

6 Korpuserstellung

6.1 Zur Bedeutung von Korpora für die Diskursanalyse

Für eine Diskursanalyse im hier verfolgten Sinne, die sich nicht nur für „die Analyse der Seltenheit“ (Foucault 1973: 182), sondern auch für „die Analyse der Häufung“ (ebd.) interessiert, sind Korpora eine entscheidende Größe. Da Diskurse zwar als solche wahrgenommen, jedoch nicht unmittelbar überprüfbar analysiert werden können, bieten Textkorpora die Möglichkeit, zumindest „Teilmengen von Diskursen für die linguistische Analyse verfügbar zu machen“ (Niehr 2014: 33). Strenggenommen wird in einer Diskursanalyse also nicht unmittelbar der Diskurs analysiert, sondern ein Korpus, das möglichst zuverlässige Rückschlüsse auf einen bestimmten Diskurs zulässt. Dadurch entsteht eine gewisse Doppeldeutigkeit des Diskursbegriffs, die auch bei Foucault schon zu finden ist, wenn dieser etwa vom *Diskurs* im Singular (vgl. Foucault 1974a: 10), aber auch von *Diskursen* im Plural spricht (vgl. ebd.: 26): einerseits der Diskurs als omnipräsente Struktur der sozialen Konstruktion von Realität(en), andererseits Diskurse im synekdochischen Sinne totum pro parte für thematisch bestimmte Textgeflechte, als „Diskursausschnitte“ (Felder 2012a: 116) oder „Teildiskurse“ (Niehr 2014: 36). In diesem zweiten Sinne wird der Ausdruck *Diskurs* fortan verwendet.

Forschungspraktisch lassen sich Diskurse in diesem Sinne als virtuelle Textkorpora verstehen (Busse & Teubert 1994: 14), insofern ein analysierter Diskurs durch thematische, zeitliche und sonstige Definitionen (ebd.) ein Konstrukt des jeweils analysierenden Menschen darstellt, der somit „sein Untersuchungsobjekt selbst konstituiert“ (Niehr 2014: 32). Im Gegensatz zum konkreten Korpus werden in einem solchen virtuellen Korpus alle Diskursbeiträge im definierten Zeitraum, zum definierten Thema etc. mitgedacht. Allerdings lassen sich Diskurse nur in konkreten Korpora tatsächlich untersuchen, was diese einerseits unverzichtbar macht (Niehr 2014: 31). Andererseits zeichnen bei der Korpuserstellung zum Zwecke der Diskursanalyse die Interessen und Vorannahmen der Analysierenden die spätere Analyse zu einem gewissen Grad bereits vor (vgl. Niehr 2014: 30 f.; Busse & Teubert 1994: 14), was bei späteren Analyseschritten immer wieder kritisch reflektiert werden muss.

Aus diesem Grund gibt es gewisse Handreichungen für die Erstellung eines Korpus zum Zwecke der Diskursanalyse. Da ein Korpus als „Stichprobe für den gesamten Diskurs“ (Niehr 2014: 33) gelten soll, fordert z. B. Gries (2009: 7) vom konkreten Korpus eine gewisse Repräsentativität für den entsprechenden Diskurs. Um Verwirrungen mit Repräsentativität im statistischen Sinne zu vermeiden, die in diskurslinguistischen Untersuchungen nicht das Ziel ist (Niehr 2014:

34 f.), soll hier jedoch mit Busch (2007: 151 ff.) eher von Generalisierung oder Generalisierbarkeit statt von Repräsentativität gesprochen werden. Dabei geht es darum, im Korpus das Musterhafte des Diskurses greifbar zu machen, indem es viele und auch typische Fälle des zu untersuchenden Phänomens enthält und somit plausible Generalisierungen vom Korpus auf den Diskurs zulässt (Busch 2007: 153). Als weitere Gütekriterien nennt Busch (2007: 151 ff.) Validität und Reliabilität der Korpuserstellung und schließlich auch der Diskursanalyse. Diese Kriterien zielen darauf ab, die einzelnen Schritte der Korpuserstellung (s. 6.2–6.4) sowie der Diskursanalyse (s. 7) offenzulegen (Reliabilität) und dadurch überprüfbar, nachvollziehbar und replizierbar zu machen (vgl. Niehr 2014: 36), was in den folgenden Abschnitten und Kapiteln geschehen soll.

In der vorliegenden Analyse ergeben sich aus den theoretischen Forschungsinteressen und dem methodischen Ausgangspunkt spezielle Anforderungen an die Korpuserstellung. Die theoretische Fragestellung nach dem progressiven und subversiven Potenzial einzelner Ausdrücke im Diskurs (s. Teil II) sowie die gegenstandsbezogene Fragestellung nach der jeweiligen Rolle verschiedener Kuratoren und Kurationskulturen bei der intersubjektiven Aushandlung kulturellen Erbes (s. 5) lassen sich für die Analyse zusammenführen zur Frage nach dem kontingenten wechselseitigen Einfluss verschiedener „Diskursdomänen“ (Vogel et al. 2020: 207) aufeinander. Um die hier relevanten Diskursdomänen (bzw. Kurationskulturen; s. 5.1) Politik, Recht und Gesellschaft/Öffentlichkeit analysieren und miteinander vergleichen zu können, müssen also verschiedene Korpora für die diskursiven Kämpfe ums kulturelle Erbe ›Ehe‹ (und speziell für die Verdunkelung des Bedeutungsaspekts ›Verschiedengeschlechtlichkeit‹) in Politiksprache, Rechtssprache und Gemeinsprache erstellt werden. Damit wird die Korpuserstellung Gries' (2009: 7) Forderung gerecht, dass „the different parts of the linguistic variety you're interested in are all manifested in the corpus“ – wie etwa auch variationslinguistische Analysen verschiedene Varietäten über verschiedene Korpora miteinander vergleichen (Felder, Müller & Vogel 2012: 8). Kurz sei in diesem Zusammenhang bereits die besondere Rolle von Social-Media-Daten hervorgehoben (s. Twitter-Korpus in 6.2). Während die Bevorzugung leichter erschließbarer Korpora oftmals zu einer Konzentration auf die professionalisierte Presse- und Literatursprache sowie auf akademische Diskursausschnitte führt (Felder, Müller & Vogel 2012: 18), lassen sich durch Social-Media-Daten auch niederschwelligere Diskursbeiträge außerhalb der üblichen professionalisierten und akademischen Daten finden. Derartige Daten versprechen einen besseren analytischen Zugang zu den erwähnten Bottom-Up-Prozessen von Kuration und Aushandlung kulturellen Erbes, bei denen nicht-professionalisierte Diskursakteure noch nicht etablierte Perspektiven und Deutungen im Diskurs durchzusetzen versuchen (s. Barkhausen, Bloching & Maasböl 2021 in 5.1).

