

Janja Polajnar, Joachim Scharloth, Tanja Škerlavaj

Intertextualität und (digitale) Diskurse

1 Einleitung

Schon in frühen Formulierungen des Programms einer linguistischen Diskursanalyse wurden Diskurse forschungspraktisch als thematisch bestimmte Korpora von Einzeltexten definiert, deren Besonderheit darin besteht, dass die in ihnen versammelten Texte intertextuell aufeinander Bezug nehmen (vgl. Teubert 1998: 148). Intertextualität wurde so zu einem Bestimmungsmerkmal von Diskursen. In der Forschungspraxis wurde freilich wegen der Vielzahl an Korpus-texten und der aufwändigen qualitativen Methoden zumeist auf den expliziten Nachweis intertextueller Bezüge verzichtet. Die intertextuellen Bezugnahmen der Korpus-texte blieb eher ein Interpretament, das mit dem Hinweis auf die thematische Verknüpftheit notdürftig mit Evidenz versehen wurde.

Diskurse in digitalen Medien ermöglichen nun einen neuen Zugang zur intertextuellen Verknüpftheit und zwar auf zwei Ebenen. Zum einen auf der Ebene der medialen Konstitutionsbedingungen: Diskurse in digitalen Medien enthalten häufig technisch erzeugte Spuren intertextueller Referenz und die Analyse dieser Spuren kann für die Rekonstruktion der Genese eines Diskurses fruchtbar gemacht werden. Zum anderen auf der methodischen Ebene: Datengeleitete maschinelle Verfahren zur Aufdeckung latenter semantischer Strukturen und zur Erkennung von Paraphrasen eröffnen auch für große Textmengen die Möglichkeit der Identifizierung intertextueller Bezugnahmen.

Nach einer kurzen Einführung in unterschiedliche Konzeptualisierungen von Intertextualität (Abschnitt 2) nehmen die folgenden Teile (in Abschnitt 3) unterschiedliche Ebenen und Formen von Intertextualität in den Blick, die für die Analyse von Diskursen im Medium des Digitalen, darüber hinaus aber auch in medial hybriden Diskursen relevant sein können. Diese liegen sowohl im Bereich der latenten semantischen Bezüge (diskurssemantische Grundfiguren, Topoi, Paraphrasen) sowie der expliziten wie impliziten, teils medial-technischen Verknüpfungen (Zitat, Link, Mention, Aggregation etc.). Letztere werden anhand von Beispielen aus öffentlichen Diskursen erläutert. Ansätze der digitalen Diskursanalyse zur Aufdeckung latenter Textbezüge werden anhand zweier datengeleiteter Methoden illustriert (Abschnitt 4).

2 Intertextualität: Deskriptive vs. ontologische Bestimmung

Die Vielzahl der Ansätze zur Bestimmung des Phänomens der Intertextualität lässt sich grob in deskriptive und ontologische Theorien gliedern.

Deskriptive Intertextualitätstheorien fassen nur solche Beziehungen zwischen Texten als Intertextualitätsphänomene, die sich als intentionale Referenzen eines Autors/ einer Autorin auf einen vorgängigen Text deuten lassen. Für die Verfechter dieser Theorierichtung sind Zitat, Übersetzung und Anspielung, Parodie und Travestie sowie der Kommentar, die Interpretation und die Kritik typische Formen deskriptiver Intertextualität (vgl. etwa Stierle 1984: 147). Abhängig vom Einfluss des intertextuellen Bezugs auf die Gestaltung des beziehnehmenden Textes unterscheiden deskriptive Intertextualitätstheorien eine Produktions- und Rezeptionsintertextualität. Während bei letzterer zwar die Sinnkonstitution latent beeinflusst, die Oberfläche jedoch nicht maßgeblich mitgeprägt wird (etwa durch Verweise), wird bei der ersteren die Textoberfläche von der intentionalen Bezugnahme mitstrukturiert.

Ontologische Intertextualitätstheorien modellieren intertextuelle Bezüge auf einer abstrakteren Ebene. Sie gehen von der Einsicht aus, dass sich in sprachliche Ausdrücke „Gebrauchswerte“ eingeschrieben haben, d. h. Bedeutungsdimensionen, die sich aus den historischen Verwendungsweisen in konkreten Kontexten ergeben, wie es Michail Bachtin in seiner Theorie der Dialogizität formulierte:

Jedes Mitglied eines Sprechkollektivs findet das Wort nicht als ein neutrales Wort der Sprache vor, das von fremden Bestrebungen und Bewertungen frei ist, dem keine fremden Stimmen innewohnen. Nein, es empfängt das Wort von einer fremden Stimme, angefüllt mit dieser fremden Stimme. In seinen Kontext kommt das Wort aus einem anderen Kontext, durchwirkt von fremden Sinngebungen. Sein eigener Gedanke findet das Wort bereits besiedelt. (Bachtin 1990: 130)

Sprechen ist damit nicht ein intentionaler Akt autonomer Sinnproduktion im Medium der Sprache, sondern ein Dialog mit früheren Verwendungen des sprachlichen Materials und den durch sie mitgeprägten Verstehenshorizonten der Adressatinnen und Adressaten. Auf Basis dieser Überlegungen entwickelte die französische Literaturwissenschaftlerin Julia Kristeva ihren Begriff der Intertextualität. So wie bei Bachtin Sprache bereits sozial geprägt ist, so ist Intertextualität die Seinsform von Texten und nicht nur die intendierte Bezugnahme auf einen Prätext: „[...] jeder Text baut sich als Mosaik von Zitaten auf, jeder Text ist Absorption und Transformation eines anderen Textes“ (Kristeva 1978: 391).

Der ontologische Intertextualitätsbegriff fügt sich dem Programm der post-strukturalen Literaturwissenschaft insofern ein, als er die Vorstellung von Subjektautonomie und auktorialer Intentionalität als Quelle der Bedeutung eines Textes obsolet macht. Dies sind Eigenschaften, die er mit dem diskursanalytischen Textbegriff teilt, weswegen der ontologische Intertextualitätsbegriff für die Diskursanalyse von besonderer Relevanz ist.

Die Diskursanalyse hat unterschiedliche Kategorien entwickelt, mit der sie ontologisch-intertextuelle Bezüge analytisch erschließbar machen will, von denen im Folgenden die wichtigsten prototypisch vorgestellt werden sollen. Die mit ihnen erfassten intertextuellen Bezüge sind teilweise ein Effekt des Medienwandels (etwa im Fall von Link oder Aggregat), teilweise sind sie durch Methoden der digitalen Diskursanalyse erst maschinell erschließbar geworden (diskurssemantische Grundfiguren, Paraphrasen).

3 Dimensionen von Intertextualität in der (digitalen) Diskursanalyse

3.1 Diskurssemantische Grundfiguren

Für die Analyse der grundlegenden, kategorial-epistemischen Dimensionen der Intertextualität hat Dietrich Busse die Kategorie der diskurssemantischen Grundfiguren vorgeschlagen. Diese bestimmt er einerseits hinsichtlich ihrer Funktion für die Konstitution eines Diskurses, andererseits auch hinsichtlich ihrer Repräsentation in Texten. Diskurssemantische Grundfiguren sind nicht an einen Diskurs gebunden, sondern können in unterschiedlichen Diskursen auftreten. In funktionaler Perspektive ordnen sie

textinhaltliche Elemente, steuern u. U. ihr Auftreten an bestimmten Punkten des Diskurses, bestimmen eine innere Struktur des Diskurses, die nicht mit der thematischen Struktur der Texte, in denen sie auftauchen, identisch sein muß, und bilden ein Raster, das selbst wieder als Grundstruktur diskursübergreifender epistemischer Zusammenhänge wirksam werden kann.
(Busse 1997: 20)

Bei ihnen kann es sich um Präsuppositionen im Sinn der linguistischen Pragmatik oder um durch Inferenzen zu erschließende Teile des Implizierten handeln, sie können als semantische Merkmale auftreten oder auch zur lexikalischen Bedeutung von Wörtern gehören (vgl. Busse 2003: 30). Diskurssemantische Grundfiguren sind Teil der epistemisch-kognitiven Grundausstattung der Textproduzenten (vgl. Busse 1997: 19) und stiften so intertextuelle Homologien. Beispiele sind

etwa die Unterscheidung „natürlich“ vs. „geköstelt“ in der Ästhetik des 18. Jahrhunderts oder „kommerziell“ vs. „authentisch“ in populärkulturellen Diskursen bis in die Gegenwart. Als Wahrnehmungsschemata sind sie grundlegende Elemente der Organisation des umfassenden semantischen Systems, in dem Aussagen überhaupt erst sinnhaft werden (vgl. Scharloth 2005). Sie lassen sich je nach Granularität mit Verfahren der distributionellen Semantik (vgl. Abschnitt 4) in einer digitalen Diskursanalyse erschließen.

