Ruth M. Mell & Cerstin Mahlow

Digitale Diskursanalyse: Annotation und formale Modellierung von Diskursen

Abstract

Dieser Beitrag widmet sich einer Auslegung des Begriffs *digitale Diskursanalyse* und fokussiert dabei auf den Aspekt des Digitalen. Wir argumentieren, dass Digitalität sich nicht auf Diskursmedium und -material oder verwendete Analysewerkzeuge bezieht, sondern für einen epistemologischen Ansatz steht, der es erlaubt, bislang eher vage und narrativ formulierte Elemente von Diskursen zu explizieren. So ist es möglich, Diskurse tatsächlich zu modellieren und empirisch fundierte Aussagen abzuleiten. Wir stellen den Prozess der Annotation von Diskursen auf verschiedenen Ebenen ins Zentrum und gelangen so zu einer adäquaten Sicht von digitaler Diskursanalyse als wissenschaftlich explizite und reproduzierbare Modellierung.

1. Einleitung – Problematik

Das Digitale ist eine zentrale Komponente der heutigen Gesellschaft. *Das Digitale* bzw. *Digitalität* ist nach einer eher allgemeinen Lesart lediglich Ausdruck unserer Lebensart und damit Chiffre des 21. Jahrhunderts, des Zeitalters der *Digitalisierung*. Konzeptionell unterscheiden wir im Englischen *digitization* und *digitalization*. Im Deutschen wird beides als *Digitalisierung* bezeichnet, es handelt sich jedoch um verschiedene Konzepte (vgl. Brennen & Kreis 2014, Chapco-Wade 2018, Mahlow & Hediger 2019). *Digitization* ist der Prozess, um ein physisches Objekt (etwa einen gedruckten oder handgeschriebenen Text) in ein elektronisches Medium (also maschinenlesbaren Text) zu wandeln. So werden kulturelle Artefakte, wie Bilder oder Texte, maschinell verarbeitbar und in digitalen Repräsentationsformen zugänglich gemacht, also *digitalisiert*.

Erst solcherart digitalisierte Objekte erlauben Digitalisierung im Sinne von *digitalization*: Durch spezifische Aufbereitungs- und Anreicherungsprozesse (Annotation) werden sie vielfältig *digital* befragbar, womit auch digitale Analyseverfahren an Bedeutung gewinnen (Rammerstorfer 2019). Hierbei handelt es sich um die Übertragung von Prozessen in das digitale Medium, also die Be- oder Verarbeitung digitaler Daten in Information. Dabei wird davon ausgegangen, dass diese digitalisierten Methoden und Prozesse Vorteile gegenüber herkömmlichen analogen Methoden bieten: So ist etwa die automatische Frequenzanalyse spezifischer Wörter als Grundlage diskurslinguistischer Betrachtungen schneller und weniger fehleranfällig als die manuelle und erlaubt das Einbeziehen sehr großer Textmengen. Digitalisierung verändert und beeinflusst demnach geisteswissenschaftliches Arbeiten: Methoden, Arbeitsabläufe und Gegenstandsbereiche verändern sich unter dem Vorzeichen des Digitalen.

Digitale Transformation meint schließlich die grundlegende Veränderung von Prozessen durch die Entwicklung und den Einsatz nicht nur digitalisierter, sondern vollkommen neuer, auf digitalen Daten operierender Methoden und Werkzeuge (Chapco-Wade 2018, Mahlow & Hediger 2019). Essentiell sind Abstraktion und Modellierung, die neue Kompetenzen und Erkenntnisse bedingen und ermöglichen. Die Sprachen, die in dieser Welt gesprochen werden, scheinen denen zu entsprechen, die wir aus der analogen Welt kennen.

Mehr und mehr entdeckt die Linguistik diese Welt als Forschungsgegenstand und als Forschungsraum. *Digitale Diskurse*, *digitale Linguistik*, *digitale Diskurslinguistik*, also digitale linguistisch motivierte Diskursanalyse, – diese Begriffe sind omnipräsent, eine allgemein akzeptierte Bestimmung oder Definition des *Digitalen* fehlt jedoch weiterhin. Studien zu Internetlinguistik, Diskursen in Massenmedien oder sozialen Medien, Untersuchungen von Online-Diskussionen operieren mit digitalen Texten, also mit digitalen Forschungsobjekten. Zugriff auf Texte und Analysen von linguistischen und sozialen Phänomenen erfolgen mittels digitaler Werkzeuge. Aber reicht dies aus, um *digitale Diskursanalyse* definierend und differenziert zu beschreiben und sie etwa von der Diskursanalyse allgemein bzw. der Diskurslinguistik abzugrenzen? Beobachten wir das Entstehen eines neuen Teilbereichs der

Linguistik bzw. der Diskursanalyse oder sind wir nur Zeugen der Adaption zeitgenössischer Forschungsmethoden, die zwar neue *Tools* für datenreichere und umfangreichere Diskursanalysen verfügbar machen, aber die etablierten *Methoden* der Diskursanalyse/Diskurslinguistik unangetastet lassen? Wie genau also ist „das Digitale" in der digitalen Diskursanalyse sprachlich zu fassen? Referiert der Begriff rein auf die technischen Affordanzen von Daten und Analysezugriff oder haben wir es bei der digitalen Diskursanalyse mit der Etablierung eines neues Forschungsparadigmas zu tun?

Um diese Frage zu klären, werden wir zunächst in Abschnitt 2 verschiedene Sichtweisen auf und Interpretationen von digitaler Diskursanalyse darstellen, um unseren Fokus zu verdeutlichen. In Abschnitt 3 gehen wir auf Ursprung und Ansätze der Methode „Diskursanalyse" ein, die Ausgangspunkte sind, um in Abschnitt 4 unsere Vorstellung von Diskursanalyse als Modellierung zu präsentieren. Annotation in Abschnitt 5 – sowohl als Prozess (das Annotieren von textuellen Daten) wie auch als Produkt (die Annotation(en) von textuellen Daten auf verschiedenen linguistischen Ebenen) – erlaubt uns anschließend eine holistische Sicht auf die digitale Diskursanalyse. Diese bietet die Grundlage für unseren abschließenden Vorschlag in Abschnitt 6 einer Definition von „digitaler Diskursanalyse" als Methode im Sinne einer digital transformierten Analyse und Interpretation von Diskurs verstanden als sprachliche Äußerungen im Kontext mit dem Fokus auf Abstraktion, Modellierung und Reproduzierbarkeit.

2. Verwandte Ansätze: Analyse von digitalen Diskursen vs. digitale Analyse von Diskursen

Im folgenden Abschnitt stellen wir unterschiedliche Lesarten des *Digitalen* mit Blick auf die Linguistik im Allgemeinen und der Diskursanalyse im Besonderen vor. Die Ansätze dieser ‚digitalen (Teil-)Disziplinen' referieren dabei auf digitale Daten oder auf digitale Werkzeuge.

Die Korpuslinguistik etwa sammelt Daten dieser digitalen Welt in großen Mengen, konserviert sie und macht sie für Analysen nutzbar. Solche Analysen sind digitalisierte traditionelle Untersuchungsmethoden wie die oben

genannte Frequenzanalyse oder völlig neu entwickelte Formen wie *topic modeling* (für einen Überblick Blei 2012).

Eigene Forschungsbereiche etablieren sich, die sich mit computervermittelter Kommunikation beschäftigen, wobei hier der Fokus auf dem Computer als *Medium* liegt. Einflussreich war bei der Benennung dieses Teilgebietes der linguistischen Forschung nicht zuletzt Crystals „Internet Linguistics" (2011), welches sich der Untersuchung dieser neuen Form der Kommunikation von seiner vermittelnden Instanz aus nähert. Auch im deutschsprachigen Raum wird die Untersuchung dieses digitalen Phänomenbereiches der Sprache immer häufiger als „Internetlinguistik" (etwa Marx & Weidacher 2014) bezeichnet. Denn obgleich es eine neue digitale Internetsprache nicht gibt (Marx & Weidacher 2014: 91, Crystal 2011: 2), die es zu untersuchen gelte, finden sich kommunikative wie stilistische Merkmale in der digitalen Kommunikation, welche in ihrer Modalität sowie in Lexik und Stil in dieser Form und Zusammenstellung von der Face-to-Face-Kommunikation oder Kommunikationsformen in etablierten Massenmedien wie der Zeitung abweichen und daher einer gesonderten Betrachtung bedürfen. Hier liegt der Fokus also auf dem Kommunikationsmedium bzw. auf dem *Ort* und dessen Spezifika, an dem Kommunikation stattfindet und der diese Kommunikation beeinflusst. Es geht um Kommunikation, um Sprache im Internet.

