

Universität des Saarlandes
Fakultät P: Sprach-, Literatur- und Kulturwissenschaften
Fachrichtung Germanistik

linguisten mist

**Eine korpuslinguistische Untersuchung und
optimalitätstheoretische Modellierung von
Schreibvarianten in wortbildungsmorphologisch
komplexen Substantiven**

Inauguraldissertation zur Erlangung des Doktorgrades der
Philosophie

Natascha Christine Immesberger aus Ottweiler

Erstgutachter: Prof. Dr. Ingo Reich
Zweitgutachter: Prof. Dr. Augustin Speyer
Drittgutachter: Prof. Dr. Peter Gallmann



Saarbrücken, den 13. Oktober 2020 eingereicht, 27.10.2021 verteidigt unter Dekan
Prof. Dr. Augustin Speyer

Vorwort

Laut einem afrikanischen Sprichwort braucht es ein ganzes Dorf, um ein Kind aufzuziehen. Ebenso verhält es sich mit der Arbeit an einer Dissertation. Mein größter Dank für seine Unterstützung und Förderung in jeder Phase der Promotion gilt meinem Doktorvater und Mentor *Prof. Dr. Ingo Reich*, der immer an mich geglaubt hat und mich nicht nur fachlich, sondern auch persönlich nachhaltig geprägt hat. Bei *Prof. Dr. Augustin Speyer* bedanke ich mich herzlich für seine regelmäßige und gewinnbringende Rückmeldung und das Zweitgutachten. *Prof. Dr. Dr. hc. Peter Gallmann* gilt mein Dank für das Drittgutachten und sein Kompliment zu meiner Zeichensetzung. Der *Hanns-Seidel-Stiftung* danke ich für die finanzielle und ideelle Förderung, die dieses Promotionsvorhaben mit ermöglicht hat.

Ein besonderer Dank gilt darüber hinaus meiner Familie. Meine Eltern *Christine* und *Peter Scherer* haben nicht nur viel Zeit, Liebe, Energie und Geld in meine Erziehung und Ausbildung investiert, sondern mir auch ihre positive Einstellung zu Arbeit und Leistung mitgegeben, und so den Grundstein für meine erfolgreiche Promotion gelegt. Meinem Mann *Matthias Immesberger* danke ich für die Ermutigung, dieses großes Projekt anzugehen, meiner Freundin *Dr. Silke Bayer* für ihr Vorbild und wünsche meiner Schwester *Katharina Scherer* viel Erfolg für ihr eigenes Promotionsvorhaben. Dankbar bin ich auch für die Unterstützung meiner Schwiegereltern *Anneliese* und *Axel Immesberger* bei Kinderbetreuung und Organisation des Alltags. Meinen Kindern *Matilda*, *Jakob* und *Karl* ist es zu verdanken, dass ich nie vergessen habe, dass es auch noch ein Leben außerhalb der Arbeit gibt.

Meine große Linguistenfamilie hat mein Projekt nicht nur inhaltlich bereichert, sondern auch entscheidenden Anteil daran gehabt, dass ich die Arbeit daran genießen konnte. *Anna Mönnich* und *Dr. Nele Hartung* waren die besten Schreibtischnachbarinnen, die ich mir hätte wünschen können. *Robin Lemke* und *Dr. Lisa Schäfer* danke ich insbesondere für ihre Geduld und Unterstützung in allen methodischen Fragen. *Dr. Philipp Rauth* danke ich für Alles abseits der leeren Kekspackungen und allen anderen KollegInnen für ihre Unterstützung, die nette Gesellschaft und die Versorgung mit Nervennahrung.

Die Arbeit wurde mit dem *Dr.-Eduard-Martin-Preis* der Universitätsgesellschaft des Saarlandes für die besten Doktorarbeiten des Jahrgangs 2020/2021 ausgezeichnet.

Inhaltsverzeichnis

Vorwort	3
1. Einleitung	11
2. Zentrale Begriffe	15
2.1. Das Wort	15
2.1.1. Morphologisch komplexe Wörter	20
2.1.2. Eingrenzung des Untersuchungsgegenstandes	22
Beschränkung auf wortbildungsmorphologische Komplexität .	23
Beschränkung auf komplexe Substantive	25
2.2. Wortinterne Schreibvariation	26
2.3. Die Textform	27
3. Theoretische Grundlagen	31
3.1. Graphematik	31
3.1.1. Buchstaben	34
3.1.2. Alphabetische Grapheme	36
Nicht-alphabetische Grapheme	38
3.1.3. Graphematische Silben	45
Allgemeines graphematisches Silbenbaugesetz	45
Die Silbenstruktur der graphematischen Silbe	48
Exkurs: Silbifizierung der graphematischen Silbe	50
3.1.4. Graphematische Wörter	55
Kernbereich oder markiert – eine überarbeitete Klassifikation .	56
Kernbereich und markiert – nach Fuhrhop (2008)	58
3.2. Orthographie	61
3.2.1. Getrennt- und Zusammenschreibung	61

3.2.2.	Schreibung mit Bindestrich	63
3.2.3.	Groß- und Kleinschreibung	66
3.3.	Die wortinternen Varianten	67
3.3.1.	Zusammenschreibung	68
3.3.2.	Bindestrichschreibung	69
3.3.3.	Getrenntschreibung	81
3.3.4.	Schreibungen mit Binnenmajuskel	84
3.3.5.	Schreibungen mit Apostroph	86
4.	Methodik	89
4.1.	Forschungshypothesen	90
4.1.1.	Einfluss der Wortbildungsart und Lexikalisierung	90
4.1.2.	Einfluss der Bestandteile	91
4.1.3.	Einfluss der Textform	91
4.1.4.	Sonstige Faktoren	92
4.2.	Korpora	93
4.2.1.	Werbeanzeigen-Korpus	96
4.2.2.	Mail-Korpus	99
4.2.3.	SMS-Korpus	101
4.2.4.	Zeitungs-Korpus	102
4.2.5.	Fachzeitschriften-Korpus	104
4.2.6.	Kriminalroman-Korpus	106
4.2.7.	Plenarprotokoll-Korpus	108
4.2.8.	Lexikoneinträge-Korpus	110
4.2.9.	Diskussionsbeiträge-Korpus	111
4.3.	Recherche	113
4.3.1.	Virtuelle Korpora	114
4.3.2.	Stichproben 25.000 und kleine Korpora	116
4.3.3.	Stichproben 5.000	118
4.3.4.	Falsch positive Treffer	119
4.4.	Belegsammlung und Annotation	122
4.4.1.	Wortbezogen	122
	Wortbildung	123
	Lexikalisierung	123

Quelle	124
4.4.2. Morphembezogen	124
Graphematische Markiertheit	124
Eigennamen	125
Fremdsprachlichkeit	125
4.4.3. Grenzbezogen	126
Fugenelemente	126
Markierte Buchstabenfolgen	127
Schreibvarianten	127
5. Ergebnisübersicht zu den 5.000er-Stichproben	129
5.1. Schreibvarianten	130
5.2. Wortbildungen	131
5.2.1. Komposita	133
5.2.2. Derivationen	134
5.2.3. Konversionen	134
5.2.4. Sonstiges	135
5.3. Textformen	136
6. Theoretische Modellierung	145
6.1. Optimalitätstheorie und Graphematik	147
6.2. Modellierung für Kompositionsgrenzen in Determinativkomposita . .	152
6.2.1. Die Zusammenschreibung als Normalfall	154
Zusammen- vs. Getrennschreibung	154
Zusammenschreibung vs. Schreibung mit Interpunktionszeichen	157
Großschreibung auf der Wortebene	160
6.2.2. Systematische Variation	182
Markierte graphematische Wörter als Teilausdrücke	183
Lexikalisierung	186
Logistische Regression	189
Eigennamen als Teilausdrücke	195
Fremde Teilausdrücke	197
Fugenelemente	202
6.2.3. Modell für Kompositionsgrenzen in Determinativkomposita .	205

6.2.4.	Übertragbarkeit auf ähnliche Wortbildungen	207
	Phrasenkomposita	207
	Affixoidbildungen	210
	Inversionskomposita	210
	Konfixkomposita	210
	Zusammenfassung	211
6.3.	Modellierung für Derivationsgrenzen in Suffigierungen	212
6.3.1.	Modell für Suffigierungsgrenzen	216
6.3.2.	Übertragbarkeit auf ähnliche Wortbildungen	218
	Präfigierungen	218
6.3.3.	Modell für Derivationsgrenzen	223
6.4.	Zusammenführung der Modelle	224
6.5.	Textformspezifik	226
6.5.1.	Werbeanzeigen	227
6.5.2.	Mail	230
6.5.3.	SMS	231
6.5.4.	Zeitung	234
6.5.5.	Fachzeitschriften	237
6.5.6.	Kriminalromane	239
6.5.7.	Plenarprotokolle	240
6.5.8.	Lexikoneinträge	241
6.5.9.	Diskussionsbeiträge	243
6.5.10.	Textformspezifische Tableaus	245
	Werbeanzeigen	245
	Mail	246
	SMS	246
	Zeitung	247
	Fachzeitschriften	248
	Kriminalromane	249
	Plenarprotokolle	249
	Lexikoneinträge	250
	Diskussionsbeiträge	251
6.5.11.	Zusammenfassung	252

6.6. Anwendbarkeit des Modells auf komplexe Substantive im Allgemeinen	256
7. Abschließende Diskussion	271
7.1. Wortbildungsart und Lexikalisierung	273
7.2. Einfluss der Bestandteile	275
7.2.1. Graphematisch markierte Wörter	276
7.2.2. Eigennamen	276
7.2.3. Fremdsprachlichkeit	277
7.2.4. Fugenelemente	278
7.3. Textform	278
7.4. Wortstruktur	279
7.5. Fazit	280
A. Vollständige Liste der optimalitätstheoretischen Beschränkungen	289

Tabellenverzeichnis

3.1. Merkmale segmentaler Mittel der Schrift nach Bredel (2011:9)	39
3.2. Operationsdomänen von Wortzeichen nach Buchmann (2015: 278) . .	41
4.1. Korpusübersicht	95
4.2. Aufbau Zeitungs-Korpus	103
4.3. Stichproben	115
4.4. Treffermenge virtuelle Korpora	115
4.5. Treffermenge 25.000er: Binnenmajuskel und manuell	117
4.6. Treffermenge 25.000er: Bindestrich und Apostroph	118
4.7. Treffermenge 5.000er	119
5.1. Belege 5.000er-Stichprobe nach Frequenz	129
5.2. Belege 5.000er-Stichprobe nach Wortbildungsgrenzen	130
5.3. Belege 5.000er-Stichprobe nach Schreibvarianten	131
5.4. Belege 5.000er-Stichprobe nach Wortbildungsarten	132
5.5. Belege 5.000er-Stichprobe nach Wortbildungsarten: Komposition . . .	133
5.6. Belege 5.000er-Stichprobe nach Wortbildungsarten: Derivation	135
5.7. Belege 5.000er-Stichprobe nach Wortbildungsarten: Konversion	135
5.8. Belege 5.000er-Stichprobe nach Wortbildungsarten: Sonstiges	136
5.9. Treffer und Belege 5.000er-Stichprobe nach Textformen	138
5.10. Determinativkomposita und Schreibvarianz 5.000er-Stichprobe	140
6.1. Datenausschnitt: Modellierung Kompositionsgrenzen nach Textform .	153
6.2. Datenausschnitt: Regression Kompositionsgrenzen	190
6.3. Datenausschnitt: Verteilung der binären Prädiktoren	192
6.4. Finales Modell der logistischen Regression	193
6.5. Verteilung fremder Teilausdrücke mit anteiliger Zusammenschreibung	199
6.6. Datenausschnitt: Modellierung Derivationsgrenzen	213

6.7. Modellierungsgrundlage: Werbeanzeigen	227
6.8. Modellierungsgrundlage: Mail	230
6.9. Modellierungsgrundlage: SMS	232
6.10. Modellierungsgrundlage: Zeitung	234
6.11. Modellierungsgrundlage: Fachzeitschriften	237
6.12. Modellierungsgrundlage: Kriminalromane	239
6.13. Modellierungsgrundlage: Plenarprotokolle	240
6.14. Modellierungsgrundlage: Lexikoneinträge	242
6.15. Modellierungsgrundlage: Diskussionsbeiträge	243
6.16. Textformspezifisch finales Modell	253
6.17. Textorientierte vs. interaktionsorientierte Schreibhaltung	255
6.18. Noch nicht erfasste Belege nach Wortbildungsarten	258
6.19. Noch nicht erfasste Belege nach Textform	259
6.20. Noch nicht erfasste Belege nach Wortbildungsart und Optimalität . .	260

Abbildungsverzeichnis

3.1. Sprache als modalitätsabhängiges System nach Bredel (2008)	33
3.2. Vertikale Ausdehnung von Minuskeln (Primus 2006: 13)	35
3.3. Die Längenhierarchie (Fuhrhop & Peters 2013: 217)	47
3.4. Einheiten der Graphematik nach Berg, Primus & Wagner (2016: 351) .	49
5.1. Belege der 5.000er-Stichprobe nach Textform	137
5.2. Anteil der Determinativkomposita und Suffigierungen nach Textform	139
5.3. Anteil der Schreibvarianten in Determinativkomposita nach Textform	141
6.1. Zusammenschreibung und Deckungsgrad nach Textformen	254

1. Einleitung

Eine Besonderheit der deutschen Sprache besteht darin, dass es für ihre Sprecher bzw. Schreiber möglich ist, durch eine Verkettung von Wortbildungsprozessen regelrechte Wortungetüme zu bilden. Mehr als 20 Zeichen oder drei bis vier Kompositionsglieder sind zwar nicht prototypisch für Wörter des Deutschen, aber gerade in der substantivischen Wortbildung eine regelmäßige Erscheinung, wie die Beispiele unter (1) verdeutlichen.

- (1) Allgemeinverbindlicherklärung (PBR), Einzelhandelsverkaufspreise (W), Konsumgenossenschaftsbewegung (WDD₁₁), Arbeitnehmerüberlassungsgesetz (PBR), Telekommunikationskonglomerat (DIV)¹

Diese Buchstabenbandwürmer scheinen den Schreibern selbst aber nicht immer ganz geheuer zu sein, zumindest beobachten aufmerksame Linguisten und Sprachkritiker wie beispielsweise Barz (1993), Starke (1993) oder Stein (1999) schon seit den 90er Jahren eine Zergliederung von morphologisch komplexen Wörtern (insbesondere von Komposita in Werbekontexten) durch wortinterne Schreibvariation an morphologischen Grenzen.

Bis heute dokumentiert und/oder analysiert wurden graphische Mittel wie Bindestriche (2a), Getrenntschreibung (2b), Binnenmajuskeln (2c), Apostrophierung (2d), Klammern (2e) oder typographische Mittel wie Majuskelschreibung, Kursivierung, Fettdruck oder Schriftgröße und -farbe (2f).

Regelkonform nach den Vorgaben des Rates für deutsche Rechtschreibung sind nur einige der Bindestrichschreibungen (vgl. 3.2), während sich der Großteil der beobachtbaren Variationen ohne Rücksicht darauf entwickelt zu haben scheint.

¹ Die angeführten Belege entstammen dem Korpus, das eigens für diese Untersuchung zusammengestellt wurde. In Klammer angegeben ist jeweils das Teilkorpus, dem die Belege extrahiert worden sind. Zur Auflösung der Abkürzungen vergleiche 4.2.

- (2) a. Öko-Attac-Bono-Charity-Gutmenschen-Weltrettungs-Celebrity (DIV)
- b. 6 wöchiges (Mail), Bonus DVD (WDD₁₁), Apple Logo (W)
- c. PremiumDepot (W), CousCoussalat (Mail), InterRegio-Zugverbindung (WPD₁₁)
- d. Urgroßmutter's (WPD₁₁), WWWER's (WDD₁₁)
- e. FPSO-(Floating Production, Storage and Offloading)-Schiffe (VDI09)
- f. AKWs (PBR), *Homo-sapiens*-Männchen (SPK), erstklass^{ich} (W)

Die graphischen und typographischen Phänomene im Einzelnen sind in den letzten Jahrzehnten immer wieder Gegenstand der Forschung und inzwischen mehr oder weniger systematisch untersucht. Einen Überblick bietet Kapitel 3.3. Was noch fehlt, ist eine übergreifende linguistische Analyse, die sich nicht auf willkürliche Belegsammlungen zu Einzelphänomenen stützt, sondern systematisch den Sprachgebrauch auch dahingehend untersucht, wie relevant die oben genannten Phänomene im Deutschen überhaupt sind, in welchen Kontexten und unter welchen Bedingungen sie tatsächlich auftreten und ob, bzw. welchen Regeln sie auch unabhängig vom Regelwerk zur deutschen Rechtschreibung folgen. Eine solche Metaanalyse auf der Basis einer systematischen Korpusuntersuchung zu leisten, ist das übergeordnete Ziel dieser Arbeit.

Um das zu erreichen, wird zunächst in den Vorbemerkungen (Kapitel 2) geklärt, was im Sinne dieser Arbeit unter morphologisch komplexen Wörtern, wortinterner Schreibvariation und Textformen (als Begriff für die Kontexte) zu verstehen ist. Damit einher geht eine begründete Einschränkung des Untersuchungsgegenstandes auf morphologisch komplexe Substantive. Darauf folgt eine Einführung in das zugrundeliegende Graphematikverständnis (Kapitel 3.1). Die komplexe Fragestellung bedingt eine umfassende Datenerhebung, die in Kapitel 4 im Anschluss an die Formulierung der Forschungshypothesen dokumentiert ist. Empirische Grundlage ist eine synchrone Auswertung von Korpusdaten zu neun verschiedenen Textformen (Textsorten, Textklassen und Kommunikationsformen, vgl. Kapitel 4.2), aus der eine große Belegsammlung hervorgeht. Die zentralen, quantitativen Ergebnisse auf Basis dieser systematischen Belegsammlung werden in Kapitel 5 dargestellt. Entsprechend der Ausgangsfrage werden dabei insbesondere die wortinternen Schreibvarianten, die Wortbildungsart und die Textform, in der die Belege vorkommen, berücksich-

tigt. Anschließend werden in Kapitel 6 die Belege für wortbildungsmorphologisch komplexe Substantive im Rahmen der Optimalitätstheorie analysiert, bevor eine abschließende Diskussion in Kapitel 7 die Forschungshypothesen wieder aufgreift und die Ergebnisse der Untersuchung zusammenfasst.

2. Zentrale Begriffe

Vor den Ausführungen zum theoretischen Hintergrund, dem Einstieg in die Korpusanalyse und der anschließenden Auswertung der Belege müssen in diesem Kapitel einige grundlegende Fragen beantwortet werden, um den Untersuchungsgegenstand und die zu analysierenden Einflussfaktoren klar einzugrenzen. Diese Fragen lassen sich recht einfach formulieren:

- Was ist ein Wort?
- Was ist ein morphologisch komplexes Wort?
- Was ist wortinterne Schreibvariation?
- Was ist unter Textform zu verstehen?

Die Antworten darauf sind schwieriger zu finden und häufig abhängig von der Perspektive, die der Antwortende einnimmt, bzw. dem Zweck, zu dem sie beantwortet werden. Dementsprechend sollen die folgenden Teilkapitel und die Ausführungen zu den theoretischen Grundlagen in Kapitel 3 auch keine erschöpfenden und allgemein gültigen Antworten auf diese Fragen liefern, sondern lediglich den Rahmen festlegen, in dem sich die Untersuchung bewegt.

2.1 Das Wort

Es hat wohl jeder Sprecher oder Schreiber eine mehr oder weniger intuitive Vorstellung davon, was ein Wort ist. Für die systematische Untersuchung wortinterner Schreibvariation an morphologischen Grenzen braucht man allerdings vielmehr eine handfeste Definition als ein vages Verständnis. Nötig ist eine Abgrenzung mit möglichst validen Unterscheidungskriterien zwischen Wörtern, die Gegenstand der Untersuchung werden, und Nicht-Wörtern, die für die Untersuchung nicht von Belang

sind. Die Wortbestimmung muss sich außerdem auf die in den Korpora auftretenden Wortformen (also das geschriebene Wort) anwenden lassen und sich aufgrund der Ausgangsfragestellung an morphologischen und graphematischen Kriterien orientieren.

Diesen Ansprüchen recht nahe kommt die Arbeit von Wurzel (2000), der sich der Wortbestimmung zunächst über eine Zusammenschau von Wortkriterien aus unterschiedlichen linguistischen Teildisziplinen (Graphematik, Phonologie, Semantik, Syntax) nähert, um schließlich anhand von vier syntaktischen und morphologischen Kriterien eine morphologische Wortdefinition in Abgrenzung zu syntaktischen Konstruktionen zu erarbeiten:

- (A') Kriterium der Nichtunterbrechbarkeit
Echte morphologische Wörter haben die Eigenschaft, daß sie nicht durch lexikalisches Material unterbrechbar sind; typische Phrasen haben diese Eigenschaft nicht.
- (B') Kriterium der einheitlichen Flexion
Echte morphologische Wörter, die flektierbar sind, haben die Eigenschaft, daß sie über eine einheitliche Flexion verfügen; typische Phrasen haben diese Eigenschaft nicht.
- (C) Kriterium des syntaktischen Status
Morphologische Wörter sind grammatische Einheiten, die auch syntaktisch Wortstatus haben.
- (D1) Definition von echten morphologischen Wörtern
Grammatische Einheiten, die die Kriterien (A') und (B') erfüllen, sind echte morphologische Wörter; grammatische Einheiten, die nicht beide Kriterien erfüllen, dagegen nicht.
- (D2) Definition von morphologischen Semiwörtern
Grammatische Einheiten, die das Kriterium (C) und eines der Kriterien (A') und (B') erfüllen, sind morphologische Semiwörter.

(Wurzel 2000: 39-40)

Aus der konsequenten Anwendung dieser Kriterien ergibt sich nicht nur nach (D1) eine Einteilung in „echte“ morphologische Wörter (*Mittwoch, stattfinden*)¹, die kohärent und einheitlich flektiert sind, einerseits, und Phrasen als unterbrechbare und

¹ Zu den echten morphologischen Wörtern nach Wurzel (2000) gehören also auch die zusammengesetzten Formen trennbarer Verben, während deren unterbrochene Formen (*Das Seminar findet*

uneinheitlich flektierte Produkte syntaktischer Prozesse (*großer Kloß, des Mannes*) andererseits, sondern auch zwei Zwischenkategorien. Zum einen ist das nach (D2) die Kategorie der morphologischen Semiwörter wie *Langeweile, Bundesgericht* und (*sie*) *fahren Rad*, die entweder kohärent oder einheitlich flektiert sind, aber auf jeden Fall syntaktischen Wortstatus haben (Wurzel 2000: 40-41); zum anderen die Gruppe der Phrasen mit partiellen Worteigenschaften, die Konstruktionen umfasst, die nur das Kriterium der Nichtunterbrechbarkeit (A') erfüllen wie *grüner Kloß* oder nur einheitlich flektiert (B') sind wie *der Frau*, aber keinen syntaktischen Wortstatus haben (Wurzel 2000: 41). Aufgrund der überwiegend phrasalen Eigenschaften werden letztere ebenso wie prototypische Phrasen nicht Bestandteil der Untersuchung sein.

Von Interesse sind dagegen nicht nur die prototypischen (echten morphologischen) Wörter, sondern auch die Gruppe der Semiwörter. Um diese Kategorien und die Grenze zwischen ihnen zu beschreiben, lohnt sich ein genauerer Blick auf die Kriterien der Nichtunterbrechbarkeit (Kohärenz) und der einheitlichen Flexion. Zumindest bei der Nichtunterbrechbarkeit handelt es sich um eine „graduelle Eigenschaft“ in Abhängigkeit von der Art der Unterbrechung (Wurzel 2002: 208).

- (1) a. Unterbrechung durch grammatische Morpheme:
tiefe – tief-er-e – tief-st-te; (sie) leben – (sie) leb-t-en
- b. Unterbrechung durch Derivationsmorpheme:
versichert – ver-un-sichert, enden – end-ig-en
- c. Unterbrechung von Derivaten durch Koordination und Tilgung:
Christentum – Christen- und Judentum, Urgeschichte – Ur- und Frühgeschichte, überschreiten – über- und unterschreiten
- d. Unterbrechung von Komposita durch Koordination und Tilgung:
Rotwein – Rot- und Weißwein, Himbeeren – Him- und Brombeeren
- e. Unterbrechung durch Stämme:
Holzrahmen – Holz-tür-rahmen, Bundesgericht – Bundes-verfassungs-gericht
(vgl. Wurzel 2002: 207-208)

Als Verstoß gegen das Kriterium der Nichtunterbrechbarkeit wertet Wurzel (vgl.

morgen statt.) zu den morphologischen Semiwörtern gezählt werden. Er wendet seine Kriterien und die damit verbundenen Kategorien also nicht auf abstrakte Lemmata, sondern auf konkrete Wortformen an.

2002: 208) allerdings nur die Unterbrechbarkeit durch Stämme wie in (1e). Dadurch rutscht ein Großteil der deutschen Komposita in die Kategorie der Semiwörter. Komposita wie in (1d), die durch Koordination und Tilgung unterbrochen werden können, zählen folglich dennoch zu den prototypischen Wörtern ebenso wie die Derivate, die durch Koordination und Tilgung (1c) oder weitere Derivationsmorpheme (1b) unterbrochen werden können. Die wohl schwächste Form ist die Unterbrechbarkeit durch grammatische Morpheme wie in (1a), die entsprechend auch nicht zu einer Aberkennung des Status als echtes morphologisches Wort führt.

Beim Kriterium der einheitlichen Flexion ist eine solche graduelle Abstufung nicht zu beobachten. Verlangt wird nicht nur die Flexion nach einheitlichen grammatischen Kategorien, sondern auch, dass die Flexion formal einheitlich auf das gesamte Wort bezogen umgesetzt wird (vgl. Wurzel 1984: 36). Dieses Kriterium bestehen also auch die Komposita, die durch Stämme unterbrochen werden können, problemlos, da sie nur wortfinal einheitlich flektieren (*Gastprofessor – Gastprofessoren – *Gästeprofessoren*). Analytische Verbformen wie *ich werde schlafen* flektieren dagegen nach verschiedenen Kategorien (1. P. Sg. Präs. Ind. vs. Infinitiv) und erfüllen das Kriterium somit eindeutig nicht (Wurzel 2002: 209).² Nach Wurzel (2002: 209) gibt es auch einzelne Fälle, die syntaktischen Wortstatus haben, kohärent sind, aber keine einheitliche Flexion (einheitliche Kategorie, aber separate Marker) aufweisen:

- (2) a. (das) Hohelied – (des) Hohenliedes
 b. (die) Langeweile – (der) Langenweile (vgl. Wurzel 2002: 209)

Die doppelte Flexion der beiden Beispiele scheint sich aber in einem Abbauprozess zu befinden. So gibt der Duden (2013: s.v. Hohelied, Langeweile) für die beiden Beispiele aus (2) eine einheitliche Flexion vor ((*das*) *Hohelied* – (*des*) *Hohelied*[e]s). Aus dem wortinternen Flexionsmarker sei also schon ein Fugenelement ohne grammatischen Gehalt geworden. Im Fall von *Langeweile* ist auch diese mögliche Zwischenstufe schon überwunden ((*die*) *Lang*[e]weile – (*der*) *Lang*[e]weile). Folglich müssen diese Beispiele aktuell sogar als echte morphologische Wörter klassifiziert werden.

Das Kriterium des syntaktischen Wortstatus, nach dem Wörter Konstituenten der

² Da sie außerdem durch Stämme zu unterbrechen sind (*ich werde lange schlafen*) und mehreren Konstituenten auf der X⁰-Ebene entsprechen, erfüllen sie auch die Kriterien der Kohärenz und des syntaktischen Wortstatus nicht, so dass sie als prototypische Phrasen gelten müssen.

X⁰-Ebene entsprechen müssen (vgl. Wurzel 2000: 39), trennt Phrasen mit partiellen Worteigenschaften von Semiwörtern. So können auch einheitlich flektierende, aber nicht kohärente Ausdrücke wie *der Tür* als Phrasen erkannt werden, da der Artikel, auch wenn er allein den Flexionsmarker trägt, und das Nomen jeweils eine eigene Konstituente auf der X⁰-Ebene sind. Ähnlich verhält es sich mit Ausdrücken wie *Schwarzes Brett*, die nicht über eine einheitliche Flexion verfügen, aber kohärent sind (**Schwarzes großes Brett*). Auch sie werden als Phrasen mit partiellen Worteigenschaften klassifiziert, weil es sich auf der X⁰-Ebene um mehrere Konstituenten handelt.

Mit Hilfe von Wurzels (2000) Kriterien der Nichtunterbrechbarkeit, der einheitlichen Flexion und des syntaktischen Status lassen sich also syntaktische Konstruktionen (prototypische Phrasen und Phrasen mit partiellen Worteigenschaften) aus der Analyse ausschließen. Zusätzlich bieten sie darüber hinaus die Möglichkeit, zwischen morphologischen Wörtern und morphologischen Semiwörtern, die auch phrasale Eigenschaften besitzen, zu unterscheiden.³

Wie gewünscht lässt sich Wurzels (2000) Wortbestimmung auf Wortformen anwenden. Die Einteilung nach Nichtunterbrechbarkeit, einheitlicher Flexion und syntaktischem Wortstatus führt sogar dazu, dass, bei „‘unfesten’ verbalen Zusammensetzungen“ mit Substantiven, Adjektiven oder Adverbien, die nicht-unterbrechbaren Formen gemäß (D1) echte morphologische Wörter sind (*stattfinden*, *Schwein haben*), während die unterbrechbaren nach (D2) als morphologische Semiwörter klassifiziert werden (*sie fahren Rad*)⁴, sofern sie nicht tatsächlich unterbrochen sind wie im Fall von *sie fahren oft Rad* (Wurzel 2000: 40). Je nach Wortform werden diese verbalen Zusammensetzungen also verschiedenen Kategorien zugeordnet.

Für einen rein graphematischen Wortbegriff, der Wörter als Einheiten zwischen zwei Leerzeichen oder einem Leerzeichen und einem Satzzeichen, ohne dass sie von einem Spatium unterbrochen werden, definiert (z. B. Fuhrhop & Peters 2013: 251), sind abtrennbare Verbzusätze dagegen ein Problem (vgl. Boettcher 2009: 2-3). Auch Komposita mit internem Spatium wie *Apple Logo* (W) gelten damit graphematisch als zwei Wörter, auch wenn sie nach dem besprochenen, morphosyntaktischen Wort-

³ Daraus ließe sich wiederum die Frage ableiten, ob es Unterschiede hinsichtlich der wortinternen Schreibvariation zwischen echten morphologischen Wörtern und morphologischen Semiwörtern gibt, die hier aber vorerst nicht beantwortet werden kann.

⁴ Wurzel (2000: 40) sieht hier durch die vertauschte Reihenfolge der Konstituenten keine Veränderung des syntaktischen Status.

begriff von Wurzel (2000) in die Kategorie der Semiwörter fallen und damit morphosyntaktisch nur ein Wort sind.

Graphematisch wird das Wort also nicht durch inhärente Eigenschaften, sondern allein durch seine graphische Umgebung bestimmt. Was dazu führt, dass bei strikter Anwendung auch Zahlenreihen (... 82889/987298 ...), Folgen anderer Zeichen (... (...)) oder zufällige Aneinanderreihungen von Graphemen (... *jhfa-euzfa'ig* ...) als graphematische Wörter gelten müssen, so lange die Zeichenketten zwischen zwei Spatien oder einem Spatium und einem Satzzeichen auftreten. Trotzdem ist gerade in der Korpuslinguistik diese graphische Definition grundlegend, weil Spatien bei der maschinellen Verarbeitung von Sprache eine einfache Möglichkeit sind, einen Text in (potentielle) Wörter aufzugliedern (vgl. Perkuhn, Keibel & Kupietz 2012: 26-27). Gerade wenn es darum geht, große Korpora zusammenzustellen und zu annotieren, liegt es nahe, der Einfachheit halber eine gewisse Fehlerquote zu akzeptieren.

Für die Entscheidung, ob ein Beleg als Wort relevant für die Untersuchung ist, wird aber in erster Linie der morphosyntaktische Wortbegriff entscheidend sein. Die Tatsache, dass für die Korpuslinguistik und speziell für die maschinelle Recherche der graphematische Wortbegriff vorrangig ist, darf dennoch nicht außer Acht gelassen werden, wenn man keine Verzerrung der Rechercheergebnisse in Kauf nehmen möchte (ausführlich dazu Kapitel 4).

2.1.1 Morphologisch komplexe Wörter

Morphologische Komplexität beginnt dort, wo sich mindestens zwei einfache morphologische Konstituenten nach einem bestimmten Muster zu einer morphologischen Einheit verbinden. Das heißt, aus dieser Untersuchung werden zunächst Simplicia ausgeschlossen, die per Definition aus einer einzelnen Konstituente bestehen und dementsprechend nicht zerlegt werden können, wie z. B. *Hof*, *Mund*, *Glück*.

Darüber hinaus bleiben auch Kombinationen außen vor, die zwar in zwei Bestandteile segmentiert werden können, von denen aber nur einer den Status einer morphologischen Einheit hat, während der andere funktional so schwach ist, dass er schon fast in die morphologische Einheit integriert ist (vgl. Eisenberg 2013: 209). Als Beispiele für solche Kombinationen führt Eisenberg (2013: 209) das nicht mehr produktive Affix *-t* wie in *Fahrt* an sowie die „Pseudoaffixe“ *-e* und *-en*, in *Treppe* und *Streifen* (aber: *Treppchen*, *Streifchen*). Diese „Pseudoaffixe“ lassen sich aber auch da-

durch erklären, dass man *Treppen* und *Trepp* bzw. *Streifen* und *Streif* als Allomorphe desselben Morphems betrachtet, die prosodisch motiviert sind, da die Alternation zwischen den Allomorphen in diesen Beispielen sowohl für die Simplizia als auch für ihre Ableitungen die im Deutschen favorisierte trochäische Silbenstruktur gewährleistet. Auch Wurzel (2002: 205-206) wertet das *-e* bei deutschen Femina wie *Treppe* als Morphem. Auch wenn es „keine Bedeutung im eigentlichen Sinne“ trage (Wurzel 2002: 205), geht er mit Eisenberg (2013) davon aus, dass es in Verbindung mit einem Basismorphem *feminine*, schwach flektierende Nominalstämme bildet. Allerdings führt auch Wurzel (2002: 205) mit *Tante*, *Katze* und *Rose* nur Beispiele an, bei denen, wie auch im Fall von Eisenbergs *Treppe*, nicht ersichtlich ist, warum nicht auch schon die Basismorpheme selbst (auch in der Variante ohne *-e*) die Eigenschaften nominal, feminin und schwach flektiert tragen sollen. Dementsprechend werden diese Pseudoaffigierungen in der Untersuchung unberücksichtigt bleiben.

Abseits der Simplizia können Kombinationen von wortfähigen Stämmen untereinander (3a), mit Wortbildungs- (3b) und Flexionsaffixen (3c) oder Konfixen (3d) sowie Zusammensetzungen aus mindestens zwei Konfixen (3e) oder einem Konfix und mindestens einem Affix (3f) als morphologisch komplex angesehen werden.

- (3)
- a. Baum-stamm, Eiche-n-holz-tür, Baum-haus-bau-satz
 - b. häus-lich, Häus-lich-keit, un-häus-lich
 - c. haus-en, ge-haus-t, Häus-er
 - d. Haus-techn-ik, schoko-phil, Wasch-o-mat
 - e. Bio-log-ie, biblio-phil, Bio-rhythm-us
 - f. fanat-isch, invest-ieren, un-dynam-isch

Im Gegensatz zu Stämmen sind Affixe für sich alleine nicht wortfähig (**ig*, **ver*), sondern treten immer gebunden auf (*mutig*, *verzehren*). D. h. sie brauchen einen Stamm (*schulisch*) oder ein Konfix (*politisch*), mit dem sie sich zu einem komplexen Wort verbinden können. Eine Wortbildung durch die Kombination von Affixen ist nicht möglich (**Verlichkeit*). Innerhalb der Gruppe der Affixe kann unterschieden werden zwischen Flexions- und Wortbildungsaffixen. Während durch das Anhängen von Wortbildungsaffixen neue Wörter (Lexeme) entstehen wie in den Beispielen unter (3b), werden Flexionsaffixe genutzt, um neue Wortformen zu bilden wie in (3c).

Eine Art Zwischenstatus nehmen die Konfixe ein. Sie sind wie Affixe alleine nicht

wortfähig (**polit*), können aber teils wie Stämme untereinander oder mit einem Stamm zu komplexen Wörtern kombiniert werden (*Bibliothek, Politbüro*), als Derivationsbasis dienen (*fanatisch*), oder beide Aufgaben erfüllen (*Biograph, graphisch*), da sie wie Stämme auch eine lexikalische Bedeutung tragen. Die Beispiele machen deutlich, dass Konfixe vor allem in der Fremdwortbildung eine Rolle spielen. Es gibt aber auch einige indigene Konfixe, die in der sprachlichen Vergangenheit ursprünglich Stämme waren, heute aber ihre Wortfähigkeit verloren haben wie *stief-* oder *-falt* in *Stiefsohn* oder *Einfalt* (Fleischer & Barz 2012: 63-64).

Ob darüber hinaus zwischen freien Grundmorphemen und Affixen eine Mischkategorie (Halbaffixe/Affixoide) anzunehmen ist, wird in der Forschungsliteratur kontrovers diskutiert.⁵ Aufgrund der marginalen Rolle der Kategorie erfolgt hier keine ausführliche Darstellung des Forschungsstandes, sondern lediglich der exemplarische Verweis auf die Veröffentlichungen von Munske (2002), Schmidt (1987), Stein (2008) und Zifonun (2012) ohne Anspruch auf Vollständigkeit.

2.1.2 Eingrenzung des Untersuchungsgegenstandes

Eine weitere Eingrenzung des Untersuchungsgegenstandes war ursprünglich nicht vorgesehen, hat sich aber aufgrund erster Rechercheergebnisse gleich in zweierlei Hinsicht als sinnvoll erwiesen.

Im ersten Schritt wird der Fokus nur auf wortbildungsmorphologisch komplexe Wörter gerichtet, die aus mindestens zwei (oder mehr) Wortbildungsmorphemen bestehen und zusätzlich Flexionsmorpheme enthalten können wie die Beispiele unter (4a). Das führt zu einer Ausklammerung rein flexionsmorphologisch komplexer Wörter, also Wörtern, die zwar insgesamt zwei oder mehr Morpheme enthalten, aber nur einen wortbildungsmorphologisch relevanten Stamm, der durch ein Flexionsmorphem erweitert wird wie in (4b).

- (4) a. Baum-stamm, häus-lich-e, Bio-log-ie, Eiche-n-holz-tür-en, Herr-lich-keit
 b. Häus-er, haus-t-e, ge-haus-t, schön-er, jüng-st-en

Im zweiten Schritt beschränkt sich die Analyse der wortbildungsmorphologisch

⁵ Auch die Konfix-Kategorie wird aufgrund ihrer Inhomogenität gegenwärtig in Frage gestellt Fleischer & Barz (vgl. 2012: 64).

komplexen Wörter zudem nur auf substantivische Wortbildungen.

Die Gründe für diese Eingrenzung des Untersuchungsgegenstandes werden im Folgenden näher erläutert:

Beschränkung auf wortbildungsmorphologische Komplexität

Die Entscheidung, rein flexionsmorphologisch komplexe Wörter nicht näher zu betrachten, beruht auf ersten Rechercheergebnissen, die gezeigt haben, dass für diese Gruppe kaum Schreibvarianten an flexionsmorphologischen Grenzen zu beobachten sind.⁶ Das Prinzip der Zusammenschreibung wirkt an den Grenzen zu Flexionsmorphemen offenbar so stark, dass im gesamten Korpus, das mehr als drei Millionen laufende Wortformen umfasst, nur 20 Belege zu finden sind, die eine davon abweichende Schreibung aufweisen.⁷ In allen 20 Fällen wird dabei die Grenze zwischen der Flexionsendung -s und der Flexionsbasis markiert. Bei zwölf Belegen wird die Flexionsendung durch einen Apostroph abgetrennt wie in *Urgroßmutter's* (WPD11), in den anderen acht Fällen wird die Grenze durch den Wechsel von Majuskel zu Minuskel markiert z. B. bei AKWs (PBR).

Zweimal wird das Genitiv-Singular-s per Apostroph von Substantiven abgetrennt, die personennamenähnliche Eigenschaften aufweisen. In einem Fall von einem Substantiv, das zwar kein Eigenname ist, aber ebenfalls eine Person bezeichnet, und einmal von einer Abkürzung, die aber im Kontext eindeutig als Pseudonym für eine Person zu verstehen ist (vgl.(5a)). In allen anderen Fällen handelt es sich um -s als Pluralflexiv. Acht davon mit apostrophierten -s bei Abkürzungen (5b). Dazu kommen die Pluralform eines Einzelbuchstabens und eine Interferenzerscheinung mit Plural-s in wortspielerischem Kontext (5c).

Bei den acht verbleibenden Belegen handelt es sich durchweg um in Großbuchstaben geschriebene Abkürzungen mit einem nahtlos angefügten aber kleingeschriebenen Pluralflexiv -s (5d). Die Grenzmarkierung findet hier also lediglich durch den Wechsel von Majuskel zu Minuskel statt.

⁶ Nähere Informationen zum Korpus und zur Recherche nach morphologisch komplexen Wörtern finden sich im Kapitel zur Methodik unter 4.

⁷ Selbst wenn man Personennamen, die in dieser Untersuchung eigentlich nicht berücksichtigt werden sollen, mit betrachtet, erhöht sich die Zahl der Belege nur um drei (*Sissy's* (NUZ09), *Beck's*, *Schweighofer's* (WPD11), obwohl in diesem Kontext eine Schreibung mit Apostroph schon in den amtlichen Regelungen erwähnt wird. Vgl. Rat für deutsche Rechtschreibung (2006: 99)).

- (5) a. Urgroßmutter's (WPD₁₁), Hickhack um Wdh's Beiträge (WPD₁₁)
 b. WWWER's, HP's, KZ's (2 mal), MG's, Ref's PDF's (2 mal) (WDD₁₁)
 c. O's, die dann Drawings und Plan's create(n) (WDD₁₁)
 d. LEDs (WPD₁₁, 2 mal), PEGs (W), AKWs (PBR, 3 mal), FPSOs, PCs (VDI₀₉)

Das Phänomen der wortinternen Schreibvariation an morphologischen Grenzen in rein flexionsmorphologisch komplexen Wörtern ist also gleich in mehrfacher Hinsicht beschränkt. Abgesehen davon, dass es relativ zur Gesamtgröße des Untersuchungskorpus weniger als zehn Treffer pro Million Textwörter gibt, bewegen die sich auch noch hinsichtlich ihrer Bestandteile und Umgebung in einem sehr eng gesteckten Rahmen. Wie oben schon festgestellt werden nur Grenzen zum Flexionssuffix-s überhaupt markiert. Nur in zwei Fällen wird es dabei als Genitiv-Singular-Marker verwendet, während es sonst zur Pluralbildung dient. Zusätzlich zur Beschränkung auf ein einzelnes Flexionssuffix in zwei verschiedenen Funktionen gibt es auch einen engen Rahmen für mögliche Flexionsbasen. Den größten Anteil haben Abkürzungen (17 von 20) ergänzt durch nur wenige Spezialfälle (*O's*, *Plan's*, *Großmutter's*).

Schließlich verteilen sich die Belege auch nicht über das gesamte Untersuchungskorpus, sondern finden sich hauptsächlich (14 von 20) in den Lexikoneinträge- und Diskussionsbeiträge-Korpora, die auf einem Online-Lexikon beruhen. Einzelne Belege finden sich auch in den Werbeanzeigen- und Fachzeitschriften-Korpora sowie in den Plenarprotokollen. Im Umkehrschluss heißt das aber auch, dass in allen vier anderen Teilkorpora keine Belege für eine Schreibvariation abseits der Zusammenschreibung in flexionsmorphologisch komplexen Wörtern zu finden sind.

Abschließend kann also festgehalten werden, dass es im untersuchten Korpus kaum Variation in der Schreibung an rein flexionsmorphologischen Morphemgrenzen deutschsprachiger Wörter gibt und sich die wenigen Fälle in einem eng gesteckten Kontext bewegen, was das verwendete Flexionselement, die Flexionsbasis und die Textform, in der sie auftreten, betrifft. Eine ausführliche, korpusbasierte Untersuchung zur Apostrophierung an morphologischen Grenzen („morphographische Apostrophen“) bietet Scherer (2013). Sie stützt ihre Untersuchung auf die Ergebnisse einer systematischen Analyse eines mehr als 375 Millionen Textwörter umfassenden Korpus (Scherer 2010), ergänzt durch unsystematische Online-Belegensammlungen

von Sprachkritikern. Nübling (2014) wirft einen diachronen Blick auf Apostrophierungen und ihre Funktion.

Beschränkung auf wortbildungsmorphologisch komplexe Substantive

Hinsichtlich der zu analysierenden Wortarten wird eine stärkere Einschränkung der Belegmenge vorgenommen. Im Fokus der Untersuchung werden lediglich Substantive stehen.

Eine Variation in der Schreibung an wortinternen, morphologischen Grenzen kommt zwar auch bei anderen Wortarten vor – z. B. trennbare Verben (Partikelverben (6a), Substantiv-Verb-Verbindungen (6b), Adjektiv-Verb-Verbindungen (6c)) Adverbien (6d) und Präpositionen (6e) – aber dort beschränkt sie sich ausschließlich auf den Wechsel zwischen Getrennt- und Zusammenschreibung (bei Beteiligung nominaler Konstituenten kommt es darüber hinaus bei Getrenntschreibung häufig zu einer Alternation von Varianten mit und ohne initialer Majuskel).

- (6) a. abwarten vs. sie wartet ab, umfallen vs. er fällt um
- b. radfahren vs. Rad fahren vs. rad fahren
- c. sie hat das Glas leergetrunken vs. leer getrunken
- d. nach Hause vs. nachhause vs. nach hause
- e. aufgrund vs. auf Grund vs. ?auf grund

Eine größere Bandbreite der Variation zeigt sich abseits der Substantive (7b) nur bei den Adjektiven (7a):

- (7) a. Beck'sche (WPD11), schwarz-rot-quergestreifte (DIV), familienfreundlichem (RHZ09), BMW typische (W)
- b. Bonus DVD (WDD11), PremiumDepot (W), InterRegio-Zugverbindung (WPD11), FPSO-(Floating Production, Storage and Offloading)-Schiffe (VDI09), *Homo-sapiens*-Männchen (SPK)

Deshalb wurden in der Korpusrecherche zunächst adjektivische und substantivische Belege erhoben, mit dem Ergebnis, dass die Substantive in der Belegsammlung deutlich in der Überzahl sind (auf 12.739 substantivische Belege kommen nur 1.491 adjektivische).⁸ Da die erhobenen Adjektive also nur grob 10% der Belegmenge aus-

⁸ Dieses Ergebnis erklärt sich auch teilweise dadurch, dass Substantive nicht nur den Hauptteil des

machen, wird sich die Analyse auf die Substantive konzentrieren, die Adjektive aber auch mit im Blick behalten.

2.2 Wortinterne Schreibvariation

Die zahlreichen Beispiele haben sicherlich schon einen recht guten Eindruck davon vermittelt, was unter *wortinterner Schreibvariation* zu verstehen ist. Dennoch wird der Begriff hier noch kurz erläutert.

Wortintern bedeutet zunächst, dass ausschließlich Phänomene untersucht werden, die innerhalb morphosyntaktischer Wörter beobachtet werden können, wie die Zusammenschreibung, Getrennschreibung oder die Verwendung von wortinternen Majuskeln oder Syngraphemen. Ausgeklammert werden dadurch beispielsweise Varianten innerhalb phrasaler Verbindungen wie engen Appositionen (*Herr Müller, Flughafen Köln-Bonn, VW Golf*) oder die wortinitiale Groß- und Kleinschreibung.

Darüber hinaus ist Schreibvariation hier als graphematisches Phänomen zu verstehen. Statt orthographische (und nicht-orthographische) Varianten einzelner Lexeme einander gegenüberzustellen, möchte diese Untersuchung unabhängig von Einzellexemen anhand der Korpusbelege systematische Prozesse der Variation erfassen. Fokussiert wird dabei eine Abhängigkeit von Wortbildungsart und der Textform, ohne Rücksicht darauf, wie eine Variante im Einzelfall orthographisch zu bewerten ist.⁹ Es geht also nicht darum, die Einhaltung der explizit kodifizierten Norm zu überprüfen, sondern darum offenzulegen, welche Prinzipien, die sich aus dem tatsächlichen Gebrauch ableiten lassen, dieser Schreibvariation zugrunde liegen.¹⁰

Kernwortschatzes ausmachen, sondern auch eine Hauptrolle bei der Wortbildung spielen was Produktivität und Vielfalt angeht (vgl. Fleischer & Barz 2012: 117-118).

⁹ Zum Verständnis von Graphematik und ihrer Rolle innerhalb der Grammatik bzw. Linguistik sowie dem Verhältnis zur Orthographie vgl. Kapitel 3.

¹⁰ Für eine solche Überprüfung sei auf Kopf (2017: 182-184) verwiesen, die für N+N-Komposita aus dem COW-Korpus in Bindestrichschreibweise (n=937) auf eine Normabweichung von 29% kommt.

2.3 Die Textform

Die Bezeichnung *Textform* ist der Versuch, einen neutralen, nicht textlinguistisch vorgeprägten Oberbegriff für die Textsorten, Publikationsformen und Kommunikationsformen als unterschiedlichen Kontexte der Korpusbelege und ihre spezifischen Eigenarten zu wählen.¹¹

Die ursprüngliche Idee war es, die wortinterne Schreibvariation ausschließlich an textsortenspezifischen Korpora zu untersuchen, um Einflussfaktoren auf die wortinterne Schreibung möglichst differenziert ermitteln zu können. Textsorten sind nämlich per definitionem „mehrfach spezifizierte [...] Muster [...] komplexer kommunikativer Interaktionen“, die jeweils typische sprachliche Merkmale aufweisen, ein bestimmtes Thema behandeln und dieselbe kommunikative Funktion erfüllen (Adamzik 2016: 62). Diese Charakterisierung und die Beispiele unter (8) verdeutlichen, wie stark diese Muster letztendlich ausdifferenziert sind.

(8) Gebrauchsanleitung, Märchen, Rechnung, Wettervorhersage

Während es sich bei einem Märchen um einen narrativen Fließtext mit literarischer Funktion handelt, sind Gebrauchsanleitungen konkrete Handlungsanweisungen mit appellativem Charakter, die teilweise im Listenmodus verfasst sind und jeweils ein Gerät, seinen Aufbau und seine Funktionen thematisieren. Bei einer Wettervorhersage steht klar die Informationsfunktion im Vordergrund. Entsprechend dem eng eingegrenzten Wetterthema, ist das Vokabular recht eingeschränkt und die Dichte formelhafter Wendungen wie beispielsweise *heiter bis wolkig* ist hoch. Für Rechnungen, die ebenfalls appellativen Charakter haben, sind gewisse Bestandteile (z.B. Summe, Datum, Steuernummer) sogar zwingend notwendig. Der Textumfang ist dabei eher minimal mit Listencharakter und bedarf anders als beispielsweise das Märchen

¹¹ Die Textklassifikation und die damit verbundenen Bezeichnungen für nach unterschiedlichen Kriterien zusammengefassten Gruppen von Texten sind in der Textlinguistik nach wie vor ein schwieriges Feld, gespickt mit einer Vielzahl nicht immer einheitlich verwendeter Begrifflichkeiten (z. B. Medium, Typ, Muster, Gattung, Kommunikation) (vgl. Habscheid 2011). Zur Klärung dieser Uneinigkeit wird diese Arbeit kaum etwas beitragen können, ist aber im Wissen um diese Problematik bemüht, die verwendeten Begriffe möglichst transparent zu machen und, soweit für die Untersuchung notwendig, trennscharf abzugrenzen.

keiner linearen Rezeption.

Eine textsortenspezifische Untersuchung setzt dann natürlich voraus, dass aus einer größeren Menge solcher Korpora eine sinnvolle Auswahl hinsichtlich der Textsorten getroffen werden kann, um eine möglichst große Bandbreite kommunikativer Interaktion abzudecken.

Solche Spezialkorpora sind aber trotz gegenläufiger Tendenzen leider noch immer nur eingeschränkt verfügbar. Denn es liegt gewissermaßen in der Natur der Sache, dass Korpora eher Publikationsformen entsprechen, da sie vom Rechteinhaber als Publikation in Gänze zur Verfügung gestellt werden. Die Bezeichnung der jeweiligen Publikationsform bezieht sich dabei auf das Trägermedium und lässt nur begrenzt Rückschlüsse auf die darin enthaltenen Textsorte(n) zu. Adamzik (2016) nennt als Beispiele für Publikationsformen:

(9) Buch, Flugblatt, Zeitung, Schild, Film, Plakat, Website

Für die traditionellen und in großen Sammlungen noch immer dominanten Zeitungskorpora bedeutet das, dass das jeweilige Korpus alle Texte der zur Verfügung gestellten Ausgaben (der Publikationsform) umfasst und damit im Fall einer Zeitung ganz verschiedene Textsorten (Bericht, Kommentar, Wettervorhersage, Todesanzeige, Veranstaltungshinweise, Interviews...).¹² Textsortenspezifische Korpora ergeben sich daher nur in einzelnen Fällen von selbst, in denen die Publikationsform entweder nur einen Text umfasst oder einer Sammlung von Texten derselben Textsorte entspricht. Ansonsten müssen sie gezielt zusammengestellt werden, indem beispielsweise Kontaktanzeigen als Textsorte zunächst aus einer oder mehreren Publikationsformen (Zeitungen, Zeitschriften, Webseiten) extrahiert werden, um sie anschließend zu einem textsortenspezifischen Korpus zusammenzufassen.

In der Sammlung des Deutschen Referenzkorpus (Institut für Deutsche Sprache 2013), die immerhin mehr als 350 Korpora umfasst, finden sich unter den frei verfügbaren Korpora lediglich fünf Korpora, die jeweils nur eine Textsorte umfassen:

¹² Es gibt durchaus Bemühungen, innerhalb publikationsspezifischer Korpora die jeweiligen Textsorten zu erkennen und zu annotieren. In der Dokumentation des COSMAS-II-Projektes zum Deutschen Referenzkorpus wird aber beispielsweise darauf hingewiesen, dass das Inventar der verwendeten Textsorten historisch gewachsen sei und keiner linguistisch definierten Systematik folge (Programmbereich Korpuslinguistik 2012).

ein Märchenkorporus (GRI), ein Handbuchkorporus (H88), ein Belletristikkorpus mit Romanen (DIV), die Plenarprotokolle verschiedener Parlamente (P) und die Wikipediaartikel (WPD₁₁).

