

7 Methoden der quantitativen und qualitativen Diskursanalyse

Im vorigen Kapitel 6 wurden die Korpora vorgestellt, die verschiedene qualitative und quantitative Analysen und darüber möglichst generalisierbare Erkenntnisse über die abstrakte Größe des Diskurses um die gleichgeschlechtliche Ehe ermöglichen sollen. Wollen Erkenntnisse über den Diskurs gewonnen werden, gelingt dies nur über eine an der sprachlichen Oberfläche ausgerichteten Analyse dieser Korpora, denn schon bei Foucault muss der Nachvollzug der Wissenskonstitutionen und sonstigen Wirkmächte des Diskurses an den wahrnehmbaren Größen nachgezeichnet werden, in denen sie wirken: „Eine diskursive Formation zu analysieren, heißt also, eine Menge von sprachlichen Performanzen auf der Ebene der Aussagen und der Form der Positivität, von der sie charakterisiert werden, zu behandeln“ (Foucault 1973: 182). Und so halten auch Felder, Müller & Vogel (2012: 12) für ihre korpuspragmatische linguistische Diskursanalyse fest: „In Diskursanalysen nach Foucault geht es meist darum, soziokognitive Gegebenheiten wie kollektive Einstellungen, Wissensformationen, und Mentalitäten oder gesellschaftliche Asymmetrien an der Analyse sprachlicher Muster auszuweisen“. In diesem Kapitel soll entsprechend ein diskursanalytisches Untersuchungsprogramm vorgestellt werden, mit dem unter Zuhilfenahme verschiedenster diskurs- und korpuslinguistischer Methoden derartige sprachliche Muster induktiv erarbeitet (s. 7.1 und 7.2), kontextspezifisch im Hinblick auf ihre pragmatische Charakteristik analysiert (s. 7.3) und anschließend in Form kategorialer Typen von sprachlichen Mustern, die für ein bestimmtes thematisches (s. 8), pragmatisches oder auch metapragmatisches (s. 9–10) Phänomen stehen (s. 7.4), entlang verschiedener Parameter wie Zeit, Diskursdomäne und Akteursgruppe quantifiziert werden können (s. 7.5).

Abbildung 8 veranschaulicht die Zusammenhänge der einzelnen Schritte der hier gewählten Methodik, die in den folgenden Teilkapiteln einzeln genauer vorgestellt werden. Die klassischen Unterscheidungen zwischen korpusbasiert (bzw. *corpus driven*) und korpusgestützt (bzw. *corpus based*) sowie vor allem zwischen qualitativ und quantitativ erweisen sich hier oftmals als schwer haltbar. Denn bei der konkreten Analyse greifen stets beide Dimensionen ineinander, sodass die eine ohne die andere gar nicht erkennbar wird. Etwa richten sich auch quantitative Verfahren immer auf Elemente einer bestimmten, hermeneutisch erschlossenen Qualität und stellen auch auf der Makroebene durchaus hermeneutische Verfahren dar (vgl. Bender & Jacob 2024). Umgekehrt sind auch die hier durchgeführten qualitativen Analysen stets „quantitativ-informiert“ (Bubenhof 2013: 110), behandeln selbst

eine bestimmte Quantität von Belegen und werden nach Möglichkeit noch weiter quantifiziert (s. 7.5). Die im Folgenden vorgestellten einzelnen Analyseschritte sind also in gewissem Maße das Ergebnis einer künstlichen Trennung, die die vielschichtige Analyse nachvollziehbarer machen soll, bei der letztlich immer sowohl die qualitative als auch die quantitative Dimension, wenn auch in je unterschiedlichem Maße, relevant sind.

Insgesamt soll dieses Bündel an einander ergänzenden Verfahren Erkenntnisse darüber liefern, welche sprachlichen Muster und kommunikativen Routinen im Diskurs um die gleichgeschlechtliche Ehe besonders relevant sind und wie diese mit bestimmten den Diskurs prägenden Wissensbeständen, politischen Haltungen, Bewertungen, Präsuppositionen, Wahrnehmungs-, Denk-, und Argumentationsmustern etc. zusammenhängen, um in diesem Komplex wiederum die diskursive Rolle und das Wirkpotenzial einzelner umkämpfter Wörter einordnen zu können.

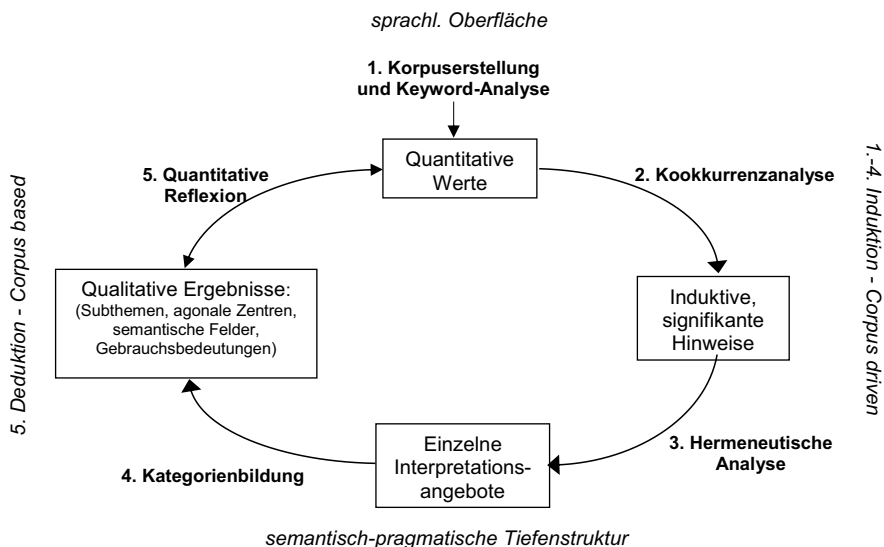


Abbildung 8: Methodische Schritte der quantitativ-qualitativen Diskursanalyse.

7.1 Induktive Genese quantitativer Werte

In diesem ersten Analyseschritt wird induktiv und abseits theoretischer Vorannahmen (die jedoch bereits in die Korpuserstellung geflossen sind; vgl. 6.1) an die themenspezifischen Korpora und Subkorpora herangetreten.⁵⁶ Eine solche induktive Herangehensweise an das Korpus wird auch als *korpusbasiert* oder *corpus-driven* bezeichnet (vgl. Tognini-Bonelli 2001; Bubenhofer 2009; Spitzmüller & Warnke 2011: 38 f.). Standardmethoden hierfür sind semi-automatische Analysen von

- Frequenzlisten (s. 7.1.1),
- Keywordlisten (s. 7.1.2) und
- N-Gramm-Listen (s. 7.1.3).

Sie finden sowohl in der linguistischen Korpus- und Diskursanalyse Anwendung (vgl. Teubert 2005; Felder 2012a: 135; Bubenhofer 2015: 487 ff.; Felder 2018b: 30 f.) als auch in der Lexikometrie, die sich auf die Diskurstheorie von Mouffe und Laclau (s. 3.2.1, 3.3) bezieht (vgl. Glasze 2008: 201 f.; Amri-Henkel 2021: 110 f.). Im Folgenden soll vorgestellt werden, wie diese Verfahren allgemein funktionieren und in der konkreten Analyse an den zu untersuchenden Korpora angewandt wurden.

7.1.1 Absolute Frequenzen

Das basalste quantitative Verfahren für einen corpus-driven Zugang zum Diskurs ist die Analyse von Frequenzlisten, also der häufigsten Wörter im jeweiligen Korpus. Mit CQPweb lassen sich solche Frequenzlisten für die Korpora *G-Konkret*, *G-Twitter*, *G-Sub-Twitter*, *P-Bundestag* und *R-Entscheidungen* automatisiert erstellen und lemmatisieren. Da Cosmas II jedoch keine vergleichbare Funktion bereithält, musste für das Korpus *G-DeReKo* eine solche Frequenzliste mit einem aufwändigeren Verfahren erstellt werden. Dengler (2022) hat eine Regex-basierte Suchanfrage für Cosmas entwickelt, die jede Wortform im Korpus findet.^{vi} „Durch Zusammenfügen der nach Frequenz sortierten Ergebnisse nach Wort-Types mit aktivierter Lemmatisierung erhält man so die entsprechende Frequenzliste“ (Dengler 2022: 16). Die somit für alle fünf Korpora erstellten Frequenzlisten wurden jeweils bis zum einhundertsten Treffer auf autosemantische und potenziell relevante Lexeme abgesehen. Für die Entscheidung, ob ein Lexem relevant in dem Sinne ist, dass es ein

⁵⁶ Für die sorgsame Genese von Frequenz-, Keyword- und N-Gramm-Listen danke ich Janine Dengler.

semantisches oder gar pragmatisches Phänomen indizieren könnte, ist wiederum eine „kontextsensitive[] Interpretation“ (Felder 2018b: 31) derselben in Form von hermeneutischen Analyse (s. 7.3) und ggf. Kookkurrenzanalysen (s. 7.2) vonnöten.