Die Art der Zusammenstellung der untersuchten Korpora hat also Einfluss darauf, was sich in der induktiven Analyse (s. 10) als signifikante Hinweise auf sprachliche Muster und Handlungsrouitinen ergeben wird. Zur Validität und Reliabilität der Analyse seien im Folgenden deren zugrundeliegenden Korpora und Subkorpora vorgestellt. Umfänglich wurde die Erstellung thematischer Korpora für das Teilprojekt *Ehe und Familie als kulturelles Erbe* bereits in (Bloching & Felder 2025) dargelegt. An dieser Stelle sollen darüber hinaus diejenigen Korpora vorgestellt werden, die für die vorliegende Analyse des Diskurses speziell um die gleichgeschlechtliche Ehe erstellt wurden, sowie deren jeweilige Relevanz für die vorgestellten Forschungsinteressen.

6.2 Gemeinsprachliche Korpora

Für die wohl am wenigsten spezifische und eher ex negativo verstandene Diskursdomäne der (im Gegensatz zu Politik und Recht nicht-institutionalisierten) „Gesellschaft/Öffentlichkeit“ wurden drei verschiedene gemeinsprachliche Korpora erstellt, die jeweils öffentlichkeitswirksame, doch im Vergleich zu Politik und Recht weniger professionalisierte und autorisierte Textformen enthalten. Die Kategorie *Gemeinsprache* kann mit Hilfe von Felders (2016) Modell greifbar gemacht werden als diejenige Varietät, die sich erstens durch einen Standarddekt auszeichnet, also keinen bestimmten Dialekt oder Regiolekt beinhaltet (vgl. Felder 2016: 134). Zweitens zeichnet sich die Gemeinsprache durch eine Alltagssemantik aus (vgl. Felder 2016: 135 f.) – im Gegensatz z. B. zur Fachsemantik der richterlichen Entscheidungstexte (s. 6.4). Beide definitorischen Merkmale müssen jedoch für das hier zugrundeliegende Verständnis von Gemeinsprache eher als hinreichende denn als notwendige Bedingungen verstanden und dahingehend flexibilisiert werden. So werden sowohl dialektale Elemente auf Twitter (bei Felder 2016: 135: „Alltagssprache“) als auch vermittlungssemantische Elemente in Zeitungstexten (bei Felder 2016: 134: „Standardsprache“) in die gemeinsprachlichen Korpora aufgenommen und als Teil der Diskursdomäne „Gesellschaft/Öffentlichkeit“ untersucht.

6.2.1 Benutzerdefiniertes Korpus im Deutschen Referenzkorpus (*G-DeReKo*)

Für Fragen nach gesellschaftlichen metasprachlichen Aushandlungen sowie nach der potenziellen Diffusion diskursiv durchgesetzter Ausdrucksformen von der Gemeinsprache in Politik- und Rechtssprache sind verschiedene Korpora veröffentlichten, nicht-reglementierten Sprachgebrauchs vonnöten. Das umfangreichste

solche Korpus stellt für die geschriebene deutsche Sprache das Archiv W des Deutschen Referenzkorpus (DeReKo) auf Cosmas II dar.⁴³ Das darin erstellte benutzerdefinierte Korpus sollte zweierlei Anforderungen genügen:

1. Textsorten-Spezifik: Das Korpus soll auf die diskursive Domäne der Gemeinsprache im hier verstandenen Sinne zugeschnitten sein – das heißt, es soll möglichst alle gemeinsprachlichen Texte des Archivs enthalten (Zeitungstexte, Wikipedia-Artikel, Sachbücher etc.) und möglichst keine nicht-gemeinsprachlichen Texte (Plenarprotokolle oder belletristische Texte).
2. Thematische Spezifik: Das Korpus soll auf das Thema ›gleichgeschlechtliche Ehe‹ zugeschnitten sein – das heißt, es soll möglichst alle Texte des Archivs enthalten, die sich mit diesem Thema befassen, und möglichst keine Texte enthalten, die dieses Thema nicht behandeln.