Aus diesem Bestand konnten aufgrund dieser und weiterer Einschränkungen (vgl. Kapitel 4) nur drei textsortenspezifische Korpora mit Kriminalromanen, Plenarprotokollen des Bundesrates und Onlinelexikoneinträgen genutzt werden.

Um den Untersuchungsbereich dennoch nicht zu stark einzuschränken, habe ich mich daher zunächst entschieden, auch zwei Publikationsformen (Fachzeitschriften und Tageszeitungen) mit allen darin enthaltenen Textsorten in die Untersuchung miteinzuschließen und ein zusätzliches textsortenspezifisches Korpus mit Print-Werbeanzeigen, in denen ein Schwerpunkt wortinterner Schreibvarianten zu erwarten ist, selbst zu erstellen. Im Vergleich zu den Textsorten sind die Publikationsformen zwar weniger stark spezifiziert, es gibt aber verbindende Elemente innerhalb der Publikationsformen über die darin enthaltenen Textsorten hinweg. Beispielsweise ein redaktioneller Entstehungsprozess der Zeitungstexte und Texte der Fachzeitschriften, das Fachvokabular der fachsprachlichen Texte oder die Nähe der Sprache der Zeitungstexte zur Standardsprache.

Komplettiert wird die Auswahl an Textformen schließlich durch drei Kommunikationsformen (Diskussionsbeiträge, private E-Mails und SMS), um ein Übergewicht redaktionell geprägter Textformen zu vermeiden und den Bereich der computervermittelten Kommunikation hinlänglich zu berücksichtigen.

Kommunikationsformen sind im Gegensatz zu Textsorten ausschließlich definiert über die „*situative Konstellation inklusive [der Verwendung] eventueller technischer Apparate*“, während Trägermedium, sprachliche Merkmale, behandelte Themen und kommunikative Funktion keine entscheidende Rolle spielen (Adamzik 2016: 62). Weitere Beispiele für Kommunikationsformen sind:

(10) Brief, Telefonat, Chat, Face-to-Face, Newsgruppen

Kommunikationsformen sind damit nicht mehr als ein situativ medialer Rahmen, innerhalb dessen sich verschiedene Textsorten als kommunikative Handlungsmuster realisieren lassen (vgl. Dürscheid 2005: 16).

Der Begriff der *Textform* dient damit in dieser Untersuchung als neutraler Sammelbegriff für diese Textsorten, Publikations- und Kommunikationsformen, um die

wortinterne Schreibvariation in Abhängigkeit von der Textform bzw. deren typischen Merkmalen zu erfassen. Eine ausführlichere Besprechung der im einzelnen untersuchten Textsorten, Publikations- und Kommunikationsformen sowie eine detailliertere Begründung der Auswahl finden sich in Kapitel 4.

3. Theoretische Grundlagen

In diesem Kapitel wird das theoretische Fundament der Arbeit gelegt. Der Schwerpunkt liegt dabei auf dem zugrundeliegenden Graphematikverständnis, das im Teilkapitel 3.1 detailliert ausgeführt wird anhand der graphematischen Einheiten von Buchstaben, über Grapheme und graphematische Silben bis hin zu graphematischen Wörtern.

Darauf folgt ein Blick auf die im Sinne der Untersuchung relevanten orthographischen Normen unter 3.2 sowie eine ausführlichere Diskussion der wortinternen Schreibvarianten (Zusammenschreibung, Bindestrichschreibung, Getrennschreibung, Schreibung mit Binnenmajuskel und Schreibung mit Apostroph) im Einzelnen in 3.3.

Dieses Kapitel legt damit den Grundstein für die Formulierung der Untersuchungshypothesen im darauf folgenden Methodikkapitel unter 4.1.

3.1 Graphematik

Graphematik beschäftigt sich mit dem System der Schrift, also schriftlich fixierter Sprache von Texten bis hin zu ihren kleinsten Bestandteilen. Deutlich abzugrenzen ist sie von der Orthographie, die als anerkannte Norm vorgibt, wie geschrieben werden soll.

Die vorliegende Untersuchung versteht sich als graphematische Arbeit. Da aber Korpusbelege untersucht werden, darf die Orthographie als Einflussgröße auf die Schriftsprache auch nicht außer Acht gelassen werden. Die relevanten Aspekte werden daher in Kapitel 3.2 näher beleuchtet.

Zunächst soll es aber darum gehen, welches Verständnis von Graphematik dieser Untersuchung zugrundeliegt, bzw. wo die geschriebene Sprache innerhalb der

linguistischen Theoriebildung verortet wird.

Der Status der Schriftsprache ist bis in die aktuelle Forschung hinein durchaus umstritten und reicht von der Vorstellung, dass die geschriebene Sprache einseitig aus der gesprochenen abgeleitet wird (Dependenzhypothese), bis hin zur Idee, dass es sich bei Laut- und Schriftsprache um voneinander unabhängige Systeme handelt (Autonomiehypothese).¹ Gemäß der Autonomiehypothese müssen Laut- und Schriftsprache also überspitzt als sich gegenseitig fremde Sprachen mit unabhängiger Grammatik aufgefasst werden (vgl. Glück 2016a), während die vor allem in der älteren Schriftlinguistik vertretene Dependenz- bzw. Abhängigkeitshypothese der Schriftsprache jegliche Autonomie abspricht und sie als reines Derivat der gesprochenen Sprache betrachtet.

Beide Positionen sind als Extreme problematisch. So wird die Dependenzhypothese der Tatsache nicht gerecht, dass es nicht nur Phänomene in der Schriftsprache gibt, beispielsweise im Bereich der Interpunktion, die keine Entsprechung in der Lautsprache haben, sondern darüber hinaus auch optisch-visuelle Eigenschaften der Schrift auf die geschriebene Sprache zurückwirken können, wie im Fall von Spelling Pronunciation: *EU* als Abkürzung für *Europäische Union* wird nicht *eu-u*, sondern *e-u* ausgesprochen. (vgl. Dürscheid 2016: 40). Die Autonomiehypothese in ihrer extremen Form verkennt dagegen die offensichtlichen Gemeinsamkeiten von Laut- und Schriftsprache, die gerade in Alphabetsprachen ein gemeinsam zugrundeliegendes System nahelegen.²

Zwischen diesen extremen Positionen sind die Interdependenzhypothese und die Korrespondenzhypothese als gemäßigtere Perspektiven auf das Verhältnis von Schrift- und Lautsprache zu verorten.

Nach der Interdependenzhypothese ist die geschriebene Sprache zwar als autonomer Forschungsgegenstand zu begreifen, sie geht aber dennoch davon aus, dass die gesprochene Sprache „*stets das Modell für die Verschriftlichung darstellt*“ (Glück 2016b: 301). Dementsprechend kann die Interdependenzhypothese als gemäßigte Dependenzhypothese aufgefasst werden.

¹ Die Diskussion, die sich in Folge von Primus (2000) entwickelt, zeigt am Beispiel des <h> als Dehnungszeichen im Deutschen wie sich eher autonomieorientierte (Primus (2000), Neef & Primus (2001)) und eher dependenzorientierte Perspektive (Ossner (2001a), Ossner (2001b)) gegenüberstehen.

² Zur ausführlicheren Diskussion der Argumente für und wider Dependenz- und Autonomiehypothese vgl. Dürscheid (2016: 36-42).

Die Korrespondenzhypothese dagegen ist eine Zwischenposition mit Nähe zur Autonomiehypothese. Die Mäßigung beruht auf der Grundannahme, dass sich die Parallelen in Laut- und Schriftsprache (Korrespondenzen) dadurch ergeben, dass sich beide aus dem selben, modalitätsunabhängigen System ableiten; also Laut- und Schriftsprache eine gemeinsame Grundlage haben, durch die sich wechselseitige Korrespondenzen ergeben, sie sich aber auch unabhängig voneinander entwickeln, wodurch jeweilige Eigenheiten der Laut- und Schriftsprache zu erklären sind.

Abbildung 3.1 nach Bredel (2008) verdeutlicht, dass gemäß der Korrespondenzhypothese nicht nur Laut- und Schriftsprache, sondern auch Zeichensprache und Blindenschrift als verschiedene Oberflächenformen eines modalitätsunabhängigen Systems aufgefasst werden können, das für die Gemeinsamkeiten zwischen diesen Oberflächenstrukturen verantwortlich ist, während individuelle Eigenheiten dem Sprachmedium (Material, Produktion, Rezeption und Kommunikationssituation) geschuldet sind.

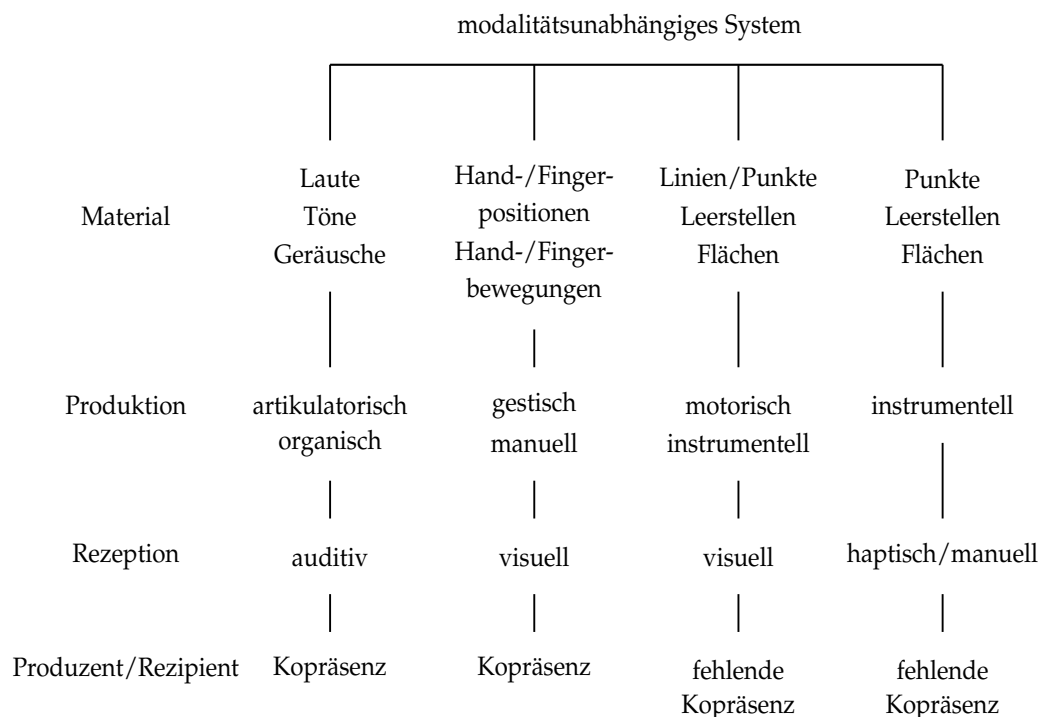


Abbildung 3.1.: Sprache als modalitätsabhängiges System nach Bredel (2008)

In der vorliegenden Untersuchung werde ich der Korrespondenzhypothese folgen, wie sie Primus (2003) skizziert hat, die ihren Ursprung aber schon in ihrem Beitrag zu den Dehnungszeichen im Deutschen nimmt (Primus 2000).

Für den damit einhergehenden Autonomiegrad, der der Schriftsprache so zugesprochen wird, spricht auch, dass die Schriftsprache sich zwar historisch nach der Lautsprache entwickelt hat und diese Reihenfolge in der Regel auch beim Spracherwerb eingehalten wird, aber das Erlernen der Schriftsprache, wie die Arbeit mit gehörlosen Kindern oder das Erlernen der lateinischen Sprache zeigt, keineswegs voraussetzt, dass die Lautsprache beherrscht wird (vgl. Dürscheid 2016: 39).

Damit ist auch klar, dass die Graphematik hier als eigenständige linguistische Teildisziplin neben Phonologie, Morphologie, Syntax, Semantik und Pragmatik wahrgenommen wird. Ihr autonomer Untersuchungsgegenstand ist die Schriftsprache als Oberflächenstruktur eines dynamischen Sprachsystems. Dabei ist es selbstverständlich, dass diese Oberflächenstruktur zunächst graphematisch zu analysieren ist, aber auch phonologische, morphologische, syntaktische und semantische Informationen abbildet.³ Im Sinne dieser graphematischen Analyse der Oberflächenstruktur werden im Folgenden die Einheiten der Schriftsprache ausgehend vom Buchstaben bis hin bis zum Wort besprochen.

3.1.1 Buchstaben

Buchstaben sind die kleinsten segmentalen Einheiten der Schriftsprache. Gerade in Druckschriften sind sie anhand kleiner Zwischenräume zweifelsfrei als einzelne Einheiten zu identifizieren.

Die neuere Forschung geht allerdings noch einen Schritt weiter und fasst die Buchstaben des Modernen Römischen Alphabets selbst als komplexe Einheiten auf, die wiederum auf einer subsegmentalen Ebene aus mehreren (funktionalen) Elementen (Kopf und Koda) bestehen.⁴ Eine systematische Darstellung dieses Konzepts inklusive einer ausführlichen Besprechung vorangegangener Arbeiten zur merkmalsbasierten Formanalyse von Buchstaben, die hier nicht geleistet werden kann, bietet Primus (2004, 2006).

Grundlegend ist die Beobachtung, dass die Minuskeln des lateinischen Alphabets in gedruckten Antiquaschriften kompositionell sind und jeweils aus einem Kopf und

³ In Kapitel 6 zeigt sich, dass dieses Graphematikverständnis auf Grundlage der Korrespondenzhypothese auch das Kandidatenkonzept im Zuge der optimalitätstheoretischen Modellierung beeinflusst.

⁴ Eine ähnliche Herangehensweise und Analyse findet sich bei Bredel (2008) für die Interpunktionszeichen.

einer Koda bestehen. Köpfe können neben dem dem Mittelband, das alle Buchstaben ausfüllen, auch ins Ober- oder Unterband hineinragen. Abbildung 3.2 zeigt neben einem kompakten Buchstaben |a|, dessen gebogener Kopf nicht über das Mittelband hinausragt, auch exemplarisch die Buchstaben |b| mit einem Kopf, der ins Oberband hineinragt, und |p|, dessen Kopf sich über Mittelband und Unterband erstreckt.

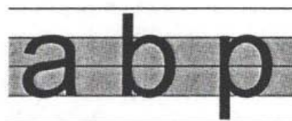


Abbildung 3.2.: Vertikale Ausdehnung von Minuskeln (Primus 2006: 13)

Kopf und Koda lassen sich anhand zweier einfacher Regeln unterscheiden:

1. Lange Elemente sind Köpfe; Kodas können nicht lang sein.
2. In den anderen Fällen ist der Kopf der Bestandteil, der das Mittelband auf dem kürzesten Weg ausfüllt.

(Fuhrhop & Peters 2013: 193)

Primus (2004) geht auf Basis der Form allerdings einen entscheidenden Schritt weiter, indem sie Buchstabenmerkmalen Lautmerkmale bzw. Lautklassen zuordnet und damit die Analyse um einen funktionalen Aspekt erweitert. So ist beispielsweise die Artikulationsart durch die Länge der Buchstabenköpfe visualisiert: Plosiven entsprechen Buchstaben mit langen Köpfen |p, t, k, b, d|, schräge oder nicht gerade lange Köpfe |f, v, w, s| entsprechen Frikativen, während konsonantische Sonoranten mit kurzen Köpfen und oberem Koda-Anschluss |m, n, r| korrespondieren und Vokalen kompakte Buchstaben |a, e, o, i, u| zugeordnet werden (vgl. Berg et al. 2016: 341).

Die systematische Einteilung der Buchstaben nach der Länge ihrer Köpfe (und dem Koda-Anschluss) in lang |b, d, k, h, t, f, p, q, j, g|, schräg |z, x, v, w, s|, kurz und gerade mit oberem Koda-Anschluss |n, m, r|, kurz und gerade ohne oberen Koda-Anschluss |i, u| und kurz und gebogen |a, e, o| bildet schließlich die Grundlage für das Konzept der graphematischen Silbe (vgl. Kapitel 3.1.3).

Für eine detailliertere Betrachtung einzelner Buchstaben und ihrer Segmente inklusive diskussionswürdiger Fälle wie |c, y, ß| sei auf Fuhrhop & Peters (2013: 191-200) verwiesen.

3.1.2 Alphabetische Grapheme

Im Sinne des oben erläuterten Graphematikverständnisses werden alphabetische Grapheme, die häufig, wie auch hier, einfach als Grapheme bezeichnet werden, unabhängig von ihrer lautlichen Entsprechung als kleinste bedeutungsunterscheidende Elemente der Schriftsprache definiert (vgl. Günther 1988: 77).⁵ Diese Funktionalität unterscheidet die Grapheme von Buchstaben, die nicht alle bedeutungsunterscheidende Funktion haben.⁶

Das Mittel der Wahl zur Bestimmung von alphabetischen Graphemen einer Sprache ist die Bildung von Minimalpaaren, bei denen durch den Austausch eines Graphems ein neues Lexem entsteht, wie bei *Mut* – *Wut*. Dabei wird deutlich, dass im Deutschen nicht jeder Buchstabe einem Graphem entspricht. Als problematisch erweisen sich insbesondere |c| und |q|.

So hat |q| im Deutschen alleine keinen Graphemstatus, weil es im Kernwortschatz kein Minimalpaar gibt, das eine bedeutungsunterscheidende Funktion von |q| belegt. Erst in Kombination mit |u| ergibt sich das komplexe Graphem <qu> (*quer* – *wer*).⁷ Ähnlich verhält sich |c|. Dem Buchstaben folgen abseits von Fremdwörtern ausschließlich |h| oder |k|. Dementsprechend finden sich neben Minimalpaaren, die |ch| und |ck| als Grapheme nahelegen wie *wachen* – *waren* und *Hecke* – *Hexe* auch solche, in denen sich |c| ersetzen lässt wie in (1).

- (1) a. wecken – welken; Bäcker – Bänker
b. Rechen – Reihen; verzechen – verzeihen

⁵ Wird das Graphem allgemeiner definiert, wie beispielsweise bei Gallmann (1985: 10-14), müssen auch Interpunktionszeichen, Leerzeichen, Ziffern und Sonderzeichen als Grapheme in Sinne diskreter, funktionaler Einheiten gelten. Vgl. dazu den folgenden Abschnitt zu nicht-alphabetischen Graphemen.

⁶ Ein weiterer Ansatz wäre die Definition des Graphems über den Bezug zum korrespondierenden Phonem, der sich allerdings nicht mit dem hier vertretenen Graphematikverständnis verträgt und auch darüber hinaus problematisch ist. Vgl. z. B. Günther (1988: 71-79).

⁷ Günther (1988: 82) verweist zwar auf das Minimalpaar *Quelle* – *Duelle* und Fuhrhop & Peters (2013: 204) auf *Qual* – *Dual*, allerdings handelt es sich bei *Duelle* und *Dual* nicht um Bestandteile des Kernwortschatzes und es gibt kein Lexem, geschweige denn Minimalpaar, dass ein Vorkommen von |q| ohne nachfolgendes |u| belegt.

- c. Backe – Bache; nickt – nicht; nackt – Nacht⁸

Anders als bei <qu> lässt sich, wie Beispiel (1c) zeigt, aber auch der zweite Buchstabe austauschen, zumindest insofern, als es Minimalpaare gibt, die sich nur durch die Buchstaben |h| und |k| unterscheiden. Das spricht meiner Meinung nach aber nicht dafür |c| als eigenständiges Graphem anzunehmen, sondern stützt die Annahme zweier weiterer komplexer Grapheme <ch> und <ck> neben <qu>. Die Beispiele unter (1) sprechen nämlich dafür, dass zwar <k> und <h> Grapheme sind, |c| im Deutschen aber nur in Kombination Graphemstatus erlangen kann als <ch> oder <ck>.

Damit sind Grapheme im Unterschied zu Buchstaben suprasegmentale Einheiten der Schriftsprache und es ergibt sich für das Deutsche folgendes Grapheminventar:

- (2) einfache Grapheme:

a, b, d, e, f, g, h, i, j, k, l, m, n, o, p, r, s, t, u, v, w, x, z, ß, ä, ö, ü

komplexe Grapheme:

qu, ch, ck

Aufgrund des Verhaltens bei der Worttrennung am Zeilenende werden auch graphematische Diphtonge (vgl. Berg et al. 2016: 350) und |sch| (vgl. Fuhrhop & Peters 2013: 205-206), die dort eben nicht getrennt werden können, als Grapheme diskutiert, obwohl entsprechende Minimalpaare existieren:

- (3) a. Keule – Kehle, Leute – Laute, Kiefer – Kiffer, Tier – Teer
b. Masche – manche, Esche – Elche – Espe

Buchmann (2015) liefert allerdings noch ein weiteres Argument, welches das Ergebnis der Minimalpaaranalyse stützt. So werden für Abkürzungen ausschließlich Grapheme verwendet und keine Buchstabenverbindungen – auch wenn sie am Zeilenende nicht trennbar sind – wie z. B. *EU* und nicht *EuU* für *Europäische Union* und *NVS* für die *Nahverkehr Schwerin GmbH* statt *NVSCH*.

⁸ Die Austauschmöglichkeiten von |c| sind in allen gezeigten Fällen beschränkt. Während das bei den Minimalpaaren unter (1c) daran liegt, dass die Grapheme <ch> und <ck> miteinander kontrastieren, führen silbische Beschränkungen dazu, dass sich |c| vor |k| nur durch Konsonanten (1a), vor |h| aber nur durch Vokale, womöglich sogar nur durch |i| ersetzen lässt (1b).

Vor der Rechtschreibreform gab es auch die Option, |ck| am Zeilenende zu trennen wie in *Zuk-ker*. In Anbetracht der Minimalpaaranalyse unter (1) war es aber nur folgerichtig diese Norm zu verwerfen und ist ein weiterer Anlass, die Worttrennung am Zeilenende als Indikator für das Erkennen von Graphemen kritisch zu überdenken.

Dennoch wird in Kapitel 3.1.3 besprochen, wie sich die Untrennbarkeit von Schreibdiphthongen und |sch| erklären lässt, ohne dass man den Buchstabenverbindungen einen Graphemstatus einräumt, den man ihnen aufgrund der Minimalpaarbildung nicht zusprechen kann. Die Beispiele unter (3) zeigen nämlich, dass jeweils beide Graphembestandteile austauschbar sind.

Nicht-alphabetische Grapheme

Neben den oben besprochenen alphabetischen Graphemen gibt es noch weitere „diskrete graphische Einheiten“ der Schrift (Gallmann 1996: 1456). Beispielsweise Ziffern (0,3, 17,459), die keine bedeutungsunterscheidende Funktion haben, sondern selbst Bedeutung tragen und miteinander kombinierbar sind, Sonderzeichen (% , & , #), die wie Ziffern unmittelbar Bedeutung tragen, aber nicht kombinierbar sind, und Interpunktions- sowie Leerzeichen, die im Folgenden noch näher betrachtet werden (vgl. Bredel 2011: 7-9).

Nach Gallmann (1996: 1456) sind all diese Zeichen als Segmente,⁹ also diskrete graphische Mittel der geschriebenen Sprache, mit bestimmbarer Funktion auch Grapheme,¹⁰ wobei „Buchstaben“ [sic!] das Grundgerüst alphabetischer Sprachen bilden.¹¹

Innerhalb der Grapheme nach Gallmann sind damit Untergruppen anhand ihrer Funktion zu unterscheiden: Ziffern und Sonderzeichen tragen unmittelbar Bedeutung, alphabetische Grapheme haben bedeutungsunterscheidende Funktion, und

⁹ Daneben gibt es auch suprasegmentale Einheiten in der Schrift, beispielsweise Umrandungen, Absätze oder Unterstreichungen, mit denen Hervorhebungen vorgenommen werden können, die aber nicht dazu dienen, Einheiten des Sprachsystems zu visualisieren (vgl. Bredel 2008: 22).

¹⁰ Gallmann (1985: 11) fasst Buchstaben, Hilfszeichen (Syngrapheme), Leerzeichen, Ziffern, Sonderzeichen und diakritische Zeichen sogar als „Grapheme (im engeren Sinn)“ zusammen in Abgrenzung zu sonstigen graphischen Mitteln und vertritt damit einen weiten Graphembegriff.

¹¹ Dass eine Gleichsetzung von Buchstaben und Graphemen problematisch ist, wurde oben schon thematisiert, und soll hier nicht weiter diskutiert werden.

Syngrapheme dienen nach ihm ebenso wie Leerzeichen zur Erleichterung des Lesens, indem sie segmentieren und/oder klassifizieren.

Bredel (2008, 2011) dagegen klassifiziert die segmentalen Mittel der Schrift mit Hilfe der Kriterien Darstellbarkeit, Verbalisierbarkeit, Kombinierbarkeit und Zweielementigkeit in Buchstaben, Ziffern, Sonderzeichen, Interpunktionszeichen und Leerzeichen.¹²

Tabelle 3.1.: Merkmale segmentaler Mittel der Schrift nach Bredel (2011:9)

	Buchstaben	Ziffern	Sonderz.	IP-Zeichen	Leerzeichen
darstellbar	+	+	+	+	–
verbalisierbar	+	+	+	–	–
kombinierbar	+	+	–	–	–
zweielementig	+	–	–	–	–

Leerzeichen erfüllen keines der Kriterien, da sie im Gegensatz zu den anderen segmentalen Mitteln nicht ohne graphischen Kontext dargestellt oder erfasst werden können. Alle weiteren Gruppen sind mit oder ohne Kontext darstellbar. Interpunktionszeichen tragen ausschließlich das Merkmal der Darstellbarkeit unabhängig vom Kontext, während Sonderzeichen zusätzlich auch verbalisiert werden können. Ziffern und Buchstaben lassen sich darüber hinaus miteinander kombinieren, Ziffern zu Zahlen und Buchstaben zu Wörtern. Das Alleinstellungsmerkmal von Buchstaben, die alle Kriterien erfüllen, ist damit die Zweielementigkeit, die paradigmatisch zu verstehen ist. Nur Buchstaben haben jeweils zwei mögliche Formen, den Groß- und den Kleinbuchstaben (vgl. Bredel 2011: 7-9).¹³

Anders als Gallmann (1996) macht Bredel (2008) keine ausdrücklichen Angaben dazu, ob bzw. welche segmentalen Mittel der Schrift Grapheme sind. Statt des Begriffes *Syngraphem* verwendet sie für ein anders definiertes, aber nahezu identisches Inventar an Zeichen den neutraleren Begriff des *Interpunktionszeichens*.¹⁴

¹² Bredel (2008: 22-23) berücksichtigt auch noch Diakritika, die sie anhand des Alleinstellungsmerkmals der Additivität von den anderen segmentalen Mitteln unterscheidet.

¹³ Bis vor Kurzem gab es unter den Buchstaben des Alphabets mit |ß| eine Ausnahme hinsichtlich dieses Merkmals. Mit der Veröffentlichung des aktualisierten Regelwerks, das |ß| auch als Großbuchstaben zulässt, ist von normativer Seite hier nun ebenfalls von einem zweielementigen Paradigma auszugehen (vgl. Rat für deutsche Rechtschreibung 2018: 15).

¹⁴ Aber auch Bredel (2008, 2011) verwendet zur Angabe des Inventars die Graphemschreibweise in

(4) a. Gallmann (1996: 1456):

Syngrapheme sind diskrete graphische Einheiten, deren Funktion das Segmentieren und/oder Klassifizieren ist.[...] Syngrapheme sind damit eine Subklasse der Grapheme.[...]und dienen zur] Erleichterung des Lesens

Inventar: <. ? ! : ; , - () „“ - ' >

b. Bredel (2011: 9):

Ein erstes Ergebnis dieser Analyse ist, dass die Interpunktionszeichen nun definiert werden können als darstellbare, nicht-verbalisierbare, nicht-kombinierbare, einelementige Segmente der Schrift.

Inventar: <. ? ! : ; , - () „“ - ' ...>

Trotz unterschiedlicher Herangehensweise – Gallmann definiert funktional (zum Segmentieren/Klassifizieren und zur Erleichterung des Lesens), Bredel durch ein ganzes Bündel von Eigenschaften (darstellbar, nicht-verbalisierbar, nicht-kombinierbar, einelementig) – unterscheidet sich das Inventar der Syngrapheme und Interpunktionszeichen also nur bezüglich <...>.¹⁵

Darüber hinaus diskutiert Bredel (2008: 24) noch den Asterisk und den Schrägstrich als potentielle Interpunktionszeichen. Beide könnten aber in bestimmten Kontexten verbalisiert werden (*22.11.2015: *geboren am 22.11.2015; Schüler/innen: Schüler und/oder Schülerinnen*), so dass sie bei einer strengen Auslegung des Kriteriums der Verbalisierbarkeit nicht zu den Interpunktionszeichen zu rechnen seien, ungeachtet dessen, dass der Schrägstrich auch ohne Verbalisierung verwendet wird: *Bündnis90/Die Grünen*. Beide graphischen Segmente wären demnach den Sonderzeichen zuzuordnen.

Für den Asterisk gibt Bredel (2008) leider keinen Kontext an, in dem er nicht verbalisiert wird, weshalb nicht ganz klar wird, warum er überhaupt als Interpunktionszeichen diskutiert werden muss, denkbar wäre aber die Verwendung des Asterisks zur Kennzeichnung von Fußnoten.

Im Folgenden wird hier der Begriff des *Interpunktionszeichens* konsequent dem des *Syngraphems* vorgezogen, um begriffliche Unklarheiten zu vermeiden. *Graphem* wird

spitzen Klammern.

¹⁵ Bredel (2008: 23-24) verweist noch auf eine ganze Reihe weiterer Inventarisierungen, die zum Teil auch paarige Kommata und Gedankenstriche, den Schrägstrich oder das Spatium als Interpunktionszeichen zählen, ist aber von den Kriterien, nach denen diese Inventare zusammengestellt werden, nicht überzeugt.

hier, wie oben schon angekündigt, im ganz engen Sinne für alphabetische Grapheme verwendet, ohne dadurch suggerieren zu wollen, dass andere segmentale Einheiten der Schrift keinen Graphemstatus haben.

Für die wortinterne Schreibvariation an morphologischen Grenzen komplexer Wörter ist nur ein Teil der Interpunktionszeichen relevant.

Buchmann (2015: 90-93) unterscheidet das Inventar der Interpunktionszeichen von Bredel (2008, 2011) in diesem Zusammenhang nach ihrem Wirkungsbereich in Satz- und Wortzeichen. Zu den Wortzeichen zählt sie dabei ausschließlich den Abkürzungspunkt, den Divis und den Apostroph < - ' > (vgl. Buchmann 2015: 90).

(5) Abk., Uni-Bibliothek, Susi's

Buchmann (2015: 93) gewinnt als zentrale Einsicht auf Basis ihrer Datenanalyse die Erkenntnis, dass diese Wortzeichen sich nicht nur durch ihre Form und Funktion, sondern auch durch ihre grammatischen Operationsdomänen unterscheiden (vgl. Tabelle 3.2).

Alle Wortzeichen operieren als graphische Zeichen auf der graphematischen Ebene (Graph); der Abkürzungspunkt allerdings ausschließlich auf der graphematischen Ebene. Er beende graphische Einheiten, die nicht ausgesprochen werden, wodurch Abkürzungen zunächst graphische Synonyme ihrer Langform seien (vgl. Buchmann 2015: 93).

Tabelle 3.2.: Operationsdomänen von Wortzeichen nach Buchmann (2015: 278)

	Graph	Phon	Flex	Der	Komp	Syn
Abkürzungspunkt	+	-	-	-	-	-
Apostroph	+	+	+	+	-	-
Trennstrich	+	+	-	+	+	-
Ergänzungsstrich	+	+	-	+	+	-
Bindestrich	+	+	-	+	+	+

Der Apostroph wirkt darüber hinaus auch auf der phonologischen Ebene (Phon) wie bei *Käpt'n* oder in bestimmten Kontexten auf den Ebenen der Flexion (Flex) bzw. Derivation (Der) wie bei *Lisa's Laden* oder *Grimm'sche Märchen*. Die Operationsdomäne der Flexion wird sogar exklusiv nur durch den Apostroph bedient, da auch die

verschiedenen Ausprägungen des Divis hier keine Verwendung finden.

Den breitesten Anwendungsbereich hat der Divis als Bindestrich zu bieten. Abseits der Flexionsmorphologie lässt er sich in allen Operationsdomänen anwenden, einschließlich Kompositionsmorphologie (*Linguistik-Seminar*) und der Syntax („*Jetzt-ist-Schluss-mit-lustig*“-*Ansage*). Ergänzungsstrich und Trennstrich operieren nicht auf der syntaktischen Domäne, bespielen aber beide durchgängig die Domäne der Kompositionsmorphologie (*Abend- und Mittagessen*, *Mittag-essen*) und operieren auch auf der Ebene der Derivationsmorphologie, dort aber nur eingeschränkt. So führen vokalisch anlautende Derivationsuffixe dazu, dass der Trennstrich aufgrund der Silbenstruktur nicht in Übereinstimmung mit der morphologischen Fuge gesetzt, sondern vorgezogen wird: *kin-disch* (**kind-isch*) (vgl. Buchmann 2015: 94). Bredel (2008: 112) weist zudem mit dem Beispiel **üb- und lieblich* darauf hin, dass auch nicht alle Derivationen einen Ergänzungsstrich an der Morphemgrenze erlauben. Nach welchen Kriterien der Einsatz eines Ergänzungsstriches bei Derivaten möglich bzw. unmöglich ist, wurde bis heute nicht abschließend beantwortet und kann auch hier nicht ausführlich diskutiert werden. Für einen ersten Versuch der Systematisierung sei auf Höhle (1982: 89-92) verwiesen, der diese Fähigkeit von Derivationsaffixen im Vergleich zu Kompositionsgliedern diskutiert.

Alle drei Zeichen operieren also auf der Wortebene, aber teilweise in verschiedenen Operationsdomänen. Darüber hinaus sind sie auch eingeschränkt hinsichtlich ihrer Position. So wird der Abkürzungspunkt ausschließlich in einer finalen Position eingesetzt. „*Er beendet graphische Einheiten, die nicht ausgesprochen werden*“ (Buchmann 2015: 93). Final heißt in diesem Zusammenhang aber nicht unbedingt, dass ein Leerzeichen folgt, sondern auch ein Bindestrich wie in *Abk.-Verz.* oder ein Groß- bzw. Kleinbuchstabe (*akt.Aufl., u.a.*) möglich sind (vgl. Buchmann 2015: 200).¹⁶ Entgegen der ersten Intuition sind auch in diesen Kontexten die Abkürzungspunkte final, wenn man akzeptiert, dass ein Bindestrich zwei graphisch in sich geschlossene Einheiten verbindet, und sich vor Augen führt, dass es sich bei den Abkürzungen

¹⁶ Buchmann (2015: 200) geht zudem davon aus, dass links vom Abkürzungspunkt obligatorisch ein Buchstabe steht. Der Punkt tritt abseits seiner Funktion als Satzzeichen aber auch in Kombination mit Ziffern zur Kennzeichnung von Ordinalzahlen wie 1., 2., 3. (*erstens, zweitens, drittens/erster, zweiter, dritter*) in einem Kontext auf, wo man ihn auch als Wortzeichen bzw. Abkürzungspunkt auffassen könnte, da er sich durchaus ähnlich verhält.

wie *a.Aufl.*, *u.a.* zumindest auch bei der Auflösung *aktualisierte Auflage* und *unter anderem/und andere* Leerzeichen an eben den Stellen in Erscheinung treten, wo sie innerhalb der Abkürzungen in der Schreibpraxis entgegen der Norm gerne ausgelassen werden. Das Regelwerk äußert sich zwar nicht explizit dazu, aber in allen Abkürzungsbeispielen in § 101 folgt auf den Punkt entweder ein Leerzeichen oder ein Bindestrich (Rat für deutsche Rechtschreibung 2018):

(6) lfd. Nr. (laufende Nummer), z. B. (zum Beispiel), Abt.-Leiter (Abteilungsleiter)

Die verschiedenen Funktionen des Divis als Bindestrich, Trennstrich oder Ergänzungsstrich bedingen ebenfalls verschiedene mögliche Positionen.¹⁷ Als Ergänzungsstrich besetzt er die finale oder initiale Position graphischer Einheiten in Kombination mit Leerzeichen oder Komma als Satzzeichen (am Zeilenende ist natürlich auch ein zusätzlicher Zeilensprung möglich):

- (7)
- a. Kinder- und Jugendfreizeit
 - b. Lebensmittelproduktion und -verarbeitung
 - c. Baby-, Kleinkind-, und Schulkindphase

Als Trennstrich oder Bindestrich wird er dagegen ausschließlich wortintern verwendet. Dabei sind Bindestriche auf morphologische Grenzen innerhalb von Wörtern beschränkt (*Notfall-Nummer*), während eine Worttrennung am Zeilenende mittels Divis nicht nur an morphologischen (*Not-fall-Nummer*), sondern auch an silbischen Grenzen erfolgen kann wie z. B. in *Notfall-Num-mer* und zwangsweise mit einem Zeilensprung kombiniert ist. Anders als der Ergänzungsstrich können der wortinterne Trennstrich und Bindestrich nicht mit einem Satzzeichen kombiniert werden. Der Bindestrich tritt aber in Kombination mit anderen Interpunktionszeichen auf, sofern diese Bestandteile der durch den Bindestrich verknüpften Wortteile sind:

- (8)
- a. Abt.-Leiter
 - b. „Wenn’s unbedingt sein muss“-Zusage

Die Beispiele ab (7) machen auch noch einen weiteren Unterschied deutlich, nämlich, dass nur der Bindestrich im Gegensatz zum Ergänzungsstrich und Trennstrich –

¹⁷ Zur Graphotaktik des Divis im Detail vgl. Buchmann (2015: 279-285).

unabhängig davon, ob der an einer morphologischen oder silbischen Grenze gesetzt wird – das nachfolgende Graphem zugänglich macht für die Substantivgroßschreibung.¹⁸

Ein ähnliches Bild bezüglich der Verknüpfung von Funktion und Position zeigt sich auch beim Apostroph. Klein (2002) unterscheidet zwischen dem Elisionsapostroph, der eine Auslassung markiert, und dem Stammformapostroph, der an einer morphologischen Grenze gesetzt wird. Während der Elisionsapostroph (9a) entsprechend seiner Funktion alle Positionen besetzen kann (initial, intern, final), tritt der Stammformapostroph (9b) nur wortintern auf:¹⁹

- (9) a. So 'n Mist!; Käpt'n; Ich geb' dir was ab.
b. Antonio's Pizzeria; Beck'sche Reihe

Abseits des Elisions- und Stammformapostrophes werden auf der Grundlage einzelner Belege von Scherer (2013: 83-84) noch weitere Funktionen diskutiert, beispielsweise ein Kompositionsapostroph wie in *SCHOKO'GIRL*, *großstadt'träume* oder *Fußball'news*. Man muss aber (noch) davon ausgehen, dass es sich hier eher um Extrembeispiele als um eine regelmäßige oder gar etablierte Funktion handelt, da die Belege nicht einer systematischen Korpusrecherche, sondern willkürlichen Belegsammlungen entstammen.²⁰

In dieser Untersuchung wird aus der Gruppe der Wortzeichen nach Buchmann (2015) also nur der Divis in seiner Funktion als Bindestrich relevant sein, da lediglich der Bindestrich eine gut belegte Variationsmöglichkeit an wortbildungsmorphologischen Grenzen im Wortinneren bietet.

¹⁸ Vgl. dazu den Absatz *Großschreibung auf der Wortebene* in Kapitel 6.

¹⁹ Die Apostrophierung der Genitivformen von *Eigennamen*, deren Grundform [...] auf einen s-Laut [...] endet, wie bei *Aristoteles' Schriften*, *Felix' Vorschlag* oder *Heinz' Geburtstag* (Rat für deutsche Rechtschreibung 2018: § 96 (1)) rechne ich ebenfalls dem Elisionsapostroph zu. Zwar wird hier parallel zum Stammformapostroph die Apostrophierung stammformfinal vorgenommen, aber die folgende Auslassung der Flexionsendung -s spricht für die Elisionsfunktion.

²⁰ In der vorliegenden Untersuchung finden sich im Korpus nur einige wenige Beispiele für den Stammformapostroph und kein einziges für einen Kompositionsapostroph oder eine andere Funktion. Eine vollständige Auflistung der im Korpus gefundenen Apostrophierungen an flektionsmorphologischen Grenzen und deren Diskussion findet sich in Teilkapitel 2.1.2 unter dem Punkt *Beschränkung auf wortbildungsmorphologische Komplexität*.

Der Apostroph ist, von einigen Extrembeispielen abgesehen, auf einen kleinen Teilbereich der Flexion und Derivation beschränkt. Der Ergänzungsstrich tritt nicht als wortinterne Variationsmöglichkeit auf, sondern ersetzt innerhalb von Koordinationen einzelne Wortbestandteile komplexer Wörter, während der Trennstrich zwar wortintern verwendet wird, aber auch abseits morphologischer Grenzen, und ausschließlich am Zeilenende in Kombination mit einem Zeilensprung auftritt. Die Variation ist also extern determiniert durch die Position des Wortes am Zeilenende und ist abseits dieser Position gar nicht möglich. Der Abkürzungspunkt ist aufgrund seiner eigentlich wortfinalen Position nicht relevant. Kommt er innerhalb von Kompositionen wortintern vor, wird er von einem Bindestrich begleitet, ist selbst aber keine mögliche Variante an wortbildungsmorphologischen Grenzen im komplexen Wort, sondern Bestandteil einer Konstituente, in der er die finale Position einnimmt.

3.1.3 Graphematische Silben

Aus Graphemen setzt sich schließlich die nächste Einheit der Schriftsprache, die graphematische Silbe, zusammen.

Das Konzept der graphematischen Silbe geht zurück auf das Konzept der „*Schreibsilbe im Deutschen*“ nach Eisenberg (1989) und wurde seitdem stetig weiterentwickelt. Die Grundidee dahinter ist, parallel zur phonologischen Silbe anzunehmen, dass es mit der graphematischen Silbe auch in der Graphematik eine Einheit zwischen Graphem und graphematischem Wort gibt.²¹

Allgemeines graphematisches Silbenbaugesetz und Längenhierarchie

Während in der Phonologie die Sonoritätshierarchie die Silbenstruktur vorgibt, gilt für graphematische Silben die ebenfalls strukturalistisch hergeleitete Längenhierarchie:²²

Das Allgemeine Graphematische Silbenbaugesetz [AGS] lautet: Die Köpfe der Buchstaben innerhalb einer Silbe werden zum Silbenkern hin kontinuierlich kompakter. Im Silbenkern steht ein kompakter Buchstabe. Das Allgemeine Phonologische Silbenbaugesetz hingegen lautet: ‚Die bevorzugte Silbe ist so gebaut, dass die CS [Consonantal

²¹ In der Literatur, z. B. Evertz (2016) und Fuhrhop & Peters (2013), wird außerdem der graphematische Fuß als Einheit zwischen graphematischer Silbe und graphematischem Wort diskutiert.

²² Zur Herleitung der Längenhierarchie und des Allgemeinen graphematischen Silbenbaugesetzes vgl. Fuhrhop & Peters (2013: 218-219).

Strength/Konsonantische Stärke, N.F.;F.B.] im Anfangsrand und im Endrand zum Nukleus hin monoton abnimmt und im Nukleus ihr Minimum erreicht.' (Vennemann 1982, 283)

(Fuhrhop & Buchmann 2016: 225)

Diese Längenhierarchie wiederum fußt auf einer systematischen Analyse der Minuskeln des lateinischen Alphabets in gedruckten Antiquaschriften von Primus (2004, 2006),²³ die zeigt, dass Buchstaben kompositionell jeweils aus einem Kopf und einer Koda bestehen, und die Köpfe neben dem Mittelband, das alle Buchstaben ausfüllen, auch ins Ober- oder Unterband hineinragen können.

Durch seine Formulierung, *Die Köpfe der Buchstaben innerhalb einer Silbe werden zum Silbenkern hin kontinuierlich kompakter*, impliziert das Allgemeine graphematische Silbenbaugesetz schon, dass eine graphematische Silbe nicht nur aus einem Buchstaben bestehen kann und zumindest einen kompakten Buchstaben als Kern aufweisen muss. Fuhrhop & Peters (2013: 216) weisen in diesem Zusammenhang darauf hin, dass graphematische Silben die kleinste Einheit seien, über die Grammatikalitätsurteile getroffen werden können, beispielsweise dass die Graphemfolge <rt> keine wohlgeformte Silbe ist, <rot> dagegen schon.

Auch graphematische Silben lassen sich wie phonologische in Anfangsrand, Kern und Endrand unterteilen. Während der Kern aus einem oder zwei kompakten Buchstaben besteht, weisen die Ränder in flektierten Formen bis zu fünf nicht-kompakte Buchstaben auf (vgl. Fuhrhop & Peters 2013: 217). Die Graphemfolge <rt> ist also wohl mangels Silbenkern keine wohlgeformte graphematische Silbe.

- (10) a. grob, krank, platt, hohl
 b. krass, braun, die, grau
 c. ernst, auf, weil, halb
 d. neu, aus, zwei, wenn

Das Ideal einer deutlich bikonkaven Struktur durch Längen im Ober- oder Unterband wie in (10a) wird dabei nicht immer realisiert. Vielmehr gibt es auch graphematische Silben, die diese Überlängen nur im Anfangs- (10b) oder Endrand (10c) aufweisen, oder aber sogar gänzlich im Mittelband realisiert werden (10d).

Dennoch wird in allen Fällen unter (10) das Allgemeine Graphematische Silben-

²³ Vgl. dazu Teilkapitel 3.1.1.

baugesetz eingehalten. Zwischen langen Köpfen und kompakten, gebogenen Köpfen ordnen sich nämlich noch schräge und kurze, gerade Köpfe in die Skala der Längenhierarchie zwischen Länge und Kompaktheit ein, wie in Abbildung 3.3 dargestellt.

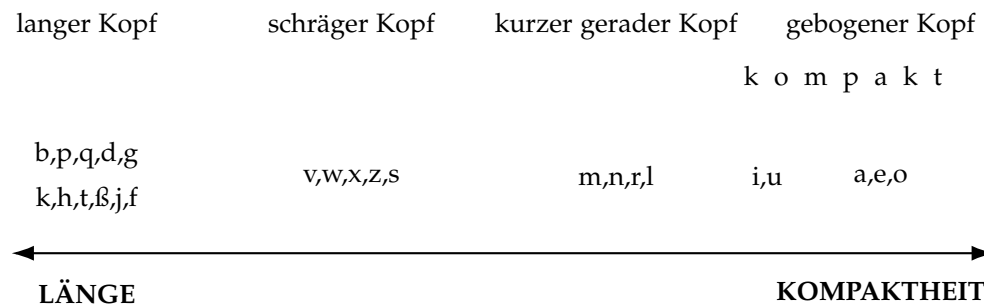


Abbildung 3.3.: Die Längenhierarchie (Fuhrhop & Peters 2013: 217)

Das Allgemeine Silbenbaugesetz hat dabei keinen Absolutheitsanspruch. Vielmehr ist es so, dass graphematische Silben bevorzugt danach aufgebaut werden, also Silben, die dem Allgemeinen graphematischen Silbenbaugesetz entsprechen, „besser“ sind, als solche, die das nicht tun (Fuhrhop & Peters 2013: 218).

Verstöße gegen das AGS kommen vor, sind aber im Deutschen (und auch im Englischen und Französischen) regelmäßig auf einzelne Buchstaben bzw. Buchstabenkombinationen beschränkt (ausführlich dazu Fuhrhop & Buchmann 2016: 362-368). Dazu zählten neben Verstößen durch |c| und |y|, die keine Grapheme des Deutschen sind, auch solche durch |s| im Anfangs- (|sp|, |st|, |sch|) oder Endrand (nach fast allen längenhaltigen Buchstaben), das |h| im Silbenn Kern oder |tz| als Endrand.

Ein extrasyllabisches |s| im Endrand (*montags*) sowie |tz| und silbenintiales oder Dehnungs-*h* in einsilbigen Formen (*Netz*, *fliehst*, *dehnt*) seien in der Regel aber funktionalisiert und zeigten die morphologische Struktur an. Diese morphologische Funktion erfüllen Verstöße durch die Kombinationen |sp|, |st| und |sch| im Anfangsrand nicht. Unproblematisch für das AGS wären sie nur, wenn man sie als eine Einheit betrachtet, die als solche das Merkmal eines langen Kopfes aufweist. Für den Trigraph |sch|, der im Deutschen genau einem Phonem /ʃ/ entspricht, liegt diese Lösung nahe.²⁴

²⁴ Ähnliches gilt für |ch|, das im Anlaut aber nur in Fremdwörtern wie z. B. *chauvinistisch* zu erwarten ist. Zu den Besonderheiten von |c| auch abseits dieser Kombination vgl. Fuhrhop & Buchmann (2016: 366f.).

Für |s| in |sp| und |st| im Anfangsrand der graphematischen Silbe argumentiere ich aber für einen funktionalisierten Verstoß als Dekodierungshilfe. Die Verletzung der Längenhierarchie durch |s| in |sp| und |st| im Onset markiert nämlich genau die Fälle, in denen |s| standardsprachlich als /ʃ/ zu lesen ist, und spiegelt zudem den Sonoritätsverstoß auf phonologischer Ebene.

Mittlerweile stützen auch erste linguistische Experimente die Konzepte der Längenhierarchie und der graphematischen Silbe. So deutet ein Leseexperiment mit 27 Probanden von Drews (2011) darauf hin, dass die Anwesenheit von langköpfigen Buchstaben und damit einer deutlichen graphematischen Silbenstruktur das Lesen von drei- und viersilbigen Wörtern beschleunigt; außerdem machten die Probanden weniger Fehler bei längenhaltigen Wörtern, wenn sie entscheiden mussten, ob es sich um Pseudowörter oder tatsächliche Wörter des Deutschen handelt.

Domahs, de Bleser & Eisenberg (2001) konnten zudem in einem Experiment mit einem an Oberflächendysgraphie leidenden Patienten feststellen, dass dieser neben zu erwartenden phonographischen Schreibungen auch nachweislich die graphematische Silbenstruktur berücksichtigt. Dafür sprechen beispielsweise die Verwendung des silbeninitialen und Dehnungs-*h* im richtigen Kontext.

Echte graphematische Grundlagenarbeit hinsichtlich des Aufbaus von Morphem im Deutschen und Englischen leistet Berg (2019) mit einer sehr detaillierten Untersuchung der Wortschreibung anhand der CELEX-Datenbank. Dabei untersucht er auch den silbischen Bau von Stämmen, listet Anfangs- und Endränder sowie Kerne und intervokalische Konsonanten(cluster) akribisch auf und analysiert sie. Obwohl er bevorzugt mit dem Begriff der Symmetrie arbeitet, nimmt er auch das Allgemeine Graphematische Silbenbaugesetz kritisch unter die Lupe und kommt für beide Sprachen zu dem Schluss, dass es sich als anwendbar erweist und Verstöße dagegen eine Systematik erkennen lassen (vgl. Berg 2019: 101-107).

Es lässt sich also festhalten, dass sich die graphematische Silbe und das Allgemeine Graphematische Silbenbaugesetz über die Buchstabenmerkmale, also rein graphematisch, bestimmen lassen und einzelsprachspezifische Verstöße in der Regel funktionalisiert sind.

Die Silbenstruktur der graphematischen Silbe

Berg et al. (2016: 351) zeigen ein Konzept, wie sich die graphematische Silbe (und daraus auch das graphematische Wort) aus den bisher besprochenen Einheiten der

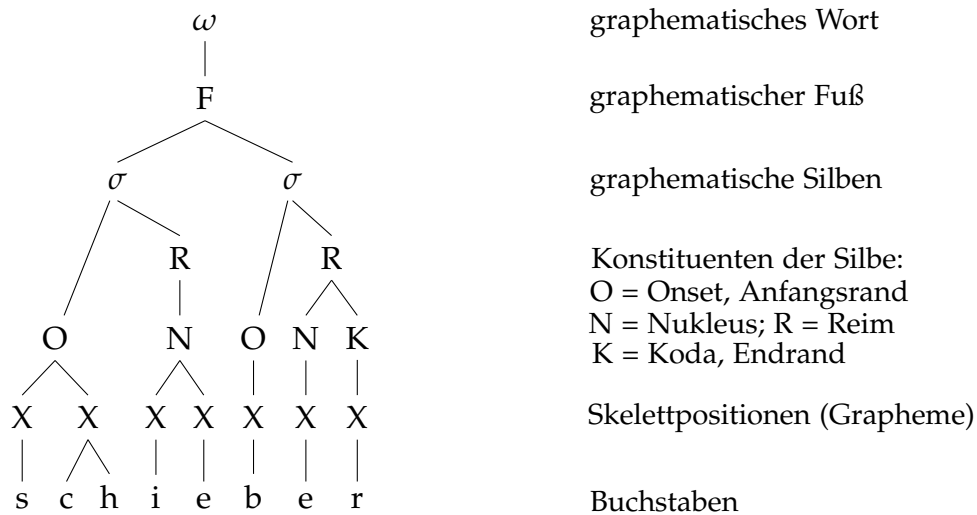


Abbildung 3.4.: Einheiten der Graphematik nach Berg et al. (2016: 351)

Schriftsprache hierarchisch konstituieren (vgl. Abbildung 3.4).²⁵

Über der segmentalen Ebene der Buchstaben bilden hier die Grapheme die Skelettpositionen, können also auch als „die kleinsten suprasegmentalen Einheiten der graphematischen Silbenstruktur“ definiert werden (Berg et al. 2016: 350). Darüber hinaus bietet sich aufgrund der Silbenstruktur eine Erklärungsmöglichkeit für die Untrennbarkeit der Buchstabenverbindungen |sch| und |ie|. Auf der segmentalen Ebene der Buchstaben bestehen sie zwar aus drei bzw. zwei Einheiten und besetzen jeweils zwei Skelettpositionen als Grapheme (<s> und <ch> sowie <i> und <e>), aber die Buchstabenverbindungen entsprechen jeweils genau einer Silbenkonstituente: |sch| dem Onset und |ie| dem Nukleus.

