

Tim Feldmüller / Marc Kupietz / Andressa Costa / Birgit Rapp / Annette Leßmöllmann /
Peter Fischer / Rameela Yaddehige

FISCHBIOLOGIE, GENDERSHYCE UND TRANSLOBBYISTEN

KORPUSAUFBAU UND ERSTE ANALYSEN ZU SPRACHLICHEN PRAKTIKEN DER POLARISIERUNG IM DISKURS UM GESCHLECHT UND GENDER

Tim Feldmüller, Marc Kupietz, Peter Fischer und Rameela Yaddehige sind Mitarbeiter*innen im Programmbe-
reich Korpuslinguistik am Leibniz-Institut für Deutsche Sprache sowie Teil des KoKoKom-Teams.

Annette Leßmöllmann ist Professorin für Wissenschaftskommunikation mit dem Schwerpunkt Linguistik im Department für Wissenschaftskommunikation am KIT und Principal Investigator des Projekts KoKoKom.

Andressa Costa ist wissenschaftliche Mitarbeiterin und Postdoktorandin im Department für Wissenschaftskommunikation am KIT und Teil des KoKoKom-Teams.

Birgit Rapp ist Doktorandin an der Universität Tübingen und Teil des KoKoKom-Teams.

„Zweigeschlechtlichkeit“ versus „Vielgeschlechtlichkeit“, „Gendersprache“ und „Genderverbot“: Der öffentliche Diskurs über Fragen von Geschlecht und Gender ist hitzig und von sprachlichen Praktiken durchzogen, die eine Lagerbildung hervorrufen können. Wie ist es in einer solchen Diskursituation noch möglich, einen *Common ground* zu schaffen, im Gespräch zu bleiben und einen fruchtbaren Austausch über Forschung und Wissenschaft zu gestalten? Das Projekt „Konflikt und Konsens in der Wissenschaftskommunikation: Über Geschlecht und Gender streiten“ (KoKoKom, <www.kokokom.de>) analysiert polarisierende sprachliche Praktiken aus linguistischer und rhetorischer Perspektive und ergründet die Rolle von *Common ground* in der Wissenschaftskommunikation. Das Projekt wird vom Bundesministerium für Forschung, Technologie und Raumfahrt im Rahmen der Initiative „Wissenschaftskommunikation“ gefördert und verknüpft Forschung mit der Entwicklung depolarisierender Diskursformate.

KAMPF UM GESCHLECHT UND GENDER WIRD ZU KAMPF UM GUTE WISSENSCHAFT

Um Polarisierung besser zu verstehen, analysieren Forschungsteams aus der Rhetorik (Lehrstuhl von Olaf Kramer, Universität Tübingen), der Linguistik (Lehrstuhl Susanne Winkler, Universität Tübingen), der Wissenschaftskommunikation und Linguistik (Lehrstuhl Annette Leßmöllmann, KIT) und der Korpuslinguistik (Marc Kupietz, Leibniz-Institut für Deutsche Sprache, Mannheim) ein im Projekt aufgebautes Korpus aus Medienbeiträgen und User-Kommentaren, das ein außerordentlich kontroverses öffentliches Diskursgeschehen kartiert: Als die Biologie-Doktorandin Marie-Luise Vollbrecht im Juli 2022 einen öffentlichen Vortrag zu den „zwei Geschlechtern in der Biologie“ ankündigte, führte das zu Protesten, die ihren Ansatz als trans- und queerfeindlich kritisierten. Der Vortrag wurde von der Universität abgesagt, was wiederum Debatten über Wissenschaftlichkeit und Wissenschaftsfreiheit hervorrief. Auch ein Meinungsbeitrag in der WELT, an dem Vollbrecht beteiligt war und der den Öffentlich-Rechtlichen Rundfunk für seine Darstellung von Geschlecht und Gender kritisiert, führte zu Auseinandersetzungen. Die Kontroverse wurde in vielen traditio-

nellen und sozialen Medien geführt und dient dem Projekt als Fallstudie für die Untersuchung sprachlicher Ausprägung von Polarisierung, etwa durch rhetorische Fragen, Bedeutungsverschiebungen oder Abwertung anderer. Wie Korpora zu solchen Diskursen in Forschungsprojekten wie dem KoKoKom-Projekt konzeptioniert, aufgebaut und ausgewertet werden können, möchten wir im Folgenden zeigen.

Korpora für die Diskursanalyse

In einem programmatischen Aufsatz von 1994 gehen die Linguisten Dietrich Busse und Wolfgang Teubert nicht nur auf die von ihnen in der Überschrift gestellte Frage ein, ob „Diskurs ein sprachwissenschaftliches Objekt [ist]“ (Busse/Teubert 1994, S. 10). Sie rücken auch den Zusammenhang von Diskurs und Untersuchungskorpus in den Fokus und unterscheiden dabei zwei Ebenen. Diskurse selbst definieren sie in Hinblick auf wissenschaftliche Untersuchungen als „virtuelle Textkorpora“ (ebd., S. 14), die drei Kriterien genügen müssen:

Zu einem Diskurs gehören alle Texte, die

- sich mit einem als Forschungsgegenstand gewählten Gegenstand, Thema, Wissenskomplex oder Konzept befassen, untereinander semantische Beziehungen aufweisen und/oder in einem gemeinsamen Aussage-, Kommunikations-, Funktions- oder Zweckzusammenhang stehen,
- den als Forschungsprogramm vorgegebenen Eingrenzungen in Hinblick auf Zeitraum/Zeitschnitte, Areal, Gesellschaftsschnitt, Kommunikationsbereich, Texttypik und andere Parameter genügen,
- und durch explizite oder implizite (text- oder kontextsemantisch erschließbare) Verweisungen aufeinander Bezug nehmen bzw. einen intertextuellen Zusammenhang bilden. (Busse/Teubert 1994, S. 14)