In Zeiten zunehmender Kommunikation und Information über verschiedenste Kanäle, die jeden Einzelnen ohne viel Aufwand und Ausbildung sowohl zu Sender wie zu Empfänger machen, interessieren vor allem Antworten auf Fragen, wie wir etwa Gesagtes oder Geschriebenes, d.h. Informationen, verarbeiten und interpretieren, was wir als valide Information oder Täuschungsversuch, mit anderen Worten: was wir als Faktum, als wahre Aussage, einstufen und wie dies unsere Sicht auf die Welt beeinflusst und verändert. Es ist dabei eine der zentralen Fragen der Diskursanalyse, warum welche Aussagen zu einem gegebenen Zeitpunkt getätigt werden und wie durch die Praxis des Aussagens und kommunikativen Aushandelns die Gegenstände in der Welt konstituiert werden. Dieses Interesse ist nicht neu:

> Et si ces plans sont reliés par un système de rapports, celui-ci n'est pas établi par l'activité synthétique d'une conscience identique à soi, muette et préalable à toute parole mais par la spécificité d'une pratique

> discursive. On renoncera donc à voir dans le discours un phénomène d'expression – la traduction verbale d'une synthèse opérée par ailleurs ; on y cherchera plutôt un champ de régularité pour diverses positions de subjectivité. Le discours, ainsi conçu, n'est pas la manifestation, majestueusement déroulée, d'un sujet qui pense, qui connaît, et qui le dit : c'est au contraire un ensemble où peuvent se déterminer la dispersion du sujet et sa discontinuité avec lui-même. (Foucault 1969: 74)

Gemeint ist die Hervorbringung von Wissen, von dem, was eine Gesellschaft und ihre Mitglieder als Kenntnisse, Erkenntnisse, Bildung usw. anerkennen, d.h. kommunikativ aushandeln und damit als wahr und akzeptiert anerkennen via Äußerungen einzelner Akteure. Wissen über die Welt wird im Diskurs einerseits durch Sprache zugänglich, andererseits konstruieren wir mit Sprache überhaupt erst die Dinge, von denen wir sprechen. Nur durch die Sprache ist uns ein Zugriff auf unsere Wirklichkeit möglich, nur so können wir über das reden, was uns umgibt. Auf diese Weise ist auch unsere Wirklichkeit diskursiv konstruiert (Teubert 2010). Für uns ist die Wirklichkeit der kommunikativen Aushandlung im Sprechen darüber eben die einzige Wirklichkeit, über die wir verfügen und auf die wir aus linguistischer Perspektive einen wissenschaftlichen Zugriff haben.

Neben zunächst vor allem pragmatischen Ansätzen hat die digitale linguistische Diskursanalyse zu Beginn der 2010er Jahre Interesse an digitalen Untersuchungsgegenständen. Also zunächst an Diskursen, die im Internet entstehen und dort stattfinden, es geht um die Analyse digitaler Diskurse. Wir finden in einzelnen Studien Einschränkungen der untersuchten Diskurse, die sich durch die untersuchten Medien oder Bereiche innerhalb des Internet ergeben. Neben der Bezeichnung des digitalen Diskurses hat sich der Begriff des *Online-Diskurses* bzw. die Formulierung ‚Diskurse in den sozialen Medien' etabliert (vgl. hierzu u.a. Fraas et al. 2013, Sommer et al. 2013). Hiermit sind jedoch immer Diskursanalysen gemeint, deren Korpusdaten genuin digital (born digital) sind, also z.B. Facebook-, Twitter- oder Wikipediadaten (etwa Gredel 2018).

Eine andere Lesart von *digitaler Diskursanalyse* als der Analyse digitaler Untersuchungsgegenstände schließt digitalisierte Textdaten ein – also *nach* ihrer Entstehung in maschinenlesbare Form gewandelte und damit digital bear-

beitbare Texte –, die mit entsprechenden digitalen Werkzeugen und Methoden der Korpuslinguistik analysiert werden. Hier ist der Fokus ganz klar auf den Methoden, die digitale Daten als Grundlage voraussetzen, jedoch deren ursprüngliche Herkunft und den Ort des Diskurses außer Acht lassen.

Beiträge in etablierten Medien wie Zeitungen, Zeitschriften, Fernseh- oder Hörfunk sind jedoch weiter wesentlicher Bestandteil von Diskursthemen. Ein häufiges Muster ist dabei, dass diese ‚traditionellen', nicht-digitalen Medien eingehen in ein digitales Archiv und im weiteren Diskursverlauf im digitalen Medium verfügbar sind und referenziert werden. Auf diese Weise entstehen hybride Diskurse (Gloning 2019), deren Komponenten teilweise ‚born digital' sind, die aber auch in traditionellen Medien veröffentlicht werden und dann zusätzlich in den digitalen Verfügungsraum einmünden und zum Teil ähnliche Eigenschaften wie ‚traditionelle' Diskurse aufweisen (vgl. Mell & Gredel 2021).

Diese Verständnisse eröffnen unterschiedliche Perspektiven auf die Bezeichnung *digitaler* Diskurse, die zum einen digitale bzw. digitalisierte Daten, zum anderen digitale Werkzeuge als Referenzpunkt annehmen. Beide Lesarten sind jedoch noch nicht auf der Ebene der *Digitalen Transformation* angelangt; es werden lediglich etablierte Forschungsprozesse in zeitgenössisch adäquate Formen übertragen.

Ein dritter Ansatz bezieht sich auf einen Diskurs *über* digitale Objekte und meint damit „Digitalisate oder Repräsentationen von physischen Objekten als (neue) Entitäten" (Bender, Kollatz & Rapp 2018: 107). Dabei stellt sich die Frage, inwiefern die korpuslinguistische Praxis mit Blick auf ihre epistemologischen Prozesse selbst als diskursiv bezeichnet werden kann. Hierfür werden Annotations- und Vernetzungspraktiken in informationstechnisch gestützten Analyseverfahren betrachtet, die einen Metadiskurs konstruieren, in den ausdrücklich auch nichtsprachliche Zeichen einbezogen werden (Bender, Kollatz & Rapp 2018: 116).

Die etablierte Diskursanalyse wurde mit Busse und Teubert (1994) eingeleitet und hat sich in der Form des transtextuellen Mehrebenenmodells DIMEAN (u.a. Spitzmüller & Warnke 2011) im deutschsprachigen Raum besonders nachhaltig durchgesetzt. Die gerade beschriebenen Verwendungs- und Definitionsansätze bilden die Grundlage, um zu bestimmen, wie sich epistemolo-

gische Grundannahmen der „Diskurslinguistik nach Foucault“ (Warnke 2007) mit einer Theorie der Annotation als epistemologisch informierte diskursive und digitale Praxis verbinden lassen. Der Fokus liegt nicht auf der Natur des Diskurses, sondern auf der Methode der Analyse und den Konsequenzen, die sich aus einer derartigen Begriffsdefinition der digitalen Diskursanalyse für die Forschungspraxis ergeben.

3. Methoden der Diskursanalyse: Historischer Abriss

Bevor wir die Bedeutung des Digitalen in der digitalen Diskursanalyse genauer untersuchen, zeigen wir in diesem Abschnitt verschiedene Verständnisse von Diskursanalyse: als ursprünglich textgebundene Methode, als qualitativ-hermeneutische Methode der Wissensgenese und als Spielart der Korpusanalyse. Diese bilden die Grundlage für ein neues Verhältnis von Diskurs und Korpus und damit ein neues Verständnis von Diskursanalyse und deren Digitalität.

3.1. Ursprung der Diskursanalyse als textgebundene Methode

Der Begriff *discourse analysis* wird von Harris bereits 1952 als Bezeichnung für die systematische Analyse der in einem Text (egal ob gesprochen oder geschrieben) enthaltenen Elemente und der Art ihrer Verknüpfung vorgeschlagen. Die so explizit gemachten Verbindungen und Muster erlauben eine neue Sicht auf einen Text und damit Interpretationen und Schlussfolgerungen: „We may not know just WHAT a text is saying, but we can discover HOW it is saying” (Harris 1952: 1), die über Satzgrenzen hinausgehen, jedoch klar *innerhalb* eines Textes – verstanden als miteinander verbundene Äußerungen – bleiben. Wichtig ist hier bereits die systematische und formale Analyse, die die Extraktion von Grammatiken erlaubt und die Verwendung von Computern bedingt, die zu dieser Zeit gerade erst entwickelt und deren Eignung für Untersuchungen im Bereich Kommunikation und Information von Shannon und Weaver (1949) und Wiener (1948) untersucht und vorhergesagt werden. Wenn Bar-Hillel (1962) später von Mechanisierung linguistischer Untersu-

chungen spricht, nennt er *discourse analysis* von großen Korpora und die Verwendung von KWIC-Indizes (Keyword-in-Context) als Beispiel.

Erst mit der größeren Verfügbarkeit von Computern allgemein und mit der Akzeptanz automatischer Analyse sprachlicher Daten als wissenschaftsrelevante Voraussetzung linguistischer Untersuchungen entwickelte sich das Feld der Korpuslinguistik. Die bereits von Harris (1952) beschriebenen Beziehungen zwischen linguistischen Elementen oder Sequenzen von Elementen über Satzgrenzen hinaus unter Einbezug von außertextlichen Klassifizierungskriterien, wie Distributionsklassen (Harris 1952: 5), werden heute als Vektormodelle im sogenannten *topic modeling* (Blei 2012) verwendet. Die heute relativ einfache Erzeugung solcher Modelle und die graphisch ansprechende Darstellung entsprechender Beziehungen und KWIC-Indizes werden von modernen korpuslinguistischen Werkzeugen erwartet, um Grundlagen für die Beantwortung der eigentlichen Forschungsfragen zu schaffen. Diskursanalyse in diesem Sinn ist eine Voraussetzung für linguistische und soziolinguistische Forschung.