So argumentieren Berg et al. (2016: 351) dafür, dass die Untrennbarkeit von |ie| und anderen Schreibdiphthongen darauf gründet, dass jeweils beide Konstituenten von der Silbenkonstituente des Nukleus dominiert sind und diese Struktur die Trennung am Zeilenende unmöglich macht, und definieren in diesem Zusammenhang Schreibdiphthonge übersprachlich als „zwei kompakte[...] Grapheme[...] im Kern derselben

²⁵ Im Original binden Berg et al. (2016: 351) auch noch die subsegmentale Ebene der Buchstabenmerkmale für |e| als unterste Hierarchiestufe mit ein. Ursprünglich entwickelt hat das Modell Primus (2010). Zum graphematischen Fuß, der hier nicht näher thematisiert wird, sei auf Evertz (2016) verwiesen. Anders als bei Berg et al. (2016) wird |sch| hier nicht als eines, sondern gemäß des oben festgelegten Grapheminventars als zwei Grapheme (<s> und <ch>) analysiert.

graphematischen Silbe“.²⁶

Ich möchte nun vorschlagen, diese Argumentation, die Berg et al. (2016) ausschließlich auf die vom Nukleus dominierten Konstituenten anwenden, auch auf die anderen Silbenkonstituenten zu übertragen. So lässt sich nämlich auch die Untrennbarkeit von Buchstabenverbindungen wie |sch| in bestimmten Kontexten auf der Basis dieser Silbenstruktur erklären. Dadurch, dass die beiden Grapheme <s> und <ch> in Beispielen wie *rauschen* durch eine Silbenkonstituente dominiert werden, hier durch die Onset-Konstituente der zweiten Silbe, ist eine Trennung nur zwischen den Silben möglich (*rau-schen*), während eine Trennung von <s> und <ch> bei *Gläs-chen* aufgrund der Zugehörigkeit zu verschiedenen Silben kein Problem ist.²⁷

Exkurs: Silbifizierung der graphematischen Silbe

Allerdings ist die Basis für eine Entscheidung, welche Grapheme in graphematisch mehrsilbigen Wörtern zur ersten bzw. den weiteren Silben gehören noch nicht ausgereift. In der Phonologie ist hier neben der Sonoritätshierarchie das Prinzip der Onsetmaximierung und die morphologische Struktur relevant.

Dass die morphologische Struktur auch für die graphematische Silbe von Belang ist, ist wohl zumindest an Kompositionsgrenzen unstrittig. Auch wenn die Längenhierarchie es zulassen würde, ist also beispielsweise bei *Haustür* nicht von einer Silbifizierung in *Haust-ür*, sondern von *Haus-tür* auszugehen. Wenn man die Worttrennung am Zeilenende als Kriterium für die graphematische Silbifizierung berück-

²⁶ Im Folgenden analysieren Berg et al. (2016: 351-353) die Gruppe der Schreibdiphthonge, die im Deutschen neben <ie> auch <aa>, <ee>, <oo>, <ai>, <au>, <ei> und <eu> umfasst, noch ausführlicher und stellen unter anderem fest, dass Schreibdiphthonge, deren zweite Kernposition ein Graphem mit gebogenem Kopf besetzt (<ie>, <aa>, <ee>, <oo>) phonologisch einem Monophtong entsprechen, Diphthonge mit Graphemen mit geradem Kopf in der zweiten Nukleusposition dagegen mit phonologischen Diphthongen korrespondieren.

²⁷ Die Buchstabenkombinationen |st| und |sp| scheinen nur wortinitial (ia), im Fall von |st| auch wortfinal (ib), bzw. an Kompositions- oder Präfigierungsgrenzen (ic) einer Silbenkonstituente entsprechen zu können, während sie abseits davon (id) zu getrennten Silben gehören:

- (i) a. Stern, Spiel, Streu-ung, Sprung
- b. Nest, Rast
- c. be-ste-hen, Vor-sprung, Foul-spiel, Abend-stern, Nest-flucht, Rast-Platz
- d. ras-ten, Ris-pe, ras-peln, Nes-ter

sichtigt²⁸ – wie beispielsweise Fuhrhop & Peters (2013: 228) es als Option aufzeigen – gilt dies zwar auch für Präfigierungen (*be-enden*), aber nicht unbedingt für die Grenzen zu Derivations- und Flexionssuffixen. Im Einklang mit der Längenhierarchie wäre bei *bergig* sonst *berg-ig* möglich, tatsächlich getrennt wird aber entgegen der morphologischen Struktur *ber-gig*, was aber wiederum der aus der aus der Phonologie bekannten Onsetmaximierung entspricht. Tatsächlich entspricht die graphematische Silbengrenze wohl nur bei Suffixen, die mit einem kompakten Buchstaben beginnen (11a), nicht der morphologischen Grenze, während sie bei Suffixen, die mit nicht-kompakten Buchstaben beginnen (11b), eingehalten wird:

- (11) a. Fla-sche, hö-ren, seh-nig, Freun-din, nei-disch, Le-sung, Bän-ker
 b. mach-bar, Gläs-chen, herr-lich, ernst-haft, Schuld-ner, Wüst-ling

Die Onsetmaximierung scheint also, zumindest auf Basis der Worttrennung, kein uneingeschränkt wirksames Prinzip innerhalb der Graphematik zu sein. Fuhrhop & Peters (2013: 228) sprechen auch nicht von einer graphematischen Onsetmaximierung, sondern konstituieren am Beispiel <Ru-he> lediglich, dass zweite Silben in der Graphematik eher bedeckt seien als in der Phonologie, wofür auch die Beispiele unter (11a) sprechen.²⁹

Mit Blick auf die Worttrennung am Zeilenende finden sich im § 110 des Regelwerks gleich mehrere Beispiele für Worttrennungen, die keinen maximalen Anfangsrand der zweiten Silbe aufweisen, sondern eine Verteilung der Konsonantenbuchstaben zwischen zwei kompakten Kernen auf beide Silben (Rat für deutsche Rechtschreibung 2018):

- (12) a. wid-rig, eif-rig, knusp-rig, Kanz-ler
 b. imp-fen, Karp-fen, dunk-le

Bei den Suffixderivaten (12a) wird nicht nur entgegen der morphologischen, sondern auch wider die phonologische Struktur genau dort getrennt, wo auf Basis der

²⁸ Evertz (2018: 91) behauptet dagegen, dass die Segmente, die bei einer Trennung am Zeilenende entstehen, zwar häufig („in many instances“) aber nicht zwingend graphematischen Silben entsprechen. Leider fehlen konkrete Beispiele für entsprechende Segmente, die seine Aussage stützen.

²⁹ Initiale Silben sind bei Annahme des glottalen Verschlusslautes in der Phonologie immer bedeckt (vgl. Fuhrhop & Peters 2013: Kap. I.5.3).

Längenhierarchie eine neue Silbe beginnen muss, da |r| kompakter als |d|, |f| oder |p| ist und ||| kompakter als das vorausgehende |z|. Dasselbe gilt bei den einfachen Wörtern unter (12b) für *dunkle*, während bei *impfen* und *Karpfen* zwischen zwei Buchstaben getrennt wird, die mit ihren langen Köpfen auf derselben Stufe der Längenskala stehen.

An diesen beiden Beispielen lässt sich sehr gut verdeutlichen, dass das Prinzip der Onsetmaximierung sowohl in der Phonologie als auch in der Graphematik wirkt und Unterschiede in der Silbifizierung in beiden Bereichen dadurch erklärt werden können, dass es in der Phonologie auf der Sonoritätshierarchie fußt, in der Graphematik aber auf der Längenhierarchie. So liegt die phonologische Silbengrenze bei *impfen* und *Karpfen* jeweils vor dem [p], da [pf] in beiden Wörtern das maximal mögliche Onset ist, ohne gegen die Sonoritätshierarchie zu verstoßen, auf der [m] und [r] höher stehen. Die graphematische Silbengrenze liegt dagegen in beiden Fällen erst nach dem |p|, da |p| und |f| auf derselben Stufe innerhalb der Längenhierarchie stehen und das Prinzip des maximalen Onsets in solchen Fällen dazu führt, dass jeder Silbe eine dieser gleichrangigen Einheiten zugeteilt wird. Als Parallelbeispiel dazu aus der Phonologie sei auf *kommen* verwiesen, wo die Onsetmaximierung ebenfalls dafür sorgt, dass die Silbengrenze zwischen den beiden Silbenkonstituenten liegt, die das ambisilbische [m] besetzt.

Das Prinzip der Onsetmaximierung spielt dementsprechend für die graphematische Silbe eine Rolle, wenn auch eine untergeordnete. Grundlegend für die Silbifizierung ist dagegen die nach Längenhierarchie mögliche Struktur, wobei davon auszugehen ist, dass morphologischen Grenzen (Grenzen zu Suffixen mit initialem, kompakten Buchstaben (*-isch*, *-ig*, *-er*, *-ung*, *-en*) aufgrund der Onsetmaximierung ausgenommen) Silbengrenzen entsprechen.

Unter Einbezug des Regelwerks zur Worttrennung am Zeilenende und der in der Phonologie wirksamen Prinzipien möchte ich daher im Folgenden einen praktischen Leitfaden zur graphematischen Silbifizierung skizzieren, der auf den bisherigen Beobachtungen beruht, aber ohne breite empirische Überprüfung zunächst nur ein Vorschlag bleibt.

Da ein kompakter Kern ein essentieller Bestandteil graphematischer Silben ist, bieten kompakte Buchstaben eine erste Orientierung bei der Silbifizierung. Ein einzelner kompakter Buchstabe wie |a| in *krank* ist ein eindeutiger Indikator dafür,

dass nur eine Silbe vorliegt, während beim Vorliegen zweier kompakter Buchstaben, sofern dazwischen mehrere, nicht-kompakte Buchstaben liegen, von einer graphematischen Zweisilbigkeit auszugehen ist, wie im Fall von *kränken*. Vereinzelt Ausnahmen davon sind Wörter mit mehreren kompakten Buchstaben, die wegen eines dazwischenliegenden nicht-kompakten Buchstabens (*Abend*, *Idee*) oder einer Aneinanderreihung von drei kompakten Buchstaben (*Kleie*) nicht nur einem gemeinsamen Kern entsprechen können. Eine Silbengrenze würde hier aber dazu führen, dass ein einzelner kompakter Buchstabe abgetrennt werden müsste.³⁰ Nach § 107 E1 ist die Abtrennung einzelner Vokalbuchstaben am Wortanfang oder -ende allerdings unzulässig (**A-bend*, **I-dee*, **Klei-e*).³¹ Legt man die Worttrennung am Zeilenende also als Kriterium für die Silbifizierung zugrunde, muss man *Abend*, *Idee* und *Kleie* als graphematische Einsilber analysieren. Da eine Einsilbigkeit im Widerspruch zur Annahme zweier Silbenkerne steht, muss man dann allerdings die kompakten Buchstaben an Wortanfang oder -ende als extrasilbisch analysieren, ähnlich wie ein |s| nach Buchstaben mit langen Köpfen in der Koda (z. B. in *montags*, siehe oben).³² Diese Einsilber sind aber dann als markiert zu klassifizieren, da sie nicht dem Allgemeinen Graphematischen Silbenbaugesetz entsprechen.

Ein Schreibdiphthong aus zwei kompakten Buchstaben kann dagegen regelkonform abgetrennt werden. So findet sich beispielsweise *Ei-er* unter den Beispielen nach § 107. Auf Basis der Worttrennung ist also davon auszugehen, dass eine gra-

³⁰ Das gleiche Problem ergibt sich bei Fremdwörtern wie *Dia* oder *bio*. Zwar können theoretisch zwei kompakte Buchstaben als Schreibdiphthong gemeinsam den Nukleus einer graphematischen Silbe besetzen, die Kombinations- und Besetzungsmöglichkeiten der beiden Positionen sind aber im Deutschen begrenzt (siehe unten).

³¹ § 107 zur Worttrennung am Zeilenende (Rat für deutsche Rechtschreibung 2018) unterscheidet dabei wortintern zwischen Kompositionsgrenzen und anderen morphologischen Grenzen. Während E1 die Abtrennung von Einzelbuchstaben auch innerhalb von Komposita am Beispiel *Ju-li-abend* (also nicht: *Ju-li-a-bend* oder *Julia-bend*) ausschließt, finden sich unmittelbar zuvor Beispiele für Fremdwörter mit „abtrennbaren“ Einzelbuchstaben: *eu-ro-pä-i-sche*, *Ru-i-ne*, *na-ti-o-nal*. Bei einer Worttrennung am Zeilenende wird es auch in diesen Fällen dennoch nicht dazu kommen, dass ein Einzelbuchstabe stehen bleibt, da ja immer nur an einer Stelle getrennt wird, im Fall von *europäische* also entweder *europä-ische* oder *europäi-sche*. Hier wäre es interessant, herauszufinden, ob in der Schreibpraxis tatsächlich beide Varianten belegt sind, und falls ja, in welchem Verhältnis.

³² Für eine Analyse kompakter Buchstaben abseits des Silbenkerns als extrasilbisch sprechen auch die in § 107 vorgesehenen Trennmöglichkeiten für *eu-ro-pä-i-sche*, *Ru-i-ne* und *na-ti-o-nal*.

phematische Silbe auch ausschließlich aus dem Silbenkern bestehen kann, unter der Voraussetzung, dass beide Nukleuspositionen besetzt sind. Ein einfach besetzter Silbenkern kann, wie oben gezeigt, nicht alleine eine graphematische Silbe bilden, wohl aber in Kombination mit einem nicht-kompakten Buchstaben in Onset (*ru-hen*) oder Koda (*er-ben*). Daraus lässt sich schließen, dass eine graphematische Silbe aus mindestens zwei Buchstaben besteht, von denen zumindest einer als kompakter Buchstabe eine Kernposition besetzen muss.

Bei zwei unmittelbar aufeinanderfolgenden kompakten Buchstaben ist nicht sofort ersichtlich, ob es sich um eine oder zwei graphematische Silben handelt, da der Kern graphematischer Silben wie bei *laut* regelhaft mit zwei kompakten Buchstaben besetzt werden kann. Die Kombinationsmöglichkeiten von kompakten Buchstaben im Silbenkern sind aber beschränkt. Im Deutschen sind neben *|au|/|äu|* auch *|ai|, |ei|, |eu|* sowie *|aa|, |ee|, |oo|* und *|ie|* als Schreibdiphthonge innerhalb eines Silbenkerns möglich. Demzufolge ist bei allen anderen Kombinationen, die sich zumeist in Fremdwörtern finden, von einer dazwischenliegenden Silbengrenze auszugehen wie in *Ritu-al, Du-ell* oder *Spi-on*, während bei den Kombinationen die Schreibdiphthongen entsprechen nur aufgrund der morphologischen Struktur ein Zugehörigkeit zu verschiedenen Silben möglich ist:

(13) be-enden, Famili-en, Muse-um, re-investieren

Im Falle von drei (oder gar vier) aufeinanderfolgenden kompakten Buchstaben ist wiederum sicher von einer Silbengrenze nach dem ersten oder zweiten kompakten Buchstaben auszugehen (im Falle von vier kompakten Buchstaben nach dem zweiten). Auch hier bestimmen das Inventar an möglichen Schreibdiphthongen und morphologische Grenzen im Einzelfall, wo die Silbengrenze liegt:

(14) a. brei-ig, See-igel, Hai-angriff, be-eindrucken
b. Schnee-eule, Zoo-auto

Ist eine graphematische Mehrsilbigkeit wie bei *kränken* aufgrund zweier Silbenkerne eindeutig zu erkennen, stellt sich dennoch die Frage, wo zwischen diesen beiden Kernen die Silbengrenze zu verorten ist. Da hier keine Kompositions- oder Präfigierungsgrenze vorliegt, wird die Silbengrenze durch das Zusammenspiel von Längen-

hierarchie und dem Prinzip der Onsetmaximierung bestimmt. Beim vorliegenden Beispiel *kränken* steigt die Länge der Buchstabenköpfe vom Kern |ä| über |n| und |k| kontinuierlich an, bis sie durch den zweiten Kern |e| wieder auf ein kompaktes Niveau fällt. Dementsprechend wären die Trennungen *kränk-en* und *krän-ken* mit der Längenhierarchie kompatibel, da sich in beiden Fällen Silben ergeben, in denen die Kompaktheit zum Kern hin zunimmt und vom Kern aus abnimmt. Dass tatsächlich *krän-ken* präferiert wird, ist auf den Einfluss der Onsetmaximierung zurückzuführen, die hier dafür sorgt, dass |k| das Onset der zweiten Silbe bildet. |n| verbleibt in der Koda der zweiten Silben, da die Längenhierarchie nicht durch die Onsetmaximierung ausgehebelt werden kann und eine Silbe *nken* der Hierarchie deutlich sichtbar zuwider laufen würde. Dementsprechend schlage ich folgende Prinzipienhierarchie für eine Silbifizierung in graphematische Silben vor:

1. Jede graphematische Silbe umfasst einen Kern.
2. Wortbildungsmorphologische Grenzen (ausgenommen: Grenzen zu Suffixen mit initialem, kompakten Buchstaben) sind Silbengrenzen.
3. Innerhalb graphematischer Silben gilt die Längenhierarchie.
4. Auch graphematische Silben folgen dem Prinzip der Onsetmaximierung.

Diese Rangfolge ist im Sinne einer optimalitätstheoretischen Hierarchisierung zu verstehen. Das bedeutet, dass Buchstabenkombinationen die untergeordnete Beschränkungen wie Onsetmaximierung und Längenhierarchie verletzen, dennoch optimal im Sinne graphematischer Silben sein können, so lange die übergeordneten Prinzipien eingehalten werden.

Diese Prinzipienhierarchie ist damit der Versuch, die Worttrennung am Zeilenende rein graphematisch zu erklären unter der Annahme, dass es sich bei den Segmenten, die sich bei einer Trennung am Zeilenende ergeben, regelmäßig um graphematische Silben handelt. Dieses Konzept steht damit nicht in direkter Konkurrenz zur orthographischen Silbentrennung, die nach Primus (2010: 19) ebenfalls durch hierarchisch geordnete Regeln (morphologisch, graphematisch, phonologisch) bestimmt ist, sondern nimmt vielmehr einen andere Perspektive bezüglich desselben Phänomens ein.

Eine detaillierte Ausformulierung und kritische Überprüfung dieses Vorschlags zur graphematischen Silbifizierung ist wünschenswert und könnte zur Weiterent-

wicklung des Konzeptes der graphematischen Silbe beitragen, kann aber im Rahmen dieser Arbeit nicht geleistet werden.

3.1.4 Graphematische Wörter

Wie bereits erwähnt, ist der Begriff des graphematischen Wortes für diese Arbeit von Bedeutung, auch wenn er kein hartes Kriterium für die Belegauswahl ist (vgl. 2.1). Für die Untersuchung sind nämlich gerade auch die Belege von Interesse, die zwei oder sogar mehr graphematische Wörter umfassen, morphosyntaktisch aber dennoch nur ein einziges Wort sind:

- (15) acht Uhr Vorstellung (SMS), Toyota Vertragshändlern (W), Sekundarstufe II-Abschluss (WPD₁₁), das ADAC Ruhr-Pokal-Rennen, (RHZ₀₉) Joe Cocker-Imitationen (M₀₉)

Formal sind graphematische Wörter nur über Leerzeichen definiert: „*Das graphematische Wort steht zwischen zwei Leerzeichen und enthält intern keine Leerzeichen*“ (Fuhrhop 2008: 193). Unter diesen Begriff fallen damit nicht nur einzelne Bestandteile komplexer morphosyntaktischer Wörter aufgrund ihrer wortinternen Getrennschreibung wie in (15), sondern auch Beispiele wie in (16):

- (16) a. USA, Kfz, usw.
 b. 100, 1,7, 80 000 000, die 80er
 c. geht's, Groß- und Kleinschreibung, Auf-besseres-Wetter-Hoffen, Früh/Spätaufsteher

Per Definition, die keine Aussagen zur internen Struktur graphematischer Wörter macht, ist also *ooo* ebenso ein graphematisches Wort wie *Kfz*, *geht's* und *Haus*. Da das mit der Intuition von Schreibern, die *459* vielleicht eher als Zahl, *Groß-* in *Groß- und Kleinschreibung* möglicherweise als Wortbestandteil identifizieren würden, so nicht übereinstimmt, schlage ich hier basierend auf Fuhrhop (2008) eine Differenzierung innerhalb dieser heterogenen Gruppe der graphematischen Wörter vor.

Kernbereich oder markiert – eine überarbeitete Klassifikation graphematischer Wörter

Graphematische Wörter bleiben, wie oben besprochen, grundlegend als ununterbrochene Zeichenketten zwischen zwei Leerzeichen definiert. Die Art der Zeichen (Buchstaben, Ziffern, Sonderzeichen, Interpunktionszeichen) ist dabei nicht relevant. Das graphematische Wort ist nur über seine Grenzen definiert.

- (17) Haus, 127, Kfz-Mechanikerin, €-Rettungsschirm, (Un)Schuld, 2fach, ABC-Alarm, Goethe-Haus, läuft's, Schul-/Arbeitswoche, LehrerInnen, voll, A

Der Kernbereich graphematischer Wörter ist zusätzlich übereinstimmend mit Fuhrhop (2008) darüber definiert, dass sie ausschließlich aus Buchstabengraphemen bestehen (und zwar mindestens aus zwei Graphemen) und diese graphematische Silben nach dem Allgemeinen Graphematischen Silbengesetz bilden. Die inhaltlich unspezifische Definition wird also für den Kernbereich um inhärente Strukturmerkmale erweitert.

- (18) Kunst, Sommerpause, flott, gehen, ernst, fröhlich, Buch, Flur

Markierte graphematische Wörter sind – im Gegensatz zur Definition von Fuhrhop (2008) über die Wortzeichen – *ex negativo* über den Kernbereich definiert. Das heißt, ich teile die graphematischen Wörter anhand der Definition des Kernbereichs binär in die Kategorien Kernbereich und markierte graphematische Wörter ein. Zu den markierten graphematischen Wörtern gehören damit neben graphematischen Wörtern mit Interpunktionszeichen (19a) auch solche mit Ziffern oder Sonderzeichen (19b) sowie graphematische Wörter aus Buchstabengraphemen ohne graphematische Silbenstruktur (19c).

- (19) a. Goethe-Haus, (Un)Schuld, läuft's, Schul-/Arbeitswoche, z. B.
 b. 7, 2fach, M@iladresse
 c. Kfz, NATO, DGfS, BahnCard

Damit gehören zu den markierten graphematischen Wörtern alle graphematischen Wörter mit einer gestörten Silbenstruktur. Diese Störungen sind darauf zurückzuführen, dass markierte graphematische Wörter allesamt Segmente enthalten, auf die

sich das Allgemeine Graphematische Silbenbaugesetz nicht anwenden lässt, weil es sich entweder nicht um Buchstabengrapheme handelt oder um interne Majuskeln, auf die die Längenhierarchie nicht angewendet werden kann.

Einzelne Bestandteile morphologisch komplexer, markierter graphematischer Wörter mit Wortzeichen, Nicht-Buchstabengraphemen oder internen (und finalen) Majuskeln können dabei eine graphematische Silbenstruktur aufweisen und so graphematischen Wörtern des Kernbereichs entsprechen (z. B. *Goethe* und *Haus* in *Goethe-Haus*),³³ sind aber in der Regel nicht bzw. nicht ausschließlich durch Leerzeichen, sondern auch durch Wortzeichen, Majuskeln oder Nicht-Buchstabengrapheme begrenzt, so dass sie nicht unter die Definition graphematischer Wörter fallen.

Auch wenn diese Begriffsbestimmungen von graphematischen Wörtern des Kernbereichs und markierten graphematischen Wörtern als Subkategorien graphematischer Wörter maßgeblich für diese Arbeit sind, soll im Folgenden noch der Ansatz von Fuhrhop (2008), auf dem diese Lösung beruht, kritisch gewürdigt werden.

Kernbereich und markierte graphematische Wörter nach Fuhrhop (2008)

Angesichts der oben gezeigten Heterogenität versucht Fuhrhop (2008: 194) erstmalig einen Kernbereich mit typischen graphematischen Wörtern auszumachen, indem sie annimmt, dass typische graphematische Wörter aus einer oder mehreren graphematischen Silben bestehen, ununterbrochene Graphemketten sind und höchstens initial einen Großbuchstaben aufweisen.

Anhand strukturinterner Merkmale präzisiert sie also den Begriff des *Graphematischen Wortes* für einen Kernbereich und führt markierte graphematische Wörter als Teilgruppe der graphematischen Wörter ein. Fuhrhops Ansatz und Ausführungen sind zwar richtungsweisend, sie verpasst aber die Gelegenheit, graphematische Wörter des Kernbereichs und markierte graphematische Wörter (jeweils als Teilgruppe graphematischer Wörter) anhand der diskutierten Kriterien abschließend sauber voneinander zu trennen.

Ihre Zusammenfassung beginnt zunächst mit einer unstrittigen Abgrenzung der drei Kategorien, die graphematische Wörter des Kernbereichs und markierte graphematische Wörter als Teilmenge der graphematischen Wörter definiert:

³³ Dass das nicht für alle Bestandteile markierter graphematischer Wörter gilt, zeigen Beispiele wie *läuft's*, bei dem zwar der Bestandteil *läuft* einem graphematischen Wort des Kernbereichs entspricht, aber dies mangels graphematischer Silbenstruktur nicht für *s* gilt.

Das graphematische Wort steht zwischen zwei Leerzeichen. Graphematische Wörter des Kernbereichs bestehen aus ununterbrochenen Graphemketten und sie enthalten intern keine Majuskeln. Graphematische Wörter mit Wortzeichen sind markiert, aber nichtsdestotrotz sind sie graphematische Wörter: insbesondere solche mit Bindestrich, Apostroph und Abkürzungspunkt.

(Fuhrhop 2008: 224-225)

Dementsprechend muss jede Zeichenkette, die zwischen zwei Leerzeichen steht, als eigenständiges graphematisches Wort gezählt werden. Bei *acht Uhr Vorstellung* (SMS) handelt es sich also um drei graphematische Wörter (die einem morphosyntaktischen entsprechen) und auch *80 000 000* ist zwar eine Zahl, umfasst aber gemäß der Definition drei graphematische Wörter.

Obwohl die Definition hier ganz klar ist, wendet Fuhrhop (2008: 222) sie nicht ganz konsequent an. In ihrer Tabelle *Wortbestimmungen im Vergleich* klassifiziert sie *Carl von Oss. Universität* graphematisch als „4 Wörter“, zögert aber *127 538* trotz des Leerzeichens als zwei graphematische Wörter zu klassifizieren (womöglich weil sie für das Leerzeichen zwischen Ziffern eine abweichende Funktionalität annimmt) und kommentiert stattdessen: „?1 Wort, mark.“ mit der Anmerkung „Leerzeichen!“.³⁴

Hier trotz internem Leerzeichen nur ein graphematisches Wort anzunehmen, führt die Definition ad absurdum, so dass es sinnvoller ist, zu akzeptieren, dass eine Zahl ebenso wenig genau einem graphematischen Wort entsprechen muss wie ein morphosyntaktisches Wort.

Graphematische Wörter des Kernbereichs müssen der Definition entsprechend darüber hinaus aus „ununterbrochenen Graphemketten“ – Graphem ist hier eindeutig im Sinne von alphabetischem Graphem zu verstehen – bestehen und dürfen keine internen Majuskeln aufweisen. Dass damit auch finale Majuskeln ausgeschlossen werden sollen, wird durch die eindeutigere Formulierung in der vorausgehenden These klar: „Das typische graphematische Wort enthält höchstens am Wortanfang einen Großbuchstaben.“ (Fuhrhop 2008: 194).

Diese Einschränkungen führen dazu, dass aus dem Kernbereich alle Zeichenket-

³⁴ Auch das Beispiel *weich kochen* annotiert Fuhrhop (2008: 220) zwar als zwei graphematische Wörter, setzt aber ein Fragezeichen dazu, kategorisiert *leerkaufen* als „1 [graphematisches] Wort?“ und bezeichnet auch *brustschwimmen* als Zweifelsfall. Von Zweifelsfällen kann hier aber nur morphosyntaktisch die Rede sein, graphematisch handelt es sich eindeutig um ein Wort, sofern die Wortform zusammengeschrieben wird, oder um zwei Wörter, wenn die Wortform getrennt geschrieben wird.

ten ausgeschlossen werden, die Syngrapheme, Ziffern oder Sonderzeichen enthalten. Beispiele für diesen Kernbereich nach Fuhrhop (2008) sind damit:

- (20) Haus, ein, Langeweile, Blumenkohlsuppe, anfangen, aufgrund, krankschreiben, im

Auch *Kfz* listet sie entgegen der Definition als graphematisches Wort. Die fehlende Silbenstruktur (kein kompakter Buchstabe) wird nur in der Bemerkungsspalte erwähnt (vgl. Fuhrhop 2008: 221).

Markierte graphematische Wörter dagegen sind darüber definiert, dass sie Interpunktionszeichen enthalten, wie die Beispiele nach Fuhrhop (2008):

- (21) gibt's, 's, Dtschld., Abk., Software-Entwicklung, Schönes-Wochenende-Ticket, Über-das-Wetter-Meckern, 1,7, bzw.³⁵

Uneindeutig bleibt Fuhrhop (2008) bezüglich interner Majuskeln. Per definitionem schließt sie diese zwar aus dem Kernbereich graphematischer Wörter aus, scheut aber auch davor zurück, sie den markierten graphematischen Wörtern zuzurechnen. So klassifiziert sie *BahnCard*, *NATO* und *DGfS* jeweils als graphematisches Wort ohne den Zusatz *markiert* und verweist nur in der Spalte „gr. wohlgeformt?“ auf die Binnengroßschreibung (Fuhrhop 2008: 220f.).

Damit stellt sich unmittelbar die Frage, ob graphematische Wörter neben dem Kernbereich und den markierten graphematischen Wörtern noch eine weitere Gruppe umfassen, bzw. ob es einen Graubereich innerhalb der graphematischen Wörter gibt, der weder von der Bestimmung des Kernbereichs noch von der Definition markierter graphematischer Wörter erfasst wird.

Nach Fuhrhop (2008) muss man wohl eine dritte Teilgruppe graphematischer Wörter annehmen, die zwar nicht näher thematisiert wird, aber als Sammelbecken für graphematische Wörter dient, die weder zum Kernbereich noch zur markierten Gruppe gehören. Dazu müssten dann neben den gerade schon thematisierten graphematischen Wörtern mit Binnenmajuskeln (Ausschlusskriterium für Kernbereich) und ohne Wortzeichen (Kriterium für die markierte Gruppe nicht erfüllt) wie

³⁵ Für die Beispiele *Ludwig-Maximilians-Univ.*, *usw.* und *komm'*, denen in der Tabelle *Wortbestimmungen im Vergleich* trotz Wortzeichen, der Verweis auf die Markiertheit fehlt (Fuhrhop 2008: 221f.), gehe ich von einem Versehen aus und zähle sie ebenfalls zu den markierten graphematischen Wörtern.

die Beispiele in (22a) auch graphematische Wörter mit Ziffern (Ausschlusskriterium für Kernbereich) und ohne Wortzeichen (Kriterium für die markierte Gruppe nicht erfüllt) wie in (22b) zählen.

- (22) a. BahnCard, NATO, DGfS
b. 5fach, 80er, 859

Wie die Beispiele zeigen, kann aber auch in diesen Fällen eine graphische Markiertheit nicht geleugnet werden. Sowohl die Majuskeln außerhalb der initialen Position des graphematischen Wortes als auch die Ziffern, stören das prototypische Bild eines graphematischen Wortes.

Deshalb wird hier die oben besprochene modifizierte Version der Einteilung in markierte graphematische Wörter einerseits und einen Kernbereich andererseits anhand des Kriteriums der Silbenstruktur bevorzugt, wodurch alle graphematischen Wörter ohne Graubereich in beiden Unterkategorien erfasst werden können.

3.2 Orthographie

Auch wenn es sich hier um eine graphematische Untersuchung handelt, kann nicht ignoriert werden, dass es für das Deutsche eine detailliert ausgearbeitete Orthographie gibt, die durch ihre Institutionalisierung in Rechtschreibduden, Rat für deutsche Rechtschreibung und Deutschunterricht den Schreibern sehr präsent ist.

Daher ist es natürlich wichtig, sich auch mit der Norm auseinanderzusetzen, wenn man das eigentliche System dahinter untersuchen möchte. Da eine fixierte Norm Träger ist als das wesentlich dynamischere sprachliche (Teil-)System, das sie abzubilden versucht, ist dabei nicht mit einer 1:1-Entsprechung zu rechnen. Spannend wird es besonders dort, wo der Sprachgebrauch der Norm nicht folgt, und sich die Frage stellt, ob es sich in solchen Fällen um einzelne Fehler oder systematische Abweichungen handelt.

Also muss man auch für eine graphematische Untersuchung den Einflussbereich der orthographischen Norm genau kennen. Bezüglich der Schreibvariation innerhalb komplexer Wörter sind dafür im Wesentlichen die Paragraphen zur Getrennt- und Zusammenschreibung (§§ 36-38), zur Schreibung mit Bindestrich (§§ 40-52) und zur

Groß- und Kleinschreibung (§ 55) und die jeweiligen Vorbemerkungen zu diesen drei Bereichen relevant (vgl. Rat für deutsche Rechtschreibung 2018).³⁶

Die folgenden Ausführungen basieren auf der aktuellsten Version des Regelwerks zur deutschen Rechtschreibung, das vom Rat für deutsche Rechtschreibung 2018 herausgegeben wurde. Auch wenn die untersuchten Korpora unter dem Einfluss der vorausgegangenen Version entstanden sind, bietet es sich an, die aktuelle als Grundlage zu nutzen, da sich für die hier relevanten Aspekte keine wesentlichen Veränderungen ergeben haben. Wo sich im Detail Unterschiede ergeben, werden sie in der folgenden Diskussion angesprochen.

3.2.1 Getrennt- und Zusammenschreibung

Gleich die erste Vorbemerkung zur Getrennt- und Zusammenschreibung macht deutlich, dass wortintern eine Zusammenschreibung zu erwarten ist, während Leerzeichen Wörter innerhalb von Wortgruppen abgrenzen.³⁷

(1) Die Getrennt- und Zusammenschreibung betrifft Einheiten, die im Text unmittelbar benachbart und aufeinander bezogen sind. Handelt es sich um die Bestandteile von Wortgruppen, so schreibt man sie getrennt. Handelt es sich um die Bestandteile von Zusammensetzungen, so schreibt man sie zusammen.

(Rat für deutsche Rechtschreibung 2018: 33)

Dementsprechend fordern auch die spezifischen §§ 36 für adjektivische und 37 für substantivische Zusammensetzungen, die Zusammenschreibung der jeweiligen Bestandteile. Die Getrenntschreibung ist laut § 36 (2) nur in den Fällen eine Option, in den ein Ausdruck entweder als syntaktische Fügung oder Zusammensetzung interpretiert werden kann. Die Beispiele unter (23) sind also nicht verschiedene graphische Varianten eines sprachlichen Zeichens, sondern verdeutlichen auf graphischer Ebene Unterschiede in der Morphosyntax (und Phonologie) zweier verschiedener sprachlicher Zeichen.³⁸

(23) a. die Rat suchenden/ratsuchenden Bürger

³⁶ Aufgrund der vorausgegangenen Einschränkung der Untersuchung auf Substantive (und Adjektive) werden Paragraphen, die sich auf andere Wortarten beziehen, hier nicht berücksichtigt.

³⁷ Ich verstehe dabei Zusammensetzungen synonym zu komplexen Wörtern, denen Wortbildungsprozesse zugrunde liegen, und Wortgruppen als syntaktische Verknüpfung von Wörtern.

³⁸ Zum Begriff des *sprachlichen Zeichens* vgl. Kapitel 6.

- b. schwer verständlich/schwerverständlich
- c. eine nicht öffentliche/nichtöffentliche Sitzung
(Rat für deutsche Rechtschreibung 2018: 39)

Etwas irritierend ist dabei die Konkurrenz zwischen den Teilen (1.5) und (2.2) des § 36:

(1) Es wird zusammengeschrieben, wenn [...]

(1.5) der erste Bestandteil bedeutungsverstärkend oder bedeutungsabschwächend ist. Mit Bestandteilen dieser Art werden zum Teil lange Reihen gebildet, zum Beispiel: *bitter-* (*bitterböse, bitterernst, bitterkalt*), *brand-*, *dunkel-*, *erz-*, *extra-*, *früh-*, *gemein-*, *grund-*, *hyper-*, *lau-*, *minder-*, *stock-*, *super-*, *tod-*, *ultra-*, *ur-*, *voll-* Zu adjektivischen Bestandteilen siehe § 36 (2.2).

(2) Zusammen- wie auch getrennt geschrieben werden kann, wenn der entsprechende Ausdruck sowohl als Zusammensetzung als auch als syntaktische Fügung angesehen werden kann. [...]

(2.2) Verbindungen mit einem einfachen unflektierten Adjektiv als graduierender Bestimmung, zum Beispiel: *allgemein gültig/allgemeingültig, eng verwandt/engverwandt, schwer verständlich/schwerverständlich, schwer krank/schwerkrank*

Für adjektivische Bestandteile wird unter (1.5) explizit auf (2.2) verwiesen und damit auf die Möglichkeit der Getrennschreibung, sofern der Ausdruck als syntaktische Fügung interpretiert wird. Allerdings finden sich auch unter den Beispielen für „erste Bestandteile“ in (1.5) mit *extra-*, *super-* und *voll-* einige Formen, die als unflektierte Adjektive mit graduierender Wirkung interpretiert werden können.

Zudem wird in der Wortliste mit *schwerreich* auch eine Zusammensetzung mit dem Verweis auf § 36 (1.5) gelistet, die dem Beispiel *schwer krank/schwerkrank* zu § 36 (2.2) entgegensteht. Die beiden Regelteile scheinen also das System nicht optimal zu erfassen, da die Unterscheidung in erste Bestandteile, die bedeutungsverstärkend oder bedeutungsabschwächend sind, und unflektierte Adjektive als graduierende Bestimmung nicht trennscharf ist, und unabhängig davon die Art der Verbindung (syntaktisch oder morphologisch) das entscheidende Kriterium bleibt.

Ausnahmen von der Zusammenschreibung zusammengesetzter Adjektive und Substantive werden in den Paragraphen §§ 40-51 zur Bindestrichschreibung festgehalten. Eine Getrennschreibung wie bei Wortgruppen ist nur bei einigen aus dem Englischen stammenden Zusammensetzungen aus Adjektiv und Substantiv normgerecht, wenn der Hauptakzent nicht auf dem ersten Bestandteil liegt (§ 37 E4).

3.2.2 Schreibung mit Bindestrich

Die Vorbemerkungen und §§ 40-52 zur Schreibung mit Bindestrich beziehen sich explizit nur auf den Bereich der Zusammensetzungen und Ableitungen, schließen also implizit den Bindestrich zwischen Bestandteilen von Wortgruppen aus. Für Fremdwörter folge die Schreibung mit Bindestrich den Regeln für das Deutsche, während sich die Schreibung bei Eigennamen im Zweifelsfall nach der amtlichen Regelung richtet. In den Regeln des Rats für deutsche Rechtschreibung wird daher unterschieden zwischen Zusammensetzungen und Ableitungen mit oder ohne Eigennamenanteil und außerdem zwischen obligatorischen und fakultativen Bindestrichkontexten.

Dem Bindestrich wird dabei durchweg eine strukturanzeigende Funktion im Sinne des Lesers zugeschrieben:

(1) Der Bindestrich bietet dem Schreibenden die Möglichkeit, anstelle der sonst bei Zusammensetzungen und Ableitungen üblichen Zusammenschreibung die einzelnen Bestandteile als solche zu kennzeichnen, sie gegeneinander abzusetzen und sie dadurch für den Lesenden hervorzuheben.

(Rat für deutsche Rechtschreibung 2018: 45)

Ein Bindestrich ist obligatorisch bei Zusammensetzungen mit Abkürzungen, Einzelbuchstaben oder Ziffern (*A-Dur*, *km-Bereich*, *3-Tonner* vgl. § 40), bei einer Ableitung von Einzelbuchstaben (*die n-te Potenz* vgl. § 41) und nach suffigierten Ziffern als Erstbestandteil (*ein 100stel-Milimeter*, *die 80er-Jahre* vgl. § 42), in substantivisch gebrauchten Aneinanderreihungen (*das Entweder-oder*, *das In-den-Tag-Hineinträumen* vgl. § 43) und zwischen Bestandteilen mehrteiliger Zusammensetzungen, die eine Wortgruppe oder eine Zusammensetzung mit Bindestrich umfassen (*A-Dur-Tonleiter*, *Berg-und-Tal-Bahn*, *Arzt-Patient-Verhältnis*, *September-Oktober-Heft* vgl. § 44), sowie zwischen gleichrangigen, nebengeordneten Adjektiven, die eine „unübersichtliche“ Zusammensetzung bilden.

Die Formulierung von § 44 ist in zweifacher Hinsicht irritierend. Zum einen werden hier in einer Regel zwei Fälle zusammengefasst, die bis auf den normativ obligatorischen Bindestrich nichts gemeinsam haben, zum anderen sind Zusammensetzungen aus gleichrangigen, nebengeordneten Adjektiven nach § 44 wohl automatisch unübersichtlich. Zumindest machen die Beispiele deutlich, dass schon Zusammensetzungen aus zweien dieser Adjektive als unübersichtlich gelten:

- (24) der wissenschaftlich-technische Fortschritt, ein lateinisch-deutsches Wörterbuch, deutsch-österreichische Angelegenheiten, manisch-depressives Verhalten, physikalisch-chemisch-biologische Prozesse

Beim Beispiel *September-Oktober-Heft* in § 44 findet sich auch ein Verweis auf eine alternative Schreibweise mit einem Schrägstrich (*September/Oktober-Heft*) nach § 106 der besagt, dass der Schrägstrich als Kennzeichnung für zusammengehörige Wörter, Zahlen oder Ähnliches fungiert im Sinne der Angabe „*mehrerer (alternativer) Möglichkeiten im Sinne einer Verbindung mit und, oder, bzw., bis oder dergleichen*“. Der Schrägstrich kann also im Kontext komplexer Zusammensetzungen, die koordinierte Bestandteile enthalten, zwischen diesen Bestandteilen äquivalent zu einem Bindestrich verwendet werden.

Fakultativ ist ein Bindestrich nach § 45 zur Hervorhebung einzelner Bestandteile (*dass-Satz, die Hoch-Zeit, be-greifen*) in unübersichtlichen Zusammensetzungen (*Arbeiter-Unfallversicherungsgesetz*)³⁹, zur Vermeidung von Missverständnissen (*Musiker-Leben, Musik-Erleben, re-integrieren*) oder beim Zusammentreffen von drei gleichen Buchstaben (*Kaffee-Ersatz, Bett-Tuch*). Außerdem wird in zwei Ergänzungen (E1 und E2) zu § 45 eine fakultative Schreibung mit Bindestrich für Substantiv-Substantiv-Verbindungen und substantivierte Verb-Adverb-Verbindungen aus dem Englischen empfohlen, ohne die Zusammenschreibung auszuschließen (*Standby, Stand-by; Sciencefiction, Science-Fiction*). In seinem Bericht für die Periode 2011-2016 verweist der Rat für deutsche Rechtschreibung im Zusammenhang mit § 45 (1) darauf, dass bei den Einträgen *Exkaiser* und *Coautor/Koautor* im aktualisierten Wörterverzeichnis nun auch die Möglichkeit einer Schreibung mit Bindestrich (*Ex-Kaiser, Co-Autor/Ko-Autor*) angeführt wird (vgl. Güthert 2016: 10-11).

Für Zusammensetzungen und Ableitungen mit Eigennamen (§§ 46-52) als Bestandteil ist der Bindestrich bis auf wenige als Gattungsbezeichnungen lexikalisierte Ausnahmen wie *Heulsuse* nach § 47 oder geografische Eigennamen wie *Neubrandenburg* oder *Sankt Georgen* nach § 46 E2 immer dann obligatorisch, wenn es sich beim zweiten Bestandteil um einen Eigennamen handelt:

- (25) Frau Müller-Weber, Annaberg-Buchholz, baden-württembergisch, die kant-

³⁹ Die Beispiele zu diesem Teilaspekt von § 45 lassen vermuten, dass Unübersichtlichkeit hier drei oder mehr Kompositionsglieder erfordert.

laplacesche Theorie

Für Zusammensetzungen mit einem Eigennamen als Erstbestandteil ist ein Bindestrich nur „zwischen allen Bestandteilen mehrteiliger Zusammensetzungen, deren erste Bestandteile aus Eigennamen bestehen“ (§ 50), obligatorisch. Die folgenden Beispiele, wie *Albrecht-Dürer-Allee*, *La-Plata-Mündung* und *Fidel-Castro-freundlich*, sprechen dafür, die Regel Formulierung so zu interpretieren, dass sie sich nur auf Zusammensetzungen mit Eigennamen, die selbst aus mehreren eigentlich durch Leerzeichen getrennten Teilen bestehen, bezieht.

Möglich, aber nicht zwingend ist ein Bindestrich außerdem zur Hervorhebung des Eigennamens oder bei komplexen zweiten Bestandteilen (*Goethe-Ausgabe*, *Elbe-Wasserstandsmelder* vgl. § 51).

Bemerkenswert ist § 52, da diese Regel Formulierung sich nicht, wie aufgrund der Vorbemerkungen zu erwarten, auf Zusammensetzungen und Ableitungen, also die Wortbildung als Domäne bezieht, sondern den Bindestrich zwischen zwei eigentlich syntaktisch verbundenen Einheiten lizenziert:

§ 52

Wird ein geografischer Eigenname von einem nachgestellten Substantiv näher bestimmt, so kann man einen Bindestrich setzen.

Beispiele:

Frankfurt Hauptbahnhof/Frankfurt-Hauptbahnhof, München Ost/München-Ost

Diese Variation spricht dafür, dass hier benachbarte Einheiten ähnlich wie bei den Beispielen unter (23) mal als syntaktisch, mal als morphologisch verbunden wahrgenommen und in Abhängigkeit davon verschriftlicht werden. Abweichend davon werden die unterschiedlichen Interpretationsmöglichkeiten, aber nicht durch den Kontrast von Getrennt- und Zusammenschreibung, sondern durch eine Dichotomie zwischen Bindestrich und Getrennschreibung abgebildet.

3.2.3 Groß- und Kleinschreibung

Aus dem Regelwerk zur Groß- und Kleinschreibung, das die §§ 55-58 umfasst, werden hier nur die Aspekte diskutiert, die relevant sind für die wortinterne Schreibvariation.

Grundlegend für die Groß- und Kleinschreibung ist § 55 („Substantive schreibt man

groß.“), der durch § 56 eingeschränkt wird hinsichtlich von Wörtern, die formgleich zu Substantiven sind, aber keine substantivischen Merkmale aufweisen, wie *angst* in *Mir wird angst*. Als substantivische Merkmale führt § 55 neben einem festen Genus und der Flektierbarkeit nach Numerus und Kasus auch die Fähigkeit an, Gegenstände, Lebewesen und abstrakte Begriffe zu bezeichnen.

Wortintern ist vor allem § 55 (2) relevant:

Großschreibung gilt auch [...] für Substantive – auch Initialwörter (§ 102(2)) und Einzelbuchstaben, sofern sie nicht als Kleinbuchstaben zitiert sind – als Teile von Zusammensetzungen mit Bindestrich, zum Beispiel:

die Ad-hoc-Entscheidung, der A-capella-Chor [...], das In-den-Tag-hinein-Leben [...], der Trimm-dich-Pfad, die X-Beine, die S-Kurve

Das heißt, die Großschreibung von Substantiven greift nicht nur wortinitial nach Leerzeichen, sondern auch wortintern nach Bindestrichen.

3.3 Die wortinternen Varianten

Die Norm sieht also für die Schreibung wortbildungsmorphologisch komplexer Substantive neben der Zusammenschreibung in einigen wenigen Fällen auch eine Bindestrichschreibung vor, während andere Varianten nicht lizenziert sind.

Dass sich im Sprachgebrauch aber daneben auch wortinterne Leerzeichen, Majuskeln oder Apostrophierungen beobachten lassen, stößt nicht nur Sprachkritikern auf z. B. Starke (1993) und Heller (2000), sondern ist auch schon längst als linguistischer Forschungsgegenstand etabliert.

Scherer (2012) nimmt in ihrem Artikel zu N+N-Komposita sogar eine Gewichtung der möglichen Grenzmarkierungen vor:

Was die Stärke der Grenzmarkierung angeht, so kann das Spatium zweifellos als stärkste Form der Markierung angesehen werden, da es die Grenze zwischen zwei graphischen Wörtern darstellt. Bindestrich und Apostroph wiederum nehmen eine deutlichere Grenzmarkierung vor als Majuskeln, da Wortzeichen in die Graphemkette eingefügt werden und den Graphemfluss unterbrechen. Bei der Verwendung von Majuskeln hingegen handelt es sich um die schwächste Form der Grenzmarkierung, da hier lediglich gegen kontextuelle Beschränkungen bei der Wahl des Allographen für ein Graphem verstoßen wird. (Scherer 2012: 66)

Im Zusammenhang damit ist allerdings nicht davon auszugehen, dass die Binnenmajuskel als schwächste Grenzmarkierung auch die akzeptabelste bzw. häufigste ist, sondern sicherlich eine Schreibung mit Bindestrich (auch durch die Rückendeckung der Norm) prominenter ist. Nach Scherer (2012: 69) ist eine Abbildung dieser Hierarchie in der Schreibung auch nur in mindestens dreigliedrigen Komposita zu erwarten und zwar derart, dass die Grenze zwischen den unmittelbaren Konstituenten stärker markiert wird als die Grenzen zwischen mittelbaren Konstituenten wie in ihren Beispielen: *Dinkel Butter-Gebäck*, *Dinkel ButterGebäck*, *Dinkel-Buttergebäck*.

Die folgende Zusammenstellungen zur Zusammenschreibung, Bindestrichschreibung, Getrenntschreibung, Schreibung mit Binnenmajuskel oder Apostroph erheben keinen Anspruch auf Vollständigkeit im Sinne eines umfassenden Forschungsüberblicks, sondern sind der Versuch einer Konzentration auf wesentliche Aspekte im Sinne der Untersuchung.

3.3.1 Zusammenschreibung

Egal, ob es sich um einfache Stämme wie *Haus* oder wortbildungsmorphologisch komplexe Wörter wie *Haustürschloss* oder *Häuslichkeit* handelt, Wörter werden im Deutschen grundsätzlich zusammengeschrieben.⁴⁰

Graphematisch sind Wörter sogar über ihre Zusammenschreibung definiert als Zeichenkette zwischen zwei Leerzeichen ohne interne Leerzeichen (vgl. Fuhrhop 2008: 193). Ein Leerzeichen zwischen Zeichenketten zeigt damit an, dass es sich nicht um Wortbestandteile handeln kann, sondern viel mehr von syntaktischen Einheiten auszugehen ist. Während nach Fuhrhop (2007) das morphologische Prinzip für die Zusammenschreibung von zwei oder mehr Stämmen sorgt, sofern sie aufgrund einer Wortbildung miteinander verbunden sind, sorgt das syntaktische Prinzip dafür, dass syntaktisch analysierbare Einheiten, die in syntaktischen Beziehungen zu anderen Einheiten im Satz stehen, durch Leerzeichen abgetrennt werden. Dass die Unterscheidung zwischen den Bestandteilen eines Wortes und syntagmatisch verknüpften Einheiten nicht immer so trivial ist, wie es im ersten Moment scheint, zeigt Fuhrhop (2007) mit der Diskussion zahlreicher Zweifelsfälle zwischen Wort und Syntagma wie beispielsweise trennbarer Verben (*radfahren*, *eislaufen*) oder Verbindungen mit

⁴⁰ Die Zusammenschreibung ist auch gemäß der Schreibnorm der Standardfall. Vgl. Teilkapitel 3.2 *Orthographie*.

dem Partizip I (*die Bier trinkenden Fans/ die biertrinkenden Fans*). Der hier fokussierte substantivische Bereich gehört dagegen größtenteils zu den unproblematischen Fällen, zu denen Fuhrhop (2007: 167-172) neben Affigierungen auch Substantivkomposita zählt.

Zu sehr ähnlichen Ergebnissen zur Zusammen- und Getrennschreibung kommt Jacobs (2005). Er wählt dafür aber den Weg einer Optimalitätstheoretischen Beschränkungshierarchie. Grundlegend für die Zusammen- und Getrennschreibung sind dabei die zwei Beschränkungen ZUS-MORPH und GETR-AUSDR, die deutliche Parallelen zu den Prinzipien nach Fuhrhop (2007) aufweisen.

(26) ZUS-MORPH

ZUSAMMENSCHREIBUNG VON MORPHOLOGISCHEN BILDUNGEN

Wenn X und Y Teilausdrücke eines morphologisch gebildeten Zeichens sind, gibt es zwischen <X> und <Y> kein Spatium. (Jacobs 2005: 34)

(27) GETR-AUSDR

GETRENNTSCHREIBUNG VON TEILAUSTRÜCKEN

Wenn X und Y Teilausdrücke sind, liegt zwischen <X> und <Y> mindestens ein Spatium. (Jacobs 2005: 97)

Diese und weitere Beschränkungen nach Jacobs (2005) spielen auch eine wichtige Rolle für die Optimalitätstheoretische Modellierung der wortinternen Varianten in Kapitel 6 und werden dort ausführlicher diskutiert.

Hier bleibt festzuhalten, dass die Zusammenschreibung für wortbildungsmorphologisch komplexe Wörter der absolute Normalfall ist, sie sogar im Vergleich zu syntaktischen Verbindungen definiert.

Wenn im Folgenden von Zusammenschreibung die Rede ist, ist damit in Abgrenzung zu den unten erläuterten, anderen wortinternen Varianten nicht nur das Fehlen wortinterner Leerzeichen, sondern auch die Abwesenheit von wortinternen Interpunktions- oder sonstigen Zeichen und Majuskeln gemeint. Auch wenn Bindestrichschreibungen und Schreibungen mit Binnenmajuskeln (*Bindestrich-Schreibung*, *BinnenMajuskel*) eigentlich auch insofern Zusammenschreibungen sind, als es sich um Zeichenketten zwischen zwei Leerzeichen handelt, wird in der Folge der Begriff der Zusammenschreibung enger gefasst und, wo nicht anders angegeben, aus-

schließlich auf Wörter oder Wortgrenzen angewandt, die keine wie auch immer geartete sonstige Variation aufweisen.

3.3.2 Bindestrichschreibung

Unter den Abweichungen von der Zusammenschreibung nimmt die Schreibung mit Bindestrich eine Sonderrolle ein. Neben der Zusammenschreibung ist der Einsatz des Bindestrichs für einen begrenzten Bereich die einzige weitere orthographisch legitimierte Variante (Vgl. Teilkapitel 3.2 *Orthographie*).

Graphematisch haben sich bisher insbesondere Bredel (2008), Fuhrhop (2008) und Buchmann (2015) mit dem Bindestrich beschäftigt.⁴¹

Bredel (2008: 115-117) verortet nicht nur den Bindestrich klar in morphologisch komplexen Ausdrücken zwischen zwei lexikalischen Wörtern (*See-Elefant*), sondern setzt sich auch mit seiner Funktion auseinander. Der Bindestrich solle dafür, „*dass vor der Anwendung von Wortbildungsregeln jeder separierte Wortbestandteil einzeln lexikalisch verarbeitet wird*“ (Bredel 2008: 115). Für die vereinzelte lexikalische Verarbeitung spreche der Umstand, dass der Bindestrich bei Kompositionen mit Fugenelement weniger akzeptabel sei, weil das Fugenelement bei einer isolierten lexikalischen Verarbeitung nicht interpretiert werden könne. Aber auch die Tatsache, dass bei einer Bindestrichschreibung die Großschreibungsmechanismen auf alle mit Bindestrich getrennten Substantive wirken können.