7.1.2 Relative Frequenzen: Keywords

Die absolute Häufigkeit zeugt jedoch noch nicht von der „diskursspezifischen Auffälligkeit eines Lexems“ (Felder 2012a: 144), die für einen quantitativ-induktiven Zugang von Interesse ist. Für solche lexikalischen „Charakteristika“ (Glasze 2008: 202) eines Diskurses sind relative Häufigkeiten in Form von Keywords von Belang.⁵⁷ Keywords sind „Ausdrücke, die im Vergleich zu einem Referenzkorpus signifikant häufiger vorkommen und außerdem als diskurs-charakterisierend eingeschätzt werden“ (Felder 2012a: 135). Diese Signifikanz wird im Log-Likelihood-Value angegeben.⁵⁸ Zur Erstellung von Keyword-Listen sind also Vergleichskorpora nötig, die die errechnete Relevanz der Keywords und somit deren „Keyness“ entscheidend mitbeeinflussen. Wenn zusätzlich zum themenspezifischen (Sub-)Korpus ein großes Korpus zur Verfügung steht, das alle sonstigen Variablen des untersuchten Korpus teilt (Textsorte, Zeitraum etc.), nicht aber die zu untersuchende Themenspezifika, so bietet es sich an, dieses Korpus als Referenzkorpus zu nehmen. Die themenspezifischen Subkorpora in vollständig verfügbaren Gesamtkorpora konnten zu diesem Zweck invertiert werden, sodass komplementär zum untersuchten Korpus alle übrigen Texte des Gesamtkorpus als Referenzkorpus dienen. Wo dies nicht möglich war, wurde möglichst große und unspezifische Korpora als Referenzkorpora gewählt. In Tabelle 11 sollen die Referenzkorpora transparent gemacht werden, die für die errechneten Keywords der (Sub-)Korpora jeweils den Vergleichswert lieferten.

57 Felder (2012a: 135) spricht in diesem Zusammenhang auch von „Schlüsselwörter[n]“. Dieser Ausdruck wird jedoch auch in verschiedenen politolinguistischen Bedeutungen gebraucht (vgl. Spitzmüller & Warnke 2011: 142), die mit dem Terminus technicus *Keywords* vermieden werden. Im Einklang mit der Terminologie englischsprachiger Korpusanalyse-Programme wie CQPweb, aber auch mit englisch- sowie deutschsprachigen Korpusanalysen (vgl. Bubenhofer & Schneider 2010; Gries 2021; Felder & Stegmeier 2012) – auch speziell zu Diskursen um die *Gay Marriage* (vgl. Bachmann 2011; Kania 2020) – soll im Folgenden also ausschließlich von *Keywords* gesprochen werden.

58 Dieses Maß dient hier jedoch nur zur groben Orientierung und zur relativen Sortierung von Treffern, da reine Signifikanzwerte im statistischen Sinne vor allem bei großen Korpora schnell in die Irre führen können: „The level of significance itself is less useful as language is not completely random“ (Manning & Schütze 2003: 166). Im Zuge einer hermeneutisch ausgerichteten Diskursanalyse ist also vielmehr eine Sortierung nach relativem Log-Likelihood-Value insofern von Bedeutung, als sie relative Hinweise auf Musterhaftigkeit und Generalisierbarkeit gibt (vgl. 6.1).

Tabelle 11: Referenzkorpora für die Berechnung von Keywords der jeweils untersuchten Korpora.

untersuchtes Korpus	Referenzkorpus
<i>G-DeReKo</i>	<i>DeReWo</i> ⁵⁹
<i>G-Konkret</i>	invertiertes Subkorpus
<i>G-Twitter</i>	<i>Public frequency list: Referenzkorpus Deutschland</i>
<i>G-Sub-Twitter</i>	invertiertes Subkorpus
<i>P-Bundestag</i>	invertiertes Subkorpus
<i>R-Entscheidungen</i>	<i>Public frequency list: Referenzkorpus Deutschland</i>

Für die Korpora *G-Twitter* und *R-Entscheidungen* ergibt der Vergleich mit einem Referenzkorpus, das Texte einer anderen Textsorte enthält (nämlich ausschließlich Zeitungstexte) strenggenommen eine Ungenauigkeit. Hierdurch werden in der Keyword-Liste nicht nur solche Lexeme als signifikant ausgegeben, die charakteristisch für das Diskursthema sind, sondern auch solche, die charakteristisch für die jeweilige Textsorte (Tweet respektive Entscheidungstext) sind.⁶⁰ Derartige Keywords müssen bei der Durchsicht der Keyword-Listen und der jeweiligen Belege ausgeschlossen werden.

Detaillierter ist ferner die Referenz für Lexem-Häufigkeiten in *G-DeReKo* zu beschreiben. Da Cosmas II für die (benutzerdefinierten) Korpora weder Frequenz- noch Keyword-Listen bereitstellt, wurde die nach Dengler (2022) selbst erstellte Frequenzliste (s. o.) stattdessen mit der korpusbasierten Wortgrundformenliste *DeReWo* verglichen, die auf dem gesamten *DeReKo* basiert (zu diesem Verfahren vgl. Felder & Stegmeier 2012: 382 f.). Hierfür wurde zunächst für jedes Lexem der Frequenzliste eine Häufigkeitsklasse errechnet, die die Frequenz des jeweiligen Lexems in Relation setzt zur Frequenz des häufigsten Lexems im selben Korpus und somit zur Korpusgröße.⁶¹ Die so errechnete Häufigkeitsklasse wurde für jedes Wort im Korpus abgeglichen mit dessen Häufigkeitsklasse im *DeReWo*. Für derart lange Wortlisten war ein solcher exhaustiver Vergleich nur durch ein maschinelles Ver-

⁵⁹ Korpusbasierte Wortgrundformenliste *DeReWo*, v-ww-bll-320000g-2012-12-31-1.0, mit Benutzerdokumentation, <http://www.ids-mannheim.de/derewo>, © Institut für Deutsche Sprache, Programmbereich Korpuslinguistik, Mannheim, Deutschland, 2013.

⁶⁰ Bspw. erzielen für *R-Entscheidungen* auch Token wie *Abs.*, *Kläger*, *Art.*, *GG*, *Urteil* etc. und für *G-Twitter* auch Token wie *via*, *ich* sowie diverse Satz- und Sonderzeichen jeweils hohe Loglikelihood-Values.

⁶¹ Die genaue Häufigkeitsklasse (kh) N wird dabei aus der absoluten Frequenz (f) errechnet und wie folgt definiert: $N = \text{hk}(\text{wort}) = \lfloor \log_2(f(\text{häufigstes_wort})/f(\text{wort})) + 0,5 \rfloor$ (*DeReWo* 2013: 13).

fahren anzustellen.⁶² Da mit diesem Verfahren nicht nur für einen Teil (vgl. Felder & Stegmeier 2012: 382 f.), sondern für alle Wörter im Korpus die jeweilige diskursive Charakteristik errechnet wird, lassen sich so auch stärker induktiv, ohne theoretische Vorannahmen diskursspezifische Lexeme erschließen – ähnlich wie mit automatisch ermittelten Keywords.⁶³ Zur Einheitlichkeit werden daher im Folgenden auch die Lexeme, die in diesem Verfahren als diskursiv charakteristisch ermittelt wurden, als *Keywords* bezeichnet. Da Keywords in Hinsicht ihrer diskurscharakteristischen Verweisfunktion aussagekräftiger sind als bloße Frequenzlisten, wurden für jedes Korpus die 250 Keywords mit dem höchsten Log-likelihood-Value (resp. der niedrigsten Häufigkeitsklasse im Vergleich zum DeReWo, s. o.) und einer Mindestfrequenz von 3 gesichtet und nach potenziell relevanten Lexemen abgesehen.

7.1.3 N-Gramme

Einen weiteren induktiven, jedoch translexikalischen Zugang zum Diskurs bietet die Analyse von N-Grammen. N-Gramme sind Syntagmen aus unmittelbar aufeinanderfolgenden Wörtern, die aus 2-n Wörtern bestehen können (Bigramme, Trigramme, ..., n-Gramme) (vgl. Bubenhofer 2009: 118). Wenn ein solches N-Gramm besonders häufig in einem Korpus auftaucht, lassen sich darüber musterhafte Strukturen im Korpus und somit im Diskurs ausfindig machen (Bubenhofer 2009: 111). In der vorliegenden Analyse konnten frequente N-Gramme auf CQPweb durch einfache Suchanfragen für eine Folge von beliebigen Token („[]“) induktiv erschlossen werden.⁶⁴

⁶² Das hierfür geschriebene simple Script sucht die beiden Lemmalisten (die selbsterstellte sowie die Liste des DeReWo) nach übereinstimmenden Lemmata ab. Diese werden in einer einzelnen Spalte aufgelistet, daneben ihre jeweilige Häufigkeitsklasse im themenspezifischen Korpus und im DeReWo. Aus letzteren wird schließlich die Differenz errechnet, anhand derer die so entstandene Liste sortiert werden kann (z. B. in Excel).

⁶³ Sowohl für die so erstellten Lemmalisten mit Häufigkeitsklassen als auch für die von CQPweb ausgegebenen Keyword-Listen muss jedoch bedacht werden, dass sie auf der automatischen Lemmatisierung des jeweiligen Analyseprogramms aufbauen. Da diese nicht unfehlbar ist, können sich so auch fehlerhafte Lemmata und andere Wortformen in den Keyword-Listen ergeben, die manuell ausgeschlossen werden müssen.

⁶⁴ Da auf Cosmas II nicht nach beliebigen Token gesucht werden kann, konnte das Verfahren dort nicht angewandt werden. Hier müssen die über CQPweb induktiv erarbeiteten n-Gramme in einem späteren Schritt deduktiv untersucht werden (s. 7.5).

Bigramme: [][]

Trigramme: [][][]

...