3.2 Topoi

In digitalen massenmedialen Diskursen beteiligen sich Akteure (noch mehr als im Fall „klassischer“ Diskurse) an den Diskussionen oft aktiv, indem sie beispielsweise in Form von Kommentaren zu Online-Zeitungsartikeln oder zu Posts in Foren oder auf sozialen Netzwerken für oder gegen das im Ausgangstext Genannte argumentieren. Dies tun sie gezielt und strategisch, „um von ihren Positionen zu überzeugen, diese im Meinungskampf durchzusetzen und ihre Ziele zu erreichen“ (Spieß 2011: 524). Im Rahmen einer Topos-Analyse, die auf Überlegungen von Kienpointner (1992), Kopperschmidt (1989) und Wengeler (2003) zurückführt, geht man davon aus, dass diese Grundpositionen mit ihren Argumenten die sog. wesentlichen Argumentationsstränge darstellen, „auf die innerhalb des Diskurses immer wieder rekurriert wird“ (Spieß 2011: 465). Solche komplexen Argumentationen manifestieren sich sprachlich in Form wiederkehrender Muster, der sog. Argumentationstopoi.

Die Toposanalyse gehört nach Spitzmüller und Warnke (2011: 191) zu den zentralen und erfolgreichen Konzepten der transtextuell (aber auch – wie weiter unten gezeigt wird – der transnational sowie der multimodal) orientierten Linguistik. Da in digitalen Diskursen aufgrund des Merkmals der Partizipation (vgl. Androutsopoulos 2010: 421) also noch mehr argumentiert wird als in „klassischen“ Diskursen und da sich in solchen argumentativen Mustern gesellschaftliches Wissen einer Zeit verdichtet (vgl. Ziem 2005: 321), stellt die Toposanalyse einen wichtigen Bestandteil der Analyse digitaler Diskurse dar.

Bei Topoi als Ausdruck der Intertextualität handelt es sich somit, ähnlich wie bei diskurssemantischen Grundfiguren, nicht um explizite intertextuelle Bezüge auf einen Referenztext im Sinne von Gattungsreferenzen oder Einzeltextreferenzen (vgl. Pfister 1985 sowie Holthuis 1993); vielmehr sind Topoi – im Sinne einer ontologischen Intertextualitätsauffassung – als intertextuelle Bezüge auf einer abstrakteren Ebene zu verstehen. Solche Argumentationsmuster,

die zwar oft kontextspezifisch¹ sind, können nämlich aus anderen Kontexten übernommen werden, d. h. sie können in unterschiedlichen Diskursen vorkommen. So kann z. B. der Topos vom ökonomischen Nutzen bzw. der sog. Ökonomie-Topos sowohl im Migrationsdiskurs als auch beispielsweise im Bologna-Diskurs vorkommen. Dieser Topos lautet: *Weil eine Handlung unter wirtschaftlichen Gründen einen Nutzen bzw. Schaden erbringt, soll sie ausgeführt/ nicht ausgeführt werden* (vgl. Wengeler 2003: 316 sowie Polajnar & Škerlavaj 2018a: 203). So argumentiert z. B. der folgende User/ die folgende Userin in seinem/ ihrem Kommentar zu einem Online-Artikel über Zuwanderer-Mythen mit dem Topos vom ökonomischen Nutzen gegen die Zuwanderung:

(1) **Digitalis.purpurea**¹

#16 – 18. August 2015, 18:08 Uhr

Das ist natürlich stichhaltig ... nicht

Nur mal um zwei Beispiele rauszugreifen:

5. Die Flüchtlinge nehmen uns die Arbeitsplätze weg

Wir müssen in Deutschland sogar weniger werden! Es werden gut 20 Millionen Arbeitsplätze wegfallen: [http://www.ingenieur.de/Them ...](http://www.ingenieur.de/Them...) .

Genau in Bezug auf Arbeitsplätze muss die die Bevölkerung schrumpfen. Und komme bitte keiner damit, dass da Ärzte, Krankenschwestern und Ingenieure einwandern. Die meisten sind noch chancenloser als die Bildungsversager hier, wegen der mangelnden Deutschkenntnisse.

8. Wer vom Balkan kommt, ist sowieso ein Wirtschaftsflüchtling

Die Auskünfte von Lobbyisten wie Pro Asyl sind natürlich auch sehr vertrauenswürdig ...

Ich könnte den ganzen Artikel auf die Art und Weise zerlegen, da reiht sich eine Propaganda an die nächste.

(Kommentar zum Artikel *Haben wir wirklich keinen Platz mehr in Deutschland?*, Zeit Online, 18. August 2015)

Die Autorin/ der Autor des obigen Kommentars ist also der Meinung, neue Zuwanderer*innen könnten den Deutschen noch diejenigen Arbeitsplätze wegnehmen, die es noch gibt, und unterstützt ihre/ seine Argumentation mit einem Zitat bzw. der Angabe eines Hyperlinks, der die Leserin/ den Leser zu einem Online-Artikel über die vierte industrielle Revolution weiterleitet.²

Während der Topos vom wirtschaftlichen Nutzen (als Argumentationsmuster pro bzw. contra Zuwanderung) bereits seit dem Anfang der 60er Jahre im deut-

¹ Das bedeutet, sie sind an bestimmte Diskurse und damit an bestimmte Themen gebunden (vgl. Spieß 2011: 2019).

² Mehr zu Hyperlinks siehe Abschnitt 3.1 (Explizite Verknüpfungen).

schen Migrationsdiskurs beobachtet wird (vgl. Wengeler 2003: 346f.), lässt sich dieser spätestens seit dem Jahr 1999³ auch im sogenannten Bologna-Diskurs konstatieren. Im folgenden Kommentar zu einem Online-Artikel über Verantwortliche für die Mängel der Bologna-Reform argumentiert die Userin/ der User (unter anderem) mit dem Topos vom ökonomischen Nutzen gegen die Reform – ihrer/ seiner Meinung nach sind Bachelorabsolventinnen und -absolventen „halbfertige Akademiker“, die auf dem Arbeitsmarkt „schlechte Karten haben“:

(2) ofx83

#2 – 24. November 2009, 19:59 Uhr

Der Sparzwang ist Schuld an der Misere

Der Bachelor wurde als Möglichkeit genutzt, die Kosten für Bildung zu senken.

Statt wie bisher in 8–9 Semestern, soll ein Student jetzt in 6 Semestern die Hochschulen verlassen.

Folge: Oberflächlicher Stoff, keine Zeit für Praktika und/oder Nebenjobs. Eine wertige akademische Ausbildung ist in 3 Jahren nicht vermittelbar. Den Master möchte man nicht mehr für alle anbieten um Geld zu sparen. Die Folge sind halbfertige Akademiker die auf dem Arbeitsmarkt schlechte Karten haben. Wir haben in Deutschland das System der dualen Berufsausbildung, einen Bachelor braucht in Deutschland kein Mensch.

Der Master sollte weiterhin der Regelabschluss sein, das kostet jedoch Geld (warum auch in Bildung investieren?). Man hat Geld für Autoabwrackprämien, Shuttleservices für Politiker, Banken, Militär etc. doch für Bildung scheint kein Platz zu sein.

Master für Alle oder Bachelor verlängern (8–9 Semester) ...

(Kommentar zum Artikel *Studentenproteste: Uni-Rektoren geben Ländern die Schuld an Bachelor-Chaos*, Zeit Online, 24. November 2009)

Da ein und derselbe Topos (wie hier etwa der Topos vom ökonomischen Nutzen) in verschiedenen Diskursen verwendet werden kann, kann hierbei von Intertextualität im weiteren Sinne gesprochen werden.

Ein weiterer interessanter Aspekt ist, dass Topoi nicht nur aus anderen Kontexten übernommen werden und somit in verschiedenen Diskursen vorkommen können, vielmehr sind diese oft auch als sprach- und kulturübergreifend zu verstehen. Nach Gür-Şeker (2012: 305) verfügen Diskurse nämlich über die Eigenschaft

³ Im Juni 1999 wurde in Bologna von 29 europäischen Bildungsministern eine Erklärung unterschrieben, in der die Schaffung eines *Europäischen Hochschulraums* bzw. eines international wettbewerbsfähigen *Europas des Wissens* bis zum Jahr 2010 vereinbart wurde. Ob die Umsetzung der Bologna-Erklärung sinnvoll erfolgte, ist seitdem Gegenstand zahlreicher Diskussionen in den öffentlichen Medien, die den sogenannten Bologna-Diskurs konstituieren (vgl. Polajnar & Škerlavaj 2018a und 2018b).

