) Materialität

„Spuren", schreibt Krämer, „treten gegenständlich vor Augen; ohne physische Signatur keine Spur."[9] Spuren sind immer und notwendigerweise Konfigurationen des Materiellen in der Welt, sie verweisen erstens auf ein der Wahrnehmungsgegenwart vorgängiges physisch-körperliches Ereignis und zweitens auf den biophysischen Wahrnehmungsprozess selbst: Spuren müssen gesehen, ertastet, gehört, gerochen werden. Beschreibt man Medien in kommunikativen Prozessen als Spuren, so tritt diese unvorgreifliche Materialität der Kommunikation offen zutage. Diese aber ist ein zentraler Gegenstand der aktuellen Medialitätsdebatte.

b) Semiotizität

Spuren werden erst zu Spuren, wenn sie als solche gedeutet werden. Das erfordert Bedingungen: Es muss ein als rekonstruierbar geltendes Ereignis als Ursache der Spur angenommen werden; es braucht eine Person, die willens und in der Lage ist, eine spezielle materielle Konfiguration als Spur zu deuten und es muss ein Deutungskontext existieren, in dessen Rahmen man die Beziehung zwischen dem der Spurhaftigkeit verdächtigen Konfigurat und einem für ihn ursächlichem Ereignis annehmen kann.

c) Kontextualität

[7] Es war Sybille Krämer die unter Berufung auf u. a. Carlo Ginzburg die alte Historikeridee der Spur für die Kulturwissenschaft wiederentdeckt und weiterentwickelt hat, besonders prominent in Sybille Krämer, Werner Kogge & Gernot Grube (Hgg.): Spur. Spurenlesen als Orientierungstechnik und Wissenskunst. Frankfurt a. M.: Suhrkamp 2007. Vgl. dazu die Kapitel 2.4. und 2.5 in Marcus Müller, *Sprachliches Rollenverhalten* [wie Anm. 4] sowie Marcus Müller: *„Symbols grow". Korpuspragmatik und Wirklichkeit.* In: Claudia Brinker von der Heyde, Nina Kalwa, Nina Maria Klug & Paul Reszke (Hgg.): Eigentlichkeit. Zum Verhältnis von Sprache, Sprechern und Welt. Berlin / Boston: De Gruyter 2015, S. 137–157.

[8] Die folgende Aufzählung ist Müller, *Sprachliches Rollenverhalten* [wie Anm. 4], S. 54 ff. entnommen.

[9] Sybille Krämer: *Was also ist eine Spur? Und worin besteht ihre epistemologische Rolle? Eine Bestandsaufnahme.* In: Sibylle Krämer, Werner Kogge & Gernot Grube: Spur. [wie Anm. 7], S. 11–36, hier S. 15.

Spuren verweisen mindestens in einem Sinne auf soziale Praktiken, nämlich auf diejenigen, derer es bedarf, um die Spur als Spur zu erkennen und zu deuten. Werden Medien in Gefügen kommunikativer Handlungen als Spuren gedeutet, wird außerdem das Verursacherereignis als soziale Praktik ausgewiesen. Daneben verweist der Ausdruck *Spur* darauf, dass die Deutung des materiellen Konfigurats als Spur vom Verursacher eben nicht intendiert wurde: „Spuren werden nicht gemacht, sondern unabsichtlich hinterlassen. [...] Und umgekehrt: Wo etwas als Spur bewusst gelegt und inszeniert wird, da handelt es sich gerade nicht mehr um eine Spur.“[10] Kontextualisierungstheorie hat nachdrücklich darauf aufmerksam gemacht, dass dort, wo Zeichen intentional produziert werden, immer auch nicht-intendierte Nebeneffekte ausgelöst werden, die ihrerseits eine wichtige Rolle in Deutungsverfahren spielen und in Gesprächen und Diskursen auch interaktiv bedeutsam werden (s.u.).

Eine Trennung der Beobachtungsebene von der Beschreibungsebene der Sprachforschung kann nur in methodischer Absicht simuliert werden – der Forschende bleibt dabei immer in seinen Gegenstand verstrickt. Der im Folgenden dargestellte Ansatz, Kontextualisierungsprozesse mit korpuslinguistischen Mitteln nachzuvollziehen und in ihrer Systematik zu begreifen, ist als ein Versuch zu verstehen, die Verstrickung des Forschenden in seinen Gegenstand anzuerkennen und mit einem Verfahren des methodisch kontrollierten Nachvollzugs von Verständigungshandlungen zu begegnen. Als Leser*in oder Hörer*in konstituiert der Forschende die zu analysierenden Sprachdaten als Kommunikate, indem er genau die Kontextualisierungsverfahren anwendet, die zu beschreiben er beabsichtigt. Es wird also ein Verfahren benötigt, mit dem der Forschende sich und anderen die Verallgemeinerbarkeit seines Verstehens verdeutlichen kann. Die im Folgenden beschriebenen sprachstatistischen Verfahren erlauben dem Forschenden in diesem Sinne den Zugriff auf eine *augmented reality* der Kommunikation, indem er sprachliche Muster *messen* und damit die Serialität der Kontextualisierungsangebote in jeweils ähnlichen Sprachverwendungen relativ zu Kontexten *ermessen* kann.

III. Kontextualisierung

Ein Theorierahmen, innerhalb dessen diese Anschauung entwickelt wurde, ist die Kontextualisierungstheorie, die eigentlich in ihren verschiedenen Ausprägungen vielmehr ein Theoriegeflecht darstellt.[11]

[10] Ebd., S. 16, so auch Werner Kogge: *Spurenlesen als epistemologischer Leitbegriff. Das Beispiel Molekularbiologie.* In: Sibylle Krämer, Werner Kogge & Gernot Grube (Hgg.), Spur. [wie Anm. 7], S. 182–221, hier S. 118.

[11] ‚Kontextualisierung' hat in mehreren Bereichen der empirischen Linguistik große Bedeutung erlangt. Die Kontextualisierungstheorie betont, dass Kontexte dynamische Gebilde sind, die von den Teilnehmern der

Mit dem Terminus ‚Kontextualisierung' werden dabei Verfahren beschrieben, mittels derer Äußerungen im Zuge der interaktionalen Praxis in einen von den Interaktanten selbst konstituierten Kontext gestellt werden. Kontextualisierung bedeutet dabei das Herstellen einer Verbindung zwischen einem „empirisch gegeben (beobachtbaren) Datum, das der kontextualisierende Teilnehmer aus einem Zeichenvorrat sprachlicher oder nichtsprachlicher Art auswählt", und einer „Komponente des Hintergrundwissens".[12] Das indexikalisch wirksame Datum wird dabei mit Gumperz „contextualisation cue", dt. „Kontextualisierungshinweis", genannt.[13] Grundlegend ist die Annahme, dass Hintergrundwissen in Form von Schemata organisiert ist.[14] Auer unterscheidet Schemata auf fünf verschiedenen Ebenen:

> das generelle Schema des fokussierten Interagierens, die Schemata des ‚turntaking' (also Rezipient, Zuhörer, Sprecher, Adressat), Handlungsschemata, thematische Schemata sowie Beziehungsschemata. Den einzelnen Ebenen lassen sich die folgenden interaktiven Problemstellungen zuordnen: 1) Reden wir gerade miteinander? 2) Wer spricht (gerade) mit wem? 3) Was tun wir (gerade)? 4) Worüber sprechen wir (gerade)? Sowie 5) Wie stehen wir (gerade) zueinander?[15]

