

5 Korpus und Methodik

5.1 Korpuszusammenstellung: Textsorten zwischen konzeptioneller Mündlichkeit und Schriftlichkeit

Die empirische Untersuchung orientiert sich am tatsächlichen Sprachgebrauch und ist strikt korpusgestützt. Die Auswertungen und Beispielanalysen basieren auf einer Materialauswahl unterschiedlicher Textsorten, da „nur breitangelegte, an verschiedenen Textsorten orientierte Untersuchungen [...] zeigen [können], welchen Stellenwert die einzelnen Wortbildungstypen tatsächlich haben“ (Elsen & Michel 2007: 9). Für die Erforschung der medialen Varietäten Mündlichkeit und Schriftlichkeit hat sich in der bisherigen Forschung das Nähe-Distanz-Modell von Koch & Oesterreicher (1985) bewährt, weshalb auf dieses zurückgegriffen wird (vgl. Kapitel 3.1). Dass die Unterscheidung in konzeptionelle Mündlichkeit und Schriftlichkeit bzw. in kommunikative Nähe und Distanz wichtig ist, um die Zusammenhänge zwischen Wortbildung und mündlichen sowie schriftlichen Texten aufzudecken, heben Elsen & Michel (2011: 4) deutlich hervor:

Empirisch lässt sich dieser ko- und kontextuelle Einfluss auf Wortbildungseinheiten und -arten durch die Analyse von Textsorten bzw. Gesprächen, die das gesamte Spektrum zwischen konzeptioneller Nähe bzw. Distanz [...] abdecken, bestimmen. Dabei erweist es sich als unerlässlich, repräsentative Korpora zusammenzustellen, die hinsichtlich der geltenden ko- und kontextuellen Faktoren eindeutig bestimmt sind [...].

Eigenschaften und Funktionen von Wortbildungskonstruktionen im Sprachgebrauch lassen sich nur dann angemessen untersuchen, wenn sie in Beziehung mit den jeweiligen Kommunikations- und Produktionsbedingungen geschriebener und gesprochener Sprache gesetzt werden, um sie anhand der je spezifischen kommunikativen Parameter gewichten und zwischen den Polen verorten zu können. Entscheidendes Kriterium bei der Auswahl der Textsorten, die in **Kapitel 7** erforscht werden, ist, dass diese sich von medial schriftlichen/konzeptionell geschriebenen und medial mündlichen/konzeptionell geschriebenen bis hin zu medial schriftlichen/konzeptionell gesprochenen und medial mündlichen/konzeptionell gesprochenen Textsorten erstrecken. Sinnvoll erscheint dabei,

zunächst von solchen Textexemplaren und Textsorten auszugehen, die ideale Ausprägungen von Texteigenschaften aufweisen, d. h. die (nahe) am Distanz- bzw. am Nähe-Pol stehen – wie es in der Forschung meistens praktiziert wird, wenn Analysen zur gesprochenen Sprache bevorzugt an Alltagsgesprächen und Analysen zur geschriebenen Sprache bevorzugt an Pressetexten durchgeführt werden. (Stein 2003: 31–32)

Mit Blick auf das Ziel der Arbeit ist die Auswahl von Textsorten, die an den beiden Polen von Mündlichkeit und Schriftlichkeit zu situieren sind, insbesondere deshalb zweckmäßig, da dort die deutlichsten Ausprägungen und somit auch die deutlichsten Gegensätze zu erwarten sind. Aus diesem Grund dient das **Alltagsgespräch** als Beispiel für eine prototypische medial mündliche und konzeptionell gesprochene Textsorte. Mit einer gewissen Abstufung kann auch die **Mitarbeiterbesprechung** als konzeptionell gesprochen angesehen werden. Der **Zeitungsbereich** stellt demgegenüber eine Textsorte dar, die medial schriftlich und in hohem Maße konzeptionell geschrieben ist. Darüber hinaus werden dem Alltagsgespräch und dem Zeitungsbereich mit **WhatsApp-Freizeitchat** und **Tagesschau-Bericht** jeweils zwei Textsorten als Pendants gegenübergestellt, die trotz ihrer medialen Verschiedenheit aus konzeptioneller Perspektive miteinander vergleichbar sind (Alltagsgespräch vs. WhatsApp-Freizeitchat, Tagesschau-Bericht vs. Zeitungsbereich). Zusätzlich wird in Anlehnung an Koch & Oesterreicher (1985: 18) mit der **Verwaltungsvorschrift** eine Textsorte in den Blick genommen, die im Vergleich mit Zeitungsbereichen noch näher am Distanzpol zu verorten ist. Neben Textsorten, die an den Extrempolen kommunikativer Nähe und Distanz anzusiedeln sind, werden zudem die konzeptionellen Zwischenstufen systematisch miteinbezogen (**Hochschulprüfungsgespräch**, **YouTube-Kommentar**, **Wikipedia-Löschdiskussion**, **Bundestagsrede**). Die zugrunde gelegten Textsorten weisen einerseits eine weite Streuung im Kontinuum zwischen konzeptioneller Mündlichkeit und Schriftlichkeit auf; andererseits ist durch die Wahl funktional ähnlicher Textsorten, die sich jeweils den beiden extremen Polen zuordnen lassen, eine gewisse Vergleichbarkeit gegeben. Es kann somit zwischen folgenden Abstufungen differenziert werden:

- **prototypisch konzeptionell gesprochen:** privates Alltagsgespräch
- **konzeptionell gesprochen:** WhatsApp-Freizeitchat, Mitarbeiterbesprechung
- **im mittleren Bereich, Tendenz eher in Richtung Mündlichkeit:** YouTube-Kommentar, Hochschulprüfungsgespräch, Wikipedia-Löschdiskussion
- **im mittleren Bereich, Tendenz eher in Richtung Schriftlichkeit:** Bundestagsrede
- **konzeptionell geschrieben:** Tagesschau-Bericht, Zeitungsbereich
- **prototypisch konzeptionell geschrieben:** Verwaltungsvorschrift

Die drei im mittleren Bereich angeordneten Textsorten YouTube-Kommentar, Hochschulprüfungsgespräch und Wikipedia-Löschdiskussion sind jeweils unter anderem durch die folgenden Kommunikations- und Produktionsbedingungen sowie Produktionsstrategien und Rezeptionsbedingungen gekennzeichnet (vgl. Stein 2003: 26–28), weshalb sie als eher konzeptionell gesprochen klassifiziert werden:

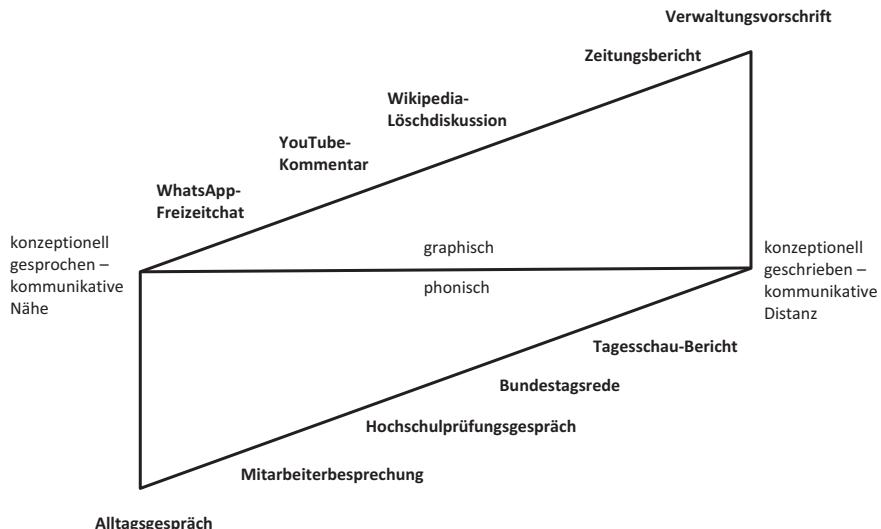
- **YouTube-Kommentar:** zum Teil Dialogizität, Spontaneität, geringe Normverbindlichkeit und größere Normtoleranz, mehr oder weniger freie Themaentwicklung, starke emotionale Beteiligung
- **Hochschulprüfungsgespräch:** Dialogizität, i. d. R. Vertrautheit der Kommunikationspartner(innen), zeitliche und/oder räumliche Kopräsenz von Produzent(in) und Rezipient(in), Kontexteinbindung/Situationsverschränkung, nicht-öffentliche Kommunikation, Einsatz zusätzlicher Kommunikationskanäle wie Prosodie, Mimik, Gestik und Proxemik, geringer Grad der Planung und geringe Planungszeit, Vorläufigkeit, Gleichzeitigkeit von Produktion und Rezeption, einmaliges Hören
- **Wikipedia-Löschdiskussion:** Dialogizität, eher nicht-öffentliche Kommunikation, eher geringerer Grad der Planung und geringe Planungszeit (aufgrund der Tatsache, dass eine Diskussion nur eine bestimmte Zeit lang geöffnet ist)

Die **Mitarbeiterbesprechung** wird gegenüber dem Hochschulprüfungsgespräch näher am Pol der konzeptionellen Mündlichkeit angesiedelt, da die gesprochensprachlichen Merkmale zum einen stärker ausgeprägt sind (so kann z. B. angenommen werden, dass die Mitarbeiter[innen] einer Firma oder eines Theaters vertrauter miteinander sind als Prüfer[in] und Prüfling, zwischen denen zusätzlich ein hierarchisches [Abhängigkeits-]Verhältnis herrscht), und da sich zum anderen einige Charakteristika diametral gegenüberstehen (z. B. [eher] spontan, geringere Normverbindlichkeit/größere Normtoleranz, freiere Themaentwicklung [Mitarbeiterbesprechung] vs. [eher] reflektiert, größere Normverbindlichkeit/geringere Normtoleranz, stärkere Themafixierung [Hochschulprüfungsgespräch]).

Die **Bundestagsrede** wird unter anderem aufgrund folgender Merkmale als eher konzeptionell geschrieben klassifiziert: Monologizität, Reflektiertheit, große Normverbindlichkeit und geringe Normtoleranz, Themafixierung, öffentliche Kommunikation, i. d. R. geringe emotionale Beteiligung, Produktorientierung, hoher Grad der Planung und große Planungszeit, Endgültigkeit, mehrmaliges Lesen möglich (da die Reden protokolliert werden).