Die beiden Bedingungen wurden in zwei entsprechenden Schritten der Korpuserstellung erfüllt:

1. Erstellung eines benutzerdefinierten Korpus im Archiv W unter Ausschluss von Plenarprotokollen und belletristischen Quellen.⁴⁴
2. Thematisch bestimmte Suchanfrage innerhalb dieses gemeinsprachlichen Korpus. Dabei ergeben sich die einzelnen Ausdrücke der Suchanfrage in einem ersten Schritt aus der intuitiven Einschätzung potenziell relevanter Schlagwörter (vgl. Felder 2012a: 135) zweitens aus einer Erweiterung dieses Wortschatzes mithilfe des Dornseiff-Wörterbuches (Dornseiff 2020; vgl. Bubenhofer 2015: 499) und drittens aus einer Sichtung der Textbelege dieser ersten Suchwörter, die zu einer induktiven Erweiterung und Präzisierung der Suchanfrage beiträgt (vgl. Vogel et al. 2020: 212). In der Suchanfragesprache von Cosmas II lautet diese Suchanfrage¹ wie folgt:

Suchanfrage 1: ((Ehe ODER Ehen) /+1w für /+1w (alle ODER Alle)) ODER (Öffnung /+w1 der /+w1 Ehe) ODER Homosexuell*Ehe+ ODER Homosexuell*ehe+ ODER Homo+Ehe+ ODER Homo+ehe+ ODER Schwul*Ehe+ ODER Schwul*ehe+ ODER Lesb*Ehe+ ODER Lesb*ehe+ ODER ((Ehe ODER Ehen ODER Heirat* ODER heirat*) /+w0:4,s0 (*Gleichgeschlechtliche* ODER *gleichgeschlechtliche* ODER *Homosex* ODER *homosex* ODER *Lesb* ODER *lesb* ODER *Schwul* ODER *schwul*)) ODER ((*Gleichgeschlechtliche* ODER *gleichgeschlechtliche* ODER *Homosex* ODER *homosex* ODER *Lesb* ODER *lesb* ODER *Schwul* ODER *schwul*) /+w0:4,s0 (Ehe ODER Ehen ODER Heirat* ODER heirat*)) ODER ((Eingetragene*

⁴³ Dessen genaue Zusammensetzung ist einzusehen unter <https://www2.ids-mannheim.de/cosmas2/projekt/referenz/archive-alt/archive-umfang.html> (zuletzt aufgerufen am 16.01.2022).

⁴⁴ Konkret wurden all jene Quellen exkludiert, die auf Cosmas mit einem der folgenden Kürzel beginnen: BIO, DCK, DIV, GOE, GRI, GRI, KJL, LIM/L10, MK1/LBC-TPM, MK2/ERL-WF1, MLD, PB_, PH_, PMV, PNI, PNO, PNW, PRP, PSH, PSL, PSN, PST, PTH.

ODER eingetragene* ODER Anerkannte* ODER anerkannte*) /+1w (*Partnerschaft* ODER *partnechaft* ODER *Lebens+gemeinschaft*)⁴⁵

Aus den Texten innerhalb des gemeinsprachlichen benutzerdefinierten Korpus (s. o. 1.), die mindestens einen Treffer dieser Suchanfrage enthalten, wurde für die Analyse ein thematisch spezifisches Korpus mit folgenden zentralen Metadaten erstellt (s. Tabelle 4).

Tabelle 4: Metadaten des benutzerdefinierten Korpus im DeReKo (G-DeReKo).

Texte	26.007
Tokens	25.714.806
Zeitraum	1952–2021
häufigste Quellen	die tageszeitung, Süddeutsche Zeitung, Spiegel-Online, Die Presse, Rhein-Zeitung, Tages-Anzeiger, NZZ, Berliner Morgenpost ... ⁴⁶

Aus einigen der häufigsten Quellen des Korpus (*Die Presse*, *Tages-Anzeiger*, *NZZ*) ist bereits ersichtlich, dass DeReKo und so auch G-DeReKo nicht nur Zeitungstexte aus Deutschland, sondern auch deutschsprachige Texte aus Österreich und der Schweiz sowie auch aus Luxemburg enthält. Wie sich im empirischen Teil der Arbeit immer wieder zeigen wird (s. Teil IV), wirkt sich dieser Aspekt der Korpuszusammenstellung insbesondere auf die quantitativen Ergebnisse der Analyse aus. Zwar hätten die Texte aus Österreich, Luxemburg und der Schweiz auch aus dem Korpus ausgeschlossen werden können, um eine bessere Vergleichbarkeit mit den politik- und rechtssprachlichen Korpora gewährleisten zu können, die ebenfalls nur Texte aus Deutschland enthalten. Da ein solcher Ausschluss jedoch bei der Erstellung des Korpus G-Twitter nicht möglich war (s. u.), wurden diese Texte in die Untersuchung miteingeschlossen, um stattdessen eine Vergleichbarkeit zwischen diesen beiden gemeinsprachlichen Korpora zu ermöglichen. Dies ermöglichte ferner auch Vergleiche zwischen den Diskursen um die gleichgeschlechtliche Ehe in den verschiedenen deutschsprachigen Ländern (s. z. B. 8.2).

Abschließend soll eine methodische Schwäche dieses Korpus in Bezug auf dessen spätere Analyse reflektiert werden. Für die induktive Genese von Subthemen und agonalen Zentren (s. 7.4) sind Frequenz- und Keywordlisten ein unab-

⁴⁵ Bei dieser Suche muss auf Cosmas II die Suchmodalität „Diakritische Zeichen beachten“ ausgewählt werden sowie beide Suchmodalitäten zu „Groß- / Kleinschreibung beachten“.

⁴⁶ Die genaue Zusammensetzung kann wiederum unter <https://www2.ids-mannheim.de/cosmas2/projekt/referenz/archive-alt/archive-umfang.html> eingesehen werden. Für genauere Untersuchungen der Korpus-Zusammensetzung sowie für einen Nachvollzug der Analysen kann das erstellte Korpus auf Anfrage für jedem Cosmas II-User zugänglich gemacht werden.

dingbarer quantitativer Zugang zum Korpus und darüber zum Diskurs. Cosmas II stellt jedoch kein Tool für die Erstellung solcher Listen zur Verfügung, wie z. B. Sketch Engine oder CQPweb (s. u.) dies tun. Daher mussten mit einem aufwendigen Verfahren Frequenzlisten erstellt und aus diesen wiederum Keywordlisten errechnet werden, wie in Kapitel 7.1.2 erläutert werden soll.