Drei Aspekte werden dabei gesondert herausgehoben: Erstens betont Auer, dass die von ihm postulierten Rahmenfragen im Gespräch von den Teilnehmern ständig neu zu beantworten sind, und zwar potenziell unter Einbeziehung aller ihnen verfügbarer Wahrnehmungsdaten. Zweitens wird gesagt, dass die Schemata mindestens zum Teil voneinander abhängen, dass also die Beantwortung beispielsweise der Frage nach dem Interaktionsmodus eine Voraussetzung zur Beantwortung der Frage nach der Beziehung der Interagierenden darstellt. Drittens wird deutlich gemacht, dass Interpretationsschemata und Interaktionsverlauf in einem ständigen Wechselverhältnis zueinander stehen. So kann ein neues Thema eine Reinterpretation des Interaktionsmodus durch die Gesprächspartner mit sich bringen und damit auch eine neue Beziehung zwischen den Partnern konstituiert werden. In Auer wird dieses Modell erweitert: Neben dem Interaktionsmuster („context brought about") werden die sozialstruktuelle Dimension der Interaktion (v. a. soziale Rollen) und die der Interaktion gegeben physikalischen Kontextaspekte wie Zeit, physische Umgebung usw. („context brought along") thematisiert.[16]

---

sprachlichen Interaktion selbst konstituiert und immer wieder neu modelliert werden. Die Kontextaspekte, die in dieser Anschauung relevant werden, sind vor allem diejenigen, welche den reibungslosen Ablauf der sprachlichen Interaktion garantieren, nämlich Partnereinschätzungen, Situationseinschätzung und Themaeinschätzung. Dieses Modell ist in der interaktionalen Soziolinguistik von John Gumperz vorgeschlagen worden, in Deutschland ist es zuerst durch Peter Auer im Rahmen der Konversationsanalyse aufgegriffen worden [Literaturangaben in Anm. 12 und 13].

[12] Peter Auer: *Kontextualisierung.* In: Studium Linguistik 19, 1986, 22–47, hier S. 24.

[13] John J. Gumperz: *Discourse strategies.* Cambridge: University Press 1982.

[14] Peter Auer: *Kontextualisierung.* [wie Anm. 12], S. 24.

[15] Ebd., S. 27.

[16] Peter Auer: *Introduction: John Gumperz' approach to contextualisation.* In: Peter Auer & Aldo Di Lu-

Eine zweite Fassung des Kontextualisierungsgedankens ist in der Diskursanalyse zu verorten und fasst Kontext allein als kognitive Kategorie auf. Vor allem die jüngsten Arbeiten van Dijks sind hier zu nennen. Van Dijk definiert ‚Kontext' als „the structured set of all properties of a social situation that are possibly relevant for the production, structures, interpretation and functions of text and talk".[17] Die Schnittstelle zwischen Diskurs und Kontext ist hier, wie auch schon bei Auer, die Kognition – Kontexte werden nicht als objektive, stabile Gegebenheiten diskurskonstitutiv sondern als mentale „context models".[18] Diese werden als erkenntnisleitende Wissenskonfigurationen systematisch auf allen relevanten Ebenen beschrieben:

> Such context models are stored in episodic memory, just like the event models are used to represent what a discourse is about. Context models, thus, represent how participants in a communicative event see, interpret and mentally represent the properties of the social situation that are now relevant for them.[19]

Da van Dijk aber eine am Sozialen interessierte Diskursanalyse propagiert, kann das Konzept rein subjektiver mentaler Kontextmodelle nicht ausreichen. Daher führt van Dijk das aus der Sozialpsychologie bekannte Modell der „social cognition" ein:

> It is important to stress that even unique, subjective models of specific events are not entirely personal. They also have important social, intersubjective dimensions. Because of earlier interaction and communication, and more generally due to their socialisation, language users have acquired various kinds of shared knowledge and other beliefs. After generalisation and abstraction, such shared general and social beliefs influence the construction of new models whose intersubjective dimensions enable interaction and mutual understanding in the first place.[20]

Drittens ist der Kontextualisierungsbegriff in der Tradition strukturaler Sprachanalysen funktionalisiert worden. Vor allem Feilke hat diesen entwickelt und in seine Theorie der Sozialität als Ort der sprachlichen Musterprägung eingebettet:[21] Die Bedeutung sprachlicher Zeichen ist damit in der nach

zio (Hrsg.), Contextualization of language, Amsterdam: John Benjamins 1992, 1–37. Die Unterscheidung „brought along" vs. „brought about" stammt aus Volker Hinnenkamp: *Foreigner talk, code–switching and the concept of trouble.* In: Karlfried Knapp, Werner Enninger, & Annelie Knapp-Potthoff (Hgg.): Analyzing intercultural communication, Berlin, New York: de Gruyter 1987, S. 137–181, hier S. 143: „I would like to introduce here the dualistic concept of taking into account what is locally *brought about* through joint effort against and complementary to what is *brought along* in terms of their emergent and prestructured groundedness into the actual encounter".

[17] Teun A. Van Dijk: *Ideology. A multidisciplinary approach.* London: Routledge 1998, S. 211.

[18] Ebd., S. 212.

[19] Ebd.

[20] Ebd. S. 6.

[21] Helmut Feilke: Sprache als soziale Gestalt. Ausdruck, Prägung und die Ordnung der sprachlichen Typik. Frankfurt a. M.: Suhrkamp 1996.

sozialen Regeln ablaufenden Iteration erworbenen Fähigkeit, aktuelle Ausdrucksketten zu indizieren, zu sehen. Kontext ist hier vor allem kommunikativ geprägter *Kotext*. Dieser strukturell inspirierte Kontextbegriff ist insbesondere für die Korpuslinguistik von Bedeutung.

Für eine Methodologie der Digitalen Diskursanalyse geht es darum, den strukturellen Kontextbegriff im Sinne von ›Kotext‹ und den soziolinguistischen Kontextbegriff im weiten Sinne von ›selbstreflexiv verfügbares soziokommunikatives Orientierungswissen‹ methodisch miteinander zu korrelieren. Ich habe an anderer Stelle einen integrativen Arbeitsbegriff von Kontext als theoriegeleitete Grundlegung einer Methodologie der Digitalen Diskursanalyse ausgearbeitet, den ich hier kurz darstelle.[22] Insofern er in methodologischer Absicht entwickelt wurde, ist er perspektivisch und selektiv angelegt. **Ein Kontext ist demnach die deutungsrelevante Umgebung einer sprachlichen Fokuskonstruktion.** Unter einer Fokuskonstruktion verstehe ich eine Form-Funktions-Einheit im Sinne der Konstruktionsgrammatik[23], die typische Zusammenhänge ihres Gebrauchs indiziert. Eine Fokuskonstruktion kann aus einem sprachlichen Ausdruck, einer lexikalisch voll- oder teilspezifischen Phrase (*So ein Mist!; So ein x!*) oder einem lexikalisch unspezifischen grammatischen Muster (*Part Det N!*) bestehen. Die Indexikalität der Fokuskonstruktion hinsichtlich typischer Kontexte ergibt sich für die Kommunikationspartner aus ihrer habitualisierten Sprachgebrauchserfahrung und für den Kontextualisierungsforscher aus korrelativen Befunden aus der Analyse von Serien der Fokuskonstruktion.