„Transnationalisierung“, die unter anderem auch durch Diskursphänomene wie etwa Topoi konstituiert werden kann.⁴ So wird beispielsweise mit dem Topos vom ökonomischen Nutzen nicht nur im Deutschen, sondern auch in anderen Sprachen für oder gegen den Bologna-Prozess und seine Ziele sowie Maßnahmen argumentiert.⁵ Topoi in solchen global geführten Diskursen (wie etwa im Migrationsdiskurs, Bologna-Diskurs oder Corona-Diskurs) können zudem für kultur-kontrastive Analysen fruchtbar gemacht werden: Indem thematisch und zeitlich ähnliche Diskurse aus zwei oder mehreren Sprach- und Kulturgemeinschaften z. B. im Hinblick auf Argumentationsmuster miteinander verglichen werden, können wichtige Konvergenzen und Divergenzen in Denkstilen verschiedener Kulturen konstatiert werden.

Schließlich soll darauf hingewiesen werden, dass die aus bestimmten Topoi schöpfenden Argumente nicht an die semiotische Ressource „Sprache“ gebunden sind. Vielmehr können Topoi „auch durch Argumente anderer Zeichenmodalitäten, z. B. in Form von Bildargumenten, realisiert werden“ (Klug 2016: 180). So argumentieren beispielsweise im aktuellen Corona-Diskurs verschiedene Akteure (z. B. Wissenschaftler*innen, Politiker*innen, Verschwörungstheoretiker*innen usw.) mit sprachlichen, sprachlich-visuellen oder rein visuellen Argumenten für oder gegen die Covid-19-Pandemie und die damit verbundenen Schutzmaßnahmen. Im folgenden Facebook-Post der slowenischen Gruppe „Corona – največja prevara stoletja“ („Corona – der größte Betrug des Jahrhunderts“, T.Š.) argumentieren beispielsweise die Corona-Skeptiker*innen mit einer Gegenüberstellung von zwei betitelten Bildern dafür, dass es sich im Fall von Covid-19 um keine richtige Pandemie handelt (vgl. Abb. 1):

⁴ Mehr zum Phänomen der Transnationalisierung und die sie prägenden Konstituenten siehe im Abschnitt 2.4 (Zitate). Vgl. dazu auch den Beitrag von Gredel & Mell in diesem Band.

⁵ Für eine ausführlichere Auseinandersetzung mit Topoi im deutschen und slowenischen Bologna-Diskurs siehe Polajnar & Škerlavaj 2018a, Polajnar & Škerlavaj 2018b, Polajnar 2017 sowie Škerlavaj 2017.



Abb. 1: Visuelle Argumentation; FB-Gruppe „Corona – največja prevara stoletja“ („Corona – der größte Betrug des Jahrhunderts“).

Die Kern-Botschaft über die nicht existierende Pandemie wird im obigen Post durch ein visuell-verbales Argument gestützt (vgl. Klug 2016: 181), das seine Schlusskraft aus dem sogenannten Topos der Widerspruchsfreiheit (Wengeler 2003: 320f.) gewinnt. Dieser Topos, der zu den Gegensatzschemata gehört, und zwar zu den Versionen, in denen mit Gegensätzen argumentiert wird, „um einen Widerspruch nachzuweisen“ (Kienpointner 1996: 116, vgl. auch Wengeler 2003: 320), lautet: *Weil in der Vergangenheit eine Gruppe oder Person in einer Situation/ einem Zustand bestimmte Handlungen ausgeführt hat bzw. bestimmte Handlungen stattgefunden sind, sollten in vergleichbaren Situationen/ Zuständen auch aktuell diese Handlungen ausgeführt werden bzw. stattfinden.*

Im obigen Facebook-Post, der slowenischen Verschwörungstheoretiker*innen zuzuordnen ist, wird dem Bild mit der Überschrift *Pandemic*, das in seiner apokalyptischen Bildsprache auf eine Pandemie als ein Ereignis verweist, bei dem so viele Menschen den Tod finden, dass man sie nicht mehr einzeln bestatten kann, ein Bild aus einer vermeintlichen Gegenwart gegenübergestellt, auf

welchem medizinische Fachangestellte mit Mundschutzmasken tanzen und dabei offensichtlich viel Spaß haben. Die Überschrift des zweiten Bildes, im Gegensatz zum oberen Bild in bunten Farben und nicht schwarzweiß dargestellt, lautet ebenfalls „*Pandemic*“ – das Wort *Pandemic* ist allerdings in Anführungszeichen gesetzt, was darauf hindeuten soll, dass die Bedeutung des Wortes nicht wörtlich zu nehmen ist. Die zu entschlüsselnde Aussage des Posts ergibt sich also aus dem Zusammenspiel zwischen beiden Texten und Bildern und lässt sich folgendermaßen ausformulieren: „*In einer wahren Pandemie sterben und nicht tanzen die Menschen, daher ist Covid-19 keine Pandemie*“.⁶

Solche visuellen Argumentationen kommen in digitalen Diskursen wie etwa in Memes, GIFs und anderen viralen multimodalen Kommunikaten auf sozialen Netzwerken, aber auch in Kommentaren zu Online-Zeitungsartikeln sowie in Foren- und Blogbeiträgen besonders häufig vor.

3.3 Paraphrasen

Im Gegensatz zu diskurssemantischen Grundfiguren und Topoi werden Paraphrasen gemeinhin als Phänomen der Einzeltextreferenz bestimmt. Ganz gleich, ob man Paraphrasen als intentionale und zumindest implizit markierte Reformulierungen im Sinne deskriptiver Intertextualität (Pfister 1985: 23) oder als rein sprachliche Homologien konzeptualisiert (Steyer 1997), so kommen die Definitionen des Paraphrasenbegriffs doch darin überein, dass der Bezug auf die Transformierbarkeit von Teilen des Referenztextes in die Paraphrase ein konstitutives Merkmal von Paraphrasen ist (vgl. hierzu ausführlicher Scharloth et al. 2019).

Kathrin Steyer etwa fordert für die linguistische Intertextualitätsforschung eine Fokussierung auf die sprachproduktbezogene Intertextualität. Eine analytische Beschreibung müsse „auch ohne Instrumentarien der Psychologie, kognitivistischer Theorien oder gar der Literaturwissenschaft geleistet werden“ können (Steyer 1997: 86). Daher müsse das Augenmerk der Distribution textueller Muster in größeren Korpora und ihrer diachronen Entwicklung gelten. Der Musterbegriff impliziert dabei, dass Referenztext und Paraphrase auf dieselbe abstrakte Repräsentation zurückgeführt werden können und diese abstrakte Repräsentation als das Vehikel ihrer wechselseitigen Transformation gedacht werden kann.

⁶ Dabei ist davon auszugehen, dass die Botschaft des Posts unter der Voraussetzung bestimmter Allgemeinkenntnisse bei der Rezipientin/ dem Rezipienten (über die Covid-19-Pandemie) auch nur anhand der Bilder, d. h. ohne den sprachlichen Teil verständlich wäre.

Verfahren der Identifizierung von Paraphrasenkandidaten kommen etwa bei der Plagiatserkennung, der Alignierung von Korpora und der Textklassifikation zum Einsatz. In diskursanalytischer Sicht richtet sich die Paraphrasenforschung auf ein den Referenztext (und mit ihm die Paraphrase) strukturierendes System von Topoi, Aussagen, Motiven oder Codes. Anders als in der traditionellen, vor allem mit qualitativen Methoden arbeitenden Diskursanalyse wird dieses System allerdings nicht als bloßes Postulat universeller Intertextualität gedacht. Die Methoden der digitalen Diskursanalyse (vgl. Abschnitt 4) zielen vielmehr darauf ab, Diskurse, verstanden als Texte strukturierende Systeme, in den sprachlichen Oberflächen- oder Tiefenstrukturen eben dieser Texte empirisch nachweisbar zu machen.

3.4 Zitate

Zitate stellen das prototypische Beispiel der textoberflächenstrukturellen Einzeltextreferenz (vgl. Pfister 1985, Janich 2019) bzw. der Text-Text-Beziehungen (vgl. Holthuis 1993, Fix 2000) dar. Darunter sind vollständige, syntaktisch und lexikalisch identische Übernahmen von Referenztext(teil)en zu verstehen, die im Phänotext⁷ entweder (un)vollständig markiert oder unmarkiert vorkommen (vgl. Janich 2019). Die Funktion von Einzeltextreferenzen ist im Allgemeinen darin zu sehen, einen „sachbezogenen Bezug zu Vortexten“ (z. B. eine Online-Rezension zum Buch) oder „Texten-in-Relation“ (z. B. Kommentarforum zu einem Online-Zeitungstext) herzustellen, um dadurch die Bedeutung des Phänotextes zu erweitern. Die gewählten Beispiele zeigen eindrucksvoll, dass durch Zitate heute oft Verknüpfungen zwischen Kommunikationsangeboten im Offline- und Online-Bereich, d. h. intermedial, hergestellt werden. Zudem stehen in digitalen Medien und Diskursen neben etablierten Zitatformen neue technisch-mediale Repräsentationen von Zitaten (z. B. zitierte Links und Hashtags) zur Verfügung, die im folgenden Kapitel (3.5.1) näher betrachtet werden. Hierbei wird Intertextualität neben Multimodalität⁸ und Heteroglossie zu den „charakteristischen Diskursprozessen des Web 2.0“ (Androutsopoulos 2010: 431) gezählt. Im Folgenden soll der Fokus auf einige zentrale Aspekte von Zitaten in digitalen Diskursen gerichtet werden, Multimoda-

⁷ Im Hinblick auf die terminologische Vielfalt in der literatur- und sprachwissenschaftlichen Intertextualitätsforschung sind in Anlehnung an Janich (2019: 178, 182) *Phänotexte* diejenigen Texte, die sich auf andere Texte beziehen, und *Referenztexte* dagegen Texte, auf die Bezug genommen wird.