Insgesamt ergibt sich die Einordnung der ausgewählten Textsorten im Modell von Koch & Oesterreicher (1985) aus dem Abgleich der Textsorten mit den Parametern für konzeptionelle Mündlichkeit und Schriftlichkeit nach Stein (2003: 26–28), wobei für die Verortung die Parameter Kommunikations-/Produktionsbedingungen, Produktionsstrategien und Rezeptionsbedingungen ausschlaggebend sind (vgl. Übersicht 5-1). Wie in Kapitel 3.1.3 beschrieben, kann es sich hierbei lediglich um eine grobe Zuordnung handeln und es sollte trotz der wohlüberlegten Auswahl der Textsorten stets ein wichtiger Aspekt bedacht werden: Die in der vorliegenden Studie herausgearbeiteten Befunde zur Distribution und Funktion von Wortbildungskonstruktionen sind stets

an die Texte und Textsorten geknüpft, an denen sie gewonnen wurden, und lassen sich nur bedingt verallgemeinern, d. h. auf umfangreichere Felder des konzeptionellen Raumes übertragen, obwohl es plausibel erscheint, vergleichbare Ausprägungen bei verwandten mündlichen und schriftlichen Textsorten anzunehmen. (Stein 2003: 32)



Übersicht 5-1: Verteilung der ausgewählten Textsorten im Nähe-Distanz-Modell.

Für die ausgewählten Textsorten, die Gegenstand von Kapitel 7 sind, werden jeweils so viele Einzeltexte zusammengestellt, dass das Korpus für jede Textsorte circa 10.000 Wörter/Tokens⁸⁶ umfasst (vgl. Übersicht 5-2). Durch diese quantitative Ausgewogenheit zwischen den einzelnen Teilkorpora ist die Vergleichbarkeit der verschiedenen Textsorten im Hinblick auf ihre Wortbildungstypik gewährleistet. Dennoch bestehen quantitative Unterschiede in Bezug auf die Größe der

⁸⁶ Die Auszählung der Texttokens erfolgte mit der automatischen Zählfunktion von Word („Wörter zählen“). Es muss ausdrücklich hervorgehoben werden, dass die Token-Anzahl der einzelnen Textsorten als Annäherung an 10.000 Wörter zu verstehen ist. Denn bei der automatischen Word-Auszählung der Tokens treten diverse Schwierigkeiten auf, insbesondere wenn es sich um Transkriptionen gesprochener Sprache oder internetbasierte Kommunikation wie beispielsweise WhatsApp-Nachrichten und YouTube-Kommentare handelt (z. B. werden in den WhatsApp- und YouTube-Texten zum Teil Wortbildungen auseinandergeschrieben, die aber von Word automatisch als zwei Wörter gezählt werden). Es konnte sich außerdem nicht auf die Token-Angaben verlassen werden, die in der DGD zu den einzelnen Transkriptionen zu finden sind, da die DGD – so ist zu vermuten – auch sprachliche Einheiten und Äußerungen als Wörter zählt, die keine Wörter im engeren Sinne sind. Aus diesem Grund wurde eine manuelle Bereini-

Übersicht 5-2: Textsortenauswahl und -zusammensetzung.

| | Textsorte | Textexemplare | Tokens |
|---|---------------------------|---|---------------|
| Medial schriftliche Textsorten | Verwaltungsvorschrift | Zwölf Verwaltungsvorschriften der Bundesregierung; verschiedene Normgeber (z. B. Auswärtiges Amt, Bundesministerium der Finanzen, Bundesministerium der Justiz, Bundesministerium für Gesundheit) von 2001 bis 2015 | 10.215 |
| | Zeitungsbereicht | 27 Berichte aus den Bereichen Politik, Wirtschaft, Gesellschaft, Kultur, Sport und Panorama vom 27.05.2018 und 28.05.2018 (zehn FAZ, neun SZ, acht Die Welt) | 10.642 |
| | Wikipedia-Löschdiskussion | 34 Diskussionsverläufe vom 24.01.2015, 27.09.2015 und 28.09.2015 | 10.100 |
| | YouTube-Kommentar | Acht Kommentarverläufe zu Videos verschiedener Kanäle (z. B. LeFloid, ApeCrimeTV, Made My Day, TopZehn); Eingrenzung auf die Top-Kommentare und deren Antworten, die auf der ersten Seite zu sehen sind (Stand 06.07.2016) | 10.466 |
| | WhatsApp-Freizeitchat | Sechs private Chatverläufe (in Ausschnitten) von Personen unterschiedlichen Alters (21, 22, 23, 25, 41, 43 Jahre), beider Geschlechter und verschiedener Wohnorte (Dresden, Altdorf, Essen, Mülheim) vom Juni/Juli 2013 und Juni/Juli/August/September/Oktober/November/Dezember 2014 | 11.096 |

Übersicht 5-2 (fortgesetzt)

| | Textsorte | Textexemplare | Tokens |
|------------------------------------|---------------------------|---|----------------|
| Medial mündliche Textsorten | Tagesschau-Bericht | 53 Berichte aus acht Tageschau-Sendungen vom 01.09.2017 bis 08.09.2017 | 10.049 |
| | Bundestagsrede | 13 Reden zu unterschiedlichen Themen (z. B. Integrationsgesetz, Milchkrise, Essensversorgung in Kitas und Schulen) vom 03.06.2016 und 04.06.2016 (vier CDU/CSU, drei SPD, drei Bündnis 90/Die Grünen, drei Die Linke) | 10.392 |
| | Hochschulprüfungsgespräch | Drei Prüfungsgespräche (Literaturwissenschaft, Literaturdidaktik, Sprachwissenschaft, Sprachdidaktik); drei und sechs Gesprächsteilnehmer(innen) vom 08.11.2010 und 13.05.2011 | 12.795 |
| | Mitarbeiterbesprechung | Drei komplette Besprechungen und zwei Besprechungen in Auszügen aus verschiedenen Berufssektoren (Soziale Einrichtung, Theater, Wirtschaftsunternehmen); zwei, fünf und sieben Gesprächsteilnehmer(innen) vom 26.02.2009, 20.03.2013, 25.03.2014 und 10.07.2016 | 10.334 |
| | Alltagsgespräch | Drei komplette Gespräche und ein Gespräch in Auszügen verschiedener Alltagssituationen (Paargespräch, WG-Gespräch, Gespräch unter Studenten, Tischgespräch); zwei und vier Gesprächsteilnehmer(innen) vom 26.06.2009, 23.09.2009, 20.11.2009 und 30.07.2014 | 10.931 |
| | Insgesamt | | 107.020 |