Die innere Schicht eines Kontextes bildet der sprachliche Kotext (einschließlich Phänomenen seiner Performanz, Medialität, Prosodie oder Typographie) (S_1). Die Konfiguration dieser inneren Kontextschicht gibt nicht nur Hinweise auf die Interpretation der Fokuskonstruktion, sondern indiziert auch seine eigene Musterhaftigkeit (Text- und Interaktionsmuster). Muster kommunikativer Gattungen wiederum verweisen auf die Typik ihrer Gebrauchssituationen. Damit sind hier insbesondere Aspekte wie der zeitliche Rahmen der Kommunikation, Nähe oder Distanz der Kommunikationspartner[24], konstellative Muster im Raum sowie typische physische Umgebungen angesprochen. Die situative Konstellation der Interaktionsteilnehmer und deren Verhalten indizieren ihre soziale Rolle. In der Gesamtschau ergibt die Musterhaftigkeit von Konstruktionen, Kotexten, Situationen und sozialen Rollenkonstellationen Hinweise auf ihre Eingebundenheit in thematische Kontexte. Aus dieser äußeren Kontextschicht lassen sich dann ggf. Schlüsse auf tiefensemantische Figuren oder Episteme, bezogen auf Gruppen

[22] Marcus Müller: *Kontexte in Korpora* [wie Anm. 4].

[23] Zur Konstruktionsgrammatik vgl. Kerstin Fischer & Anatol Stefanowitsch (Hgg.): *Konstruktionsgrammatik. Von der Anwendung zur Theorie.* Tübingen: Stauffenburg 2007.

[24] Vgl. zu dem Konzept Peter Koch & Wulf Oesterreicher: *Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte.* In: Romanistisches Jahrbuch 36, 1986, S. 15–43.

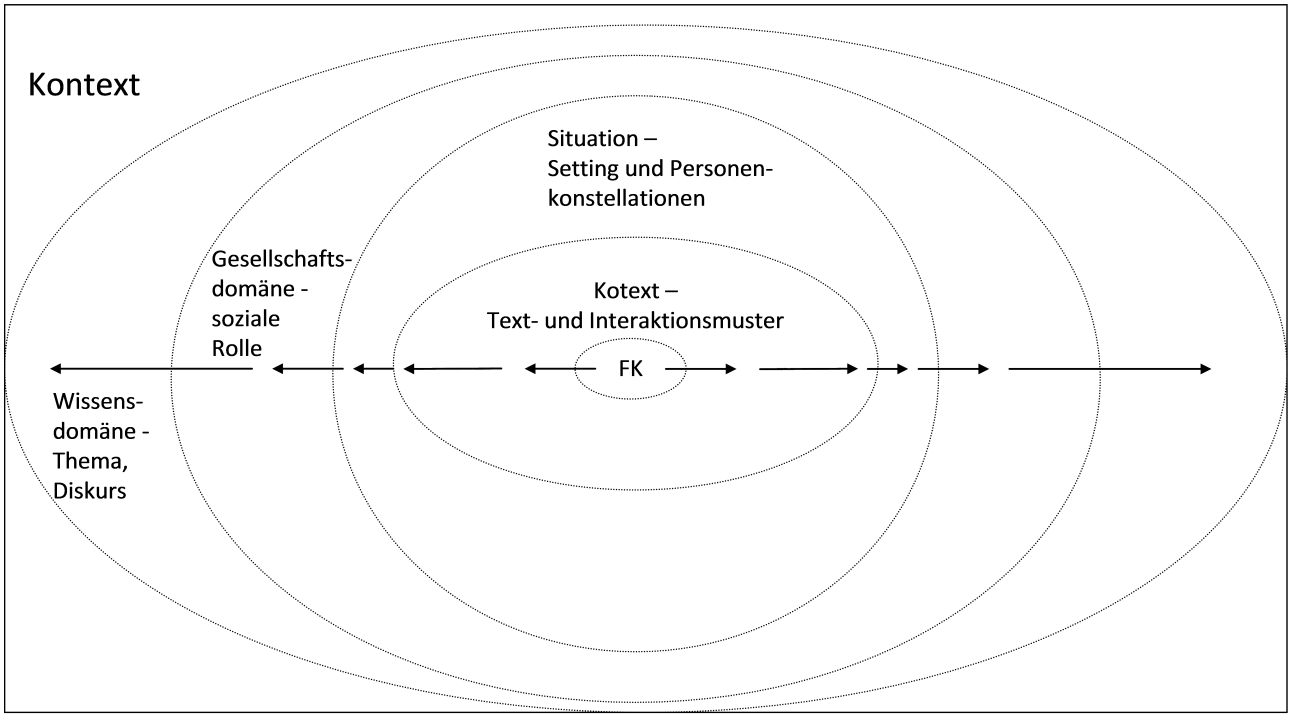

Abbildung 1: Das Zwiebelmodell der Kontextualisierung – FK = Fokuskonstruktion.
(Aus Marcus Müller: *Kontexte in Korpora* [wie Anm. 4], S. 50.)

oder Epochen, ableiten. Die Gesamtheit aller Kontextschichten verweist also auf den oben exponierten weiten Sprachbegriff S_2. So ergibt sich ein Zwiebelmodell der Kontextualisierung (Abb. 1).[25] Alle genannten Kontexte werden als und über Diskurszusammenhänge konstituiert. In unserem Beispiel indiziert die Fokuskonstruktion MP MP, sofern sie in einer bestimmten Konfiguration und einem bestimmten syntaktischen Kotext auftritt (s.u.), den Sprechhandlungstyp ‚Moderieren', den situativen Kontext ‚Mediendiskussion' sowie die soziale Sprecherrolle ‚Journalist'.

Kontextualisierungshinweise dienen lokal dem Verstehen und strukturell dem Aufbau von sozial strukturierter Sinnwelt. Wenn man Serien solcher Kontextualisierungsverfahren nachvollzieht, kann man in der Analyse solcher Verweispotenziale eine diskursgeschichtliche Systematik rekonstruieren.

Wir müssen dabei Kontextualisierungen erster Ordnung (emisch) und Kontextualisierungen zweiter Ordnung (etisch) unterscheiden.[26] Dabei gilt es zu beachten, dass es die Korpuslinguistik immer

[25] Zum Zwiebelmodell der Kontextualisierung ist zu sagen, dass es sich natürlich nicht um den Versuch handelt, die psychologische Realität des Verstehensprozesses zu beschreiben. Ein psychologisch realistisches Modell müsste auf Schichten verzichten, die eine Hierarchisierung der verschiedenen Kontextdimensionen suggerieren. Vielmehr müssten die unterschiedlichen Kontextdimensionen als gleichrangige Aspekte des Verstehensprozesses modelliert werden, die sich ständig wechselseitig bedingen und beeinflussen. Außerdem müsste die hier vorgeschlagene Kontextualisierungsrichtung, die einen Bottom-up-Prozess nahelegt, durch Pfeile in die umgekehrte Richtung als Markierungen eines Top-down-Prozesses ergänzt werden. Das Zwiebelmodell ist bewusst selektiv und reduktionistisch angelegt, um deutlich zu machen, welche Aspekte des Kontextualisierungsverfahrens für eine Methodologie der korpuslinguistischen Kontextualisierungsforschung operationalisierbar sind.