⁸ Zur Multimodalität digitaler Diskurse siehe auch den Beitrag von Bender, Mell & Wildfeuer in diesem Band.

lität, Multilingualität, Interaktivität und Komplexität, die allerdings nicht nur im digitalen Raum vorzufinden sind.

Aus diskurslinguistischer Sicht ist zunächst die diskurskonstitutive Funktion von Zitaten hervorzuheben, denn Zitate verknüpfen einzelne Diskursfragmente miteinander und explizieren so verschiedene mediale Thematisierungsorte dieser Diskursfragmente im Offline- und Online-Bereich. Durch Zitate in Phänotexten eines thematisch orientierten Diskurses wird also angezeigt, „welche Referenztexte mit in die transtextuelle Analyse einbezogen werden müssen, um sie (be-)deutbar bzw. verstehbar zu machen“ (Klug 2018: 115). Für Diskurse ist folglich „ein dichtes Netz intertextueller Bezugnahmen charakteristisch“ (Klug 2018: 109), das, wie bereits erwähnt, nicht an den Grenzen des Digitalen aufhört, denn Diskurse sind heute „mehrheitlich medial hybrid organisiert“ (vgl. den Beitrag von Gloning in diesem Band). Hierbei wird der digitale Raum mit seinen Zitaten aus dem Online- und Offline-Bereich nicht nur auf ein Medium für die Anschlusskommunikation reduziert, sondern spielt als eigenständiges Diskursmedium eine Rolle, denn auch seine Text(teil)e werden in den Traditionsmedien aufgegriffen. Deshalb ist bei der Beschreibung und Klassifizierung von Zitaten des Digitalen wichtig, den Ort der Referenztexte zu berücksichtigen. Obwohl Zitate in diskurslinguistischen Analysen relativ selten ins Zentrum des Forschungsinteresses rücken und der Fokus meist auf der Beschreibung semantischer Beziehungen (Metaphern, Topoi, Frames u. a.) liegt (vgl. ebd.), kommt ihnen im Zusammenhang mit Textketten und Textnetzen eine wichtige Rolle zu (vgl. Girnth 1996; Schröter 2017; Klug 2018). Klug (2018) zeigt anhand von Bild-Zitaten und weiteren Bezügen auf zwei Fotografien von Aylan Kurdis (das im September 2015 aufgefundene tote Kind am Strand in Bodrum), welche Schlüsselstellung Referenzbilder innerhalb eines hybriden Diskurses (hier des Flüchtlingsdiskurses) einnehmen können: Bilder rücken ins Zentrum des Netzes „intra- und intermodaler Bezüge“ (Klug 2018: 120–127) und können einen Diskurs initiieren oder wesentlich vorantreiben. Zitierte Bilder, Videos usw. werden im digitalen Raum nicht nur gebraucht, um Texte zu ergänzen, sondern können diese ersetzen, wodurch sie Sprachhandlungscharakter erhalten (vgl. Androutsopoulos 2010: 434).

Referenztext(teil)e, -bilder und -musikstücke, die in digitalen Diskursen zitiert werden, können zudem Ausgangspunkt für (kultur-)kontrastive Diskursanalysen darstellen (vgl. Gredel 2018), denn Zitate beschränken sich in Diskursen des Digitalen oft nicht auf Einzelsprachen. Durch Zitate werden intertextuelle Bezüge nicht nur zu lokalen und nationalen, sondern auch zu globalen Referenztexten hergestellt (vgl. Androutsopoulos 2010: 431). Gür-Şeker (2012: 305) führt in diesem Zusammenhang das Konzept der „Transnationalisierung“ von Diskursen ein, die auch in digitalen Diskursen intertextuell durch Verlinkungen bzw. „Verweise auf fremdsprachige Quellen bzw. Medien [...] durch die Verwendung von Zi-

tatwörtern oder zitierter Formulierungen fremdsprachiger Diskurse“ (Gür-Şeker 2012: 307) aber auch durch fremdsprachige Hashtags etabliert wird. Transnationalisierung wie auch Multilingualität spielen für digitale Plattformen wie etwa die Wikipedia eine zentrale Rolle (vgl. den Beitrag von Gredel & Mell in diesem Band), indem Wikipedia-Autorinnen und -Autoren beispielsweise über Bildzitate (Abb. 2) Bezüge zwischen den Sprachversionen der Online-Eycklopädie herstellen. Folgende Bildschirmaufnahme einer Bildinventar-Gegenüberstellung unterschiedlicher Sprachversionen eines Wikipedia-Artikels (Abb. 2), ermittelt mit dem Tool „Wikipedia Cross-Lingual Image Analysis“, suggeriert die Bedeutung von Bildzitate[n] für transnationale Diskurse.



Abb. 2: Bildinventar-Gegenüberstellung am Beispiel deutscher (de), englischer (en) und spanischer (es) Sprachversionen des Wikipedia-Artikels „Bologna-Prozess“.

Im digitalen Raum können Diskurse zudem durch interaktionsorientiertes Diskutieren in Foren, Kommentarforen unter Online-Zeitungsartikeln oder YouTube-Videos, auf Twitter usw. konstituiert werden, wo interaktionales Zitieren eine wesentliche Rolle für eine gelungene Kommunikation und somit die Fortführung des Diskurses spielt. Hierbei sind Texte nicht mehr nur unidirektional aufeinander bezogen, sondern kommunizieren „auf vielfältige Weise wechselseitig miteinander“ (Berghofer 2009: 71). Gray behandelt beispielsweise die Intertextualitätsbezüge in Kommentarforen im Rahmen eines „voll interaktive[n] Intertextualitätsmodell[s]“ (Gray 2006: 20 ff.). In Postings eines Kommentarforums unter Zeitungsartikeln sind nämlich neben Zitaten aus dem kommentierten Online-Artikel oder einer dritten Quelle relevante Zitate aus früheren Postings vorzufinden, die bewertet und kommentiert werden (vgl. Polajnar 2019). Auf diese Weise können mehrere Stellen in einem Referenztext bzw. -Posting oder auch in mehreren Referenztexten bzw. -Postings kommentiert werden, wodurch nicht nur sehr genaue, sondern auch sehr dichte Bezüge zwischen einzelnen Postings entstehen. Beim Zitieren bereits kommentier-

ter Referenztexte können so komplexe Zitierstrukturen zustande kommen (Zitat im Zitat), die zur Explizierung der Referenzbezüge und zur Themenentwicklung beitragen (vgl. Polajnar 2019). Durch das Einbinden von umfangreichen Zitaten wird an exponierten Stellen von Postings zudem die Struktur der Phänopostings visuell stark beeinflusst und somit der Intertextualitätsgrad erhöht (textstrukturierende Funktion). Beispielhaft für komplexe Zitierstrukturen sind zudem Retweets, die an sich direkte Zitat-Texte darstellen, zudem aber oft noch einen im Originaltweet zitierten Hyperlink und/ oder Hashtag enthalten.

3.5 Technisch-mediale Repräsentationen von Intertextualität in Onlinediskursen

In digitalen Diskursen lassen sich aufgrund technisch-medialer Affordanzen (vgl. Pentzold, Fraas & Meier 2013 sowie den Beitrag von Bender, Mell & Wildfeuer zu Affordanzen in diesem Band) zum einen explizite Verknüpfungen (z. B. Hyperlinks, Hashtags und der @-Operator) und zum anderen implizite Verknüpfungen (wie Newsfeeds und andere Formen der Aggregation) vorfinden, die im Folgenden als Erscheinungsformen der Intertextualität näher betrachtet werden.

3.5.1 Explizite Verknüpfungen

Bei expliziten Verlinkungen, die in der computervermittelten Kommunikation typischerweise durch Hyperlinks und Hashtags realisiert werden, handelt es sich entweder um Zitate (z. B. zitierte Hyperlinks, Hashtags, @-Adressierung) oder indizierende Bezugnahmen durch verlinkte Zeichenketten. Dang-Anh, Einspänner und Thimm (2013: 140f.) zählen #, @ und http:// zu Kommunikationsoperatoren, die je spezifische operative und kommunikative Funktionalitäten erfüllen, welche allerdings „Aspekte eines gemeinsamen Vollzugs“ darstellen.