einzelnen Textsortensammlungen, weshalb in der Analyse stets relative Zahlen- und Prozentwerte angegeben werden.

Die empirische Praxis zeigt, dass das Kernkorpus mit 107.020 Tokens eine ausreichende Größe besitzt, um die Distribution von Wortbildungsphänomenen in verschiedenen Textsorten sowie in den medialen Varietäten Mündlichkeit und Schriftlichkeit erforschen zu können (vgl. Kapitel 7). Die Korpusgröße mag zunächst im Vergleich mit anderen datengeleiteten Wortbildungsstudien gering erscheinen; der grundlegende und entscheidende Unterschied zu anderen Arbeiten liegt aber darin, dass in der vorliegenden Untersuchung alle nominalen, adjektivischen und verbalen Wortbildungen in den Texten identifiziert und annotiert werden (vgl. Kapitel 5.2). In vielen Arbeiten richtet sich der Blick demgegenüber auf eine spezifische Wortbildungsart oder ein bestimmtes Wortbildungsmodell. Solche Arbeiten benötigen logischerweise ein (viel) größeres Textkorpus, um statistisch valide Aussagen über das fokussierte Einzelphänomen treffen zu können. So erforscht Scherer (2005: 42) beispielsweise die nominale *-er*-Derivation im Deutschen auf der Grundlage von 1.031.000 Tokens (Mainzer Zeitungskorpus). Hartmann (2016: 131) legt seiner Untersuchung von Nominalisierungen auf *-ung* und von Infinitivnominalisierungen zwei Korpora (Mainzer Frühneuhochdeutschkorporus und GerManC-Korpus) mit einem Gesamtumfang von circa 1.026.000 Tokens zugrunde. Die Analyse von Kempf (2016: 107) zu adjektivischen Suffixderivaten auf *-lich*, *-ig* und *-isch* fußt auf dem Bonner Frühneuhochdeutschkorporus und dem GerManC-Korpus mit einer Gesamttokenzahl von 890.292. In Studien, die sich dediziert mit den Beziehungen zwischen Wortbildung und Text beschäftigen, finden sich häufig gar keine Angaben zum Wortumfang des Textkorpus. Zum Beispiel führt Siebold (2000: 255–257) in seiner Untersuchung von Wortneubildungen in der Science-Fiction-Literatur nur die Seitenzahl der verwendeten Romane an. Peschel (2002: 86) nennt in ihrer Studie zu okkasionellen Suffixierungen auf *-ig* und Präfigierungen auf *ver-* sowie Substantiv-Substantiv- und Substantiv-Adjektiv-Komposita lediglich die analysierten Zeitungen (z. B. Süddeutsche Zeitung, Die

gung der Transkriptionen durchgeführt und die Wortanzahl der gesprochensprachlichen Texte eigenständig ermittelt. Dabei wurden beispielsweise gefüllte Pausen wie *äh*, *ähm*, *öh* oder *öhm* und Hörersignale wie *hm* oder *hm_hm* ausgeschlossen sowie Zusammenziehungen wie *gibt_s* (*gibt es*), *ich_n* (*ich einen*) oder *zu_ner* (*zu einer*) als zwei Wörter gezählt. Zudem wurden Wortabbrüche nicht berücksichtigt (*so frau weber sie sind ges gesund*) und transkribierte Kurzwörter wie *ef de pe* als ein Wort gezählt. Nicht zuletzt ist zu bedenken, dass Partikelverben zwar als ein Wort bzw. eine Wortbildungskonstruktion angesehen werden, diese in konkreten Texten aber häufig in Form von zwei getrennten Textwörtern realisiert sind und somit von Word als zwei einzelne Wörter gezählt werden (z. B. *Die Redaktion habe im Internet Hasskommentare erhalten, teilte das Magazin am Montag auf seiner Website mit.*).

Zeit) und die literarischen Quellen (z. B. Dürrenmatt „Die Panne“, Jelinek „Die Klavierspielerin“), ohne die Gesamtwortanzahl transparent zu machen.