3.2. Diskursanalyse als qualitativ-hermeneutische Methode der Wissensgenese

Seit Ende der 1980er Jahre wird *Diskursanalyse* im deutschsprachigen Raum in Anlehnung an Foucaults Konzeptionen der Hervorbringung von Wissen als „Methode der historischen Wissensanalyse" (Busse 1987: 251) verstanden. Wissen ist in diesem Sinne alles, „was in einer Gesellschaft *Bedeutung* hat" (Busse 1987: 256–257, Hervorhebung im Original). Damit wird eine diskurslinguistische Methodologie begründet, welche später durch Busse und Teubert (1994) weiter maßgeblich beeinflusst wird. Diese umfasst eine breit gefächerte sprachwissenschaftlich deskriptive sowie qualitativ-hermeneutische Bearbeitung von Äußerungszusammenhängen in transtextuellen Strukturen.

Insofern damit Wissen und Bedeutung in eins fallen, wählt Busse für seine Form der Diskursanalyse die Bezeichnung „Diskurssemantik", welche er als Erweiterung einer linguistisch fundierten Wort- und Begriffsgeschichte

fasst. Zur Rekonstruktion sprachlich vermittelten Wissens werden unter der Perspektive des Diskurses einzelne signifikante Elemente an der Textoberfläche, d.h. Aussagen, identifiziert, aus denen Daten selektiert sowie in Bezug auf ihre Bedeutungsaspekte analysiert und kategorisiert werden. Dies erfolgt unter der Vorannahme, dass alle Aussagen eines Diskurses als gleichwertige Teilsequenzen verstanden werden, die zusammen ein intertextuelles Verweisnetz schaffen (vgl. u.a. Teubert 2010; 2013), aus dem Bedeutung und Wissen rückführbar sind (Mell 2015, Gredel & Mell 2018).

Für dieses Verständnis von Diskursanalyse sind ab dann nicht mehr einzelne *Texte* (oder Textsammlungen) als Grundeinheit zwingend die zentralen Analyseeinheiten von Sprache, insofern die linguistische Diskursanalyse oder auch Diskurslinguistik in ihren Analysen im Sinne des Erweiterungspostulats auf transtextuell bzw. intertextuell verbundene *Aussagen* zurückgreift: Es geht vielmehr nun um einzelne Wörter oder Phrasen, die in Sprachgebrauchssituationen, d.h. in ihrem Kontext und in ihrer grammatischen Funktion, analysiert werden. Die daraus gezogenen Schlüsse werden – in soziologischer Tradition – und in Anlehnung an Berger & Luckmann ([1966] 2009) als Rekonstruktion von Wirklichkeit verstanden, als ein durch Äußerungen vermitteltes Abbild der Welt und ihrer Artefakte. Dafür werden die Äußerungen, welche in (digitalen) Fundstellen manifest werden, unabhängig von der Einheit Text, miteinander in Beziehung gesetzt. Es entsteht ein intertextuelles Netz an Aussagen über Textsorten hinweg; Janich (2008) spricht von „Text(sorten)vernetzung". Spitzmüller und Warnke (2011: 22–23) sehen in der Diskurslinguistik, verstanden als „transtextuelle Sprachanalyse", eine Erweiterung der Textlinguistik und schlagen eine argumentative Brücke vom „Erweiterungspostulat" bei Heinemann und Viehweger (1991: 22) zu den transtextuellen Sprachstrukturen, welche sie im Konzept der „Historischen Semantik" von Busse (1987; 1988) verorten. Texte gewinnen dann jedoch für die Diskursanalyse allerdings als Objekte der Korpusanalyse wieder an Bedeutung.

3.3. Diskursanalyse als Korpusanalyse

Korpuslinguistik als wissenschaftlicher Ansatz ist ganz klar motiviert durch das Bedürfnis, linguistische Theorien empirisch zu überprüfen und Aussagen zu Sprachentwicklung allgemein, wie zu lexikalischen Entwicklungen im Besonderen sowohl diachron wie auch synchron empirisch abzustützen. Kilgarriffs Arbeiten zu lexikographischen Erkenntnissen, gestützt auf systematische Auswertung spezifischer Korpora wie auf den systematischen Vergleich verschiedener Korpora (Kilgarriff 2001), sind die Grundlage moderner Korpuslinguistik. Die Software-Entwicklungen von Kilgarriff et al. zu geeigneten Werkzeugen zur Zusammenstellung, Aufbereitung und Exploration von Korpora führten schließlich zu Sketch Engine (Kilgarriff et al. 2014, https://www.sketchengine.eu), dem ersten mit umfassenden computerlinguistischen und lexikographischen Methoden ausgestatteten Korpuswerkzeug.

Das Aufkommen von frei verfügbaren maschinenlesbaren Texten in großem Umfang, d.h. die verbreitete Nutzung des Web und später entsprechender Suchmaschinen wie Yahoo oder Google, brachte der empirisch gestützten linguistischen Forschung einen großen Entwicklungsschub. Der noch vor einigen Jahren offenbar notwendige Hinweis, systematisch und reproduzierbar vorzugehen und sich nicht auf undokumentierte Angebote Dritter zu verlassen (Kilgarriff 2007) scheint heute nicht mehr notwendig. Zudem sind Korpuserstellung und -aufbereitung für spezifische Fragestellungen oder hinsichtlich verschiedenster Kriterien ausgewogene Daten-Repräsentation von *Sprache* mittlerweile etablierte Grundlagenmethoden der Korpuslinguistik. Damit eröffnen sich Möglichkeiten, Korpora für weitere linguistische Fragen über lexikalische Aspekte hinaus zu nutzen – eben auch für die Diskursanalyse.

Folgt man Busse und Teubert (1994: 14) in ihrer Definition, sind zu Beginn der 1990er Jahre *Diskurse* konkret als *virtuelle Textkorpora* und – abzüglich des Konzepts des Virtuellen – als identisch aufzufassen bzw. zumindest als partiell identisch zu denken: der Diskurs ist das Korpus. *Virtuell* überdies verweist auf die – rein auf den Analyseabsichten des Forschenden – beruhende Textauswahl, welche dann als ein homogenes Korpus an Texten verstanden wird, das die Grundlage der Analyse darstellt. Damit hat dieser Kor-

pusbegriff einen wesentlich auf die Inhalte der Textsammlung bezogenen Impetus; *virtuell* ist hier keinesfalls mit *digital* gleichzusetzen.

Bubenhofer (2018: 214) stellt fest, dass ein zu untersuchender Diskurs nicht zwingend durch ein *vorab* klar definiertes und ausgewogenes Korpus im korpuslinguistischen Sinne repräsentierbar ist, da Diskurse als Aussagensystem begriffen werden müssen, die quer zu Texten liegen können (vgl. auch Spitzmüller & Warnke 2011). Es ist die Aufgabe des Diskurslinguisten/der Diskurslinguistin, die Texte, welche das Korpus *bilden*, so auszuwählen und einzugrenzen, dass sie das Diskursobjekt möglichst adäquat *abbilden*. Gefordert wird also eine *Modellierung* des Diskurses als Korpus, ohne diesen Begriff explizit zu verwenden. Diese Eingrenzung ist häufig als Schwäche diskursanalytischer Arbeiten beschrieben worden. Kritisiert wird zumeist, dass der gewählte Querschnitt an Sprache zu stark von der Forschungsfrage des Wissenschaftlers/der Wissenschaftlerin abhängig sei. Dem wird üblicherweise mit der Verwendung von dynamisch erzeugten Subkorpora entsprechend Spezifikationen inhaltlicher (Domäne) oder formaler (Textsorte, Zeitraum, Länge, etc.) Art aus sehr großen thematisch nicht eingeschränkten Korpora wie DeReKo (Kupietz et al. 2018) oder Swiss-AL (Krasselt et al. 2020) begegnet.

Die mittlerweile verbreiteten KWIC-Analysen in Korpuswerkzeugen und -abfragesystemen wie CQPWeb (http://cwb.sourceforge.net/cqpweb.php), COSMAS (https://www2.ids-mannheim.de/cosmas2/), CorpusWorkBench (http://cwb.sourceforge.net) oder KorAP (https://github.com/KorAP) für einzelne Wörter, Phrasen oder syntaktische Muster und vor allem ihre Darstellung erlauben einen schnellen Eindruck über die unmittelbare Umgebung der Fundstellen in allen Texten des aktuell betrachteten (möglicherweise dynamisch zusammengestellten) Korpus, lassen jedoch den Diskurs im Sinne Harris' als *Struktur* und *Muster* innerhalb eines Textes über Satzgrenzen hinweg völlig verschwinden.

Was dagegen erst so möglich wird, ist das Identifizieren und Visualisieren transtextuell und intertextuell verbundener Aussagen. Die übliche Aufbereitung von Korpora durch automatische Annotation einfacher morphosyntaktischer Angaben erlaubt zwar die Suche nach syntaktischen Mustern und damit mehr als die Suche nach Zeichenketten auf der sprachlichen Oberfläche; der Zusammenhang von Bedeutung und Verwendung sprachlicher Zei-

chen ist so jedoch nicht automatisch ermittelbar und damit nicht visualisierbar und bleibt mithin unsichtbar. Korpuslinguistische Methoden bilden daher lediglich eine quantitative Grundlage zur Unterstützung qualitativ-hermeneutischer Methoden nah am Text.