[26] Das Begriffspaar ‚emisch' und ‚etisch' wird hier im Sinne der linguistischen Anthropologie und der empiri-

mit Serialisierungen von Sprache zu tun hat, Meinen und Verstehen aber eine Angelegenheit des Individuums ist. Der Schluss von einem allgemeinen Befund auf einen individuellen ist unzulässig. Die Korpuspragmatik hat daher zwei Möglichkeiten: Entweder deutet sie ihre semiotischen Befunde, die in der Regel Häufigkeiten des Auftretens einer Fokuskonstruktion relativ zu einem Verwendungszusammenhang betreffen, als probabilistische Möglichkeitsrahmen des Meinens und Verstehens im jeweils adressierten Zusammenhang. Oder aber sie nutzt die seriellen Befunde, um in qualitativen Analysen Interpretationen individueller Meinens- und Verstehensprozesse zu unterstützen. Der so verstandene Kontextualisierungsbegriff soll also ein Verfahren der linguistischen Spurensuche legitimieren, mit dem von korpuslinguistisch auffindbaren Kotexten auf soziopragmatische Kontexte geschlossen wird.

emisch:	intrakommunikative Kontextualisierung durch die Kommunikationspartner
	intuitive Heuristik
	wahrnehmungsbasiert
	dynamische Kontextmodelle
	„online“[27]
etisch:	systematisierender Nachvollzug durch den Sprachforscher
	forensische Heuristik
	korpusbasiert
	statische Kontextparameter
	rekonstruktiv

Nun ist in der Diskurslinguistik die Metapher einschlägig, Diskurse als zerdehnte Gespräche aufzufassen.[28] Beispielsweise begegnet öfter der Ausdruck ‚Gesellschaftsgespräch‘ als Synonym für Diskurse.[29] Erst die Korpuslinguistik aber ist nun in der Lage, die methodische Leitmaxime der Aufweisbar-

schen Sozialwissenschaften gebraucht, vgl. zur entsprechenden Diskussion um die Begriffe die Beiträge in Thomas N. Headland, Kenneth Pike & Marvin Harris (Hgg.) (1990): Emics and etics. The insider/outsider debate. Newbury Park u. a.: SAGE 1990. Sie beziehen sich auf die Perspektivik der Kategorienbildung im Forschungsprozess: ‚Emisch‘ bedeutet also ›aus der Perspektive der Partizipanten eines Sozialgefüges bzw. einer Interaktion‹, ‚etisch‘ ›aus der Perspektive und Begriffslogik des Forschers‹.

[27] Peter Auer: *Online-Syntax – oder: was es bedeuten könnte, die Zeitlichkeit der mündlichen Sprache ernst zu nehmen.* In: Sprache und Literatur 85, 2000, S. 43–56.

[28] Nämlich im Fahrwasser von Ehlichs Begriff der ‚zerdehnten Kommunikationssituation‘; vgl. Konrad Ehlich: *Zum Textbegriff.* In: Annely Rothkegel & Barbara Sandig (Hrsg.), Text – Textsorten – Semantik. Linguistische Modelle und maschinelle Verfahren. Hamburg: HBV 1984, S. 9–25.

[29] Vgl. Sven Kersten Roth: *Interpersonale Diskursrealisationen – Überlegungen zu ihrer Integration in die diskurssemantische Forschung.* In Ingo Warnke & Jürgen Spitzmüller (Hrsg.), Methoden der Diskurslinguistik.

keit von anzusetzenden Kontextualisierungsverfahren, so wie sie die Gesprächsforschung hervorgebracht hat, mit einem kognitiv beherrschbaren Anteil an Metaphorik bei der Analogiebildung auf die Diskursanalyse zu übertragen. Dies geschieht, indem durch den Nachweis der Serialisierungen sprachlicher Konstruktionen in ähnlichen Kontexten plausibel gemacht wird, dass spezifische Gebrauchsaspekte der Konstruktionen von den Diskursakteuren als Kontextualisierungshinweise angeboten, affirmiert und ratifiziert wurden.

Für die diskursanalytische Kontextualisierungsforschung ergeben sich damit folgende methodologischen Grundsätze:

a) Meinen und Verstehen kann korpuslinguistisch nicht untersucht werden.

Emische Kontextualisierungsverfahren erster Ordnung sind individuelle und dynamische psychologische Prozesse. Sie können korpuslinguistisch direkt nicht untersucht werden. Dazu eignen sich experimentelle Verfahren und solche der teilnehmenden Beobachtung. Dafür können der Korpusanalyse aber etische Kontextualisierungsverfahren zweiter Ordnung vollzogen werden. Sie basieren auf Korrelationsanalysen zwischen sprachlichen Mustern und typischen Zusammenhängen ihrer Verwendung.

b) Digitale Kontextualisierungsforschung muss viable Ergebnisse haben.

Wenn ein sprachliches Phänomen als Kontextualisierungshinweis angesprochen wird, dann muss dies eine Begründung in den untersuchten Sprachdaten haben. Sie gilt als gegeben, wenn ein sprachliches Phänomen (Fokuskonstruktion) in einem bestimmten Verwendungszusammenhang signifikant häufig auftritt. Ein so eruierter Kontextualisierungshinweis zweiter Ordnung darf nicht einfach als Hinweis erster Ordnung ausgegeben werden. Er kann aber als Deutungshintergrund für die qualitative Text- und Gesprächsinterpretation gelten.

Kontextualisierungsforschung in diesem Sinne braucht Korpora, deren Sprachdaten mit Metadaten zu den genannten Kontextebenen ‚Situation', ‚soziale Sprecherrolle' und ‚thematischer Diskurs' ausgezeichnet sind, oder aber mindestens solche, die hinsichtlich einer Kontextdimension zusammengestellt worden sind. So verstandene Kontextualisie-rungsforschung basiert immer auf einem Vergleich der sprachlichen Verhältnisse in zwei oder mehreren Vergleichskorpora.

Sprachwissenschaftliche Zugänge zur transtextuellen Ebene, 2008, S. 323–358. Berlin, New York: de Gruyter.

IV. Korpuslinguistik und Diskursanalyse

Die Beobachtung, mit der ich diesen Text eingeleitet habe, entstammt dem Kontext einer Studie zu sprachlichem Rollenverhalten in der Bioethikdebatte.[30] In diesem Zusammenhang haben wir ein Korpus aus zeitgenössischen Äußerungen zur Bioethikdebatte erstellt. Es versammelt transkribierte mündliche Äußerungen (Gattungen: Diskussion, Interview, Statement, Call-In, Umfrage), schriftliche Äußerungen (Gattungen: Internetforum, policy paper, schriftlicher Debattenbeitrag, Leserkommentar) und Wortlautprotokolle (Gattungen: Vortrag, Anhörung, Parlamentsdebatte, Fernsehdebatte). Das Korpus umfasste zum Zeitpunkt der hier präsentierten Analysen ca. 3,8 Millionen Wortformen.[31]

Die Forschungsfrage zu der Studie, aus der die Belege im Eingang dieses Textes stammen, lautete: Gibt es sprecherrollenspezifische Gebrauchsmuster von Modalpartikeln im Bioethikdiskurs? Modalpartikeln sind unflektierbare Modifikatoren von Äußerungen wie *denn, doch, eben, ja, halt, mal* oder *wohl*. Sie haben Homonyme in anderen Wortarten und treten stets in unbetonter Position auf. Modalpartikeln markieren Sprechereinstellungen zu Propositionen im Hinblick auf Interaktionspartner, daher indizieren sie

- Partner- und Rollenhypothesen,
- Situationshypothesen,
- Sprecherhypothesen über gemeinsames verstehensrelevantes Wissen und
- Sprecherhypothesen über die Einbettung von Äußerungen in Diskurse.

Methodisch stellt sich das Problem, dass Modalpartikeln nicht automatisch in Korpora auffindbar sind, weil sie von Part-of-Speech-Taggern nicht als solche klassifiziert werden können. Unser Korpus ist, wie viele andere Korpora des Deutschen, mit dem Stuttgart-Tübingen-Tagset getaggt.[32] Dort fallen

[30] Marcus Müller: *Sprachliches Rollenverhalten* [wie Anm. 4].