DISKURSE WERDEN DURCH ‚KONKRETE KORPORA‘ UNTERSUCHBAR GEMACHT

Solche Korpora sind bereits deswegen virtuell, da kein*e Forscher*in alle Texte zu einem Thema sammeln kann. Dies ist schon bei öffentlich verfügbaren Texten aufgrund von finanziellen und anderen praktischen Limitationen kaum möglich – eine große Zahl von Texten ist aber überhaupt

nicht erfassbar, da sie etwa in privater Kommunikation ausgetauscht werden oder in einem flüchtigen Medium verfasst sind. In Abgrenzung zu diesem virtuellen Korpus führen Busse und Teubert daher den Begriff des „konkret[en] Korpus“ (1994, S. 14) ein. Solche konkreten Korpora entsprechen den in einzelnen Untersuchungen verwendeten Daten und stellen „Teilmengen der jeweiligen Diskurse“ (ebd.) dar. Bei der Zusammenstellung ist laut den Autoren zum einen eine inhaltliche Relevanz der Texte für den Diskurs Voraussetzung; zum anderen auch deren forschungspraktische Verfügbarkeit. Letztlich bedürfe es einer „Prüfung der Eignung der einzelnen Texte“ (ebd.). Hier zeigt sich, dass die beiden Autoren eher kleine Textsammlungen vor Augen haben. Wenn jeder Text gelesen und nach seiner thematischen Einschlägigkeit bewertet werden muss, bleibt die Menge einzubeziehender Texte notwendigerweise eher gering. Dies steht einerseits im Widerspruch zur Forderung, auch aus der Diskurslinguistik, den Untersuchungsgegenstand möglichst repräsentativ abzubilden, wofür große Korpora besser geeignet sind als kleine (Scharloth/Eugster/Bubenhof 2013). Andererseits führt dies auch zu einem Problem, das insbesondere von Bubenhof (2009, S. 36) hervorgehoben wird: Nur solche Texte in das Korpus einzubeziehen, die den Forschenden als relevant und für einen Diskurs zentral erscheinen, kann dazu führen, die eigenen Ergebnisse vorzuformen und nur das Offensichtliche abzubilden. Dies führt Bubenhof dazu, thematische Diskursdefinitionen abzulehnen und mit einem themenunspezifischen Korpus zu arbeiten. Diskurse definiert er stattdessen entlang von geteilten sprachlichen Mustern, sogenannten ‚Sprachgebrauchsmustern‘ (ebd., S. 37).

In der Praxis finden sich verschiedene Herangehensweisen zwischen diesen Polen. Ein häufiges Vorgehen besteht z. B. darin, ein oder mehrere für den zu untersuchenden Diskurs typische Suchwörter zu definieren und alle Texte aus einem thematisch heterogenen Korpus aufzunehmen, die diese Suchwörter enthalten. Auch wenn dieses Vorgehen Gefahr läuft, solche Texte zu verpassen, die auf untypische Art und Weise auf ein Thema Bezug nehmen, ist es für thematisch eng gefasste und zu einem zeitlichen Punkt akute Diskurse ein probates Mittel. So stellen z. B. Ziem/Pentzold/Fraas (2018) ein Korpus zur Snowden-Affäre anhand der Suchwörter

Snowden, *NSA* und *data* zusammen und Baker et al. (2020) bilden ein themenbezogenes Subkorpus zum Framing von Übergewicht in der britischen Presse anhand der Suchwörter *obese* und *obesity*. Ein anderer Ansatz besteht darin, das Korpus nicht thematisch zu filtern, sondern Themen entlang der Verteilung von Wörtern im Korpus zu rekonstruieren. Gemeinsam vorkommende und in ähnlichen Kontexten vorkommende Wörter dienen dabei als Indikatoren für thematische Diskurse (Feldmüller 2026; Feldmüller/Schweden i. Ersch.). Dieses Vorgehen wurde bislang aber eher zur Beschreibung sehr breit diskutierter und über einen langen Zeitraum aktueller Themen angewandt.

‘DAS EINE’ KORPUS KANN ES NICHT GEBEN

Im KoKoKom-Projekt haben wir als interdisziplinäres Team unterschiedliche Fragestellungen und Ansprüche an das Korpus. Der Aufbau eines einzigen ‚konkreten Korpus‘ für eine einzelne Fragestellung kommt daher nicht in Frage. Selbst ein einziges ‚virtuelles Korpus‘ können wir nicht als Ziel festlegen, da für einzelne Fragestellungen auch der Einbezug von Quellen von Interesse ist, die nicht im engeren Sinn den Vollbrecht-Diskurs widerspiegeln, sondern Teil des weiteren Diskurses um Geschlecht und Gender sind. Wir verfolgen daher das Urstichprobendesign (Kupietz/Längen/Diewald 2023), das dem Deutschen Referenzkorpus (DeReKo) (Kupietz et al. 2018) zu Grunde liegt. Die Idee besteht darin, nicht ‚das eine‘ Korpus zur Verfügung zu stellen, das in Hinblick auf gewisse Kriterien und Forschungsfragen ausgewogen ist, sondern die Bearbeitung verschiedener Fragestellungen zu ermöglichen. Dazu wird eine recht breite Menge an Texten über die Korpusplattform bereitgestellt. Die Texte sind jedoch mit Metadaten versehen. Mit Hilfe der Metadaten können – abhängig vom konkreten Forschungsinteresse – Subkorpora gebildet werden, etwa, wenn nur Texte aus einem bestimmten Zeitraum und Medium untersucht werden sollen.