4. Modellierung von Diskurs

Allen oben gezeigten Ansätzen zur Diskursanalyse gemein ist der Aspekt des Aufspürens von Mustern, der Erkenntnisgewinn durch *Abstraktion* und *Modellierung*, auch wenn diese Begriffe nicht expliziert werden. Die Sicht von Analyse als Prozesse der Modellierung, also der Erstellung eines Modells des Diskurses, drängt sich auf. Stachowiak konstatiert

> Hiernach *ist alle Erkenntnis Erkenntnis in Modellen* oder *durch Modelle*, und jegliche menschliche Weltbegegnung überhaupt bedarf des Mediums 'Modell'. (Stachowiak 1973: 56, Hervorhebung im Original)

Wir zeigen daher also zunächst, wie wir generell Modellierung verstehen und fokussieren dann auf die Verbindung von linguistischer und soziologischer Modellierung von Diskursen.

Im Bereich der Linguistik gelingt Modellierung nur auf expliziten Daten, die Formalisierungen im Sinne von Gladkij und Mel'čuk ([1969] 1973: 15) als folgerichtige, eindeutige und explizite Beschreibung linguistischer Daten erlauben:

> Die formale Beschreibung eines Objektes ist notwendigerweise mit einer Schematisierung und Vergröberung der untersuchten Fakten verbunden. […] Ein kompliziertes Objekt zu erkennen bedeutet nichts anderes, als die Gesetzmäßigkeiten seines Aufbaus festzustellen, d.h. seine Grundkomponenten herauszuarbeiten und Regeln zu formulieren, nach denen sich diese Komponenten miteinander verbinden. (Gladkij & Mel'čuk [1969] 1973: 16)

Die Autoren weisen zu Recht darauf hin, dass dies für die Linguistik eine (zumindest in der Mitte des 20. Jahrhunderts) ungewohnte Methode ist, jedoch diese Vergröberungen durch fortschreitende Annäherung der Be-

schreibung immer geringer wird und es sich dabei um einen potentiell endlosen Prozess handelt (Gladkij & Mel'čuk [1969] 1973: 16, Fußnote 2). Je mehr wir also implizites Wissen explizit machen können, je präziser wir den Kontext von Äußerungen, und damit hier von *Diskurs*, explizieren können, desto näher kommen wir einer Analyse im Sinne einer Bestimmung struktureller Beziehungen.

Damit sind die Voraussetzungen gegeben, um Modelle nach Stachowiak (1973) unter Berücksichtigung der Merkmale von Abbildung, Verkürzung und Pragmatik konstruieren zu können:

> Modelle sind stets Modelle *von etwas*, nämlich Abbildungen, Repräsentationen natürlicher oder künstlicher Originale, die selbst wieder Modelle sein können. […] Modelle erfassen im allgemeinen *nicht alle* Attribute des durch sie repräsentierten Originals, sondern nur solche, die den jeweiligen Modellerschaffern und/oder Modellbenutzern relevant erscheinen. […] Modelle sind ihren Originalen nicht per se eindeutig zugeordnet. Sie erfüllen ihre Ersetzungsfunktion a) für *bestimmte* – erkennende und/oder handelnde, modellbenutzende – *Subjekte*, b) innerhalb *bestimmter Zeitintervalle* und c) unter Einschränkung auf *bestimmte gedankliche oder tatsächliche Operationen*. (Stachowiak 1973: 131–133, Hervorhebungen im Original)

Ein Modell als

> un intermédiaire à qui nous déléguons la fonction de connaissance, plus précisément de réduction de l'encore-énigmatique, en présence d'un champ d'études dont l'accès, pour des raisons diverses, nous est difficile (Bachelard 1979: 3)

erlaubt dann zu verstehen, *wie* Erkenntnisse gewonnen werden können. Und erst solche Modelle ermöglichen tatsächlich, das Versprechen einzulösen, dass „Discourse analysis tells […] how a discourse can be constructed to meet various specifications" (Harris 1952: 30).

Sowohl Pêcheux (1975) wie Foucault übernehmen die Begrifflichkeit (Diskursanalyse) und Methodologie von Harris, fokussieren jedoch den ideologischen Aspekt und lassen die von Harris (1952) in den Vordergrund gerückten linguistischen Aspekte außer Acht (Hodge 2017: 524). Auffällig ist,

dass Foucault nie Bezüge zu Harris' Arbeiten und Ideen explizit macht, obwohl er sich dessen sehr wohl bewusst war (Deleuze [1985] 2018).

Harris betont die Notwendigkeit, über Satzgrenzen hinaus zu analysieren. Dabei umfasst dies explizit nicht nur den gesamten Text, sondern ebenso außertextliche Bezüge: „social and interpersonal situation in which speech occurs" (Harris 1952: 2). Es geht also um den *Kontext* von Äußerungen, jedoch nur um den Kontext des Senders, nicht um den Kontext des Empfängers. Der Soziologe Foucault fokussiert später genau darauf und entwickelt ein dezidiert soziologisch perspektiviertes Modell von Diskurs (Foucault 1972) – die linguistischen Elemente ignoriert er dabei.

Foucault beschreibt Diskurse als Orte von Wissen (Foucault 1969). Seine Diskurstheorie beruht auf der These, dass Erkenntnis und Wissen stets abhängig von den soziokulturellen Codes betrachtet werden müssen, denen sich der Diskursakteur gegenübersieht:

> Les codes fondamentaux d'une culture – ceux qui régissent son langage, ses schémas perceptifs, ses échanges, ses techniques, ses valeurs, la hiérarchie de ses pratiques – fixent d'entrée de jeu pour chaque homme les ordres empiriques auxquels il aura affaire et dans lesquels il se retrouvera. (Foucault 1966: 11)

Spitzmüller & Warnke (2011: 69) spezifizieren das systematische Netz der Ordnungsstrukturen für die Diskurslinguistik, also die linguistisch motivierte Diskursanalyse, indem sie sie auf die sprachlichen Zeichen beziehen, welche zueinander in Beziehung gesetzt werden. Damit führen sie einen linguistischen Bezug ein, den Foucault selbst so nie explizit formuliert, aber wohl gemeint hat.

Das Konzept der starken Strukturabhängigkeit von Diskursen ist eng verbunden mit dem Aspekt der Wissensgenerierung als diskursive Praktik. Diese Dualität erlaubt es, Struktur – verstanden als Ausdrucksmuster – und Wissensgenerierung im Foucault'schen Sinne sowie im Sinne der linguistischen Diskursanalyse als zentrales Merkmal aufzufassen.

Eine entsprechende Wissens-Diskurs-Analyse zielt vor allem darauf ab, den Kontext von Äußerungen präzise zu bestimmen, die Grenzen des Gesagten und Gemeinten zu fixieren. Foucault benennt die relevanten Faktoren:

> L'analyse de la pensée est toujours *allégorique* par rapport au discours qu'elle utilise. Sa question est infailliblement : qu'est-ce qui se disait donc dans ce qui était dit ? L'analyse du champ discursif est orientée tout autrement ; il s'agit de saisir l'énoncé dans l'étroitesse et la singularité de son événement ; de déterminer les conditions de son existence, d'en fixer au plus juste les limites, d'établir ses corrélations aux autres énoncés qui peuvent lui être liés, de montrer quelles autres formes d'énonciation il exclut. (Foucault 1969: 40)

Er zeigt in seinen Arbeiten und Vorträgen nie konkret, wie entsprechende Modelle konstruiert und wie solche Systeme verwendet werden könnten. Die heutigen Möglichkeiten erlauben es jedoch, sich dem zu nähern und den nur beschriebenen Ansatz in eine funktionierende wissenschaftliche Methode zu transformieren. Im Folgenden unternehmen wir eine erste Annäherung daran und schlagen systematische Annotation als wesentliches Element der Diskursanalyse und damit der Modellierung vor.

Durch korpuslinguistische Verfahren können diskurstypische Sprachgebrauchsmuster aufgedeckt werden. Durch Annotation ist es dann genau möglich, diese Muster zu explizieren und damit später wieder auffindbar zu machen, die das Wesen der Diskurse als geordnete Systeme maßgeblich bestimmen. Mit dem Konzept des Musters tritt die korpusbasierte Diskurslinguistik der Vagheit des Foucault'schen Ordnungsbegriffs entgegen und macht ihn für die diskurslinguistische wie diskurssemantische Praxis in der Tradition Busses anwendbar. Wenn es der Diskurslinguistik nämlich um „Regeln der Wissenkonstitution und -strukturierung" (Busse 1987: 233) geht, ist jene sprachliche Konstituierung von Wissen, welche durch die Annotation geschieht, als diskursiver Prozess zu bezeichnen. Die Annotation bildet hier die Möglichkeit, die Ordnung des Diskurses, verstanden als Muster des Sprachgebrauchs, zu identifizieren und daraus Schlüsse zu ziehen: Die Ordnung des Diskurses findet sich in der Annotation als Diskursmuster. In diesem Sinne kann selbst das Auffinden des Musterhaften durch automatisierte Annotation als digitaler Diskurs verstanden werden.