Im Hinblick auf Hyperlinks können intertextuelle Verlinkungen durch zitierte Hyperlinks entstehen oder aber es werden Zeichenketten im Phänotext mit einem externen oder internen Link⁹ unterlegt und auf diese Weise direkt mit dem verstehensrelevanten Referenztext verbunden. Im letzten Fall spricht

⁹ Marx und Weidacher (2014: 184–186) verweisen darauf, dass bei Texten im Internet „die Grenzen zwischen intra- und intertextuellen Verweisen nicht mehr in jedem Fall eindeutig zu ziehen“ sind, denn durch die intertextuellen Hyperlinks werden bei der Rezeption die Texte eng miteinander verzahnt; die Unterschiede zwischen verlinkten Modulen innerhalb eines Textes und zwischen verschiedenen Texten verschwimmen.

Klug (2018: 120–121) von „indizierenden intertextuellen Bezugnahmen“. Im Phänotext selbst dienen Hyperlinks oft der argumentativen Expansion und zugleich der Textentlastung, denn auf diese Weise wird der Phänotext nicht durch längere Zitate unterbrochen. Auf der transtextuellen Ebene konstituieren sie ein Konstrukt explizit erkennbarer, diskursrelevanter Vernetzungen. Goman und Gür-Şeker (2020: 137f.) erarbeiten anhand von thematischen Hyperlinks in themengebundenen Online-Artikeln eine 3-Ebenen-Analyse von Textsorten-in-Online-Vernetzung. Mit der Bestimmung von Link-Strukturen auf 3 Ebenen können auf jeder Ebene Textsortennetze zu Ausgangstexten identifiziert werden, die ein diskursives Gesamtkonstrukt bilden (Goman & Gür-Şeker 2020: 139). Der Link-Ansatz von Goman und Gür-Şeker ermöglicht nicht nur die verlinkten Online-Texte textsortenspezifisch zu kategorisieren und ihr Vorkommen im Zeitverlauf quantitativ zu erfassen, sondern kann auch „das Ensemble von Textsorten eines festgelegten Untersuchungszeitraums“ aber auch Diskurses sichtbar machen (Goman & Gür-Şeker 2020: 142). Allerdings ist die Extraktion und Dokumentation von multimodalen Korpora mit aufbewahrten Linkstrukturen mit erheblichen methodischen Herausforderungen verbunden. Goman und Gür-Şeker schlagen vor, neben Bildschirmaufnahmen auch Videoaufnahmen analysierter Objekte zu machen, die dann mit Tools wie Maxqda annotiert und analysiert werden können. Solche umfangreichen Datenmengen werfen die Fragen des Forschungsdatenmanagements, vor allem der Langzeitarchivierung und der nachhaltigen Nutzung von Forschungsdaten auf.

Hashtags, durch ein Rautezeichen (#) eingeleitete Zeichenketten, dienen auf der technisch operativen Ebene der expliziten Verlinkung von zunächst für sich stehenden Beiträgen auf Instagram oder Twitter, die mit dem gleichen Hashtag versehen wurden. Für digitale Diskurse ist es relevant, dass durch die Vorgabe von Hashtags vonseiten eines diskursrelevanten Einzelakteurs oder einer diskursrelevanten Einzelakteurin (z. B. Eventveranstalter*in oder Argumentationslinie von Politikerinnen und Politikern usw.) oder durch Aushandlung von Hashtags durch User*innen digitale Diskurse initiiert, etabliert und konstituiert werden (vgl. Dang-Anh, Einspänner & Thimm 2013). Auf diese Weise erhöhen User*innen aber auch die Sichtbarkeit eigener Beiträge für andere User*innen, die nach aktuellen diskursrelevanten Hashtags suchen (z. B. Trending Topics bei Twitter), um Diskurse zu verfolgen. Mit getaggten Schlagwörtern werden also nicht nur themenspezifische Postings, sondern auch deren Verfasser*innen eruiert, was für die Akteursanalyse von Relevanz ist (vgl. den Beitrag von Schnick zu Näheeffekten in diesem Band). Die zitierten Hashtags lassen nämlich Rückschlüsse auf die Akteurinnen und Akteure zu, die diese Hashtags zitieren und durch ihre individuelle Perspektive anreichern. Da ein Hashtag der Sammlung aller gleichgetaggten Beiträge zu einem diskursrelevanten Thema

dient, wird durch ein Hashtag das Diskursthema symbolisiert. Bei manchen Diskursthemen werden mehrere Hashtags vorgeschlagen, die neben dem Hauptdiskurs auch weitere Themenstränge abdecken und verschiedene Aspekte eines Themas behandeln. Über Hashtags werden Diskurse also konstituiert, strukturiert sowie identifiziert. In einem Beitrag werden am Ende oft mehrerer Hashtags angehäuft, wodurch der Tweet oder der Instagram-Post zu anderen Hashtags Bezug aufnimmt. Auf diese Weise werden in einer globalisierten Welt häufig transnationale, sprachübergreifende Diskurse konstituiert; so betrachten Goman und Gür-Şeker Hashtaganhäufungen als ein „Panoramafenster für transnationale Analysen“¹⁰ von digitalen Diskursen. Betrachtet man Hashtags wie hier als Zitate, die User*innen entweder extraponieren oder in die syntaktische Struktur eigener Beiträge einbetten können, um auf die Referenzbeiträge zu verweisen, so sind bei der Analyse ihre unterschiedlichen kommunikativen Funktionen zu berücksichtigen, z. B. argumentatierende, kommentierende, vorschlagwortende Funktion oder thematische Konkretisierung (vgl. Dang-Anh, Einspänner & Thimm 2013).

Eine weitere Möglichkeit explizite Verlinkungen herzustellen erfolgt mittels des @-Zeichens, das aus der E-Mail-Kommunikation remediatisiert wurde. Auf der Nachrichtenplattform Twitter wird durch @ + Accountname beispielsweise der Adressat durch den Sprecher zunächst genannt, zugleich „wird per Verlinkung auf ihn gezeigt“ (Dang-Anh 2019: 137). Eine ähnliche Funktion erfüllen im Allgemeinen auch @-replies, denn in beiden Fällen werden dadurch soziale Beziehungen hergestellt (Dang-Anh 2019: 138). Neben der explizit ausgewiesenen personen- bzw. accountgebundenen Adressierung ist es zudem möglich, einen Account per @ im Verlauf des Postings nur zu erwähnen (sog. „@-mention“) (vgl. Dang-Anh 2019: 139). Letztere haben einen höheren Grad an Öffentlichkeit als @-Adressierungen oder @-replies, denn sie werden nicht nur „in den Timelines von Abonnenten beider Interaktanten“ von @-Postings angezeigt (Dang-Anh 2019: 139). In digitalen Diskursen ist die Relevanz von Verlinkungen per @ vor allem im Hinblick auf die Akteursperspektive zu sehen, denn dadurch werden zwei Formen der personalen Referenz realisiert, Adressierung und Erwähnung, die auf diskursrelevante Akteure verweisen.

¹⁰ PowerPoint zum Vortrag „Kontrastive Diskurslinguistik und methodische Perspektivierungen auf Online-Diskurse“ am 02.11.2020 im Rahmen der Ringvorlesung Diskursdynamiken an der Philosophischen Fakultät der Universität Ljubljana.

3.5.2 Implizite Verknüpfungen

Im Gegensatz zu Links und Hashtags verläuft die Verknüpfung durch **Aggregatoren** implizit, indem die digitalen Inhalte (z. B. Nachrichten, Bilder, Filme oder Musik) beispielsweise durch eine Software zu einer Liste thematisch ähnlicher (oder auf eine andere Art und Weise miteinander verbundener) Inhalte verknüpft werden. Das Sammeln und die Zusammenstellung von Inhalten kann maschinell (anhand bestimmter Algorithmen) oder manuell durch verschiedene Bearbeitungsprogramme erfolgen. Als Beispiele für Nachrichten-Aggregatoren sind etwa Google News oder Yahoo Nachrichten zu nennen, als Aggregator kann aber auch Twitter dienen (allerdings handelt es sich hier – anders als bei sogenannten **Newsfeeds** – um explizite Verknüpfungen durch Hashtags). Darüber hinaus gibt es verschiedene Aggregatoren im Bereich der Blog- und Foren-Kommunikation, der Rezensionen oder verschiedener Dienstleistung.