[31] Es handelt sich Äußerungen um mündliche und schriftliche öffentliche Beiträge zur Bioethikdebatte in deutscher Sprache aus den Jahren 2000–2010. Das Korpus liegt im XML-Format in CQPWeb@discourselab vor. Für jede Einzel-Äußerungen (Print- oder elektronischer Text, Blogpost, mündlicher Gesprächsbeitrag) sind die Kontextschichten soziale Sprecherrolle, thematischer Kontext, Gattung und Medium neben den spezifischen Informationen zur Situation als Metadaten hinterlegt, so dass zu diesen Kontextaspekten Korrelationsstudien angestellt werden können. Erfasst sind die thematischen Kontexte ‚grüne Gentechnik', ‚allgemeine Bioethik', ‚Sterbehilfe', ‚Lebensbeginn', ‚rote Gentechnik', ‚Stammzellenforschung', ‚Transplantationsmedizin', ‚Reproduktionsmedizin/ PID'. Eine genaue Korpusbeschreibung findet sich in Müller, *Sprachliches Rollenverhalten* [wie Anm. 4] S. 125–149. Unter `https://discourselab.de/resources/#heideko5` sind die einzelnen Datenquellen der Korpustexte aufgelistet.

[32] Vgl. Anne Schiller, Simone Teufel, Christine Stöckert & Christine Thielen: *Guidelines für das Tagging deutscher Textcorpora mit STTS*. 1999. Online unter: `http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf` [zuletzt am 12.11.2017]. Die Grobkörnigkeit des STTS ist vor allem beim Tagging mündlicher Sprachdaten ein Problem. Vgl. dazu sowie zu einem Entwicklungsansatz zum Tagging von Gesprächsparti-

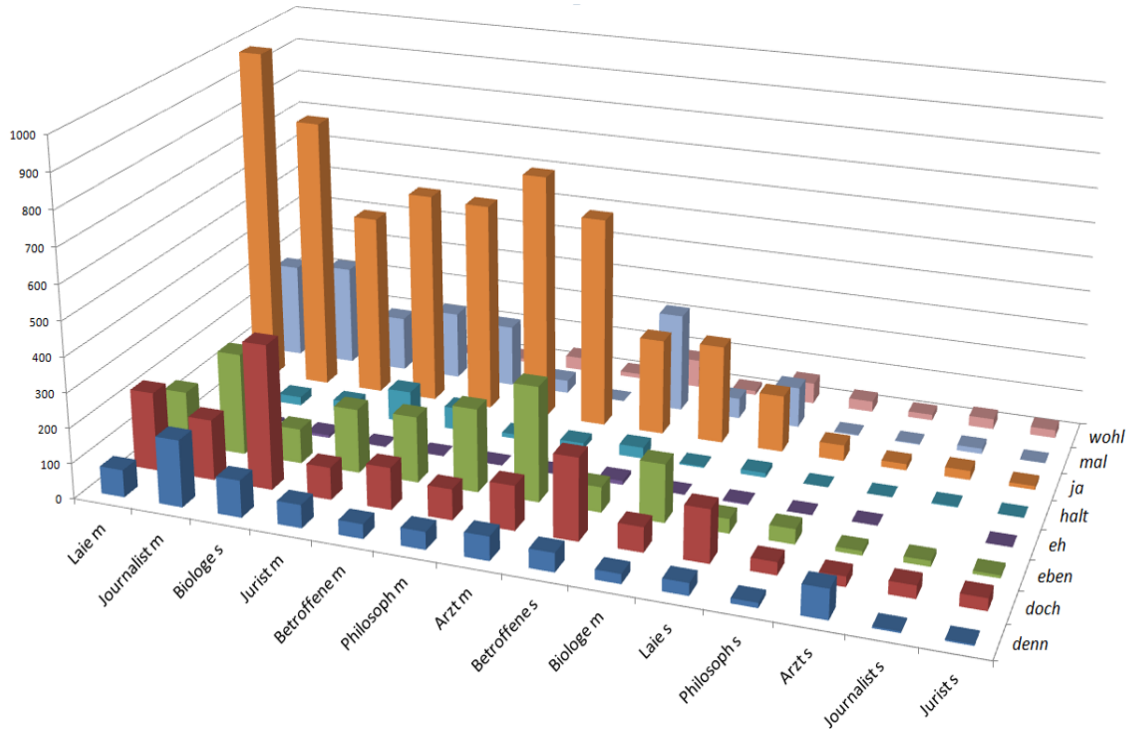

Abbildung 2: Verteilung ausgewählter Modalpartikeln im Bioethik-Korpus relativ zu sozialen Sprecherrollen und Medium.

Modalpartikeln unter die Klasse ADV, also „Adverb“. ADV ist eine Art Restkategorie für nicht flektierbare Wörter, bei denen es sich nicht um Adpositionen, Konjunktionen oder leicht identifizierbare Partikeln (wie z.B. Negationspartikeln) handelt. Um Modalpartikeln zu annotieren, bin ich stufenweise vorgegangen: Suche mit einer Liste der Hauptklasse von Ausdrücken, die in der Forschung als Modalpartikeln beschrieben und vom TreeTagger als ADV klassifiziert werden: *denn, doch, eben, halt, ja, mal, wohl.*[33] Die sortierten Konkordanzen wurden auf der Basis von Stellungseigenschaften der Modalpartikeln vorbereinigt und schließlich manuell annotiert. Auf der Basis ergab sich die in Abb. 2 gegebene Verteilung von Modalpartikeln, differenziert nach sozialen Sprecherrollen und Medium (m steht für mündlichen und s für schriftlichen Sprachgebrauch).

Aus der Distributionsanalyse lassen sich folgende Schlüsse ziehen: Die mediale Kontextdomäne von Modalpartikeln ist, wie in der Forschung schon oft festgestellt, die Mündlichkeit. Die Gebrauchsdomänen hinsichtlich der sozialen Sprecherrollen in der mündlichen Interaktion sind ‘Journalist*in’ und ‘Laie’. Die soziale Rollendomäne in der schriftlichen Interaktion ist die Biologenrolle. Das ist leicht

keln Swantje Westpfahl & Thomas Schmidt: *POS für(s) FOLK – Part of Speech Tagging des Forschungs- und Lehrkorpus Gesprochenes Deutsch.* In: JLCL, Band 28 (1) 2013, S. 139–153.

[33] Es fehlt die Modalpartikel *schon*, da eine hinreichend intersubjektiv stabile Lesartendisambiguierung in Testläufen mit mehreren Annotator*inn*en nicht gewährleistet werden konnte. Die übrigen Modalpartikeln können auf der Basis eines Merkmalkatalogs eindeutig bestimmt werden.

dadurch erklärbar, dass Biolog*inn*en im Korpus vor allem in der Forenkommunikation zur Grünen Gentechnik engagiert sind, wo der stark interaktive Kontext und der agonale Interaktionsmodus zum relativ gesehen häufigen Gebrauch von Modalpartikeln führen. Bemerkenswert ist die negative Korrelation zwischen den Modalpartikeln *doch* und *eben*. Mit *doch* wird typischerweise operiert, wenn man aus einer als schwächer bzw. ohnmächtig empfundenen Diskursposition heraus den Geltungsanspruch einer Aussage bekräftigen will – hier ist die Gebrauchsdomäne die Biologenrolle in der Forenkommunikation im thematischen Kontext Grüne Gentechnik:

Jetzt beginnen sie doch mal selber weiterzudenken ...
Nehmen sie doch Wissenschaft endlich Ernst!
Machen sie doch endlich mal die Augen auf!