Das Aufspüren von Sprachgebrauchsmustern, die Annotation von außertextuellen Bezügen und deren Kategorisierung ist die Abstraktion von einem konkreten Diskurs hin zu einer Verallgemeinerung in Form von Diskurs-

mustern. Dieser Abstraktionsschritt wiederum ist ein wesentliches Element der Modellierung im Sinne Stachowiaks. Das Ziel ist die Konstruktion eines Modells des Phänomens Diskurs im linguistischen wie im soziologischen Sinne, das durch seine pragmatische Vereinfachung gegenüber dem Original formalisiert wird und damit die weitere maschinelle Analyse des Diskurses als maschinelle Bearbeitung seines Abbilds erlaubt.

Piotrowski (2019: 12) stellt fest, dass bereits Korpora selbst Modelle im Sinne von Stachowiak (1973) sind; sie sind also Modelle *von etwas* und *für einen bestimmten Zweck*. Er weist darauf hin, dass in der Korpuslinguistik Korpora erstellt werden, um eine Sprache zu untersuchen; in anderen Bereichen wie der Geschichtswissenschaft – und eben auch in der Diskursanalyse – dient ein Korpus dagegen als Modell einer *außersprachlichen* Wirklichkeit. Diesen Aspekt gilt es, dem Verhältnis von Korpus und Diskurs, wie es von Busse und Teubert (1994) dargelegt und bis heute in seinen Grundsätzen beibehalten wurde, aus unserer Sicht unbedingt zu ergänzen: Diskurse lassen sich als Modelle darstellen, die Modellierung von Diskursen entspricht der Analyse von Diskursen.

Diese Sicht von Diskursanalyse als Modellierung des Diskurses erlaubt eine fundierte Entgegnung auf die oben skizzierte Kritik: Das Explizieren der Modelleigenschaften als Abbildung unter Berücksichtig bestimmter Attribute für einen bestimmten Zweck ist gute wissenschaftliche Praxis, die reproduzierbar und erklärbar ist.

Die Analyse des Diskurses verstanden als Modellierung des Diskurses, wie wir sie vorschlagen, bedingt also Formalisierung und Explizitheit. Beide Elemente sind ebenso Voraussetzung für maschinelle Verarbeitung von Daten, also für digitalisierte Prozesse. Umgekehrt gilt: erst die Digitalisierung ermöglicht es, Modelle zu kreieren, die Diskurse umfassend abbilden und mittels informatischer Methoden analysiert werden können. Im Sinne der oben beschriebenen digitalen Transformation ist die Entwicklung der digitalen Diskursanalyse als formale und explizite Modellierung von Diskursen also tatsächlich eine Transformation, nicht nur die Umsetzung etablierter Prozesse in ein digitales Medium. Wir haben Annotation bereits als essentielles Element dieser Modellierung genannt; im folgenden Abschnitt erläutern wir dies näher.

5. Annotation

Wir haben oben bereits festgehalten, dass KWIC-Analysen einen schnellen Eindruck über die unmittelbare Umgebung der Fundstellen in allen Texten eines Korpus geben können. Dabei haben wir herausgestellt, dass dies den Diskurs im Sinne Harris' als *Struktur* und *Muster* innerhalb eines Textes über Satzgrenzen nicht abbildet. Somit muss auch die Nützlichkeit solcher Resultate für die Bestimmung eines Diskurses im Sinne Foucaults in Zweifel gezogen werden, da der Zusammenhang von Bedeutung und Verwendung sprachlicher Zeichen so nicht automatisch ermittelbar ist und unsichtbar bleibt. Dieser muss als intellektuelle Leistung der Betrachtenden entsprechender Resultate hinzugefügt werden oder bedarf vorgängiger tieferer Annotation.

Insofern wir mit Spitzmüller und Warnke (2011: 46) die Konstituierung von Wissen und Wirklichkeit, also von Erkenntnis, als sozialen und vor allem diskursiven Prozess begreifen, verstehen wir Annotation selbst zunächst als diskursiven Prozess, der das Explizieren von Kontext meint.

Annotation bezeichnet sowohl den *Prozess* des Anreicherns und Auszeichnens von Text wie auch diese zusätzlich hinzugefügten Informationen selbst, also das *Produkt* des Annotierens. Diese Sichtweise ebnet den Weg für die Etablierung des Verständnisses eines digitalen Diskurses als korpuslinguistische Annotationspraxis, d.h., die Analyse ist selbst ein Diskurs.

Betrachten wir zunächst Annotation als Prozess und zwar einerseits als (digitale) Untersuchungspraxis und dynamischen, womöglich kollaborativen, Akt des Erkenntnisgewinns und andererseits als systematisches, womöglich automatisiertes, Sichtbarmachen von impliziten Strukturen und Anreichern mit außertextlichen Informationen.

5.1. Annotation als Prozess (Annotieren)

Das Annotieren von Text ist etablierte Praxis in verschiedenen geisteswissenschaftlichen Feldern: Der Leser oder die Leserin hebt ‚wichtige' Wörter, Sätze oder Passagen hervor. Dies geschieht über eine Markierung direkt im

Text (etwa Unterstreichen) oder über Zeichen und textuelle Anmerkungen in Marginalien oder in einer zusätzlichen Ebene, die als Folie über den Text gelegt werden kann. Diese Hervorhebungen machen etwas explizit, das bereits vorhanden, aber nicht offensichtlich ist. Sie dienen einem späteren weiteren Leser als Fingerzeig oder erlauben der Leserin selbst bei einer nochmaligen Lektüre eine andere durch ihre eigene explizit gemachte Struktur und Leseführung beeinflusste und somit gerichtete Aufnahme von Informationen. Markierte und dadurch hervorgehobene Textteile sollen etwa später zusammengetragen, miteinander verglichen und interpretiert werden, um Muster zu erkennen oder eine gezielte Frage zu beantworten. Die Ebene der Annotation ist abhängig vom Forschungsinteresse, wie etwa Jacke und Gius (2016) für das Annotieren literarischer Texte zeigen. Das Annotieren trägt dazu bei, implizite Attribute eines Originals zu explizieren und so Modellierung zu ermöglichen.

Die Bedeutung von Markierungen wird bei Annotation für den Privatgebrauch oder ein wissenschaftliches ad-hoc-Bedürfnis oft nicht explizit und vorab definiert: Was auf welche Art hervorgehoben wird, ist der Leserin-Annotatorin klar – weil sie sich bestimmte Praktiken angewöhnt hat – oder erschließt sich durch konsistente Verwendung von Farben, Symbolen, etc. Markierungen können so nur für sich selbst stehen – etwa eine Unterstreichung bzw. die Tatsache, dass und was unterstrichen wurde – oder durch eine zusätzliche Information erweitert werden, etwa ein Symbol oder eine kurze Bemerkung.

Für die gezielte Lektüre, für die gemeinsame rezeptive Arbeit an Texten und Dokumenten, verstanden als Forschungsdaten, für die Bearbeitung größerer Mengen von Texten werden Richtlinien (Guidelines) und spezifische Vereinbarungen für die Verwendung getroffen oder mindestens im Verlauf des Annotierens entwickelt. Hier wird festgelegt, welche Phänomene wie gekennzeichnet, welche Information hinzugefügt werden. Existieren solche Konventionen nicht oder werden sie nicht offengelegt und zugänglich gemacht, sind die Resultate des Annotierens gesamthaft nicht nachvollziehbar und vor allem nicht reproduzierbar:

> Only if we take both sampling and annotation seriously and make all the decisions involved in them explicit we can perform rigorous experiments with reproducible results. (Lüdeling 2011:223)

Für Zweifelsfälle – soll eine Textstelle annotiert werden oder nicht, in welche Kategorie fällt diese, etc. – bieten Annotationsrichtlinien eine Grundlage, Entscheidungen zu treffen oder nachzuvollziehen. Auch hier: das Annotieren ist eine wesentliche Grundlage für eine Modellierung des Originals entsprechend bestimmter Forschungsfragen.

Das Anwenden und in noch stärkerem Maße das Aushandeln solcher Richtlinien ist bereits Teil wissenschaftlicher Auseinandersetzung und wesentlicher Aspekt epistemologischer Praxis (vgl. Lüdeling 2011, 2017). Bei diesem kollaborativen Prozess kann selbst von einer diskursiven Aushandlung von Wissenselementen gesprochen werden, welche mit den Textgegenständen der Annotation verknüpft ist.