Laut Bredel (2008: 116) organisiert der Bindestrich (wie alle Divisvorkommen) die Teilaktivitäten der Einzelworterkennung und -verknüpfung beim Leser komplexer morphologischer Wörter. Das Lesen des Bindestrichs führt zunächst zum Abbruch der morphologischen Verkettung und zur Ablage des Eingelesenen im Arbeitsspeicher, bevor eine Verknüpfung mit der Folgeeinheit stattfindet. So kommt es zu einer Trennung von Worterkennung und Wortverknüpfung durch den Bindestrich beim Leseprozess. Dadurch sei der Bindestrich umso akzeptabler, „*je eher die Einzelbestandteile morphologisch erkennbar und einer autonomen Interpretation zugänglich sind*“ (Bredel 2008: 116).

Geilfuß-Wolfgang (2007b: 81) verweist auf eine Blickbewegungsstudie von Pfeiffer (2002), deren Ergebnisse diese Annahme stützen, da sie dafür sprechen, dass die

⁴¹ Zur formalen Analyse des Bindestriches, die hier nicht vertieft wird, sei auf Bredel (2008: 24-30) verwiesen.

initiale Sakkade ins Wort bei Komposita mit Bindestrich weiter vorne landet als in Komposita ohne Bindestrich (3,59 Buchstaben vor der Wortmitte im Vergleich zu 2,62 Buchstaben vor der Wortmitte). Diese Beobachtung ließe den Schluss zu, dass sich die initiale Sakkade nicht nach dem Gesamtwort, sondern dem ersten Teilwort richtet. Auf dieselbe Studie bezieht sich auch Geilfuß-Wolfgang (2013: 138-140) und erläutert dort, dass Pfeiffer (2002) anhand zweigliedriger Komposita wie *Gartenarbeit* und *Gartenfrucht* mit Blickbewegungsstudien untersucht, wie Komposita mit Bindestrich gelesen werden. Seine Grundannahme dabei ist, dass Komposita lexikalisch auf zwei Wegen verarbeitet werden können: Häufigere Komposita wie *Gartenarbeit* erhalten einen lexikalischen Eintrag aus dem die Gesamtbedeutung abgerufen werden kann, während seltene, aber transparente Komposita ohne eigenen Eintrag wie *Gartenfrucht* über die Einzelbedeutungen lexikalisch verarbeitet werden müssen. Im zweiten Fall von Komposita ohne Lexikoneintrag wie *Gartenfrucht* sollte sich also ein positiver Effekt durch einen Bindestrich ergeben, weil die Analyse durch die Grenzmarkierung beschleunigt wird, während für lexikalisierte Komposita wie *Gartenarbeit* durch das Einfügen eines Bindestrichs ein negativer Effekt zu erwarten ist, weil der Zugriff auf den vorhandenen Lexikoneintrag erschwert wird. Diese Annahmen kann Pfeiffer (2002) in seiner Studie bestätigen. Bei gängigeren Komposita vom Typ *Gartenarbeit* sorgt ein Bindestrich zwischen den Konstituenten für höhere Blickzeiten und eine erhöhte Gesamtlesezeit, die wohl in einer zweiten Fixation auf dem Zweitglied nach dem Bindestrich begründet sind, die bei einer Schreibung ohne Bindestrich gar nicht erforderlich ist. Bei selteneren Komposita des Typs *Gartenfrucht* hat ein Bindestrich keinen Einfluss auf die Blickzeiten, aber die Gesamtlesezeit verkürzt sich. Ein Bindestrich zwischen den Konstituenten eines selten belegten oder okkasionellen Kompositums beschleunigt also wohl die lexikalische Verarbeitung.

Die Domäne des Bindestrichs nach Bredel (2008) sind Kompositionsglieder und bestimmte Konfixe (*Bio-Masse*), während Flexions- oder Derivationsmorpheme nicht durch den Bindestrich abgetrennt werden. Bezüglich der Kontexte, in denen ein Bindestrich gesetzt wird, trifft Bredel (2008) die generelle Annahme:

Je autonomer die Einzelbestandteile und je weniger fest/klar (bzw. je ambiguitätserzeugender) die Verknüpfungsstruktur ist, desto wahrscheinlicher ist der Bindestrich. (Bredel 2008: 116)

Im Folgenden exemplifiziert sie die zentralen Begriffe dieser Annahme:

Autonomie von Einzelbestandteilen liegt bei heterogenen semiotischen Basen vor (3/4-Takt, *b-Moll*)¹⁷[Fußnote: *Heterogenität kann auch bei Flexionsmorphemen zur Bindestrichschreibung führen (das x-te Mal).*], nicht-feste Verknüpfungsstruktur liegt vor, wenn die zu verknüpfenden Ausdrücke nicht durch einen regulären Wortbildungsprozess zustande gekommen sind (sog. Durchkoppelungsbindestrich: *das Auf-der-Hut-Sein*). Unklare Verknüpfungsstrukturen ergeben sich bei der Komposition von abgekürzten Einheiten (**Dipling, Dipl.-Ing., aber: Diplomingenieur*). (Bredel 2008: 116)

Darüber hinaus hält sie fest, dass kopulative Kompositionsstrukturen wie *süßsauer* oder *Magen-Darm-Grippe* für Bindestriche zugänglicher sind als determinative, und führt dies darauf zurück, dass bei kopulativen Strukturen die einzelnen Bestandteile semantisch intakt bleiben, während bei determinativen Strukturen ein semantisches Modifikationsverhältnis zwischen den Strukturen besteht.

Fuhrhop (2008: 203-205) beschäftigt sich noch ausführlicher damit, in welchen Komposita Bindestriche auftreten, und macht neun Gruppen aus:

- (28)
- a. Tee-Ei, Armee-Einsatz
 - b. Goethe-Forschung, Eisenberg-Grammatik, Schröder-Biografie
 - c. China-Restaurant – Deutschlandbild
 - d. Den-Job-Wechseln, Schönes-Wochenende-Ticket, Auf-das-Wetter-Schimpfen
 - e. rot-grün, amerikanisch-israelisch
 - f. *isch*-Adjektive, *ge*-Partizipien, *zu*-Infinitive, *Ja*-Wort
 - g. US-Regierung, SPD-Vorsitzender, IP-Nummer
 - h. Software-Entwicklung, Online-Markt, Info-Highway
 - i. Hair-Stylist, Dow-Jones-Index, Know-How, Joint-Venture

Neben kopulativen Strukturen (28e) und dem Durchkoppelungsbindestrich (28d), die auch Bredel (2008) erwähnt, nennt sie auch die Aufeinanderfolge dreier gleicher Buchstaben (28a) und die Kombination zweier fremder Bestandteile (28i).

Die übrigen fünf Gruppen nach Fuhrhop (2008) könnten auch unter den Begriff der Heterogenität von Bredel (2008) fallen. Egal, ob Namen (28b),⁴² Objektsprache (28f), Akronyme (28g) oder ein fremder erster Bestandteil (28h), in allen Fällen un-

⁴² Mit dem Beispiel (28c) macht Fuhrhop (2008) darauf aufmerksam, dass es bei Ländernamen Unterschiede im Vergleich zu Personennamen geben kann.

terscheiden sich die Wortbestandteile aufgrund ihrer Semantik, Pragmatik, Graphematik oder Herkunft, so dass sie eine Bindestrichschreibung begünstigen.

Nach Buchmann (2015: 46-48) ist eine systematische Unterscheidung in morphologisch und graphematisch motivierte Bindestrichschreibungen angebracht. Zu den morphologisch motivierten Schreibungen zählt sie dabei Einheiten mit Eigennamen (*Goethe-Forschung*), Einheiten, die insgesamt Syntagmen sind (*Den-Job-Wechseln*)⁴³ oder syntagmenhaltige Vorderglieder aufweisen (*Schönes-Wochenende-Ticket*) und adjektivische Kopulativkomposita (*rot-grün*). Die graphematisch motivierten Schreibungen umfassen Schreibungen mit Akronymen (*US-Regierung*) und solche mit Objektsprache (*isch-Adjektive*).

Darüber hinaus bietet Buchmann (2015: 45-57) einen ausführlichen und gut sortierten Überblick über den Forschungsstand zu Form und Funktion des Bindestriches sowie zu den Forschungsbeiträgen aus graphematischer, wortbildungsmorphologischer, sprachhistorischer und orthographischer Perspektive.

Der Kern der Arbeit von Buchmann (2015) ist aber die Korpusanalyse eines großen Zeitungskorpus zu Wortformen mit Wortzeichen (Abkürzungspunkt, Bindestrich und Apostroph). Die Bindestrichschreibung systematisiert und diskutiert sie als graphisch markierte Schreibungen (mit Majuskelschreibungen, Interpunktions-, Sonder- und Wortzeichen sowie Ziffern) und graphisch unmarkierte Schreibungen nach den morphologischen Kriterien Fremdwörter, Eigennamen, Kurzwörter, Konfixe und fremde Präfixe.

Ihre Analyse beruht auf ca. 270.000 qualitativ untersuchten Bindestrichschreibungen, von denen eine repräsentative Stichprobe von 1350 Schreibungen händisch mit Informationen zu Wortart sowie graphischen und morphologischen Eigenschaften der Bestandteile getaggt wurde (vgl. Buchmann 2015: 216). Der Großteil dieser Schreibungen weist nur einen Bindestrich auf.⁴⁴ Determinativkomposita umfassen maximal zwei Bindestriche, häufigere Vorkommen entfallen auf syntagmenhaltige Wortformen (vgl. Buchmann 2015: 219). Dementsprechend analysiert sie zunächst die Schreibungen mit nur einem Bindestrich, bevor sie sich Schreibungen mit meh-

⁴³ Buchmann (2015), Fuhrhop (2008) und Jacobs (2005) analysieren diese Einheiten nicht als Wortbildungsprodukt (Konversion) wie Fleischer & Barz (2012).

⁴⁴ In einer Abbildung zur Verteilung der Bindestrichschreibungen wird die Zahl 248.000 für Schreibungen mit einem Bindestrich genannt, allerdings ergeben sich in Summe aller Kategorien über 275.000 Bindestrichschreibungen (Buchmann 2015: 219).

reren Bindestrichen zuwendet.

- (29) a. A-Jugend, BMX-Fahrrad, BRIGITTE-Diät
 b. enjoy-Tarif, ph-Wert
 c. BayernLB-Chef, FlowTex-Betrugsfall

Zunächst diskutiert Buchmann (2015: 220-225) Majuskelschreibungen mit Bindestrich. Dazu zählt sie, wie die Belege unter (29) zeigen, neben Beispielen mit (einzelnen,) wortinternen Majuskeln (29c) und Konstituenten, die durchgängig in Majuskeln geschrieben sind (29a), auch Belege, die eine erste Konstituente in durchgängiger Minuskelschreibung aufweisen trotz Substantivstatus des Wortes (29b). Die Einheiten mit Binnenmajuskel(n) oder durchgängiger Majuskel- bzw. Minuskelschreibung sind nicht nur graphematisch markiert, sondern auch morphologisch meist Sonderfälle. Es handelt sich in der Regel um Einzelbuchstaben (*a*), Akronyme (*BMX*, *ph*) oder Eigennamen (*enjoy*, *BayernLB*, *FlowTex*, *BRIGITTE*). Aufgrund dieser Beobachtungen formuliert Buchmann (2015: 224) die These, dass markierte Schreibungen „mit einem auffälligen oder fremden grammatischen Verhalten“ (beispielsweise veränderte Lautstruktur und Flexionsverhalten bei Kurzwörtern) einhergehen. Darüber hinaus nimmt sie (vermutlich auf Basis der Eigennamen) an, dass graphische Auffälligkeiten auch bewusst vom Schreiber geschaffen werden, um eine erhöhte Aufmerksamkeit vom Leser zu erzwingen.

Unter ihren Belegen für Bindestrichschreibungen finden sich auch 66 Beispiele, bei denen auf ein Akronym plus Bindestrich kein lexikalischer Stamm sondern entgegen Bredel (2008) ein Affix (*-ler*) folgt: *FH-ler*, *ABC-ler*.

Abschließend hält Buchmann (2015: 225) zu den Majuskelschreibungen fest, dass die Markierung eines Wortbestandteils durch Majuskelschreibung dazu führt, dass weitere Bestandteile mit einem Bindestrich angefügt werden und vermutet, dass der Schreiber damit „ein graphisches Verbindungselement zwischen einer graphisch markierten Form und einer Form in Normalschreibung“ setzen will.

Ebenso graphisch markiert, wenn auch seltener belegt als Majuskelschreibungen, sind Wortbestandteile mit Interpunktions-, Sonder- oder Wortzeichen wie in (30) (vgl. Buchmann 2015: 225-233).

- (30) a. Messdiener/innen-Treffen

- b. „Big Brother“-Kandidat
- c. Tel.-Nummer
- d. Geburtstags-Rock'n'Roll

Die Funktion des Schrägstrichs als Sonderzeichen liegt nach Buchmann (2015: 225) in der Koordination zweier Einheiten. Neben einer Genderfunktion wie in *Messdiener/innen-Treffen*, die zwei Geschlechter koordiniert, sei auch eine Koordination von Gegenständen, Gruppen oder Größen in kopulativen Komposita möglich (*Herbst-Winter-Kollektion*). Der Bindestrich ist dabei nur im zweiten Fall eine mögliche Alternative.

Unter den Belegen für Bindestrichschreibungen mit Anführungszeichen finden sich nach Buchmann (2015: 227-228) neben einigen Beispielen für den modalisierenden Gebrauch (*Fußball-„Fans“*) hauptsächlich solche für den konventionellen Gebrauch in Form von Zitaten und Titelnennungen (*„Was-willst-du-werden-wenn-du-mal-groß-bist“-Frage*, *„Harry Potter“-Romane*). Auch hier sorgt die graphische Markierung durch die Anführungszeichen für eine Bindestrichschreibung der Komposita.

Im Zusammenhang mit Beispielen wie *„Was-willst-du-werden-wenn-du-mal-groß-bist“-Frage* und *„Harry Potter“-Romane* weist Buchmann (2015: 229) darauf hin, dass Anführungszeichen die Definition eines graphematischen Wortes außer Kraft setzen können. Für die Definition des graphematischen Wortes nach Fuhrhop (2008) als Zeichenkette zwischen zwei Leerzeichen ohne interne Leerzeichen sind die wortinternen Leerzeichen in der Anführungszeichenumgebung tatsächlich zunächst problematisch. Allerdings ist davon auszugehen, dass die Anführungszeichen auf linear-suprasegmentaler Ebene (wie beispielsweise auch eine Kursivierung: *Harry Potter-Romane*) dazu führen, dass alles dazwischen und die Anführungszeichen selbst als Einheit aufgefasst werden, so dass Leerzeichen innerhalb von Anführungszeichen kein Grund sind, mehrere graphematische Wörter anzunehmen.

Auch weitere Satz- oder Sonderzeichen, die inhärent bei vereinzelt Kompositionsgliedern sind, sorgen in diesen Fällen für eine graphische Markiertheit und machen eine Schreibung mit Bindestrich nötig, wie die Beispiele mit Abkürzungspunkt und Apostroph unter (30c) und (30d) oder die Beispiele unter (31) zeigen.

(31) m:con-Mitarbeiter, P+R-Anlage, C&A-Sprecher, ver.di-Landesverband

Die letzte Gruppe graphisch markierter Schreibungen nach Buchmann (2015: 233-235) überschreibt sie *Ziffern in Zahlen*. Auch hier ergibt sich das Problem potentiell wortinterner Leerzeichen bei Schreibungen wie *35 000-Marke*. Wie schon bei den Anführungszeichen ist aber dennoch aufgrund der unterschiedlichen Segmentklassen von einer Einheit des aus Ziffern aufgebauten Wortbestandteils auszugehen (vgl. Buchmann 2015: 234). In diesen Fällen wird der Bindestrich also als Verbindungszeichen zwischen Zeichen unterschiedlicher Segmentklassen gesetzt. Eine Kombination von Ziffern und Buchstaben ohne Verbindungszeichen ist nur in Beispielen wie *0190er-Nummern* in Kombination mit einem Affix zu beobachten. Buchmann (2015: 235) sieht darin einen Reflex der orthographischen Norm, die bei Ableitungen abseits von Einzelbuchstaben als Basis die Zusammenschreibung fordert, während die ebenfalls, aber seltener belegte Schreibung mit Bindestrich *0190-er-Nummern* das Produkt des graphematischen Systems sein könnte.

Zu den graphisch markierten Schreibungen hält sie abschließend fest, dass alle besprochenen, graphisch markierten Einheiten ein Verbindungselement brauchen, um mit einer weiteren Einheit verknüpft werden zu können (vgl. Buchmann 2015: 235).

Unter *graphisch unmarkierten Schreibungen* diskutiert Buchmann (2015: 236-253) dann Bindestrichschreibungen mit Fremdwörtern, Eigennamen, Kurzwörtern, Konfixen und fremden Präfixen sowie zwei Problemfälle. Bei diesen Problemfällen handelt es sich um Rektionskomposita (*Geldwäscherfahndung*), die sich in der Diskussion als unproblematisch herausstellen, und native Einheiten mit und ohne Fugenelement (*Arbeits-Urlaub*, *Abgas-Grenzwert*) (vgl. Buchmann 2015: 249-252). Letztere sollten nach der Theorie von Bredel (2008) mangels Unterschiedlichkeit tatsächlich nicht mit einem Bindestrich vorkommen. Buchmann (2015: 250) verweist daher für die sieben betroffenen Lexeme, die fast alle häufiger als Zusammenschreibung belegt sind, auf die Möglichkeit eines pragmatischen Kontextkriteriums. Da sich unter den Belegen aber auch Fälle mit mehr als zwei Kompositionsgliedern befinden (*Weihnachtsmarkt-Häuschen*), bringt sie auch die Idee eines möglichen Längeneffekts ins Spiel, der für eine Trennung verantwortlich sein könnte, allerdings ohne diesen weiter zu diskutieren.

Fremdwörter werden als Wortbestandteile nach Buchmann (2015: 237) „*bevorzugt mit dem Bindestrich segmentiert*“, und zwar unabhängig vom Zeitpunkt der Entlehnung und dem Fortschritt des Integrationsprozesses. Aus einer Übersicht zu

den Bindestrichschreibungen mit Fremdwörtern (Buchmann 2015: 239) geht hervor, dass von zwölf betreffenden Lexemen (32) bis auf drei Ausnahmen (*Online-Haus*, *Orchester-Ersatz*, *Panorama-Glasschiebedach*) alle anderen auch in Zusammenschreibung belegt sind, z. B. *Alumni-Tag*.

- (32) a. Alumni-Tag, Cartoon-Zeichner, Online-Haus, Orchester-Ersatz,
Panorama-Glasschiebedach, Misch-Typen, Bürger-Büro
b. Traditions-Termin, Theater-Projekte, Telefon-Terror
c. musical-erfahrene, yoga-ähnliche

Gerade die Belege mit zwei fremden Bestandteilen unter (32b) sind durchweg in der zusammengescriebenen Variante zehn Mal so häufig wie die Bindestrichschreibung, während die Belege mit einem fremden Anteil (32a) nicht oder etwa genauso häufig in Zusammenschreibung belegt sind. Aufgrund dieses Befundes ist Buchmann (2015: 238) zu relativieren, die mit Blick auf die Beispiele in (32a) und (32b) die These von Bredel (2008) in Frage stellt, dass der Bindestrich lediglich zwei grammatisch ungleiche Einheiten miteinander verbinde. Zwar zeigen die Belege (32b), dass ein Bindestrich auch zwischen zwei fremden Einheiten realisiert wird, dort aber eher die Ausnahme ist, da alle beobachteten Lexeme deutlich häufiger zusammengescrieben werden, während bei den Belegen mit nur einem fremden Bestandteil (32a) die Bindestrichschreibung überwiegt.

Passender wäre also die Feststellung, dass Belege mit einem fremden Bestandteil bevorzugt mit dem Bindestrich segmentiert werden, während beim Aufeinandertreffen zweier fremder Bestandteile der Bindestrich auch möglich ist, aber die Zusammenschreibung bevorzugt wird.

Die zweite große Gruppe von Bindestrichschreibungen ohne graphische Markierungen nach Buchmann (2015) machen Belege mit Eigennamen aus. Wie schon die Fremdwörter unterscheiden sich auch Eigennamen signifikant vom Kernwortschatz (Monoreferenz, Artikellosigkeit, stark eingeschränkte Pluralbildung). Aus ihrer Datensammlung gehen acht Lexeme (33) für eine Bindestrichschreibung bei Eigennamen hervor (Buchmann 2015: 241).

- (33) a. Amerika-Reise, Casablanca-Bühne, Mexiko-Stand, Olympia-Gastgeber,
Werder-Angreifer

b. israel-feindliche, ferrari-roten, europa-skeptische

Dabei sind für alle adjektivischen Belege sowie *Amerika-Reise* und *Olympia-Gastgeber* auch Zusammenschreibungen belegt (*Amerikareise*, *Olympiagastgeber*). Nach Buchmann (2015: 241-242) kann diese Verteilung zufällig sein. Dort wo der Bindestrich gesetzt wird, führt sie ihn auf eine Wortschonungsstrategie nach Nübling (2014) zurück, die dafür sorgt, dass Eigennamen in ihrer grammatischen Struktur (phonologisch, morphologisch, graphematisch) möglichst konstant gehalten werden.

Den nächsten Abschnitt widmet Buchmann (2015) den graphisch unmarkierten Kurzwörtern.

(34) Alu-Leiter, Abi-Feier, Abo-Büro, Dia-Schau, Uni-Sprecher, Lok-Führer, Zoo-Sprecherin

Alle Beispiele für Bindestrichschreibungen unter (34) sind im Korpus auch (meist vergleichbar häufig) als Zusammenschreibungen belegt (vgl. Buchmann 2015: 242). Sie weist zurecht darauf hin, dass die Erstbestandteile in der Regel nicht nur Kurzwörter sind, sondern auch aus Fremdsprachen stammen, sich also gleich zwei morphologische Kriterien überschneiden, die im Zusammenhang mit Bindestrichschreibungen stehen.⁴⁵ Außerdem könnten Kurzwörter wie *Uni* auch als Eigennamen verwendet werden, die wie Fremdwörter eine Tendenz zur Bindestrichschreibung fördern (vgl. Buchmann 2015: 245-246). Womit sich meiner Meinung nach in letzter Konsequenz die Frage stellt, ob tatsächlich davon auszugehen ist, dass Kurzwörter als morphologische Kategorie einen Effekt auf die Schreibung haben, oder ob sich dieser Effekt nicht eher durch ihre graphematischen Eigenschaften, den fremden Ursprung oder Eigennamenstatus erklären lässt. Womöglich ist diese Kategorie ein bloßer Reflex auf § 40 der orthographischen Norm, der einen Bindestrich bei Zusammensetzungen mit Abkürzungen fordert.

Weitere acht Belege bespricht Buchmann (2015: 246-249) unter dem Titel *Konfixe*

⁴⁵ In einem kleinen Exkurs betrachtet sie außerdem den Unterschied zwischen graphematisch unmarkiert(er)en (*Pin*, *Aids*, *Pkw*, *Lkw*, *Kfz*) Kurzwörtern und ihren markierten Pendanten (*PIN*, *AIDS*, *PKW*, *LKW*, *KFZ*) hinsichtlich ihrer graphischen Anknüpfung in komplexen Wörtern mit dem Ergebnis, dass eine Zusammenschreibung nur in Einzelfällen, bei den unmarkierten Kurzwörtern, die einen kompakten Vokalbuchstaben enthalten (*Pin*, *Aids*) und damit einer regulären graphematischen Silbe entsprechen, belegt ist (vgl. Buchmann 2015: 243-245).

und fremde Präfixe.

- (35) Ex-Freundin, Mini-Club, Öko-Steuer, Bio-Produkte, Extra-Klasse, Super-Sommer, Anti-Held, Turbo-Abitur

Die Klassifikation der Erstbestandteile unter (35) ist durchaus umstritten. Da die Entscheidung für das Schreibverhalten wohl keine Rolle spielt, wird die Diskussion hier nicht weiter vertieft, sondern auf die Zusammenstellung der relevanten Literatur durch Buchmann (2015: 246) verwiesen. Zu allen Beispielen unter (35) gibt es bindestrichlose Varianten, die zumeist sogar häufiger belegt sind. Allein *Ex-Freundin* (759:61) und *Bio-Produkte* (114:82) sind häufiger mit dem Bindestich zu finden (vgl. Buchmann 2015: 247). Allen ersten Einheiten ist gemein, dass sie fremden Ursprungs sind, also schon dadurch eine Bindestrichschreibung begünstigen. Daneben diskutiert Buchmann (2015: 247-248) völlig zurecht, dass die Beispieleinheiten, anders als von Prä- oder Konfixen zu erwarten, durchaus wortfähig sind. Sie können alleine stehen (*Ich bin ein verkappter Bio.*), entwickeln Flexionsverhalten (*Wir sind Ökos.*) oder erfahren eine Genuszuweisung (*Er trifft immer noch seine Ex.*).

Neben der Einschätzung, dass der Bindestrich hier als Markierung für Einheiten funktioniert, deren Verhalten morphologisch nicht eindeutig ist (vgl. Buchmann 2015: 249), muss auch berücksichtigt werden, dass dieser potentielle Effekt immer mit dem der Fremdsprachlichkeit kombiniert ist.

Zusammenfassend gilt für die einfachen Bindestrichschreibungen aufgrund morphologischer Kriterien nach Buchmann (2015), dass es anders als bei den graphisch markierten Einheiten immer auch die Möglichkeit der Zusammenschreibung gibt. Zudem ist eine Kombination der diskutierten morphologischen Kriterien in einer Einheit möglich, wie die Diskussion der Kriterien *Kurzwörter* und *Konfixe und fremde Präfixe* gezeigt hat.

Wortformen mit mehreren Bindestrichen beobachtet Buchmann (2015: 254-259) bei Substantiven im Fall von kopulativen Strukturen wie *Mutter-Kind-Kur*, bei syntaktischen Wörtern, die aus Syntagmen bestehen (*Hans-Dampf-in-allen-Gassen*), und Syntagmen als Erstgliedern von Determinativkomposita wie in *Trimm-Dich-Pfad*.

Am Beispiel *Patenkind-Turnier/Patenkindturnier* im Vergleich zu *Paten-Kind-Turnier* illustriert Buchmann (2015: 255-256), dass bei *Paten-Kind-Turnier* der Bindestrich zwischen den ersten beiden Bestandteilen dafür sorgt, dass hier nur die Lesart möglich

ist, dass bei dem Turnier Paten und Kinder teilnehmen, während die Zusammenschreibung oder die Schreibung mit einem Bindestrich vor *Turnier* auch die Lesart zulässt, dass sich nur die Kinder im Turnier messen. Es bietet sich also die These an, dass der Bindestrich kopulative Strukturen anzeigt.⁴⁶

- (36) Nach-Hause-Weg, Trimm-Dich-Pfad, Rund-um-die-Uhr-Bereitschaft, Von-Haus-zu-Haus-Geschäft

Die Bindestriche innerhalb der syntaktischen Strukturen machen auch innerhalb der Wortbildung deutlich, dass es sich an dieser Stelle eigentlich um eigenständige morphosyntaktische Wörter in syntaktischen Beziehungen handelt. Bei den Belegen für syntagmenhaltige Vorderglieder (36) ist festzustellen, dass die ersten beiden auch mit einem („determinativen“) Bindestrich (*Nachhause-Weg*, *Trimmdich-Pfad*) und als Zusammenschreibungen (*Nachhauseweg*, *Trimmdichpfad*) vorliegen (Buchmann 2015: 257). Dabei sind *Nachhauseweg* und *Trimm-Dich-Pfad* jeweils die häufigsten Varianten. Buchmann vermutet hier die stärkere Lexikalisierung der Syntagmen als Ursache für die Möglichkeit oder den Vorrang der Zusammenschreibung.

Ganz ähnlich ist ihr Befund bei Syntagmen als syntaktischen Wörtern:

- (37) Hans-Dampf-in-allen-Gassen, Sehr-früh-mit-allem-Anfangen, Sich-selbst-auf-die-Schulter-Klopfen

Auch hier ersetzt der Bindestrich auf Wortebene die Leerzeichen der zugrundeliegenden syntaktischen Struktur und die Großschreibung der syntaktischen Ebene bleibt wortintern erhalten (vgl. Buchmann 2015: 258).

Mehr als ein Bindestrich pro Wortform in Substantiven findet sich in den Daten von Buchmann (2015) also nur zur Anzeige kopulativer Strukturen als Vorderglied eines determinativen Kompositum, dessen Kopf ebenfalls durch einen Bindestrich abgetrennt ist, oder als Kennzeichen ursprünglich syntaktischer Strukturen auf der Wortebene. Der Bindestrich zwischen Syntagma und Wortbildungskopf ist dabei obligatorisch, die Bindestriche innerhalb des Syntagmas können durch andere Mittel der Schrift (z. B. Anführungszeichen, Kursivierung) ersetzt werden (vgl. Buchmann

⁴⁶ Dies ist auch bei Adjektiven zu beobachten: Während *blaugrün* determinativ als bläuliches Grün und kopulativ als blau und grün interpretiert werden kann, legt *blau-grün* die kopulative Lesart nahe (Buchmann 2015: 256).

2015: 259).

Mit dieser Beispielgruppe schließt Buchmann (2015) ihre Datenanalyse und widmet sich abschließend einer Gewichtung der Kriterien:

Graphematische Kriterien sind grundsätzlich stärker als morphologische, denn bei den graphematischen Kriterien gibt es in der Regel keine bindestrichlose Alternativschreibung. (Buchmann 2015: 260)

Einer weiteren Gruppe von Bindestrichschreibungen, die nicht prominent in den Daten von Buchmann (2015) in Erscheinung tritt,⁴⁷ widmet sich Geilfuß-Wolfgang (2013). Er untersucht mit Hilfe einer kleinen Fragebogenstudie (Bewertung anhand einer Achter-Skala) die Setzung von Bindestrichen in dreigliedrigen Determinativkomposita anhand okkasioneller Komposita unter dem Gesichtspunkt, ob dort durch Bindestriche die Verzweigungsstruktur angezeigt wird. Die Möglichkeiten der Verzweigung und die damit zusammenhängende Interpretation der dreigliedrigen Komposita erläutert er am Beispiel *Hammerklaviersonate*, das als Linksverzweigung (*[Hammerklavier]sonate*) als Sonate fürs Hammerklavier interpretiert wird, als Rechtsverzweigung (*[Hammer[klaviersonate]]*) aber als Klaviersonate, die der Hammer ist (vgl. Geilfuß-Wolfgang 2013: 135). In solchen Fällen könne ein Bindestrich also eine disambiguierende Funktion einnehmen, indem er die Haupttrennfuge zwischen den beiden unmittelbaren Konstituenten markiert. *Hammerklavier-Sonate* wäre dann eindeutig als Sonate fürs Hammerklavier zu interpretieren und *Hammer-Klaviersonate* als Klaviersonate, die der Hammer ist.

Geilfuß-Wolfgang (2013: 74) nimmt also an, dass unabhängig von der Verzweigungsrichtung eine Schreibung mit dem Bindestrich an der Haupttrennfuge besser bewertet wird als eine Zusammenschreibung, während eine Schreibung mit Bindestrich zwischen den mittelbaren Konstituenten im Vergleich zur Zusammenschreibung schlechter bewertet wird.⁴⁸ Für die rechtsverzweigenden Komposita bestätigen

⁴⁷ Unter den Belegen von Buchmann (2015) finden sich zwar einige dreigliedrige Determinativkomposita, sie werden aber nicht als eigene Gruppe, sondern aufgrund unterschiedlicher Erstglieder in verschiedenen Gruppen diskutiert.

(i) Abgas-Grenzwert, Wein-Präsentkorb, Weihnachtsmarkt-Häuschen, Panorama-Glasschiebedach, BMX-Fahrrad, FlowTex-Betrugsfall, ADAC-Fachmann, BBC-Fernsehfilm

⁴⁸ Die Blickbewegungsstudie von Betram, Kuperman, Baayen & Hyöna (2011) konnte für dreiglied-

sich seine Annahmen. Bei den linksverzweigten werden zwar die Bindestriche zwischen den mittelbaren Konstituenten schlechter bewertet als die Zusammenschreibung, die Schreibungen mit Bindestrich zwischen den unmittelbaren Konstituenten aber nicht besser als die Zusammenschreibung (Geilfuß-Wolfgang 2013: 146). Diese Asymmetrie erklärt er dadurch, dass linksverzweigende dreiteilige Komposita im Deutschen sehr viel häufiger sind als rechtsverzweigende, also den unmarkierten Fall darstellen und daher ein Bindestrich an der Haupttrennfuge nicht zu einer besseren Bewertung führt (vgl. Geilfuß-Wolfgang 2013: 152).

3.3.3 Getrennschreibung

Eine Getrennschreibung morphologisch komplexer Substantive wie *linguisten mist* (SMS) ist für das Deutsche nicht nur außerhalb der Norm, sondern ein gravierender Verstoß gegen die Grundlagen des graphematischen Systems.

Wie schon unter 3.3.1 diskutiert, ist es die Funktion des Leerzeichens innerhalb des Schriftsystems, eben gerade anzuzeigen, dass es sich bei den Einheiten links und rechts davon um syntaktisch und nicht morphologisch verknüpfte Einheiten handelt. Ein Leerzeichen zwischen den Konstituenten eines morphologisch komplexen Wortes führt den Leser also potentiell in die Irre. Dennoch sind Getrennschreibungen morphologisch komplexer Wörter so häufig, dass sich nicht nur Sprachkritiker

rige links- oder rechtsverzweigende Komposita im Niederländischen und Finnischen einen Einfluss der Bindestrichsetzung auf die Gesamtlesezeiten zeigen. Niederländische und finnische Komposita mit drei Gliedern werden demnach langsamer verarbeitet (höhere Gesamtlesezeit), wenn sie einen Bindestrich zwischen den mittelbaren Konstituenten aufweisen, als wenn sie zusammengesrieben werden oder einen Bindestrich zwischen den unmittelbaren Konstituenten aufweisen. Betram et al. (2011: 541) führen dies darauf zurück, dass ein Bindestrich zwischen den mittelbaren Konstituenten zunächst zu einer Fehlinterpretation der morphologischen Struktur führt. Ein Bindestrich zwischen den unmittelbaren Konstituenten ist dagegen zweischneidig: Er erleichtert die Strukturanalyse, verstößt aber gegen die niederländische und finnische Schreibnorm. In den Experimenten wird das dadurch abgebildet, dass diese Bindestriche unabhängig von der Verzweigungsrichtung zu einer kürzeren Blickdauer (*subgaze duration*) auf der Konstituente links von der Haupttrennfuge im Vergleich zur Zusammenschreibung führen, was die Autoren als Zeichen dafür bewerten, dass die Struktur einfacher erkannt wird. Die späteren Fixationen und ihre Dauer sprechen dann aber dafür, dass der Verarbeitungsprozess aufgrund des unüblichen Bindestrichs ins Stocken gerät im Vergleich zu den Zusammenschreibungen. Vor diesem Hintergrund ist eine Übertragbarkeit der Ergebnisse aufs Deutsche kritisch zu hinterfragen.

beispielsweise unter *deppenleerzeichen.de* (zuletzt eingesehen am 30.05.2020) daran stören, sondern sie auch Gegenstand der linguistischen Forschung sind.

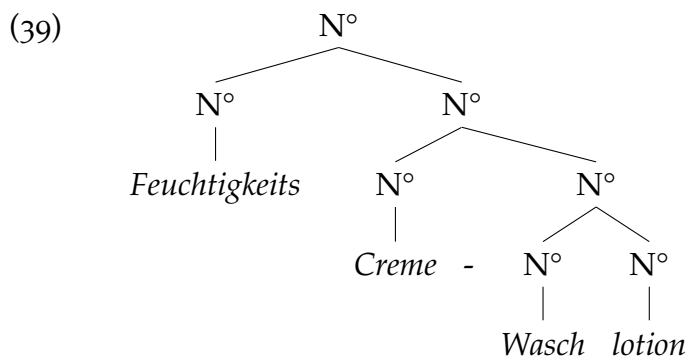
- (38)
- a. Der Apfel Birnen Kürbis Mann (Prestel-Bilderbuchtitel 1999)
 - b. Super Sommer Spar Menü (Plakatwerbung von McDonalds, August 1999)
 - c. Nuss Nougat Creme (Nutella 2006)
 - d. An der Gebäude Außenwand keine brennbaren Stoffe lagern (Warnschild am Bauhaus Mannheim 2006)
 - e. Feuchtigkeits Creme-Waschlotion (Tubenaufschrift Dove 2006)

Die Beispiele unter (38) und die dazugehörigen Quellen, die ein Auszug aus einer Zusammenstellung von Donalies (2011: 44) sind, vermitteln einen ersten Eindruck davon, in welchen Kontexten Getrenntschreibungen besonders häufig zu beobachten sind. All diesen Beispiele und auch allen weiteren von Donalies (2011) gelisteten ist gemein, dass sie als Aufschriften auf Produkten oder Schildern und als Titel plakativen Charakter haben. Nach Bredel (2008: 32-33) entsprechen diese Kontexte alle eher einer Schreibung im Listen- als im Textmodus. Der Listenmodus sei dadurch gekennzeichnet, dass die Zeile syntaktische Steuerungsfunktion übernimmt, so dass Satzzeichen fakultativ seien, der Punkt häufig ganz fehle und es auch für den Bindestrich zu abweichenden Verwendungen komme (inwiefern führt sie dabei leider nicht aus). Führt man den Gedanken von Bredel (2008) weiter, liegt es aber nahe, dass der Listenmodus sich nicht nur auf die Setzung von Interpunktionszeichen, sondern auch auf die Verwendung von Leerzeichen auswirken könnte. Wenn die Zeile als linear-suprasegmentaler Raum syntaktische Funktion übernimmt, befreit sie dadurch eventuell das Leerzeichen von seiner eigentlichen Aufgabe, die Grenze zwischen syntaktisch verknüpften Einheiten zu markieren, und macht so den Weg frei für einen Einsatz des Leerzeichen zur Grenzmarkierung zwischen morphologisch verknüpften Einheiten. Dementsprechend wäre zu erwarten, dass Getrenntschreibungen in Texten oder Textteilen im Listenmodus häufiger sind.

Darüber hinaus ist bei den Beispielen unter (38) zu beobachten, dass es sich um mindestens dreigliedrige Komposita handelt,⁴⁹ die nach dem, was unter 3.3.2 zu den

⁴⁹ Donalies (2011) führt auch zwei Belege mit nur zwei Kompositionsgliedern an: den Gedichttitel *Meeres Stille* und die Schildaufschrift *Reinigungs Annahme*.

Bindestrichschreibungen besprochen wurde, zumindest teilweise mit einem oder mehreren Bindestrichen zu erwarten gewesen wären, beispielsweise aufgrund der kopulativen Strukturen in *Apfel Birnen Kürbis Mann* oder *Nuss Nougat Creme* oder der orthographischen Empfehlung unübersichtliche Zusammensetzungen durch Bindestrichsetzung zu gliedern. Besonders irritierend ist das Leerzeichen trotz Fugenelement im Beispiel *Feuchtigkeits Creme-Waschlotion*. Nach der oben erläuterten Hypothese von Scherer (2012: 69) dient das Leerzeichen in Verbindung mit dem Bindestrich an der zweiten und der Zusammenschreibung an der dritten Kompositionsgrenze hier vielleicht der Anzeige der morphologischen Struktur. Die morphologische Struktur von *Feuchtigkeits Creme-Waschlotion* wäre demnach wie in (39) zu analysieren, da Scherer (2012) davon ausgeht, dass ein Leerzeichen als stärkste Grenzmarkierung die Haupttrennfuge zwischen den unmittelbaren Konstituenten markiert, während der Bindestrich eine hierarchisch darunter einzuordnende Grenze kennzeichnet, die aber wiederum höher einzustufen ist als diejenige, an der zusammengeschrieben wird.



Semantisch liegt damit die Interpretation nahe, dass es sich um eine besonders cremige Waschlotion handelt, die Feuchtigkeit spendet. Ohne das wortinterne Leerzeichen (*Feuchtigkeitscreme-Waschlotion*) oder bei einer durchgängigen Zusammenschreibung (*Feuchtigkeitscremewaschlotion*) läge es sicherlich näher, eine kopulative Verbindung der beiden lexikalisierten Komposita *Feuchtigkeitscreme* und *Waschlotion* anzunehmen. Das Leerzeichen blockiert also möglicherweise die semantische Interpretation, dass es sich um eine Waschlotion handelt, die gleichzeitig auch eine Feuchtigkeitscreme ist.

Diese kurze Betrachtung einiger zufälliger Beispiele soll nicht darüber hinwegtäuschen, dass eine Untersuchung von Getrennschreibungen komplexer Wörter, die

ebenso ausführlich und systematisch ist wie vorliegende Arbeiten zur Schreibung mit Bindestrich, bis heute meines Wissens nach leider nicht vorliegt, aber ein echtes Forschungsdesiderat ist.

Inhoff, Radach & Heller (2000) haben experimentell untersucht, welche Auswirkungen das Auslassen oder Setzen von Leerzeichen zwischen den Konstituenten morphologisch komplexer Wörter hat. Dazu haben sie dreigliedrige, deutsche Determinativkomposita mit und ohne interne Leerzeichen mit Hilfe von Benennungstests und Blickbewegungsstudien untersucht. Sie kommen dabei zu einem ganz ähnlichen Ergebnis wie Betram et al. (2011) für den Bindestrich zwischen den unmittelbaren Konstituenten dreigliedriger Komposita im Finnischen und Niederländischen: Auch Leerzeichen⁵⁰ sorgen zunächst für eine schnellere Benennung und eine kürzere initiale Fixation, führen aber zu Verzögerungen bei den Endfixationen (und danach gelagerten Fixationen bei den eingebetteten Stimuli). Diese Parallelität ist gerade deshalb spannend, weil der Bindestrich in den niederländischen und finnischen Komposita genauso wenig der orthografischen Norm entspricht wie die Setzung eines Leerzeichens innerhalb komplexer Wörter im Deutschen. Während Betram et al. (2011) das als potentielle Ursache für die Verzögerung der Bedeutungszuweisung annehmen, vermuten Inhoff et al. (2000: 46), dass die wortinternen Leerzeichen die Konstruktion der Gesamtbedeutung aus den Einzelbestandteilen erschweren.⁵¹

Obwohl dem deutschsprachigen Leser also durch das Leerzeichen in seiner Grundfunktion signalisiert wird, dass die Einheit Wort abgeschlossen ist und die nächste Einheit damit syntaktisch zu verknüpfen ist, sind komplexe Komposita mit Leerzeichen an den Kompositionsgrenzen also erstaunlich gut les- und interpretierbar.

⁵⁰ An dieser Stelle muss erwähnt werden, dass tatsächlich nur Leerzeichen eingesetzt wurden. Einer Zusammenschreibung wie *Datenschutzexperte* wurde also die Getrennschreibung *Daten schutz experte* gegenübergestellt, also wirklich eine bloße Getrennschreibung ohne eine Berücksichtigung der Substantivgroßschreibung bei den dann graphisch selbständigen Konstituenten *schutz* und *experte* (vgl. Inhoff et al. 2000: 25). Darüber hinaus wurde eine Schreibung mit Binnenmajuskeln (*DatenSchutzExperte*) als dritte Variante getestet. Ausführlicher dazu 3.3.4.

⁵¹ In den Items von Inhoff et al. (2000) wurden aber auch im Vergleich zur Zusammenschreibung immer beide Kompositionsgrenzen durch Leerzeichen oder Majuskeln markiert, während Betram et al. (2011) nur eine Grenze variieren.

3.3.4 Schreibungen mit Binnenmajuskel

Die Binnenmajuskel hat nicht zuletzt durch die Diskussion um gendergerechte Sprache an Popularität gewonnen. Dennoch sind Ausdrücke wie *SchülerInnen* nicht relevant im Sinne der Untersuchung. Da es ihre Aufgabe ist, statt eines Lexems gleich zwei zu kodieren (*Schüler* und *Schülerinnen*), ist nicht davon auszugehen, dass es sich um eine Variante an einer rein wortbildungsmorphologischen Grenze handelt, sondern die Majuskel der Hinweis auf das zweite Lexem und gleichzeitig dessen Initiale innerhalb des komplexen Ausdrucks ist. *SchülerInnen* ist demnach nicht als ein komplexes morphosyntaktisches Wort zu erfassen, sondern vielmehr die graphisch reduzierte Variante einer Koordination von *Schüler* und *Schülerinnen*.

Aber auch abseits des Binnen-*I* in gegenderten Formen finden sich ohne lange Suche wortinterne Majuskeln wie in den Beispielen unter (40).

(40) HP Laserjet, InterCityExpress, PrivatLeasing-Vertrag, CenterParcs

Die genannten Beispiele entstammen dem Aufsatz von Stein (1999: 261-262), der sich wohl als Erster um eine systematische Betrachtung der Binnenmajuskel im Deutschen bemüht hat. Schon im Zuge der Auflistung seiner zahlreichen Beispiele weist er darauf hin, dass Schreibungen mit Binnenmajuskel(n) zum Großteil Werbetexten entstammen und es sich häufig um Produkt- oder Firmennamen handelt. Die Domäne der Binnenmajuskel ist, wie schon beim Bindestrich, die Komposition, insbesondere die substantivische.

Aus sprachhistorischer Perspektive sind Binnenmajuskeln nichts Neues (ausführlicher dazu: Stein 1999: 263). Als Vorreiter für die verstärkte Ausbreitung einer Schreibung mit interner Majuskel vermutet Stein (1999: 264) das oben angesprochene Binnen-*I* bei Personenbezeichnungen.

Funktional vergleicht Stein (1999: 270) die Binnenmajuskel mit dem Bindestrich, geht also davon aus, dass die Binnenmajuskel wie der Bindestrich signalisiert, dass es sich bei der Einheit, die sie beinhaltet, um ein Wort handelt, und anzeigt, wo eine Konstituenten- oder Morphemgrenze liegt. Die Experimente von Inhoff et al. (2000) haben allerdings für dreigliedrige Komposita mit Binnenmajuskeln (*DatenSchutz-Experte*) keine signifikanten Effekte im Vergleich zur Zusammenschreibung (*Datenschutzexperte*) gezeigt, also kann zumindest bezweifelt werden, ob diese Grenzmar-

kierung Auswirkungen auf das Erkennen der Wortstruktur und die Bedeutungsweisung hat.

Unbestritten ist die Binnenmajuskel dagegen als Möglichkeit, von den üblichen graphischen Wortstrukturen abzuweichen. Stein (1999: 273) geht von einer intendierten Abweichung von der Normalform aus, die dadurch auch einen funktionalen Mehrwert im Vergleich zu dieser hat. Durch eine Binnenmajuskel werde der Erwartungshorizont des Rezipienten gesprengt und so Aufmerksamkeit erregt, was ihre Beliebtheit in der Werbung erklären könnte.

3.3.5 Schreibungen mit Apostroph

Aus der Gruppe der Apostrophierungen sind im Sinne dieser Untersuchung nur morphographische Apostrophe relevant, also Apostrophierungen, die einer Markierung der morphologischen Struktur auf Schriftebene dienen wie in den Beispielen unter (41).

- (41) a. Lisa's Lädchen
 b. Merkel'sche Politik
 c. Fußball'news

Im Vergleich zur phonographischen Funktion des Apostrophs als Auslassungszeichen sind diese Vorkommen deutlich seltener (vgl. dazu die korpuslinguistische Untersuchung von Scherer (2010) zu 's.). Stammformapostrophe wie in *Merkel'sche Politik* und die Apostrophierung der Genitivendung -s kommen mit einer gewissen Regelmäßigkeit vor und sind durch die orthographische Norm (§ 97E) gedeckt:

Von dem Apostroph als Auslassungszeichen zu unterscheiden ist der gelegentliche Gebrauch dieses Zeichens zur Verdeutlichung der Grundform eines Personennamens vor der Genitivendung -s oder vor dem Adjektivsuffix -sch

(Rat für deutsche Rechtschreibung 2018: 99)

Für diese Untersuchung sind allerdings beide Fälle nicht unmittelbar von Bedeutung, weil hier der Fokus nicht nur auf wortbildungsmorphologischen Grenzen, sondern auch auf dem Bereich der substantivischen Wortbildung liegt.⁵²

⁵² Scherer (2013: 98-102) diskutiert auch apostrophierte Adverbien wie *dienstag's*, *halbttag's* und *link's*.

So bleibt also nur der Apostroph zwischen Kompositionsgliedern wie in *Fußball'news*, ein Phänomen, das Klein (2002: 180), auf den auch dieses Beispiel zurückgeht, das erste Mal beschreibt. Scherer (2013: 95) verweist aber schon darauf, dass sich selbst in Onlineforen, die sich explizit der Sammlung von unorthografischen Apostrophierungen widmen, kaum vergleichbare Beispiele finden, während Beispiele für einen sogenannten fugenmarkierenden Apostroph wie in *Bauer'n Hof* und *Bahnhof's Gaststätte* etwas häufiger sind. In diesen Fällen ist die Kompositionsgrenze allerdings durch ein Leerzeichen markiert, während der Apostroph das Fugenelement vom Erstglied abgrenzt, so dass auch diese Fälle nur als Getrennschreibung – unabhängig davon, ob diese auf die vorausgehende Apostrophierung zurückzuführen sein mag – für die vorliegende Untersuchung relevant wären.

Dementsprechend ist davon auszugehen, dass auch die angestrebte Korpusuntersuchung kaum Belege für Schreibungen mit einem Apostroph an Wortbildungsgrenzen liefern wird.

4. Methodik

Eine umfassende Analyse von Phänomenen des Sprachgebrauchs setzt eine Erhebung relevanter linguistischer Daten voraus, die auf die zu untersuchenden Hypothesen abgestimmt ist. In der Linguistik haben sich neben der Beobachtung vor allem die Arbeit mit Korpora und experimentellen Designs als empirische Methoden etabliert. Um verschiedene Schreibvarianten an Morphemgrenzen in komplexen Wörtern untersuchen zu können, ist eine Vielzahl schriftlich fixierter Belege erforderlich. Wenn die Belegmenge darüber hinaus, wie anvisiert, noch ausreichen soll, um auch Aussagen über die Varianz in Abhängigkeit vom Wortbildungstyp des komplexen Wortes zu machen, muss sie entsprechend noch größer sein. Durch den Einbezug des Kontextkriteriums, also der Untersuchung der Textformspezifik der Varianten, ist die Korpusrecherche die beste Wahl.

Die deutschsprachige Korpuslandschaft ist mittlerweile soweit gediehen, dass nicht nur Zeitungskorpora oder Sammlungen literarischer Texte verfügbar sind, sondern eine große Bandbreite von Spezialkorpora zu den verschiedensten Textformen. Der technische Fortschritt hat es zudem möglich gemacht, dass Korpora heute Größen erreichen, aus denen auch bei komplexeren Fragestellungen tausende Belege extrahiert werden können. Diese Belege sind nicht nur als realsprachliche Daten unverfälscht im Sinne der Untersuchung, sondern ermöglichen es im Sinne der Untersuchung auch – anders als beispielsweise gezielte Sammlungen besonders auffälliger Schreibvarianten – die Relevanz der Phänomene und ihr Verhältnis zueinander realistisch einzuschätzen.

Die auf die Formulierung der Forschungshypothesen (4.1) folgenden Teilkapitel dokumentieren im Sinne eines transparenten Vorgehens die Zusammenstellung des Korpusmaterials zu den verschiedenen Textformen in 4.2, das Vorgehen bei der Recherche nach wortbildungsmorphologisch komplexen Substantiven (und Adjektiven) unter 4.3 und die Entstehung der systematischen Belegsammlung inklusive der

Annotationskategorien in 4.4.

4.1 Forschungshypothesen

Auf Grundlage der bisherigen Beobachtungen zu den wortinternen Schreibvarianten werden nun die Hypothesen formuliert, die in der anstehenden Korpusuntersuchung überprüft werden sollen.

Die Grundannahme ist zunächst, dass wortinterne Schreibvarianten in komplexen Wörtern keinesfalls willkürlich sind, sondern systematisch. Das mag im ersten Moment redundant erscheinen, da die Schreibung komplexer Wörter im Deutschen durch die Regeln zur Rechtschreibung einer starken Normierung unterliegt, aber einerseits ist die Norm nicht immer systematisch und erfasst andererseits auch nicht alle besprochenen Varianten.

Unabhängig davon, ob es sich um normgerechte oder nicht normgerechte Schreibungen handelt, sollten sich also zumindest 90% der Schreibungen an wortbildungsmorphologischen Grenzen in komplexen Wörtern in einem optimalitätstheoretischen Modell als optimale Kandidaten erfassen lassen.

4.1.1 Einfluss der Wortbildungsart und Lexikalisierung

Die Ausführungen zu Schreibungen mit Bindestrich, Getrenntschreibungen und Binnenmajuskeln haben gezeigt, dass insbesondere Komposita zugänglich für wortinterne Varianten an morphologischen Grenzen abseits der Zusammenschreibung sind. Daher wird angenommen, dass an Kompositionsgrenzen im Vergleich zu Derivationsgrenzen nicht nur mehr Belege zu finden sind, die von einer Zusammenschreibung abweichen, sondern auch die Bandbreite der belegten Varianten größer ist.

In diesem Zusammenhang soll auch der Einfluss der Lexikalisierung auf die Ausprägung von Schreibvarianten überprüft werden. Wie oben schon besprochen, konnte Pfeiffer (2002) in seinen Experimenten zeigen, dass ein Bindestrich in selten belegten Komposita einen positiven Effekt auf den Leseprozess hat, während er in lexikalisierten Belegen zu Verzögerungen führt. Lexikalisierte Belege sollten entsprechend

seltener Varianten jenseits der Zusammenschreibung aufweisen als Belege, die nicht als lexikalisiert gelten können.

Denkbar ist auch eine Interaktion zwischen Wortbildungsart und Lexikalisierungsstatus, da im Deutschen die Substantive und dort insbesondere die Determinativkomposita das produktivste Feld der Wortbildung sind (vgl. Fleischer & Barz 2012: 117-118), so dass hier ein signifikanter Anteil nicht lexikalisierter Belege zu erwarten ist. Ich vermute also für den Bereich der Komposition einen geringeren Lexikalisierungsanteil (und damit eine höhere Neigung zu Varianten abseits der Zusammenschreibung) als für den Bereich der Derivation.

4.1.2 Einfluss der Bestandteile

Neben einem Einfluss von Wortbildungsart und Lexikalisierung haben insbesondere die Ausführungen zu den Bindestrichschreibungen gezeigt, dass einzelne Wortbestandteile einen Einfluss darauf haben, ob und wie ihre Grenzen zu anderen Bestandteilen eines Wortes markiert werden.

In der Tradition von Bredel (2008), die von einer Markierung beim Aufeinandertreffen von zwei heterogenen Einheiten ausgeht, soll untersucht werden, welchen Einfluss fremdsprachige Einheiten, Eigennamen und graphisch markierte Einheiten haben. Dabei nehme ich an, dass sie allesamt die Markierung durch eine Abweichung von der Zusammenschreibung an Wortbildungsgrenzen befördern und graphisch markierte Einheiten nach Buchmann (2015) eine Zusammenschreibung sogar ausschließen.