In der restlichen Arbeit wird dieses Korpus auch mit *G-DeReKo* (für *gemeinsprachliches Korpus im Deutschen Referenzkorpus*) abgekürzt.

6.2.2 Die Zeitschrift *konkret* (G-Konkret)

Einen besonderen Bereich der gemeinsprachlichen Datengrundlage stellt ein eigens erstelltes Korpus aus Artikeln der Zeitschrift *konkret* dar. Die separate Untersuchung der *konkret* ist dadurch bedingt, dass die Zeitschrift auf Cosmas II nicht verfügbar ist und daher als eigenes Korpus analysierbar gemacht werden musste. Ferner erscheint ihre separate Analyse in Bezug auf die politische Perspektive auch lohnend: Da die *konkret* sich selbst als „einzige linke Publikumszeitschrift Deutschlands“⁴⁷ versteht, vom Verfassungsschutz (2004: 203) hingegen „im Spektrum des undogmatischen Linksextremismus“ verortet wird, ermöglicht ein Korpus ihrer Texte die Analyse auch innerhalb progressiver Diskursakteure einzigartiger Diskurspositionen.

Durch Sammlung von *konkret*-Texten, Annotation der Metadaten, Tokenisierung, Lemmatisierung und Part-of-speech-Tagging entsteht ein Gesamtkorpus mit den folgenden zentralen Metadaten (s. Tabelle 5).

Tabelle 5: Metadaten des Gesamtkorpus der Zeitschrift *konkret*.

Texte	15.500
Tokens	25.330.954
Zeitraum	1974–2009
Quelle	monatliche Ausgaben + Sonderausgaben der <i>konkret</i>

Aus diesem Gesamtkorpus an *konkret*-Ausgaben wurde auf der Korpusanalyse-Plattform CQPweb wiederum ein themenspezifisches Subkorpus erstellt, das möglichst alle Texte zum Thema ›gleichgeschlechtliche Ehe‹ enthält. Die hierfür verwendete Suchanfrageⁱⁱ, die mit der Suchanfrage zur Erstellung des Korpus auf Cosmas II möglichst gleichbedeutend sein sollte, stellt sich in der CQP-Language wie folgt dar:

⁴⁷ <https://www.konkret-magazin.de/ueber> (zuletzt aufgerufen am 16.01.2023).

Suchanfrage 2: ([word="Ehe.?"] [word="für"] [word="alle"]) | (([lemma="Ehe"] [lemma="für | zwischen"] [lemma="."*gleichgeschlechtlich.*"%c"]) | (([lemma="Ehe"] [lemma="für | zwischen"] [lemma="."*homosex.*"%c"]) | (([lemma="Ehe"] [lemma="für | zwischen"] [lemma="."*lesb.*"%c"]) | (([lemma="Ehe"] [lemma="für | zwischen"] [lemma="."*schwul.*"%c"]) | ((word="Homo.? Ehe.?"%c)) | (word="Homosex.+Ehe.?"%c)) | ([lemma="homosexuell"] [lemma="Ehe"]) | ([lemma="gleichgeschlechtlich"] [lemma="Ehe"]) | ([lemma="schwul.+Ehe"%c] | ([lemma="schwul"] [lemma="Ehe"]))) | ([lemma="lesb.+Ehe"%c] | ([lemma="lesbisch"] [lemma="Ehe"]))) | ([lemma="eingetragen | anerkannt"%c] [lemma="."*partnerschaft"%c]) | ([lemma="eingetragen | anerkannt"%c] [lemma="Lebensgemeinschaft"%c]) | ([lemma="Öffnung"] [word="der"] [lemma="Ehe"])

Aus den Texten des Gesamtkorpus, die mindestens einen Treffer dieser Suchanfrage enthalten, wurde ein thematisch spezifisches Subkorpus mit den folgenden zentralen Metadaten erstellt (s. Tabelle 6).

Tabelle 6: Metadaten des themenspezifischen Subkorpus der Zeitschrift konkret (G-Konkret).

Texte	30
Tokens	54.426
Zeitraum	1986–2009
Quelle	monatliche Ausgaben + Sonderausgaben der konkret

Auch für das Gesamt- und das Subkorpus aus Texten der *konkret* müssen abschließend deren Schwachstellen für die korpuslinguistische Untersuchung festgehalten werden. Zunächst fällt der begrenzte zeitliche Umfang der verfügbaren Texte ins Auge: Da ab 2009 keine Artikel der *konkret* mehr verfügbar sind, lässt sich der (Teil-)Diskurs um das Gesetz zur Einführung des Rechts auf Eheschließung für Personen gleichen Geschlechts (in der Gemeinsprache eher bekannt als *Ehe für alle*) mithilfe dieses Korpus nicht untersuchen. Zumindest der (Teil-)Diskurs um das Lebenspartnerschaftsgesetzes 2001 dürfte mit diesem Zeitraum jedoch abgedeckt werden. Ferner muss für quantitative Vergleiche der sehr geringe Umfang des thematisch spezifischen Subkorpus bedacht werden: Nur 30 der Texte scheinen sich mit dem Thema ›gleichgeschlechtliche Ehe‹ zu befassen. Aufgrund seines geringeren Umfangs und des begrenzten verfügbaren Zeitraums werden sich in der Analyse für dieses Korpus meist nur wenige Treffer finden, weshalb es meist nicht für statistische Auswertungen zu Rate gezogen wird, sondern stattdessen für die qualitative Analyse einzelner Textbelege (s. etwa 9).