Das kollaborative Annotieren birgt, wie Jacke und Gius (2016: 235) konstatieren, eine Möglichkeit, zur Überwindung des *subjectivity bias*. Indem mehrere Sichtweisen kombiniert werden und in die Annotation einfließen, können aus der „Pluralität von Perspektiven und Meinungen" (Jacke & Gius 2016: 235) überindividuelle Erkenntnisse generiert werden, welche wir insofern als diskursive Praxis auffassen können, als auch hier Aussagen unterschiedlicher Annotationsakteure gemeinsam betrachtet werden müssen, um dann zu Erkenntnis zu gelangen. Diese Erkenntnis muss als Wissen bzw. eine Wahrheit, verstanden werden, auf die man sich gemeinsam geeinigt hat. Damit ist diese Art von *digitalem Diskurs* sehr nahe beim Verständnis der Hermeneutik, wie es Caputo (2018) darlegt:

> In hermeneutics, we defend the idea that there are no pure facts. Behind every interpretation lies another interpretation. We never reach an understanding of anything that is *not* an interpretation. We can never peel away the layers to get to some pure, uninterpreted, naked facto of the matter. No matter how loudly you proclaim you are just sticking to the facts, you are only raising the volume of your own interpretation. In hermeneutics, I like to say, interpretation goes all the way down. (Caputo 2018: 4)

Der Vorgang des Annotierens ist ein expliziter Prozess: es wird jeweils eine *eindeutige* Entscheidung getroffen: etwas wird annotiert oder eben nicht. Die Möglichkeit einer nur vagen Auszeichnung existiert nicht. Unsicherheit dagegen ist über einen solchen Prozess sehr wohl abbildbar und damit auch nachvollziehbar, ebenso wie verschiedene Sichtweisen und Interpretationen der Annotationsrichtlinien. Verschiedene AnnotatorInnen können konkurrierende Markierungen der gleichen Ebene vornehmen. Lüdeling (2011) argumentiert, dass hier elektronische Werkzeuge einen entscheidenden Vorteil bieten im Vergleich zu Korpusarbeit im 20. Jahrhundert: konkurrierendes Annotieren auf verschiedenen Ebenen ist möglich, für spezifische Forschungsfragen können in der Explorations- und Interpretationsphase gezielt spezifische Annotationen verwendet werden (und andere ausgeblendet oder unterdrückt und damit ignoriert werden). Lüdeling et al. (2016) zeigen den praktischen Einsatz von Mehrebenenannotation, die mehrheitlich automatisch erfolgt.

Annotieren bildet hierbei sehr gut ab, was Caputo (2018) für die Hermeneutik beschreibt: Es gilt nicht die Idee, dass es so etwas wie *die reinen Fakten* gibt. Durch Annotationsebene um Annotationsebene, welche miteinander verglichen, diskutiert und in Beziehung gesetzt – d.h. diskursiv verhandelt – werden, kommt man dem Verständnis des Textes und seiner Aussagen auf die Spur – und zwar unabhängig von der eigenen Interpretation. Als durch diese Prozeduren ‚gesichertes' und in diesem Sinne als wahr anerkannt, wird nur, was von den meisten Annotations-Diskurs-Akteuren anerkannt wird.

Annotieren als wissenschaftliche und diskursive Praxis kann durchaus explorativ beginnen, wird mit zunehmender Korpusgröße und Reflexion des Annotierens jedoch zielgerichteter und nähert sich in Absicht und Ausgestaltung den Gegebenheiten der zentralen Forschungsfrage an. Während eine rein linguistische Annotation auf Ebene der Morphologie und Syntax unabhängig von Domäne, Textsorte und Informationsabsicht der AutorInnen erfolgen kann, ist eine Annotation auf textlinguistischer oder soziolinguistischer Ebene von anderen Faktoren beeinflusst.

Der zu berücksichtigende Kontext schließt einerseits die Entstehung von Texten ein und versucht so, autorInnenimmanente Faktoren zu explizieren. Zudem ist der Kontext auf RezipientInnenseite zu berücksichtigen: Wie wird

ein Text präsentiert (etwa auch: mit welchen anderen Beiträgen ist er in einem größeren Zusammenhang zu sehen auf einer Zeitungsseite, in einem Blog, etc.) und wahrgenommen (sind Reaktionen darauf in anderen Medien oder als explizite Kommentare bekannt)? Zu diesem Kontext gehören weiterhin Fragen der gegenseitigen Wahrnehmung von Äußerungen: Ist ein Text X eine direkte oder indirekte Reaktion auf einen Text Y oder sind beide höchstwahrscheinlich unabhängig voneinander entstanden? Solche Kontexte können rekonstruiert und damit über weitere Äußerungen und Diskursbeiträge in dritten Medien expliziert werden. Annotation erfordert als Ausgangspunkt etwas Manifestes, hier einen Text oder eine Sammlung von Texten, dem weitere Ebenen der Interpretation und Explizierung hinzugefügt werden. Je komplexer der Diskurs, je vielschichtiger die (möglichen) Beziehungen der Akteure und ihrer Äußerungen, umso schwieriger wird die Explizierung via Annotation. Aber nur durch Annotation werden solche Beziehungen und Prozesse überhaupt sichtbar und damit modellierbar im Sinne Stachowiaks.

Annotieren bedeutet also einen Wissens- und Erkenntniszuwachs durch das Explizieren selbst und findet in den kollaborativen Diskurshandlungen der AnnotatorInnen statt, welche als Diskursakteure durch ihre Annotation Zuschreibungen zu den Diskurselementen, den Äußerungen und Paraphrasen, in den Ausgangstexten vornehmen und welche sie dann diskursiv aushandeln. Nur diejenigen Annotationslabel (tags), die übereinstimmend, man spricht hier vom Inter-Annotator-Agreement, von der Mehrheit der AnnotatorInnen im Text angebracht werden und auf die sich in nachfolgenden Aushandlungsprozessen geeinigt werden kann, werden in Annotationskriterien zusammengefasst und als Richtlinien festgehalten. Diese sind wiederum von immenser Relevanz für die spätere Nutzung – etwa für die Annotation weiterer Texte oder um eine maschinelle Weiterverarbeitung zu ermöglichen. Digitale Werkzeuge, wie etwa CATMA (https://catma.de/) oder INCEpTION (https://inception-project.github.io/) bieten hierfür die entsprechenden technischen Voraussetzungen und Möglichkeiten zu annotieren, Werkzeuge wie ANNIS erlauben das Explorieren solcher Mehrebenenannotationen.

5.2. Automatisches vs. manuelles Annotieren

Maschinelle Annotation ist mittlerweile als ein zentrales Werkzeug für korpusbasierte linguistische Diskursanalysen zu bezeichnen. Korpuslinguistische Werkzeuge bieten Suchfunktionen an, die auf automatisch erzeugten Annotationsebenen basieren: etwa Wortarten (Part-of-Speech) und deren morphosyntaktische Eigenschaften, Lemmatisierung oder die Kennzeichnung und Typisierung von Eigennamen (Named Entity Recognition). Je nach Sprache, Domäne und Textsorte sind diese Bestimmungen recht zuverlässig. Mehrebenenannotation erlaubt hier auch konkurrierende Annotation unter Explizierung von Ambiguitiät oder Unterspezifikation oder durch verschiedene computerlinguistische Werkzeuge. Darüberhinausgehende Annotation etwa von Argumentstrukturen oder von *topics* ist wesentlich ungenauer, was in der weiterführenden Analyse als Interpretation berücksichtigt werden muss. In jedem Fall finden wir hier systematische Unsicherheit und Unterspezifikation.

Manuelles Annotieren wie in Abschnitt 5.1 gezeigt ist dagegen epistemgetrieben mit zunächst eher vagen Entscheidungskriterien. Kollaboratives Annotieren und konkurrierendes Annotieren bietet hier eine gute Möglichkeit, mehr als eine subjektiv-individuelle Perspektive auf einen Forschungsgegenstand, ein Wort, Mehrworteinheiten oder einen Text abzubilden. Soll daher die Annotation einem spezifischen Erkenntnisinteresse dienen, werden unterschiedliche Perspektiven diskutiert und die Ergebnisse in *Annotationsguidelines* festgehalten. Diese enthalten beispielsweise Definitionen über verwendete Kategorien, sie geben Hinweise zur Operationalisierbarkeit und dienen allen Annotierenden als gemeinsame Grundlage für ihre kollaborative Arbeit. Das Erarbeiten solcher Richtlinien ist also bereits ein Schritt des Modellierens.

Gleichzeitig ist dieser Prozess ein Aushandeln von Wissen: nur Paraphrasen, welche im Diskurs von der Mehrheit akzeptiert werden, werden zu gültiger Erkenntnis und damit zum Wissen in einer Diskursgemeinschaft. Die individuelle Annotation, der individuelle Gedanke zählt nicht, sondern nur das Gesagte, was immer wieder paraphrasiert, permutiert wird, wird intertextuell mit bereits Gesagtem verknüpft. In diesem Sinne kann es in einem Dis-

kurs nie neues Wissen geben – und dies gilt auch für das Annotationswissen. Was an dieser Stelle gleichsam neu hinzutritt, sind die Informationen, welche über das Diskursobjekt in den Diskurs eingebracht werden. In der manuellen Annotation entsteht ein Metadiskurs, welcher bei künftigen Diskursanalysen – als Teil des Diskurses – wieder mitgedacht werden muss.

Für die Praxis des manuellen Annotierens sind also diskursive Handlungen seitens der Akteure zu konstatieren. Welches sind also die Resultate des Annotierens?