Bei der Materialzusammenstellung wird einerseits auf bestehende Korpora zurückgegriffen, andererseits werden eigene Textsortensammlungen angefertigt:

- Die **Alltagsgespräche**, **Mitarbeiterbesprechungen** und **Hochschulprüfungs-gespräche** stammen aus der Datenbank für Gesprochenes Deutsch (DGD) des Leibniz-Instituts für Deutsche Sprache (IDS),⁸⁷ innerhalb derer Audio- und Videoaufnahmen sowie Transkriptionen verschiedener Gesprächssorten vorhanden sind.
- Mit dem Projekt „What's Up Deutschland“⁸⁸ befindet sich zurzeit zudem eine Sammlung von **WhatsApp-Kommunikationsverläufen** im Aufbau. In Absprache mit dem Projektleiter Prof. Dr. Beat Siebenhaar (Universität Leipzig) wurden mir die vorhandenen Daten freundlicherweise bereits vor der Veröffentlichung zur Verfügung gestellt.
- Eine Sammlung an **Wikipedia-Löschdiskussionen** erhielt ich dankenswerterweise von meiner Kollegin Dr. Monika Hanauska (Karlsruher Institut für Technologie) (vgl. Hanauska 2018a, 2018b).
- Die Zusammenstellung der **Bundestagsreden** basiert auf den Bundestagsprotokollen, die über die Homepage des Deutschen Bundestags frei zugänglich sind.⁸⁹
- Die **Verwaltungsvorschriften** sind der Datenbank „Verwaltungsvorschriften im Internet“⁹⁰ der Bundesregierung Deutschlands entnommen. Auf der Homepage sind aktuelle Verwaltungsvorschriften der obersten Bundesbehörde aufrufbar.
- Der Norddeutsche Rundfunk hat mir die Untertitel von **Tagesschau-Berichten** bereitgestellt. Diese mussten jedoch nochmals sorgfältig mit dem Originalton abgeglichen und überarbeitet werden, da die Untertitel in der Regel von der Audiospur im Hinblick auf ihre exakte lexikalische und morphosyntaktische Realisierung abweichen.⁹¹
- Die Teilkorpora zu **YouTube-Kommentaren** sowie **Zeitungsbereichten** sind manuell zusammengestellt worden. Dabei wurde auf eine ausgewogene und

⁸⁷ dgd.ids-mannheim.de/dgd/pragdb.dgd_extern.welcome (23.01.2020).

⁸⁸ tu-dresden.de/tu-dresden/newsportal/news/whatsup (23.01.2020).

⁸⁹ www.bundestag.de/protokolle (23.01.2020). Die stenographischen Berichte wurden exemplarisch mit dem Originalton abgeglichen. Dabei wurden mit Blick auf die Wortbildung keine gravierenden Unterschiede zwischen den Reden und den stenographischen Berichten festgestellt.

⁹⁰ www.verwaltungsvorschriften-im-internet.de/ (23.01.2020).

⁹¹ Ausgespart werden in der Untertitelung unter anderem häufig Attribute (z. B. Adjektivattribute wie *das [kommunistische] Regime*; Genitivattribute wie *Entschärfung [der Bombe]*; Appositionen wie *angeordnet von [Staatschef] Kim Jong Un*), fakultative Satzglieder (z. B. *bekamen die Nachricht [per Sondersendung] im Staatsfernsehen; zeichneten aber massive Erschütterung [nahe*

repräsentative Auswahl geachtet (YouTube-Kommentare zu Videos verschiedener bekannter Kanäle wie LeFloid oder ApeCrimeTV; online abrufbare und aktuelle Berichte der auflagenstärksten überregionalen Zeitungen wie FAZ, SZ und Die Welt).

Mit WhatsApp-Freizeitchat, YouTube-Kommentar und Wikipedia-Löschdiskussion werden gezielt (mehr oder weniger) konzeptionell gesprochene Textsorten analysiert, die der internetbasierten Kommunikation entstammen. Dies trägt zum einen der rasanten Entwicklung der Kommunikation in digitalen Medien Rechnung. Zum anderen hat die Fokussierung auf diese Textsorten auch forschungspraktische Gründe. Denn die Sprache des Web 2.0 bietet sich „in besonderem Maße für die empirische Forschung von Wortbildungsphänomenen der gesprochenen Sprache, welche sich üblicherweise mit enormem Erhebungsaufwand konfrontiert sieht, an“ (Elsen & Michel 2010: 34).

Das Beispielmaterial zur Veranschaulichung der Funktionen von Wortbildungen in Einzeltexten (vgl. **Kapitel 6.2**) geht über die in Kapitel 7 behandelten Textsorten hinaus. Es werden zusätzliche sowohl geschriebenssprachliche als auch gesprochensprachliche Texte exemplarisch herangezogen, die unterschiedlichen Kommunikationsbereichen zuzuordnen sind und die verschiedene Funktionalstile umfassen (z. B. Zeitungskommentar, Sportlerinterview, Werbeanzeige, Battle-Rap, Beitrag aus der Satire-Sendung Extra3). Der exemplarischen Untersuchung von Textkomplexen/Puzzletexten liegen Transkriptionen von Tagesschau-Beiträgen zugrunde, die sich aus verschiedenen Textsorten zusammensetzen (Anmoderation, Bericht, Interview, Kommentar, Abschlussworte) (vgl. **Kapitel 6.3**).