In der restlichen Arbeit wird dieses Korpus auch mit *G-Konkret* (für *Gemeinsprachliches Korpus der Zeitschrift konkret*) abgekürzt.

6.2.3 Twitter (*G-Twitter*)

Für eine möglichst umfassende Analyse der in einem Diskurs zirkulierenden Meinungsäußerungen und sprachlichen Muster stellen Daten aus den sozialen Medien eine wichtige Ergänzung dar. In 6.1 wurde bereits erwähnt, dass Korpora aus Social-Media-Posts Zeitungskorpora entscheidend ergänzen können. Dies betrifft nicht zuletzt die Textsorte und die Art der Textgenese, die sich in sozialen Medien in aller Regel dialogischer gestaltet als in Zeitungstexten. Derartige Kommentarverläufe (Threads) sind vor allem in Bezug auf die Frage nach der potenziellen Wirkung bestimmter Äußerungen oder einzelner Ausdrücke auf Rezipierende von Interesse. Indem verschiedene User auf denselben Post antworten, geben sie ihr jeweiliges Verständnis des Posts zu erkennen und machen so im Thread verschiedene Lesarten derselben Äußerung auf der Sprachoberfläche analysierbar.⁴⁸ Während bei Zeitungs- und ähnlichen Texten die Dimension der subjektiven Sinnstiftung durch den Rezipienten (vgl. 2.1.2, 2.2) der Analyse verborgen bleibt, lässt sich in den Kommentarstrukturen sozialer Medien die kommunikative (also interaktive) Sinnstiftung mit Verweis auf die subjektiven Eindrücke der Kommunikationspartner selbst analysieren – ganz im Sinne der „procedural consequentiality“ in der Gesprächsanalyse (vgl. Deppermann 2001: 70). Somit kann die Analyse von Twitter-Daten und insbesondere von interaktionalen Kommentar-Threads einen geforderten „Impuls [...] für eine stärker pragmatisch-interaktionale Ausrichtung der Diskurssemantik insgesamt leisten“ (Roth 2018: 364).

Speziell die Kurztexte (sog. Tweets) des Mikroblogging-Dienstes Twitter⁴⁹ eröffnen dabei auch neue Perspektiven auf politolinguistische Fragestellungen, indem sie die politischen Texte von Profis aus Politik und Journalismus um Diskursbeiträge der wählenden Bevölkerung (vgl. Burkhardt 1996: 96) – und auch von politischen Laien (vgl. Eckerlin 2021: 16) – ergänzen. Dabei ist festzuhalten, dass Tweets keineswegs eine belanglose oder politisch epiphänomenale Textsorte darstellen. Für den politischen Einfluss von Twitter (bzw. die politische Einflussnahme über Twitter) finden sich zahlreiche Beispiele wie Obamas Twitter-

⁴⁸ Vgl. hierzu Kommentare als „Sekundärtexte“ im Sinne von (Fiske 1991) in (Bloching 2020).

⁴⁹ Da die Daten vor der Umbenennung von Twitter in *X* gesammelt und analysiert wurden, wird hier weiterhin der Name *Twitter* verwendet, um so der Zeitlichkeit der Daten gerecht zu werden.

Kampagne 2008 und Trumps PR auf Twitter 2016 (Jungherr 2015: 16), aber auch Twitters (umstrittene) Rolle bei subversiven Bottom-Up-Bewegungen wie dem arabischen Frühling 2011 (Pfaffenberger 2016: 14). Speziell in Bezug auf die Frage nach individuellen lexikalischen Einflussmöglichkeiten auf diskursiven und kulturellen Wandel scheint Twitter somit vielversprechend.

Vor allem aber machen Tweets die Perspektiven einer großen Menge von Diskursakteuren der Analyse zugänglich, die in Zeitungstexten aus verschiedenen Gründen nicht vertreten sind. Das stark akademisch geprägte journalistische Feld, in dem „soziale Herkunft eine bedeutende Rolle“ spielt (Lueg 2012: 41), kann somit durch Perspektiven aus einem Feld ergänzt werden, in dem weit weniger durch soziales Kapital und Gate-Keeping vorsortiert wird. Die Verbindung aus dieser geringeren Professionalisierung und der Möglichkeit, anonym die eigene Perspektive in den Diskurs einzubringen, erhöht im Vergleich zu überarbeiteten Zeitungstexten die Wahrscheinlichkeit, auch potenziell stigmatisierte Positionen abseits der Mehrheitsmeinung zu finden (vgl. 3.3). Andererseits ist dieser Nutzen von Twitter-Daten für die Analyse auch wieder einzuschränken, da auch Twitter-Nutzende keineswegs einen repräsentativen Querschnitt durch die Bevölkerung darstellen. Hölzig bezeichnet Twitter-User vielmehr als eine „meinungsstarke Minderheit“ (Hölzig 2018) und fasst die Ergebnisse seiner Befragung derselben wie folgt zusammen:

Die Ergebnisse deuten darauf hin, dass sich aktive Twitterer in demographischer, politischer und vor allem in Hinsicht der Persönlichkeitsmerkmale vom Durchschnitt der Internetnutzer unterscheiden. Das Stimmungsbild auf Twitter wird im Vergleich eher von Persönlichkeiten geprägt, die höhere Werte in der Tendenz zum Narzissmus aufweisen, die persönlichkeitsstärker, extrovertierter und weniger ängstlich sind. Eine Orientierung an den Meinungen und Trends auf Twitter kann daher verzerrte Relevanzrahmen und Stimmungsbilder suggerieren, welche mit denen der Gesamtbevölkerung nur wenig zu tun haben. (Hölzig 2018: 140)