Die Rollendomäne von *eben* dagegen ist die ‚Arzt'-Rolle in mündlichen Medienäußerungen. Mit dem Wörtchen *eben* bekräftigt man den mit einer Aussage verbundenen Wahrheitsanspruch, indem man wie nebenbei auf einen vorausgesetzten, aber nicht explizit gemachten Common Sense verweist:

... ich kann eben nicht nur pränatal während der Schwangerschaft therapieren ...
... das noch Wichtigere ist eben die Beratung ...
... die Prognoseeinschätzung ist eben manchmal gar nicht so einfach ...

Die Modalpartikel *eben* indiziert hier also einen Expertengestus, der von Vertreter*innen der ‚Arzt'-Rolle in öffentlichem Raum oft mit Verweis auf den Praxisbezug einer Expertise eingenommen wird.

Am systematisch interessantesten ist hier aber die Journalistenrolle in der Mündlichkeit – und zwar, weil hier erstens die gleichmäßigste Verteilung der Modalpartikeln vorliegt und weil zweitens im Vergleich zur Gesamtverteilung signifikant häufig Doppelokkurrenzen von Modalpartikeln vorkommen. Aus diesem Zusammenhang entstammten die Belege im Eingang dieses Textes. Ich möchte jetzt eine kurze Gesprächssequenz geben und diese kurz besprechen. Es handelt sich um ein Call-In Format, die Sequenz folgt unmittelbar auf eine Überlappungssequenz mit einem durchgestellten Anrufer:

Moderatorin: *jammern sie nicht auf hohem nivEAU (.) uns geht es* ***doch eigentlich*** *sehr GUT (.) sie haben vorhin gesagt wir haben keine wartezeiten (.) schauen wir nach großbritAnnien (.) WIR gu-*

cken NICHT ob man einer fünfundachtzigjährigen noch ein hüftgelenk einbauen kann oder nicht

Experte: *herr weber hat ja durchaus EINgeräumt dass er privat versichert ist aber wir müssen einfach FESTstellen (.) in der bundesrepublik sind fast neunzig prozent eben NICHT privat versichert*[34]

Die Moderatorin nimmt den Einwand eines durchgestellten Anrufers auf und gibt ihn an den Studioexperten weiter. Die Modalpartikelkonstruktion, die uns hier interessiert, unterläuft ihr im zweiten Äußerungssegment ihres Gesprächsbeitrags. Die Modalpartikeln *doch eigentlich* erfüllen an dieser Stelle eine ganze Reihe an Funktionen: Sie dienen erstens als anaphorischer Index, indem sie den Rückbezug auf die vorherige Äußerungseinheit steuern: die Äußerung der Modalpartikeln ergibt nur Sinn, wenn sie auf eine Voräußerung Bezug nimmt. Zweitens fungieren die Partikeln aber auch als kataphorischer Index, indem sie Erwartungsrahmen zur Ratifikation des Common Ground der Äußerungen eröffnen: Die Sprecherin macht mit den Partikeln deutlich, dass sie erwartet, dass der Geltungsanspruch der gemachten Äußerung vom Gesprächspartner geteilt wird. Sowohl durch die rückverweisende als auch durch die vorverweisende Indexikalität kommt der Partikelkonstruktion eine wichtige Rolle beim Aufbau von Gesprächskohärenz zu. Durch die Modalpartikel *doch* wird drittens der Schärfegrad des Gesprächs forciert, indem die vorgetragene Äußerung als Gegenposition markiert und damit das Gespräch in einen konfrontativen Interaktionsmodus überführt wird. Gleichzeitig wird viertens durch *eigentlich* die Äußerung abgetönt, ins Prinzipielle gehoben: Die Moderatorin leitet das konfrontative Moment, welches durch *doch* markiert wird, von der interpersonalen Gesprächsdimension weg und zur inhaltlichen hin. Damit leistet sie prospektiv Imagearbeit: Sie verhindert, dass der Gesprächspartner die modalisierte Äußerung als persönlichen Angriff deutet und in der Folge evtl. seine Rolle als Experte in dieser öffentlichen Situation in Gefahr gerät. Die Konstruktion dient ferner der konzeptuellen Orientierung der Äußerung, indem diese fünftens (mit *doch*) als Gegenthese markiert wird und sechstens (mit *eigentlich*) im Common Ground verortet wird: Der Aussage, „es geht uns sehr gut“ wird also einerseits unterstellt, sie sei Konsens, andererseits wird sie als Gegenposition exponiert.

Ein naheliegendes Argument gegen diese Analyse lautet, die Partikeln seien jeweils gesondert zu analysieren, es liege keine Notwendigkeit vor, sie als eine syntaktische Einheit aufzufassen. Dagegen spricht allerdings die Beobachtung, dass diese Partikeldopplung in Äußerungen von Journalist*inn*en

[34] Radiosendung zum Thema *Krankenhäuser in Not, Fühlen Sie sich gut versorgt?* Gesendet am 19.08.2008 auf BR 2.

in konfrontativen Mediendiskussionen musterhaft sind. Das Muster besteht eben darin, dass jeweils die erste Modalpartikel eine forcierende und die zweite eine abtönende Funktion hat. Da dieses Muster ein klares Kontextualisierungspotenzial hinsichtlich Sprecherrolle und Situationstyp aufweist, kann man von einer Konstruktion im Sinne der Konstruktionsgrammatik sprechen. Ich habe sie an anderer Stelle die *Eiertanz-Konstruktion* genannt, weil sie den kommunikativen Eiertanz von Journalist*inn*en markiert, einerseits im Gespräch für ausreichend Dynamik und Kontrastierung von Positionen zu sorgen, andererseits aber es zu ermöglichen, dass alle Gesprächspartner ihr Gesicht wahren können.[35] Weitere Beispiele für diese Konstruktion:

dem kann man ***doch eigentlich*** *nicht kommen und sagen du musst . . . .*

dann müssen wir uns ***doch eigentlich*** *selbst die frage stellen wollten und könnten wir . . . .*

dann müssen sie ***doch eigentlich*** *die debatte und die ernsthafte debatte die . . .*

letztlich geht es ***natürlich auch*** *um risiken . . .*

gibt es ***denn auch*** *eine POSitive seite an diesem neuen gesetz für sie. . .*

dass höhere tabaksteuer dann ***eben auch*** *fließt. . .*

die frage ist ***doch eigentlich*** *gar nicht so falsch gestellt ob das nicht . . .*

Es finden sich auch Expansionen der Konstruktion im Korpus, bei denen die abtönenden und die forcierenden Elemente jeweils wiederholt werden:

aber wir können ***dann eben halt doch*** *hoffen dass*

das war meines erinnerns im reichsstrafgesetzbuch von achtzehnhunderteinundsiebzig ***eben dann doch eben schon*** *biologistisch festgelegt*

Die mediale Domäne der Konstruktion ist also die Mündlichkeit, ihre Rollendomäne 'Journalist*in'. Die Zuordnung zur Journalistenrolle ist wohlgemerkt nicht eindeutig, sondern probabilistisch. Natürlich finden sich auch Belege von Sprecher*inne*n anderer Rollen. Ein thematisches Kontextualisierungspotenzial innerhalb der im Korpus untersuchten Themen besteht nicht, es gibt auch keine Hinweise darauf, dass Konstruktion typisch für die Themen der Bioethikdebatte wäre, auch wenn das hier nicht gezeigt werden kann. Allerdings gibt es ein diskursstrukturelles Merkmal der Bioethikdebatte,