des nordkoreanischen Testgeländes] auf) oder auch Erst- und Zweitglieder in Komposita (z. B. [Unruhe-]Provinz; [Nuklear-]Test; [Bundes-]Kanzlerin; [Weltkriegs-]Bombe; Premier[-Minister]; Themen[-Komplexe]; Evakuierung[s-Aktion]). Zudem kommt es nicht selten zur Substitution einzelner Wörter bzw. Wortgruppen (z. B. Isolierung Nordkoreas statt Isolierung des Nachbarlandes; Kinder illegal Eingereister statt Söhne und Töchter illegal Eingereister; die Frage, die alle interessiert statt die Frage, die ganz Amerika bewegt), zur Komprimierung eines Syntagmas in Form einer Wortbildungskonstruktion (z. B. Kohleexporte statt der Export von Kohle; Wahlkampfhöhepunkt statt Höhepunkt des Bundestagswahlkampfes; Wasser-Beförderung statt Beförderung auf dem Wasser), zur Verwendung von Kurzwörtern statt der entsprechenden Langformen (z. B. EU statt Europäische Union; Formel-1-WM statt Formel-1-Weltmeisterschaft) sowie zur Änderung des Tempus (z. B. Tausende Einwohner der Hauptstadt verließen in Panik ihre Häuser statt Tausende Einwohner der Hauptstadt hatten in Panik ihre Häuser verlassen) oder zur Umgestaltung der Satzstruktur (z. B. Es gibt keine unabhängige Bestätigung statt Eine unabhängige Bestätigung gibt es nicht; Die USA wollen Nordkorea von seinen Handelspartnern abschneiden statt Die USA kündigten Maßnahmen an, um Nordkorea von seinen bisherigen Handelspartnern abzuschneiden).

Das Kapitel zu Wortbildungen in interaktionaler Kommunikation (vgl. **Kapitel 8**) stützt sich primär auf die Analyse der in Übersicht 5-2 gelisteten Alltagsgespräche, Mitarbeiterbesprechungen, Hochschulprüfungsgespräche und WhatsApp-Freizeitgespräche. Dabei wurde die Anzahl an konkreten Einzelgesprächen/Textexemplaren um insgesamt circa 10.000 Tokens vergrößert, um im Rahmen der qualitativen Untersuchung Aussagen auf der Grundlage einer breiteren Datenbasis treffen zu können (vgl. zum analytischen Vorgehen ausführlich Kapitel 8.2.1).

Die Studie zu Passe-partout-Konstruktionen, die in **Kapitel 9** im Zentrum steht, basiert auf einer Zusammenstellung aller konkreten Realisierungen dieser teillexikalisierten Konstruktion, die in der DGD zu finden sind. Zudem werden auch Daten aus dem Deutschen Referenzkorpus (DeReKo)⁹² ausgewertet. Genauere Informationen zur Datengrundlage können den Kapiteln 9.2.3.1 und 9.2.4.1 entnommen werden.

Die in der Arbeit angeführten Transkriptionsauszüge, die der DGD entstammen, sind ursprünglich nach den Konventionen für Minimaltranskriptionen nach GAT-2 (vgl. Selting et al. 2009) gestaltet, in denen die Prosodie nicht verzeichnet wird. An Stellen, an denen prosodische Auffälligkeiten bei der Analyse der hier behandelten Wortbildungskonstruktionen eine Rolle spielen, werden in Anlehnung an das GAT-2-Basistranskript unter anderem Intonationsphrasen, der Fokusakzent und die Tonhöhenbewegungen am Ende von Intonationsphasen eigenhändig ergänzt. Darüber hinaus werden einige Transkriptionsauszüge aus Gründen der besseren Übersichtlichkeit formal überarbeitet (z. B. Turneinteilungen und -nummerierungen, Sprecher[innen]kürzel) sowie teilweise Korrekturen von Transkriptionsstellen vorgenommen, wenn diese im Abgleich mit der Audiospur unsauber oder fehlerhaft erschienen. Die Transkriptionen derjenigen Beispieldokumente, die nicht der DGD entnommen sind (z. B. Battle-Rap-Texte), orientieren sich ebenfalls an den GAT-2-Konventionen.

5.2 Methodisches Vorgehen: korpusbasierte und computergestützte Analyse

Die empirische Studie ist durch eine computergestützte Vorgehensweise gekennzeichnet. Die Korpusanalyse erfolgt mithilfe einer vom Trier Center for Digital Humanities sowie vom Servicezentrum eSciences der Universität Trier entwickelten virtuellen Forschungsumgebung für die Geistes- und Sozialwissenschaften na-

⁹² cosmas2.ids-mannheim.de/cosmas2-web/ (23.01.2020).

mens „Forschungsnetzwerk und Datenbanksystem“ (FuD).⁹³ FuD ermöglicht es, Daten zu sammeln, zu dokumentieren, zu klassifizieren und zu quantifizieren. Die einzelnen Texte und Transkriptionen der zehn ausgewählten Textsorten sind in FuD eingespeist und mit Metadaten (z. B. Dokumenttitel, Veröffentlichungsdatum/Zeitpunkt des Gesprächs, Wortanzahl) versehen. Die Texte werden digital verwaltet und langzeitarchiviert. Das FuD-System stellt durch eine TEI-konforme XML-Exportschnittstelle die Nachnutzbarkeit der Daten sicher.