5.3. Annotation als Produkt

Auch wenn die Korpuslinguistik *Text* als *Folge von sprachlichen Zeichen* versteht und bearbeitet, die sich zu Wörtern, Phrasen und Sätzen verbinden, werden Texte in der Regel als *Dokumente* produziert und rezipiert. Dokumente als explizit strukturierte Texte mit einer klaren Abgrenzung zu anderen Texten bzw. zu nicht zugehörigem Text führen LeserInnen und unterstützen so AutorInnen in deren Absicht, eine Information zu übermitteln. Typographische Gestaltung, Hierarchie von Abschnitten, bereits von AutorInnen hervorgehobene Passagen (entsprechend üblicher Markierung etwa von sprachlichen Beispielen, der Einführung neuer Begriffe oder über typographische Hervorhebung für emphasierte Wörter oder Phrasen um Topikalisierung und Fokus zu verdeutlichen) sind für die menschliche Rezeption von Texten essentiell und steuern damit den Diskurs. Die Behandlung von Dokumenten als Text, also als Folge sprachlicher Zeichen, wie üblicherweise in der Korpuslinguistik, die auf lexikographische oder syntaktische Phänomene abzielt, entfernt solche diskurssteuernden Elemente, die durch AutorInnen oder Publikationsorgan gezielt eingesetzt wurden. Daher müssen diese in Metadaten und als Strukturebene festgehalten werden.

Grundlegende syntaktische Strukturen sind bereits auf der Oberfläche von Texten und Dokumenten sichtbar und durch Interpunktion markiert, etwa Satzgrenzen. Weitere linguistische Strukturen wie Phrasen und die Beziehungen der darin enthaltenen Wörter sind dem Leser, der Leserin nicht

direkt zugänglich aber via generellem Sprachverständnis bewusst. Diese können ebenfalls explizit gemacht werden.

Solche textinhärenten Strukturen werden üblicherweise durch automatische Annotation erkannt und markiert, die Markierung und Klassifizierung erfolgt in der Regel als zusätzliche Ebene, die technisch ausgewertet, als Filter verwendet und etwa für Visualisierung nutzbar gemacht werden kann. Manuelle epistemologische Annotationen werden ebenfalls als zusätzliche Ebenen festgehalten, für die dann die gleichen Möglichkeiten zur Verwendung als Filter oder für Visualisierung genutzt werden können.

Annotationsebenen, die außertextliche Beziehungen festhalten, also der Kontext der Autor/-in und Rezipient/-in, bilden Diskurs im Foucault'schen Sinn ab. Auch diese werden als eigenständige und möglicherweise konkurrierende Ebenen festgehalten – Äußerungen mögen von einer Autorin anders ‚gemeint' sein, als sie von einer Leserschaft wahrgenommen werden – und sind damit wie rein automatische morphosyntaktische Annotation Prozessen und Werkzeugen zum Suchen, Filtern und Visualisieren zugänglich.

6. Das Digitale digitaler Diskursanalysen

Wie wir gezeigt haben, ist für die Definition des Begriffs *digitale Diskursanalyse* nicht die Natur der Untersuchungsgegenstände oder der Ort ihrer Entstehung entscheidend. Die definitorische Schärfung des Konzepts der *digitalen Diskursanalyse* in Bezug auf ihre Funktion als Forschungsparadigma bedingt die Bestimmung des spezifisch *Digitalen*. Unser Verständnis der digitalen Diskursanalyse als Modellierung von Diskurs im Sinne einer Explizierung und Formalisierung hat Konsequenzen für die Forschungspraxis.

Die digitale Diskurslinguistik als Modellierung von Diskursen, mit dem Ziel einer Abbildbarkeit als Repräsentation unter Berücksichtigung spezifischer Attribute zu einem bestimmten Zweck, lässt sich mithin als Erweiterung und Operationalisierung der Diskurslinguistik nach Foucault begreifen: Sowohl die Auswahl der für einen Diskurs relevanten Texte in einem Korpus wie auch die Annotation von textinhärenten Merkmalen und Elementen zusammen mit außertextlichen Kontexten zur Entstehung und Rezeption

dieser Texte, die den Diskurs bilden, erlauben es, das Musterhafte der Sprache im Sinne des Foucault'schen Ordnungsbegriffes zu identifizieren und mit Blick auf die Konstitution von Wissen und Erkenntnis zu interpretieren. Sie ist damit Voraussetzung für transdisziplinäre Anwendungen zur Analyse öffentlicher Kommunikation (Dreesen & Stücheli-Herlach 2019).

Modellierung bedingt Explizierung und Formalisierung. Die begrifflich wie strukturell unklaren ‚Ordnungen' in Foucaults Theorie können als linguistische Diskursanalyse via Annotation ihre intertextuellen und semantischen Verweissysteme offenlegen. Gleichzeitig ist die Praxis der manuellen epistemologischen Annotation ein diskursives Aushandeln von Wissen und Erkenntnis durch die AnnotatorInnen selbst, insofern die wissenschaftliche Praxis des Annotierens den Regeln der diskursiven und sprachlichen Konstruktion von Wissen und Erkenntnis im Sinne der linguistischen und wissenssoziologischen Diskursanalyse folgt. Der Diskurs der AnnotatorInnen ist als digitaler Diskurs aufzufassen, wenn wir den Diskurs im Sinne Foucaults und der Tradition der epistemologisch interessierten Diskurslinguistik auffassen; hierfür sprechen seine Systematik sowie die Rolle der Akteure. Somit sind alle Annotationsdiskurse als Modellierung von Diskursen aufzufassen.

So werden die bei Foucault noch unsichtbar gebliebenen Ordnungen des Diskurses durch das Annotieren sichtbar gemacht. Sie werden nicht nur in ihrer Semantik, sondern auch in ihrer Grammatik nachvollziehbar und reproduzierbar. Annotation ist explizit und eindeutig. Sie kann manuell oder automatisiert erfolgen und ist eine Erweiterung des ursprünglichen Textes, eine zusätzliche Ebene, eine Folie, durch die der Text rezipiert werden kann. Durch Annotation kann das Musterhafte der Sprache im Sinne des Foucault'schen Ordnungsbegriffes identifiziert und interpretiert werden. In diesen Strukturen und im Annotieren selbst spiegelt sich der Diskurs wider, aus dem Wissen und Bedeutung rekonstruierbar werden. Annotieren, verstanden als korpuslinguistische Praxis, ist somit Brennglas und Schnittpunkt bei der Engführung von Korpuslinguistik und Diskursanalyse und selbst digitaler Diskurs. Sie folgt selbst den Regeln der diskursiven und sprachlichen Konstruktion von Wissen und Erkenntnis.

Die Resultate konstituieren ein explizites Modell des Diskurses: Sie liegen als folgerichtige, eindeutige und explizite Daten im Sinne einer formalen Beschreibung von textuellen und außertextuellen Fakten vor und erlauben so Abstraktion und Modellierung, um den damit beschriebenen Diskurs zu analysieren und zu interpretieren. Das Digitale digitaler Diskuranalysen besteht also in der Modellierung als Sichtbarmachen der Ordnung des Diskurses wie auch in der daraus resultierenden Systematisierbarkeit und Reproduzierbarkeit der Ergebnisse.

7. Literatur

Bachelard, Suzanne (1979): Quelques aspects historiques des notion de modèle et de justification des modèles. In Pierre Delattre & Michel Thellier (Hrsg.), *Élaboration et justification des modèles*: Vol. I, 3–19. Paris: Maloine.

Bar-Hillel, Yehoshua (1962): Theoretical Aspects of the Mechanization of Literature Searching. In Walter Hoffman (Hrsg.), *Digitale Informationswandler*, 406–443. Braunschweig: Friedrich Vieweg & Sohn.

Bender, Michael, Thomas Kollatz & Andrea Rapp (2018): Objekte im digitalen Diskurs – epistemologische Zugänge zu Objekten durch Digitalisierung und diskursive Einbindung in virtuelle Forschungsumgebungen und -infrastrukturen. In Markus Hilgert, Kerstin P. Hofmann & Henrike Simon (Hrsg.), *Objektepistemologien. Zur Vermessung eines transdisziplinären Forschungsraums* (Berlin Studies of the Ancient World 59), 107–132. Berlin: Edition Topoi.

Berger, Peter L. & Thomas Luckmann ([1966] 2009): *Die gesellschaftliche Konstruktion der Wirklichkeit. Eine Theorie der Wissenssoziologie*. 19. Aufl. Frankfurt a. M.: Fischer.

Blei, David M. (2012): Probabilistic topic models. *Communications of the ACM* 55 (4), 77–84.

Brennen, J. Scott & Daniel Kreiss (2016): Digitalization, In Klaus B. Jensen, Eric W. Rothenbuhler, Jefferson D. Pooley & Robert T. Craig (Hrsg.), *The International Encyclopedia of Communication Theory and Philosophy*, 1–11. Hoboken, NJ: Wiley.

Bubenhofer, Noah (2018): Diskurslinguistik und Korpora. In Ingo H. Warnke (Hrsg.), *Handbuch Diskurs, Sprachwissen*, 208–241. Berlin, New York: De Gruyter.

Busse, Dietrich (1987): *Historische Semantik. Analyse eines Programms*. Stuttgart: Klett-Cotta.