Das Vorkommen eines Fugenelementes an einer Kompositionsgrenze zwischen zwei Bestandteilen eines Wortes lässt dagegen eher eine Zusammenschreibung erwarten, weil es nach Bredel (2008: 115) durch eine Markierung (die Setzung eines Bindestrichs) *„bei einer isolierten lexikalischen Interpretation nicht ausgewertet werden“* kann.

4.1.3 Einfluss der Textform

Der Einfluss der Textform hat sich oben insbesondere bei den Schreibungen mit Binnenmajuskel und den Getrenntschreibungen gezeigt, die beide jeweils im Werbekontext oder auf Beschilderungen im weitesten Sinne prominent sind. Daher ist

anzunehmen, dass die Häufigkeit und Bandbreite der wortinternen Schreibvarianten an wortbildungsmorphologischen Grenzen in Abhängigkeit von der Textform variiert.

In Kapitel 4.2 werden im Zuge der Vorstellung der Teilkorpora jeweils spezifische Erwartungen hinsichtlich der wortinternen Schreibung an wortbildungsmorphologischen Grenzen für jede der Textformen formuliert. Grob lassen sich diese Erwartungen aber abseits der Werbeanzeigen in zwei Gruppen einteilen: In den Zeitungstexten, den Texten der Fachzeitschriften, den Kriminalromanen, Plenarprotokollen und Lexikoneinträgen ist von einer standardnahen Schreibung auszugehen. Das heißt, dass neben Zusammenschreibungen nur Bindestrichschreibungen als alternative Varianten zu erwarten sind. Ein breiteres Spektrum mit zusätzlichen Getrennschreibungen, Binnenmajuskeln und andere Varianten ist abseits des Werbeanzeigen-Korpus dagegen eher in den bisweilen auch stärker nächsprachlich und quasi-synchron geprägten Kommunikationsformen (Mail, SMS und Diskussionsbeiträge) zu erwarten. Der wesentliche Unterschied zwischen beiden Gruppen ist dabei die textorientierte bzw. interaktionsorientierte Schreibhaltung nach Storrer (2014). Während beim textorientierten Schreiben der Text als Produkt des Prozesses im Vordergrund steht wie z. B. beim Verfassen dieser Dissertation liegt der Fokus beim interaktionsorientierten Schreiben auf der Interaktion mit dem Kommunikationspartner bzw. den Kommunikationspartnern, so dass Korrektur- und Überarbeitungsprozesse zu Gunsten einer schnellen Reaktion und erfolgreichen Kommunikation vernachlässigt werden (Storrer 2014: 172). Die Werbeanzeigen fallen bei dieser Einteilung also aus dem Rahmen, weil sie einen besonderen Variantenreichtum und eine Häufung von Schreibungen abseits der Zusammenschreibung erwarten lassen, obwohl sie das Ergebnis einer textorientierten Schreibhaltung sind.

Ein allgemeines optimalitätstheoretisches Modell für die Schreibung an wortinternen, wortbildungsmorphologischen Grenzen im Deutschen ist also für einzelne Textformen wahrscheinlich weniger aussagekräftig, womöglich ist sogar eine gruppenspezifische oder textformspezifische Modellierung nötig.

4.1.4 Sonstige Faktoren

Neben den allgemeinen Einflüssen von Wortbildungsart, Textform und der Natur der Bestandteile sollen noch zwei weitere potentielle Faktoren untersucht werden:

die Hierarchie der Wortbildungsstruktur bei Komposita mit mehr als zwei Kompositionsgliedern und ein Effekt von markierten Buchstabenfolgen an der Wortbildungsgrenze.

Die orthographische Norm (§ 45) erlaubt einen Bindestrich beim Aufeinandertreffen dreier gleicher Buchstaben. Ob dieser tatsächlich regelmäßig gesetzt wird und darüber hinaus auch das Auftreten zweier gleicher Buchstaben wie in *Tauffeier* oder eine Kombination von zwei Vokalbuchstaben (*Kloampel*, *Miniaffe*) eine Grenzmarkierung begünstigt, soll hier ebenfalls untersucht werden.

Zudem stellt sich Buchmann (2015: 250) angesichts der Bindestrichschreibungen von komplexeren Komposita (*Weihnachtsmarkt-Häuschen*) die Frage, ob die Wortlänge oder Buchstabenanzahl ein Kriterium für die Setzung eines Bindestrichs ist. Auch das soll hier empirisch überprüft werden.

Ich vermute allerdings, dass die bloße Buchstabenanzahl kein entscheidender Faktor ist, sondern gehe mit Geilfuß-Wolfgang (2013) nicht nur davon aus, dass es bei Komposita mit drei (oder mehr Gliedern) eine Tendenz gibt, die Haupttrennfuge graphisch zu markieren, um die Strukturanalyse zu erleichtern, sondern nehme auch mit Scherer (2012) an, dass in dieser Kategorie von Wortbildungen die Konstituentenstruktur durch eine Rangfolge der Stärke der Grenzmarkierung verdeutlicht werden kann, die Fugen vorausgegangener Kompositionsschritte also schwächer markiert werden als die terminaler Schritte.

4.2 Korpora

Auch wenn die Begriffe Korpus und Textsammlung häufig synonym verwendet werden, ist nicht jede Textsammlung automatisch auch ein Korpus. Ein Korpus zeichnet sich gegenüber einer zufälligen Textsammlung vor allem dadurch aus, dass es planmäßig konzipiert wurde. Wichtig sind nicht nur eine auf den Untersuchungsgegenstand abgestimmte Größe und die Beständigkeit des Korpus, die eine Überprüfung der Ergebnisse möglich macht, sondern vor allem die Repräsentativität im Sinne der Untersuchung (vgl. Scherer 2006: 5). Diese Repräsentativität wird im Idealfall dadurch erreicht, dass Korpora individuell für die jeweilige Untersuchung zusammengestellt werden und eine zufällige Stichprobe der Grundgesamtheit darstellen. Die

Grundgesamtheit im Sinne dieser Untersuchung sind schriftsprachliche Äußerungen im Gegenwartsdeutschen. Da diese Grundgesamtheit als Ganzes nicht erfasst, bzw. nicht erfassbar ist, und dementsprechend nicht einfach eine Zufallsstichprobe gezogen werden kann, ist die beste Option eine Annäherung über ein möglichst breit aufgestelltes Korpus. Dabei ist es von Vorteil, dass komplexe Substantive, die hier untersucht werden, im Deutschen in eigentlich allen Texten zu erwarten sind (vgl. Fleischer & Barz 2012: 117-118). Das Untersuchungskorpus muss also nicht extrem groß sein, um eine hinreichende Menge an komplexen Substantiven zu erfassen, sollte sich aber nicht nur auf eine einzelne Textform beschränken.

Vielmehr ist ein Korpus erforderlich, das eine größere Menge an Textformen umfasst, die wiederum jeweils möglichst spezifisch sind in ihren Merkmalen, um auf dieser Basis einerseits allgemeingültige, textformübergreifende Aussagen zur wortinternen Schreibvariation im Deutschen zu treffen, und andererseits Unterschiede zwischen den einzelnen Textformen herauszuarbeiten und im Idealfall an spezifischen Eigenarten der Textform festzumachen. Da der Aufbau von Korpora sehr zeitintensiv ist, wurde, soweit möglich, auf bereits bestehende Korpora zurückgegriffen.

Grundlage dieser Untersuchung ist daher ein Korpus, das gut drei Millionen laufende Wortformen aufgeteilt auf neun verschiedenen Teilkorpora umfasst. Sechs dieser Teilkorpora basieren auf dem Deutschen Referenzkorpus (DeReKo) des Instituts für deutsche Sprache in Mannheim (Institut für Deutsche Sprache 2013). Mit Hilfe der zugehörigen Korpusmanagement- und Recherchesoftware COSMAS II wurden aus dem über 24 Milliarden Textwörter umfassenden DeReKo (Kupietz & Lungen 2014: 2379) sechs virtuelle Korpora im willkürlich festgelegten Umfang von etwa 500.000 laufenden Textwörtern zusammengestellt:¹ ein Korpus mit Zeitungstexten unterschiedlicher Tageszeitungen, eines mit Texten zweier fachspezifischer Zeitschriften, eines auf der Basis von Kriminalromanen, eines bestehend aus Plenarprotokollen des Bundesrates, ein weiteres aus Lexikoneinträgen einer Online-Enzyklopädie und schließlich ein Korpus aus den Diskussionsbeiträgen zu solchen Lexikoneinträgen.

Um möglichst viele Textformen untersuchen zu können, werden diese Korpo-

¹ Ein Abweichen der Größe um +/- 10% wird toleriert, so dass sich alle tatsächlichen Korpusgrößen zwischen 450.000 und 550.000 Textwörtern bewegen.

Tabelle 4.1.: Korpusübersicht

Korpus	Teilkorpora	Quelle	Alter	Textwörter
Werbeanzeigen	W	selbst erstellt	2014	22.699
Mail	Mail	Schnitzer (2012)	2011	23.150
SMS	SMS	Schnitzer (2012)	2011	26.080
Zeitung	BRZ09, HMP09, HAZ09, M09, NUZ09, RHZ09	DeReKo	2009	509.312
Fachzeitschriften	SPK,VDI09	DeReKo	2009-2010	502.085
Kriminalromane	DIV	DeReKo	2011	514.675
Plenarprotokolle	PBR	DeReKo	2011-2012	516.795
Lexikoneinträge	WPD11	DeReKo	2003-2011	453.161
Diskussionsbeiträge	WDD11	DeReKo	2003-2011	484.908
Gesamt				3.052.865

ra durch drei weitere, bestehend aus privaten SMS, privaten E-Mails bzw. Printwerbeanzeigen aus Zeitschriften ergänzt. Diese drei von Einzelpersonen erstellten Korpora umfassen jeweils nur 25.000 laufende Textwörter, die Untersuchungen haben aber gezeigt, das auch dieser reduzierte Umfang repräsentative Ergebnisse hervorbringt.² SMS- und E-Mail-Korpus werden zur Verfügung gestellt von Caroline-Victoria Schnitzer, die im Rahmen ihrer Dissertation *Linguistische Aspekte der Kommunikation in den neueren elektronischen Medien, SMS – E-Mail – Facebook* per Fragebogen gleich mehrere Korpora für diese Kommunikationsformen zusammengestellt hat. Mit ihrem Einverständnis werden hier das SMS-Korpus 2011 und das E-Mail-Korpus verwendet.³

Da keine entsprechenden Korpora aus dem Werbebereich zur Verfügung stehen, gerade dort aber eine große Bandbreite und überdurchschnittliche Häufigkeit von wortinternen Schreibvariationen zu erwarten ist, wurde eigens für diese Untersuchung ein Korpus aus Printwerbeanzeigen zusammengestellt, dessen Größe sich an SMS- und E-Mail-Korpus orientiert.

² Vgl. dazu 4.3 *Recherche*.

³ Zur Erhebungsmethode und internen Struktur der Korpora vgl. Schnitzer (2012).

Da die wortinternen Schreibvarianten synchron untersucht werden, sind alle Korpora möglichst aktuell gehalten. Zielsetzung war eine Datierung nach 2008, die – abgesehen von den wikipediabasierten Korpora aufgrund ihrer ungewöhnlichen Textgenese – auch eingehalten werden konnte.⁴

Im Folgenden werden die Teilkorpora jeweils kurz vorgestellt. Dabei erfolgt ausdrücklich keine systematische textlinguistische Betrachtung der jeweiligen Textformen, sondern lediglich eine Zusammenstellung von Informationen zu Textgrundlage und Umfang, eine kurze Charakterisierung der Textform anhand exemplarischer Ausschnitte sowie Vorhersagen bezüglich der zu erwartenden wortinternen Varianten.

Eine Übersicht über das Gesamtkorpus und seine Teile bietet Tabelle 4.1.

4.2.1 Werbeanzeigen-Korpus

Textgrundlage und Umfang Dieses Korpus aus Print-Werbeanzeigen wurde eigens für die Untersuchung zusammengestellt. Die Anzeigen entstammen 11 Publikumszeitschriften (Jahrgang 2014), jeweils in zwei Ausgaben (Ende März und Ende Juni) aus dem „Lesezirkel-Geschäftswelt-Paket“, das darauf abgestimmt ist, geschlechts- und altersübergreifend alle Zielgruppen zu erfassen (Lesezirkel 2014). Bei den zugrunde liegenden Zeitschriften handelt es sich im einzelnen um:

- auto motor und sport (2014, 6 & 13)
- BUNTE (2014, 10 & 24)
- FOCUS (2014, 9 & 24)
- FürSie (2014, 10 & 14)
- Gala (2014, 10 & 25)
- manager magazin(2014, 3 & 6)
- Men’sHealth (2014, 3 & 6)
- petra (2014, 4 & 7)
- SCHÖNER WOHNEN (2014, 3 & 6)
- DER SPIEGEL (2014, 10 & 21)
- stern (2014, 10 & 24)

⁴ Die Wikipedia-Lexikoneinträge und zugehörigen Diskussionen sind zwar auf dem Stand von 2011. Textteile einzelner Lexikoneinträge sind aber bereits 2003 entstanden und wurden seitdem über die Jahre ergänzt, überarbeitet und eben diskutiert.

Aus diesen Zeitschriften wurden alle deutschsprachigen Werbeanzeigen für Produkte erfasst, die sich den vier Bereichen Mobilität (Auto, Motorrad, Bahn, Airline), Finanzdienstleistungen (Banken, Versicherungen, Geldanlage), Kosmetik (Pflegeprodukte, Make-Up) und Arzneimittel (Apothekenvertrieb) zuordnen lassen. Diese Gruppen finden sich mit unterschiedlicher Schwerpunktsetzungen in fast allen Zeitschriften und Ausgaben. Nach diesem System habe ich 199 Print-Werbeanzeigen erfasst.

Da die automatische Texterkennung bei Werbeanzeigen schnell an ihre Grenzen stößt, mussten große Teile des Textaufkommens nicht nur nachbearbeitet, sondern manuell digitalisiert werden. Dabei habe ich Logos, Produktaufschriften und Gütesiegel bewusst ausgelassen bzw. durch Platzhalter ([LOGO], [PRODUKT]) ersetzt, weil davon auszugehen ist, dass es sich dabei jeweils um eine eigene Textsorte mit spezifischen Beschränkungen handelt. Bildliche Gestaltungsmittel oder die Beziehung von Text und Bild in den Werbeanzeigen sind zudem kein Gegenstand dieser Arbeit.

Aus den erfassten Anzeigen resultiert ein textsortenspezifisches Werbeanzeigen-Korpus im Umfang von 22.690 Textwörtern.

Charakteristik der Textform Als Textsorte sind die Werbeanzeigen dadurch charakterisiert, dass allen Texten dieselbe appellierende kommunikative Funktion gemein ist. Thematisch behandeln sie in der Regel ein Produkt, beispielsweise ein Auto wie in (1) und (2) und nutzen dabei typische sprachliche Merkmale wie Marken- und Produktnamen oder wertende Begriffe.

Auch aus textlinguistischer Sicht handelt es sich bei den Textanteilen in Werbeanzeigen aus dem Printbereich nicht unbedingt um prototypische Texte. Werbeanzeigen sind beispielsweise nicht immer linear und arbeiten mitunter sehr stark mit der Beziehung zwischen Text und Bild. Adamzik (2012: 123) verweist zudem darauf, dass sich Werbetexte auch untereinander stark unterscheiden können. So zeigen die beiden Anzeigenbeispiele aus dem Korpus unter (1) und (2), dass es auch deutliche Unterschiede zwischen einzelnen hier erfassten Anzeigen gibt, obwohl schon eine Verengung zu Gunsten von Werbeanzeigen in Printmedien stattgefunden hat und es sich in beiden Fällen um Produktwerbung für ein Auto handelt.

(1) *Hochdruckgebiet.*

Das neue Audi A3 Cabriolet.

Jetzt entdecken auf www.audi.de/a3cabriolet [BILD]

Kraftstoffverbrauch in l/100 km: innerorts 7,8-5,1; außerorts 5,0-3,7; kombiniert 6,0-4,2; CO₂-Emissionen in g/km: kombiniert 140-110.

Audi

Vorsprung durch Technik

(2) *HONDA The Power of Dreams*

GRÖSSE ZEIGEN. DER NEUE CIVIC TOURER.

Im neuen Civic Tourer wächst der Civic über sich hinaus: mit starken und gleichzeitig sparsamen Motoren, modernen Fahrerassistenzsystemen sowie bis zu 1.668 Liter Gepäckvolumen, dem größten seiner Klasse. Dynamisch, intelligent und unendlich vielseitig! Die besten Voraussetzungen, um selbst aktiv zu werden und Größe zu zeigen.

Fahren Sie den neuen Civic Tourer Probe – bei Ihrem Honda Händler!

Mehr Infos unter www.civic-tourer.de

Kraftstoffverbrauch in l/100 km: innerorts 8,9-4,2; außerorts 5,5-2,6; kombiniert 6,6-3,8; CO₂-Emission in g/km: 155-99. (Alle Werte gemessen nach 1999/94/EG.) Abbildung zeigt Sonderausstattung.

Während (1) mit einem stichpunktartigen Minimum an Text auskommt, umfasst (2) auch mehrere ausformulierte Sätze. Gemein ist beiden Anzeigen die Angabe zum Kraftstoffverbrauch, die für Autoanzeigen so vorgeschrieben ist und damit zum spezifischen Muster dieses Typs gehört.⁵

Der Vergleich der Anzeigen zeigt, dass es Werbeanzeigen wie (1) gibt, die komplett im Listenmodus nach Bredel (2008) verfasst sind, während andere, wie Beispiel (2), auch Passagen im Textmodus aufweisen.

Eine Abweichungen von der Zusammenschreibung findet sich allerdings mit *Civic Tourer* auch im Textmodus-Anteil des Beispiels (2). Dabei handelt es sich aber um einen Produktnamen, die wiederum äußerst zugänglich für Varianten sind.⁶

⁵ Zu den typischen Bausteinen von Werbeanzeigen vgl. Janich (2013: 53-75).

⁶ Hierzu sei auf die Ausführungen von Stein (1999: 266-267) im Rahmen seiner Analyse von Binnenmajuskeln verwiesen.

Erwartungen hinsichtlich der wortinternen Schreibung Schon bei der Diskussion der wortinternen Varianten unter 3.3 ist deutlich geworden, dass die Werbesprache besonders variantenreich ist. Neben einer Hervorhebung von Wortbestandteilen oder eine Segmentierung durch einen Bindestrich, Leerzeichen oder eine Binnenmajuskel, werden in dieser Funktion in Werbetexten auch verstärkt typographische Mittel wie Schriftfarbe, -größe oder -stärke genutzt (vgl. Ewald 2012: 4). Ewald (2012: 5-8) argumentiert dabei dafür, dass gerade durch von der (orthographischen) Norm abweichende Schreibungen gezielt die Aufmerksamkeit des Rezipienten erregt wird. Von einer unbeabsichtigten Schreibvariation im Rahmen von Werbeanzeigen, die das Ergebnis redaktioneller Arbeit sind und damit auch den Korrekturprozessen einer textorientierten Schreibhaltung unterliegen, ist dagegen nicht auszugehen.

4.2.2 Mail-Korpus

Textgrundlage und Umfang Das Mail-Korpus geht zurück auf ein von Caroline-Victoria Schnitzer im Rahmen ihrer Dissertation *Linguistische Aspekte der Kommunikation in den neueren elektronischen Medien, SMS – E-Mail – Facebook* erstelltes Korpus mit E-Mails aus dem Bereich der privaten Kommunikation. Von den ursprünglich 501 E-Mails, die Schnitzer 2011 per Fragebogen erfasst und nach Geschlecht, Alter und Bildungsabschluss gruppiert hat (vgl. Schnitzer 2012: 18-23), wurden 15 ausgeschlossen, da sie komplett in einer Fremdsprache verfasst und damit für diese Untersuchung nicht relevant sind. Die verbleibenden 486 deutschsprachigen E-Mails konstituieren ein Korpus im Umfang von 23.150 Textwörtern.

Charakteristik der Textform Anders als das Werbeanzeigen-Korpus, das nur eine einzelne Textsorte umfasst, repräsentiert das Mail-Korpus eine Kommunikationsform. Diese Kommunikationsform ist nach Dürscheid (2005: 16) lediglich ein situativ medialer Rahmen, in dem sich verschiedene kommunikative Handlungsmuster realisieren lassen. Im Fall der Kommunikation per Mail bedingt dieser Rahmen ein technisches Endgerät (Computer, Laptop, Tablet, Smartphone). Das Schreiben erfolgt in der Regel über eine Tastatur, die allerdings je nach Endgerät variieren kann zwischen einer analogen, externen Tastatur bis hin zur digitalen Einblendung auf dem Bildschirm von Smartphone oder Tablet. Das verwendete Mailprogramm oder die gewählten Benutzereinstellungen des Endgerätes können dabei einen Einfluss auf das Geschriebene haben durch unterschiedliche Worterkennungs- oder Korrek-

turprogramme.⁷

Thema und Funktion sind dagegen prinzipiell frei, bei der Korpuszusammenstellung durch Schnitzer (2012) aber auf den privaten Bereich beschränkt. Die folgenden Beispiele unter (3) bis (5) zeigen nicht nur, dass die Länge der Mails stark variieren kann, sondern machen auch deutlich, dass Kommunikationsformen wie Textsorten gewisse Muster erkennen lassen. Ein Beispiel hierfür sind die Anrede- und Grußformel in (5) und die sehr reduzierte Version davon (LG) in (3). Während die Beispiele (4) und (5) eint, dass ihr Fokus funktional ein Kontakterhalt ist, ist (3) lediglich eine Nachfrage als Ausschnitt eines wahrscheinlich längeren Diskurses. Damit wird auch deutlich, dass die Kommunikation per Mail im privaten Bereich zumeist dialogisch erfolgt.

(3) *Wie hast du des gelesen?*

LG

(4) *hast Du meine letzte mail bekommen? Ich hatte Dir noch vom So. nach der Filmvorführung geschrieben. Was macht die Wohnungssuche? Im letzten Kreisboten waren zwei Zwei-Zi. Wohnungen in X [Majuskeln, Abkürzung] drin mit ca. 47 qm, eine Dachwohnung und eine mit Garten und Terrasse, beide mit TG und nicht teuer, aber eben klein.*

(5) *Hi X, ja irgendwie dachte ich, ich hätte als Hausfrau ganz viel Zeit.*

Aber so ein Vormittag ist ja schnell rum und dank des schönen Wetters bin ich auch ziemlich viel mit dem Radl unterwegs. Ich habe dann auch noch einen kleinen Ackerranteil in Y, wo es immer was zu Tun gibt (z. B. zur Zeit viele Gurken und Salat anpflanzen...). Ja, und die Freitage sind natürlich auch dank unseres Jour Fix gut ausgebucht. Übrigens, falls Du diese Woche Zeit hast, wir treffen uns um 8.00 Uhr im Z [Majuskeln].

Am Freitag Nachmittag fahren wir dann Richtung W, wo wir in den Pfingstferien die Insel erkunden. Also wie gesagt, die Zeit vergeht wie im Flug.

Viele liebe Grüße, vielleicht sehen wir uns am Freitag.

A

⁷ Zur Frage, wie die Nutzung einer Tastatur das Schreiben offline und online beeinflusst, sei auf Dürscheid & Brommer (2009) verwiesen.

Erwartungen hinsichtlich der wortinternen Schreibung Solche getippten Dialoge in den neuen Medien sind häufig Gegenstand sprachkritischer Betrachtungen und Äußerungen (vgl. Dürscheid & Brommer 2009: 8-14), die hier aber nur bedingt relevant sind. Der entspanntere Umgang mit der orthographischen Norm könnte allerdings dazu führen, dass mehr oder ungewöhnlichere Varianten an wortbildungsmorphologischen Grenzen zu beobachten sind.

Falls dem so ist, können diese Varianten als Folge einer interaktionsorientierten Schreibhaltung gelten, die nicht den Text als Schreibprodukt fokussiert (textorientierte Schreibhaltung), sondern die laufende Interaktion, die unter Umständen mehr von schnellen Reaktionen als von der Elaboriertheit des Textes profitiert (vgl. Storrer 2014: 172).

4.2.3 SMS-Korpus

Textgrundlage und Umfang Auch das SMS-Korpus basiert auf einem von Schnitzer (2012) erstellten Korpus, parallel zum Mail-Korpus.⁸ Im Sinne der Untersuchung wurden hier ebenfalls Nachrichten aussortiert, die ausschließlich in einer Fremdsprache verfasst sind oder lediglich Ziffern und/oder Sonderzeichen enthalten. Von den 2020 erhobenen Nachrichten verbleiben damit 1973 SMS mit einem Gesamtumfang von 26.080 laufenden Wortformen.

Charakteristik der Textform Auch das SMS-Korpus entspricht damit einer Kommunikationsform. Der situativ mediale Rahmen ist dadurch gekennzeichnet, dass ein Handy oder Smartphone als Endgerät genutzt wird und die zugelassene Zeichenmenge pro Text eingeschränkt ist. Entsprechend kurz sind die einzelnen Texte wie die Beispiele unter (6) bis (7) zeigen. Thematisch und funktional besteht auch hier dagegen Freiheit. Eine Anrede oder Grußformel wie in (7) kann vorkommen, ist aber nicht obligatorisch. Typisch für SMS ist unter anderem die Verwendung von Emoticons wie :) in (8). *spiel* in (8) illustriert dagegen, dass die Groß- und Kleinschreibung im SMS-Korpus nicht so strikt eingehalten wird wie in anderen standardnahen Textformen wie z. B. Zeitungstexten (vgl. Schnitzer 2012: 149-158).

(6) *Bin noch auf Arbeit*

(7) *Servus X. Wir sind dabei. Bis Samstag*

⁸ Einzelheiten zu Datensammlung und zum Korpusaufbau vgl. (Schnitzer 2012: 8-18).

- (8) *Nicht wirklich, bin echt platt :) war ganz schön anstrengend. Fahren auch jetzt erst los und ich wollte noch das spiel schauen. Kommst vorbei?*

Ein vergleichender Blick auf das Mail-Beispiel unter (3) macht deutlich, dass es zwischen Mail und SMS im Bereich der privaten Kommunikation einen fließenden Übergang gibt und (3) ebenso gut in den situativ medialen Rahmen der SMS gepasst hätte.

Stand 2020 haben längst Instant-Messenger wie *WhatsApp* im privaten Bereich wohl nicht nur der SMS, sondern durchaus auch der E-Mail den Rang abgelassen. Schnitzer (2012: 196-203) weist aber bereits in einem Exkurs darauf hin, dass sich die Kommunikation per SMS und *WhatsApp* hinsichtlich sprachlicher Besonderheiten nicht wesentlich unterscheidet.

Erwartungen hinsichtlich der wortinternen Schreibung Die Gemeinsamkeiten von Mail und SMS als dialogische, interaktionsorientierte Kommunikationsform unter dem Einfluss von (verschiedenen) Textverarbeitungsprogrammen lassen also auch Gemeinsamkeiten hinsichtlich der Schreibung an wortbildungsmorphologischen Grenzen erwarten. Allerdings zeigt sich in den Beispiel-SMS, aus denen einzig *Sams-tag* relevant im Sinne der Untersuchung ist, schon, dass wohl wenige Belege für wortbildungsmorphologisch komplexe Substantive im SMS-Teilkorpus zu erwarten sind.

4.2.4 Zeitungs-Korpus

Textgrundlage und Umfang Beim Zeitungs-Korpus handelt es sich wie bei allen nun folgenden Korpora um ein virtuelles Korpus, das auf dem DeReKo basiert und mit Hilfe von COSMAS II zusammengestellt und durchsucht worden ist. Es beschränkt sich auf den Jahrgang 2009, der die gewünschte Aktualität wahrt und auch im mit dem Tree-Tagger morpho-syntaktisch annotierten T-Archiv verfügbar ist, so dass sich zusätzliche Recherchemöglichkeiten im Vergleich zu ungetaggten Korpora ergeben. Im Sinne der Repräsentativität für deutschsprachige Zeitungstexte dienen sechs verschiedene Regionalzeitungen als Grundlage.

Um die Zielgröße von etwa 500.000 laufenden Wortformen zu erreichen, habe ich die Korpusgröße mit Hilfe von COSMAS II auf eine Zufallsstichprobe von 5% des eigentlichen Gesamtumfangs reduziert. Da der Textumfang der Zeitungsjahrgän-

Tabelle 4.2.: Aufbau Zeitungs-Korpus

Zeitung	Kürzel	Textanteil
<i>Braunschweiger Zeitung</i>	(BRZ09)	25%
<i>Hannoversche Allgemeine</i>	(HAZ09)	9,2%
<i>Mannheimer Morgen</i>	(M09)	14,4%
<i>Hamburger Morgenpost</i>	(HMP09)	7%
<i>Nürnberger Zeitung</i>	(NUZ09)	6,2%
<i>Rhein-Zeitung</i>	(RHZ09)	38,2%

ge stark variiert (zwischen gut 4 Millionen Textwörtern in der Rhein-Zeitung und gut 600.000 in der Hamburger Morgenpost) ergibt sich auch im reduzierten Korpus ein Ungleichgewicht bezüglich der Textanteile. Daher wird in Tabelle 4.2 neben dem Kürzel der jeweiligen Quelle auch der prozentuale Textanteil am Zeitungs-Teilkorpus angegeben.

Im Zuge einer weiteren pseudorandomisierten Stichprobe mit dem Ziel, einen Datenausschnitt im Umfang von etwa 5.000 Textwörtern aus den Teilkorpora zu erheben, wurde allerdings auf ein ausgewogenes Verhältnis hinsichtlich des Textwortanteils geachtet, so dass alle sechs Regionalzeitungen mit gut 800 Textwörtern vertreten sind. Diese kleinste Stichprobengröße liefert die Datengrundlage, die es ermöglicht, die Schreibvarianten an wortbildungsmorphologischen Grenzen hinsichtlich ihrer Häufigkeit ins Verhältnis zueinander zu setzen. Vgl. dazu auch das Kapitel 4.3 zur *Recherche*.

Charakteristik der Textform Da das Korpus nicht nur einzelne Artikel sondern ganze Ausgaben der Zeitungen umfasst, handelt es sich wie bei den Kommunikationsformen Mail und SMS auch hier nicht um ein textsortenspezifisches Korpus. Das Korpus entspricht vielmehr der Publikationsform der Zeitung, die verschiedene Textsorten (Artikel, Kommentare, Rezensionen usw.) umfasst. Thematisch gibt es dementsprechend kaum Einschränkungen, höchsten einen mehr oder weniger ausgeprägten regionalen Fokus und auch funktional gibt es eine Durchmischung (informieren, appellieren, unterhalten), gemein ist aber allen Texten die Art der Veröffentlichung in der Zeitung.

Der Ausschnitt aus der *Braunschweiger Zeitung* in (9) zeigt einen prototypischen Absatz (nur Fließtext) eines Berichtes über das lokale Geschehen (hier zu den Pla-

nungen bezüglich eines Radweges). Schon dieses kurze Beispiel lässt erkennen, dass in den Zeitungstexten im Gegensatz zu den SMS-Nachrichten eine Vielzahl morphologisch komplexer Substantive zu finden ist. Allein die erste Zeile liefert mit *Ringgleis*, *Anschluss* und *Gartenstadt* gleich drei entsprechende Belege.

- (9) *Es hakt zurzeit auf dem Ringgleis: Der Anschluss der Gartenstadt verzögert sich, und am Werksteig an der Celler Straße nehmen Radfahrer lieber Umwege. Doch die Verwaltung beteuert auf Anfrage: Geduld, alles werde sich richten. Festes Ziel in diesem Jahr ist: Braunschweigs schönster und beliebtester Fuß- und Radweg soll weiter wachsen. So sollen noch in diesem Jahr die letzten Arbeiten für den Wegebau des rund 1,4 Kilometer langen Teilstücks zwischen Arndtstraße und Harzstieg abgeschlossen werden. Dann wird der 2001 begonnene Ringgleisweg zwischen Oker und Harzstieg insgesamt 5,7 Kilometer lang sein. „Das Ringgleis sei immerhin in diesem Jahr vom Bundesministerium für Verkehr, Bau und Stadtentwicklung als Beispiel für gelungene Quartiersaufwertung ausgezeichnet worden“, sagte Stadtbaurat Wolfgang Zwafelink. Das Ringgleis belebe ein von Strukturwandel betroffenes Stadtgebiet und schaffe gleichzeitig einen zentrumsnahen attraktiven „Grünen Ring“ als Wegeverbindung und Freizeitanlage für alle Bürger der Stadt. Als Grünanlage und schnelle Radwege-Verbindung habe das Ringgleis auch eine ökologische Funktion.*

Erwartungen hinsichtlich der wortinternen Schreibung Wie schon bei den Werbeanzeigen handelt es sich auch hier um redaktionelle Texte. Im Gegensatz zum Mail- und SMS-Teilkorpus (und den Diskussionsbeiträgen) ist also von einer textorientierten Schreibhaltung auszugehen, die Überarbeitungs- und Korrekturprozesse umfasst und sich entsprechend stärker an Sprachnormen orientiert.⁹

Zeitungstexte sind in der Korpuslinguistik sehr prominent und gelten als die Textform, die am ehesten der Standardschriftsprache entspricht. Dementsprechend sind in den Zeitungstexten kaum wortinterne Varianten abseits der Zusammenschreibung und der Bindestrichschreibung zu erwarten. Tatsächlich findet sich im Ausschnitt unter (9) nur eine Bindestrichvariante (*Radwege-Verbindung*), die nach § 45 der Norm als orthographisch korrekt gilt.

⁹ Vgl. dazu auch die Ausführungen von Buchmann (2015: 58-59) zum Zeitungskorpus, das ihrer Untersuchung zugrunde liegt.

4.2.5 Fachzeitschriften-Korpus

Textgrundlage und Umfang Wie das Zeitungs-Korpus basiert auch das Fachzeitschriften-Korpus auf dem T-Archiv. Textgrundlage sind der Jahrgang 2010 von *Spektrum* (SPK), einem Wissenschaftsmagazin mit naturwissenschaftlichem Schwerpunkt, und die August- bis Dezember-Ausgaben der *VDI nachrichten* (VDI09), der Wochenzeitung des Vereins Deutscher Ingenieure, aus dem Jahr 2009, die eher berufsbezogene und technische Themen behandelt.

Um die Zielgröße von 500.000 Textwörtern zu erreichen, wurde dieses Korpus auf eine Größe von 56% reduziert. Ähnlich wie beim Zeitungs-Teilkorpus ergibt sich auch hier ein Ungleichgewicht bezüglich des Textanteiles. Der gesamte Jahrgang 2010 von *Spektrum* hat nur gut 30% Textanteil, während knapp 70% auf die August- bis Dezember-Ausgaben der *VDI nachrichten* entfallen. Wie beim Zeitungs-Teilkorpus wurde aber auch hier für eine pseudorandomisierte, ausbalancierte Stichprobe hinsichtlich des Anteils an Textwörtern beider Fachzeitschriften im Umfang von 5.000 Textwörtern gezogen. Vgl. dazu auch das Kapitel 4.3 zur *Recherche*.

Charakteristik der Textform Wie beim Zeitungs-Korpus handelt es sich auch hier um ein Korpus, das einer Publikationsform, den Fachzeitschriften, entspricht und ganz ähnliche Textsorten umfasst wie die Zeitungstexte. Thematisch ist der Fokus aber etwas eingeschränkter als in den Tageszeitungen, und zwar auf die jeweiligen naturwissenschaftlichen bzw. beruflichen Schwerpunkte, wobei Innovationen auf den jeweiligen Gebieten im Vordergrund stehen. Die Textausschnitte aus den Fachzeitschriften *Spektrum* (10) und *VDI nachrichten* (11) zeigen wie zu erwarten Berichte, die unter der jeweiligen Überschrift als Fließtexte verfasst sind und die thematischen Schwerpunkte widerspiegeln.

(10) *Pluto ändert sein Aussehen*

*Neu ausgewertete Bilder des Weltraumteleskops Hubble ergaben jetzt die bislang detailreichste Karte des Pluto. Die Aufnahmen zeigen helle braune und dunkle Flächen sowie weiße Areale auf dem Zwergplaneten. Über deren geologische Struktur ist bislang noch nichts bekannt. Allerdings belegen Spektren, dass es auf Pluto Ablagerungen von Wassereis, Kohlendioxid und Kohlenmonoxid gibt und er zudem von einer äußerst dünnen Atmosphäre aus Stickstoff und Methan umgeben ist.

 <!-- Bild 1 --> Vergleicht man die neuen Karten mit den gröbereren Aufnahmen, die im Jahr*

1994 vom Weltraumteleskop Hubble aufgenommen wurden, so zeigt sich, dass Pluto derzeit sehr viel rötler erscheint als damals, berichtet ein Forscherteam um Marc Buie vom Southwest Research Institute in Boulder im US-Bundesstaat Colorado. Sie führen die rötliche Farbe auf organische Stoffe zurück, die durch die Aufspaltung von Methan durch die ultraviolette Strahlung der Sonne entstehen.

(11) *Trennung: GE und Fanuc haben Joint Venture beendet*

Am 11. Dezember haben GE und Fanuc in Chicago offiziell ihr Joint Venture beendet. GE Intelligent Platforms wird die Entwicklung von Unternehmenssoftware und intelligenten Plattformen vorantreiben, während sich Fanuc Inc., Fujo/Japan, auf die CNC-Technik konzentriert. Bei der Antriebssteuerung werde weiter zusammengearbeitet.

Laut Roelcke (2020: 112-113) ist für fachsprachliche Texte¹⁰ nicht nur eine größere Zahl, der eh schon sehr produktiven Komposita zu erwarten, sondern es ist auch davon auszugehen, dass die Zahl ihrer Kompositionsglieder im Vergleich zur Standardsprache höher ist. Beides sei darauf zurückzuführen, dass sich die Fachsprache um Deutlichkeit und Spezifität in ihren Ausdrücken bemüht. So finden sich schon in den kurzen Textbeispielen komplexe Komposita wie *Weltraumteleskop* und *Unternehmenssoftware*.

Erwartungen hinsichtlich der wortinternen Schreibung Dieses Bemühen und die Menge auch komplexerer Kompositionen können ein begünstigender Faktor für Abweichungen von der Zusammenschreibung sein, beispielsweise zur Anzeige der Wortstruktur.

Schon der Titel der Fachzeitschrift *VDI nachrichten* ist der erste Beleg für eine Getrennschreibung im Fachsprachekontext, sogar kombiniert mit der Kleinschreibung eines dadurch graphisch selbständigen Substantivs *nachrichten*. Die Recherche und Analyse wird zeigen, ob es sich dabei um einen Vorboten für weitere (ungewöhnliche) Varianten an wortbildungsmorphologischen Grenzen handelt.

¹⁰ Für eine ausführliche Diskussion des Begriffes *Fachsprache* und der Besonderheiten dieses Sprachregisters sei auf Roelcke (2020) verwiesen. Hier findet der Begriff Anwendung, weil sich das Korpusmaterial aus zwei Zeitschriften zusammensetzt, die einen thematischen Schwerpunkt in den Bereichen Natur- bzw. Ingenieurwissenschaft haben und sich damit an ein begrenztes Publikum wenden, das diesen Bereichen angehört oder sich dafür interessiert.

4.2.6 Kriminalroman-Korpus

Textgrundlage und Umfang Auch das Kriminalroman-Korpus basiert auf den Textsammlungen des IDS. Allerdings wurde hier auf das ungetaggte Archiv W zurückgegriffen.¹¹ Aus dem DIV-Korpus, das Belletristik des 20. und 21. Jahrhunderts beinhaltet, wurden zufällig acht vollständige, 2011 erschienene Kriminalromane von sieben verschiedenen AutorInnen ausgewählt und zu einem benutzerdefinierten Korpus zusammengefügt. Bei den Kriminalromanen handelt es sich im Einzelnen um:

- Gude, Christian: *Binärcode*
- Graf, Edi: *Leopardenjagd*
- Imbsweiler, Marcus: *Schlussakt*
- Keiser, Gabriele: *Gartenschläfer*
- Lascaux, Paul: *Wursthimmel*
- Pundt, Hardy: *Deichbruch*
- Schneider, Harald: *Ernteopfer* und *Schwarzkittel*¹²

Der Umfang der Romane variiert zwischen den beiden Extremen von etwa 35.000 und knapp 100.000 Textwörtern, während das breite Mittelfeld zwischen 55.000 und 70.000 Textwörter umfasst. Wie schon bei den Zeitungs- und Fachtexten wurde bei der kleinsten Stichprobe daraus, im Umfang von 5.000 Textwörtern, auf ein ausgewogenes Verhältnis geachtet.

Charakteristik der Textform Durch die Beschränkung auf Kriminalromane handelt es sich hier wie bei den Werbeanzeigen um ein textsortenspezifisches Korpus. Es handelt sich um literarische Texte, die sich thematisch mit einem oder mehreren Kriminalfällen befassen und der Unterhaltung dienen. Sie sind das Ergebnis eines textorientierten Schreibprozesses und als lange Fließtexte in Kapitel gegliedert.

¹¹ Leider ist dieses und auch alle folgenden auf dem DeReKo basierenden Korpora nicht in einer morphosyntaktisch getagkten Version verfügbar. Dementsprechend entfällt bei der Recherche mit COSMAS II die Möglichkeit einer Einschränkung auf Substantive und Adjektive, was zu einer größeren Menge an falsch positiven Treffern führt. Vgl. dazu auch 4.3.1.

¹² Anders als beim Zeitungs- und Fachzeitschriften-Korpus ist nicht unmittelbar nachvollziehbar, aus welchem Kriminalroman ein Beleg stammt, da für alle Werke im Belletristik-Korpus beim Ergebnisexport das Kürzel *DIV* als Quellenangabe verwendet wird.

- (12) *«Und was gab's bei der PK?», fragte Babs und ignorierte die nervige Tonfolge ihres Telefons, die nach dem fünften Signal auf den Apparat im Sekretariat umgeleitet wurde. «Noch nicht viel Neues. Ich hab ein paar O-Töne für die Zwölf-Uhr-Sendung. Das Wesentliche hatten wir heute früh in den News. Was spannend war, ist die Vermutung der Polizei, dass die Frau, die mit ihrem Fahrzeug zwischen Bebenhausen und Tübingen verunglückt ist – Mona Rieger heißt sie, glaube ich – am Tatort gewesen sein könnte. Man muss noch ein paar Spuren auswerten und Bodenproben an ihren Schuhen untersuchen, aber der Pressesprecher sprach von einem interessanten Zusammenhang. Außerdem ist der Zeugin, die die Leiche gefunden hat, auf dem Weg zum Tatort eine Fußgängerin begegnet, deren Beschreibung auf Mona Rieger passt. Klingt ziemlich spannend, nicht?» Babs wollte gerade etwas erwidern, als erneut das Telefon, diesmal auf Lindas Apparat, klingelte.*

Der Textausschnitt aus einem der Kriminalromane unter (12) umfasst zwar nur einen kurzen Dialog, durch Lexeme wie *Polizei*, *Tatort*, *Spuren* oder *Zeugin* ist aber der thematische Rahmen eines Kriminalfalles bereits deutlich.

Erwartungen hinsichtlich der wortinternen Schreibung Was die sprachliche Gestaltung anbelangt, gibt es kaum Hinweise darauf, verstärkte Abweichungen von der Standardsprache anzunehmen. *PK* und *News* sind wohl dem umgangssprachlichen Ton des Dialogs geschuldet. Diese dialogische Form ist ein verbindendes Element zu den schon besprochenen Kommunikationsformen Mail und SMS. Kriminalromane sind aber anders als diese interaktionsorientierten Formen ohne Zweifel das Ergebnis einer textorientierten Schreibhaltung (wie auch die bereits besprochenen Werbeanzeigen, Zeitungs- und Fachzeitschriftentexte), dementsprechend ist davon auszugehen, dass sich die Variation an wortbildungsmorphologischen Grenzen auch hier auf die Zusammenschreibung und Bindestrichschreibungen beschränken wird.

4.2.7 Plenarprotokoll-Korpus

Textgrundlage und Umfang Ebenfalls textsortenspezifisch ist das Plenarprotokoll-Korpus. Für die Zielgröße von ca. 500.000 Textwörtern wurden die aktuellsten Plenarprotokolle aus dem PBR-Korpus des Archivs W des IDS mit Hilfe von COSMAS II zu einem benutzerdefinierten Korpus zusammengestellt. Das Protokollkorpus umfasst damit 23 vollständige Plenarprotokolle des

Bundesrates aus den Jahren 2011 und 2012.

Charakteristik der Textform Bei diesen Plenarprotokollen handelt es sich um eine Textsorte als hochspezifiziertes Muster der Kommunikation. Im Sinne einer Dokumentation und Archivierungen wird dabei die eigentlich mündliche, kommunikative Interaktion des Parlamentes eins zu eins in eine Schriftform gebracht, eben protokolliert. Beispiel (13) zeigt, dass dabei nicht nur der Inhalt erfasst wird, sondern auch eine Sprecherzuordnung erfolgt.

(13) Präsidentin Hannelore Kraft

Vielen Dank, Herr Ministerpräsident Beck!

Nächste Wortmeldung: Parlamentarischer Staatssekretär Dr. Kues (Bundesministerium für Familie, Senioren, Frauen und Jugend).

Dr. Hermann Kues, Parl. Staatssekretär bei der Bundesministerin für Familie, Senioren, Frauen und Jugend: Frau Präsidentin! Sehr geehrte Damen und Herren! In der Tat geht mit der Aussetzung der Wehrpflicht zur Mitte des Jahres und damit auch des Zivildienstes, [...], eine Ära zu Ende.

(14) Weitere Wortmeldungen liegen nicht vor.

Wir kommen zur Abstimmung. Hierzu liegen Ihnen die Empfehlungen des Wirtschaftsausschusses vor, den Vermittlungsausschuss aus zwei Gründen anzurufen. Wer ist allgemein für die Anrufung des Vermittlungsausschusses? Handzeichen bitte! – Das ist eine Minderheit.

Damit hat der Bundesrat den Vermittlungsausschuss nicht angerufen.

Auch die institutionalisierte Struktur der sprachlichen Handlungen des Parlamentes spiegelt sich in den Protokollen wider, beispielsweise der Abstimmungsverlauf in (14). Zwischen solchen Abschnitten gibt es längere Passagen einzelner Sprecher, die jeweiligen Wortmeldungen zu einem Thema, mit argumentativem Charakter.

Erwartungen hinsichtlich der wortinternen Schreibung Da die Protokolle ein fester Bestandteil des politischen Prozesses sind und einer breiten Öffentlichkeit zur Verfügung gestellt werden, ist davon auszugehen, dass die Sprachstandards, beispielsweise die orthographische Norm, sehr genau eingehalten werden und ent-

sprechend nur wenige, orthographisch korrekte Bindestrichschreibungen als Abweichungen von der Zusammenschreibung an wortinternen morphologischen Grenzen zu beobachten sind.

Diese Schreibungen sind dann aber aufgrund der Besonderheiten der Protokoll-Situation nicht auf den Urheber der sprachlichen Äußerung zurückzuführen, sondern eine Art graphische Interpretation der eigentlich nur lautlichen Äußerung durch den Schreiber des Protokolls.

4.2.8 Lexikoneinträge-Korpus

Textgrundlage und Umfang Das Korpus aus Online-Lexikoneinträgen basiert wie das Korpus mit den dazugehörigen Diskussionsbeiträgen auf der Wikipedia-Plattform. Die Zielgröße von 500.000 laufenden Textformen ist hier aufgrund der immensen Größe der Korpora WPD₁₁ (ca. 575.000.000 Textwörter) und WDD₁₁ (ca. 246.000.000 Textwörter) in einem zweistufigen Verfahren erreicht worden. Mit dem Ziel, den inhaltlichen Zusammenhang zwischen Lexikonartikeln und Diskussionsbeiträgen durch eine reine Zufallsstichprobe nicht ganz aufzugeben, sind zunächst aus beiden Korpora parallel gut 60 Dokumente (Untereinheiten der Textorganisation im Korpus) ausgewählt worden, aus denen erst im zweiten Schritt eine Zufallsstichprobe gezogen wurde.

Im Fall des Lexikoneinträge-Korpus ist so über den Zwischenschritt einer manuellen Auswahl anhand der alphabetisch geordneten Dokumente im Umfang von gut 21.000.000 Textwörtern und deren zufällige Reduzierung auf nur noch 2% des Umfangs das virtuelle Korpus entstanden.

Charakteristik der Textform Anders als beim thematisch dadurch eng verwandten Diskussionsbeiträge-Korpus, das eine Kommunikationsform repräsentiert, handelt es sich beim Lexikoneinträge-Korpus wie beim zuvor charakterisierten Protokollkorpus um ein textsortenspezifisches Korpus. Jeder einzelne Artikel innerhalb des Online-Lexikons beschäftigt sich mit einem Einzelthema und folgt in der Regel einer einheitlichen Struktur von Lemma, über Inhaltsverzeichnis, den entsprechend gegliederten Artikelkern, sowie die Verweise auf Quellen und verwendete Literatur, die sich auch im Ausschnitt unter (15) zeigt. Funktional steht die möglichst neutrale Informationsübermittlung im Fokus.

- (15) *Georg Decker (*7. Dezember 1818 in Pest; †13. Februar 1894 in Wien) war ein österreichischer Aquarellmaler und Lithograf.*

Leben

Georg Decker war der Sohn des Malers Johann Stephan Decker und Bruder der Maler Albert Decker und Gabriel Decker. Er kam 1821 zusammen mit seiner Familie nach Wien und war zunächst Schüler seines Vaters. Anfang der 1840er Jahre kann er dann als Student der Wiener Akademie nachgewiesen werden. Um 1860 leitete er eine private Malschule und war ab 1861 Mitglied des Künstlerhauses. Der Plan des Wiener Hofes, Decker eine Schule für Pastellmalerei einzurichten, kam nicht zustande.

Georg Decker war von 1851-1860 in erster Ehe mit Ottilie von Sobek verheiratet, nach deren Tod heiratete er 1861 Josefine Helene von Lucam. Er wurde nach seinem Tode auf dem Wiener Zentralfriedhof bestattet. 1909 wurde die Deckergasse in Wien-Meidling dem Künstler zu Ehren benannt.

Zu den Unterschieden zwischen Artikeln in gedruckten Enzyklopädien und *Wikipedia* verweise ich auf Fandrych & Thurmair (2011: 104-108), die auch Lexikonartikel allgemein als Textsorte im Hinblick auf ihre spezifischen Muster ausführlich diskutieren. Hier sei nur erwähnt, dass in Online-Lexikonartikeln mangels der Notwendigkeit sich kurz zu fassen, weniger Abkürzungen und keine so ausgeprägte nominale Verdichtung wie in gedruckten Enzyklopädien zu erwarten ist.

Erwartungen hinsichtlich der wortinternen Schreibung Grundsätzlich ist trotz des Online-Formates auch im Wikipedia-Lexikon eine standardnahe Sprache zu erwarten, da es sich an eine breite Öffentlichkeit richtet und die Artikel das Ergebnis eines speziellen redaktionellen Prozesses sind, der es prinzipiell jedem interessierten Leser, egal ob Spezialist oder Laie hinsichtlich des betreffenden Themas, erlaubt einen Artikel zu bearbeiten.¹³

Dementsprechend ist auch hier davon auszugehen, dass sich die wortinternen Varianten hauptsächlich auf die Zusammenschreibung und Bindestrichschreibung begrenzen.

¹³ Vgl. dazu auch Storrer (2014: 181-183).

4.2.9 Diskussionsbeiträge-Korpus

Textgrundlage und Umfang Auch das Korpus mit den Diskussionsbeiträgen basiert auf der Wikipedia-Plattform. Wie schon unter 4.2.8 beschrieben, ist das Diskussionsbeiträge-Korpus parallel zum Lexikoneinträge-Korpus entstanden. Durch die Dokumentenauswahl hat hier eine erste Reduktion auf 8,5 Millionen Textwörter stattgefunden, von denen das virtuelle Korpus dann noch 5% umfasst. Eine enge Abstimmung der Lexikonartikel und Diskussionsbeiträge, so dass zum Artikel im Lexikon-Korpus die Diskussionsbeiträge, die den Entstehungs- und Überarbeitungsprozess des jeweiligen Artikels dokumentieren, ins Diskussionsbeiträge-Korpus aufgenommen werden, war leider nicht möglich. Die parallele Auswahl der alphabetisch sortierten Dokumente für die Lexikonartikel- und Diskussionsbeiträge-Korpora verhindert aber zumindest eine inhaltlich gänzliche Entkopplung.

Charakteristik der Textform Anders als das textsortenspezifische Lexikoneinträge-Korpus entspricht das Diskussionsbeiträge-Korpus einer Kommunikationsform, gemein ist beiden aber die thematische Grundlage.

Die Kommunikationsform des Diskussionsforums weist dabei einige Gemeinsamkeiten mit der Kommunikation per Mail oder SMS auf. In allen drei Fällen ist von einer interaktionsorientierten Schreibhaltung im Zusammenhang mit der dialogischen Form auszugehen. Sie unterscheidet sich dagegen trotz der thematischen Überschneidung stark von den Lexikoneinträgen der Wikipedia-Enzyklopädie. Storrer (2014: 181-185) präsentiert empirische Befunde zu einem Vergleich der Artikelseiten (Lexikoneinträge) und Diskussionsseiten des jeweils selben Lemmas und diagnostiziert zwei verschiedene Schreibwelten. Während die Lexikoneinträge auf den Artikelseiten monologisch und sachlich neutral formuliert den Schriftnormen entsprechen, ergibt sich für die Diskussionsbeiträge eher ein Bild, das mit der Kommunikation per SMS und Mail zu vergleichen ist.

Das bestätigen auch die Beispiele (16) und (17) als Ausschnitte der Diskussionsseite zu *Elizabeth Bowes-Lyon (Queen Mum)*. In (16) findet unter dem Thread *Tod* ein kurzer (in Bezug auf die Beitragslänge) Wortwechsel statt, der auch in ein SMS-Format gepasst hätte. Die Angaben zu Verfasser, Datum und Uhrzeit, die wie die Aufteilung in Threads zum typischen Muster von Diskussionsforen gehören, machen aber deutlich, dass sich dieser Dialog zeitlich über fast ein Jahr erstreckt hat.

Der Diskussionsbeitrag (17) ist schon etwas umfangreicher und weist sogar eine

Grußformel auf. An dieser wird deutlich, dass sich die Diskussionsbeiträge anders als die meisten Mails und SMS nicht an eine Einzelperson, sondern häufig an die ganze Gruppe von Mitdiskutierenden richten.