Um in **Kapitel 7** einen möglichst differenzierten Einblick in die unterschiedliche Verwendung und quantitative Verteilung von Wortbildungen in Bezug auf die zugrundeliegende Textsorte zu erhalten, werden die einzelnen Wortbildungen einer detaillierten Analyse unterzogen, die verschiedene Beschreibungsebenen berücksichtigt (vgl. Kapitel 2.3.5). Als Hintergrundfolie lassen sich die Modellierungsschritte heranziehen, die Fleischer & Barz (2012: 73) „[f]ür die Zuordnung einer Wortbildung zu einem Modell“ anführen. Die von Fleischer & Barz (2012) angeführten Modellierungsschritte werden zum Teil um weitere Analysekategorien wie die Anzahl der enthaltenen Morpheme/Komponenten ergänzt. Insgesamt ergeben sich somit folgende Parameter, die bei der Beschreibung der Wortbildungen einbezogen werden:

- 1) Grundform (d. h. Tilgung von möglichen Flexionsendungen der Textwörter)
- 2) Wortbildungsart (vgl. Kapitel 2.2.5.2)
- 3) Wortart (Substantiv, Adjektiv, Verb) (vgl. Kapitel 2.2.4)
- 4) Wortbildungsstruktur (binär – nicht-binär) (vgl. Kapitel 2.3.3)
- 5) Anzahl der in der Wortbildung enthaltenen (Grund- und Wortbildungs-)Morpheme bzw. Bestandteile/Komponenten (Komplexität)⁹⁴
- 6) Struktur der unmittelbaren Konstituenten (d. h. Bestimmung der Wortbildungseinheiten) (vgl. Kapitel 2.2.5.1)
- 7) Lexikalisierungsgrad (usuell – okkasionell) (vgl. Kapitel 2.2.9)

⁹³ fud.uni-trier.de/ (27.04.2020).

⁹⁴ Mithilfe dieses Parameters wird die Komplexität der Wortbildungskonstruktionen im Kontinuum von Mündlichkeit und Schriftlichkeit gemessen. Es werden diejenigen Bestandteile/Komponenten gezählt, die an der Oberfläche der Wortbildungskonstruktion erkennbar sind, wobei es sich hierbei nicht immer um klassische Morpheme handeln muss. Als jeweils ein Bestandteil werden folgende Elemente aufgefasst: Grundmorphem (*Haus₁-tür₂, [das] Lauf₁[en], Hass₁-ist₂-krass₃-Lieb₄-e₅-ist₆-krasser₇-Sticker₈*), Wortbildungsaffix (*Schön₁-heit₂, be₁-rat₂[en], plan₁-mä₂sig₃, Ge₁-red₂-e* [das Zirkumfix wird als ein zusammengehöriger Bestandteil angesehen]), Konfix (*Obdukt₁-ion₂, inform₁-ier₂[en], Bio₁-log₂-ie₃*), Kurzwort (*Pkw₁-Maut₂, Schoko₁-Kuchen₂*), Buchstabe (*C₁-Klasse₂, Doppel₁-E₂*), Fremdwort (*An₁-biet₂-er₃-Homepage₄, Bund₁[es]-trainer₂, Bord₁-restaurant₂*) sowie (fremdsprachliche) Produkt-/Firmennamen und Personennamen (*Ray-Ban₁-Sonnen₂-brille₃, Erdogan₁-An₂-häng₃-er₄*).

Wie ersichtlich ist, wird auf eine systematische Annotation der Wortbildungsbedeutung, die ebenfalls Bestandteil des Modells von Fleischer & Barz (2012) ist, verzichtet (vgl. hierzu auch Kapitel 2.3.4). Dies geschieht aus zwei Gründen:

1. Bei der Arbeit mit authentischen Textdaten stößt man schnell an Grenzen, wenn man gewillt ist, für jedes Wortbildungsprodukt eine zugrundeliegende Bedeutung anzugeben. Die Charakterisierung der Wortbildungsbedeutungen stellt im Allgemeinen sowie in Anbetracht der Vielzahl und Heterogenität von Wortbildungen (Komposition, explizite Derivation, Konversion, Kurzwortbildung usw.) im Besonderen schlichtweg eine kaum zu bewältigende Aufgabe dar. Eine angemessene und systematische Modellierung sowie Beschreibung von Wortbildungsbedeutungen bezeichnet Ortner (1997: 40) nicht umsonst als die „Hölle der Semantik“. Erschwerend kommt hinzu, dass bei vielen der identifizierten (okkasionellen) Wortbildungen nicht auf bereits bestehende abstraktere semantische Paraphrasen aus der einschlägigen Wortbildungsliteratur zurückgegriffen werden kann und deshalb individuelle Bedeutungsumschreibungen formuliert werden müssten.
2. Mit Blick auf die Fragestellung der Arbeit ist es nicht notwendig, für alle identifizierten Wortbildungen Bedeutungsangaben zu machen. Da es um die Wortbildungstypik und -spezifik in mündlichen und schriftlichen Textsorten geht, wäre es kaum sinnvoll, die Semantik von solchen Wortbildungen zu beschreiben, deren Vorkommen für eine Textsorte nicht typisch ist.

Deshalb wird folgendermaßen vorgegangen: Die Wortbildungssemantik wird zwar im Rahmen der quantitativen Korpusanalyse nicht gesondert annotiert, in Kapitel 7 werden aber ausgehend von der Datengrundlage bei der Darstellung der textsortentypischen Wortbildung beispielhaft auch semantische Aspekte derjenigen Wortbildungarten berücksichtigt, die sich im Vergleich mit anderen Textsorten sowie mit konkurrierenden Wortbildungarten als besonders frequent und somit als charakteristisch für eine bestimmte Textsorte herausstellen. Dabei wird zudem eine konsequente Verbindung hergestellt zwischen dem gehäuften Gebrauch bestimmter Wortbildungskonstruktionen und den Merkmalen der zugrundeliegenden Textsorte bzw. der (konzeptionellen) Kommunikationssituation. Bei der Bedeutungsbeschreibung der ausgewählten Beispiele wird unter anderem auf die semantischen Darstellungen in Fleischer & Barz (2012: Kapitel 2.1.3 und 2.2.3) zurückgegriffen.⁹⁵

⁹⁵ Besondere Beachtung findet die Wortbildungsbedeutung in Kapitel 9, in dem Wortbildungen mit dem Ansatz der Konstruktionsgrammatik holistisch (d. h. morphologisch, semantisch und pragmatisch) analysiert und modelliert werden.