Busse, Dietrich (1988): Kommunikatives Handeln als sprachtheoretisches Grundmodell der historischen Semantik. In: Ludwig Jäger (Hrsg.), *Zur historischen Seman-*

tik des deutschen Gefühlswortschatzes. Aspekte, Probleme und Beispiele seiner lexikographischen Erfassung, 247–272. Aachen: Rader Verlag.

Busse, Dietrich & Wolfgang Teubert (1994): Ist Diskurs ein sprachwissenschaftliches Objekt? Zur Methodenfrage der historischen Semantik. In Dietrich Busse, Fritz Hermanns & Wolfgang Teubert (Hrsg.), *Begriffsgeschichte und Diskursgeschichte. Methodenfragen und Forschungsergebnisse der historischen Semantik*, 10–28. Opladen: Westdeutscher Verlag.

Caputo, John D. (2018): *Hermeneutics. Facts and Interpretation in the Age of Information*. Penguin UK.

Chapco-Wade, Colleen (2018): *Digitization, Digitalization, and Digital Transformation: What's the Difference?*, Online: https://medium.com/@colleenchapco/ digitization-digitalization-and-digital-transformation-whats-the-difference-eff1d002fbdf (06.06.2020).

Crystal, David (2011): *Internet Linguistics. A Student Guide*. London, New York: Routledge.

Deleuze, Gilles (2018): *Foucault: Lecture 3, 05 November 1985*. Purdue University Research Repository. https://doi.org/10.4231/R76T0JVM.

Dreesen, Philipp & Peter Stücheli-Herlach (2019): Diskurslinguistik in Anwendung: ein transdisziplinäres Forschungsdesign für korpuszentrierte Analysen zu öffentlicher Kommunikation. *Zeitschrift für Diskursforschung* 7 (2), 123–162.

Foucault, Michel (1966*): Les mots et les choses*. Paris: Gallimard.

Foucault, Michel (1969): *L'archéologie du savoir*. Paris: Gallimard.

Foucault, Michel (1972): *L'ordre du discours: leçon inaugurale au Collège de France prononcée le 2 décembre 1970*. Paris: Gallimard.

Fraas, Claudia, Stefan Meier & Christian Pentzold (Hrsg.) (2013): *Online-Diskurse. Theorien und Methoden transmedialer Online-Diskursforschung* (Neue Schriften zur Online-Forschung 10). Köln: Herbert von Halem.

Gladkij, Aleksej V. & Igor A. Mel'čuk (1973): *Elemente der mathematischen Linguistik*. Berlin: VEB Deutscher Verlag der Wissenschaften (Autorisierte Übersetzung aus dem Russischen, Original 1969, Moskau: Nauka).

Gloning, Thomas (2019): Die Kölner Silvesternacht als Diskursthema. Vortrag vom 08.11.2019. *6. Netzwerktreffen des DFG-Netzwerks ‚Diskurse – digital'*. Universität Mannheim.

Gredel, Eva (2018): *Digitale Diskurse und Wikipedia: Wie das Social Web Interaktion im digitalen Zeitalter verwandelt* (Dialoge 1). Tübingen: Narr Attempto.

Gredel, Eva & Ruth M. Mell (2018): Die narrative Dimension digitaler Diskurse: Zum Einsatz digitaler Tools und Ressourcen für die Analyse internetbasierter Kommunikation am Beispiel der Wikipedia. *Zeitschrift für Literaturwissenschaft und Linguistik* 48 (2), 331–355.

Harris, Zellig S. (1952). Discourse analysis. *Language* 28 (1), 1–30.

Heinemann, Wolfgang & Dieter Viehweger (1991): *Textlinguistik. Eine Einführung*. Tübingen: Niemeyer.

Hodge, Bob (2017): Discourse Analysis. In Tom Bartlett & Gerard O'Grady (Hrsg.), *The Routledge Handbook of Systemic Functional Linguistics. Routledge Handbooks in Linguistics*, 520–532. Taylor & Francis.

Jacke, Janina & Evelyn Gius (2016): Kollaboratives Annotieren literarischer Texte. *DHd 2016 Modellierung, Vernetzung, Visualisierung: die Digital Humanities als fächerübergreifendes Forschungsparadigma: Konferenzabstracts*, 240–43.

Janich, Nina (2008): Intertextualität und Text(sorten)vernetzung. In Nina Janich (Hrsg.), *Textlinguistik. 15 Einführungen*. 177–198. Tübingen: Narr.

Kilgarriff, Adam (2001): Comparing Corpora. *International Journal of Corpus Linguistics* 6 (1), 97–133.

Kilgarriff, Adam (2007): Googleology is Bad Science. *Computational Linguistics* 33 (1), 147–151.

Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel (2014): The Sketch Engine: Ten Years on. *Lexicography* 1 (1), 7–36.

Krasselt, Julia, Philipp Dreesen, Matthias Fluor, Cerstin Mahlow, Klaus Rothenhäusler & Maren Runte (2020): Swiss-AL: A Multilingual Swiss Web Corpus for Applied Linguistics. *Proceedings of the 12th Language Resources and Evaluation Conference, European Language Resources Association*, 4138–4144.

Kupietz, Marc, Harald Lüngen, Pawel Kamocki & Andreas Witt (2018): The German Reference Corpus DeReKo: New Developments – New Opportunities. *Proceedings of the 11th Language Resources and Evaluation Conference, European Language Resources Association*, 4353–4360.

Lüdeling, Anke (2011): Corpora in Linguistics: Sampling and Annotation. In Karl Grandin (Hrsg.), *Going Digital. Evolutionary and Revolutionary Aspects of Digitization* (Nobel Symposium 147), 220–243. New York: Science History Publications.

Lüdeling, Anke, Julia Ritz, Manfred Stede & Amir Zeldes (2016): Corpus Linguistics. In Caroline Fery & Shinishiro Ishihara (Hrsg.), *OUP Handbook of Information Structure*, 599–617. Oxford: Oxford University Press.

Lüdeling, Anke (2017): Variationistische Korpusstudien. In Marek Konopka & Angelika Wöllstein (Hrsg.), *Grammatische Variation. Empirische Zugänge und theoretische Modellierung. IDS Jahrbuch 2016*, 129–144. Berlin: de Gruyter.

Mahlow, Cerstin & Andreas Hediger (2019): Digital Transformation in Higher Education—Buzzword or Opportunity? *eLearn Magazine* 2019 (5), Article 13 https://doi.org/10.1145/3329488/3331171.

Marx, Konstanze & Georg Weidacher (2014): *Internetlinguistik. Ein Lehr- und Arbeitsbuch*. Tübingen: Narr.

Mell, Ruth M. (2015): *Vernunft, Mündigkeit, Agitation. Eine diskurslinguistische Untersuchung zur Generierung und Strukturierung von Wissen über das Konzept 'Aufklärung 1968'*. Bremen: Hempen.

Mell, Ruth M. & Eva Gredel (2021): Diskurse – digital: Theorien, Methoden, Fallstudien Ein von der DFG gefördertes Netzwerk zur Analyse digitaler Diskurse (2016–2020). *Zeitschrift für Diskursforschung* 6 (1), 103–106.

Pêcheux, Michel (1975): *Language, Semantics and Ideology*. Springer.

Piotrowski, Michael (2019): Historical Models and Serial Sources. *Journal of European Periodical Studies* 4 (1), 8–18.

Rammerstorfer, Lydia (2019): *Digitalität. Daten. Digital Humanities*. https://www.hsozkult.de/event/id/termine-40808 (28.08.2019).

Shannon, Claude Elwood & Warren Weaver (1949): *A Mathematical Theory of Communication (PDF)*. Illinois: University of Illinois Press.

Sommer, Vivien, Claudia Fraas, Stefan Meier & Christian Pentzold (2013): Qualitative Online-Diskursanalyse. Werkstattbericht eines Mixed-Method-Ansatzes zur Analyse multimodaler Deutungsmuster. In Claudia Fraas, Stefan Meier & Christian Pentzold (Hrsg.), *Online-Diskurse. Theorien und Methoden transmedialer Diskursforschung* (Neue Schriften zur Online-Forschung 10), 258–284 Köln: Herbert von Halem.

Spitzmüller, Jürgen & Ingo H. Warnke (2011): *Diskurslinguistik. Eine Einführung in Theorien und Methoden der transtextuellen Sprachanalyse*. Berlin, Boston: de Gruyter.

Stachowiak, Herbert (1973): *Allgemeine Modelltheorie*. Wien, New York: Springer.

Teubert, Wolfgang (2010): *Meaning, Discourse and Society*. Cambridge: Cambridge University Press.

Teubert, Wolfgang (2013): Die Wirklichkeit des Diskurses. In Dietrich Busse & Wolfgang Teubert (Hrsg.), *Linguistische Diskursanalyse: neue Perspektiven* (Reihe Interdisziplinäre Diskursforschung), 55–145. Wiesbaden: Springer VS.

Warnke, Ingo H. (Hrsg.) (2007*): Diskurslinguistik nach Foucault. Theorie und Gegenstände*. Berlin, New York: de Gruyter.

Wiener, Norbert (1948): *Cybernetics or Control and Communication in the Animal and the Machine*. Massachusetts: MIT Press.