Heptagramme: [][][][][][]

Mit diesen Suchanfragen wurden für jedes themenspezifische Korpus – außer G-DeReKo – Listen mit den häufigsten Wortfolgen bestehend aus 2–7 Wörtern generiert. Die Beschränkung auf maximal 7 aufeinanderfolgende Wörter (sog. Heptagramme) ergibt sich aus der begrenzten Rechenleistung von CQPweb, die ab einer gewissen Korpusgröße keine umfangreicheren Suchanfragen verarbeiten kann. In jeder solchen Liste wurden die 250 häufigsten N-Gramme auf Indizien hin überprüft, sofern diese eine Mindestfrequenz von 5 Vorkommen hatten. In einem zweiten Schritt wurden die jeweiligen N-Gramm-Listen (also alle Bi- bis Heptagramme) nach den häufigsten Folgen von Wortarten (kurz auch POS für „part of speech“) ausgegeben. Mit solchen POS-Grammen lassen sich auch abstraktere Muster morphosyntaktischer Art aufspüren (Bubenhofer 2015: 487 f.; vgl. auch jüngste Ansätze der Diskursgrammatik in Bloching, Kiemes & Kückelhaus 2022). In jeder solchen POS-Gramm-Liste unterschiedlicher Länge für jedes Korpus wurden jeweils die fünf häufigsten POS-Gramme auf ihre konkreten lexikalischen Ausformulierungen hin untersucht – also je eine Liste für die fünf häufigsten POS-Bigramme, eine für die fünf häufigsten POS-Trigramme usw. Ein Beispiel zur Veranschaulichung: Eines der häufigsten POS-Pentagramme im Korpus *G-Twitter* ist eine Wortfolge mit den folgenden POS-Tags ART NN APPR ART NN (△Artikel, Nomen, Präposition, Artikel, Nomen). Wird nach dieser Abfolge von Wortarten gesucht, ergibt sich eine Liste an frequenten Pentagrammen, die diese syntaktische Gemeinsamkeit haben, jedoch semantisch sowie pragmatisch je ganz unterschiedliche Hinweise liefern können.⁶⁵ Mit diesen N-Gramm-Listen wurde wiederum verfahren wie mit den syntaktisch unbestimmten Wortfolgen aus dem ersten Schritt: Die 100 häufigsten N-Gramme wurden nach Indizien durchsucht, sofern sie eine absolute Frequenz von mindestens 5 Vorkommen hatten.

Wie für Frequenzlisten bereits erwähnt, ist auch für die Entscheidung der potenziellen Relevanz eines N-Gramms oder Keywords i. d.R. eine „kontextsensitive[] Interpretation“ (Felder 2018b: 31) von dessen Gebrauch im Diskurs vonnöten. Jedoch indizieren v. a. längere N-Gramme, aber auch einige Keywords recht unmittelbar und sehr zuverlässig bestimmte Themenbereiche oder Subthemen im Diskurs, sodass durch eine Sichtung der Belege (s. 7.3) metaphorische Verwendungsweisen und

⁶⁵ Um nur einige Beispiele zu nennen: *eine Niederlage für die Menschheit, die Debatte um die Homo-Ehe, die Vergewaltigung in der Ehe, ein Recht auf ein Kind, ein Aufruf an die Kirchen, eine Schande für das Haus, die Homo-Ehe in der Rechtsordnung ...*

Ähnliches schnell ausgeschlossen werden können. Andere Keywords und kurze N-Gramme können hingegen stärker ambig sein, sodass ihre verschiedenen Bedeutungspotenziale über ihre häufigsten Verwendungsweisen erschlossen werden müssen. Hinweise auf Musterhaftigkeiten im disambiguierenden Kotext⁶⁶ bietet hier eine quantitative Analyse desselben – in Form einer Kookkurrenzanalyse.

7.2 Kookkurrenzanalyse

Bereits in den theoretischen Ausführungen (v. a. 2) wurde problematisiert, dass sich die Bedeutung eines Wortes nicht einfach aus der Wortform erschließen oder gar stabil in dieser fassen lässt. Dasselbe gilt also auch für die in 7.1 ermittelten diskursspezifischen Indikatoren, aber auch für metadiskursiv umkämpfte Ausdrücke sowie alle anderen im Zuge der Diskursanalyse untersuchten lexikalischen Phänomene. Die theoretische Folgerung aus dieser immerwährenden Kontingenz von Wortbedeutung war, dass diese sich im konkreten Gebrauch, durch kotextuelle sowie weitere kontextuelle Faktoren ergibt. Um ebendiesen bedeutungsstiftenden Gebrauchskontext linguistisch untersuchen zu können, muss ein Zugang zu diesem gefunden werden, der sich nicht auf Willkür, sondern auf Signifikanz und Musterhaftigkeit stützt. Einen solchen Zugang bieten Verfahren der Kookkurrenzanalyse.

Als Kookkurrenzen eines Suchwortes X wird das Ko-Vorkommen anderer Ausdrücke bezeichnet, die in einem Korpus in der unmittelbaren Nähe von X auftauchen. Von analytischem Interesse wird die Kookkurrenzen zweier Ausdrücke aber i. d.R. erst dann, wenn sie „häufiger zu beobachten ist, als bei einer Zufallsverteilung aller Wörter zu erwarten wäre“ (Engelberg 2009: 18). Wenn im Folgenden von Kookkurrenzen die Rede ist, so sind also signifikant auftretende Kookkurrenzen gemeint. Genau wie die Signifikanz einer bestimmten Worthäufigkeit in einem Korpus wird auch die Signifikanz der Häufigkeit einer Kookkurrenz zweier Wörter im Log-Likelihood-Value angegeben (Bubenhofer 2015: 495 f.). Bei überzufällig häufig kookkurrierenden Wörtern wird oftmals auch von *Kollokationen* gesprochen. In einem engeren Verständnis zeichnen sich Kollokationen jedoch auch durch eine bestimmte syntaktische Beziehung der beiden Kollokationen zueinander aus (vgl. Lemnitzer & Zinsmeister 2006: 148; Engelberg 2009: 18;

⁶⁶ Unter *Kotext* wird im Folgenden die Art von Kontext verstanden, die sich auf der rein sprachlichen Wortumgebung befindet – in Abgrenzung etwa zu Situationskontext, historischem Kontext etc. (vgl. Reisigl & Wodak 2009: 93; Spitzmüller & Warnke 2011: 105).

Bubenhofers 2009: 114), sodass hier der Oberbegriff *Kookkurrenz* vorzuziehen ist (vgl. Felder 2012a). Die Wortumgebung, in der nach Kookkurrenzpartnern gesucht wird, kann entsprechend des Erkenntnisinteresses frei gewählt werden. Meist wird der Suchradius auf 3, 5 oder 10 Wörter links oder rechts vom Suchwort festgelegt (vgl. Baker 2006: 100), oder auch durch syntaktische Einschränkungen wie die Satz- oder gar Phrasenebene bestimmt (vgl. Bubenhofers 2015: 494).

Für das Erkenntnisinteresse der vorliegenden Untersuchung ist die Kookkurrenzanalyse deshalb von hervorgehobener Bedeutung, weil sie hilft, die „musterhafte Verwendungsweise“ (Bubenhofers 2015: 486) eines Wortes im jeweiligen Korpus zu ermitteln. Wie in den theoretischen Ausführungen (s. 2) immer wieder betont wurde, ergibt sich die Bedeutung eines Wortes aus seinem konkreten Gebrauch, über den sie, sofern dieser regelmäßig genug ist, auch konventionalisiert wird. Mit Saussure gesprochen: „Was ein Zeichen an Vorstellung oder Lautmaterial enthält, ist weniger wichtig als das, was in Gestalt der anderen Zeichen um dieses herum gelagert ist“ (zitiert nach Hofbauer 1995: 138). Daraus folgt, dass über die Kookkurrenzen eines sprachlichen Zeichens analysiert werden kann, in welcher Bedeutung dieses Zeichen im jeweiligen (Sub-)Korpus konkret verwendet wird (vgl. Bubenhofers 2015: 491 ff.). Können solche Bedeutungsfixierungsversuche (s. 3.2.2) für verschiedene Akteursgruppen, Zeiträume etc. unterschieden werden, so lässt sich aufzeigen, dass dieselbe Zeichenform nicht von jedem Akteur zu jedem Zeitraum in derselben Bedeutung verwendet wird (vgl. Felder 2017: 47).

Vor allem wegen dieses letzten Zusammenhangs ist die Kookkurrenzanalyse nicht nur vor dem Hintergrund (post-)strukturalistischer Theorien im Allgemeinen von besonderem Interesse, sondern auch für die Anwendung der Diskurstheorie nach Laclau & Mouffe im Speziellen (s. 3.2.1.4). Artikulationen nach Laclau & Mouffe (1991: 55) stellen Bedeutungsfixierungsversuche in dem Sinne dar, dass beim Artikulieren das „Gleiten der Bedeutung“ (Münker & Roesler 2012: 180) auf der Makroebene des Sprachsystems für die Mikroebene des Sprachgebrauchs kurzfristig angehalten werden muss, um „überhaupt eine Sache vertreten zu können“ (ebd.: 181). „Die Unmöglichkeit einer endgültigen Fixierung bedeutet gleichzeitig, dass immer neue partielle, temporäre Fixierungen möglich werden“ (Glasze 2008: 190). Diese temporäre Fixierung der Bedeutung (eines Wortes) vollzieht sich lokal über ein paradigmatisches, aber auch über ein syntagmatisches In-Bezug-Setzen, das sich im komplexeren referenziellen Kontext, aber auch im unmittelbaren sprachlichen Kontext des Wortes entfaltet. Glasze (2008: 195) schlägt daher zur korpusanalytischen Operationalisierung der Diskurstheorie nach Laclau und Mouffe „lexikometrische Verfahren“ vor, um dieses kotextuelle In-Bezug-Setzen auf der sprachlichen Oberfläche analysieren und quantifizieren zu können. Methodisch sind lexikometrische Verfahren dabei mit Kookkurrenzanalysen im Allgemeinen zu vergleichen, ihr Erkenntnisinteresse bezieht sich jedoch speziell auf Bedeutungsfixierungen des un-

tersuchten Wortes, indem sie „untersuchen, wie Bedeutungen durch Beziehungen zwischen lexikalischen Elementen hergestellt werden“ (Glasze 2008: 195). Für die Untersuchung von Bedeutungskonkurrenzen (s. 3.2.2) lohnen sich vor diesem Hintergrund insbesondere lexikometrische Analysen der Gebrauchsweisen eines umkämpften Ausdrucks zu verschiedenen Diskurszeiträumen, um das diachrone Gleiten der Bedeutung nachzuzeichnen (vgl. Glasze 2008: 200), sowie bei verschiedenen Diskursakteuren, um das Besetzen der Begriffe aus verschiedenen politischen Haltungen herauszuarbeiten (vgl. Glasze 2008: 203).