Inhalt des Korpus

Für das Projekt wurde ein Textkorpus erstellt, das unterschiedliche Kommunikationsformate und Plattformen umfasst. Dazu gehören Beiträge und Kommentare aus sozialen Medien (Twitter/X, Reddit, YouTube, Facebook und Insta-

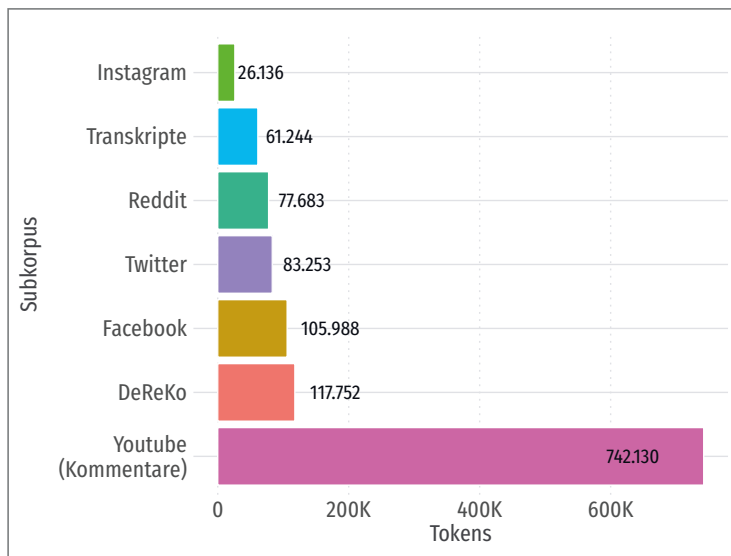


Abb. 1: Zusammensetzung des Korpus nach Quellen

gram), Transkripte von YouTube-Videos sowie Texte aus dem DeReKo. Alle Texte des Korpus sind Teil des Diskurses um Geschlecht und Gender, wobei sich die meisten Beiträge dem engeren Diskurs um den abgesagten Vortrag Marie-Luise Vollbrechts zuordnen lassen. Der Erhebungszeitraum startete am 1. Juni 2022 mit der Publikation des WELT-Artikels von Vollbrecht und ihren Co-Autor*innen als Diskursereignis. Als Ende wurde das Jahresende 2022 festgelegt, um eine verlässliche Identifikation diskurszugehöriger Texte mithilfe des Suchwortes *Vollbrecht* zu gewährleisten. Zusätzlich wurden alle vorliegenden Kommentare, etwa zu YouTube-Videos, erhoben, sofern der ursprüngliche Beitrag im Untersuchungszeitraum veröffentlicht wurde. Es ergibt sich so insgesamt eine Publikationszeitspanne von Juni 2022 bis Januar 2025 für die im Korpus enthaltenen Texte. Abbildung 1 gibt einen Überblick über die Zusammensetzung des Korpus.

Für die Datenerhebung wurden gezielt Beiträge in sozialen Medien gesucht, die den Ausdruck bzw. Hashtag *Vollbrecht* enthalten. Um die verschiedenen Social-Media-Plattformen systematisch zu durchsuchen, kamen unterschiedliche Tools zum Einsatz.¹ Von Reddit wurden deutschsprachige Beiträge über die Suche nach *Vollbrecht* extrahiert, sofern sie im Zeitraum 1. Juni bis 31. Dezember 2022 erstellt wurden. Für das YouTube-Kommentare-Subkorpus wurden zunächst mit dem Suchwort *Vollbrecht* Videos aus dem Untersuchungszeitraum gesammelt und alle zugehörigen Kommentare erhoben. Anschließend wurden Kommentare zu Videos außerhalb des genannten Zeitraums zusammengetragen, deren Transkripte für den Diskurs als relevant betrachtet wurden. Die Auswahl der transkribierten Videos erfolgte auf Basis der Forschungsinteressen im Projekt, sodass z. B. Videos mit relevanten rhetorischen Fragen ausgewählt wurden. Diese Videos wurden zunächst automatisch transkribiert und die so entstandenen Transkripte in mehreren Schritten von verschiedenen Personen überprüft.²

Ein Blick in den Maschinenraum: Aufbereitung der Ausgangsdaten

Für die Aufbereitung der in Tabellenform (als Data Frame oder CSV-Datei) quellenweise erhobenen Ausgangsdaten wurden R-Skripte entwickelt, die diese jeweils in eine TEI-I5-XML-Datei (Lüngen/Sperberg/McQueen 2012) umwandeln, wodurch Metadaten nach einem standardisierten Schema eingefügt werden. Die Konvertierung in einen von der Korpusplattform (siehe nächster Abschnitt) lesbaren Index, d. h. ein Verzeichnis, das es ermöglicht, sehr schnell nach Wörtern oder sprachlichen Strukturen im Korpus zu suchen, erfolgt über die Pakete *korapxmltool* und *Krill-Indexer*.³

KORPUSDATEN AUF KNOPFDRUCK

Die Verarbeitung ist durch GNU Make und einen sogenannten *Continuous Deployment (CD)* GitLab-Workflow automatisiert: Dadurch können die einzelnen Verarbeitungsschritte (von der Erzeugung und formalen Überprüfung der XML-Dateien bis zur Aktualisierung des Index) nach einer Aktualisierung der Korpusdaten per Knopfdruck in der Git-Lab-Oberfläche gestartet werden.

Zugang zum Korpus

Das KoKoKom-Korpus wird primär über eine Instanz der Korpusanalyseplattform KorAP (Bański et al. 2012) bereitgestellt. KorAP ist ein webbasiertes Tool, das den Zugriff auf eine Vielzahl von linguistischen Korpora ermöglicht, darunter auch das Deutsche Referenzkorpus DeReKo (Kupietz et al. 2018).

ÜBER KORAP WIRD DAS KORPUS FÜR DAS GESAMTE PROJEKT ONLINE BEREITGESTELLT UND ANALYSIERBAR

KorAP bietet eine benutzerfreundliche Oberfläche für die Suche in und die Analyse von Texten und unterstützt die Definition von Subkorpora anhand von Metadaten, Suchen mit regulären Ausdrücken und alle Arten linguistischer Annotationen. Für das KoKoKom-Korpus werden Annotationen zu Wortart, Lemma, Morphosyntax und Dependenz mit dem Tool *Udpipe2* (Straka 2018) erzeugt. Außerdem bietet

KorAP für statistische Analysen, wie z. B. Kookkurrenzanalysen, und alle Arten von Analysen, die reproduzierbar sein sollten, Bibliotheken für die Programmiersprachen R und Python an (Kupietz/Diewald/Margaretha 2020).