Einige der am meisten verbreiteten Formate, anhand von welchen das Sammeln von Inhalten in Nachrichten-Aggregatoren (Newsfeeds) erfolgt, sind beispielsweise RSS („Really Simple Syndication“ bzw. „Rich Site Summary“) und Atom (vgl. dazu Wittenbrink 2005). Solche Newsfeeds dienten ursprünglich als Wegweiser zu HTML-Seiten, denn sie erlauben es, „die Headlines und Inhalte einer Seite als Teaser in andere Seiten einzubauen“ (Wittenbrink 2005). Heutzutage ist die Hauptfunktion von Aggregatoren, die Internetnutzer*innen über die wichtigsten Neuigkeiten auf einem bestimmten Gebiet kurz und knapp zu informieren (bzw. im Fall der Dienstleistungen: diese anzubieten, für sie zu werben) und den Usern und Userinnen zugleich Zeit zu ersparen. Im Unterschied zu Medienformaten wie etwa Newsletter, die abonniert werden können und beispielsweise per E-Mail gleichzeitig an verschiedene Rezipienten/ Rezipientinnen gesendet werden („Push-Medien“), kann der Empfänger/ die Empfängerin bei Newsfeeds als typischen „Pull-Medien“ selbst entscheiden, welche Nachrichten er / sie wann abrufen will.

Das folgende Beispiel veranschaulicht das Prinzip des Newsfeeds von Google (vgl. Abb. 3):

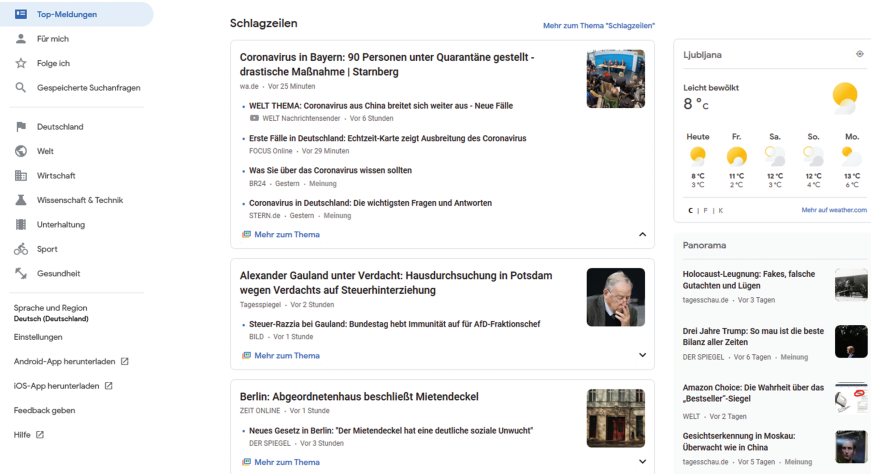


Abb. 3: Ausschnitt aus Google News.

In der Mitte befindet sich eine Zusammenstellung von Schlagzeilen mit den wichtigsten Nachrichten zu aktuellen Themen aus verschiedenen Online-Zeitungen bzw. -Portalen. Die Schlagzeilen fungieren gleichzeitig als Links – wenn man diese anklickt, gelangt man zum ursprünglichen Artikel. Auf der linken Seite befindet sich die Navigationsleiste mit verschiedenen Einstellungen und Themenbereichen („Deutschland“, „Welt“, „Wirtschaft“ etc.), die den Nutzern/ Nutzerinnen weitere „Filter“ im Hinblick auf die erwünschte Thematik der Nachrichten ermöglichen. Auf der rechten Seite können sich die Nutzer*innen über das aktuelle Wetter informieren bzw. sich im Bereich „Panorama“ einen Überblick über weitere, zum Teil ältere meist gelesene Beiträge verschaffen.

Dabei werden die Funktionen von Newsfeeds – wie die Funktionen digitaler Kommunikationsangebote schlechthin – nicht nur durch die Modalität „Sprache“ bzw. „geschriebener Text“ erfüllt, vielmehr sind solche Feeds hochgradig multimodal gestaltet: Ihr kommunikativer Sinn ergibt sich also aus dem Zusammenspiel von Schrift und anderen semiotischen Ressourcen wie etwa Layout, Typographie, Farbe und Bild.

4 Methoden

Im Folgenden sollen exemplarisch zwei methodische Zugänge zur Intertextualität vorgestellt werden, die in der computergestützten linguistischen Diskursanalyse

verwendet werden. Diese kommen insbesondere bei der Aufdeckung latenter Textbezüge zum Einsatz, etwa bei der Identifizierung von diskurssemantischen Grundfiguren (vgl. Abschnitt 3.1) und Paraphrasen (vgl. Abschnitt 3.3).

Sie sind so gewählt, dass sie in doppelter Hinsicht beispielhaft sein können, einerseits indem sie die zentralen intertextualitätstheoretischen Dimensionen der System- und der Einzeltextreferenz abdecken, zum anderen intransparente Black-box- und transparente Whitebox-Algorithmen als prototypische methodologische Konstellationen vorstellen.

4.1 N-Gramm-basierte Zugänge zur Intertextualität als Paraphrase

N-Gramme sind Einheiten, die aus n Elementen bestehen. Normalerweise werden n -Gramme als Folge von Wortformen operationalisiert. Im Rahmen einer n -Gramm-Analyse werden dann alle im Korpus vorkommenden n -Gramme berechnet, wobei vorher bestimmte Parameter wie Länge der Mehrworteinheit (aus zwei, drei oder mehr Wörtern bestehend) oder Spannweite (sind Lücken zwischen den Wörtern erlaubt?) festgelegt werden und Kennzahlen wie die Frequenz oder die Festigkeit der Verbindung der Elemente berechnet werden.

Kommt ein n -Gramm in Prä- und Folgetexten vor, ggf. in umgekehrter Reihenfolge oder anderen leichten Abwandlungen, dann kann dies als Paraphrase gedeutet werden. Eine nur Oberflächenstrukturen berücksichtigende n -Gramm-Analyse ist jedoch nur sehr bedingt für die Identifizierung von Paraphrasenkandidaten geeignet.

Um auch lexikalisch-semantische oder morphosyntaktische Abwandlungen des Referenztextes zu erfassen, müssen vielmehr einzelne oder alle Elemente eines n -Gramms durch alternative Repräsentationen (wie etwa interpretative linguistische Kategorien) ersetzbar sein (vgl. Scharloth et al. 2019). Dies können zum einen Elemente sein, die sich auf die Tokenebene beziehen und die Wortform funktional oder semantisch deuten (als Repräsentant einer Wortart oder als Teil einer semantischen Klasse). Zum anderen aber auch Elemente, die über die Tokenebene hinausgreifen, etwa das Tempus oder die Modalität einer Äußerung (direkte vs. indirekte Rede). Welche Elemente welcher interpretativer Dimensionen in die Analyse mit einbezogen werden, hängt einerseits von der jeweiligen Forschungsfrage ab, andererseits forschungspraktisch auch davon, welche Ressourcen für die Annotation des Korpus zur Verfügung stehen.

Ein Beispiel: Ein Satz wie „Angela Merkel ist die erfolgreichste Regierungschefin in Europa.“ kann beispielsweise maschinell mit interpretativen Kategorien auf folgenden Ebenen repräsentiert werden: Auf der Token-Ebene, also auf der Ebene von Wörtern und Satzzeichen, sind dies Informationen zum Lemma

	Angela	Merkel	ist	die	erfolgreichste	Regierungschefin	in	Europa	.
Lemma	Angela	Merkel	sein	d	erfolgreich	Regierungschefin	in	Europa	.
POS	NE	NE	VAFIN	ART	ADJA	NN	APPR	NE	\$.
NP	[Angela	Merkel]		[die	erfolgreichste	*Regierungschefin]	[in	*Europa]	
NER	[NE	Person]						NE Ort	
Hyperonym					gut, positiv	Regierungsbeamter, Verantwortlicher, Leiter		Erdteil, Kontinent	
Synonym					gelingen, sieghaft	Kanzler, Präsident		Abendland, Okzident	

und zur Wortart des jeweiligen Token (Tokenisierung, Lemmatisierung und POS-Annotation erfolgten mittels des Tree-Taggers), sowie unterschiedliche semantische Informationen, wie das Hyperonyme (aus der semantischen Taxonomie GermaNet) und Synonyme (aus Open Thesaurus). Tokenübergreifend stehen Informationen zu sog. *Named Entities* und ihrer Klassifikation als Ort, Organisation, Person und anderes zur Verfügung, ebenso wie eine grundlegende Gliederung des Satzes in Teil-Nominalphrasen. Die folgende Tabelle gibt einen (um der Übersichtlichkeit willen vereinfachten) Überblick über die für die jeweiligen Token zur Verfügung stehenden Informationen.

Komplexe n-Gramme lassen sich nun einerseits durch syntagmatische Parameter wie die Festlegung von Grenzen (Länge des n-Gramms) und die Möglichkeit von Auslassungen, andererseits durch paradigmatische Ersetzungen erzeugen. Ein durch die Operationen der Begrenzung, Auslassung und Ersetzung erzeugtes n-Gramm ist eine abstrakte Repräsentation des Referenztextes. Wie die n-Gramme jeweils gebildet werden sollen, wird durch die Festlegung der Parameter des Berechnungsalgorithmus gesteuert.