Vor diesem Hintergrund kritisiert Hölzig die Tendenz in Politik und Journalismus, Twitter „als informelles Stimmungsbarometer öffentlicher Meinung“ zu nutzen (Hölzig 2018: 140). Auch die bloßen Zahlen der Twitter-Nutzenden mahnen zur Vorsicht: In Deutschland hat Twitter etwa 8 Millionen User und steht damit weit hinter den meist genutzten sozialen Medien Facebook (47 Mio.) und Instagram (32 Mio.) zurück – aber auch hinter Plattformen wie Xing, Pinterest, Tiktok, Reddit und LinkedIn.⁵⁰

Der große forschungspraktische Vorteil von Twitter gegenüber anderen sozialen Medien ist jedoch, dass Twitter einen *Academic Research access* gewährt,

50 <https://www.blog2social.com/de/blog/social-media-nutzer/>.

der es erlaubt, über eine API (eine Programmierschnittstelle) bis zu 10 Millionen Tweets samt Metadaten im Monat herunterzuladen. Mit Hilfe dieser API wurden in der gesamten Datenbank von Twitter alle deutschen Tweets gesucht und heruntergeladen, die eine Nähe zum Thema ›gleichgeschlechtliche Ehe‹ aufweisen. Da die Twitter-API mit einer eigenen Suchanfragesprache operiert, wurden die Suchanfragen von Cosmas II bzw. CQPweb (s. o.) möglichst gleichbedeutend umformuliert⁵¹:

Suchanfrage 3: (((Ehe OR Ehen) für alle) OR (Homo-Ehe OR Homo-Ehen OR Homoehe OR Homoehen OR "Homo Ehe" OR "Homo Ehen") OR (Homosexuellen-Ehe OR (homosexuelle OR homosexuellen OR homosexueller) (Ehe OR Ehen)) OR Homosexuellenehe) OR (gleichgeschlechtliche OR gleichgeschlechtlichen OR gleichgeschlechtlicher (Ehe OR Ehen)) OR (((schwule OR schwulen OR schwuler)(Ehe OR Ehen)) OR (Schwulen-Ehe OR Schwulen-Ehen) OR (Schwulenehe OR Schwulenehen)) OR (((lesbische OR lesbischen OR lesbischer) (Ehe OR Ehen))) OR (Lesben-Ehe OR Lesben-Ehen) OR (Lesbenehe OR Lesbenehen) OR ((eingetragene OR eingetragenen OR eingetragener OR anerkannte OR anerkannten OR anerkannter)(Lebenspartnerschaft OR Lebenspartnerschaften OR Partnerschaft OR Partnerschaften OR Lebensgemeinschaft OR Lebensgemeinschaften OR Gemeinschaft OR Gemeinschaften)) OR (Öffnung der (Ehe OR Ehen))) lang:de -is:retweet⁵¹

Aus allen Tweets, die mindestens eines der gesuchten Wörter oder Phraseme enthalten und entsprechend die ›gleichgeschlechtliche Ehe‹ thematisieren, wurde ein Korpus mit den folgenden zentralen Metadaten erstellt (s. Tabelle 7).⁵²

Tabelle 7: Metadaten des themenspezifischen Korpus aus Tweets (G-Twitter).

Texte	267.918
Tokens	5.309.059
Zeitraum	Aug. 2007–Mai 2021
Quelle	einzelne deutschsprachige Beiträge des Mikroblogging-Dienstes Twitter

Wie bei G-DeReKo bereits erwähnt, setzt sich auch das Twitter-Korpus nicht nur aus Tweets aus Deutschland, sondern aus deutschsprachigen Tweets überhaupt zusammen, da sich die Suche zwar auf die automatisiert erkannte Sprache des

⁵¹ Über „lang:de“ wird die Suche auf Treffer beschränkt, die Twitter selbst (sehr zuverlässig) als deutschsprachige Treffer kategorisiert hat. Durch „-is:retweet“ werden Retweets (also bloße Kopien anderer Tweets auf die eigene Timeline), die im Korpus zu Duplikaten führen würden, ausgeschlossen.

⁵² Das Twitter-Korpus ist über die kollaborative Lehr- und Forschungsplattform *Discourse Lab* der Universitäten Heidelberg und Darmstadt zugänglich: <https://www.discourselab.de>.

Tweets, nicht aber auf Nationalgrenzen einschränken lässt. Entsprechend werden beim Twitter-Korpus auch Texte außerhalb Deutschlands, insbesondere aus Österreich und der Schweiz bei den quantitativen Analysen ins Gewicht fallen, wobei sich auch Unterschiede zwischen den Diskursen in Deutschland, Österreich und der Schweiz abzeichnen werden (s. z. B. 8.2).

In der restlichen Arbeit wird dieses Korpus auch mit *G-Twitter* (für *Gemeinsprachliches Korpus mit Daten aus Twitter*) abgekürzt.

Innerhalb dieses Korpus wurde wiederum ein Subkorpus erstellt aus Tweets derjenigen Accounts, die sich in ihrem Namen explizit als *LGBTQ+* oder *queer* bezeichnen. Hierfür wurden manuell alle relevanten im Korpus versprachlichten Nutzernamen – genauer gesagt: die einmaligen Twitter-Handles – gesammelt, die durch folgende Suchanfrage^{iv} in der CQP-Language gefunden wurden:

Suchanfrage 4: [word="@.*(LGB|queer).*" %c]