Rechtliche und ethische Aspekte

Die Nutzung von Social-Media-Daten zu Forschungszwecken berührt diverse rechtliche und ethische Fragestellungen. Im Projektverlauf wurde eingehend geprüft, ob Datenerhebung und -verarbeitung konform zu geltendem Recht und relevanten Richtlinien erfolgen. Besonderes Augenmerk lag dabei auf der Datenschutz-Grundverordnung (DSGVO) sowie den Nutzungsbedingungen der jeweiligen Plattformen.

Hinsichtlich der Urheber- und Nutzungsrechte bestand – im Unterschied zum DeReKo – keine Möglichkeit, umfassende, übertragbare Lizenzen zu erwerben. Der Zugang zum KoKoKom-Korpus erfolgt daher ausschließlich im Rahmen der Text-und-Data-Mining-Schranke (§ 60d UrhG), die die Nutzung auf „einen bestimmt abgegrenzten Kreis von Personen für deren gemeinsame wissenschaftliche Forschung“ beschränkt, der in diesem Fall die Projektmitarbeitenden umfasst. Für Kooperations- und Überprüfungs Zwecke kann auf Anfrage auch externen Forschenden Zugang gewährt werden (vgl. Raue 2022).

Analysen

Im Folgenden möchten wir schlaglichtartig Einblicke in erste Analysen geben, die wir bereits auf Basis des KoKoKom-Korpus durchgeführt haben und die aus den unterschiedlichen Perspektiven der Projektmitglieder auf die Daten zugreifen.

Bedeutungswandel

Als erstes möchten wir eine word-embedding-basierte Analyse vorstellen (vgl. Feldmüller/Kupietz i. Beg.). Mit Word Embeddings wird die Semantik von Wörtern in Vektoren, d. h. „Koordinaten“, die sie in einem gemeinsamen Vektorraum verorten, abgebildet. Sie werden mithilfe eines neuronalen Netzes trainiert, das lernt, welche Wörter in ähnliche Lücken passen und daher – so die zentrale Hypothese der distributionellen Semantik – eine ähnliche Bedeutung aufweisen (Sahlgren 2008).

Foundry	Layer	dieser	ganzen	Gendershyce	geht	es
ud	l	dies	ganz	Gendershyce	gehen	es
ud	m	case:dat gender:fem num:gender:feminine prontype:dem	degree:pos number:sing	gender:fem number:sing person:3	mood:ind number:sing person:3 tense:pres verbform:fin	case:n number: gender: number: person: prontype:
ud	p	DET	ADJ	NOUN	VERB	PROI

Abb. 2: Konkordanz-Analyse zu Komposita mit dem Erstglied *Gender*

Um semantische Charakteristika des Sprachgebrauchs im Vollbrecht-Diskurs herausarbeiten zu können, berechnen wir zwei Word-Embedding-Modelle: eines auf Basis unseres KoKoKom-Korpus und eines auf Basis eines Teils des DeReKo, der ebenfalls Online-Kommunikation enthält, jedoch nicht spezifisch zum Diskurs um Sex (biologisches Geschlecht) und Gender.⁴ Anschließend vergleichen wir, welche Wörter oder Mehrworteinheiten im Vergleich der Korpora in möglichst unterschiedlichen lexikalischen Umgebungen vorkommen, d. h. welche Wörter ihre Bedeutung am meisten verändert haben könnten (vgl. zu dieser Auffassung von Bedeutungswandel bereits Busse/Teubert 1994, S. 18). Das Vorgehen wird dabei auf alle im Korpus vorkommenden Wörter angewendet und ist insofern *corpus-driven*, d. h. ohne Vorannahmen, welche Wörter ihre Bedeutung gewandelt haben könnten. Es können so Auffälligkeiten freigelegt werden, denen in detaillierteren Analysen weiter nachgegangen werden kann.

GEOMETRISCHE VERMESSUNGEN VON BEDEUTUNG: WORTVEKTOREN ALS MITTEL ZUR ANALYSE VON SEMANTISCHEM WANDEL

Als Top 15 Ergebnisse erhalten wir die folgenden Ausdrücke: „XY, XX, *Berliner_Zeitung*, *Capriccio*, *Böhmermann*, bzw. *Süddeutsche_Zeitung*, zwei, *Biologie*, *Marie*, jetzt, einfach, Mann, *“. Bereits an diesem Ergebnis lässt sich der thematische Fokus des KoKoKom-Korpus ablesen. So finden sich eine Reihe von Ausdrücken, die auf die Aushandlung von Geschlecht (XY, XX, Mann) und (Non-)Binarität (zwei, *, *Biologie*) verweisen. In anderen Fällen wie *Berliner_Zeitung* oder *Süddeutsche_Zeitung* ist nicht ohne Weiteres ersichtlich, worin die unterschiedlichen Verwendungsweisen be-

Kollokat links	LogLikelihood	zwei	Kollokat rechts	LogLikelihood
nur	4345,94		Geschlechter	8151,91
genau	318,45		biologische	1191,04
Biologie	225,53		Geschlechtern	402,82
gibt	197,06		unterschiedliche	181,13
exakt	93,08		geschlechter	163,40

Tab. 1: Kollokationsanalysen zu *zwei* im KoKoKom-Korpus mit Kollokationsfenstern von einem Wort zu beiden Seiten

stehen könnten. Besonders hier ist eine Sichtung von Korpusbelegen notwendig, um die Ergebnisse interpretieren zu können. Die Konkordanzen in den jeweiligen Korpora zeigen, dass die Namen der Zeitungen im KoKoKom-Korpus fast ausschließlich als Teil von Facebook-Kommentaren auftreten, in denen sie den Kommentar einleiten und den Account, dem geantwortet wird, anzeigen, so z. B. im folgenden Beleg:

- (1) Süddeutsche Zeitung Wissenschaftliche Fakten sind weder tolerant noch intolerant [KFC/006.000076]⁵

Inhaltlich relevanter ist z. B. der Ausdruck *zwei*, der im Vollbrecht-Diskurs eine semantische Verengung erfährt. Während das Wort im DeReKo eine bestimmte Anzahl von Objekten, die potenziell eine beliebige Ausprägung annehmen könnte, bezeichnet, wird es im KoKoKom-Korpus musterhaft in Bezug auf Geschlecht gebraucht. Dies zeigt sich auch im engen Umfeld des Wortes, für das wir mit einer Kollokationsanalyse ermitteln, welche Wörter unmittelbar vor und nach dem Ausdruck *zwei* vorkommen. Die Ergebnisse veranschaulichen die Bedeutungsverschiebung von einem gewöhnlichen Zahlwort hin zu einer Versprachlichung von Binarität, in der es [nur/ genau/ exakt] *zwei* [unterschiedliche/ biologische] *Geschlechter* gibt (siehe Tab. 1).

Polarisierung durch rhetorische Fragen

Eine weitere Analyse (vgl. Rapp i. Beg.) befasst sich mit dem polarisierenden Effekt rhetorischer Fragen (RF). Unter RF werden Interrogativsätze verstanden, auf die von Sprechenden keine Antwort erwartet wird, da die Antwort den Sprechenden und den Hörenden bekannt ist (z. B. Meibauer 1986; Han 2002; Caponigro / Sprouse 2007). Sie werden als Assertionen mit Polaritätsumkehr verstanden, negative RFs also als positive Assertionen, positive RFs als negative Assertionen (z. B. Han 2002, S. 202). Das Polarisierungsverständnis orientiert sich an der von Fortuna (2019) vorgestellten Kombination von Elementen in polarisierenden Texten: Provokation, Simplifizierung und Übertreibung.

Zunächst wurde ein Subkorpus gebildet, das in Hinblick auf Positionen in der Geschlechterdebatte ausgewogen sein sollte. Einbezogen wurden YouTube-Kommentare zu einer online gehaltenen Version des abgesagten Vollbrecht-Vortrags sowie ein Video des Kanals MAITHINK X. Anschließend wurde eine Suche nach Fragezeichen durchgeführt und aus den Ergebnissen ein Sample aus jeweils 300 Fragezeichen sowie jeweils allen Kombinationen aus Fragezeichen und mindestens einem weiteren Fragezeichen oder Ausrufezeichen gebildet. Letztlich wurden die Ergebnisse nach folgenden Kategorien annotiert: ‚RF‘, ‚keine RF‘ und ‚nicht eindeutig‘ für Fälle, in denen keine eindeutige Eingruppierung möglich war.

In der Analyse konnten RF in verschiedenen syntaktischen Formen sowie unterschiedliche polarisierende Marker identifiziert werden. Ein Beispiel für eine polarisierende RF ist die fettgedruckte Frage in (2). Die schon im Original vorhandene fette Schrift ist eine Emphase durch eine Form der typografischen Intensivierung (Cotgrove 2025), durch die provoziert wird. Es handelt sich um eine durch die Partikel *denn* (z. B. Meibauer 1986) markierte Entscheidungs-RF. Ein weiteres polarisierendes Element ist das Demonstrativpronomen *die*, durch das auf eine Out-Group verwiesen wird, hier das linke Spektrum, sodass eine simplifizierende In-Group/ Out-Group-Dichotomie entsteht (siehe Fortuna 2019). Ein weiteres polarisierendes Element ist das als Übertreibung wirkende Absolut *alle* (vgl. z. B. Scott 2002; Zhao et al. 2022 für englische Äquivalente), wodurch die gesamte Out-Group als spinnend bezeichnet wird.

- (2) [...] Und hier geht es um die Anzahl möglicher Geschlechter. Warum wird dieses Thema primär aus dem linken Spektrum stammend so gehyped ? **Ja spinnen die denn jetzt alle ?** 🤔 [KYC/MAL.000294]

Diskursive Polarisierungsmuster und rhetorische Strategien

In einer weiteren Untersuchung des KoKoKom-Korpus wurden polarisierende Frames und rhetorische Strategien iden-

tifiziert und beschrieben (vgl. Costa i. Ersch.). Aufbauend auf Brüggemann und Meyer (2023) wird Polarisierung als kommunikative Divergenz verstanden, die sich auf zwei Ebenen zeigt: einerseits ideologisch als inhaltliche Meinungsgegensätze und andererseits affektiv als emotionale Abgrenzung und Konstruktion von Feindbildern gegenüber einer angenommenen Fremdgruppe. Der Fokus der vorliegenden Analyse liegt auf ideologischer Polarisierung.

WENN WORTE SPALTEN: WIE DISKURSE ÜBER GESCHLECHT ZU IDEOLOGISCHEN FRONTEN WERDEN

Die Untersuchung kombinierte computergestützte und qualitative Verfahren: Hierzu zählen Biterm Topic Modelling (Yan et al. 2013) zur Erkennung latenter Themen und Frames, Kookkurrenzanalysen (Scott 2024) zur Identifikation relevanter Kollokationen und Cluster sowie anschließende Konkordanzanalysen und manuelle Kodierungen (Curtain/Dröge 2024). Dieser Mixed-Methods-Ansatz erlaubte es, sowohl thematische Schwerpunkte als auch die sprachlich-rhetorischen Mechanismen der Polarisierung sichtbar zu machen. Die zentralen Ergebnisse dieser Analysen sollen im Folgenden an Korpusbeispielen veranschaulicht werden.