(16) *Tod*

Woran starb Queen Mum eigentlich? War es die im Artikel erwähnte Infektion der Bronchien?—Alexmagnus 00:38, 11. Jun. 2009 (CEST)

Schon mal was von Altersschwäche gehört? Sie war über 100! —195.212.29.187 13:01, 25. Feb. 2010 (CET)

Nicht jeder Tod im hohen Alter ist auf die Altersschwäche zurückzuführen.Zumal sie auf ihrem 101. Geburtstag alles andere als todkrank aussah.—Alexmagnus 19:28, 16. Apr. 2010 (CEST)

(17) *Aufbarung und Kirche*

Hallo miteinander, die Queen Mum wurde nach Ihrem Tod in einer Kirche aufgebahrt, damit die englische Bevölkerung die Möglichkeit hatte, Ihr den letzten Gruß zu erteilen. Es gab auch eine Ehrenwache des britischen Militärs am Sarg der Queen Mum in jener Kirche. Das war eine sehr große und schöne Kirche und ich dachte das diese Kirche zu Queen Mum verlinkt wurde.Leider steht im Artikel über Queen Mum nichts über den Namen dieser Kirche drin. Kennt jemand den Namen dieser Kirche und kann der in den Artikel über Queen Mum mit aufgenommen werden? —217.234.119.129 02:37, 10. Apr 2005 (CEST)

Erwartungen hinsichtlich der wortinternen Schreibung Daneben offenbart (17) einige Normabweichungen bezüglich Rechtschreibung und Zeichensetzung (*ich dachte das [...] wurde.Leider*). Für eine ausführlichere Betrachtung dieser und weiterer Besonderheiten sei auf Fandrych & Thurmair (2011: 136-152) verwiesen.

Aufgrund der Parallelen zu den anderen beiden Kommunikationsformen stellt sich also die Frage, ob sich Mail-, SMS- und Diskussionsbeiträge-Korpus auch hinsichtlich der Variation an wortinternen morphologischen Grenzen insofern ähneln, dass durch die interaktionsorientierte Schreibhaltung mehr oder ungewöhnlichere wortinterne Varianten zu beobachten sind.

4.3 Recherche

Die Korpusrecherche sollte im Sinne der Untersuchung eine repräsentative Menge von wortbildungsmorphologisch komplexen Substantiven (und Adjektiven) liefern. Um dieses Ziel zu erreichen, wurde in 2.1.1 schon definiert, was unter morphologisch komplexen Wörtern zu verstehen ist und warum sich diese Untersuchung auf wortbildungsmorphologisch komplexe Substantive beschränkt. Jetzt stellt sich die Frage, wie Belege für morphologisch komplexe Substantive systematisch aus den vorgestellten Korpora extrahiert werden.

Leider gibt es keine Korpora, die in Bezug auf morphologische Komplexität annotiert sind, so dass man gezielt danach suchen könnte. Daher habe ich drei verschiedene Verfahren genutzt und die Belege einerseits maschinell über die Schreibvarianten und die Länge der Zeichenkette erfasst und andererseits auch manuell recherchiert.

Recht unproblematisch war dabei die Suche nach Belegen, die Interpunktionszeichen wie den Bindestrich oder Apostroph enthalten, die sich auch in großen Korpora in Kombination mit Platzhaltern gut maschinell suchen und finden lassen (Vgl. dazu 4.3.1.), ähnliches gilt auch für Binnenmajuskeln (Vgl. dazu 4.3.2.). Schwieriger wurde es schon bei unauffälligen Zusammenschreibungen (Vgl. dazu 4.3.3.) und wirklich problematisch bei Getrennschreibungen (Vgl. dazu 4.3.2.).

Dementsprechend habe ich nicht nur verschiedene Suchanfragen formuliert, sondern auch im Sinne der Machbarkeit mit zwei kleineren Stichprobengrößen (25.000 und 5.000 Textwörter) gearbeitet, die dann jeweils als Grundlage für einen Teil der Suchanfragen bzw. eine händische Recherche dienten. Einen Überblick über die verwendeten Stichprobengrößen liefert Tabelle 4.3. Da besonders wortbildungsmorphologisch komplexe Substantive im Deutschen sehr verbreitet sind, liefern auch schon sehr kleine Stichproben im Umfang von nur etwa 5.000 Textwörtern eine Vielzahl von Belegen. Abhängig von der Textform kommen auf 5.000 Textwörter trotz der Beschränkung auf Substantive (und Adjektive) bis zu 800 Treffer, was fast jedem 6. Wort entspricht.

Tabelle 4.3.: Stichproben

Textform	Virtuelle Korpora	Stichprobe 25.000	Stichprobe 5000
Werbeanzeigen	–	22.699	4.753
Mail	–	23.150	5.219
SMS	–	26.080	5.351
Zeitung	509.312	24.907	5.067
Fachzeitschriften	502.085	23.966	4.795
Kriminalromane	514.675	24.891	5.237
Plenarprotokolle	516.795	24.534	5.094
Lexikoneinträge	453.161	23.550	5.057
Diskussionsbeiträge	484.908	23.708	5.015
Gesamt (Textwörter)	3.052.865	217.485	45.588

4.3.1 Virtuelle Korpora

Mittels COSMAS II sind die virtuellen Korpora nach Belegen mit Wortzeichen durchsucht worden. Mit Hilfe von Platzhaltern und unter der Berücksichtigung der verschiedenen Codierungsmöglichkeiten habe ich Treffer mit Apostrophierungen und Bindestrichen extrahiert. Um die Treffermenge im Falle der Bindestriche in einem zu bewältigenden Rahmen zu halten und falsch positive Treffer zu reduzieren, habe ich nicht nach den Wortzeichen allein, sondern nach der Kombination des Zeichens mit einem darauffolgenden Buchstaben gesucht, ganz konkret die Kombinationen *'a, 'e, 'i, 'n, 'r, 's* und *-a, -e, -i, -n, -r, -s* ohne Berücksichtigung der Groß- oder Kleinschreibung.¹⁴ Für das Zeitungs- und das Fachzeitschriften-Korpus wäre aufgrund der morphosyntaktischen Annotation zusätzlich eine Einschränkung der Suche auf Substantive und Adjektive möglich gewesen, die zu deutlich weniger falsch positiven Treffern geführt hätte. Um eine Verzerrung der Ergebnisse im Vergleich zu den anderen Teilkorpora aufgrund von Annotationsfehlern zu vermeiden, wurde aber darauf verzichtet. Tabelle 4.4 bietet einen Überblick über die Ergebnisse der Suchanfragen in den virtuellen Korpora.

¹⁴ Bei den sechs verwendeten Buchstaben handelt es sich nach Beutelspacher (2009: 10) um diejenigen, die im Deutschen unabhängig von der Position im Wort am häufigsten vorkommen.

Tabelle 4.4.: Treffermenge virtuelle Korpora

Textform	Bindestrich	Apostroph
Zeitung	2.326	77
Fachzeitschriften	2.343	33
Kriminalromane	446	0
Plenarprotokolle	809	1
Lexikoneinträge	2.538	366
Diskussionsbeiträge	1.615	254
Suchanfragentreffer	10.553	760

4.3.2 Stichproben 25.000 und kleine Korpora

Eine Stichprobengröße von 25.000 laufenden Textwörtern bietet sich unmittelbar an, da so den kleineren Korpora (Mail, SMS, und Werbeanzeigen) vom Umfang her vergleichbare Entsprechungen der DeReKo-basierten Korpora gegenüber gestellt werden können. Anders als bei den virtuellen Korpora wurde bei den Stichproben im Umfang von 25.000 und 5.000 Textwörtern darauf geachtet, dass die Zeitungs-, Fachzeitschriften- und Kriminalroman-Stichproben ausbalanciert sind hinsichtlich des Anteils der beteiligten Medien. So macht jede der sechs Tageszeitungen des Zeitungs-Korpus etwa ein Sechstel der Textwörter der 25.000er- und 5.000er-Stichproben aus. Das Verhältnis von *VDI nachrichten* und *Spektrum* ist hälftig und die acht Kriminalromane tragen jeweils etwa ein Achtel der Textwörter zum Umfang der Stichprobe bei. Dementsprechend handelt es sich bei den 25.000er- und 5.000er-Stichproben um pseudorandomisierte Stichproben.

Ebenfalls im Sinne der Vergleichbarkeit sind alle 25.000er-Stichproben mit der Freeware *AntConc* (Anthony 2014) und manuell durchsucht worden.¹⁵

Mit *AntConc* und unter Zuhilfenahme regulärer Ausdrücke habe ich in diesen

¹⁵ Da die verwendeten Korpora des DeReKo nur über die Rechercheplattform COSMAS II zugänglich sind, für die manuelle Recherche und die Arbeit mit *AntConc* aber eine Stichprobe als Textdatei notwendig war, habe ich mit COSMAS II in den virtuellen Korpora nach dem Zeichen „“ gesucht und die Treffer mit einigen Absätzen Kontext exportiert. Aus diesem Textmaterial habe ich jeweils pseudorandomisierte Stichproben (Überschneidungen ausgeschlossen und Medienanteile ausbalanciert) von 25.000 laufenden Textwörtern gezogen, die dann als Textdatei manuell und mit *AntConc* durchsucht werden konnten.

Tabelle 4.5.: Treffermenge 25.000er: Binnenmajuskel und manuell

Textform	Binnenmajuskel	manuell
Werbeanzeigen	107	428
Mail	11	76
SMS	5	110
Zeitung	9	31
Fachzeitschriften	14	58
Kriminalromane	1	16
Plenarprotokolle	12	12
Lexikoneinträge	27	148
Diskussionsbeiträge	22	54
Suchanfragentreffer	208	933

Stichproben und Korpora maschinell nach Zeichenketten mit Bindestrich, Apostroph und Binnenmajuskel gesucht und sie darüber hinaus manuell durchgesehen auf der Suche nach Getrennschreibungen oder sonstigen Auffälligkeiten an wortbildungs-morphologischen Grenzen. Manuell können auch solche morphologisch komplexen Wörter erfasst werden, die graphische Varianten, wie Fettdruck, Kursivierung oder eine farbliche Absetzung von Bestandteilen beinhalten. Ziel der manuellen Suche war es also, innerhalb einer repräsentativen aber auch überschaubaren Stichproben-größe auch all die Variationen zu erfassen, die sich entweder maschinell nicht oder nur sehr schwer finden lassen oder so unerwartet sind, dass sie gar nicht erst gezielt gesucht worden wären.

Eine maschinelle Suche nach Getrennschreibungen, also einer Lücke zwischen Bestandteilen ein und desselben Wortes, ist nicht unmöglich, aber mit so vielen Abstrichen verbunden, dass sie im Falle dieser Untersuchung nicht zielführend gewesen wäre.¹⁶ Tabelle 4.5 bietet einen Überblick über die Treffermengen der manuellen Suche und das Ergebnis der maschinellen Recherche nach Binnenmajuskeln.

Tabelle 4.6 dokumentiert die Suchanfragentreffer für Bindestriche und Apostrophe

¹⁶ Denkbar gewesen wäre eine Art Kookkurrenzanalyse mit parallelem Wortlistenabgleich, um so durch Leerzeichen getrennte Einheiten zu finden, die auch als komplexes Wort dokumentiert sind. Da es sich aber gezeigt hat, dass Getrennschreibungen hauptsächlich in nicht usualisierten Wörtern zu erwarten sind, wäre diese Methode wenig erfolgversprechend gewesen.

Tabelle 4.6.: Treffermenge 25.000er: Bindestrich und Apostroph

Textform	Bindestrich	Apostroph		
Werbeanzeigen	250		6	
Mail	123		8	
SMS	103		15	
Zeitung	117	(114)	12	(4)
Fachzeitschriften	97	(112)	3	(2)
Kriminalromane	23	(22)	0	(0)
Plenarprotokolle	61	(38)	0	(0)
Lexikoneinträge	113	(132)	5	(19)
Diskussionsbeiträge	131	(79)	21	(12)
Suchanfragentreffer	1018	(973)	70	(66)

in den 25.000er Stichproben. Um die Repräsentativität der Stichproben im Vergleich zu den größeren virtuellen Korpora zu belegen, wird bei den betreffenden Textformen in Klammern das auf die Stichprobengröße herunter gerechnete Ergebnis der Suchanfrage mit COSMAS II angegeben. Der Vergleich der Summen der Anfragetreffer über alle Textformen hinweg zwischen Stichprobe und rechnerisch zu erwartendem Ergebnis macht deutlich, wie gut die virtuellen Korpora in den Stichproben abgebildet sind. Aufgrund der geringeren Treffermengen sind dabei die Zahlen zu den Apostrophierungen in den einzelnen Textformen weniger aussagekräftig als die zu den Treffer der Bindestrichanfragen. Insgesamt ist aber auch hier eine Repräsentativität zu erkennen.

4.3.3 Stichproben 5.000

Die Stichproben im Umfang von 5.000 laufenden Textwörtern sind eine Teilmenge der 25.000 Textwörter umfassenden Stichproben bzw. der 25.000 Textwörter umfassenden Teilkorpora und dementsprechend ebenfalls ausgewogen hinsichtlich der Anteile der am Zeitungs-, Fachzeitschriften- und Kriminalromankorpus beteiligten Medien (Vgl. dazu auch 4.3.2).

Aus diesen in Summe gut 45.000 laufenden Textwörtern sind maschinell alle Zei-

chenketten mit zehn oder mehr Zeichen zwischen zwei Spatien erfasst worden.¹⁷ Dieses Verfahren geht also vom graphematischen Wortbegriff (ununterbrochene Zeichenkette zwischen zwei Leerzeichen) aus.

Im Gegensatz zur gezielten Suche nach gängigen Schreibvarianten ist diese Art der Recherche völlig ergebnisoffen und liefert neben erwartbaren Belegen, die Interpunktionszeichen (*Abi-Prüfung* (SMS)) oder Binnenmajuskeln (*PrivatKredit* (W)) aufweisen, auch unerwartetes wie *Biolectra®Magnesium243mgforte* (W) und vor allem eine Vielzahl variantenlos zusammengeschriebener Belege wie *Badezimmer* (W), so dass es aufgrund dieser Stichprobe auch möglich ist, die Zusammenschreibung und davon abweichende Varianten in ein Verhältnis zueinander zu setzen, zumindest für Zeichenketten ab zehn Zeichen aufwärts. Tatsächlich finden sich abgesehen von Apostrophierungen in diesen kleinsten Stichproben alle Varianten, die auch in den großen Korpora zu finden sind.

Die Beschränkung auf längere Zeichenketten hilft dabei, einen möglichst großen Anteil komplexer Wörter zu erfassen, ohne sich extrem viele falsch positive Treffer in Form einfacher Stämme einzuhandeln.

Die kleinste Stichprobengröße ist damit eigentlich die wichtigste für diese Untersuchung, weil nur sie es ermöglicht alle Varianten ins Verhältnis zueinander zu setzen, während über die größeren Stichproben sichergestellt wird, dass genügend Varianten abseits der im Deutschen sehr dominanten Zusammenschreibung erfasst werden, um auch diese systematisch untersuchen zu können.

4.3.4 Falsch positive Treffer

Die Recherche über alle Korpora und Stichproben hinweg liefert insgesamt 19.338 Treffer für potentiell wortbildungsmorphologisch komplexe Wörter des Deutschen, darunter ist aber auch eine größere Menge falsch positiver Treffer zu erwarten. Falsch positive Treffer sind Belege, die mit der durchgeführten Suchanfrage zwar

¹⁷ Vorbild dafür war eine Extraktion von Wörtern mit mindestens 15 Buchstaben im Rahmen einer Untersuchung zu deutschen Komposita von Gaeta & Zeldes (2012: 201). Da hier aber nicht nur Komposita, sondern komplexe Substantive allgemein betrachtet werden sollen, wurde das Längenkriterium auf mindestens zehn reduziert. Durch diese Anpassung werden auch andere Wortbildungsprodukte wie beispielsweise Affigierungen, die in der Regel etwas kürzer sind, erfasst. Bei den Zeichenketten mit nur neun oder weniger Zeichen ist der Anteil der falsch positiven Treffer dagegen so unverhältnismäßig hoch, dass eine weitere Absenkung nicht zielführend ist.

Tabelle 4.7.: Treffermenge 5.000er

Textform	>10 Zeichen
Werbeanzeigen	998
Mail	417
SMS	321
Zeitung	801
Fachzeitschriften	942
Kriminalromane	568
Plenarprotokolle	1.025
Lexikoneinträge	993
Diskussionsbeiträge	819
Suchanfragentreffer	6.884

gefunden werden, für die Untersuchung aber nicht von Interesse sind und dementsprechend ausgeschlossen werden müssen. Welche Arten von Treffern ausgeschlossen wurden, sollen im Sinne eines transparenten Vorgehens die folgenden Beispiele veranschaulichen.

- (18) a. Fremdsprachlicher Kontext:
Weblinks Texaco Cup & Anglo-Scottish Cup 1971-81 (RSSSF) (WPD11)
- b. Fremdwörter, die keine Anzeichen einer Integration in die deutsche Sprache aufweisen:
*verhindern keinen **edit war**, sondern laden weitere **edit warrior** geradezu ein* (WDD11)
- c. Nicht integrierte, fremdsprachliche Firmen- oder Produktnamen:
*Top-20-Systeme wie Obi, **Mcdonald's** oder Tschibo erlösen* (VDI09)
- d. Ortsnamen, Personennamen und Ländernamen:
Karl-Marx-Stadt (WPD11)
Junge-Reyer, Lea-Sophie
Saudi-Arabien, Deutschland (PBR)
- e. Apostroph als Auslassungszeichen:
*Zum Glück **gibt's** Erfolg* (WPD11)

- f. Apostroph als Zeichen für den glottalen Verschlusslaut:
ist in Altäthiopisch (Ge'ez) verfasst (HMP09)
- g. Konglomerate:
mit den Einladungen....Hier mein Vorschlag (Mail)
- h. Metatext:
richtige Magnesium [PRODUKT] [PRODUKT]* [PRODUKT]** Magnetrans® 375* (W)
- i. Link/URL:
jetzt bewerben auf vr.del/Erfolg Jeder Mensch hat (W)
- j. Hashtag:
MYTHOS ZWEIER. #BMWzer (W)
- k. Zahlen:
Science 327 S. 1350-1352, 2010 also Rhythmus erzeuger (SPK)
- l. Phrasen:
August 2008 wurden erstmals 5-Sterne an den Maintal-Radweg (WPD11)
- m. Wortstatus unsicher:
vor allem bei Wind Stärke drei bis vier, (Krimi)
- n. Koordinierte Komposita mit Tilgungsstruktur:
bei allen Mercedes-Benz Niederlassungen und teilnehmenden Mercedes-Benz und smart Vertragspartnern (W)
- o. nicht zu untersuchende Wortarten:
im browser manipulieren könnte, ohne die Originaldatei auch nur an zufassen. (WDD11)
was ich zwischendurch mit meinen (Mail)

Die Treffer unter (18a) bis (18c) zeigen exemplarisch, welche Gruppen von Treffern aufgrund ihrer Fremdsprachlichkeit und mangels Integrationszeichen ausgeschlossen worden, da es sich hier um eine Untersuchung der wortinternen Schreibung im Deutschen handelt. Ebenfalls ausgeschlossen wurden Ortsnamen, Personennamen und Ländernamen (18d). Berücksichtigt werden dagegen Produktnamen, Titel und Straßennamen. Für den gesamten Namenbereich gilt zwar, dass die Schreibung mitunter willkürlich festgelegt und damit nicht unmittelbar repräsentativ für das zu

untersuchende System ist, die Kategorie sollte jedoch nicht als Ganzes ausgeblendet werden. Die Apostrophierungen unter (18e) und (18f) markieren jeweils keine wortbildungsmorphologische Grenze und sind deshalb nicht relevant als Belege. Die Treffer unter (18g) bis (18l) entsprechen zwar dem graphematischen Wortbegriff als Zeichenketten zwischen zwei Leerzeichen, es handelt sich aber nicht um morphosyntaktische Wörter, sondern um Konglomerate aufgrund ausgelassener Leerzeichen (18g), Annotationen (18h), URLs (18i), Hashtags (18j), Zahlen (18k) oder Phrasen (18l). *5-Sterne* zeigt damit eine Bindestrichschreibung an der Grenze zwischen zwei morphosyntaktischen Wörtern und keine wortinterne. Zudem wurden Treffer wie *Wind Stärke* in (18m), die, wenn man die Schreibung außer acht lässt, sowohl als Phrase als auch als ein morphosyntaktisches Wort klassifiziert werden können, aussortiert. Genauso wurde mit Koordinationsstrukturen (18n) verfahren, da es sich dabei um mehrere syntaktische Wörter handelt und die Schreibung der Vollform durch die Tilgung beeinflusst sein könnte. Wie in den Vorbemerkungen zu den zentralen Begriffen schon angekündigt konzentriert sich die Untersuchung auf Substantive (und Adjektive). Dementsprechend bleiben auch andere Wortarten (18o) außen vor. Außerdem wurden Mehrfachtreffer aufgrund der mehrstufigen Recherche für ein und denselben Beleg ausgeschlossen, die unter anderem dadurch zustande kommen, dass Schreibungen mit einem Bindestrich wie *Burnout-Syndrom* (VDI09) sowohl über die Suchanfrage „*-s*“ als auch über die Erhebung aller Zeichenketten mit mehr als 10 Zeichen gefunden werden.

4.4 Belegsammlung und Annotation

Um die Ergebnisse der Recherche für die tiefergehende Analyse und Auswertung der Daten nutzbar zu machen, sind sie in einer Belegsammlung zusammengetragen und aufbereitet worden.

Diese Belegsammlung umfasste in einer ersten Version zunächst alle Treffer der Suchanfragen und manuellen Recherche inklusive Kontext und Quellenangabe. Vor der eigentlichen Analyse der Belege mussten aus dieser Grundgesamtheit von 19.338 Treffern manuell die falsch positiven Treffer aussortiert werden. Die verbleibenden 14.229 Belege für wortbildungsmorphologisch komplexe Substantive und Adjekti-

ve wurden schließlich hinsichtlich der erwarteten Einflussfaktoren auf wortinterne Schreibvarianten annotiert. Zuerst beschränkte die Annotation sich aus ökonomischen Gründen auf die Wortebene. Im nächsten Schritt wurde sie dann für Beleggruppen, die unterschiedliche Varianten aufweisen, auf der Ebene der morphologischen Grenzen spezifiziert.

4.4.1 Wortbezogen

Die Belege wurden als komplexe Wortformen im Ganzen hinsichtlich ihres Wortbildungsmusters, ihres Lexikalisierungsstatus und der Belegquelle annotiert, um auf dieser Basis den Einfluss des Wortbildungsmusters, der Lexikalisierung und der Textform auf die Ausprägung der Schreibung an wortbildungsmorphologischen Grenzen analysieren zu können.

Wortbildung

Die Klassifizierung der Wortbildungen wurde auf der Basis der grundlegenden Arbeiten zur Wortbildung von Fleischer & Barz (2012) und Elsen (2011) vorgenommen. Unterschieden wurde dabei zwischen den Wortbildungsarten Komposition, Derivation, Konversion sowie ihren jeweiligen Unterkategorien, die im Folgenden allgemein als Wortbildungsmuster bezeichnet werden. Wortbildungsmuster, die sich keiner, bzw. nicht eindeutig einer der Wortbildungsarten zuordnen lassen, wie beispielsweise Rückbildungen oder Zusammenbildungen sind nur sehr selten belegt und werden unter *Sonstiges* geführt.

Lexikalisierung

Ob eine Wortform als lexikalisiert gilt oder nicht, ist stark abhängig von der Definition, ob beispielsweise eine bestimmte Vorkommenshäufigkeit zugrunde gelegt wird oder die Anwendung in möglichst vielen Kontexten belegt sein sollte.

Für die vorliegende Untersuchung wurde ein möglichst genau auf das Korpus und die Belegsammlung passendes Verfahren gewählt, um eine Entscheidung bezüglich des Lexikalisierungsstatus der Belege zu treffen. Dazu wurden die Belege maschinell mit einer am 31.12.2014 vom Institut für Deutsche Sprache veröffentlichten Wortliste abgeglichen, die die 100.000 häufigsten Wortformen des Deutschen Referenzkorpus enthält (DeReKo-2014-II-MainArchive-STT.100000).

Dementsprechend werden Belege, die sich in identischer Schreibung auf der Wortformliste finden als lexikalisiert klassifiziert, während Belege, zu denen sich keine Entsprechung findet, als nicht lexikalisiert gelten. Es wird also ein frequenzbasierter Lexikalisierungsstatus erhoben. Aufgrund des Umfangs der gewählten Vergleichsliste von nur 100.000 Wortformen ist zudem von einem engen Lexikalisierungsbegriff auszugehen.

Da Ausschnitte des Deutschen Referenzkorpus einen Großteil des hier genutzten Untersuchungskorpus ausmachen und die Liste ebenso wie die Untersuchungsbelege nicht lemmatisiert ist, ist von einer guten Passung von Frequenzliste und Untersuchungskorpus auszugehen.

Quelle

Um einen Einfluss der Textform auf die Schreibung an wortbildungsmorphologischen Grenzen erfassen zu können, muss natürlich auch die Textform als solche annotiert werden. Das bedeutet, dass allen Belegen die Textform des entsprechenden Teilkorpus zugeordnet wird, also Werbeanzeige, Mail, SMS, Zeitung, Fachzeitschrift, Kriminalroman, Plenarprotokoll, Lexikoneintrag oder Diskussionsbeitrag.

Da die Teilkorpora mitunter mehrere Medien umfassen, wie das Zeitungs-Korpus oder das Fachzeitschriften-Korpus, ist in diesen Fällen neben der Textform auch das Teilkorpus als Kürzel (nach Vorgabe des DeReKo) annotiert. So entstammt der Beleg *Bakterien-Gemeinschaft* dem Fachzeitschriften-Korpus, genauer dessen Teilkorpus der Zeitschrift *Spektrum* (SPK). Eine Übersicht, welche Kürzel für die jeweiligen Teilkorpora verwendet werden, findet sich in Tabelle 4.1.

4.4.2 Morphembezogen

Den Forschungshypothesen entsprechend wurden für die einzelnen Bestandteile der Wortbildungen graphematische Markiertheit, Eigennamenstatus und Fremdsprachlichkeit annotiert, weil alle drei Kategorien eine Abweichung von der Zusammenschreibung begünstigen.

Graphematische Markiertheit

Als graphematisch markiert gilt ein Wortbestandteil immer dann, wenn er nicht der Definition eines graphematischen Wortes des Kernbereichs entspricht (vgl. 3.1.4).

Das bedeutet, es handelt sich zwar um ununterbrochene Zeichenketten, sie erfüllen aber zumindest eines der Kriterien nicht, über die der Kernbereich graphematischer Wörter definiert ist.

Sie weisen also entweder weniger als zwei Buchstabengrapheme auf wie z. B. $|B|$ in *B-Klasse*, bestehen nicht (2 in *2fach*) oder nicht nur aus Buchstabengraphemen (*N8* in *N8werk*) oder bestehen zwar ausschließlich aus einer hinreichenden Menge an Buchstabengraphemen, diese bilden aber keine graphematische Silbe nach dem Allgemeinen Graphematischen Silbenbaugesetz, wie z. B. *Kfz* in *Kfz-Versicherung*.

Eigennamen

Komplexe Personennamen mit Eigennamen in der morphologischen Kopfposition, die die Recherche ergeben hat, wurden bereits von der Belegsammlung ausgenommen, da ihre Schreibung stärker arbiträr ist als die des Kernwortschatzes. Als Bestandteile morphologisch komplexer Wörter kommen aber Eigennamen, darunter auch Personennamen, innerhalb der Belegsammlung als Wortbestandteile vor.

Da die Beteiligung von Eigennamen Einfluss auf die Schreibvariation an Wortbildungsgrenzen hat, wird auch deren Auftreten annotiert. Eigennamen sind dabei nach Nübling, Fahlbusch & Heuser (2012) definiert über ihre Monoreferenz, also den Bezug auf ein einzelnes Objekt, die einen schnellen und direkten sprachlichen Zugriff auf ebendieses Objekt ermöglicht. Prototypische Beispiele für Eigennamen sind dabei Personen- und Ortsnamen (vgl. Nübling et al. 2012: 17-18).

Unter (19) finden sich Belege mit Eigenamen als Bestandteil:

- (19) Opel-Krise (WDD11), Niedersachsen-Delegierten (HAZ09), Parkinsonoid (WDD11), Amerikaner (HAZ09), Galileo-Satellitennavigationssystem (DIV)

Fremdsprachlichkeit

Ab wann ein Wort oder Wortbestandteil fremd ist, ist nicht einfach zu beantworten. Auch die Einführung zusätzlicher Begriffe und Kategorien wie Lehnwort, Anglizismus oder Eindeutschung macht die Entscheidung eher komplizierter als einfacher.

Daher bleibt es in der vorliegenden Untersuchung bei einer binären Einteilung in die Kategorien fremd (exogen) und nicht fremd (nativ). Ausschlaggebend für die Entscheidung, welcher Kategorie ein Bestandteil angehört, sind die Eigenschaften des Wortbestandteils. Nach Eisenberg ist dann von einem Fremdwort zu sprechen,

„wenn ein Wort fremde Eigenschaften hat, die der Normalsprecher einer fremden Sprache zuschreibt“ (Eisenberg 2012: 29). Als Beispiel dafür führt er die Verwendung von Präfixen wie *ex-* oder *prä-* aus dem Lateinischen oder den Suffixen *-age* und *-iere* aus dem Französischen an, also morphologische Indikatoren für Fremdheit. Darüber hinaus gibt es aber auch phonologische (fremde Laute, aber auch Silbenbau und Wortakzent) oder orthographische Eigenschaften, die Fremdwörter von nativen Wörtern des Kernwortschatzes unterscheiden (vgl. Eisenberg 2012: 27-29).

Für eine Untersuchung geschriebener Sprache sind natürlich orthographische/graphematische Eigenschaften von besonderer Bedeutung als Unterscheidungskriterien. Dazu zählen neben vom Kernwortschatz abweichenden Laut-Buchstaben-Korrespondenzen auch Abweichungen bei Dehnung, Schärfung und morphologischer Schreibung und der Silbentrennung (vgl. Eisenberg 2012: 316-353).

Der Großteil dieser Kriterien lässt sich auch auf einzelne Wortbestandteile übertragen. Ein Wortbestandteil gilt daher im Folgenden als fremd, wenn er eine fremde morphologische oder graphematische Eigenschaft aufweist. Diese Definition führt zu einem recht breiten Fremdwortbegriff, ist aber im Sinne der Annotation gut anwendbar und reproduzierbar. Zur Illustration dienen die Beispiele unter (20) für Komposita mit fremdem Erstglied (20a), fremdem Zweitglied (20b) und gleich zwei fremdsprachlichen Anteilen (20c):

- (20) a. Bakterien-Gemeinschaft (SPK), Businessziele (W), Browser-Fenster (Mail)
 b. Abstellungstheater (NUZ09), Anlagekapitals (WPD11), Wind-Stadion (HMP09)
 c. Actionprämie (W), Sozialpolitik (PBR), Burnout-Syndrom (VDI09)

4.4.3 Grenzbezogen

Die folgenden Annotationskategorien beziehen sich nicht auf die jeweiligen Bestandteile von Wortbildungen in ihrer Gänze, sondern fassen nur die Grenzen dieser Bestandteile und die Phänomene, die sich dort abspielen und einen Einfluss auf die Schreibung an dieser Grenze haben (können), ins Auge.

Fugenelemente

Da Komposita mit Fugenelementen nach Bredel (2008: 115) seltener einen Bindestrich aufweisen, wird ihr Auftreten annotiert. Mit Fuhrhop (2000: 202) werden „alle Einheiten, durch die sich die Erstglieder in Komposita von den entsprechenden Nominativ-Singular-Formen unterscheiden“ als Fugenelemente verstanden. Daraus ergibt sich für das Deutsche das Inventar: *-s-*, *-es-*, *-er-*, *-e-*, *-n-*, *-en-*, *-(e)ns-*, das sich auch in den Belegen unter (21) widerspiegelt.

- (21) Vermittlungsausschuss (PBR), Vorjahresvergleich (VDI09), Länderzuständigkeiten (PBR), Fachkräftemangel (VDI09), Erstrundenniederlagen (WDD11), Personentransporter (VDI09), Glaubensreife (WDD11)

Markierte Buchstabenfolgen

Nach der orthographischen Norm ist der Einsatz eines Bindestriches in Zusammensetzungen, wenn drei gleiche Buchstaben aufeinandertreffen (*Bett-Tuch*, *Kaffee-Ersatz*) oder zur Vermeidung von Missverständnissen wie in *re-integrieren*, fakultativ (vgl. die Ausführungen zur Orthographie unter 3.2). Neben Buchstabentrippeln kann also auch schon das Aufeinandertreffen zweier Vokale zum Einsatz eines Bindestrichs führen, in *re-integrieren* wohl dadurch begünstigt, dass es sich bei den beiden Vokalen potentiell um einen Diphthong handelt.

Um ermitteln zu können, ob es in den Korpusbelegen für solche und ähnliche Fälle tatsächlich einen Effekt zur Ausprägung einer Schreibvariante abseits der Zusammenschreibung gibt, wurden alle potentiell markierten Buchstabenfolgen an Morphemgrenzen annotiert. Das schließt neben Trippeln wie in *Bett-Tuch* und *Kaffee-Ersatz* auch das Aufeinandertreffen zweier Vokale mit ein (unabhängig davon ob es sich um unterschiedliche wie in *re-integrieren* handelt oder um dieselben wie in *Messe-Empfang* bzw. *Messeempfang*). Außerdem wurden gewissermaßen als Gegenstück zu den Doppelvokalen auch die Fälle annotiert, bei denen dieselben Konsonanten an der Wortbildungsgrenze aufeinander treffen, wie in *Tauf-Feier* bzw. *TauFFEIER*, um auch für diese Kombinationen einen Effekt auf die Schreibung an der Wortbildungsgrenze zu prüfen.

Schreibvarianten

Ebenfalls grenzbezogen wurden die Schreibvarianten als zentrale Variable erfasst und zwar gleich auf zwei Arten: einmal binär als standardisierte Zusammenschreibung oder davon abweichende Variante wie in Beispiel (22) und mit einer detaillierten Aufschlüsselung der Varianten abseits der Zusammenschreibung in Bindestrichschreibung, Getrenntschreibung, die Verwendung von Binnenmajuskeln und andere Phänomene wie z. B. die Kombination eines Bindestrichs mit anderen Interpunktionszeichen wie in „*Spiegel*“-Reporter (HAZ09).

- (22) a. Gemüseauflauf (DIV), Parkinsonpatienten (SPK)
b. A-Junioren (M09), linguisten mist (SMS), „*Spiegel*“-Reporter (HAZ09), Luther-Schau (M09), PrivatKredit (W)

5. Ergebnisübersicht zu den 5.000er-Stichproben

Von den 14.229 Belegen für wortbildungsmorphologisch komplexe Wörter, die die Korpusrecherche ergeben hat, finden sich mit 3.950 Belegen etwas weniger als 30% in den kleinsten Korpusausschnitten, die jeweils etwa 5.000 Textwörter pro Textform umfassen. Unter diesen 3.950 Belegen sind 3.106 Belege für wortbildungsmorphologisch komplexe Substantive.

Die Suchanfrage nach Zeichenketten mit zehn oder mehr Zeichen in den 5.000er-Stichproben erlaubt es, für diesen Datenausschnitt, und nur für diesen, die Schreibvarianten ins Verhältnis zueinander zu setzen, weil so tatsächlich alle Varianten ergebnisoffen erfasst werden können. Dementsprechend dient dieser Datenausschnitt von 3.106 wortbildungsmorphologisch komplexen Substantiven als Basis der Analyse.

Die Belege sind nicht lemmatisiert. Ein Blick in die Frequenzen in Tabelle 5.1 macht aber deutlich, dass dennoch der weitaus größte Teil der Belege nur ein- oder zweimal vorkommt, während Ausreißer nach oben mit fünf oder mehr Instanzen selten sind.

Tabelle 5.1.: Belege 5.000er-Stichprobe nach Frequenz

Frequenz	1	2	3	4	5	6	7	8	9	>9
Belege	1.928	237	72	21	14	6	7	5	6	9

Die Belege, die häufiger als neunmal vorkommen (1), finden sich hauptsächlich in für einzelne Textformen typischen formelhaften Wendungen. So ist in Werbeanzeigen für Autos der *Kraftstoffverbrauch* zwingend anzugeben und die protokollierten Plenumssitzungen des Bundesrates sind anhand der Belege *Ausschussempfehlungen*,

Tabelle 5.2.: Belege 5.000er-Stichprobe nach Wortbildungsgrenzen

Wortbildungsgrenzen	1	2	3	4	5	6	7	8	10
Belege	908	1.260	649	203	65	12	7	1	1

Bundesregierung, Drucksache, Handzeichen, Minderheit und *Wortmeldung* regelrecht in ihrem Ablauf nachvollziehbar.

- (1) Ausschussempfehlungen (22mal), Bundesregierung (12mal), Drucksache (21mal), Gesellschaft (12mal), Handzeichen (20mal), Kraftstoffverbrauch (10mal), Minderheit (33mal), Sicherheit (11), Wortmeldung (14mal)

Vor der theoretischen Modellierung soll hier zunächst ein Einblick in diese 3.106 Belege für wortbildungsmorphologisch komplexe Substantive gegeben werden mit Rücksicht auf die realisierte wortinterne Schreibvariation (5.1), die vorkommenden Wortbildungen (5.2) und die Verteilung über die neun verschiedenen Textformen (5.3).

5.1 Schreibvarianten

Die 3.106 Belege für wortbildungsmorphologisch komplexe Substantive weisen häufig nicht nur eine, sondern gleich mehrere Wortbildungsgrenzen auf und bieten damit auch gleich mehrere Ansatzpunkte für wortinterne Schreibvarianten. Bei *Strass-Steinchen* (W) liegt z. B. jeweils eine Kompositions- und eine Derivationsgrenze vor. Während an der Derivationsgrenze zwischen *Stein* und *-chen* zusammengeschrieben wird, findet sich zwischen den Kompositionsgliedern *Strass* und *Steinchen* ein Bindestrich.

Den 3.106 Belegen aus den 5.000er-Stichproben entsprechen 6.651 Wortbildungsgrenzen, also durchschnittlich etwas mehr als zwei Grenzen pro Beleg. Wie Tabelle 5.2 zeigt, weist dabei ein Großteil der Belege (2.817 von 3.106, also gut 90%) eine bis drei wortbildungsmorphologische Grenzen auf, und nur 10% der Belege haben vier oder mehr Grenzen.

Die Übersicht über die Schreibvarianten in 5.3 umfasst also nicht die 3.106 Belege

Tabelle 5.3.: Belege 5.000er-Stichprobe nach Schreibvarianten

Schreibvariante	Wortbildungsgrenzen
Zusammenschreibung	6.382
Bindestrichschreibung	458
Getrenntschreibung	32
Binnenmajuskel	262
Sonstiges	5

als Grundgesamtheit, sondern alle 6.651 Wortildungsgrenzen dieser Belegmenge.

Erfasst wurden dabei Zusammenschreibung, Bindestrichschreibung, Getrenntschreibung, Binnenmajuskel und sonstige Varianten. Einen Ergebnisüberblick bietet 5.3. Dass dabei 7.139 Varianten bei 6.651 Wortbildungsgrenzen beobachtet wurden, ist der Tatsache geschuldet, dass es regelmäßig Kombinationen von Varianten an einzelnen Wortbildungsgrenzen gibt. Die häufigste Kombination bei gut 200 Belegen ist dabei die von Bindestrich und Binnenmajuskel wie in *Business-Klasse* (W), während sich eine Binnenmajuskel ohne vorausgehenden Bindestrich oder wortinternes Leerzeichen wie in *LautsprecherBox* (Mail) nur in 19 Fällen findet.

Wie 5.3 zeigt, ist die wortinterne Zusammenschreibung dabei so dominant, dass alle anderen Varianten, auch die Verwendung von Bindestrichen, als Randerscheinungen gelten müssen. So finden sich unter den 3.106 Belegen 2.860, die an allen wortbildungsmorphologischen Grenzen zusammengeschrieben werden, was gut 92% der Belege entspricht, und nur 47, die an keiner einzigen Wortbildungsgrenze eine Zusammenschreibung aufweisen. Dabei handelt es sich in der Regel um zweigliedrige Komposita mit Bindestrich an der Kompositionsfuge aus zwei Grundmorphemen wie *Mersenne-Zahlen* (DIV).

Die wenigen sonstigen Schreibvarianten umfassen Klammern wie in *(Kämpfer-)Lieder* (WDD11) und Anführungszeichen wie in *„Spiegel“-Reporter* (HAZ09).

Tabelle 5.4.: Belege 5.000er-Stichprobe nach Wortbildungsarten

Wortbildungsart	Belege	Anteil in %
Komposition	2.139	68,87%
Derivation	861	27,72%
Konversion	70	2,25%
Sonstiges	36	1,16%
Stichprobe 5000er	3.106	100%

5.2 Wortbildungen

Tabelle 5.4 bietet einen ersten Überblick über die Wortbildungsarten der 3.106 Belege aus den 5.000er-Stichproben, während die Tabellen 5.5, 5.6, 5.7 und 5.8 die Komposita, Derivate, Konversionen und sonstigen Typen genauer aufschlüsseln.

Die Belege, die zum Teil das Produkt einer ganzen Reihe von Wortbildungsprozessen sind, wurden dabei der Kategorie zugeordnet, die dem letzten Wortbildungsschritt entspricht. *Datenaufzeichnungen* (SPK) wird also zunächst als Beleg im Ganzen als Kompositum, bzw. Determinativkompositum, erfasst, obwohl beim Kompositionsglied *aufzeichnungen* diesem letzten Wortbildungsschritt auch schon eine Präfigierung der Derivationsbasis *zeichnen* mit *auf-* und eine Suffigierung des Ergebnisses davon (*aufzeichnen*) mit dem Suffix *-ung* vorausgegangen ist.

Hier werden also nicht alle 6.651 wortbildungsmorphologischen Grenzen in den Blick genommen, sondern nur die 3.106 Grenzen des letzten Wortbildungsschrittes, also der höchsten Hierarchiestufe. Diese Einschränkung wird aber im Sinne der Validität der Untersuchung vorgenommen, um dem Untersuchungsgegenstand treu zu bleiben. Die in Kapitel 2.1.2 erläuterte Fokussierung auf wortbildungsmorphologisch komplexe Substantive bedingt es, sich auf Prozesse zu konzentrieren, die ein substantivisches Produkt erzeugen. Dieses Kriterium erfüllt der letzte Wortbildungsschritt bei allen Belegen, während das Beispiel *Datenaufzeichnungen* zeigt, das unter den vorangegangenen Prozessen nur ein Teil relevant im Sinne der Untersuchung ist. Während die Präfigierung mit *aus-* ein Verb bildet, die Grenze also irrelevant für die eigentliche Untersuchung ist, ergibt die Suffigierung durch *-ung* ein Substantiv und ist damit potentiell auch von Interesse.

Tabelle 5.5.: Belege 5.000er-Stichprobe nach Wortbildungsarten: Komposition

Komposition	Belege	Anteil in %	Zusammenschreibungsquote
Determinativkomposita	2.055	96,07%	89,88%
Phrasenkomposita	21	0,98%	14,29%
Possessivkomposita	0	-	-
Affixoidbildungen	23	1,07%	100%
Inversionskomposita	9	0,42%	100%
Kopulativkomposita	0	-	-
Konfixkomposita	31	1,45%	77,42%
Stichprobe 5000er	2.139	100%	88,70%

Tabelle 5.4 zeigt deutlich, dass Komposita mit Abstand den größten Teil (knapp 69%) der Belege ausmachen und knapp 28% der Belege auf Derivationen zurückgehen, während Konversionen und sonstige Wortbildungsmuster nur einen minimalen Anteil von summiert gut 3% ausmachen.¹

Im Sinne der Ausgangsfrage, ob die Wortbildungsart einen Effekt auf die Ausprägung von Schreibvarianten hat, wird in den folgenden Tabellen zu den Wortbildungsarten auch die Quote der Zusammenschreibungen angegeben.

5.2.1 Komposita

Die 2.139 Belege umfassende Gruppe der Komposita wird wiederum dominiert von prototypischen Determinativkomposita, die aus zwei (*Intensivpflege* (W)) oder mehr Grundmorphemen (*Druckluftspeicher* (PBR)) aufgebaut sind und 96 % der Belege ausmachen. Daneben finden sich noch 21 Komposita mit phrasalem Erstglied (*Grand-Slam-Titel* (WPD₁₁)), neun Inversionskomposita (*Jahrhundert* (WPD₁₁)), 23 Affixoidbildungen (*Meisterwerk* (DIV)) und 31 Konfixkomposita (*Mikrosekunde* (SPK)). Possesiv- und Kopulativkomposita sind dagegen nicht belegt.

Die Zusammenschreibungsquote ist für die Komposita mit 88% etwas niedriger als die Gesamtquote von 92% und bildet vor allem die prototypischen Determinativkomposita ab, von denen gut 10% abweichende Varianten aufweisen. Wäh-

¹ Eine detailliertere Aufschlüsselung der Belege nach den Unterkategorien der Wortbildungsarten erfolgt in den Tabellen 5.5 bis 5.8

rend alle Affixoidbildungen und alle Inversionskomposita zusammengeschrieben werden, zeigen Phrasenkomposita und Konfixkomposita einen höheren Anteil von Bindestrich- oder Getrennschreibungen. Im Fall der Phrasenkomposita werden sogar nur drei der 14 Belege zusammengeschrieben (z. B. *Dreistufenmodell* (PBR)), was sicherlich darauf zurückzuführen ist, dass die Norm hier eine Bindestrichschreibung empfiehlt. Für die Konfixkomposita gibt es keine Norm, die den Einsatz eines Bindestrichs oder Ähnliches, befördern würde. Dennoch finden sich unter den 31 Belegen gleich sieben Belege mit Bindestrichschreibung, die alle dem gleichen Muster folgen. Der Bindestrich sitzt jeweils zwischen dem Konfix (*bio, ultra, mini*)² und dem darauf folgenden Grundmorphem bzw. Determinativkompositum aus Grundmorphemen (2) und alle fünf Komposita mit *bio* entstammen dem Werbeanzeigen-Korpus. Für Komposita wie *Bürokraten* (DIV), die nur aus Konfixen gebildet sind, ist dagegen nur die Zusammenschreibung in dieser Stichprobe belegt.

- (2) Bio-Pflanzenauszügen (W), Bio-Granat-Apfel (W), Bio-Fachhandel (W), Bio-Goldhirse (W), Ultra-Rechte (WPD11), Bio-Handel (W), Mini-Teens (BRZ09)

5.2.2 Derivationen

In der Stichprobe finden sich neben der großen Gruppe der Komposita auch noch 861 Derivate, bei denen es sich größtenteils um Suffigierungen (790 Belege) handelt.

Präfigierungen sind nur in 15 Fällen belegt, von denen zwölf zusammengeschrieben werden und drei eine Bindestrichschreibung aufweisen (*ANTI-OXIDANT* (W, zweimal), *Ex-Präsident* (M09)), was einer relativ niedrigen Zusammenschreibungsquote von 80% entspricht.

Anders liegt der Fall bei den Suffigierungen. Alle 790 Belege wie *Wahrscheinlichkeit* (Mail) und *Attraktivität* (PBR) werden zusammengeschrieben. Ebenso die 56 Belege für implizite Derivationen wie *Stellungnahme* (PBR) und *Zusammenkunft* (DIV). Zirkumfigierungen sind in der Stichprobe nicht belegt.

Dementsprechend liegt die Zusammenschreibungsquote für die Derivationen bei fast 100%.

² Für *mini* ist mit *Miniarmbrust* (M09) im Gegensatz zu *bio* und *ultra* auch eine Zusammenschreibung belegt.

Tabelle 5.6.: Belege 5.000er-Stichprobe nach Wortbildungsarten: Derivation

Derivation	Belege	Anteil in %	Zusammenschreibungsquote
Präfigierung	15	1,74%	80%
Suffigierung	790	91,75%	100%
Zirkumfigierung	0	-	-
Implizite Derivation	56	6,5%	100%
Stichprobe 5000er	861	100%	99,65%

Tabelle 5.7.: Belege 5.000er-Stichprobe nach Wortbildungsarten: Konversion

Konversionen	Belege	Anteil in %	Zusammenschreibungsquote
morphologisch	5	92,86%	100%
syntaktisch	65	7,14%	100%
Stichprobe 5000er	70	100%	100%

5.2.3 Konversionen

Die Konversion als Wortbildungsprozess ist nur für 70 Belege verantwortlich, die ausnahmslos alle zusammengeschrieben werden, unabhängig davon, ob es sich um syntaktische wie *Verschmelzen* (SPK) oder morphologische Konversionen wie *Unterrecht* (SMS) handelt.

Aufgrund der geringen Belegmenge und der Tatsache, dass Konversionen nicht an einer wortbildungsmorphologischen Grenze ansetzen,³ werden diese Belege in der folgenden Analyse ausgeklammert.

5.2.4 Sonstiges

Abseits von Komposition, Derivation und Konversion finden sich mit 36 Belegen nur noch wenige Belege, die aus vier anderen Wortbildungsprozessen hervorgehen. Bis auf eine Zusammenrückung (3a) und einige Zusammenbildungen (3b) werden dabei alle Belege zusammengeschrieben.

³ Das schließt natürlich Wortbildungsgrenzen innerhalb der Konversionen nicht per se aus, da dem Konversionsprozess wie im Beispiel *Verschmelzen* (SPK) auch ein anderes Wortbildungsmuster, hier eine Präfigierung, vorausgegangen sein kann.

- (3) a. würden dich und deine Familie gern für nächste Woche Sonntag zum **Kaffee trinken** einladen (SMS)
- b. Anti-Falten Pflege, Anti-Ageing Pflegesystem (2mal), Anti-Haarverlust Kollektion (2mal) (W)

Die niedrigen Belegzahlen und die damit marginale Rolle der Prozesse innerhalb dieser Stichprobe sprechen aber dafür, diese Belege von der Analyse auszunehmen.

Tabelle 5.8.: Belege 5.000er-Stichprobe nach Wortbildungsarten: Sonstiges

Sonstiges	Belege	Anteil in %	Zusammenschreibungsquote
Zusammenrückung	7	19,44%	85%
Zusammenbildung	27	75%	70,37%
Reduplikativkomposita	0	-	-
Kurzwortbildung	1	2,8%	0%
Rückbildung	1	2,8%	100%
Kontamination	0	-	-
Stichprobe 5000er	36	100%	72,22%

5.3 Textformen

Abbildung 5.1 schlüsselt die 3.106 Belege nach den verschiedenen Textformen auf, in denen sie vorkommen. Dabei gibt es trotz annähernd gleicher Stichprobengröße von etwa 5.000 Textwörtern pro Textform deutliche Unterschiede bezüglich der Belegmenge für die Teilkorpora.

Den größten Anteil an der Gesamtmenge haben die 596 Belege aus dem Plenarprotokoll-Korpus, was einem Anteil von 19% entspricht, während aus dem SMS-Korpus nur 76 Textwörter als Belege erfasst wurden, die damit weniger als 2% der Belege ausmachen. Die drei Textformen SMS, Mail und Kriminalromane mit den wenigsten Belegen tragen gemeinsam nur so viele Belege (14%) bei wie andere Textformen (Zeitung, Fachzeitschriften, Lexikoneinträge) alleine, während etwa die Hälfte der 3.106 Belege den drei Textformen Plenarprotokolle, Werbeanzeigen und Fachzeitschriften entstammt.

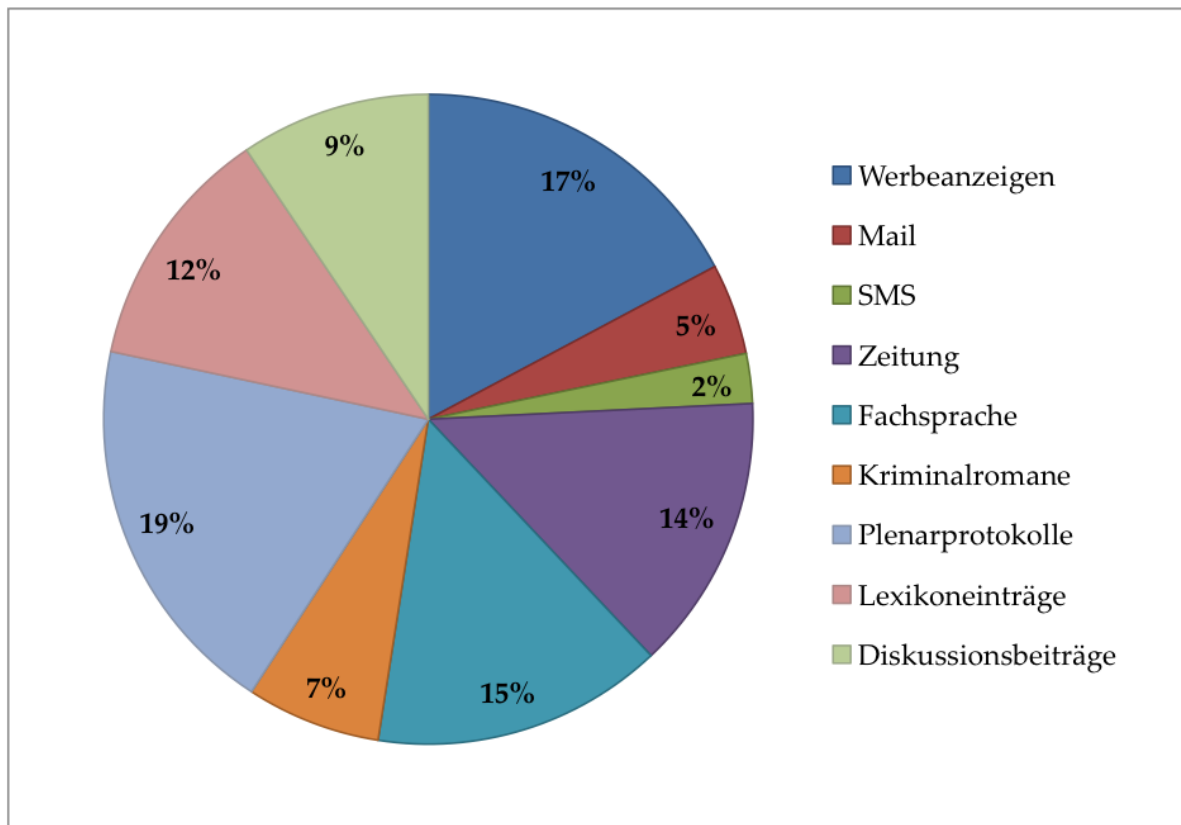


Abbildung 5.1.: Belege der 5.000er-Stichprobe nach Textform

Diese Unterschiede gehen auf die Suchanfrage nach Wörtern mit zehn oder mehr Zeichen zurück, deren Anteil je nach Textform ebenso stark variiert. Wie Tabelle 5.9 zeigt, machen wortbildungsmorphologisch komplexe Substantive durchschnittlich knapp 7% der 45.588 Textwörter aus, die die 5.000er-Stichproben für alle Textformen gemeinsam umfassen. Diese Belegquote variiert aber deutlich zwischen den verschiedenen Textformen und bestätigt auch hier die Textformen Mail (2,7%), SMS (1,42%) und Kriminalromane (3,97%) als Ausreißer nach unten und Werbeanzeigen- (11,25%) und Plenarprotokoll-Korpus (11,7%) nach oben.