Aus all den genannten Gründen spielt die Methode der Kookkurrenzanalyse auch für die vorliegende Untersuchung eine bedeutende Rolle. Konkret wurden für drei Arten von untersuchten Wörtern die Kookkurrenzen analysiert:

- Kookkurrenzen von Keywords (s. 7.2.1)
- Kookkurrenzen von adversativen und konzessiven Konnektoren (s. 7.2.2)
- Kookkurrenzen von metadiskursiv umstrittenen Bezeichnungen (s. 7.2.3)

7.2.1 Kookkurrenzen von Keywords

Wie in 7.1 beschrieben, muss die genaue Bedeutung der Keywords für den untersuchten Diskurs unter Berücksichtigung ihrer jeweiligen Kontexte ermittelt werden. Die Kookkurrenzanalyse liefert einen quantitativen Zugang zu musterhaften Kontexten, in denen das jeweilige Keyword auftaucht (vgl. Felder 2018b: 30; Felder & Stegmeier 2012: 380). Gerade bei sehr häufigen Keywords kann eine Kookkurrenzanalyse effizient Hinweise darauf geben, aus welchem Grund das jeweilige Lexem für den untersuchten Diskurs charakteristisch ist, indem die Äußerungen, in denen das Lexem musterhaft verwendet wird, aufgezeigt werden. Auf diese Weise stellen Kookkurrenzanalysen (zusammen mit qualitativen, hermeneutischen Analysen, s. 7.3) eine hilfreiche Ergänzung für Keyword-Analysen dar; sie bilden eine Brücke vom induktiven Zugang in Richtung erkenntnisstiftender Kategorien. In der vorliegenden Analyse wurden Kookkurrenzanalysen kursorisch angewandt und mit qualitativen Konkordanzanalysen (s. 7.3) kombiniert. Hierbei konnten erstens Konstruktionen für idiomatische und andere „uneigentliche“ Verwendungsweisen des Keywords aufgespürt werden, die die Zuverlässigkeit der Indikatorfunktion für eine bestimmte Kategorie unterminieren würden.⁶⁷ Zweitens lieferte die Analyse der Kookkurrenzen genauere Hinweise auf der Mehrwortebene für detailliertere

⁶⁷ Soll zum Beispiel das Keyword *Gott* später quantifizierbarer Indikator für den Themenbereich ›Religion/Moral‹ dienen (s. 7.4, 7.5), so müssen Konstruktionen wie *oh (mein) Gott, um Gottes Willen* und *Gott sei Dank* explizit ausgeschlossen werden. Entdeckt wurden diese Konstruktionen über die Analyse der Kookkurrenzen von *Gott*.

Subthemen und z. T. auch pragmatische Kategorien (s. 7.4). So kann eine Analyse der Kookkurrenzen von Keywords die N-Gramm-Analyse um weitere aussagekräftige Mehrwort-Verbindungen ergänzen – auch in Form „diskontinuierlichen N-Grammen“ (Bubenhofer 2009: 118), die beliebige Elemente zwischen den kookkurrierenden Wörtern zulassen.

7.2.2 Kookkurrenzen von adversativen und konzessiven Konnektoren

Einen besonderen Zugang zum Diskurs bietet die Analyse der Kookkurrenzpartner von adversativen und konzessiven Konnektoren, die einen inhaltlichen Kontrast zwischen den umliegenden Aussagen signalisieren.⁶⁸ Prototypische Beispiele sind *aber*, *allerdings*, *obwohl* und *trotzdem*. Auf Satzebene reflektieren adversative und konzessive Konnektoren metadiskursiv zentrale Streitfragen des Diskurses, die Felder (2012a; 2015; 2018b) *agonale Zentren* nennt. Entsprechend finden sich in der sprachlichen Umgebung der Konnektoren die beiden einander gegenübergestellten Aussagen dieser agonalen Zentren, die Felder (ebd.) wiederum handlungsleitende Konzepte nennt. Er fasst zusammen:

Durch die Untersuchung der adversativen und konzessiven Konnektoren-Kotexte nach signifikanten Lexemen entstehen extensional geprägte onomasiologische Vernetzungen, die auf handlungsleitende Konzepte zurückgeführt werden können und dadurch operationalisierbar sind. (Felder 2012a: 138)

Konkret wird hierfür eine Suchanfrage formuliert, die möglichst viele adversative und konzessive Konnektoren beinhaltet, jedoch allzu ambige Wortformen, die zu falschen Positiven führen würden, ausschließt. Felder (2012a: 154) verwendet hierzu eine Liste an adversativen und konzessiven Konnektoren von Grammis, deren potenzielle Fehltreffer er hermeneutisch ausschließt. Schedl und Münch präzisieren diese Liste jeweils, indem sie ambige Wortformen wie *wo*, *dabei*, *während*, *nur*, *dafür*, *allein*, *einmal*, *wieder* (Schedl 2017: 10 f.), *dafür dass*, *einmal* und *bald ... bald* (Münch 2018: 11) aus der Suchanfrage ausschließen und somit schon vor der hermeneutischen Analyse signifikantere Ergebnisse erzielen. Aufbauend auf ihr Verfahren lässt sich im bspw. für das themenspezifische Korpus *G-DeReKo* in der Sprache von Cosmas II die folgende Suchanfrage^{vii,viii} formulieren:

Suchanfrage 6: indessen ODER indes ODER aber ODER jedoch ODER dahingegen ODER wiederum ODER einesteils ODER andernteils ODER wohingegen ODER alldieweil ODER hinwiederum ODER hinwieder ODER einerseits ODER hingegen ODER wogegen ODER

68 Vgl. <https://grammis.ids-mannheim.de/terminologie/1865> (zuletzt aufgerufen am 16.02.2023).

derweil ODER derweilen ODER demgegenüber ODER freilich ODER andererseits ODER allerdings ODER dementgegen ODER obgleich ODER nichtsdestoweniger ODER wiewohl ODER sosehr ODER wennschon ODER obwohl ODER wenngleich ODER obzwar ODER nichtsdestotrotz ODER gleichwohl ODER dennoch ODER trotzdem ODER obschon ODER wobei ODER nichtsdestominder ODER zum einen ODER nur dass ODER nur bloß ODER wenn auch ODER unbeschadet dessen ODER dessen ungeachtet⁶⁹

Nachdem eine solche Menge an adversativen und konzessiven Konnektoren in jedem Korpus gefunden wurde, wurden aus dem Ko-Text von 10 Wörtern vor und nach dem Konnektor (vgl. Lemnitzer & Zinsmeister 2006: 147 f.) Kookkurrenzprofile erstellt (vgl. Felder 2012a: 155). Die Kookkurrenzprofile der adversativen und konzessiven Konnektoren wurden wiederum nach Hinweisen auf umstrittene Positionierungen im Diskurs (handlungsleitende Konzepte) abgesucht – in der Form signifikant häufiger Autosemantika (vgl. Felder 2012a: 156). So konnten aus den quantitativ-statistischen Hinweisen der Kookkurrenzprofile im Zuge der qualitativ-hermeneutischen Analyse der Einzelbelege (s. 7.3) schließlich handlungsleitende Konzepte und agonale Zentren als Interpretationskonstrukte (Felder 2018b) abgeleitet werden (s. 7.4).

7.2.3 Kookkurrenzen konkurrierender Bezeichnungsalternativen

Zuletzt wurden detailliert die Kookkurrenzen der verschiedenen Bezeichnungsalternativen für das Diskursthema ›gleichgeschlechtliche Ehe‹ untersucht. Diese sind für die vorliegende Untersuchung von besonderer Relevanz, weil mit ihnen das theoretische Erkenntnisinteresse an den diskursiven Möglichkeiten der Bezeichnungs- und Bedeutungsfixierung systematisch verfolgt werden kann. Im theoretischen Teil der Arbeit wurde einerseits dargelegt, wie Akteure versuchen, ihre spezifischen Perspektiven auf einen Sachverhalt als Bedeutung einer bestimmten Bezeichnung (für diesen Sachverhalt) zu fixieren (s. 3.2). Andererseits wurde gezeigt, dass die genaue Bedeutung ebendieser Bezeichnung –wie die Bedeutung eines jeden Wortes– arbiträr und konventionell ist und somit von ihrer konkreten Verwendung im Diskurs bestimmt wird (s. o.; vgl. 4). Die Kookkurrenzanalyse liefert nun einen induktiven Zugang zu Regel- und Musterhaftigkeiten in ebendieser Verwendung und somit zur konventionalisierten Bedeutung der jeweiligen Bezeichnung. Somit wird über die Untersuchung verschiedener Bezeich-

⁶⁹ Für die übrigen Korpora, die mit CQPweb analysiert wurden, wurde diese Suchanfrage in die CQP-Language übertragen. Im Einzelfall wurden hier weitere ambige Wortformen ausgeschlossen, sofern diese durch Spezifika der jeweiligen Textsorte weitere falsche Positive ergeben hätten. Die genauen Suchanfragen finden sich im Anhang in Endnote viii.

nungskonkurrenzen die paradigmatische und über die Untersuchung ihrer jeweiligen Wortumgebung die syntagmatische Dimension des Valeurs ausgeleuchtet. Die Frage, welche Bezeichnungen für den Konzeptbereich ›gleichgeschlechtliche Ehe‹ entsprechend untersucht werden sollten, lässt sich mit Hilfe der sorgfältig erstellten Suchanfragen für die Korpuserstellung beantworten (s. Suchanfrage 1 in 6.2). Da Kookkurrenzanalysen erst ab einer gewissen Treffermenge aussagekräftig sind, wurden selten verwendete Ausdrücke nicht systematisch auf ihre Kookkurrenzen untersucht. Ein Blick auf die absoluten Trefferzahlen im DeReKo – dem größten und vielfältigsten der verfügbaren Korpora – hilft hier bei der Prioritätensetzung (s. Tabelle 12).