Frames ideologischer Polarisierung

Brüggemann / Meyer (2023, S. 135) definieren Frames als interpretative Pakete, die eine zentrale Organisationsidee liefern und aufzeigen, worum es geht. Die in der Analyse identifizierten Frames zeigen, welche Aspekte der Genderdebatte von den Kommentierenden behandelt werden und welche epistemische Annahme sie hervorheben. Drei zentrale Frames prägen den Diskurs:

- 1) Das binäre Geschlechtermodell präsentiert Geschlecht als biologisch festgelegte Dichotomie (3), während Gegenpositionen auf die Vielfalt biologischer und sozialer Geschlechtsausprägungen verweisen (4).
- (3) Biologisch gesehen gibt es nur 2 Geschlechter. Man und Frau. XX und XY. [KYC/MR1.000408]
- (4) Es gibt mehr als zwei Geschlechter, weil Geschlecht ein riesiges Spektrum ist, dass mit den Kategorien Mann und Frau nicht hinlänglich beschrieben werden kann. [KYC/MR1.002123]

- 2) In der Gegenüberstellung ‚biologische versus soziale Geschlechtskonzepte‘ wird über das Verhältnis von Natur und Identität gestritten – zwischen biologischem Determinismus und sozialkonstruktivistischen Perspektiven (5).
- (5) Man darf vor Allem nicht Geschlechtsidentität mit dem biologischen Geschlecht gleichsetzen. [KYC/MAI.000494]
- 3) Schließlich zeigen epistemische Annahmen den Kampf z.B. um wissenschaftliche Autorität: Nur Biologie bzw. Naturwissenschaften gelten als legitime Autorität (6), sozialwissenschaftliche Ansätze werden hingegen als „ideologisch“ diskreditiert (7).
- (6) es gibt ja eine objektive Wahrheit und die liegt in der Wissenschaft der Biologie. Alles andere ist Wunschdenken und die Gefahr liegt darin, Kinder durch ideologischen Genderquatsch falsch zu informieren. [KFC/016.000266]
- (7) Desweiteren sind Gender-Studies keine ernstzunehmenden Wissenschaften. Es gibt zwei biologische Geschlechter. Das sind Mann und Frau. [KYC/DO1.001880]

Die Werkzeugkiste der Polarisierung

Besonders aufschlussreich sind die rhetorischen Strategien, mit denen diese Frames kommuniziert werden. Dabei werden die inhaltlichen Divergenzen durch spezifische rhetorische Mittel zusätzlich verstärkt.

VON BINÄREN FRAMES ZUR FEINDBILD-RHETORIK: DIE LINGUISTISCHE WERKZEUGKISTE DER GESELLSCHAFTLICHEN SPALTUNG

Zu den zentralen Strategien (Fortuna 2019) gehören zunächst Vereinfachungstechniken wie Dichotomien (siehe (3) „Man(n) und Frau. XX und XY.“) und „Wir-gegen-die-anderen“-Formulierungen (8). Hinzu kommen Übertreibungen und Verstärkungen der wahrgenommenen Unterschiede, die sich in Gut-Böse-Dichotomien (wie in moralisierender Sprache oder im religiösen Fundamentalismus (9) und (10) („legale Beziehungen zu stiften sind abnorme Bemühungen (...“)) sowie Übergeneralisierungen (10) manifestieren. Als Provokationstechniken dient der Einsatz emotional und mo-

ralisch aufgeladener Sprache in Form von Ad-hominem-Argumenten/Beleidigungen (siehe (9) „anti-wissenschaftlichen Spinnern“), Spott, Ironie und Sarkasmus.

- (8) Wenn **Transaktivist:innen** anfangen, biologische und evolutionäre Tatsachen zu leugnen, stehen sie auf der selben Stufe wie Kreationisten und Flat Earther. [KYC/DO1.001733]
- (9) **Das Leben von Trans-Personen ist schon schwer genug, die müssen nicht auch noch von anti-wissenschaftlichen Spinnern „vertreten“ werden.** [KYC/DO1.001733]
- (10) Ich akzeptiere diese **geringen Fälle** als gegeben, so wie andere Behinderungen auch, aber dieser Hype und alle Versuche daraus **legale Beziehungen zu stiften sind abnorme Bemühungen gottgewolltes Verhalten der Menschen zu unterwandern.** [KYC/JK1.000064]

KAMPF UM DIE ‚WAHRE WISSENSCHAFT‘: ZWISCHEN BINÄREN KATEGORIEN, MORALISCHER AUFLADUNG UND SYSTEMATISCHER DELEGITIMIERUNG DES GEGENÜBERS

Darüber hinaus werden Legitimationsstrategien eingesetzt, um die eigene Position zu stärken und gegnerische Meinungen zu diskreditieren. Hierzu zählen insbesondere die Invokation oder Negation wissenschaftlicher Autorität (11) sowie die strategische Aussage und Bewertung von Expertise.

- (11) **Die Biologin und Medizin-Nobelpreisträgerin (1995, Embryonalentwicklung) Christiane Nüsslein-Volhard** hat sich in die anhaltende Debatte um Geschlecht und Gender eingeschaltet. [KYC/US1.000144]

Fazit

Diskurse um Geschlecht und Gender werden, das ist intuitiv ersichtlich, auf eine polarisierte Art und Weise geführt bzw. sind Gegenstand der Polarisierung. Um jedoch etwas über die genauen sprachlichen Mechanismen aussagen zu können, die dabei eine Rolle spielen, benötigt es eine Datenba-

sis. Unser Beitrag hat gezeigt, wie wir in einem interdisziplinären Team aus der Rhetorik und Linguistik ein Korpus aufgebaut haben, mit dem sich solche Fragen bearbeiten lassen. An drei beispielhaften Analysen haben wir die Potenziale unserer unterschiedlichen Interessen und Herangehensweisen deutlich gemacht: Datengeleitete quantitative Verfahren können helfen, große Mengen an Korpusdaten zu überblicken und sprachliche Auffälligkeiten eines Diskurses bzw. ‚konkreten Korpus‘ freizulegen, denen anschließend in detaillierteren Analysen näher nachgegangen werden kann. Auch einzelne Strukturen wie rhetorische Fragen können mithilfe unseres Korpus untersucht und anhand der empirischen Belege im Detail beschrieben werden. Ein Mixed-Methods-Zugang aus quantitativen und qualitativen Verfahren hat wesentliche Strategien der Polarisierung herausgearbeitet, die von den Diskursteilnehmenden eingesetzt werden und Polarisierung sprachlich erzeugen. ■