In enger Abstimmung mit dem Trier Center for Digital Humanities und dem Trierer Servicezentrum eScience wurde für die Annotation der Informationen eine individuelle Erfassungsmaske erstellt, die aus der Forschungspraxis heraus entwickelt und somit exakt an die Bedürfnisse der durchgeführten textorientierten Wortbildungsanalyse angepasst worden ist. Der Entwicklungsprozess und die Funktionen der Erfassungsmaske lassen sich wie folgt beschreiben:

Das Datenmodell für die Analyse der einzelnen Wortbildungen in den Texten wurde gemeinsam mit dem Wissenschaftler entwickelt, der seine methodische Vorgehensweise anhand von Beispielen detailliert erläuterte: Es wurde dann festgelegt, dass bei seiner Vorgehensweise der Ausgangspunkt der Analyse einer Wortbildung die Bestimmung der Wortbildungsart ist. Daher wurde ein Analyseindex mit den Wortbildungsarten angelegt. Das bedeutet in der Praxis, dass er eine Wortbildung im Volltext markiert und ihr die entsprechende Analysekategorie (Wortbildungsart, z. B. Suffixderivat) zuweist. Da alle anderen Analysekategorien (z. B. Wortart, Lexikalisierung, Grundform) abhängig von der jeweiligen Analyse *einer* Wortbildung sind, wurde das Metadatenschema des Analyseobjektes um diese Analysekategorien in Form von weiteren Feldern erweitert. Auf diese Weise können die verschiedenen Ausprägungen von einzelnen Wortbildungsarten (z. B. hinsichtlich ihrer Wortart) strukturiert erfasst und später quantitativ ausgewertet werden. (Lemaire 2018: 242–243; Herv. i. O.)

Mithilfe der entwickelten Maske, die in Übersicht 5-3 zu sehen ist, lassen sich die in den Texten identifizierten Wortbildungsprodukte nach den oben genannten Parametern annotieren:⁹⁶

Übersicht 5-3: Ausschnitt aus FuD am Beispiel *quengelig* in einem Familiengespräch.

96 Die Annotation erfolgte manuell, da automatische Verfahren zur Annotation der fokussierten Analysekategorien/-parameter mit FuD (noch) nicht möglich sind.

Auf der rechten Seite befindet sich der zu untersuchende Text (in diesem Fall ein Familiengespräch). Wortbildungen, die im Text identifiziert werden (hier *quengelig*), lassen sich einer Wortbildungsart zuordnen (linke Seite), woraufhin sich ein Fenster öffnet, in dem weitere Informationen festgehalten werden können (vgl. das Fenster „Analyseobjekt bearbeiten“). Schließlich unterstützt FuD auch die Strukturierung, Sortierung und Auswertung der Datensätze. Zur genauen quantitativen Auswertung werden die in FuD annotierten und systematisierten Datensätze in Excel überführt und mithilfe von Pivot ausgewertet. Auf diese Weise können statistische Auswertungen der annotierten Wortbildungsinformationen vorgenommen werden (z. B. Frequenzanalysen in einer oder mehreren Textsorte[n] sowie Frequenzvergleiche zwischen verschiedenen Textsorten) (vgl. hierzu Kapitel 7.2.1).

Es ist zu betonen, dass in den Kapiteln 6–9 verschiedene methodische Herangehensweisen (Mischung aus qualitativer und quantitativer Korpusanalyse) und unterschiedliche theoretische Ansätze (z. B. Framesemantik, Konstruktionsmorphologie, vgl. Kapitel 4) gewählt werden, um den einzelnen Desideraten sowie den aufgeworfenen Fragestellungen gerecht zu werden (vgl. Kapitel 1.2 und 1.3). Die empirische und theoretische Auseinandersetzung mit dem Untersuchungsbereich hat gezeigt, dass sich die vielfältigen Beziehungen, die zwischen Wortbildungen und schriftlichen wie mündlichen Texten bestehen, nur mithilfe unterschiedlicher Vorgehensweisen adäquat und umfassend beschreiben lassen. So wird beispielsweise die Distribution von Wortbildungsphänomenen zwischen Nähe- und Distanzkommunikation mithilfe computergestützter und statistischer Korpusauswertungen erforscht (vgl. Kapitel 7); die Gebrauchs- und Funktionsweisen von Wortbildungen in der Interaktion werden hingegen mithilfe detaillierter qualitativer Studien von Gesprächsdaten herausgearbeitet (vgl. Kapitel 8). Darüber hinaus wird in den einzelnen Kapiteln die traditionelle Wortbildungsforschung mit den Ansätzen der Kognitiven Semantik (Kapitel 6), der Interaktionalen Linguistik (Kapitel 8) und der gebrauchsisierten Konstruktionsgrammatik (Kapitel 9) zusammengeführt und deren Fruchtbarkeit für die Wortbildungsforschung demonstriert. Die Kapitel 6–9 sind dementsprechend thematisch in sich geschlossen und können auch (weitgehend) unabhängig voneinander rezipiert werden.