Tabelle 12: Häufigkeit der verschiedenen Bezeichnungen im Archiv W des Deutschen Referenzkorpus (DeReKo) (Stand März 2023).

Nr.	Bezeichnung(+ Suchanfrage in Cosmas II) ⁷⁰	Häufigkeit im DeReKo
1	<i>Homo-Ehe</i> Homo + Ehe +	17.448
2	<i>eingetragene (Lebens-)Partnerschaft</i> eingetragene* / + 1w *Partnerschaft*	10.033
3	<i>Ehe für alle</i> (Ehe ODER Ehen) / + 1w für / + 1w (alle ODER Alle)	7.935
4	<i>Gleichgeschlechtliche Ehe</i> gleichgeschlechtlich* / + 1w (Ehe ODER Ehen)	4.925
5	<i>Schwulen-Ehe</i> Schwul*Ehe +	1.528
6	<i>Öffnung der Ehe</i> Öffnung / + 1w der / + 1w Ehe	1.478
7	<i>Homosexuellen-Ehe</i> Homosexuell*Ehe +	1.243
8	<i>Ehe für/zwischen homosexuelle(n) ...</i> (Ehe ODER Ehen) / + 1w (für ODER zwischen) / + 1w Homosex*	852

⁷⁰ Bei allen Suchanfragen wurde die Suchmodalität „Diakritische Zeichen beachten“ ausgewählt, die Suchmodalitäten zu „Groß- / Kleinschreibung beachten“ jedoch nur bei der Suchanfrage zu *Ehe für alle* (vgl. Suchanfrage 1).

Tabelle 12 (fortgesetzt)

Nr.	Bezeichnung(+ Suchanfrage in Cosmas II)	Häufigkeit im DeReKo
9	<i>Ehe für/zwischen gleichgeschlechtliche(n) ...</i> (Ehe ODER Ehen) / + 1w (für ODER zwischen) / + 1w gleichgeschlechtlich*	791
10	<i>homosexuelle Ehe</i> Homosex* / + 1w (Ehe ODER Ehen)	511
11	<i>eingetragene Lebensgemeinschaft</i> eingetragene* / + 1w Lebensgemeinschaft*	334
12	<i>Lesben-Ehe</i> Lesb*Ehe +	85
13	<i>anerkannte (Lebens-)Partnerschaft</i> anerkannte* / + 1w Lebenspartnerschaft*	62
14	<i>anerkannte Lebensgemeinschaft</i> anerkannte* / + 1w Lebensgemeinschaft*	22

Um die Analyse auf die relevantesten und ggf. auch am stärksten ausgehandelten Bezeichnungen zu konzentrieren, wurde für die systematische Kookkurrenzanalyse eine Untergrenze von 1000 Vorkommen im DeReKo gewählt. Konkret wurden also in jedem themenspezifischen Korpus Kookkurrenzprofile für folgende Ausdrücke erstellt: *Homo-ehe/Homo-Ehe*, *eingetragene Lebenspartnerschaft*, *Ehe für alle*, *gleichgeschlechtliche Ehe*, *Schwulen-Ehe*, *Öffnung der Ehe*, *Homosexuellen-Ehe*. Zuletzt wurden auch die Kookkurrenzen des Ausdrucks *Ehe* analysiert, da speziell dessen Bedeutungsfixierungsversuche untersucht werden sollen und da er schließlich als Überbegriff ebenfalls ein mögliches Referenzmittel für das Konzept ›Ehe (auch) zwischen gleichgeschlechtlichen Paaren‹ darstellt (s. 9). Hierfür wurde *Ehe*, wenn möglich, nicht nur in den themenspezifischen Korpora, sondern auch im Gesamtkorpus untersucht.⁷¹

Da die Bedeutungsfixierungsversuche rund um diese Bezeichnungsalternativen von besonderem Interesse für die Forschungsfragen sind, wurden deren Kookkurrenzen aufwändiger errechnet als für die bisherigen Suchwörter. Einerseits wurde durch eine Suche im Ko-Text von 10 Token vor und nach dem Ausdruck die

⁷¹ Da die Korpora *G-Twitter* und *R-Entscheidungen* keine Untermengen aus größeren Korpora darstellen sondern eigens für die Analyse erstellt wurden, konnte dieser Analyseschritt nur im *DeReKo*, den Bundestagsplenarprotokollen sowie im Gesamtkorpus der Zeitschrift *konkret* unternommen werden.

umfassende syntaktische und vor allem thematische Einbettung des jeweiligen Ausdrucks untersucht. Separat wurde für unmittelbare sprachliche Anschlüsse nach Kookkurrenzpartnern direkt vor und nach dem Ausdruck gesucht, die somit Bigramme ergaben. Für die Untersuchung direkter sprachlicher Zuschreibungspraktiken zu einem bestimmten Ausdruck (bzw. dessen Signifikat) wurde nach speziellen syntaktischen Kookkurrenzen (vgl. Bubenhofer 2015: 494) in Form von Attributen der Nominalphrase und von Prädikativa auf lexikalischer sowie auf Satzebene gesucht. Die Suchanfragen hierfür gestalteten sich z. B. für das Suchwort *Ehe* in *Cosmas II* wie in Tabelle 13 zu sehen.⁷²

Tabelle 13: Suchanfragen für syntaktische Kookkurrenzen.

syntaktische Kookkurrenz	Suchanfrage	Kookkurrenzfenster
Attribut mit unbestimmtem Artikel	(eine ODER einer) /w + 2:2,s0 Ehe	1 Wort rechts vom ersten Treffer
Attribut mit bestimmtem Artikel	(die ODER der) /w + 2:2,s0 Ehe	1 Wort rechts vom ersten Treffer
Prädikativ	(Ehe ODER Ehen) / + w1 (ist ODER sind)	2 Wörter rechts vom ersten Treffer
Prädikativ	(Ehe ODER Ehen) / + w1 (ist ODER sind)	20 Wörter rechts vom ersten Treffer

Aus diesem Verfahren ergeben sich für jeden Ausdruck mehrere Kookkurrenzprofile, die in weiteren Schritten thematisch geclustert werden können (vgl. Amri-Henkel 2021: 113 f.; Kania 2020: 142) zu semantischen Feldern (s. 7.4), die quantifizierbare Hinweise (s. 7.5) darauf geben sollen, in welcher Richtung die jeweilige Bezeichnung perspektiviert wird. Da Kookkurrenzen zwar quantitative Hinweise darauf geben, mit welchen Ausdrücken die untersuchte Bezeichnung in Verbindung gebracht wird, jedoch noch nicht aufzeigen, welche Qualität diese Verbindung hat (vgl. Amri-Henkel 2021: 114) – man denke z. B. nur an Möglichkeiten der Negation –, ist eine kontextsensitive Deutung einzelner Belege (vgl. Felder 2018b: 31) für die Kategorisierung unerlässlich (s. 7.3).

⁷² Für CQPweb wurden entsprechende Suchanfragen formuliert. Z. B. „MU(meet [word=„eine|einer“][word=„Ehe“] 2 2)“, „MU(meet [word=„die|der“][word=„Ehe“] 2 2)“ für Attribute und „[word=„Ehe|Ehen“][word=„ist|sind“]“ für Prädikativa. Bei letzteren konnten höchstens 10 Treffer rechts vom Ergebnis als Kookkurrenzfenster angegeben werden.

7.3 Hermeneutische Analyse

An dieser Stelle soll noch einmal in aller Kürze die zentrale Bedeutung hermeneutischer Analysen für das vorgestellte Untersuchungsprogramm betont werden. Bevor aus den induktiv erschlossenen quantitativen Hinweisen der Analyseschritte in 7.1. und 7.2. erkenntnistiftende Kategorien abgeleitet werden können, muss die konkrete Verwendung der jeweiligen Wörter und Syntagmen überprüft werden. Schließlich bedeutet eine gebrauchsbasierte, pragmatische Analyse nicht nur, dass 1. authentische, natürliche Sprachdaten als Untersuchungsgegenstand dienen und dass 2. Ko-Texte quantitativ-statistisch ausgewertet werden, sondern auch, dass 3. Sprache möglichst in der Form untersucht wird, in der sie auch rezipiert wird – unter Berücksichtigung aller verstehensrelevanten kontextuellen Faktoren (vgl. Spitzmüller & Warnke 2011: 57 ff., 151; Felder & Gardt 2015: 16).