Anmerkungen

- ¹ X (ehemals Twitter): Python-Bibliothek `ntscraper`; Facebook-Kommentare: Python-Bibliothek `facebook_scraper`; Instagram-Kommentare: Web-Scraping-Plattform `apify`; Reddit: R-Paket `RedditExtractoR` <<https://cran.r-project.org/package=RedditExtractoR>>; YouTube: R-Paket `vosonSML` (Gertzel et al. 2025); automatische Videotranskriptionen: automatische Untertitel von YouTube, Dropbox und TikTok, kostenlose Version von Happy Scribe.
- ² Wir danken unseren Hilfskräften an der Universität Tübingen, die uns bei den Transkriptionen unterstützt haben: Sarah Ebert, Leander Möllers, Amelie Rüeck, Isolde Sellin und Mira Tara.
- ³ Alle Tools sind open-source und unter einer freien Lizenz unter <<https://github.com/KorAP>> verfügbar.
- ⁴ Wir verwenden hier Daten von Twitter (Kupietz/Lüngen/Diewald 2023), das NottDeuYTSch-Korpus mit YouTube-Kommentaren (Cotgrove 2023) sowie Wikipedia-Nutzer*innen- und -Artikeldiskussionen (Margaretha et al. 2025).
- ⁵ Korpusbelege werden wie im Original abgedruckt.

Literatur

- Baker, Paul/Brookes, Gavin/Atanasova, Dimitrinka/Flint, Stuart W. (2020): Changing frames of obesity in the UK press 2008-2017. In: *Social Science & Medicine* 264, S. 1-9. <<https://doi.org/10.1016/j.socscimed.2020.113403>>.
- Bański, Piotr/Fischer, Peter M./Frick, Elena/Ketzan, Erik/Kupietz, Marc/Schnober, Carsten/Schonefeld, Oliver/Witt, Andreas (2012): The new IDS corpus analysis platform: challenges and prospects. In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul, Turkey: European Language Resources Association (ELRA), S. 2905-2911. <www.lrec-conf.org/proceedings/lrec2012/pdf/789_Paper.pdf>.
- Brüggemann, Michael/Meyer, Hendrik (2023): When debates break apart: discursive polarization as a multi-dimensional divergence emerging in and through communication. In: *Communication Theory*, 33, 2-3, S. 132-142. <<https://doi.org/10.1093/ct/qtad012>>.
- Bubenhofer, Noah (2009): Sprachgebrauchsmuster: Korpuslinguistik als Methode der Diskurs- und Kulturanalyse. (= Sprache und Wissen 4). Berlin: de Gruyter.
- Busse, Dietrich/Teubert, Wolfgang (1994): Ist Diskurs ein sprachwissenschaftliches Objekt? Zur Methodenfrage der historischen Semantik. In: Busse, Dietrich/Hermanns, Fritz/Teubert, Wolfgang (Hg.): *Begriffsgeschichte und Diskursgeschichte*. Opladen: Westdeutscher Verlag, S. 10-28.
- Caponigro, Ivano/Sprouse, Jon (2007): Rhetorical questions as questions. In: McNally, Louise/Puig-Waldmüller, Estela (Hg.): *Proceedings of Sinn und Bedeutung 11*. (= Sinn und Bedeutung 11). Barcelona: Universitat Pompeu Fabra, S. 121-133.
- Costa, Andressa (i. Ersch.): Diskursive Polarisierung in Online-Debatten über Zweigeschlechtlichkeit – Eine Analyse von Frames und Gruppenkonstruktion.
- Cotgrove, Louis (2023): The NottDeuYTSch Corpus: a corpus of German-language YouTube comments. In: *Korpora Deutsch als Fremdsprache* 3, 2, S. 225-229. <<https://doi.org/10.48694/KOR-DAF.3841>>.
- Cotgrove, Louis (2025): Entstehende Arten von Intensivierung in der digitalen Kommunikation junger Menschen. In: *Deutsche Sprache* 53, 1, S. 2-18. <<https://doi.org/10.37307/j.1868-775X.2025.01.03>>.
- Curtain, Colin/Dröge, Kai (2024): QualCoder. <<https://github.com/ccbogel/QualCoder/releases/tag/3.6>> (Stand: 15.7.2026).
- Feldmüller, Tim (2026): Das Framing von Extremismusvarianten im medialen Diskurs der Jahre 1999-2021: Eine corpus-driven Methode zur Erschließung und Visualisierung semantischer Frames. (= Frame-Based Approaches to Semantics 2). Berlin: Language Science Press.
- Feldmüller, Tim/Kupietz, Marc (i. Beg.): Word embeddings for discourse studies: distributional approaches to polarised discourses on sex and gender. In: Kalwa, Nina/Leßmöllmann, Annette (Hg.): *Epistemic wars in polarised discourse: in search of the lost common ground*. Berlin: Springer.
- Feldmüller, Tim/Schweden, Theresa (i. Ersch.): Sprachlich-konzeptueller Wandel im Behinderungsdiskurs der Jahre 1950-2023: Eine distributionell-semantische Analyse der Leitmedien ZEIT und SPIEGEL. In: *Zeitschrift für Germanistische Linguistik*.
- Fortuna, Alan (2019): Polarization: rhetorical strategies in the Tea Party Network. Berlin/Boston: de Gruyter. <<https://doi.org/10.1515/9783110664775>>.
- Gertzel, Bryan/Borquez, Francisca/Graham, Timothy/Ackland, Robert (2025): vosonSML: collecting social media data and generating networks for analysis. VOSON Lab, virtual observatory for the study of online networks. <<https://github.com/vosonlab/vosonSML>> (Stand: 15.7.2026).
- Han, Chung-hye (2002): Interpreting interrogatives as rhetorical questions. In: *Lingua* 112, 3, S. 201-229. <[https://doi.org/10.1016/S0024-3841\(01\)00044-4](https://doi.org/10.1016/S0024-3841(01)00044-4)>.
- Kupietz, Marc/Diewald, Nils/Margaretha, Eliza (2020): RKorAP-Client: an R package for accessing the German Reference Corpus DeReKo via KorAP. In: *Proceedings of the 12th Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association (ELRA), S. 7015-7021. <www.aclweb.org/anthology/2020.lrec-1.867> (Stand: 15.7.2026).
- Kupietz, Marc/Lüngen, Harald/Diewald, Nils (2023): Das Gesamtkonzept des Deutschen Referenzkorpus DeReKo: Vom Design bis zur Verwendung und darüber hinaus. In: Deppermann, Arnulf/Fandrych, Christian/Kupietz, Marc/Schmidt, Thomas (Hg.): *Korpora in der germanistischen Sprachwissenschaft*. Mündlich, schriftlich, multimedial. Berlin/Boston: de Gruyter, S. 1-28. <<https://doi.org/10.1515/978311085708-002>>.
- Kupietz, Marc/Lüngen, Harald/Kamocki, Paweł/Witt, Andreas (2018): The German Reference Corpus DeReKo: new developments – new opportunities. In: *Proceedings of the 11th International Conference on Language Resources and Evaluation*. Miyazaki, Japan: European Language Resources Association (ELRA), S. 4353-4360. <<https://aclanthology.org/L18-1689.pdf>> (Stand: 15.7.2026).