Wenn bei der Paraphrasensuche beispielsweise Personennamen durch die *named-entity*-Kategorie [NE Person] ersetzt werden, Verben als Wortform-Token erhalten bleiben, sämtliche Artikel und Adjektive ignoriert werden, alle Präpositionen durch ihren Part-of-Speech-Tag ersetzt werden und Substantive durch ihr Hyperonym, dann erhält man folgendes n-Gramm als Repräsentation des Referenztextes: [NE Person] ist [Regierungsbeamter, Verantwortlicher, Leiter] [APPR] [Erdeil, Kontinent]. Die Paraphrasensuche besteht nun darin, in potentiellen Folgetexten nach Textstellen zu suchen, deren abstrakte Repräsentation mit der des Referenztextes identisch ist. Dies wären beispielsweise Sätze wie „Putin [...] ist [...] Präsident [...] von Russland.“ oder „Alfons der Viertelvorzwölfte [...] ist [...] König [...] auf Lummerland.“ Das Beispiel zeigt, dass die gewählten Parameter den Algorithmus womöglich zu „gierig“ machen, d. h. dass zahlreiche Textstellen gefunden werden, die nur schwerlich als Paraphrasen gelten können.

Der Suchalgorithmus ist freilich hochgradig konfigurierbar und transparent. In jedem Fall ist anhand der abstrakten Repräsentation von Referenztext und Paraphrasenkandidat überprüfbar, aufgrund welcher sprachlichen Transformationen zwei Textstellen als hinreichend ähnlich aufgefasst werden können, um als Paraphrase gedeutet zu werden.

4.2 Word Embeddings und Vektorähnlichkeiten

Ein zweiter Ansatz zur Identifizierung von Paraphrasenkandidaten beruht auf computerlinguistischen Techniken der Modellierung von Sprache, die in der

distributionellen Semantik (vgl. Firth 1957) fußt. Wörter werden als Vektoren repräsentiert, wobei die einzelnen Werte dieser Vektoren sich aus der Distribution der sie umgebenden Wörter ergeben. Das semantische System einer Sprache oder auch nur eines Diskurses oder Diskursausschnittes wird somit als hochdimensionaler Vektorraum modelliert, in dem die Relationen zwischen den Einheiten als Vektordistanzen berechenbar sind. Ein Meilenstein für die breite Anwendbarkeit dieser Techniken ist der von Mikolov et al. (2013) entwickelte Word2Vec-Toolkit, der ein schnelleres Training der Modelle ermöglichte.

Intertextualität als semantische Ähnlichkeit von Texten oder Textabschnitten kann nun anhand der Transformierbarkeit des Referenztextes bzw. eines Teils daraus in einen potentiellen Folgetext modelliert werden. Für die Berechnung der Kosten der Transformation haben Kusner et al. (2015) die Word Mover's Distance (WMD) entwickelt. Sie ist definiert als die minimale kumulative Distanz, um die Wörter eines Referenztextes so zu überführen, dass der Phänotext entsteht.

Wendet man diese Methode nicht auf ganze Texte sondern auf Textsegmente an, so können in einem Phänotext unterschiedlichste intertextuelle Referenzen sichtbar gemacht werden, die als tiefensemantische, konzeptuelle oder topische Diskursstrukturen gedeutet werden können. Es lassen sich aber auch datengeleitet rekurrente Referenzen identifizieren, deren Co-Okkurrenz als Indiz für Diskursivität gelesen werden kann. So können die skizzierten Verfahren dazu dienen, Diskurse maschinell und datengeleitet zu identifizieren.

5 Fazit und Ausblick

Die vorangegangenen Abschnitte haben gezeigt, wie Intertextualität als konstitutives Merkmal von Diskursen sich im Zuge der digitalen Transformation von Diskurs und Diskursanalyse materialisiert und vom Interpretament zu einer methodisch einholbaren Kategorie diskursanalytischer Forschung wird. Zum einen ist die Digitalität der Texte eine entscheidende Voraussetzung dafür, dass Intertextualität mittels maschineller Verfahren im für Diskursanalysen nötigen Umfang methodisch erfasst werden kann. Andererseits prägen digitale Medien den Texten im Prozess ihrer Entstehung, aber auch in Anschlusskommunikationen Spuren ein, die zumindest teilweise als intertextuelle Verknüpfungen im Sinn der Diskursanalyse gedeutet werden können. Dass beide Dimensionen insofern miteinander verknüpft sind, als Plattformen die skizzierten datengeleiteten Methoden in Aggregations- oder Vorschlagsalgorithmen zur Verknüpfung digitaler Kommunikate einsetzen, ist ein Beleg dafür, dass digitaler Diskurs und digitale Diskursanalyse nicht unabhängig voneinander gedacht werden dürfen.

Literatur

- Androutsopoulos, Jannis. 2010. Multimodal – intertextuell – heteroglossisch: Sprach-Gestalten in „Web 2.0“-Umgebungen. In Arnulf Deppermann & Angelika Linke (eds.), *Sprache intermedial. Stimme und Schrift, Bild und Ton*, 419–445. Berlin & New York: De Gruyter.
- Bachtin, Michail M. 1990. *Literatur und Karneval. Zur Romantheorie und Lachkultur*. Frankfurt am Main: Fischer.
- Berghofer, Simon. 2009. *Dialogizität und [Inter]Textualität im Internet. Zur kommunikativen Textgenese von literarischen Blogs*. Magisterarbeit. Berlin. Freie Universität Berlin.
- Busse, Dietrich. 1997. Das Eigene und das Fremde. Annotationen zu Funktion und Wirkung einer diskurssemantischen Grundfigur. In Matthias Jung, Martin Wengeler & Karin Böke (eds.), *Die Sprache des Migrationsdiskurses. Das Reden über „Ausländer“ in Medien, Politik und Alltag*, 17–35. Opladen: Westdeutscher Verlag.
- Busse, Dietrich. 2003. Begriffsgeschichte oder Diskursgeschichte? Zu theoretischen Grundlagen und Methodenfragen einer historisch-semantischen Epistemologie. In Carsten Dutt (ed.), *Herausforderungen der Begriffsgeschichte*, 17–38. Heidelberg: Winter.
- Dang-Anh, Mark. 2019. *Protest twittern. Eine medienlinguistische Untersuchung von Straßenprotesten* (= Locating Medien/ Situierete Medien 22). Bielfeld: transcript.
- Dang-Anh, Mark, Jessica Einspänner & Caja Thimm. 2013. Kontextualisierung durch Hashtags: die Mediatisierung des politischen Sprachgebrauchs im Internet. In Hajo Diekmannshenke & Thomas Niehr (eds.), *Öffentliche Wörter: Analysen zum öffentlich-medialen Sprachgebrauch*, 137–159. Stuttgart: ibidem.
- Firth, John Rupert. 1957. *Papers in Linguistics, 1934–1951*. London: Oxford University Press.
- Fix, Ulla. 2000. Aspekte der Intertextualität. In Klaus Brinker, Gerd Antos, Wolfgang Heinemann & Sven F. Sager, Sven F (eds.), *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*, 449–457. Berlin & New York: De Gruyter (=Handbücher zur Sprach- und Kommunikationswissenschaft), 1. Halbband.
- Fix, Ulla. 2008. *Texte und Textsorten – sprachliche, kommunikative und kulturelle Phänomene*. Leipzig: Frank & Timme.
- Girnth, Heiko. 1996. Texte im politischen Diskurs. Ein Vorschlag zur diskursorientierten Beschreibung von Textsorten. *Muttersprache* 106(1). 66–80.
- Goman, Helge & Derya Gür-Şeker. 2020. (Multimodale) Vernetzungsstrukturen in Online-Zeitungen. Methodische und analytische Zugänge einer digitalen Linguistik. In Derya Gür-Şeker (ed.): *Wörter, Wörterbücher, Wortschätze. (Korpus-)Linguistische Perspektiven*, 135–155. Duisburg: Universitätsverlag Rhein-Ruhr.
- Gray, Jonathan. 2006. *Watching with the Simpsons. Television, parody, and intertextuality*. New York: Routledge.
- Gredel, Eva. 2018. *Digitale Diskurse und Wikipedia: Wie das Social Web Interaktion im digitalen Zeitalter verwandelt*. Tübingen: Narr Francke Attempto Verlag.
- Gür-Şeker, Derya. 2012. *Transnationale Diskurslinguistik: Theorie und Methodik am Beispiel des sicherheitspolitischen Diskurses über die EU-Verfassung in Deutschland, Großbritannien und der Türkei* (Sprache, Politik, Gesellschaft 6). Bremen: Hempen.
- Heibach, Christiane. 2003. *Literatur im elektronischen Raum*. Frankfurt am Main: Suhrkamp.
- Holthuis, Susanne. 1993. *Intertextualität. Aspekte einer rezeptionsorientierten Konzeption*. Tübingen: Stauffenburg.