Die geringen Quoten für Belege aus den SMS-, Mail- und Kriminalromankorpora korrelieren unmittelbar mit einer geringeren Treffermenge (321/417/568) für Zeichenketten mit zehn oder mehr Zeichen (vgl. 5.9). Es gibt in diesen Textformen also ganz schlicht weniger längere Zeichenketten.

Obwohl die Treffermenge, die neben den Belegen für komplexe Substantive ja auch noch falsch positive Treffer umfasst, für die übrigen Teilkorpora nur zwischen 800 und etwa 1.000 Treffern variiert, ergeben sich Abweichungen hinsichtlich der Be-

Tabelle 5.9.: Treffer und Belege 5.000er-Stichprobe nach Textformen

Textform	Umfang	Treffer	Belege	Belegquote
Werbeanzeigen	4.753	998	535	11,25%
Mail	5.219	417	141	2,7%
SMS	5.351	321	76	1,42%
Zeitung	5.061	801	427	8,44%
Fachzeitschriften	4.795	942	450	9,38%
Kriminalromane	5.237	568	208	3,97%
Plenarprotokolle	5.094	1.025	596	11,7%
Lexikoneinträge	5.057	993	379	7,49%
Diskussionsbeiträge	5.015	819	294	5,86%
Stichprobe 5000er	45.588	6.884	3.106	6,81%

legquote zwischen 7,49% bei 993 Treffern für die Lexikoneinträge und 11,7% für die 1.025 Treffer aus dem Korpus der Plenarprotokolle. Hier können die Unterschiede also nicht durch die Treffermenge an sich erklärt werden, sondern nur durch einen genaueren Blick auf die Treffer und die Kriterien, aufgrund derer nur ein Bruchteil davon als Belege Eingang in die Untersuchung findet. Am Beispiel der Textformen Lexikoneinträge und Plenarprotokolle zeigt sich dabei sehr gut, dass diese Unterschiede auf die Besonderheiten der jeweiligen Textformen zurückzuführen sind.

Von den 1.025 Treffern aus dem Korpus der Plenarprotokolle wurden fast ausschließlich Treffer aussortiert, die keine Substantive, sondern andere Wortarten (Verb, Adverb, Adjektiv) sind, nämlich 325 von 429 im Sinne der Untersuchung ungültigen Treffern. Bei den Lexikoneinträgen scheiden dagegen neben 283 Nicht-Substantiven auch 187 Treffer aus, bei denen es sich nicht um Wörter, sondern um Links, Hyperlinks oder Ziffern handelt, und weitere 52 gänzlich fremdsprachliche Treffer. Die inhärenten Eigenschaften der Textformen wirken sich also unmittelbar auf die Belegquote aus, die bei den Plenarprotokollen so hoch ist, da es sich um lineare Texte mit wenig fremdsprachlichem Anteil und vielen Substantiven handelt, während bei den Lexikoneinträgen, die einem Online-Lexikon entstammen, die Textstruktur mit zahlreichen Verweisen (Hyperlinks) und ein höherer fremdsprachlicher Anteil zu mehr falsch positiven Treffern führen.

Die niedrigsten Belegquoten haben mit Mail, SMS, Kriminalromanen und Diskus-

sionsbeiträgen diejenigen Textformen die entweder als Kommunikationsform dialogisch angelegt sind (Mail, SMS und Diskussionsbeiträge) oder einen hohen Dialoganteil im Gesamttext aufweisen (Kriminalromane). Dass unter den Treffern der Mail- und Kriminalromanstichprobe jeweils etwa 50% der falsch positiven Treffer auf Nicht-Substantive zurückzuführen sind, deutet darauf hin, dass Substantive in diesen dialogischen Formen seltener sind. Unter den Treffern aus den SMS und den Diskussionbeiträgen macht diese Kategorie mit gut 30% aber kaum mehr aus als in den Lexikonbeiträgen oder Plenarprotokollen, stattdessen gibt es dort zusätzlich verstärkt falsch positive Treffer, die keine morphosyntaktischen Wörter sind.

Im Sinne der Ausgangsfrage nach einem Einfluss von Wortbildungsart und Textform auf die Ausprägung wortinterner Schreibvarianten an Wortbildungsgrenzen sollen auch hier die Belege entsprechend aufgeschlüsselt werden. Da sich unter 5.2 anhand der Zusammenschreibungsquote schon gezeigt hat, dass Komposita zugänglicher für Varianten abseits der Zusammenschreibung sind als Derivationen, wird im Folgenden die Zusammenschreibungsquote für die einzelnen Textformen jeweils am Beispiel von prototypischen Determinativkomposita ermittelt. Dieses Vorgehen bieten sich an, weil Determinativkomposita als produktivstes Muster eine möglichst aussagekräftige Belegzahl auch für die einzelnen Teilkorpora liefern.

Wie Abbildung 5.2 zeigt, machen prototypische Determinativkomposita in allen Textformen den größten Teil der Belege aus.⁴ Im Durchschnitt über alle Textformen sind zwei Drittel der Belege prototypische Determinativkomposita. Die nächstgrößere Gruppe abseits dieser Determinativkomposita sind, wie aus Abbildung 5.2 zu erkennen, die Suffigierungen mit einem durchschnittlichen Anteil von gut 25%. Für einen Vergleich der wortinternen Schreibvarianz über die verschiedenen Textformen hinweg eignen sich die Suffigierungen allerdings nicht, da die Belege für Suffigierungen unabhängig von der Textform, aus der sie stammen, zu 100% zusammengeschieden werden. Alle anderen Wortbildungsmuster werden als Rest zusammengefasst und machen im Durchschnitt über alle Textformen weniger als 10% der Belege aus.

Tabelle 5.10 zeigt die Belegzahlen für prototypische Determinativkomposita aufgeteilt in Zusammenschreibungen und weitere Varianten (Bindestrichschreibungen,

⁴ Die genauen Belegzahlen für die Determinativkomposita in den einzelnen Textformen liefert Tabelle 5.10.

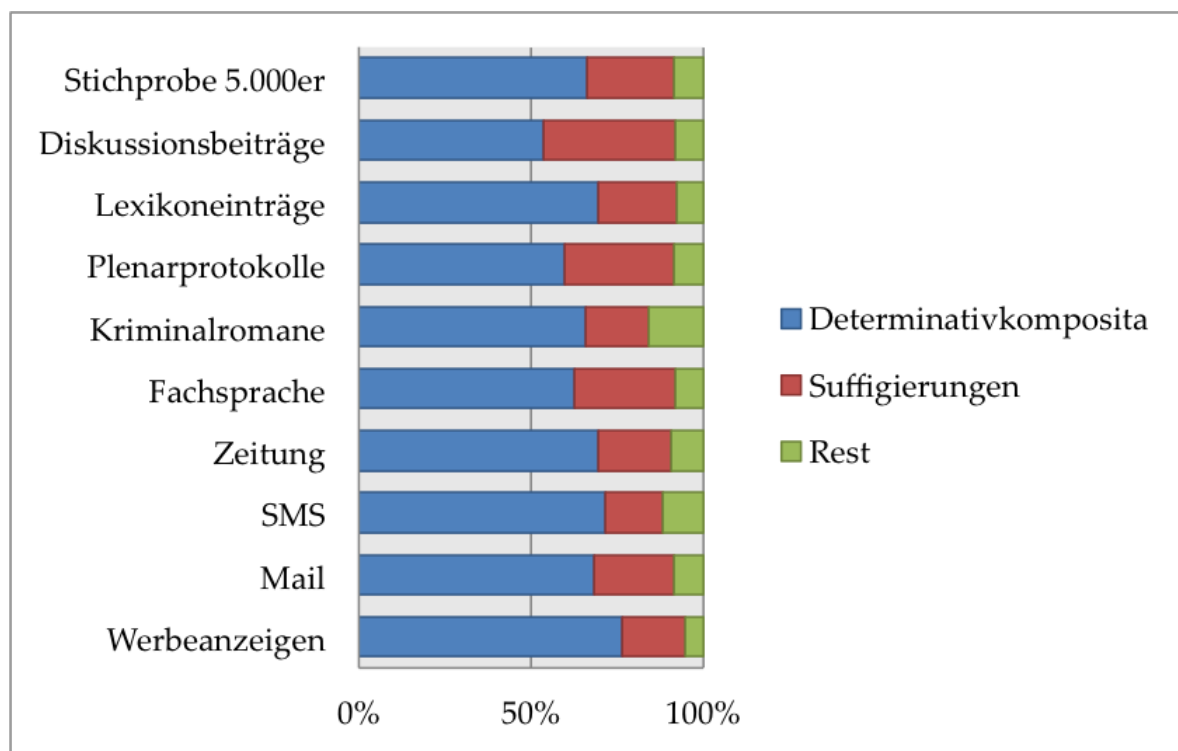


Abbildung 5.2.: Anteil der Determinativkomposita und Suffigierungen nach Textform

Getrennschreibungen, Binnenmajuskeln, usw.). Schon die Zahlen lassen erkennen, dass die Varianten abseits der Zusammenschreibung nur eine marginale Rolle spielen, obwohl die wortbildungsmorphologischen Grenzen in Determinativkomposita am zugänglichsten sind für abweichende Varianten.

Der prozentuale Anteil von Zusammenschreibung und anderen Varianten für die Gruppe der Determinativkomposita wird in Abbildung 5.3 visualisiert. Im Durchschnitt über alle Textformen werden nur 10% der Determinativkomposita nicht zusammengeschieden. Die wenigsten Abweichungen von der Zusammenschreibung finden sich unter den Belegen aus dem Plenarprotokolle-Teilkorpus. Nur sechs von 357 Belegen, also weniger als 2%, werden nicht zusammengeschieden, sondern weisen einen Bindestrich zwischen den terminalen Konstituenten der Wortbildung auf (vgl. (4)). Alle sechs Bindestrichschreibungen sind entsprechend motiviert aufgrund eines markierten Bestandteils *CCS-Technologie*, eines Eigennamens *Irland-Krise* oder der Komplexität des Kompositums *Körperschaftsteuer-Bemessungsgrundlage*.

(4) CO₂-Emissionen, Pkw-Energieverbrauchs-kennzeichnungsverordnung, CCS-

Tabelle 5.10.: Determinativkomposita und Schreibvarianz 5.000er-Stichprobe

	Belege	Zusammenschreibung	andere Varianten
Werbeanzeigen	409	347	62
Mail	96	87	9
SMS	55	51	4
Zeitung	296	245	51
Fachzeitschriften	283	259	24
Kriminalromane	137	130	7
Plenarprotokolle	357	351	6
Lexikoneinträge	264	238	26
Diskussionsbeiträge	158	140	18
Stichprobe 5000er	2.055	1848	207

Technologie, Irland-Krise, Arbeitnehmer-Entsendegesetz, Körperschaftsteuer-Bemessungsgrundlage (PBR)

Vergleichsweise viele Varianten abseits der Zusammenschreibung finden sich dagegen unter den Belegen aus den Zeitungen (17%) und Werbeanzeigen (15%). Die Werbeanzeigen entsprechen damit der Erwartung einer Variantenhäufung. Für die Zeitungstexte, die die Standardsprache repräsentieren sollen, ist diese Häufung zunächst überraschend. Aufklärung kann hier erst eine detailliertere Analyse der Belege im Zuge der Modellierung in Kapitel 6 bringen. Ein erster Blick auf die Determinativkomposita aus Werbeanzeigen und den Zeitungstexten macht aber deutlich, dass sich die Abweichungen von der Zusammenschreibung in den Zeitungsbelegen unter (5) sich fast ausschließlich auf die Bindestrichschreibung beschränken, während die Werbeanzeigen unter (6), wie erwartet, eine größere Bandbreite der Variation bieten.

In der Zeitungsstichprobe ist unter 51 Abweichungen von der Zusammenschreibung nur eine einzelne Getrennschreibung belegt, und zwar an der Kompositionsgrenze zwischen einem komplexen Eigennamen und einem fremdsprachlichen Kopf: *Hoffmann-von-Fallersleben Museum* (BRZ09). Neben Zusammenschreibungen und dieser Getrennschreibung sind sonst nur Bindestrichschreibungen belegt.

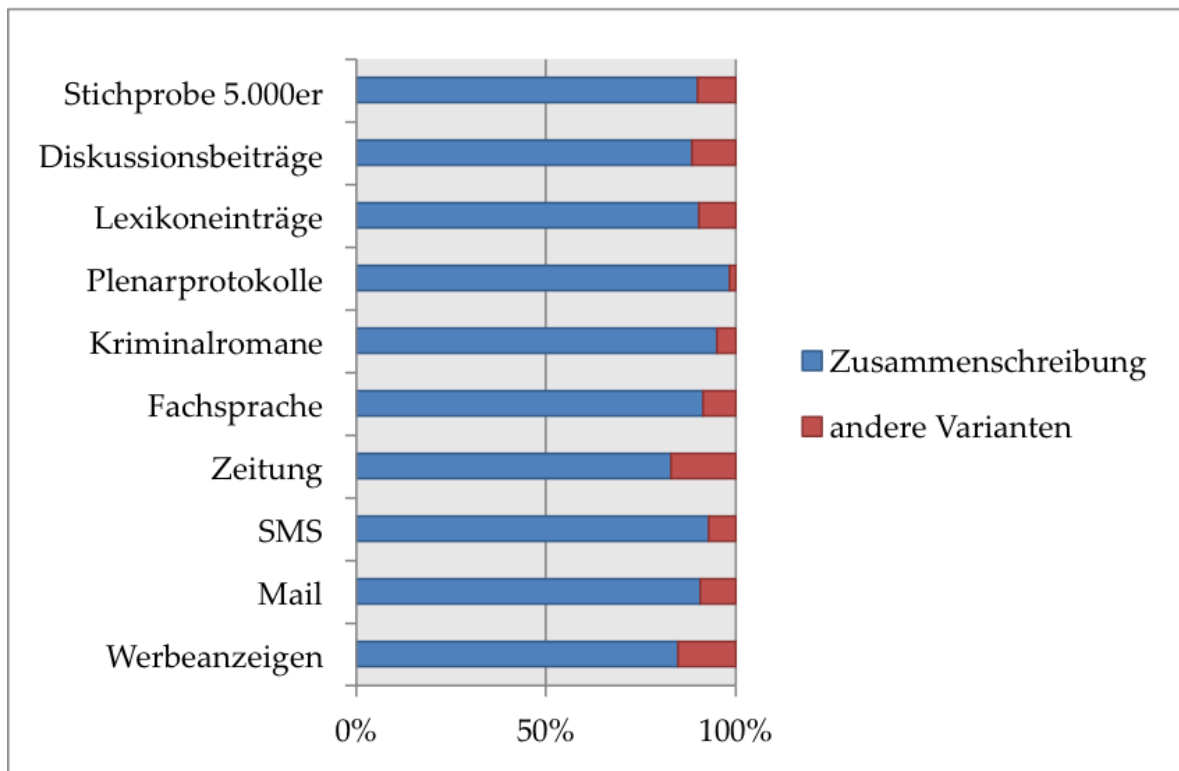


Abbildung 5.3.: Anteil der Schreibvarianten in Determinativkomposita nach Textform

- (5) a. SED-Diktatur (BRZ09), CO₂-Ausstoß (2mal), US-Quartett, FIFA-Homepage (HMP09), „Spiegel“-Reporter, SPD-Zeitung, 96-Klubchef, US-Verteidigungsminister (HAZ09), A-Junioren, D-Junioren, TSV-Amicitia-Jugend (M09) DFB-Sportdirektor (NUZ09), ETA-Mitbegründer, ETA-Ideologie, TuS-Mitglied (RHZ09)
- b. Willy-Brandt-Platz, Hertie-Stiftung, Corsa-Klasse (HMP09), Afghanistan-Konferenz, Niedersachsen-Deligierten (HAZ09), Luther-Schau, Günter-Brand-Halle, Fußball-Deutschland (NUZ09), Rapolder-Elf, Rhein-Mosel-Stadt (RHZ09)
- c. Protokoll-Abteilung, Plakat-Ausstellung, Landesliga-Volleyballerinnen (BRZ09), VERBRAUCHER-TIPP, Serien-Tickets, Ticket-Verkauf, Wind-Stadion, Senioren-Initiativen, Promi-Autopilot, Schauspielhaus-Intendant (HMP09), Abstellungs-Theater, Handball-Marathon, Zweitliga-Saison (NUZ09)
- d. Kinder-Weihnachtsfeier, Vierer-Mittelfeld, Weihnachtsmann-

- Zipfelmützen (BRZ09), Tollhaus-Stimmung, Fußball-Bund (HMP09)
- Musikschul-Geschäftsstelle (3mal) (M09), Prinzen-Landsitz (RHZ09)
- e. Murrel-Kuhle (BRZ09), Einzel-Karten (HMP09)
- f. Hoffmann-von-Fallersleben Museum (BRZ09)

Die Sortierung der Belege in (5) folgt dabei den besprochenen Einflussfaktoren auf die wortinterne Schreibung. Während die 16 Belege unter (5a) einen graphisch markierten Bestandteil aufweisen wie *96-Klubchef*, der eine Bindestrichschreibung erzwingt, sind für die zehn Bindestriche in den Belegen unter (5b) die Namen als Bestandteile verantwortlich wie in *Afghanistan-Konferenz*. Die 13 Belege unter (5c) weisen dagegen fremdsprachliche Bestandteile auf wie *Stadion* in *Wind-Stadion*. Bei sieben weiteren Belegen unter (5d) handelt es sich um vielgliedrige Komposita mit einem Bindestrich zwischen den unmittelbaren Konstituenten wie in *Weihnachtsmann-Zipfelmützen*. Allein die Bindestriche in den beiden zweigliedrigen Komposita ohne graphisch auffällige, fremde oder namenartige Bestandteile unter (5e) wirken unmotiviert.

Die Bandbreite der Variation in den Zeitungstexten ist also nicht nur fast ausschließlich auf die Zusammenschreibung und die Schreibung mit Bindestrich beschränkt, sondern die Bindestrichschreibungen sind auch (bis auf zwei Ausnahmen) durch entsprechende Faktoren motiviert.

Die 62 von einer Zusammenschreibung abweichenden Belege aus den Werbeanzeigen ergeben tatsächlich ein etwas anderes Bild hinsichtlich der Bandbreite der wortinternen Varianten und den davon betroffenen Beleggruppen unter (6).

- (6) a. 360°-Kamera, CO₂-Emissionen (7mal), CO₂-Emission (3mal), CO₂-Ausstoß, SMART-Tarif, BEST-Tarif, Methyl-4-hydroxybenzoat-Natrium, Plug-in-Hybridantrieb, Euro-5-Niveau, DAK-Spezialisten-Netzwerk, DAK-Gesundheitspaket[®] (2mal), CO₂-Emissionswert, ECE-Testzyklus
- b. Strass-Steinchen, Berenberg-Markenbotschafter
- c. ZELL-RENNANISANCE, direkt-granulat, ZELL-ERNEUERUNG, PFLEGE-SYSTEM, Apotheken-Magnesium, Park-Assistenten, Business-Klasse, Top-Konditionen, Funktionalitäts-Symbole, Qualitäts-Garantie, Top-Motorsport, Thymiankraut-Dickextrakt, Rundum-Leistungspaket, Online-Coaching, Kopfhaut-Experten

- d. PrivatKredit, FeuchtCreme
- e. Granatapfel Gesichtspflege (2mal), Commerzbank Filiale
- f. optiway-GarantiePlus-Vertrages, ForMe-Umfrage, EcoBoost-Technologie, TipMatic[®]-Getriebe, EcoBoost-Motorentechnologie, PriTarder[®]-Technik
- g. Mercedes-Benz Partner (2mal) 1.4 Turbo-Benzinmotor, Vagisan[®] Feucht-Creme (6mal), N°1 LUXUS-MARKE, Commerzbank PremiumKreditkarte, Commerzbank PremiumKonto (W)

Neben 22 Bindestrichschreibungen bei graphisch markierten Wortbestandteilen (6a), zwei bei der Beteiligung von Eigennamen (6b) und 15 mit fremdsprachlichen Bestandteilen (6c), sind auch zwei Schreibungen mit Binnenmajuskel bei fremdsprachlichen Anteilen belegt (6d), sowie drei Getrenntschreibungen (6e), die durch die höhere Zahl der Kompositionsglieder und den Eigennamen im Fall von *Commerzbank Filiale* begünstigt wurden. Weitere Binnenmajuskeln und Getrenntschreibungen zeigen sich kombiniert miteinander wie in *Commerzbank PremiumKreditkarte* oder in Kombination mit einem Bindestrich (*Mercedes-Benz Partner*, *EcoBoost-Technologie*) in (6f) und (6g) in Komposita mit mindestens drei Kompositionsgliedern. Dabei umfassen alle zwölf Belege unter (6g) entweder graphisch markierte Bestandteile oder Eigennamen. In den sechs Belegen unter (6f), sorgt jeweils die Binnenmajuskel an der mittelbaren Kompositionsgrenze dafür, dass an der unmittelbaren Grenze zwischen den Konstituenten des Kompositums ein Bindestrich eingesetzt wird. Dabei wird von allen Belegen unter (6f) und (6g), die verschiedene Varianten abseits der Zusammenschreibung aufweisen, die von Scherer (2012) angenommene Abschwächung der Markierungstärke von der Grenze zwischen den unmittelbaren Konstituenten zu den mittelbaren Grenzen eingehalten.⁵

Auch wenn die Bandbreite der wortinternen Varianten in den Belegen der Werbeanzeigen-Stichprobe, wie erwartet, größer ist als in standardnahen Zeitungstexten, sind die wortinternen Schreibungen in den Werbeanzeigen keineswegs unsystematisch. Ganz im Gegenteil können alle Abweichungen von der Zusammenschreibung auf eine Motivation durch entsprechende Wortbestandteile oder eine komplexe Kompositionsstruktur zurückgeführt werden.

Dieser kurze Einblick in einen Ausschnitt der Belege für Varianten abseits der Zu-

⁵ Vgl. dazu auch 3.3.

sammenschreibung in den 5.000er-Stichproben des Werbeanzeigen- und Zeitungskorpus liefert aber erste Evidenzen für die Annahme, dass die wortinterne Schreibung und beobachtete Varianten an entsprechender Stelle keineswegs willkürlich, sondern sehr systematisch erfolgen. Die endgültige Bestätigung wird die detaillierte Analyse und Modellierung im nächsten Kapitel leisten.

6. Theoretische Modellierung

Nach einem ersten Einblick in die quantitativen Ergebnisse im vorausgegangenen Kapitel 5 erfolgt in diesem Kapitel die systematische Analyse dieser Belege im Rahmen einer optimalitätstheoretischen Modellierung.

Die Grundidee der Optimalitätstheorie, die ich hier in Anlehnung an Müller (2000) formuliere, ist eine Konkurrenz zwischen universell gültigen, aber verletzbaren Beschränkungen und die Auflösung der Konflikte zwischen den Beschränkungen durch eine entsprechende Ordnung. Regeln, die sich nicht miteinander vereinbaren lassen, hierarchisch zu ordnen, um so eine optimale Lösung zu finden, ist dabei kein spezifisch linguistischer Ansatz, sondern eine Art universelle Lösungsstrategie. Auch im Straßenverkehr, der häufig herangezogen wird, um dieses Konzept zu erklären, wird beispielsweise die Vorfahrt nach nach diesem Prinzip geregelt. Grundsätzlich gilt *Rechts vor links!*, es sei denn die Beschilderung (z.B. *Vorfahrtstraße, Stopp*), spricht dagegen. Die Beschilderung kann wiederum ausgehebelt werden durch entsprechende Lichtzeichen (also eine rote oder grüne Ampel), die letzte Instanz wäre aber ein Polizist, der den Verkehr und damit auch die Vorfahrt entsprechend regelt. Die Beschränkungen durch die Ampel, die Beschilderung und den Grundsatz *Rechts vor links!* gelten zwar auch dann noch, der Verstoß gegen diese Beschränkungen wiegt aber geringer als der gegen die Anordnung des Verkehrspolizisten. Im Straßenverkehr gilt entsprechend die Beschränkungsordnung in (1).¹

(1) POLIZIST » AMPEL » VERKEHRSSCHILD » RECHTS VOR LINKS!

Wer tatsächlich Vorfahrt hat, ergibt sich also erst in Abhängigkeit von der Situation und der Relevanz der Beschränkungen für diese Situation. Es kann also sein, dass jemand Vorfahrt hat, obwohl er von links kommt und ein Stoppschild überfährt, weil

¹ Wesentlich detaillierter anhand von den Paragraphen der StVO entsprechend formulierten Beschränkungen führt Müller (2000: 5-8) diesen Vergleich.

ihm (und sonst keinem Kandidaten) eine grüne Ampel freie Fahrt signalisiert.

Dasselbe gilt in der Linguistik: Ein sprachlicher Ausdruck oder seine Repräsentationsebene sind nicht aufgrund inhärenter Merkmale optimal, sondern stellen sich erst im Vergleich zu anderen Kandidaten als optimale Kandidaten heraus, sofern kein anderer Kandidat ein besseres Beschränkungsprofil, also lediglich Verstöße gegen in der Hierarchie weiter unten angesiedelte Beschränkungen, aufweist.

Hier erarbeite ich dementsprechend Beschränkungen und eine hierarchische Ordnung dieser Beschränkungen mit dem Ziel, die Schreibvarianten an den wortbildungsmorphologischen Grenzen der Korpusbelege mit einem möglichst hohen Deckungsgrad als optimale Kandidaten abbilden zu können. Berücksichtigt werden dabei neben den inhärenten Merkmalen der Korpusbelege (Fremdsprachlichkeit, Wortstruktur, Wortbildungsmuster usw.), die den Kernbereich der Untersuchung ausmachen, auch der pragmatische Aspekt der Textformspezifik und der Lexikalisierungsstatus der Belege als semantisches Kriterium.

Im Folgenden thematisiere ich vor dem Einstieg in die Modellierung die Eignung eines optimalitätstheoretischen Ansatzes im Rahmen einer graphematischen Untersuchung in 6.1 und erläutere in diesem Zusammenhang auch das Kandidatenkonzept dieser Untersuchung. Da ich in Kapitel 5 zeigen konnte, dass Determinativkomposita nicht nur den größten Teil der Belege ausmachen, sondern auch mit Abstand die größte Bandbreite wortinterner Schreibvarianten bieten, konzentriere ich mich zunächst in 6.2 auf die Entwicklung eines Modells für die Kompositionsgrenze in zweigliedrigen Determinativkomposita. Anschließend nutze ich die dort formulierten Beschränkungen auch im Rahmen einer Modellierung für Derivationsgrenzen in 6.3, für die weniger Belege zur Verfügung stehen, die auch nur in Ausnahmefällen von der Zusammenschreibung abweichen, und kann unter 6.4 zeigen, dass das für die Kompositionsgrenzen entwickelte Modell sich auch für die Anwendung auf Derivationsgrenzen eignet. Darüber hinaus erweist sich das Modell in 6.5 auch – abgesehen von den SMS – für alle Textformen ohne zusätzliche Beschränkungen oder Eingriffe in die Beschränkungsordnung als tragfähig. Die Allgemeingültigkeit des Modells für den substantivischen Bereich belegt schließlich in 6.6 die Anwendung auf alle Wortbildungsgrenzen der 3.106 komplexen Substantive aus den 5.000er-Stichproben inklusive komplexere Komposita und bis dahin unberücksichtigter Wortbildungsmuster.

6.1 Optimalitätstheorie und Graphematik

Die Idee, die ursprünglich in der Phonologie begründete Optimalitätstheorie (vgl. Prince & Smolensky 1993) auch auf graphische Repräsentationen zu übertragen, ist nicht neu. Seit der Jahrtausendwende sind schon einige Arbeiten zu graphematischen oder orthographischen Phänomenen vorgelegt worden, die mit dem Prinzip der verletzbaren, aber hierarchisch geordneten Beschränkungen arbeiten, um so den besten aus einer Gruppe möglicher Kandidaten zu ermitteln.

So liegen aktuell Arbeiten zur Laut-Buchstaben-Zuordnung (Wiese (2004) mit Berücksichtigung der Stammkonstanz und Geminaten), zu den Schreibgeminaten im Besonderen (Sternefeld 2000), zur strukturellen Analyse der Buchstaben des römischen Alphabetes (Primus 2004), zur Worttrennung am Zeilenende (Geilfuß-Wolfgang 2002, 2007b),² zur Getrennt- und Zusammenschreibung (Jacobs 2005), zur typologischen und historischen Variation im System der Zeichensetzung mit Fokus auf Kommata (Primus 2007), zur satz-internen Großschreibung (Hübl & Steinbach 2011) und zur wortinternen Schreibung von N+N-Komposita (Scherer 2012) vor. Die Eignung graphematischer Themen als Anwendungsgebiet für die Optimalitätstheorie wird dabei nicht immer explizit thematisiert, sondern zumeist implizit durch die gelungene Anwendung gezeigt.

Wiese (2004) leitet seine Arbeit zur Laut-Buchstaben-Zuordnung mit der Feststellung ein, dass Schreibsystem und Orthographie in der Regel Prinzipien folgen, die miteinander in Konflikt stehen, und Optimalität daher in diesem Zusammenhang eine prominente Rolle spielt, wodurch eine optimalitätstheoretische Herangehensweise naheliegt:

Optimization in a very general and somewhat vague sense plays a prominent role in many discussions of writing systems and orthography. In theoretically oriented research, writing systems, that is, fundamental systematic properties of the written modality in a language, and orthographies, that is, the explicitly regulated details within a writing system, are often supposed to follow general preferences, though exceptions to these obviously exist, and the principles are often in conflict with each other;

(Wiese 2004: 305)

² Außerdem gibt es noch einen Aufsatz von Geilfuß-Wolfgang (2007a) zur *Stammkonstanz ohne Stützform*, der mit dem Modell der optimalen Paradigmen arbeitet.

Primus (2004: 6) schickt ihren Ausführungen zur Gestalt der Buchstaben des Modernen Römischen Alphabets und den damit verbundenen Funktionen in Übereinstimmung mit dem hier zugrundeliegenden Graphematikverständnis die grundsätzliche Annahme voraus, dass geschriebene Sprache, gesprochene Sprache und Zeichensprache denselben tiefgreifenden, sprachlichen Beschränkungen unterliegen. Diese werden lediglich durch verschiedene Systeme kodiert, so dass sich Ähnlichkeitsbeziehungen auf der Ebene der Derivate auch durch „*modalitätsübergreifende Strukturprinzipien*“ ergeben (Korrespondenzhypothese).³ Graphematische Repräsentationen sind also nicht einseitig von phonologischen abgeleitet, vielmehr müssen auch graphematisch basierte Ableitungsregeln angenommen werden „*sowie bidirektionale, d.h. eindeutige Korrespondenzen*“ zwischen den Systemen (Primus 2003: 5), so dass Merkmale der jeweiligen Repräsentationen auf ein gemeinsames Strukturprinzip oder auf Korrespondenzen zwischen den Systemen zurückgehen können.⁴ Mit der Annahme dieser Korrespondenzhypothese liegt es sehr nahe, eine Theorie, die erfolgreich auf phonologische Phänomene angewendet wurde, auch auf graphematische zu übertragen.

Am Beispiel der Buchstabengestalt des Modernen Römischen Alphabets zeigt Primus (2004) dann eingängig, dass die Optimalitätstheorie sich sehr gut eignet, um die Einflüsse der modalitätsübergreifenden Tiefenstrukturen und auch Innergraphematisches modellieren zu können, und kommt zu dem Schluss:

Optimality Theory turned out to be congenial to the purpose of revealing the systemacity of the MRA-system. The best results are obtained by competing, violable and ranked constraints. There are both idiosyncratic, i.e. inexplicable, and systematic, i.e. explicable violations. (Primus 2004: 270)

Sternefeld (2000) will ebenfalls durch die praktische Anwendung belegen, dass die Optimalitätstheorie sich für graphematische Phänomene anbietet. So zeigt er, „*dass die verschiedenen Bedingungen für die Geminatenschreibung im Deutschen gerade so organisiert sind, wie dies für die OT charakteristisch ist, und dass deshalb die Graphematik ein prototypisches Anwendungsgebiet für die OT ist*“ (Sternefeld 2000: 35). Auch er

³ Diese Hypothese entwickelt Primus (2000) bereits in ihrem Artikel zur suprasegmentalen Graphematik und Phonologie und untermauert sie später am Beispiel der Silbe in Laut-, Schrift- und Gebärdensprache (Primus 2003).

⁴ Zum dieser Arbeit zugrundeliegenden Graphematikverständnis und der Korrespondenzhypothese vgl. Kapitel 3.1.

macht deutlich, wie eng Phonologie und Graphematik miteinander verknüpft sind; beginnend mit dem Hinweis auf das Wesen von Alphabetschriften wie dem Deutschen, deren Grundprinzip es ist, lautlichen Einheiten Buchstaben oder Buchstabenkombinationen zuzuordnen und zwar mit Hilfe von Regeln zur Graphem-Phonem-Korrespondenz. Diese haben wiederum in der Optimalitätstheorie den Status von Treuebeschränkungen und entsprechen den Prinzipien der Maximalität, Dependenz und Korrespondenz (vgl. Sternefeld 2000: 36).

Diese Ableitbarkeit scheint zunächst dem Konzept von Primus (2004) zu widersprechen,⁵ bei genauerer Betrachtung sind die beiden Ansätze aber – trotz sich widersprechender Benennung – vergleichbar. Für die Phonem-Graphem-Interaktion schlägt Sternefeld zwar ein derivationales Modell vor, in dem phonologische Regeln des Typs A (Silbifizierung und Akzent), die auch graphierelevant seien, ausgehend von lexikalischen Einheiten einen „SPELLOUT“ erzeugen, aus dem dann wiederum durch nicht-graphierelevante, phonologische Regeln des Typs B (z. B. Assimilation und ?-Epenthese) die phonetische Repräsentation abgeleitet werde, oder aber mit Hilfe graphematischer Regeln die graphematische Repräsentation dieses „SPELLOUTS“ (Sternefeld 2000: 38). Diese Zwischenstufe des SPELLOUTS muss Sternefeld (2000) zu Ungunsten einer direkten Ableitbarkeit von der Schreibung aus der Lautung annehmen, weil es ebendiese im Deutschen trotz Alphabetschrift nicht gibt. Primus könnte ihm also entgegen, dass er gar kein derivationales Modell annimmt, sondern sein SPELLOUT eigentlich schlicht ein modalitätsunabhängiger Input für die verschiedenen Medien ist und die phonologischen Regeln des Typs B und die graphematischen Regeln, die er annimmt, dem Einfluss dieser Sprachmedien geschuldet sind.⁶

Jacobs (2005) formuliert den Begriff des *lautsprachlichen Zeichens* als Ausgangspunkt seiner theoretischen Modellierung. Ebenso wie der Begriff des SPELLOUTS bei Sternefeld (2000) impliziert dieser Begriff eine Ableitbarkeit der graphematischen Repräsentation, allerdings nicht nur auf Basis der phonologischen, sondern auch

⁵ Zur Diskussion um Ableitbarkeitshypothese, Autonomiehypothese und Korrespondenzhypothese (als gemäßigte Variante der Autonomiehypothese) vgl. Neef & Primus (2001).

⁶ Da Sternefeld (2000: 36) die Zuordnung von Graphemen zu Phonemen zwar als Grundprinzip einer Alphabetschrift bezeichnet, aber darüber hinaus auch rein graphematische Regeln annimmt, die am SPELLOUT ansetzen, kann man annehmen, dass er der Graphematik zumindest ein gewisses Maß an Autonomie zugesteht und damit wohl kein Verfechter einer starken Ableitbarkeitshypothese ist.

ausdrücklich anhand semantischer und morphosyntaktischer Aspekte:

Wie in der neueren Grammatikforschung üblich, gehe ich davon aus, daß ein lautsprachliches Zeichen (ein Wort, eine Phrase, ein Satz, eine Satzfolge) als Triple <PR, KR, SR> aus einer *phonologischen*, einer *kategorialen* und einer *semantischen Repräsentation* dargestellt werden kann. Die PR spezifiziert die Lautsegmentfolge, ihre prosodische Gliederung in Silben, Füße, phonologische Wörter usw. sowie die grammatisch relevanten Aspekte von Betonung und Intonation. [...]

Die KR gliedert das jeweilige Zeichen in Konstituenten, die nach syntaktisch oder morphologisch relevanten Formmerkmalen kategorisiert sind.[...]

Die SR charakterisiert die durch das Sprachsystem festgelegte Bedeutung des Zeichens durch eine Kombination elementarer Prädikat-Argument-Komplexe.⁷

(Jacobs 2005: 13-14)

Diesem Triple wird in verschrifteten Sprachen nach Jacobs (2005: 14) schließlich noch eine „*graphische Repräsentation GR*“ zugeordnet.

Diese Zuordnung und die Begriffswahl an sich implizieren wiederum eine einseitige Ableitbarkeit der graphematischen Repräsentation, ohne Einflussmöglichkeit dieser Ebene auf alle anderen. Jacobs (2005) steht damit in der Tradition der Dependenzhypothese, nimmt aber eine gemäßigte Position ein. Ein Vorteil gegenüber Sternefeld (2000) ist beispielsweise die ausdrückliche Berücksichtigung der kategorialen Ebene.

Für diese Untersuchung entwickle ich Jacobs' (2005) Konzept des lautsprachlichen Zeichens im Sinne der Korrespondenzhypothese nach Primus (2003) weiter und nehme dazu das Konstrukt eines *sprachlichen Zeichens* an, das einem Quadrupel aus phonologischer, kategorialer, semantischer Repräsentation und graphematischer Repräsentation, <PR, KR, SR, GR>, entspricht, dem dann konkrete laut-, schrift- oder auch zeichensprachliche Realisierungen zugeordnet werden können.

Im Vergleich zu Jacobs (2005) wird damit nicht nur der Begriff des *lautsprachlichen Zeichens* ersetzt, sondern auch der der *graphischen Repräsentation* durch den der *graphematischen Repräsentation*. Dies geschieht in erster Linie, um Unklarheiten zu vermeiden, da *graphisch* einerseits in der Alltagssprache häufig auch Typographisches (Schriftart, Satz, Farbe) meint, was aber unwesentlich für die Theorie eines

⁷ Ausgelassen wurde die Exemplifizierung der Repräsentationen am Beispiel der Verbalphrase *Tee trinken* nach Jacobs (2005).

sprachlichen Zeichens ist, und andererseits im wissenschaftlichen Diskurs mitunter *graphematisch* und *graphisch* analog zu *phonologisch* und *phonetisch* dazu genutzt werden, abstrakt theoretische und konkrete Aspekte voneinander zu trennen.

Die graphematische Repräsentation als Bestandteil des sprachlichen Zeichens umfasst hier die für die linguistische Theorie relevanten Informationen zu Graphemfolge (inklusive Interpunktion und Spatiensetzung) und der Gliederung in graphematische Silben.⁸

Das Konzept des sprachlichen Zeichens als <PR, KR, SR, GR>-Quadrupel bietet dann auch die theoretische Grundlage, um zu erklären, wie optisch-visuelle Eigenschaften der Schrift auf die gesprochene Sprache zurückwirken. Dürscheid (2016: 40 f.) führt unter anderem die *Spelling Pronunciation* als Hörbarmachung von Buchstaben in der Lautung bei Abkürzungen als ein solches Phänomen an (*WDR = We-de-er*, *FAZ = Ef-a-zet*).

Da ich mit Fuhrhop & Peters (2013: 207f.) davon ausgehe, dass es sich bei Groß- und Kleinbuchstaben nicht um unterschiedliche Grapheme handelt, sondern um Allographen, ist die GR der sprachlichen Zeichen in dieser Hinsicht nicht näher spezifiziert. Ob ein in der GR festgelegtes Graphem als Groß- oder Kleinbuchstabe realisiert wird, ist allerdings für jedes sprachliche Zeichen durch die morphosyntaktischen Eigenschaften in der KR bereits festgelegt.

Die Kandidaten für die optimalitätstheoretische Modellierung sind dementsprechend die schriftsprachlichen Oberflächenstrukturen der so definierten sprachlichen Zeichen. Die Kandidatenmenge ist durch die in der Belegsammlung erfassten komplexen Substantive als Ganzes bzw. ihre wortbildungsmorphologischen Grenzen vorgegeben.

Die Modellierung verfolgt dabei nicht nur das Ziel, einen optimalen Kandidaten zu ermitteln, sondern auch Abstufungen der Optimalität von Kandidaten sichtbar zu machen, so dass im Ranking auch deutlich wird, welche Varianten abseits der optimalen, am weitesten verbreitet sind. Diese Annahme einer abgestuften Optimalität steht einem Grundgedanken der Optimalitätstheorie entgegen, der Grammatikalität mit Optimalität gleichsetzt, so dass alle nicht-optimalen Kandidaten ohne Rücksicht

⁸ Bei Jacobs (2005: 14), der seine dem lautsprachlichen Zeichen zugeordnete *graphische Repräsentation* genauso definiert, könnte es also zu Verwirrungen führen, dass auch die abstrakte Struktur der Schreibsilben Bestandteil der graphischen Repräsentation ist, ohne in der Schriftsprache abgebildet zu werden.

auf die Unterschiede in den Beschränkungsprofilen als ungrammatisch gelten (vgl. Müller 2000: 9). Der Schreibung an wortbildungsmorphologischen Grenzen wird diese Schwarz-Weiß-Unterscheidung, wie die Analyse zeigen wird, aber nicht gerecht. Es gibt zwar starke Evidenzen für die jeweils optimalen Kandidaten unter den Belegen, aber auch durchaus systematisch Belege für eine zweitbeste Wahl, deren Einstufung als ebenso ungrammatisch wie alle anderen suboptimalen Kandidaten damit fraglich ist. Womöglich handelt es sich hier um eine Besonderheit im Schriftsystem, beeinflusst durch fakultativ anwendbare orthographische Normen.⁹ Im Folgenden werden anhand der Belege für die Kompositionsgrenze in zweigliedrigen Determinativkomposita die grundlegenden Beschränkungen und ein erstes Basis-ranking erarbeitet, bevor die Anwendbarkeit dieser Beschränkungsordnung auf andere Wortbildungsgrenzen, unabhängig von der Textform und auch in komplexeren Wortbildungsmustern demonstriert wird.

6.2 Modellierung für Kompositionsgrenzen in Determinativkomposita

Als empirische Basis dienen mir zunächst 1.653 substantivische Determinativkomposita mit zwei Kompositionsgliedern, die die Recherche in den 5.000er Stichproben (in Summe also ca. 45.000 Textwörter verteilt auf die neun Textformen) ergeben hat. Da für diese Stichprobengröße alle Zeichenketten mit mehr als zehn Zeichen erfasst wurden, ist es nur auf dieser Basis möglich, die dominierende Zusammenschreibung in Relation zur Schreibung mit Bindestrich und anderen Varianten zu sehen. Die Gruppe der zweigliedrigen Determinativkomposita eignet sich dabei besonders als Grundlage der Theoriemodellierung, da es sich, wie in Kapitel 5 bereits gezeigt, mengenmäßig um die größte Kompositionsgruppe handelt und zusätzlich für dieses Wortbildungsmuster der höchste Anteil an Varianten abseits der Zusammenschreibung belegt ist. Außerdem finden sich substantivische, zweigliedrige Determinativ-

⁹ Auch Scherer (2012) arbeitet im graphematischen Bereich mit einem abgestuften Optimalitätsbegriff, kommentiert dieses Vorgehen aber nicht weiter. Auch im syntaktischen Bereich wird die Optimalität mitunter in Markiertheitsstufen abgebildet. Vgl. dazu am Beispiel des Scrambling-Kriteriums Müller (2000: 241-249).

Tabelle 6.1.: Datenausschnitt: Modellierung Kompositionsgrenzen nach Textform

Textform	zweigliedrige Determinativkomposita
Werbeanzeigen	318
Mail	81
SMS	48
Zeitung	238
Fachzeitschriften	226
Kriminalromane	118
Plenarprotokolle	293
Lexikoneinträge	206
Diskussionsbeiträge	125
Stichprobe 5000er	1.653

komposita in allen neun Teilkorpora (vgl. die absoluten Häufigkeiten in Tabelle 6.1). Damit bleibt zunächst der Einfluss des Wortbildungsmusters und der Textform zugunsten einer Konzentration auf die wortimmanenten Eigenschaften außen vor. Im nächsten Schritt wird dann überprüft, inwiefern das so entstandene Ranking auf andere Wortbildungsarten und -muster übertragbar (vgl. 6.3) und für die verschiedenen Textformen passend ist (vgl. 6.5), bzw. ob der Einfluss von Textform und Wortbildungsmuster in dieses Basisranking integriert werden kann. Alternativ wäre auch eine Entwicklung des Ausgangsmodells an einer standardnahen Textform wie den Zeitungstexten denkbar gewesen. Eine textformübergreifende Modellierung bietet aber die breitere Datenbasis und darüber hinaus im Ergebnis ein potentiell auf das gesamte schriftsprachliche Deutsche anwendbares Modell.

Ich fasse im Zuge der Modellierung zunächst pro Wort nur eine einzige morphologische Grenze ins Auge, nämlich die zwischen den beiden Kompositionsgliedern als unmittelbare Konstituenten, während andere Grenzen, wie beispielsweise die Derivationsgrenze vorm Suffix *-ung* in *Genesungswünsche*, zunächst der Übersichtlichkeit halber unberücksichtigt bleiben, also auch nicht zu Verstößen gegen Beschränkungen in den Tableaus führen. Im letzten Modellierungsschritt in Kapitel 6.6 werden dann aber alle Wortbildungsgrenzen der Belege berücksichtigt.

Um zu gewährleisten, dass beim sukzessiven Aufbau der Beschränkungsordnung mit der steigenden Zahl der Beschränkungen nicht der Überblick verloren geht, fin-

det sich im Anhang A eine vollständige Liste aller Beschränkungen, die Teil der finalen Beschränkungsordnung sind. Diese Liste liegt außerdem als Einlegeblatt bei, das man beispielsweise bei der Lektüre der Tableaus zur Hand nehmen kann, um ohne umständliches Blättern die Benennungen aufzulösen oder die genauen Formulierungen parat zu haben.

Datengrundlage für die ersten Modellierungsschritte sind also 1.653 Kompositionsgrenzen zweigliedriger Determinativkomposita mit substantivischem zweiten Bestandteil. Die Morphosyntax der Erstglieder variiert. Neben N+N-Komposita wie *Priesterseminar* (WPD11), die die Mehrheit der Belege ausmachen, finden sich auch beispielsweise V+N-Komposita wie *Bedienelemente* (VDI09) oder A+N-Komposita wie *Intensivpflege* (W).

Die Belege sind nicht lemmatisiert. Die Frequenzen der Einzelbelege sprechen aber dafür, dass dadurch keine Verzerrung der Ergebnisse zu erwarten ist, weil etwa zwei Drittel der Belege singular sind und nur 13 Wortformen mehr als fünfmal belegt sind. Neben der Zusammenschreibung an 1.545 Kompositionsgrenzen in zweigliedrigen Determinativkomposita gibt es 108 Belege, die dort eine andere Variante aufweisen. Dementsprechend modelliere ich zunächst auf breiter Basis die Zusammenschreibung als Normalfall unter 6.2.1 und ergänze anschließend das so entstandene Basisranking unter 6.2.2, um auch die 108 Varianten abseits der Zusammenschreibung abbilden zu können.

6.2.1 Die Zusammenschreibung als Normalfall

Da, wie oben erwähnt, schon einige optimalitätstheoretische Arbeiten im orthographischen bzw. graphematischen Bereich vorliegen und die Beschränkungen universell gelten sollen, liegt es nahe, zu prüfen, ob bereits formulierte Beschränkungen auch auf den untersuchten Bereich angewendet werden können. Fündig geworden bin ich bei Sternefeld (2000) und Jacobs (2005).

Zusammen- vs. Getrennschreibung

Jacobs (2005) führt zur Modellierung der Getrennt- und Zusammenschreibung im Deutschen die beiden Beschränkungen ZUS-MORPH und GETR-AUSDR, die auch für die wortinterne Schreibvariation grundlegend sind.

(2) ZUS-MORPH

ZUSAMMENSCHREIBUNG VON MORPHOLOGISCHEN BILDUNGEN

Wenn X und Y Teilausdrücke eines morphologisch gebildeten Zeichens sind, gibt es zwischen <X> und <Y> kein Spatium. (Jacobs 2005: 34)

(3) GETR-AUSDR

GETRENNTSCHREIBUNG VON TEILAUSTRÜCKEN

Wenn X und Y Teilausdrücke sind, liegt zwischen <X> und <Y> mindestens ein Spatium. (Jacobs 2005: 97)

Mit *Teilausdruck* oder *Ausdruck* prägt Jacobs (2005) einen neuen Begriff als Bezugsgröße für seine Beschränkungen:

Abschnitte von lautsprachlichen Zeichen, die in einem weiten Sinne Bedeutung tragen und nach Aussage der KR Konstituenten des Zeichens sind, bezeichne ich als *Teilausdrücke* (manchmal auch einfach als Ausdrücke). [...] Z. B. hat die Verbalphrase *Tee trinken* die Teilausdrücke *Tee*, *trinken*, *trink* und *-en*.

(Jacobs 2005: 16)

Nach dieser Definition sind auch Affixe Teilausdrücke innerhalb morphologischer Bildungen, die Jacobs (2005) über morphologische Bildungsmuster definiert.

MORPHOLOGISCHE BILDUNGEN (MB)

Ein lautsprachliches Zeichen Z ist morphologisch gebildet genau dann, wenn der letzte Schritt der Derivation von Z einem morphologischen Bildungsmuster entspricht.¹⁰

(Jacobs 2005: 38)

Bildungsmuster sind dabei nach Jacobs (2005: 35) als grammatische Regeln zu verstehen, die den syntaktischen oder eben auch morphologischen Aufbau von lautsprachlichen Zeichen aus anderen lautsprachlichen Zeichen „spezifizier[en]“, indem sie die Relation zwischen <PR, KR, SR>-Tripeln charakterisieren und so die Bildung neuer Zeichen bzw. die Analyse vorhandener Zeichen im theoretischen Rahmen der „relationale[n] Morphologie“ ermöglichen. In diesem Rahmen sind Affixe und affixähnliche Ausdrücke „keine im Lexikon gespeicherten lautsprachlichen Zeichen, sondern

¹⁰ Nach Jacobs (2005: 35) umfasst die Menge der morphologischen Bildungsmuster „die traditionell der Morphologie zugeordneten Bildungsweisen der Komposition, der expliziten Derivation, der Konversion und der synthetische[n] Flexion“.

werden erst durch das Bildungsmuster eingeführt“ (Jacobs 2005: 37).¹¹

Durch die Formulierung von ZUS-MORPH und GETR-AUSDR und seine Definition von Teilausdrücken und morphologischen Bildungen macht er deutlich, dass er für den Bereich der Getrennt- und Zusammenschreibung die Getrenntschreibung von Teilausdrücken als den unmarkierten Fall betrachtet, während die auf morphologische Bildungen beschränkte Zusammenschreibung von Teilausdrücken für ihn und andere den markierten Fall darstellt (vgl. Jacobs 2005: 97).

Der historischen Entwicklung unserer alphabetischen Verschriftlichung von Sprache an deren Anfang die *Scriptio Continua* – ohne jegliche Spatien, also der Zusammenschreibung als absoluten Default – stand, trägt eine weitere Beschränkung Rechnung.

(4) *SPAT

SPARSAME SPATIENSETZUNG

Graphemfolgen enthalten kein Spatium.

(Jacobs 2005: 127)

Mit dem Ende der *Scriptio Continua* ist *SPAT keineswegs irrelevant geworden, sondern verhindert auch als vielleicht niedrigste Beschränkung innerhalb der Hierarchie noch immer das willkürliche Setzen von Spatien zwischen Graphemen abseits der Grenzen von Teilausdrücken. Lediglich für diesen Grenzbereich kann sie im Gegenwartsdeutschen durch die übergeordnete Beschränkung GETR-AUSDR ausgehebelt werden, sofern nicht ZUS-MORPH als wiederum übergeordneter Reparaturprozess greift. Aus diesen drei Beschränkungen von Jacobs (2005) ergibt sich somit ein erstes Grundgerüst für die optimalitätstheoretische Modellierung der wortinternen Schreibvariation im Deutschen:

(5) ZUS-MORPH \gg GETR-AUSDR \gg *SPAT

Angewendet auf die vier am häufigsten beobachteten Fälle wortinterner Schreibvariation an wortbildungsmorphologischen Grenzen (also die Zusammenschreibung, die Verwendung einer Binnenmajuskel und der Einsatz eines Bindestrichs oder Leerzeichens) am Beispiel eines zweigliedrigen *N+N*-Kompositums mit nativen Bestand-

¹¹ Der Begriff des *Teilausdrucks* erfasst also mehr als nur Kompositionsglieder, wie sie hier im ersten Modellierungsschritt betrachtet werden, so dass er sich auch für die Modellierung zu anderen Wortbildungsmuster bzw. -arten wie der Derivation eignet.

teilen, das im Korpus aber nur als Zusammenschreibung belegt ist, führt dieses Ranking gleich zu mehreren optimalen Kandidaten:

(6)

	ZUS-MORPH	GETR-AUSDR	*SPAT
a. ☞ Honigmilch (DIV)		*	
b. ☞ HonigMilch		*	
c. ☞ Honig-Milch		*	
d. Honig Milch	*!		*

Lediglich Kandidaten, die trotz ihres Status als Wortbildung ein Spatium zwischen den Teilausdrücken aufweisen, werden als nicht-optimale Kandidaten abgestuft. Wende ich dieses Ranking auf alle 1.653 zweigliedrigen Komposita an, die die Recherche in den 5.000er Stichproben ergeben hat und die mir nun als empirische Basis der optimalitätstheoretischen Modellierung dienen, werden nur sechs Getrenntschreibungen als nicht optimale Kandidaten aussortiert:

- (7)
- a. BWL Studiums (Mail)
 - b. Hoffmann-von-Fallersleben Museum (BRZ09)
 - c. linguisten mist, USB Verlängerung (SMS)
 - d. VDI nachrichten (VDI09 (2mal))

Vor dem Hintergrund, dass der Anteil der Zusammenschreibung ohne wortinterne Interpunktionszeichen oder Majuskeln des Typs *a. Honigmilch* unter den nicht getrennt geschriebenen Komposita bei knapp 94% liegt, bildet dieses Ranking die tatsächliche Beleglage natürlich noch nicht zufriedenstellend ab. Im Idealfall sollte für den Großteil der Belege also allein die Zusammenschreibung ohne sonstige Markierung der Grenze (*Honigmilch*) als optimaler Kandidat aus der Beschränkungshierarchie hervorgehen, während die anderen Varianten abgestuft werden.