- Lüngen, Harald / Sperberg-McQueen, Christopher M. (2012): A TEI P5 document grammar for the IDS text model. In: *Journal of the Text Encoding Initiative* 3, S. 1-18.
- Margaretha, Eliza / Lüngen, Harald / Diewald, Nils / Kupietz, Marc / Yaddehige, Rameela (2025): Building and querying Wikipedia discussion corpora using KorAP. In: Fábíán, Annamária / Trost, Igor (Hg.): *Impulses and Approaches to Computer-Mediated Communication: Proceedings of the 12th International Conference on Computer Mediated Communication and Social Media Corpora for the Humanities (CMC 2025)*. Bayreuth: Universität Bayreuth, S. 123-124.
- Meibauer, Jörg (1986): *Rhetorische Fragen*. Berlin/New York: de Gruyter. <<https://doi.org/10.1515/9783111352572>>.
- Rapp, Birgit (i. Beg.): Rhetorical questions as a trigger for polarisation: YouTube comments in the gender discourse. In: Kalwa, Nina / Leßmöllmann, Annette (Hg.): *Epistemic wars in polarised discourse: in search of the lost common ground*. Berlin: Springer.
- Raue, Benjamin (2022): Text und Data Mining in Einrichtungen des Kulturerbes. In: *Recht und Zugang* 1, 1, S. 5-18.
- Sahlgren, Magnus (2008): The distributional hypothesis. In: *Rivista di Linguistica* 20, 1, S. 33-53.
- Scharloth, Joachim / Eugster, David / Bubenhofer, Noah (2013): Das Wuchern der Rhizome: Linguistische Diskursanalyse und Data-driven Turn. In: Busse, Dietrich / Teubert, Wolfgang (Hg.): *Linguistische Diskursanalyse: neue Perspektiven*. Wiesbaden: Springer Fachmedien Wiesbaden, S. 345-380. <https://doi.org/10.1007/978-3-531-18910-9_11>.
- Scott, Mike (2024): WordSmith Tools version 9 (64 bit version) Stroud: Lexical Analysis Software. <<https://lexically.net/downloads/version9/index.html>> (Stand: 15.7.2026).
- Scott, Suzanne (2002): Linguistic feature variation within disagreements: an empirical investigation. In: *Text – Interdisciplinary Journal for the Study of Discourse* 22, 2, S. 301-328. <<https://doi.org/10.1515/text.2002.011>>.
- Straka, Milan (2018): UDPipe 2.0 Prototype at CoNLL 2018 UD Shared Task. In: *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*. Brussels, Belgium: Association for Computational Linguistics, S. 197-207. <<https://doi.org/10.18653/v1/K18-2020>>.
- Yan, Xiaohui / Guo, Jiafeng / Lan, Yanyan / Cheng, Xueqi (2013): A biterm topic model for short texts. In: *Proceedings of the 22nd international conference on World Wide Web. WWW '13: 22nd International World Wide Web Conference*. Rio de Janeiro, Brazil: ACM, S. 1445-1456. <<https://doi.org/10.1145/2488388.2488514>>.
- Zhao, Jingbo / Song, Tingting / Song, Xiaoying / Bai, Yuanmeng (2022): Analysis on the linguistic features of conflict discourse in mathematical cooperation problem solving in China. In: *Frontiers in Psychology* 13, S. 1-16. <<https://doi.org/10.3389/fpsyg.2022.945909>>.
- Ziem, Alexander / Pentzold, Christian / Fraas, Claudia (2018): Medien-Frames als semantische Frames: Aspekte ihrer methodischen und analytischen Verschränkung am Beispiel der ‚Snowden-Affäre‘. In: Ziem, Alexander / Inderelst, Lars / Wulf, Detmer (Hg.): *Frames interdisziplinär: Modelle, Anwendungsfelder, Methoden*. (= *Proceedings in Language and Cognition* 2). Düsseldorf: Düsseldorf University Press, S. 155-182. <<https://doi.org/10.1515/9783110720372>>. ■