[35] Marcus Müller: *Geisterkonstruktionen. Zum Beispiel PPER ADV ADV.* In: Alexander Ziem / Alexander Lasch (Hgg.): Konstruktionsgrammatik IV. Tübingen: Narr 2015, S. 203–221.

welches die Konstruktion triggert, nämlich ihre agonale Grundstruktur: Als situative Domäne lassen sich nämlich klar Mediendiskussionen ausmachen, in denen konträre Positionen verhandelt werden. Der dominierende Sprechhandlungskontext ist das Moderieren solcher Situationen. Ein hier wichtiger Aspekt dieser Sprechhandlung ist die Aufgabe, die Mehrfachadressierung von Äußerungen in Mediengesprächen zu managen. Damit ist gemeint, dass mit Gesprächsbeiträgen in den Medien nicht nur die Gesprächsteilnehmer adressiert werden, sondern immer auch die Rezipienten des Medienformats. Der moderierenden Person fällt damit die Aufgabe zu, einerseits Gegenpositionen herauszuarbeiten und Gespräche interessant zu halten und andererseits dafür zu sorgen, dass die Form gewahrt bleibt und niemand sein Gesicht verliert. Vor diesem Hintergrund lässt sich die kommunikative Leistung der Eiertanz-Konstruktion wie folgt zusammenfassen:

- phorische Orientierung (Rück- und Vorbezug);
- konzeptuelle Orientierung (Markierung der Gegenthese, Markierung von Common Ground);
- Ausbalancierung der emotionalen Involviertheit: Forcierung (*doch/denn*) + Abtönung (*eigentlich/auch*).

Damit zeigt sich die Konstruktion im oben eingeführten als eine Spur sozialer Praxis. An dieser im kommunikativen Alltag völlig unauffälligen Formulierungsgewohnheit lässt sich beobachten, wie soziale Erfahrung in der Serialität des Sprachgebrauchs zur grammatischen Form gerinnt.

V. Es tun: Digitale Diskuranalyse

In der hier vorgeführten Forschungsskizze habe ich auf Informationsressourcen unterschiedlichen Auflösungsgrades zurückgegriffen und diese auf meine Forschungsdaten bezogen: Dazu gehören in der Linguistik validiertes Wissen über das Sprachsystem, z. B. über die Wortartenkategorisierung des Deutschen. Hier habe ich wie viele Kolleg*inn*en auf das Klassifizierungssystem vertraut, welches dem Stuttgart-Tübingen-Tagset (STTS) zugrunde legt. Da ich an einer Wortart interessiert war, die STTS nicht eigens klassifiziert, die aber in der linguistischen Literatur eingeführt und anerkannt ist,[36] habe ich auf der Basis der mit STTS vorstrukturierten Daten eigene Klassifizierungen durchgeführt, bei denen ich mich aber vollständig auf das institutionalisierte Wissen meines Fachs verlassen habe. Dasselbe gilt für etliche hier vorausgesetzte Kategorien und Begriffe, die ich unkommentiert übernommen

[36] Auch wenn Modalpartikeln zum Teil als *Abtönungspartikeln* bezeichnet werden, vor allem in der älteren Tradition.

habe, wie ‚Phrase', ‚Äußerung', Gesprächsbeitrag' oder ‚Kohärenz'. Hinzu kommen Annahmen über Aussagekraft des untersuchten Korpus. Die Aufgabe war es, die zeitgenössische öffentliche Debatte zur Bioethik in Deutschland in einem Korpus so abzubilden, dass Aussagen darüber gemacht werden können. Es war also die Frage nach der Repräsentativität zu bearbeiten. Hier haben wir uns wiederum nach einer im Fach geltenden Standardannahme gerichtet, die lautet, dass das unter dem Begriff ‚Repräsentativität' in den empirischen Sozialwissenschaften verhandelte quantitative Verhältnis zwischen Stichprobe und Grundgesamtheit in der Korpuslinguistik gar nicht sinnvoll zu adressieren sei, weil nämlich die Grundgesamtheit (hier: die Menge aller zeitgenössischen Äußerungen in der deutschen Bioethikdebatte) auf systematische Weise unbekannt ist. Die Korpuslinguistik versteht daher unter Repräsentativität das Ausmaß der Übereinstimmung zwischen Grundgesamtheit und Stichprobe, bezogen auf die Variation der Korpusdaten: „Representativeness refers to the extent to which a sample includes the full range of variability in a population."[37] Eine Studie über die geschriebene Sprache des Gegenwartsdeutschen sollte also möglichst nicht nur Kochrezepte und Traueranzeigen enthalten, sondern das gesamte Textsortenspektrum abbilden. Nach diesem Grundsatz haben wir das Korpus zusammengestellt.[38] Zum Korpusdesign notwendig war außerdem eine aus der philosophischen Literatur bezogene Definition von ‚Bioethik'[39] sowie die in jedem Einzelfall zu treffende Frage danach, ob eine Äußerung darunter fällt oder nicht. Als Entscheidungshilfe haben wir auf der Basis der Fachliteratur und eigener Piloterhebungen ein Kategorienraster bioethisch relevanter Themen erstellt. Dann haben wir in der entsprechenden Arbeitsgruppe[40] ein Jahr lang einschlägige Sendungen aus Radio und Fernsehen aufgenommen und darin enthaltene direkt geäußerte Debattenbeiträge (O-Töne aus Reportagen und Call-In Formaten, Mediendiskussionen) transkribiert. Außerdem haben wir auf Basis der Themenliste schlagwortbasiert einschlägige Threads in Internetforen, online publizierte Positionspapiere sowie bioethisch relevante Printpublikationen erhoben. Die Zuordnung von Sprecher*inne*n zu Rollenlabels haben wir möglichst auf der Basis von Fremdzuschreibungen (z.B. durch Ankündigungen von Diskussionssendungen in Fernsehzeitschriften) oder auf der Grundlage interaktiv validierter Selbstpo-

[37] Douglas Biber: *Representativeness in Corpus Design.* In: Literary and Linguistic Computing 8(4) 1993, S. 243–257.

[38] Eine ausführliche Darstellung und Diskussion der Erstellung des Heidelberger Bioethikkorpus findet sich im Müller, *Sprachliches Rollenverhalten* [wie Anm. 4], S. 125–149.

[39] „[...] die kritische Auseinandersetzung mit den moralischen Dimensionen in den von den Biowissenschaften betroffenen Handlungskontexten Biomedizin, Biotechnologien und Ökologie." Zitiert nach Marcus Düwell, Christoph Hübenthal & Micha Werner (Hgg.): *Handbuch Ethik.* Stuttgart u. a.: J. B. Metzler 2002, S. 247.

[40] Am Aufbau und der Auszeichnung des Heidelberger Bioethikkorpus waren neben mir als Hilfskräfte und Kollegen beteiligt: Thomas Bögel, Anja Chaluppa, Johanna Emmerich, Clara Herdeanu, Franziska Köder, Anna Mattfeldt, Jörn Stegmeier und Friedemann Vogel. Außerdem die Kolleginnen und Kollegen des Medienzentrums der UB Heidelberg.

sitionierungen (z.B. der Positionierung als ‚Laie' in Onlineforen) vorgenommen. Wie alle Prozesse der Korpusbildung war auch dieser eine ständige Abfolge von strikt methodengeleitetem Vorgehen und arbeitspraktischen Entscheidungen, die nach bestem Wissen und Gewissen getroffen wurden und in Arbeitsgruppensitzungen diskutiert wurden.