Das bedeutet, dass eine rein quantitative, statistische Auswertung des Ko-Textes nicht genügt, um die Rolle eines Ausdrucks in einem bestimmten Diskurs angemessen zu analysieren, weil in die Bedeutungskonstitution auch „Relationen oberhalb der Wort- und Satzebene, häufig sogar oberhalb der Ebene einzelner konkreter Texte“ einfließen (Glasze, Husseini & Mose 2009: 293). Da von korpuslinguistisch ermittelten seriellen Phänomenen nicht auf individuelle soziopragmatische Größen geschlossen werden kann, müssen individuelle Verwendungen eines Ausdrucks oder Syntagmas qualitativ und kontextsensitiv analysiert werden (vgl. Müller 2012: 55). Dies kann z. B. in der Form von Konkordanzanalysen geschehen, bei denen viele Belege einer bestimmten Konstruktion auf einen Blick im jeweiligen Ko-Text ausgewertet werden können (vgl. Felder 2015: 104; Amri-Henkel 2021: 114). Um das richtige Verständnis der jeweiligen Gebrauchsweise zu sichern, kann der untersuchte Ko-Text bei Bedarf erweitert werden und im Zweifelsfall müssen ganze Texte gründlich gelesen werden (vgl. Kania 2020: 150). Zum verstehensrelevanten Wissen können bei der Deutung eines Belegs auch außertextuelle Kontextfaktoren wie Informationen über Akteur, Zeitraum, Kommunikationssituation, intertextuelle Bezüge etc. zählen, die über die Metadaten des jeweiligen Belegs eingesehen werden können. Wie bereits erwähnt, können für häufige Einzelexeme die typischsten und häufigsten Verwendungsweisen über die Kookkurrenzanalyse vorsortiert werden (s. 7.2.1), sodass sich die hermeneutische Analyse auf die so konventionalisierte Gebrauchsbedeutung konzentrieren kann.

Erkenntnistiftend und effizient ist also vor allem eine komplementäre Kombination von quantitativen und qualitativen, von corpus-driven bzw. korpusbasierten und corpus-based bzw. korpusgestützten Verfahren (vgl. Spitzmüller & Warnke 2011: 39; Felder 2012a: 125), wie auch Kania (2020: 141) dies speziell für den Diskurs um die gleichgeschlechtliche Ehe tut. So können qualitative Analysen dazu dienen, Erklärungen für die Häufigkeit eines bestimmten Sprachmusters zu liefern (vgl. Bubenhofer

2015: 498 f.). Bubenhofer (2013) spricht in diesem Zusammenhang auch von einer qualitativ informierten qualitativen Diskursanalyse. Dabei können auch die singulären Interpretationen einzelner sprachlicher Ereignisse bereits eine umfassende Aussagekraft für den Diskurs haben: „it may be that we can come up with findings of some considerable generality by looking at very singular, particular things. By asking what it takes for those things to have come off“ (Sacks 1992: 298). Diese Perspektive steht letztlich in der Tradition von Foucault, für den die diskursiven Regeln und Grenzen nicht etwa nur bestimmten, was gesagt werden darf, sondern auch, was in einem bestimmten Kontext mit der Sprache überhaupt ausgedrückt werden kann (Foucault 1973: 68). So kann für die vorliegende Analyse schon ein einzelner Beleg von Interesse sein, wenn mit ihm gezeigt werden kann, dass und wie es prinzipiell möglich ist, mit (oder trotz) einem bestimmten umstrittenen Ausdruck eine bestimmte sprachliche Konzeptualisierung vorzunehmen. Umgekehrt können die Ergebnisse quantitativ informierter qualitativer Analysen (s. 7.4) aber auch wieder zu sprachoberflächlichen Indizien verdichtet werden, sodass sich die so aufgedeckten Form-Funktions-Korrelationen (vgl. Felder, Müller & Vogel 2012: 17) wiederum quantifizieren lassen (s. 7.5). Die kontextsensitive qualitative Analyse von Einzelbelegen erfüllt also erstens den Zweck, die induktiv erschlossenen quantitativen Auffälligkeiten auf ihre Zuverlässigkeit als Indizien für bestimmte diskursive Phänomene zu überprüfen; zweitens liefert eine hermeneutische Analyse einzelner Belege auch weitere kontextuelle Erkenntnisse, die sich in emergenten Kategorien verdichten lassen.

7.4 Kategorienbildung

Mit den bisherigen Schritten wurden die häufigsten Ausdrücke im Diskurs aufgespürt (s. 7.1), deren musterhafte syntagmatische Kombinatorik (s. 7.2) sowie individuelle verstehensrelevante Kontextualisierungshinweise (s. 7.3). Bei den einzelnen Schritten wurde bereits angedeutet, welche Art Erkenntnisse über den Diskurs sich aus ihnen ableiten lassen könnten. Im Folgenden soll es nun darum gehen, diese induktiv erarbeiteten Hinweise heuristisch zu generalisierbaren Kategorien zu verdichten, um diese wiederum quantifizieren und so für verschiedene Zeiträume, Akteure und Diskursdomänen vergleichen zu können (s. 7.5). Abbildung 9 stellt eine Spezifikation der einzelnen Schritte in Abbildung 8 dar und zeigt, welche Art Kategorien aus welchen methodischen Verfahren gewonnen werden können.

Im Folgenden sollen die verschiedenen Typen von Kategorien grob vorgestellt werden, sodass der Weg von den induktiv erschlossenen Werten und hermeneutisch untersuchten Einzelbelegen hin zu den emergenten Kategorien prin-

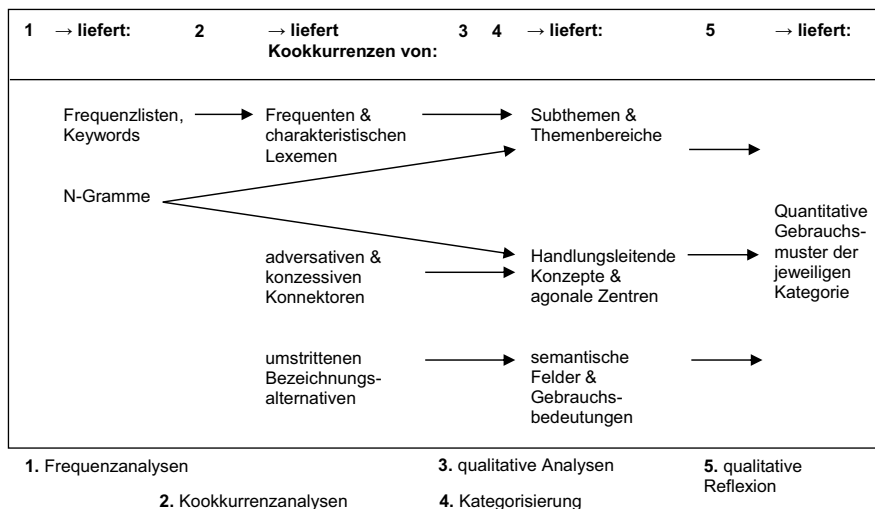


Abbildung 9: Einzelne Schritte und Zwischenergebnisse der quantitativ-qualitativen Diskursanalyse.

zipiell nachvollziehbar wird. Detaillierter hergeleitet und vorgestellt werden die einzelnen Kategorien im Analyseteil (Teil IV).

Subthemen und Themenbereiche: Werden frequente und Diskurs-charakteristische Lexeme (Keywords) sowie N-Gramme wie oben beschrieben analysiert, lassen sie sich bestimmten Subthemen im Diskurs zuordnen. Dadurch kann ihr Auftreten später als ein quantifizierbares, möglichst zuverlässiges Indiz für die sprachliche Aushandlung ebendieses Subthemas darstellen (s. 7.5). Die Kategorisierung der Subthemen orientiert sich hierbei an der pragma-semiotischen Textarbeit nach Felder (2012a; 2018b). In diesem Verfahren sind die errechneten Keywords „als Schibboleth (Erkennungszeichen und Verdichtungsmarker) zu sehen und führen gegebenenfalls zu Subthemen“ (Felder 2012a: 135). Die so hergeleiteten Subthemen können nach Felder (ebd.) mit entsprechender Fachliteratur abgeglichen werden. Für den vorliegenden Untersuchungsgegenstand bieten einschlägige Forschungen zum Diskurs um die gleichgeschlechtliche Ehe in den U. S. A. sowie in Deutschland einen hilfreichen Bezugsrahmen. So arbeiten Crehan & Rickenbacker (2006) die folgenden fünf Hauptthemen für den US-amerikanischen Diskurs heraus: 1. Religion/Moral, 2. Fortpflanzung/Kindeswohl, 3. Familienwerte, 4. (Grund-)Rechte und 5. juristischer Aktivismus. Kania (2020: 140) stellt fest, dass diese fünf Hauptthemen sich in unterschiedlichem Maße in allen bisher untersuchten kulturellen und linguistischen Zusammenhängen finden. Ihre Übertragung dieser fünf Hauptthemen auf den deutschen Diskurs um die gleichgeschlechtliche Ehe ergibt erstens, dass die

Themenbereiche ›Familienwerte‹ und ›Fortpflanzung/Kindeswohl‹ hier untrennbar verknüpft sind, und zweitens, dass ›Religion/Moral‹ sowie ›juristischer Aktivismus‹ im deutschen Diskurs „not salient at all“ sind (Kania 2020: 150). Auch für die vorliegende Analyse kann diese thematische Einteilung als flexibles Gerüst zur Kategorisierung der Subthemen dienen, um Vergleichbarkeit mit vorigen Analysen zum deutschen sowie anderen Diskursen herzustellen. Da diese Hauptthemen bei Crehan und Rickenbacker sowie bei Kania sehr allgemein gefasst und schlagwortartig formuliert sind, dienen sie hier jedoch vor allem als zusammenfassende thematische Kategorien, die fortan *Themenbereiche* genannt werden sollen. Die induktiv erarbeiteten Subthemen im Sinne Felders lassen sich spezifischer formulieren⁷³ und somit den übergeordneten Themenbereichen zuordnen. Letztere müssen hierfür gegebenenfalls ausdifferenziert, ergänzt oder gar verworfen werden (s. 8).