Zusätzlich wurde das Subkorpus induktiv um solche Accounts ergänzt, die sich im Diskurs selbst durch queer-aktivistische Arbeit hervortun und die sich auch abseits des Namens, insbesondere in der eigenen Account-Beschreibung als *queer*, *schwul* oder *lesbisch* identifizieren. Ein solches Subkorpus soll verhindern, dass die Diskursanalyse zu einem bloßen Reden-über verkommt (vgl. 4.2.2.7) und gerade die Stimmen der Betroffenen bei der Analyse in der Menge untergehen.⁵³ Ferner kann in diesem Subkorpus speziell die Perspektive einer aktivistischen queeren Szene isoliert untersucht und mit anderen Perspektiven verglichen werden. Es ist jedoch wichtig festzuhalten, dass diese Unterscheidung keineswegs eine Differenzierung zwischen heterosexuellen und nicht-heterosexuellen Akteuren darstellt. Erstens sind auch in allen anderen Korpora – rein stochastisch qua ihrer Größe – nicht nur heterosexuelle Akteure vertreten; und zweitens ist davon auszugehen, dass die in einem solchen Subkorpus analysierte aktivistische queere Community wie schon in den 90ern „immer nur einen Teil der Lesben und Schwulen“ ausmacht (Wegner 1995: 182). Um also eine essentialistische Fremdzuschreibung sexueller Orientierung zu vermeiden, beschränkt sich das Subkorpus auf Accounts mit entsprechenden Selbstzuschreibungen. Das Subkorpus verfügt über die folgenden zentralen Metadaten (s. Tabelle 8).

In der restlichen Arbeit wird dieses Korpus auch mit *G-Sub-Twitter* (für *Gemeinsprachliches queer-aktivistisches Subkorpus mit Daten aus Twitter*) abgekürzt.

⁵³ Für diesen Hinweis danke ich Ingo Warnke.

Tabelle 8: Metadaten des queer-aktivistischen Subkorpus auf Twitter (G-Sub-Twitter).

Texte	2.447
Tokens	46.377
Zeitraum	Nov. 2008–Mai 2021
Quelle	Einzelne Beiträge des Mikroblogging-Dienstes Twitter

Zuletzt seien auch für das Twitter-Korpus (samt Subkorpus) Schwächen und Herausforderungen für die Analyse reflektiert. Zunächst muss die einschränkende Umgebung von Twitter für die Analyse beachtet werden: Einzelne Tweets waren bis Ende 2017 auf 140 Zeichen begrenzt, seitdem auf 280 (vgl. Jungherr 2015: 13). Daraus ergibt sich ein besonderes Verhältnis zwischen Text- und Tokenzahl, die sich auch in der Signifikanz der Keywords niederschlagen wird. Des Weiteren lässt sich der Zeitraum um die Einführung der eingetragenen Lebenspartnerschaft 2001 auf Twitter nicht untersuchen, da Twitter erst 2006 gegründet wurde. In dieser Hinsicht ergänzen jedoch Twitter- und konkret-Korpus einander, da ersteres Diskursbeiträge ab 2007, letzteres solche bis 2009 enthält. Ein zentrales Problem in Twitter-Korpora sind Bots – Algorithmus-gesteuerte Fake-Accounts, die automatisiert und wiederholt Posts verfassen und mit keiner einfachen technischen Lösung aussortiert werden können (vgl. Jungherr 2015: 25). Im Zuge der Analyse lassen sich Bots jedoch meist aufgrund ihres Post-Verhaltens erahnen, oftmals anhand einer Beschreibung verifizieren und auf dieser Grundlage von den Ergebnissen ausschließen.⁵⁴ Eine schwerer zu umschiffende Grauzone stellen genuine Tweets (keine Retweets) von unterschiedlichen Accounts dar, die einen derart identischen Wortlaut aufweisen, dass ein Copy-and-Paste-Vorgehen naheliegt. Da derartige Tweets in gewisser Weise eine technische Fortführung der in Diskursen ohnehin vorherrschenden Musterhaftigkeit und Imitation von Sprachgebrauch darstellen und nach wie vor die Perspektive oder Diskursposition der jeweiligen Twitter-Nutzenden indizieren können, werden sie von der Analyse nicht gänzlich ausgeschlossen. Stellenweise muss dieses Phänomen jedoch für die Deutung der Aussagekraft reiner Frequenzwerte berücksichtigt werden.

⁵⁴ Z. B. finden sich im vorliegenden Twitter-Korpus 11.365 Tweets mit dem Muster „Wer ‚Homo-Ehe‘ sagt, muss auch ‚Homo-[Nomen]‘ sagen.“ Sie alle stammen von demselben Account, der in seiner Account-Beschreibung auf Twitter unter anderem das Hashtag #botALLY angibt. Auf dieser Grundlage kann der Account von der Analyse ausgeschlossen werden.

6.3 Politiksprachliches Korpus (*P-Bundestag*)

Neben der gemeinsprachlichen Diskursdomäne ist ein weiterer spezifischer Bereich des zu untersuchenden Diskurses, zumal dieser sich um politisch durchzusetzenden rechtlichen und auch kulturellen Wandel dreht, die Domäne der Politik. Durch politische Mediensprache, die in gemeinsprachlichen Korpora wie Cosmas II untersucht werden kann (Bopp 2013: 324), ist die Politiksprache im engeren Sinne auch mit einem Sprechen über Politik im Austausch (Burkhardt 1996: 81). Nach Burkhardt (1996: 79 ff.) setzt sich die Politiksprache in diesem Sinne zusammen aus der institutionsinternen Sprache in der Politik einerseits und der an die wählende Bevölkerung gerichteten Politikersprache andererseits. Auch wenn für die Fragestellung der vorliegenden Untersuchung prinzipiell beide Formen der Politiksprache von Interesse sind, wird im nachfolgend beschriebenen Korpus der Plenarprotokolle des Deutschen Bundestages vor allem die Sprache in der Politik untersucht (vgl. Bopp 2013: 322).