- Janich, Nina. 2019. Intertextualität und Text(sorten)vernetzung. In Nina Janich (ed.), *Textlinguistik. 15 Einführungen*, 177–196. 2. Aufl. Tübingen: Narr.
- Kienpointer, Manfred. 1992. *Alltagslogik. Struktur und Funktion von Argumentationsmustern*. Stuttgart-Bad Cannstatt: frommann-holzboog.
- Kienpointner, Manfred. 1996. *Vernünftig argumentieren. Regeln und Techniken der Diskussion*. Reinbek: Rowohlt.
- Klug, Nina-Maria. 2016. Multimodale Text- und Diskurssemantik. In Nina-Maria Klug & Hartmut Stöckl (eds.), *Handbuch Sprache im multimodalen Kontext* (Handbücher Sprachwissen 7), 165–189. Berlin & Boston: de Gruyter.
- Klug, Nina-Maria. 2018. Wenn Schlüsseltexte Bilder sind. Aspekte von Intertextualität in Presse und öffentlichem Raum. In: Steffen Pappert & Sascha Michel (eds.), *Multimodale Kommunikation in öffentlichen Räumen. Texte und Textsorten zwischen Tradition und Innovation* (Perspektiven germanistischer Linguistik 14), 109–131. Stuttgart: ibidem-Verlag.
- Kopperschmidt, Josef. 1989. *Methodik der Argumentationsanalyse* (Problemata 119). Stuttgart-Bad Cannstatt: frommann-holzboog.
- Kusner, Matt, Yu Sun, Nicholas Kolkin & Kilian Weinberger. 2015. From word embeddings to document distances. *Proceedings of the 32nd International Conference on Machine Learning* 37. 957–966.
- Kristeva, Julia. 1978. Bachtin, das Wort, der Dialog und der Roman. In Dieter Hillebrandt (ed.), *Zur Struktur des Romans* (Wege der Forschung 488), 388–407. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Marx, Konstanze & Georg Weidacher. 2014. *Internetlinguistik. Ein Lehr- und Arbeitsbuch*. Tübingen: Narr.
- Mikolov, Tomas, Ilya Sutskever, Kai Chen, G.s Corrado, Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. *Advances in Neural Information Processing Systems*, 26. (NIPS 2013)
- Pentzold, Christian, Claudia Fraas & Stefan Meier. 2013. Online-mediale Texte: Kommunikationsformen, Affordanzen, Interfaces. *Zeitschrift für Germanistische Linguistik* 41(1), 81–101.
- Pfister, Manfred. 1985. Konzepte der Intertextualität. In Ulrich Broich & Manfred Pfister (eds.), *Intertextualität. Formen, Funktionen, anglistische Fallstudien* (Konzepte der Sprach- und Literaturwissenschaft 35), 1–30. Tübingen: May Niemeyer Verlag.
- Polajnar, Janja. 2019. Zur Intertextualität in Postings. Eine exemplarische Analyse deutschsprachiger Postings in Kommentarforen zu Online-Zeitungsartikeln zum Bologna-Prozess. In Uršula Krevs Birk (ed.), *Wort – Text – Kultur. Studien zu Ehren von Prof. Dr. Siegfried Heusinger zum 85. Geburtstag = Beseda – besedilo – kultura: razprave v počastitev 85-letnice prof. dr. Siegfrieda Heusingerja* (Linguistica, 59), 267–285. Ljubljana: Znanstvena založba Filozofske fakultete.
- Polajnar, Janja. 2017. Kultur-kontrastiver diskursanalytischer Zugang zum Bologna-Diskurs am Beispiel vom Mobilitäts-Topos in den slowenischen und deutschen Tageszeitungen. In Petra Kramberger, Irena Samide & Tanja Žigon (eds.), *Und die Brücke hat gezogen, die vom Ost zum West sich schwingt. Literarische, kulturelle und sprachliche Vernetzungen und Grenzüberschreitungen. Festschrift für Mira Miladinović Zalaznik*, 417–430. Ljubljana: Znanstvena založba Filozofske fakultete.
- Polajnar, Janja & Tanja Škerlavaj. 2018a. Bologna-Prozess: Leichter Einstieg in den Beruf oder nur Studium mit Durchlaufgeschwindigkeit? Ökonomie-Topoi und der Finanzierungs-Topos zu Beginn des slowenischen und deutschen Bologna-Diskurses. *Muttersprache* 128(2), 192–213.

- Polajnar, Janja & Tanja Škerlavaj. 2018b. Zentrale Topoi im Bologna-Diskurs: Ein Einblick in die kultur-kontrastive Untersuchung der Argumentationsmuster im Deutschen und im Slowenischen. In Eva Gredel, Heidrun Kämper, Ruth M. Mell, Janja Polajnar (eds.), *Diskurs – kontrastiv: Diskurslinguistik als Methode zur Erfassung transnationaler und sprachübergreifender Diskursrealitäten*, 150–174. Bremen: Hempen.
- Scharloth, Joachim. 2005. Die Semantik der Kulturen. Diskurssemantische Grundfiguren als Kategorien einer linguistischen Kulturanalyse. In Dietrich Busse, Thomas Niehr & Martin Wengeler (eds.), *Brisante Semantik. Neuere Konzepte und Forschungsergebnisse einer kulturwissenschaftlichen Linguistik* (RGL 259), 119–135. Tübingen: Niemeyer.
- Scharloth, Joachim, Franz Keilholz, Simon Meier-Vieracker, Xiaozhou Yu & Roman Dorniok. 2019. Datengeleitete Kategorienbildung in den Digital Humanities: Paraphrasen aus korpus- und computerlinguistischer Perspektive. In Charlotte Schubert, Paul Molitor, Jörg Ritter, Joachim Scharloth & Kurt Sier (eds.): *Platon Digital: Tradition und Rezeption* (Digital Classics Books 3), 61–88.
- Schröter, Melani. 2017. Texte und Textsorten. In Kersten Sven Roth, Martin Wengeler & Alexander Ziem (eds.): *Handbuch Sprache in Politik und Gesellschaft* (Handbücher Sprachwissen 19), 212–233. Berlin/ Boston: De Gruyter.
- Spieß, Constanze. 2011. *Diskurshandlungen. Theorie und Methode linguistischer Diskursanalyse am Beispiel der Bioethikdebatte*. Berlin & Boston: De Gruyter.
- Spitzmüller, Jürgen & Ingo H. Warnke. 2011. *Diskurslinguistik. Eine Einführung in Theorien und Methoden der transtextuellen Sprachanalyse*. Berlin & Boston: De Gruyter.
- Steyer, Kathrin. 1997. Irgendwie hängt alles mit allem zusammen – Grenzen und Möglichkeiten einer linguistischen Kategorie, Intertextualität. In Josef Klein & Ulla Fix (eds.), *Textbeziehungen. Linguistische und literaturwissenschaftliche Beiträge zur Intertextualität*, 83–107. Tübingen: Stauffenburg.
- Stierle, Kathrin. 1984. Werk und Intertextualität. In Karlheinz Stierle & Rainer Warning (eds.), *Das Gespräch* (Poetik und Hermeneutik 11), 139–150. München: Fink.
- Škerlavaj, Tanja. 2017. International heißt die Devise: eine kultur-kontrastive Analyse der Globalisierungs-Topoi im deutschen und slowenischen Bologna-Diskurs. In Petra Kramberger, Irena Samide & Tanja Žigon (eds.), „Und die Brücke hat gezogen, die vom Ost zum West sich schwingt“: literarische, kulturelle und sprachliche Vernetzungen und Grenzüberschreitungen: *Festschrift für Mira Miladinović Zalaznik*, 431–443. Ljubljana: Znanstvena založba Filozofske fakultete.
- Teubert, Wolfgang. 1998. Korpus und Neologie. In Wolfgang Teubert (ed.), *Neologie und Korpus*, 129–170. Tübingen: Narr.
- Wengeler, Martin. 2003. *Topos und Diskurs. Begründung einer argumentationsanalytischen Methode und ihre Anwendung auf den Migrationsdiskurs (1960–1985)*. Tübingen: Max Niemeyer Verlag.
- Wikipedia Cross-lingual Image Analysis: <https://tools.digitalmethods.net/beta/wikipediaCrosslingualImageAnalysis/> (Zugriff: 04.11.2019)
- Wittenbrink, Heinz. 2005. *Newsfeeds mit RSS und Atom: Nachrichtenaustausch mit Really Simple Syndication und Atom*. Galileo Computing.
- Ziem, Alexander. 2005. Begriffe, Topoi, Wissensrahmen: Perspektiven einer semantischen Analyse gesellschaftlichen Wissens. In Martin Wengeler (ed.), *Sprachgeschichte als Zeitgeschichte. Konzepte, Methoden und Forschungsergebnisse der Düsseldorfer Sprachgeschichtsschreibung für die Zeit nach 1945*, 315–348. Hildesheim & New York: Olms.