VI. Nur ein Schritt entfernt: der Text

Das Korpus habe ich mit Standardmethoden der Korpuslinguistik bearbeitet.[41] Das wichtigste Instrument dabei ist gleichzeitig das einfachste: die Konkordanz. Die Möglichkeit, das Auftreten einer Fokuskonstruktion als Serie zu betrachten, deren syntagmatische Umgebung sich in Variation und Musterhaftigkeit beschreiben lässt, ist die Basisressource der digitalen Diskursanalyse. Die kontextsensitive Serie ist die entscheidende Größe, die den einzelnen Beleg mit statistischen Korpusbefunden verbindet. Der Befund, dass zweistellige Modalpartikelkonstruktionen im Bioethikkorpus signifikant häufiger auftreten als in Vergleichskorpora, ist weder eine linguistische noch eine diskursanalytische Aussage. Die funktional dichte Beschreibung einer Formulierung aus einem Einzelbeleg wiederum erlaubt keine Aussagen über Musterbildungen, ist damit ebenfalls weder linguistisch noch diskursanalytisch hinreichend. Auch das Verfahren, aus der Einzelbeleganalyse gewonnene Hypothesen quantitativ am Korpus zu prüfen, ist problematisch, weil nicht gezeigt werden kann, dass die Eigenschaften des Musters, auf Grund derer wir eine Nullhypothese ablehnen können, auch die textanalytisch herausgearbeiteten Eigenschaften der Einzelformulierung sind. Die Kontrolle über Gebrauchsbedingungen von Konstruktionen im Diskurs erlaubt letztlich daher nur die Konkordanz.

Auf der Idee der sprachlichen Serie im Kontext basiert im Wesentlichen die korpuslinguistische Forschung, und zwar insofern sie mit Bezug auf Firth Sprache als im Gebrauch situierte, serialisierte und kulturell kontextualisierte Ausdruckskomplexe begreift.[42] Vor diesem Hintergrund lassen sich nun aber Studien, die ein sprachstrukturimmanentes Interesse haben, von solchen unterscheiden, die Regelhaftigkeiten in den Phänomenbereichen ‚Kognition', ‚Handlung', ‚Interaktion', ‚Gesellschaft' oder ‚empirische Epistemik' auf der Basis authentisch verwendeter Sprache (gesprochen und geschrieben) erforschen. Diese Forschung haben Felder, Müller und Vogel unter dem Etikett *Korpuspragmatik* zusammengefasst:

[41] Darüber kann man sich einen Überblick verschaffen z. B. in Tony McEnery & Andrew Hardie: *Corpus linguistics: method, theory and practice.* (= Cambridge textbooks in linguistics) 1. publ., Cambridge [u.a.]: Cambridge Univ. Press 2012.

[42] John R. Firth: *Papers in Linguistics* (1934–1951). Oxford: University Press 1957. Elena Tognini-Bonelli, Elena: *Corpus linguistics at work.* Amsterdam: John Benjamins 2001, S. 157 ff.

> Unter Korpuspragmatik verstehen wir einen linguistischen Untersuchungsansatz, der in digital aufbereiteten Korpora das Wechselverhältnis zwischen sprachlichen Mitteln einerseits und Kontextfaktoren andererseits erforscht und dabei eine Typik von Form-Funktions-Korrelationen herauszuarbeiten beabsichtigt. Solche Kontextfaktoren betreffen potenziell die Dimensionen *Handlung, Gesellschaft* und *Kognition*. Die Analyse bedient sich insbesondere einer Kombination qualitativer und quantitativer Verfahren.[43]

Natürlich beschränkt sich korpuspragmatische Forschung nicht auf das Auswerten von Konkordanzen. Neben der Kookkurrenzanalyse (inklusive der Berechnung von Kookkurrenz-Netzwerken[44]) und verschiedenen Formen der Analyse spezifischer, teilspezifischer oder lexikalisch unspezifischer n-Gramme ist in der thematisch orientierten Diskursanalyse vor allem die Forschung mit Keywords einschlägig.[45] Das sind Wörter, die in einer Datenpopulation A relativ zu einer Datenpopulation B signifikant häufiger auftreten als das erwartbar wäre, wenn man annimmt, dass alle Wörter in beiden Populationen gleich verteilt sind. Außerdem wird mit automatischen, halbautomatischen und händischen Annotationen semantischer und diskurspragmatischer Kategorien (z. B. ‚Definieren' oder ‚Argumentieren') gearbeitet. Aus der französischen Tradition der digitalen Diskursanalyse stammen lexikometrische Verfahren.[46] Innovation findet man in letzter Zeit vor allem in der Verfeinerung, Rekombination, Variation und Visualisierung dieser Ansätze[47] sowie in der Integration diskurslinguistischer Modelle, z.B. aus der Frame-Semantik.[48] Gemeinsam ist diesen Zugängen, dass die digitalen Zugriffe als Teilprozesse der Analyse begriffen werden, deren Ergebnisse theoriegeleitet zu kontextualisieren sind. Das verläuft typischerweise – explizit oder implizit – über den Leitaspekt der Diskursfunktion sprachlicher Zeichen, die wiederum nur am Einzelbeleg nachzuvollziehen und an der Konkordanz in ihrer Geltung abzuschätzen ist. Deswegen ist Digitale Diskursanalyse sinnvollerweise: immer nur einen Schritt vom Text entfernt.

[43] Ekkehard Felder, Marcus Müller & Friedemann Vogel: Korpuspragmatik. Paradigma zwischen Handlung, Gesellschaft und Kognition. In: Ekkehard Felder, Marcus Müller & Friedemann Vogel (Hrsg.), Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analysen. Berlin, Boston: De Gruyter 2012, 3–30, S. 4.

[44] Da sind Netzwerke, die durch mehrfache Iteration von Kookkurrenzanalysen, ausgehend von einer Fokuskonstruktion, entstehen, vgl. dazu Friedemann Vogel: *Das LDA-Toolkit. Korpuslinguistisches Analyseinstrument für kontrastive Diskurs- und Imageanalysen in Forschung und Lehre.* In: Zeitschrift für Angewandte Linguistik 3, 2012, S. 129–165, hier S. 139, sowie Marcus Müller, *Sprachliches Rollenverhalten* [wie Anm. 4], S. 168 ff.

[45] Vgl. Paul Baker: *Using corpora in discourse analysis.* London, New York: continuum 2006.

[46] Vgl. Ronny Scholz & Annika Mattissek: *Zwischen Exzellenz und Bildungsstreik. Lexikometrie als Methodik zur Ermittlung semantischer Makrostrukturen des Hochschulreformdiskurses.* In: Martin Nonhoff u. a. (Hgg.): Diskursforschung. Ein interdisziplinäres Handbuch. Band II. Methoden und Analysepraxis. Perspektiven auf Hochschulreformdiskurse. Bielefeld: transcript 2014, S. 86–112.

[47] Einen schönen Überblick über aktuelle Tendenzen der diskursanalytisch orientierten Korpuslinguistik, insbesondere im Bereich der Visualisierung, bietet die von Noah Bubenhofer betriebene Seite: https://www.bubenhofer.com/ [zuletzt am 13.11.2017].

[48] Vgl. Marcus Müller & Jörn Stegmeier: *Investigating risk, uncertainty and normativity* [wie Anm. 3].