Handlungsleitende Konzepte und agonale Zentren: Die qualitative Analyse von N-Grammen sowie von Kookkurrenzen rund um adversative und konzessive Konnektoren ermöglicht die Identifikation zentraler Streitthemen des Diskurses, die Felder (2012a; 2015; 2018b) agonale Zentren nennt. Ein agonales Zentrum besteht bei Felder demnach aus zwei konfligierenden handlungsleitenden Konzepten (Felder 2012a: 130), die jeweils dominant versprachlichte Konzeptualisierungen eines umstrittenen Sachverhaltes darstellen (vgl. Felder 1995: 3; 2006: 18). Im theoretischen Teil (s. 3.2, 3.3) wurde bereits die besondere Rolle von Agonalität für politische Momente im Diskurs sowie für lexematische Entscheidungen von Diskursakteuren herausgearbeitet.

Speziell für Kulturkämpfe um die gleichgeschlechtliche Ehe ermöglicht die Analyse von handlungsleitenden Konzepten und agonalen Zentren, über die heuristische Größe des Subthemas hinaus die Positionen herauszuarbeiten, die verschiedene Diskursakteure zum jeweiligen Subthema einnehmen. So bleiben auch Crehan & Rickenbacker (2006) bei ihrer thematischen Einteilung des Diskurses (s. o.) nicht stehen, sondern suchen nach den genauen thematischen Verschiebungen im Wandel der gesellschaftlichen Einstellung zur gleichgeschlechtlichen Ehe (vgl. Kania 2020: 139). In Zimmermanns (2010) Analyse von Kulturkämpfen um die gleichgeschlechtliche Ehe spielen ferner konkurrierende kulturelle Selbstverständnisse als jeweils „handlungsleitendes Orientierungssystem“ (Zimmermann 2010: 130) eine Rolle. Ganz zentral geht es bei Kulturkämpfen also, wie auch in 5.1 gezeigt wurde, um konfligierende Deutungsansprüche desselben (bzw. als identisch verstandenen) kulturellen Erbes – in diesem Fall ›Ehe‹. Agonale Zentren bieten kon-

⁷³ So lesen sich die Subthemen nach Felder z. B. für den Diskurs um den Mauerfall so: „Motive für den Fall der Mauer und dessen (seine?) prognostizierte und tatsächliche Folgen“ (Felder 2012a: 136).

densierende Interpretationsgrößen dieser zentralen Streitfragen, denn sie reflektieren einen „sich in Sprachspielen manifestierenden Wettkampf um strittige Akzeptanz von Ereignisdeutungen, Handlungsoptionen, Geltungsansprüchen, Orientierungswissen und Werten in Gesellschaften“ (Felder 2012a: 136). Ebenso lassen sich mit ihnen die zu untersuchenden Transformationsprozesse kulturellen Erbes als kollektive Verschiebungen, als wechselnde Dominanz verschiedener handlungsleitender Konzepte diachron nachzeichnen: „Agonale Zentren haben also nichts Statisches, sondern sind im Gegenteil stets dem diskursiven Aushandlungsprozess ausgesetzt“ (Felder 2012a: 136). Da agonale Zentren aber auch thematisch orientiert sind, lassen sie sich ebenfalls den oben erwähnten Themenbereichen und z. T. auch einzelnen Subthemen des Diskurses zuordnen.

Zuletzt erlaubt die Erarbeitung metasprachlicher agonaler Zentren die gewinnbringende Verknüpfung des theoretischen Erkenntnisinteresses (an der diskursiven Rolle umstrittener Einzellexeme) mit der erkenntnistiftenden Analyse des Diskurses um die gleichgeschlechtliche Ehe. Denn für letzteren stellen die metasprachliche Reflexion sowie der Wandel der Bezeichnungsalternativen eine entscheidende Größe dar, wie sich in der empirischen Analyse bestätigen wird. So legt z. B. auch Vigo (2015) in seiner Arbeit zum italienischen sowie britischen Diskurs um die gleichgeschlechtliche Ehe nahe, dass ein Wandel in der lexikalischen Bezeichnung der gleichgeschlechtlichen Ehe auch einen Wandel in der gesellschaftlichen Einstellung gegenüber derselben aufdecken. Die sprachliche Aushandlung solcher lexikalischen Entscheidungen soll in der vorliegenden Arbeit in Form von metasprachlichen agonalen Zentren gefasst werden. Diese stellen also konfligierende Geltungsansprüche in Bezug auf die Bedeutung oder Angemessenheit einzelner Ausdrücke dar – sowie deontische Handlungsanleitungen zu Verwendung respektive Vermeidung bestimmter Ausdrücke.

Semantische Felder und Gebrauchsbedeutungen: Umstrittene Bezeichnungsalternativen im Sinne Vigos (2015) können auch durch eine semantische Kategorisierung ihrer Wortumgebung genauer analysiert werden. Diese soll aufdecken, in welcher Bedeutung eine Bezeichnung (in welchem Zeitraum vom welchem Akteur etc.) verwendet wird, indem sie syntagmatische – und durch den Vergleich zwischen den Bezeichnungsalternativen auch paradigmatische – Beziehungen aufdeckt und somit ihrem *Valeur* erarbeitet (vgl. Felder 2009a: 20; 31 ff.). Somit soll die Analyse der bloßen Häufigkeit einer bestimmten Bezeichnung im Sinne Vigos (2015) um qualitative Analysen musterhafter Verwendungen ergänzt werden. Hierzu werden die in Schritt 7.2.3 erarbeiteten Kookkurrenzen der häufigsten Bezeichnungsalternativen (*Homoehe/Homo-Ehe*, *eingetragene Lebenspartnerschaft*, *Ehe für alle*, *gleichgeschlechtliche Ehe*, *Schwulen-Ehe*, *Öffnung der Ehe*, *Homosexuellen-Ehe* und *Ehe*) zunächst in ihrem konkreten Gebrauch untersucht (s. 7.3). Anschließend wurden

diejenigen Kookkurrenzpartner, die möglichst zuverlässig eine bestimmte Bedeutungskonstitution indizierten, semantischen Feldern zugeordnet (vgl. „key semantic field[s]“ in Analysen zur gleichgeschlechtlichen Ehe bei Kania 2020: 142; Bachmann 2011). Diese semantischen Felder sollen aufdecken, unter welchen Bedingungen eine bestimmte Bezeichnung mit bestimmten anderen Bedeutungsaspekten – merkmalssemantisch gesprochen: mit bestimmten *Semen* – in Verbindung gebracht wird (zu Bedeutungsfeldern und syntagmatischen Bedeutungsbeziehungen vgl. auch schon Porzig 1950: 71). Auch dieses Verfahren steht in der Tradition des Übergangs vom Strukturalismus zum Poststrukturalismus (s. 2.1.2), insofern es „signifikante Komplexe“ aufdeckt, wie sie Barthes (zitiert nach Schiwy 1984: 83) schon für Foucaults Methodik beschrieben hat: „[D]as sind, wenn man will, Klassen von Signifikaten, von riesigen ›Semantemen‹, deren Signifikanten selbst veränderlich sind“ (ebd.).

In der vorliegenden Analyse orientieren sich die semantischen Felder, nach denen die Kookkurrenzen kategorisiert wurden, erstens an den Erkenntnisinteressen des methodischen Deutungsrahmens ›Kämpfe ums kulturelle Erbe‹ (s. 5.1., vgl. Harnisch, Felder & Leyboldt 2020: 1 f.) und zweitens am theoretischen Erkenntnisinteresse an der metasprachlichen Reflexion einzelner Ausdrücke. Konkret teilen sich die semantischen Felder auf oberster Ebene also in die Kategorien *politische*, *zeitliche* sowie *affektive Relationalität* und *metasprachliche Reflexion*. Diese erste Kategorisierung wurde induktiv zu einer detaillierteren Taxonomie semantischer Felder, vor allem aber zu Kategorien verschiedener Gebrauchsformen der strittigen Begriffe ausgearbeitet (s. 10.2). Parallelen zu den oben beschriebenen Themenbereichen und Subthemen können und müssen hier nicht vermieden werden. Unterschiedlich bleibt hier die analytische Perspektive: Die Themenbereiche und Subthemen fokussieren thematische Schwerpunkte des Diskurses auf Textebene (unabhängig von der konkreten Bezeichnung für die gleichgeschlechtliche Ehe); die hier beschriebenen semantischen Felder fokussieren die Bedeutungsbeziehungen, die bestimmte kontingente Bezeichnungen für die gleichgeschlechtliche Ehe auf Satzebene eingehen.

Bei der Kookkurrenzanalyse nach semantischen Feldern besteht nicht nur Interesse an bestimmten thematischen Verknüpfungen, sondern auch an den syntaktischen Beziehungen, die je eine bestimmte Konzeptualisierung (i. e. Bedeutungsfixierung) gleichsam „hinter“ der jeweiligen Bezeichnung offenlegen. Z. B. fixiert ein Syntagma wie *eine gescheiterte Ehe* das Konzept einer bestimmten, individuellen ehelichen Partnerschaft, während ein Syntagma wie *die bürgerliche Ehe* das Konzept der Institution der Zivilehe im Allgemeinen versprachlicht. Besonders unter diesem syntaktischen Gesichtspunkt lassen sich aus den quantitativ zusammengefassten semantischen Feldern auch qualitativ einzelne unterschiedliche Gebrauchsbedeutungen herausarbeiten. *Gebrauchsbedeutung* bezeichnet hier die genaue Bedeutung, die ein Ausdruck in einer konkreten Verwendung über

verschiedene Kontextfaktoren erhält. Sind die Gebrauchskontexte eines Ausdrucks ausreichend einheitlich, so kann eine konkrete Gebrauchsbedeutung konventionalisieren und (für die Sprechergruppe, die diese Konvention teilt) in die kontext-abstrakte Kernbedeutung des Wortes Einzug nehmen. Insgesamt soll durch dieses Verfahren für jede sinn- und sachverwandte Bezeichnung ein Vergleich angestellt werden zwischen den ihr metasprachlich zugeschriebenen Bedeutungspotenzialen einerseits und ihren sich syntagmatisch entfalteten Gebrauchsbedeutungen andererseits.