Die Plenarprotokolle des Deutschen Bundestages sind in grundständig annotierter Form bereits über die Forschungsplattform *Discourse Lab* zugänglich, die mit CQPweb operiert. Entsprechend konnten mit der bekannten Suchanfrage in der CQP-Language (s. o. Suchanfrage 2) all diejenigen Redebeiträge gefunden werden, die das zu untersuchende Thema ›gleichgeschlechtliche Ehe‹ behandeln oder zumindest erwähnen. Aus diesen Redebeiträgen in den Bundestags-Plenarprotokollen entsteht ein Subkorpus mit den folgenden zentralen Metadaten (s. Tabelle 9).

Tabelle 9: Metadaten des themenspezifischen Subkorpus in den Plenarprotokollen (P-Bundestag).

Texte	453
Tokens	453.713
Zeitraum	Sep. 1992–Apr. 2021
Quelle	Plenarprotokolle des Deutschen Bundestags

Im Vergleich zu den gemeinsprachlichen Korpora aus Tweets und *konkret*-Artikeln weist das Subkorpus in den Plenarprotokollen den Vorteil auf, dass es über den gesamten zentralen Zeitraum des Diskurses über die Ehe zwischen gleichgeschlechtlichen Paaren Daten bereithält.

In der restlichen Arbeit wird dieses Korpus auch mit *P-Bundestag* (für *Politiksprachliches Korpus aus Plenarprotokollen des Deutschen Bundestags*) abgekürzt.

6.4 Rechtssprachliches Korpus (*R-Entscheidungen*)

Die letzte untersuchte Diskursdomäne ist die des Rechts. Für eine erkenntnisstiftende Untersuchung kulturellen Wandels spielt diese Domäne eine besondere Rolle, da in ihr höchstrichterliche Kuratoren kulturellen Erbes den Wandel desselben mit größter denkbarer intersubjektiv anerkannter Autorität (Validität) und handlungsleitender Geltung (Faktizität) konsolidieren. Sprich: Wenn ein Konzept-, Bedeutungs- oder Kulturwandel sich in der Diskursdomäne des Rechts etabliert, kann er als gesellschaftlich durchgesetzt angesehen werden (s. 5). Auf der anderen Seite kann das Recht auch eine beschleunigende Rolle bei kulturellem Wandel spielen und Transformationsprozesse kulturellen Erbes dynamisieren (s. 5.2; vgl. Felder 2018a: 275 f.). Eine Analyse des rechtlichen Wandels in der Analyse der Rechtssprache soll entsprechend dem Umstand gerecht werden, dass „die gesellschaftliche Handlungsfähigkeit im Recht an die Sprachlichkeit des Rechts gebunden ist, weil die Rechtssprache und Rechtskommunikation die Schnittstelle zwischen rechtlichem Wissen und seinen regulativen Funktionen in Gesellschaft und Staat darstellen“ (Ekkehard und Friedemann 2017: XIF.). In Anlehnung an die GfDS soll *Rechtssprache* verstanden werden als „eine Fachsprache, die in den Bereichen Justiz, Verwaltung und privatem Rechtsverkehr zur optimalen Verständigung unter Fachleuten verwendet wird. Sie umfasst sowohl die Sprache der rechtlichen Normen [...] als auch die Verwaltungssprache“⁵⁵, wobei für die vorliegende Untersuchung insbesondere die Sprache der Rechtsprechung von Relevanz ist.

Für die Untersuchung solcher judikativen Zubereitungs- und Dynamisierungsprozesse wurde ein Korpus aus gerichtlichen Entscheidungen rund um das Thema ›gleichgeschlechtliche Ehe‹ erstellt, das auch alle hierfür relevanten Entscheidungen des Bundesverfassungsgerichts enthält. Zur Suche der entsprechenden Entscheidungen in den hierfür einschlägigen juristischen Datenbanken wurden genau diejenigen Suchausdrücke und Mehrwort-Verbindungen verwendet, die auch die bisherigen Suchanfragen (1–3) konstituierten.^v

Suchanfrage 5: "Ehe für alle" ODER "Ehe für gleichgeschlechtliche" ODER "Ehe zwischen gleichgeschlechtlichen" ODER "Ehe für Personen gleichen Geschlechts" ODER "Ehe zwischen Personen gleichen Geschlechts" ODER "Ehe für homosexuelle" ODER "Ehe zwischen homosexuellen" ODER "Homoehe" ODER "Homo-Ehe" ODER "Homosexuelle Ehe" ODER "Gleichgeschlechtliche Ehe" ODER "Öffnung der Ehe" ODER "Eheöffnung" ODER "Eheöffnungsgesetz" ODER "Eingetragene Lebenspartnerschaft" ODER "Eingetragene Lebensgemeinschaft" ODER "Anerkannte Lebenspartnerschaft" ODER "Anerkannte Lebensgemeinschaft"

⁵⁵ <https://gfds.de/verstaendliche-rechtssprache/>; zuletzt aufgerufen am 20.11.2023.

Aus denjenigen Entscheidungen, die mindestens einen dieser Treffer enthalten, wurde ein Korpus mit den folgenden zentralen Metadaten erstellt (s. Tabelle 10).

Tabelle 10: Metadaten des themenspezifischen Korpus aus Entscheidungen (R-Entscheidungen).

Texte	790
Tokens	3.554.986
Zeitraum	Jun. 1973–Aug. 2021
Quelle	Entscheidungen verschiedener deutscher Gerichte

In der restlichen Arbeit wird dieses Korpus auch mit *R-Entscheidungen* (für *Rechtssprachliches Korpus aus gerichtlichen Entscheidungen*) abgekürzt.

Damit sind für alle zu untersuchenden Diskursdomänen die analysierten Korpora vorgestellt. Im Folgenden sollen die konkreten Analyseschritte für die Untersuchung dieser Korpora aufgezeigt werden.