7.5 Deduktive Analyse von Form-Funktions-Korrelationen

In Weiterführung von 7.4 wurden die dort corpus-driven (induktiv) ermittelten Indizien für Themenbereiche und Subthemen, agonale Zentren und handlungsleitende Konzepte sowie für semantische Felder und Gebrauchsbedeutungen der Bezeichnungsalternativen gesammelt. Hierdurch ergeben sich jeweils relativ stabile Form-Funktions-Korrelationen (vgl. Felder 2015: 88; Felder & Gardt 2015: 15; Felder, Müller & Vogel 2012: 4 f.) bzw. Form-Bedeutungs-Paare im Sinne der Konstruktionsgrammatik (vgl. 2.2.2; Ziem und Lasch 2013: 1). Diese möglichst zuverlässigen sprachoberflächlichen Indizien können nun wiederum quantifiziert und corpus-based (deduktiv) in den verschiedenen Korpora untersucht werden (vgl. Abbildung 8). Durch die höhere quantitative Leistungsfähigkeit dieses verdichtenden Verfahrens können auch thematische und pragmatische Interessen kondensierter an der sprachlichen Oberfläche untersucht werden, was zu einem gewissen Grad auch einen quantitativen Vergleich dieser abstrakteren Forschungsinteressen zwischen verschiedenen Akteuren, Zeiträumen und Textsorten bzw. Diskursdomänen (s. 6: Twitter, Zeitungen, Plenarprotokolle, Gerichtsentscheidungen) erlauben soll.

Insofern ergänzt das vorgestellte Verfahren die pragma-semiotische Textarbeit um einen weiteren quantifizierenden Schritt, der die induktiv erarbeiteten Kategorien quantitativ rückbindet und in Relation setzt. Während die ersten corpus-driven Schritte der Diskursanalyse also der Frage nachgegangen sind „Wie kann man von sprachstrukturellen Befunden auf sprachpragmatische schließen [...]?“ (Felder, Müller & Vogel 2012: 17), behandelt dieser letzte Schritt die umgekehrte Frage: „Welche typisierten Sprachgebrauchs- und Handlungskonstellationen weisen bestimmte Spezifika im Sprachmaterial auf [...]?“ (ebd.). Das vorgestellte Verfahren versucht also insgesamt, die korpuspragmatische Frage zu beantworten: „Wie lassen sich Klassen funktionsäquivalenter Ausdrücke bilden und quantitativ plausibilisieren?“ (Felder, Müller & Vogel 2012: 17). Stellte eine solche Triangulation von quantitativen und qualitativen Verfahren nach Felder, Müller & Vogel (2012: 19) noch „weitgehend ein Desiderat“ dar, gibt es inzwischen zahlreiche methodische Bestrebungen, auch

abstraktere diskursive Phänomene über sprachoberflächliche Indizien quantifizierbar zu machen (vgl. z. B. die Ansätze der Tagung „Diskursgrammatik“ zusammengefasst von Bloching, Kiemes & Kückelhaus 2022). Ein Beispiel hierfür stellt das *DiskursBarometer* von Vogel et al. (2020) dar; es ergänzt qualitative Auswertungen um „quantifizierende „Diskursmetriken“, die Veränderungen typisierbarer sprachlicher Diskursstrukturen über die Zeit hinweg sichtbar machen“ (Vogel et al. 2020: 212). In eine ähnliche Richtung weisen auch Ansätze, die versuchen, „einen Zusammenhang zwischen dem eruierten Strukturmuster und dem angenommenen Sprechakttyp plausibel zu machen“ (Felder, Müller & Vogel 2012: 7) und so „Aussagen über die quantitative Typizität von Illokutionen“ (ebd.: 7) zu treffen. Ein frühes Beispiel hierfür liefern Biber, Connor & Upton (2007) mit ihrer *move-analysis*. In diesem top-down, corpus-based Verfahren (Biber, Connor & Upton 2007: 13, 23) verbinden sie die rein qualitative Analyse einer Stichprobe sowie die manuelle Kodierung von „move types“ (ebd.: 36) mit einer anschließenden Analyse der linguistischen Charakteristika der verschiedenen *moves*. Auf dieser Grundlage können prototypische sprachliche Muster zur Realisation der jew. *move types* erarbeitet und schließlich quantitativ untersucht werden (Biber, Connor & Upton 2007: 37). Im Gegensatz zur *move-analysis* geht dem hier vorgestellten Analyseschritt ein bottom-up, corpus-driven Erschließen der relevantesten Ausdrucksformen voraus, aus denen sich die Kategorien erst induktiv erschließen. Im Anschluss verfolgt Schritt 5 jedoch ebenfalls das Ziel, die sprachlichen Realisationsmöglichkeiten einer bestimmten semantischen oder pragmatischen Kategorie zu sammeln und schließlich zu quantifizieren. Dabei kann die Frequenz eines solchen Realisationsmusters einerseits die diskursive Relevanz des dahinterstehenden semantischen oder pragmatischen Phänomens anzeigen. Andererseits lässt sich durch die Quantifizierung auf der sprachlichen Oberfläche auch die Varianz zwischen Akteuren/Sprachgemeinschaften, Zeiträumen und diskursiv-thematischen Zusammenhängen untersuchen (vgl. Ziem 2018: 117).

Konkret wurden hierzu die Keywords, n-Gramme und Kookkurrenzpartner aus den Schritten 1–2 gesammelt und je so spezifisch wie möglich bestimmten Themenbereich, Subthemen, agonalen Zentren, handlungsleitenden Konzepte sowie semantischen Feldern zugeordnet (s. 7.4). Hieraus ergaben sich induktiv für jedes solche zu untersuchende Phänomen Listen ausdrucksseitiger Indizien (auf Einwort- sowie Mehrwort-Ebene). Diese konnten wiederum mit sinn- und sachverwandten Wörtern (vgl. Dornseiff 2020) ergänzt werden, sodass auch Ausdrücke genutzt werden konnten, die für sich genommen nicht signifikant häufig auftauchen, zusammengekommen jedoch ein bestimmtes Phänomen frequent ausdrücken (vgl. Kania 2020: 142; Felder 2012a: 135). Durch diese Ergänzung kann ausdrucksseitige Varianz nicht über inhaltlich saliente Phänomene hinwegtäuschen. Insgesamt gelingt es so, „rein deduktive Wortlisten-basierte Verfahren einschließlich der damit verbundenen

methodischen Probleme zu überwinden und mit induktiven Ansätzen zu kombinieren“ (Vogel et al. 2020: 212).

Die so erarbeiteten Sammlungen ausdrucksseitiger Indizien wurden schließlich zu umfassenden Suchanfragen formuliert, mit denen – je nach Fragestellung – auf zwei verschiedene Weisen corpus-based in den Korpora gesucht wurde. Die allgemein diskursiven Phänomene der Themenbereiche, Subthemen, agonalen Zentren und handlungsleitenden Konzepte wurden auf Textebene in den themenspezifischen Korpora gesucht (s. 8). Die Ergebnisse lassen sich jeweils nach verschiedenen Metadaten sortieren, sodass daraus hervorgeht, ob z. B. ein Subthema zu einer Zeit besonders oft angesprochen wurde, oder ob z. B. eine Partei ein bestimmtes agonales Zentrum besonders häufig behandelt. Die Suchanfragen zu semantischen Feldern und Gebrauchsbedeutungen wurden hingegen im unmittelbaren Kontext der zu untersuchenden umstrittenen Bezeichnungen angewandt (s. 10.2.3). Hier geht es schließlich um die Kookkurrenz eines semantischen Feldes mit einer bestimmten Bezeichnung für den konzeptuellen Komplex ›gleichgeschlechtliche Ehe‹. Auch hier können die Ergebnisse nach Metadaten sortiert und somit die Bedeutungsfixierungen in verschiedenen Diskursdomänen und durch verschiedene Akteure aufgedeckt werden. Vor allem kann durch eine diachrone Kookkurrenzanalyse (also durch einen diachronen Vergleich der Kookkurrenz einer bestimmten Bezeichnung mit einem bestimmten semantischen Feld) „die Dynamik von Bedeutung analysiert werden“ (Glasze 2008: 203; vgl. Amri-Henkel 2021: 110). Zuletzt können auch spezifischere Gebrauchsbedeutungen der jeweiligen Bezeichnung (s. 7.4) durch komplexere syntagmatische Muster (etwa (un)bestimmte Artikel, Adjektivattribute, Valenzstrukturen etc.) identifiziert und so entsprechend quantitativ ge- und untersucht werden.

Damit ist die Methodik der quantitativ-qualitativen Diskursanalyse vorgestellt: Ein Kreislauf von einem quantitativen, induktiven Zugang zum Korpus (7.1) über signifikante Hinweise auf musterhaften Wortverwendungen (7.2), qualitative Analysen solcher prototypischer Sprachgebrauchsmuster (7.3), Verdichtung der Erkenntnisse zu heuristischen Kategorien (7.4) zu wiederum quantitativen Analysen der so gewonnenen Kategorien (7.5). Im nächsten Teil der Arbeit sollen die Ergebnisse dieses Analyseverfahrens für den untersuchten Diskurs um die gleichgeschlechtliche Ehe vorgestellt und diskutiert werden.