Zusammenschreibung vs. Schreibung mit Interpunktionszeichen

Um der einfachen Zusammenschreibung als optimalem Kandidat näher zu kommen, ergänze ich das Ranking nach Jacobs (2005) um eine weitere Beschränkung. Sterne-

feld (2000) hat in seinem Artikel zu den Schreibgeminaten im Deutschen nämlich eine Beschränkung zum Verhältnis zwischen phonologischer und graphemischer Repräsentation (Graphem-Phonem-Korrespondenz) entwickelt, die sich auch hier gewinnbringend anwenden lässt:

(8) GPK

a. MAXIMALITÄT:

Jedes Segment der phonologischen Repräsentation korreliert eindeutig mit einem Segment der graphemischen Repräsentation.

b. DEPENDENZ:

Jedes Segment der graphemischen Repräsentation korreliert eindeutig mit einem Segment der phonologischen Repräsentation.

c. KORRESPONDENZ:

Jede Relation erfüllt gewisse (z. B. die in Bierwisch 1972 formulierten) Graphem-Phonem-Korrespondenzen. (Sternefeld 2000: 36)

Damit formuliert Sternefeld die einer Alphabetschrift zugrundeliegenden wechselseitigen Graphem-Phonem-Korrespondenzen als optimalitätstheoretische Treuebeschränkungen, die sicherstellen, dass sich phonologische und graphemische Ebene möglichst exakt entsprechen. Für die Untersuchung der Schreibgeminaten ist die MAXIMALITÄT von besonderem Interesse, da hier ein Phonem mit zwei Graphemen zu korrespondieren scheint (vgl. Sternefeld 2000: 37). Für die Untersuchung der wortinternen Schreibvariation an Morphemgrenzen konzentrieren wir uns aber auf die DEPENDENZ, auf die im Folgenden mit GPKB Bezug genommen wird, ohne weitere Berücksichtigung von potentiellen Verstößen gegen MAXIMALITÄT oder KORRESPONDENZ.¹²

(9) GPKB

DEPENDENZ:

¹² Bei Beispielen wie *Druckerzeugnis/Druck-Erzeugnis* aus dem amtlichen Regelwerk liegt der Verdacht nahe, dass für den Glottisverschlusslaut als Segment der phonologischen Repräsentation nach der Teilbeschränkung der MAXIMALITÄT der Bindestrich als korrelierendes Element auf der graphemischen Ebene in Frage kommt. Dem steht entgegen, dass der Glottisverschlusslaut, der im Deutschen genau dann realisiert wird, „wenn ein Fuß sonst mit einem Vokal beginnen würde, gleich ob wortinitial oder wortintern“, in aller Regelmäßigkeit nicht verschriftlicht wird (Wiese 2011: 85).

Jedes Segment der graphemischen Repräsentation korreliert eindeutig mit einem Segment der phonologischen Repräsentation.

Auch Geilfuß-Wolfgang (2007a) arbeitet mit einer GPK-Beschränkung, formuliert sie aber leider nicht explizit. Dennoch wird deutlich, dass er davon ausgeht, dass Grapheme in einer Beziehung zu Phonemen stehen und GPK-Regeln diese Beziehungen bestimmen (vgl. Geilfuß-Wolfgang 2007a: 137). Seine GPK-Beschränkung entspricht also vermutlich einer Kombination aus MAXIMALITÄT und KORRESPONDENZ nach Sternefeld (2000), während die DEPENDENZ in der Lesbarkeitsbeschränkung („PGK: Jedes Graphem soll mit einem bestimmten Phon korrespondieren.“ (Geilfuß-Wolfgang 2007a: 144)) eine Annäherung findet; mit dem wesentlichen Unterschied, dass hier die graphemische mit der phonetischen Ebene korrespondiert. Mangels expliziter Formulierung kann eine GPK-ähnliche Beschränkung nach Geilfuß-Wolfgang (2007a) hier allerdings nicht berücksichtigt werden.

Unabhängig davon hat GPKB den Vorteil, dass keine Korrelation zwischen Graphem und Phonem thematisiert wird, sondern die Beziehung von Segmenten der graphemischen und phonologischen Repräsentation Gegenstand der Beschränkung ist. Diese Formulierung lässt sich somit auch auf phonologische und graphemische Elemente anwenden, die keinen Phonem- oder Graphemstatus per Definition haben oder deren Status nicht eindeutig ist. Nach GPKB dürfte es also keine Elemente der graphemischen Repräsentationen geben, denen eine segmentale Entsprechung auf phonologischer Ebene fehlt. Genau das ist aber der Fall bei der Verwendung von wortinternen Interpunktionszeichen.

Ein Ranking von GPKB zwischen ZUS-MORPH und GETR-AUSDR wie in (10) führt demnach dazu, dass Kandidaten mit Interpunktionszeichen (z. B. einem Bindestrich¹³) ein schlechteres Beschränkungsprofil (vgl. Tableau (11)) aufweisen als solche mit Binnenmajuskel oder in unmarkierter Zusammenschreibung.¹⁴

¹³ Problematisch bleibt vorerst, dass so zunächst alle Verstöße durch ein wortinternes Interpunktionszeichen gegen GPKB zu Kandidaten führen, die ebenso zu bewerten sind, wie eine Bindestrichvariante. *Honig.Milch* oder *Honig/Milch* sind auf dieser Basis also genauso gut oder schlecht zu bewerten wie *Honig-Milch*. Mehr dazu unter 6.2.2.

¹⁴ Auch ohne dass Geilfuß-Wolfgang (2007a) GPK ausformuliert, ist klar, dass seine Version der Beschränkung dies nicht leisten könnte. Da er von einer GPK-Regel-basierten Beziehung zwischen Graphemen und Phonemen ausgeht, deren relevante Domäne Morpheme seien, sind keine Aussagen

(10) ZUS-MORPH \gg GPKB \gg GETR-AUSDR \gg *SPAT

(11)

	ZUS-MORPH	GPKB	GETR-AUSDR	*SPAT
a. ☞ Honigmilch (DIV)			*	
b. ☞ HonigMilch			*	
c. Honig-Milch		*!	*	
d. Honig Milch	*!			*

Von den 1.653 erfassten zweigliedrigen Determinativkomposita aus den 5.000er Stichproben verlieren dadurch 99 Schreibungen mit Bindestrich ihren Status als optimale Kandidaten. Es bleibt aber noch Kandidat *b. HonigMilch* mit der Binnenmajuskel entgegen der Intuition und den Ergebnissen der Korpusstudie als optimaler Kandidat. Es finden sich in dieser Stichprobe nämlich nur drei Belege mit Binnenmajuskel:

- (12) a. FeuchtCreme, PrivatKredit (W)
b. LautsprecherBox (Mail)

Großschreibung auf der Wortebene

Um den Normalfall der einfachen Zusammenschreibung ohne Binnenmajuskel als einzigen optimalen Fall zu modellieren, sind also weitere Beschränkungen nötig, die auf der graphischen Ebene operieren und die Realisierung von Graphemen als Majuskeln oder Minuskeln regeln.

Dazu wird im Folgenden zuerst die Kleinschreibung als Standardfall modelliert, um im Anschluss auch die Substantivgroßschreibung innerhalb morphosyntaktischer Wörter (*HonigMilch, Honig-Milch, Honig Milch*) ins Modell zu integrieren. Dabei wird sich zeigen, dass die Substantivgroßschreibung unabhängig vom Status dieser Einheiten als morphosyntaktisches Wort nicht nur an der initialen Position graphematischer Wörter wie in *Honigmilch* und *Honig Milch* ansetzt, sondern auch systematisch die auf die erste Position von Bestandteilen graphematischer Wörter nach einem Bindestrich wie in *Honig-Milch* zugreift, während die Majuskelschreibung oh-

über Varianten an Morphemgrenzen möglich (vgl. Geilfuß-Wolfgang 2007a: 137).

ne vorausgehendes Grenzsinal in Einzelfällen wie in *HonigMilch* unsystematisch ist. In einem kleinen Exkurs diskutiere ich außerdem, ob neben dem Bindestrich noch andere Interpunktions- oder Sonderzeichen eine graphematische Eigenständigkeit von Wortbestandteilen kodieren.

Die Kleinschreibung als Standardfall Sowohl die historische Entwicklung (mit Entstehung der Minuskeln verschwinden Majuskeln nicht einfach, sondern werden funktionalisiert), als auch die orthographische Perspektive (es gibt eher Regeln zur Groß- als zur Kleinschreibung), sprechen dafür, Minuskeln als Default für die Realisierung von Graphemen anzunehmen (vgl. Hübl & Steinbach 2011: 262). Hübl & Steinbach (2011: 262) modellieren diese Erkenntnis im Zusammenhang ihrer Untersuchung zur Großschreibung im Satz zunächst durch die Beschränkung WORT-KLEIN und verweisen darauf, dass diese „Ökonomie-Beschränkung“ die wortinitiale Großschreibung blockiere, da eine einheitliche Kleinschreibung aller Wortformen schreibökonomischer ist.

(13) WORT-KLEIN

Wortformen werden kleingeschrieben. (Hübl & Steinbach 2011: 262)

Für die Untersuchung zu wortinternen Schreibvarianten greift WORT-KLEIN aufgrund der Konzentration auf die wortinitiale Position zu kurz. Um die Kleinschreibung als Standard für alle Grapheme einer Wortform unabhängig von der Position zu definieren, formuliere ich daher mit MINUS eine erste eigene Beschränkung für den unmarkierten Fall der durchgängigen Kleinschreibung von Wörtern:¹⁵

(14) MINUS

GRAPHISCHE REALISIERUNG VON GRAPHEMEN

Grapheme werden als Minuskeln realisiert.

Da diese Beschränkung keiner der bisherigen Beschränkungen widerspricht und un-

¹⁵ Historisch gesehen war zunächst die durchgängige Majuskelschreibung der Standardfall. Da es sich aber hier um eine synchrone Untersuchung handelt, wird keine entsprechende Beschränkung formuliert. Anders als die Beschränkung *SPAT, die ursprünglich als Grundlage der *Scriptio Continua* diente und heute noch dafür sorgt, dass keine Leerzeichen zwischen den Graphemen innerhalb einzelner Teilausdrücke auftreten, wäre eine solche Beschränkung durch MINUS auch komplett aufgehoben, was der Ablösung der Großschreibung durch die Kleinschreibung als Standardfall entspricht.

abhängig von ihrer Position den Default der Kleinschreibung modelliert, stelle ich sie gemeinsam mit *SPAT auf die untersten Stufe der Hierarchie in (15). Dort sorgt sie dafür, dass innerhalb der fiktiven Beispielgruppe nur noch die Zusammenschreibung *Honigmilch* als optimaler Kandidat hinsichtlich der Schreibung an der Kompositionsgrenze verbleibt. *Honigmilch* verstößt zwar wie *HonigMilch* gegen GETR-AUSDR, hat aber aufgrund des des Verstoßes von *HonigMilch* gegen MINUS das bessere Beschränkungsprofil in (16).

(15) ZUS-MORPH \gg GPKB \gg GETR-AUSDR \gg MINUS, *SPAT

Die Beispiele in Tableau (16) verdeutlichen außerdem, dass es zumindest eine weitere Beschränkung geben muss, die als Markiertheitsbeschränkung oberhalb von MINUS nicht nur die Großschreibung von Substantiven, die graphematischen Wörtern entsprechen (wie *Honigmilch* (DIV) und *Honig Milch*) lizenziert, sondern auch die von substantivischen Bestandteilen graphematischer Wörter, die mit einem Bindestrich segmentiert sind wie in *Honig-Milch*.

Belege wie *BWL Studiums* (SMS) oder *Murmel-Kuhle* (BRZ09) deuten nämlich darauf hin, dass auch wortintern die Großschreibung greift, wenn mit dem Spatium eine graphische Grenze vorliegt oder durch einen Bindestrich gegliedert wird.

(16)

	ZUS-MORPH	GPKB	GETR-AUSDR	MINUS	*SPAT
a. ☞ Honigmilch (DIV)			*		
b. HonigMilch			*	*!	
c. Honig-Milch		*!	*	*	
d. Honig Milch	*!			*	*

Die systematische Substantivgroßschreibung Auch bezüglich der Substantivgroßschreibung bietet sich zunächst ein Blick in die Literatur an, um zu prüfen, ob dort bereits formulierte Beschränkungen für die Modellierung der wortinternen Schreibvarianten genutzt werden können. Tatsächlich formuliert Scherer (2012) in ihrer Untersuchung zur Schreibung von N+N-Komposita die Beschränkung SUBST-GROSS:

(17) GROSSSCHREIBUNG VON SUBSTANTIVEN (SUBST-GROSS)

Substantive beginnen mit einer Majuskel.

(Scherer 2012: 76)

Ihre Beispiele zu dieser Beschränkung zeigen, dass sie auch nominale Zweitglieder in Komposita trotz ihrer Gebundenheit innerhalb einer ununterbrochenen Folge von Buchstabengraphemen wie graphisch freie Substantive behandelt, so dass *ReiseZentrum* SUBST-GROSS entspricht während *Reisezentrum* aufgrund von $|z|$ dagegen verstößt (vgl. Scherer 2012: 77). Scherer (2012) führt nicht weiter aus, welches Konzept von Substantivität sie dieser Beschränkung zugrunde legt. Die Bewertung von *ReiseZentrum* als SUBST-GROSS-konformes N+N-Kompositum, während *Reisezentrum* und *Reise-zentrum* jeweils einen Verstoß darstellen, liefert hierzu auch keine hinreichenden Evidenzen.

Betrachtet man wie Scherer (2012) die Teilausdrücke individuell hinsichtlich ihrer Substantivität, muss auch klar sein, wie man mit Komposita umgeht, in denen substantivische Teilausdrücke mit nicht-substantivischen kombiniert werden. Bei einem V+N-Kompositum wie *Bedienmodule* (VDI09) ist bei einer Einzelbetrachtung für das Erstglied zunächst nicht von einer Großschreibung auszugehen, da es keinen inhärenten substantivischen Charakter hat (optimal müsste also *bedienModule* sein). Erst der gesamte Ausdruck ist durch die Vererbung des nominalen Merkmals vom Kopf *Module* ein Substantiv. Dieser substantivische Gesamtstatus wird dann aber wiederum in der Regel durch eine einzelne initiale Majuskel angezeigt, und zwar nicht am inhärent substantivischen Zweitglied, sondern am nicht-substantivischen Erstglied (*Bedienmodule*). Wenn man SUBST-GROSS konsequent auch auf graphisch gebundene Teilausdrücke anwendet, wäre danach also die Schreibung *bedienModule* optimal bzw. unter Berücksichtigung der Vererbung *BedienModule*, wenn man davon ausgeht, dass *Module* seinen Substantivstatus und dessen Anzeige durch eine Majuskel während der Vererbung behält, wofür auch Scherers (2012) Beispiele *ReiseZentrum* und *Reise-Zentrum* sprechen.

Diese beiden Möglichkeiten der inhärenten oder ererbten Substantivität stehen auch ganz direkt im Konflikt, nämlich bei N+A-Komposita wie *benutzeradaptives* (VDI09). Die Anwendung von SUBST-GROSS auf das Erstglied würde zur Großschreibung von *Benutzeradaptives* führen, während der Gesamtausdruck als Adjektiv nicht unter SUBST-GROSS fällt und dementsprechend kleinzuschreiben wäre (*benutzeradap-*

tives).

Für Scherer (2012) sind weder V+N-Komposita noch N+A-Komposita relevante Problemfälle, da sie ihren Untersuchungsgegenstand stark einschränkt und ausschließlich zweigliedrige N+N-Komposita untersucht. Für die vorliegende, breiter angelegte Untersuchung kann die Beschränkung allerdings so nicht gehalten werden.¹⁶

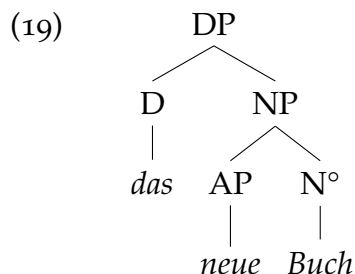
Hübl & Steinbach (2011) formulieren in ihrem Aufsatz schon ein Jahr zuvor die Beschränkung KOPF-GROSS, mit dem Ziel, in Kombination mit der oben schon besprochenen Beschränkung WORT-KLEIN die satzinterne Großschreibung zu modellieren:

(18) KOPF-GROSS

Wortformen, die den Kopf einer NP besetzen, werden großgeschrieben.
(Hübl & Steinbach 2011: 262)

Im Vergleich zu SUBST-GROSS nach Scherer (2012) sind diese beiden Beschränkungen wesentlich präziser formuliert. KOPF-GROSS legt eindeutig ein syntaktisches Konzept für die satzinterne Großschreibung zugrunde und durch die Formulierung „*Wortformen werden groß- bzw. kleingeschrieben*“ ist auch klar, dass sich die Beschränkungen jeweils nur auf den ersten Buchstaben einer Wortform und nicht darüber hinaus auch auf Anfangsbuchstaben von Teilausdrücken beziehen.

Ansatzpunkt für die satzinterne Großschreibung ist ausschließlich die Initialposition der Wortform, die „*die N°-Position innerhalb einer DP-NP-Struktur besetzt*“ (Hübl & Steinbach 2011: 259).



¹⁶ Eine alternative Modellierung unter Beibehaltung von SUBST-GROSS wäre nur durch die Formulierung einer zusätzlichen Beschränkung, die SUBST-GROSS für Fälle wie *wasserscheu* aushebeln kann, möglich. Ich zeige im Folgenden aber, dass sich die Großschreibung auch mit einer einzigen Beschränkung modellieren lässt.

In Beispiel (19) setzt KOPF-GROSS also nur an der initialen Position von *Buch* an und ist durch die Positionierung über WORT-KLEIN für das $|B|$ verantwortlich. Für die satzinterne Großschreibung ist es durchaus sinnvoll, den Wirkungsbereich beider Beschränkungen auf die erste Position einer Wortform einzugrenzen. Bezüglich der vorliegenden Untersuchung, die wortinterne Prozesse betrachtet, ist die oben formulierte, weitreichendere Beschränkung MINUS, wie bereits erläutert, besser geeignet, da MINUS auch die Kleinschreibung abseits der wortinitialen Grapheme als Default modelliert.

Mit der einfachen Ordnung von KOPF-GROSS \gg WORT-KLEIN gelingt Hübl & Steinbach (2011: 263-269) eine überzeugende Analyse für den Kernbereich satzinterner Großschreibung auf einer rein syntaktischen Basis. Eine wortinterne Großschreibung wie bei Komposita mit Bindestrich (*Reise-Zentrum*) gehört im Rahmen ihrer Untersuchung zwar nur zum Randbereich, im Sinne der Universalität optimalitätstheoretischer Ansätze machen sie sich aber dennoch die Mühe, zu überprüfen, inwiefern auch dieses und andere Phänomene mit ihrem Modell vereinbar sind (vgl. Hübl & Steinbach 2011: 269-280), und erweitern KOPF-GROSS wie folgt:

(20) KOPF-GROSS 2:

Großgeschrieben werden:

- (i) Wortformen, die den Kopf einer NP besetzen, und
- (ii) graphematisch eigenständige morphologische Köpfe einer komplexen Wortform, die den Kopf einer NP besetzen.

Durch diese Erweiterung wird auch Schreibungen wie *Reise-Zentrum* und *Füße-vom-Tisch-Ansage* Rechnung getragen. Der durch den Bindestrich graphematisch eigenständige Kopf der Wortbildung (*Zentrum*) erbt nämlich die Großschreibung des syntaktischen Kopfes *Reise-Zentrum* und die Großschreibung des ebenfalls durch Bindestriche graphematisch eigenständigen Wortbestandteils *Tisch* innerhalb des Phrasenkompositums *Füße-vom-Tisch-Ansage* ist durch KOPF-GROSS 2 (ii) lizenziert, da *Tisch* innerhalb des phrasalen Kompositionserstgliedes den Kopf einer NP besetzt.

Anhand ihrer Beispiele in (21) wird klar, dass sie eine graphematische Eigenständigkeit neben graphematischen Wörtern auch Bestandteilen markierter graphematischer Wörter zusprechen, die durch einen Bindestrich graphisch abgetrennt sind.

- (21) a. <Lotto-Annahmestelle> vs. <*LottoAnnahmestelle>
 b. <*Lotto-annahmestelle> vs. <Lottoannahmestelle>

Hübl & Steinbach (2011) verweisen bezüglich graphematisch eigenständiger Bestandteile innerhalb von graphematischen Wörtern auch auf Fuhrhop (2008) und Gallmann (1997), die sich beide dafür aussprechen, dass auch für durch Bindestriche abgegrenzte Wortteile graphematischer Wörter „graphisch ein gewisse Eigenständigkeit“ (Gallmann 1997: 11) anzunehmen ist, die dazu führt, dass auch bei Wortteilen die lexikalische Substantivgroßschreibung greifen kann (vgl. auch Fuhrhop 2008: 205-206). Die lexikalische Begründung der Substantivgroßschreibung für durch Bindestrich(e) graphematisch eigenständige Bestandteile markierter graphematischer Wörter deckt sich zwar nicht mit dem syntaktischen Ansatz zur Substantivgroßschreibung von Hübl & Steinbach (2011),¹⁷ aber in puncto graphematische Eigenständigkeit stimmen sie allesamt darin überein, dass die Großschreibung von durch Bindestriche abgetrennten Wortbestandteilen ein Indikator für die graphematische Selbstständigkeit dieser Bestandteile ist.

Schreibungen komplexer Adjektive wie *Airbus-internen* (WDD11) lassen sich aber mangels nominalen Kopfes auch mit KOPF-GROSS 2 nicht syntaktisch erklären. Hübl & Steinbach (2011: 274) verweisen zudem noch auf weitere Beispiele, die sich nicht auf Basis eines syntaktischen Großschreibungskonzeptes erklären lassen, wie *Kosten-Nutzen-Rechnung*, *Lehrer-Schüler-Verhältnis* oder *Hals-Nasen-Ohren-Arzt*, in denen als Determinans zum nominalen Kopf eine Art Aneinanderreihung vorliege. Neben den Wortformen, die den Kopf einer NP besetzen (wortinitiale Großschreibung) und den eigenständigen morphologischen Köpfen innerhalb dieser Wortformen (*Rechnung*, *Verhältnis*, *Arzt*), bleibt die Großschreibung aller anderen Bestandteile unerklärt (*Nutzen*, *Schüler*, *Nasen*, *Ohren*).

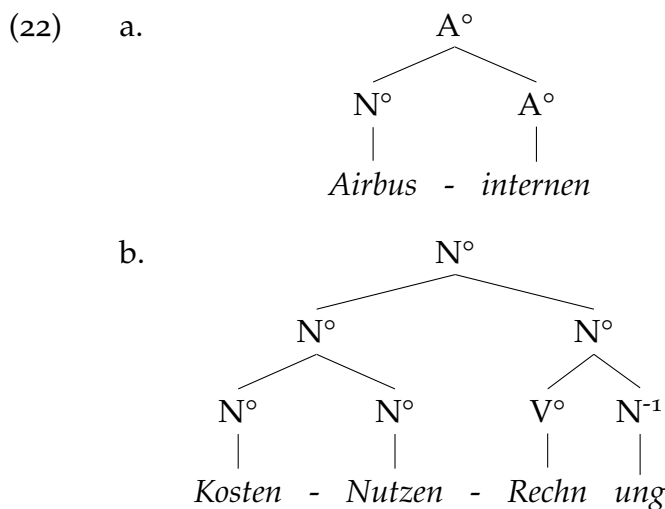
Hierarchisch strukturierte Determinativkomposita mit mehr als zwei Gliedern wie z. B. *Mindeststeuersatzproblematik* (PBR), *Hochschul-Informationen-System* (VDI09) und *Euro Behinderten-WC-Schlüssel* (RHZ09) werden von Hübl & Steinbach (2011) nicht explizit thematisiert. Aufgrund der hierarchischen morphosyntaktischen Struktur können aber alle wortinternen Großschreibungen morphosyntaktisch durch Ver-

¹⁷ Buchmann (2015: 281) geht für solchen Kontexte von einer Markierung der Wortbestandteile als potentielle Kerne von Nominalgruppen aus, und versucht so syntaktische Kriterien wortintern anwendbar zu machen.

bung erklärt werden.

Für diesen recht kleinen, verbleibenden Bereich, der nicht von KOPF-GROSS 2 erfasst wird, gehen Hübl & Steinbach (2011: 274) davon aus, dass diese Großschreibungen tatsächlich nicht syntaktisch, sondern nur lexikalisch, durch den Wortartbezug der graphematischen Wörter, beziehungsweise deren graphisch selbständiger Bestandteile, zu erfassen sind. Sie sehen darin einen Reflex der historischen Entwicklung der satzinternen Großschreibung, die einen Grammatikalisierungsprozess von der fakultativen pragmatischen über die lexikalische hin zur syntaktischen Markierung durchlaufen hat (vgl. Hübl & Steinbach 2011: 284-288).

Ich hingegen argumentiere dafür, dass auch diese potentiell lexikalischen Großschreibungen über die N°-Position erfasst werden können. Bei N+A-Komposita wie *Airbus-internen* (WDD11) und Komposita mit Aneinanderreihungen wie *Kosten-Nutzen-Rechnung* handelt es sich bei *Airbus* und *Nutzen*, deren Großschreibung von KOPF-GROSS 2 nicht erfasst wird, nämlich ebenfalls um N°-Positionen. Allerdings nicht um die N°-Position, die als Wortform nach KOPF-GROSS 2 den Kopf einer NP-Struktur besetzt, sondern um eine N°-Position unterhalb der Phrasenstruktur auf Ebene der Wortstruktur:¹⁸



Als Ganzes entsprechen die beiden Komposita *Airbus-internen* und *Kosten-Nutzen-*

¹⁸ Buchmann (2015: 281) weist darauf hin, dass es sich bei den graphematisch eigenständigen, substantivischen Erstgliedern adjektivischer Determinativkomposita wie *Airbus* in *Airbus-internen* in der Regel um Eigennamen und Fremdwörter handelt, und geht insbesondere für die Eigennamen in diesem Zusammenhang von einer Großschreibung im Sinne einer „Wortschonungsstrategie“ nach Nübling (2014) aus.

Rechnung jeweils der syntaktischen A°- bzw. N°-Kategorie, weisen aber darüber hinaus intern weitere N°-Positionen auf (vgl. (22)), da davon auszugehen ist, dass es sich auch bei Nicht-Kopf-Konstituenten in der Wortbildung um maximale Projektionen ihrer Kategorie (also X°-Positionen) handelt (vgl. Selkirk 1982: 8-9). Wie in den obigen Beispielen können also Kopf- und Nicht-Kopf-Konstituente dieselbe Kategorie aufweisen, da auf der morphologischen Ebene gilt:

$$(23) \quad X^n \rightarrow \phi Y^m \psi^{19}$$

wobei $0 \geq n$ und $n \geq m$ (Selkirk 1982: 8)

Ein morphologische Kategorie kann also zwar anders als in der Syntax nicht von einer Kategorie einer niedrigeren Stufe dominiert werden, aber von einer Kategorie derselben Stufe (und der einer höheren Stufe).

Ich formuliere daher eine neue Großschreibungsbeschränkung N°-GROSS, um den gesamten Wirkungsbereich der satz- und wortinternen Großschreibung abzudecken:

$$(24) \quad \text{N}^\circ\text{-GROSS}$$

GROSSSCHREIBUNG

Graphematisch eigenständige Wörter und Wortbestandteile, die eine N°-Position besetzen, werden großgeschrieben.

Damit fasst N°-GROSS in einer Beschränkung die Teilbereiche der satzinternen und wortinternen Großschreibung zusammen, da die beiden Teilbereiche nicht in Konkurrenz zueinander stehen, sondern sich gegenseitig ergänzen.

Ansatzpunkt für die Großschreibung ist unabhängig vom Teilbereich jeweils das initiale Buchstabengraphem eines Ausdrucks, der die N°-Position besetzt. Die Teilbereiche unterscheiden sich aber hinsichtlich der graphematischen Eigenständigkeit der großgeschriebenen Ausdrücke in N°-Positionen. Während es sich auf syntaktischer Ebene um graphematische Wörter handelt, die von Leerzeichen umgeben sind, genügt auf morphologischer Ebene schon der Bindestrich, um auch Teilausdrücke markierter graphematischer Wörter zugänglich zu machen für N°-GROSS. Diese Teilausdrücke wiederum entsprechen in der Regel graphematischen Wörtern des Kernbereichs hinsichtlich ihrer internen Struktur (ausschließlich Buchstabengra-

¹⁹ ϕ und ψ sind dabei nach Selkirk (1982: 4) „variables over category symbols“, die auch Affixe mit einschließen.

pheme, graphematische Silbe), unterscheiden sich jedoch hinsichtlich der graphischen Grenzmarkierung, die nicht beidseitig über Leerzeichen erfolgen muss, sondern auch von einem Bindestrich geleistet werden kann.²⁰

Aus N°-GROSS folgt damit unmittelbar, dass graphematische Wörter des Kernbereichs nur eine einzige Position haben, nämlich die initiale, an der diese Beschränkung ansetzen kann, wohingegen in markierten graphematischen Wörtern zusätzlich auch Positionen nach einem Bindestrich Angriffspunkte sein können. Auf Basis dieser Beschränkung lässt sich dann auch der Kontrast erklären, der durch die Beispiele unter (25) illustriert wird.

- (25) a. benutzeradaptives (VDI09)
- b. Airbus-internen (WDD11)
- c. Bedienmodule (VDI09)
- d. Abnehm-Erfolg (W)

Sowohl bei den N+A-Komposita in (25a) und (25b), als auch bei den V+N-Komposita unter (25c) und (25d) führt das Einsetzen eines Bindestriches zu Veränderungen hinsichtlich der Großschreibung nominaler Elemente.

Wie *benutzeradaptives* (VDI09) illustriert, wird das graphematische Wort des Kernbereiches als Ganzes kleingeschrieben, da es sich um eine A°-Kategorie handelt und der Teilausdruck, der eine morphologische N°-Position besetzt (*benutzer*), mangels graphematischer Eigenständigkeit keine Großschreibung lizenziert. An *Airbus-internen* (WDD11) sieht man wiederum, dass die graphematische Selbständigkeit des Teilausdrucks (*Airbus*) in der morphologischen N°-Position die Großschreibung trotz A°-Kategorie des komplexen graphematischen Wortes zulässt.

Ähnlich liegt der Fall bei den V+N-Komposita unter (25c) und (25d). In beiden Fällen wird wortinitial großgeschrieben, da das graphematische Wort als Ganzes in die N°-Kategorie fällt. Aber nur durch den Einsatz eines Bindestriches und der damit einhergehenden Eigenständigkeit des morphologischen Kopfes *Erfolg* wird eine zusätzliche Großschreibung lizenziert.

Am Beispiel *Abnehm-Erfolg* (W) zeigt sich auch, wie sich satz- und wortinterner Aspekt von N°-GROSS ergänzen, statt miteinander zu konkurrieren. *Erfolg* wird nach

²⁰ Zur Definition graphematischer Wörter und den Untergruppen der markierten graphematischen Wörter und den graphematischen Wörtern des Kernbereichs vgl. Kapitel 3.1.4.

N^o-GROSS als graphematisch selbständiger morphologischer Kopf großgeschrieben, zusätzlich zur Großschreibung des Gesamtausdrucks *Abnehm-Erfolg*, der seine Nominalität letztlich vom Kopf *Erfolg* ererbt hat, und das wortinitial durch die Großschreibung von |A| anzeigt, da es sich dabei ebenfalls um den Initialbuchstaben einer N^o-Position handelt.

Innerhalb der Beschränkungsordnung muss N^o-GROSS als Markiertheitsbeschränkung natürlich über MINUS als Beschränkung für den Standardfall gerankt werden und ist im Deutschen so stark, dass Verstöße dagegen in den Korpusbelegen kaum und nur in speziellen Kontexten belegt sind.

- (26)
- a. direkt-granulat (W)
 - b. linguisten mist (SMS)
 - c. VDI nachrichten (VDI09 (2mal))
 - d. Staats(un)wesen (WPD11)

Von den hier zugrunde liegenden 1.653 Determinativkomposita verstoßen nur fünf gegen N^o-GROSS.

direkt-granulat (W) bezeichnet ein Produkt, das beworben wird, und *VDI nachrichten* (VDI09) ist der Titel einer Fachzeitschrift, so dass die unübliche Kleinschreibung in Kombination mit Bindestrich bzw. Spatium wohl eine bewusste Entscheidung ist, um diese Namen graphisch abzusetzen. *linguisten mist* entstammt dem SMS-Korpus und ist ein gutes Beispiel dafür, dass in dieser Textform das Prinzip der Großschreibung (und der Zusammenschreibung morphologischer Bildungen) weniger dominant ist.

Wegen der geringen Zahl und Spezifität der Verstöße gegen N^o-GROSS ist anzunehmen, dass N^o-GROSS in der Hierarchie auf einer Stufe oder noch über ZUS-MORPH steht. Da gegen N^o-GROSS nur fünf Verstöße, gegen ZUS-MORPH aber sechs Verstöße belegt sind, entscheide ich mich aufgrund der Datenlage für eine Positionierung von N^o-GROSS über ZUS-MORPH (vgl. (27)), bin mir aber bewusst, dass die Einordnung auf Basis einer einfachen Mehrheit durch einen einzelnen Beleg natürlich angreifbar und eine kritische Überprüfung anhand eines größeren Datenausschnittes wünschenswert ist.

- (27) N^o-GROSS >> ZUS-MORPH >> GPKB >> GETR-AUSDR >> MINUS, *SPAT

Diese Positionierung führt dazu, dass die Beschränkungsprofile der potentiellen wortinternen Varianten *Honig-milch* und *Honig milch*, die den Beispielen unter (26) hinsichtlich ihres Verhaltens an der Kompositionsgrenze entsprechen, im Tableau unter (28) im Vergleich zu denen der anderen Kandidaten am schlechtesten sind.

(28)

	N°-GROSS	ZUS-MORPH	GPKB	GETR-AUSDR	MINUS	*SPAT
a. ☞ Honigmilch (DIV)				*		
b. HonigMilch				*	*!	
c. Honig-Milch			*!	*	*	
d. Honig Milch		*!			*	*
e. Honig-milch	*!		*	*	*	
f. Honig milch	*!	*			*	*

Wortinterne Großschreibung Scherer (2012), deren Beschränkung zur Substantivgroßschreibung oben schon diskutiert wurde, liefert mit *INT-MAJ auch eine Beschränkung, die die wortinterne Großschreibung in ihrem Ranking verbietet:

(29) VERBOT WORTINTERNER MAJUSKELN (*INT-MAJ)

Graphische Wörter enthalten keine Majuskeln. (Scherer 2012: 73)

Aus ihrer Beispielreihe (*Reisezentrum*, *ReiseZentrum*, *Reise-Zentrum*, *Reise Zentrum*, *Reise-zentrum*) verstößt nur *ReiseZentrum* gegen *INT-MAJ. Dementsprechend und auf Basis ihrer Ausführungen ist davon auszugehen, dass graphische Wörter graphematisch eigenständigen Wortteilen entsprechen (Scherer 2012: 65-66), und sich die Beschränkung nur auf wortinterne Positionen abseits der initialen bezieht; sonst müsste diese Beschränkung als Äquivalent zu MINUS verstanden werden und auch zu Verstößen bei *Reise-Zentrum* und *Reise Zentrum* führen.

Damit eignet sich diese Beschränkung grundsätzlich dazu, durch eine entsprechende Positionierung innerhalb der Beschränkungsordnung dafür zu sorgen, dass Kandidaten mit Binnenmajuskel schlechter zu bewerten sind, als solche mit einem Bindestrich an der Kompositionsgrenze. Dass ich sie dennoch nicht nutzen werde, liegt zum einen an der oben besprochenen, missverständlichen Formulierung, zum

anderen daran, dass mir eine Funktionalisierung der Beschränkung fehlt.

Scherer (2012) modelliert die Schreibung von N+N-Komposita in einem mehrstufigen Verfahren. So nimmt sie zunächst drei mögliche Schreibstrategien an: eine primär graphische, die die Worteinheit betone, eine primär morphologische, bei der die Markierung der morphologischen Wortstruktur höchste Priorität habe, und eine graphisch-morphologische Strategie, die neben der graphischen Wortgrenze zusätzlich die Grenzen einzelner untergeordneter Konstituenten markiere (vgl. Scherer 2012: 68-70). Im nächsten Schritt entwickelt sie für jede dieser drei Strategien eine eigene Beschränkungsordnung anhand unterschiedlicher Schreibungen des Lexems *Reisezentrum* (vgl. Scherer 2012: 73-80). Die Beschränkungen, die sie dazu nutzt, sind sehr eng auf das begrenzte Untersuchungsfeld abgestimmt und nur teilweise funktionalisiert. Die beiden übergeordneten Beschränkungen MORPH-WORT und *INT-MWORT, die eine Markierung morphologischer Grenzen in der Schreibung fordern, bzw. verbieten, modellieren im Sinne der primär graphischen Strategie durch die Dominanz von *INT-MWORT die Zusammenschreibung *Reisezentrum* als einzigen optimalen Kandidaten, bzw. im Sinne der primär morphologischen Strategie durch die Überordnung von MORPH-WORT die Kandidaten mit Bindestrich-, Getrennt- und Majuskelschreibung als gleichwertig optimale Kandidaten. Ausschließlich für die Modellierung der graphisch-morphologischen Strategie nutzt Scherer (2012: 73-74) drei zusätzliche Beschränkungen, die wortinterne Syngropheme wie den Bindestrich (*INT-SYNGR), Majuskeln (*INT-MAJ) und Spatien (*INT-SPAT) verbieten, um allein die Bindestrichschreibung als optimalen Kandidaten zu modellieren.

Diese Modellierung erweckt meiner Meinung nach zu Unrecht den Eindruck, dass die Schreibung an der Kompositionsgrenze von N+N-Komposita ungeachtet der beteiligten Kompositionsglieder und ihrer inhärenten Eigenschaften oder des Kontextes eine gewisse Beliebigkeit hat durch die willkürliche Wahl einer Schreibstrategie. Außerdem fehlt mir eine funktionale Verankerung von *INT-SYNGR, *INT-MAJ und *INT-SPAT. Durch die parallele Formulierung „Graphische Wörter enthalten keine“... bleibt beispielsweise verborgen, dass eine Majuskel deshalb nicht wortintern zu erwarten ist, weil sie als Initiale graphematisch eigenständiger Substantive funktionalisiert ist. Bei der Anwendung von *INT-SYNGR und *INT-SPAT ergibt sich meiner Einschätzung nach durch die Formulierung das noch wesentlich gravierendere Problem, dass Schreibungen wie *Reise-Zentrum* oder *Reise Zentrum* anders als bei

Scherer (2012) gar nicht zu Verstößen gegen *INT-SYNGR oder *INT-SPAT führen können. Wenn der Bindestrich oder das Spatium zwischen *Reise* und *Zentrum* dafür sorgen, das die beiden Bestandteile des morphosyntaktischen Wortes jeweils als eigene graphische Wörter betrachtet werden müssen, dann stehen sie zwischen zwei graphischen Wörtern, die selbst kein Syngphem oder Spatium enthalten, und sind damit nicht problematisch für *INT-SYNGR oder *INT-SPAT. Wenn Scherer (2012) dagegen die Bindestrichschreibung *Reise-Zentrum* als ein graphisches Wort betrachtet, geht damit nicht nur ein Verstoß gegen *INT-SYNGR einher, sondern auch ein von ihr nicht festgestellter gegen *INT-MAJ aufgrund der Großschreibung von *Zentrum*, da |Z| dann wie bei der Schreibung *ReiseZentrum* als Binnenmajuskel gelten muss.

Ausgehend von der Kritik an *INT-MAJ nach Scherer (2012) formuliere ich mit MAJUS in (30) selbst eine funktionalisierte Beschränkung zur Position von Majuskeln, die im Rahmen der Beschränkungsordnung zur Abstufung von Schreibungen mit Binnenmajuskeln eingesetzt werden kann.

(30) MAJUS

POSITION VON MAJUSKELN

Eine Majuskel leitet eine graphematisch eigenständige Einheit ein.

Gegen MAJUS verstoßen dann ebendiese Belege mit Binnenmajuskeln wie *ReiseZentrum*, da sie an einer Position auftreten, an der es keine graphische Grenze beispielsweise durch Spatium oder Bindestrich gibt, also kein graphematisch eigenständiges Wort oder Wortbestandteil, und damit auch keine Lizenzierung einer Substantivgroßschreibung nach N^o-GROSS möglich ist.²¹

Unter den 1.653 untersuchten Belegen bleiben nach Ausschluss aller monoreferentiellen Eigennamen nur noch die drei in (31) aufgeführten Belege mit Binnenmajuskel, unter denen sich mit *FeuchtCreme* und *PrivatKredit* gleich zwei Bezeichnungen für ein beworbenes Produkt finden.

(31) a. FeuchtCreme, PrivatKredit (W)

²¹ Ebenfalls gegen diese Beschränkung verstoßen aufgrund ihrer häufig durchgängigen Majuskel-schreibung auch Initialkurzwörter (*TÜV, KFZ*), und zwar gleich mehrfach. Vielleicht unterstützt also die Wirkung von MAJUS die von Buchmann (2015: 243-245) beobachtete Entwicklung dieser Initialkurzwörter von einer markierten durchgängigen Großschreibung hin zur unmarkierten Schreibung (*Tüv, Kfz*), die sie auf die Verwendungshäufigkeit zurückführt.

b. LautsprecherBox (Mail)

Majuskeln abseits der Initialposition graphematisch eigenständiger Wortbestandteile sind also in absoluten Zahlen noch seltener als eine Getrennschreibung (sechsmal belegt) oder eine systemwidrige Kleinschreibung (fünfmal belegt). Dementsprechend müsste diese Variante vor allem im Vergleich zur Bindestrichschreibung ein schlechteres Beschränkungsprofil aufweisen. Über die 5.000er-Stichproben hinaus ist zudem die Getrennschreibung im Gesamtkorpus mit einigem Abstand (351:81) häufiger belegt als eine Schreibung mit Binnenmajuskel. Um die Ergebnisse der Korpusanalyse abzubilden, muss MAJUS also anhand der Datenbasis über ZUS-MORPH eingeordnet werden. Die Einordnung unter N^o-GROSS in (32) erfolgt im Vergleich dazu auf einer wackeligeren Datenbasis, die sich am ehesten am Beschränkungsprofil der Kandidaten unter (33) erläutern lässt.

(32) N^o-GROSS » MAJUS » ZUS-MORPH » GPKB » GETR-AUSDR » MINUS,
*SPAT

Zwar verstoßen fünf Belege gegen N^o-GROSS (vgl. (26)) und lediglich drei gegen MAJUS (vgl. (31)), alle drei Verstöße gegen MAJUS sind aber Kandidaten auf der Stufe *d.* des Tableaus und damit genau einmal häufiger vertreten als Kandidaten der Stufe *e.* (*direkt-granulat* (W), *Staats(un)wesen* (WPD11)) und ebenso häufig wie Kandidaten der Stufe *f.* (*linguisten mist* (SMS) und *VDI nachrichten* (VDI09, 2mal)). Da es sich aber bei zwei dieser drei Kandidaten auf der niedrigsten Optimalitätsstufe um zwei Instanzen des Zeitschriftentitels *VDI nachrichten* (VDI09) handelt, für dessen suboptimale Schreibung von einer Funktionalisierung im Sinne der Aufmerksamkeitserregung auszugehen ist, sehe ich davon ab, MAJUS noch höher einzuordnen. Ein Ranking von MAJUS auf einer Stufe mit N^o-GROSS würde aber zu identischen Ergebnissen führen, da die Belege der Optimalitätstufen *d.* bis *f.*, die gegen MAJUS oder N^o-GROSS verstoßen, ihre Einstufung auf Basis der Verstöße gegen die untergeordneten Beschränkungen beibehalten würden (vgl. dazu (33)).

(33)

	N°-GROSS	MAJUS	ZUS-MORPH	GPKB	GETR-AUSDR	MINUS	*SPAT
a. ☞ Honigmilch (DIV)					*		
b. Honig-Milch				*!	*	*	
c. Honig Milch			*!			*	*
d. HonigMilch		*!			*	*	
e. Honig-milch	*!			*	*		
f. Honig milch	*!		*				*

Exkurs: Graphematische Eigenständigkeit Als graphematisch eigenständig gelten, wie oben schon erläutert, neben graphematischen Wörtern auch die Bestandteile markierter graphematischer Wörter, sofern sie durch einen Bindestrich abgegrenzt sind. Eine weiterführende Analyse ausgehend vom Beleg *Staats(un)wesen* (WPD11) hat aber ergeben, dass auch andere Zeichen eine graphematische Eigenständigkeit kodieren können, sofern sie

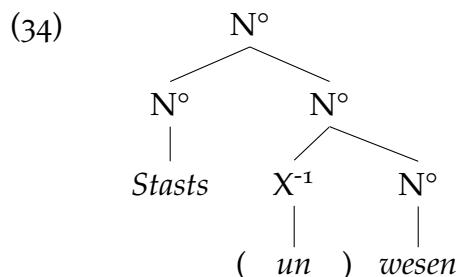
- nicht mit einem Bindestrich kombiniert werden,
- oder nicht rückstandslos durch graphematische oder morphosyntaktische Wortbestandteile ersetzt werden können,
- oder nicht zur Kodierung mehrerer Lexeme innerhalb eines graphematischen Wortes dienen.²²

Diese Definition schafft es, neben dem Divis (als Bindestrich) auch den Schrägstrich (zwischen koordinierten Teilausdrücken) als Zeichen zu legitimieren, das eine graphematische Selbständigkeit kodieren kann, sofern er in Fällen wie *Kunststoff/Metall-Strukturbauteile* (VDI09) alle Aspekte der Definition erfüllt. Alle anderen Ausprägungen des Divis (Ergänzungs- und Trennstrich) können dagegen trotz Formgleichheit

²² Scherer (2012: 66) nimmt an, dass neben Spatien auch Bindestriche und Apostrophierungen sowohl die linke als auch die rechte Grenze eines Wortes (aus dem Kontext wird deutlich, dass damit auch Kompositionsglieder gemeint sind) markieren können, exemplifiziert oder begründet ihre Annahme aber leider nicht. Für die Apostrophierungen adjektivischer Ableitungen von Eigennamen (*die darwinsche/die Darwin'sche Evolutionstheorie*) ist aber in §62 der orthographischen Norm tatsächlich zu beobachten, dass auch der Apostroph zur graphematischen Selbständigkeit von Wortbestandteilen führen kann.

aufgrund ihrer Ersetzbarkeit nicht für eine graphematische Eigenständigkeit sorgen.

Ausgangspunkt dieser Überlegungen war der Beleg *Staats(un)wesen* (WPD11). Durch die wortinterne Klammerung handelt es sich um ein markiertes graphematisches Wort, das potentiell – sofern eine Klammerung zur graphischen Eingeständigkeit von Wortbestandteilen führt – sogar gleich drei graphematisch eigenständige Teilausdrücke aufweist (*Staats, un, wesen*).



Wie die Analyse unter (34) zeigt, müssten, für den Fall, dass Klammern zur graphischen Eigenständigkeit führen, nach N°-GROSS eigentlich alle Teilausdrücke mit einer Majuskel beginnen (*Staats(Un)Wesen*), da sie allesamt N°-Positionen besetzen.

Die wortinternen Klammern, die hier letztendlich dazu dienen, innerhalb eines graphematischen Wortes gleich zwei Lexeme zu kodieren,²³ nämlich *Staatswesen* und *Staatsunwesen*, scheinen in diesem Fall also anders als Bindestriche keine Angriffspunkte für N°-GROSS zu schaffen.

Ein Blick in die Korpusbelege auch über die zweigliedrigen Determinativkomposita hinaus zeigt, dass Klammern wortintern fast immer mit Bindestrichen (oder sogar Bindestrich und Spatium) kombiniert werden:

- (35) FPSO-(Floating Production, Storage and Offloading)-Schiffe (VDI09), Food Chemical Codex (FCC)-Einheiten (W), Future-Combat-Systems (FCS-)Programm (VDI09), (Kämpfer-)Lieder (WDD11), (Verzeichnis-)Netzwerke (HAZ09), Austausch(Exportformat) (WDD11)

Diese Kombinationsmöglichkeit spricht ebenfalls dafür, dass Klammern, anders als Bindestriche, Teilausdrücken markierter graphematischer Wörter keine graphematische Eigenständigkeit verschaffen, da diese sonst entgegen jeglicher sprachlicher

²³ Ähnlich verhält es sich mit den deshalb ausgenommenen gegenderten Bezeichnungen mit Binnen-I wie *SchülerInnen*. Vgl. 3.3.4.

Ökonomiebestrebungen doppelt kodiert wäre. Die Kombinationsmöglichkeit mit einem Bindestrich ist also womöglich ein Kriterium, an dem sich ablesen lässt, ob andere Interpunktionszeichen zur graphematischen Selbständigkeit von Wortbestandteilen führen können. Lassen sich Interpunktionszeichen – wie in den obigen Beispielen die Klammer – wortintern mit dem Bindestrich kombinieren, sind sie nicht selbst in der Lage, Wortbestandteilen graphematische Selbständigkeit zu verleihen.

Eine Klammerung wird neben der Kodierung von zwei Lexemen in einer Wortform offensichtlich auch für wortinterne Einschübe genutzt: Im Beispiel *FPSO* (*Floating Production, Storage and Offloading*)-Schiffe (VDI09) dient der Einschub der Auflösung bzw. Erläuterung der vorausgehenden Abkürzung *FPSO*, während in *Future-Combat-Systems* (*FCS*-)Programm (VDI09) und *Food Chemical Codex* (*FCC*)-Einheiten (W) in Klammern die Abkürzung zur vorausgehenden Vollform eingeführt wird. Unabhängig von der Reihenfolge ist hier wortintern also dieselbe Information zweimal, wenn auch unterschiedlich, kodiert, wobei die Wiederholung durch die Klammern markiert wird. Die Schreibung der durch Bindestriche oder Spatien graphematisch eigenständigen Teilausdrücke entspricht dabei N°-GROSS.

Die verbleibenden Belege mit Klammern verhalten sich aber parallel zu *Staats(un)wesen* (WPD11) und kodieren in nur einem markierten graphematischen Wort ebenfalls zwei Lexeme (eins als Vollform mit geklammertem Teilausdruck und eines unter Auslassung desselben) mit je eigener graphematischer Oberfläche:

- (36) a. Staats(un)wesen (WPD11): Staatsunwesen, Staatswesen
 b. (Kämpfer-)Lieder (WDD11): Kämpfer-Lieder, Lieder
 c. (Verzeichnis-)Netzwerke (HAZ09): Verzeichnis-Netzwerke, Netzwerke
 d. Telekom(munikations)-Flatrate (WDD11): Telekommunikations-Flatrate, Telekom-Flatrate²⁴

Großschreibungen nach einer Klammer kommen in diesen Beispielen dann vor, wenn die Klammer mit einem Bindestrich kombiniert ist. Die Reihenfolge scheint dabei keine Rolle zu spielen, wie (36c) und (36d) verdeutlichen. Die Auflösung der markierten graphematischen Wörter mit Klammerung in die beiden Einzellexeme, die sie kodieren, macht deutlich, dass die Groß- oder Kleinschreibung nach und in

²⁴ In den Beispielen (36a) und (36d) wäre eine Großschreibung für den Wortbestandteil innerhalb der Klammern auch nicht möglich, da hier gar keine N°-Position eingeklammert ist.

einer wortinternen Klammerung vielmehr davon abhängt, welche beiden Lexeme kodiert werden sollen.²⁵

Der einzige Beleg, der für eine Anwendung von N^o-GROSS auf das initiale Graphem nach einer Klammer (ohne Kombination Bindestrich) spricht (*Austausch(Exportformat)* (WDD11)), ist gleich in mehrfacher Hinsicht unsystematisch und wird hier zur Aufklärung mit Kontext gezeigt.

(37) nur ein genormtes *Austausch(Exportformat)* und *Archivierungsformat* (WDD11)

Der Kontext spricht dafür, dass inhaltlich letztlich durch die Erstglieder *Austausch*, *Export* und *Archivierungs* auf dasselbe *Format* aus leicht unterschiedlicher Perspektive verwiesen wird. Dabei ist die Klammerung bzw. Koordination in diesem Beispiel wohl nicht ganz gelungen, da sie für die Auflösung als *Austausch*, *AustauschExportformat* spricht, und wahrscheinlich weder die Binnenmajuskel der zweiten Struktur noch der fehlende Kopf der ersten beabsichtigt ist. Geschickter wäre wohl eine Schreibung wie *Austausch-*, *Export-*, und *Archivierungsformat* gewesen, wie sie sich im Korpus auch bei ähnlichen Konstruktionen findet (z. B. *Schwerlast- und Baustellenverkehr* (W)).

Festzuhalten bleibt, dass eine Klammerung einem Teilausdruck keine graphematische Eigenständigkeit verleiht, so dass die Position nach der Klammer keinen Ansatzpunkt für N^o-GROSS bietet, sondern diese Beschränkung die Klammerung gewissermaßen ignoriert und an den entsprechenden Stellen nur greift, wenn neben der Klammer ein Bindestrich (eventuell sogar in Begleitung eines Spatiums) für graphematische Eigenständigkeit sorgt.

Parallel zu geklammerten Teilausdrücken verhalten sich auch solche, die mit Anführungszeichen eingefasst werden („*Jesus-liebt-euch*“-Gruppe (RHZ09), „*Africanthropus*“-Schädel (WPD11)), oder auch solche mit Punkten (*Dipl.-Arbeit* (WDD11)). Das liegt wohl daran, dass die Punkte und Anführungszeichen Bestandteile der jeweiligen Teilausdrücke sind, und die graphematische Eigenständigkeit

²⁵ Streng genommen sind also Belege wie *Staats(un)wesen* ebenso von dieser Untersuchung auszunehmen wie Belege mit Ergänzungsstrichen oder Binnen-I, da es sich in diesen Fällen um Koordinationen handelt und nicht, bzw. nur graphisch, um Einzelwörter.

auch hier nur durch einen Bindestrich oder ein Spatium hergestellt werden kann.²⁶

Ein anderes Bild zeigt sich bei wortinternen Schrägstrichen:

- (38) a. Olympische Sommerspiele 1976/Leichtathletik-Artikel (WDD₁₁),
"Pack-Zu/Schäfer-ShopTeamkollegen (RHZ₀₉), Kunststoff/Metall-
Strukturbauteile (VDI₀₉), Sendeler/Circle-IT-Forum (VDI₀₉), Ja/Nein-
Antworten (WPD₁₁), Ein/Aus-Schalter (VDI₀₉), Reich/Arm-Spiel
(WDD₁₁)
- b. Markit/BME-Einkaufsmanagerindex (VDI₀₉), ÖVP/FPÖ-
SympathisantInnen (WDD₁₁), CC-by-sa/3.0-Lizenzvorlagen (WDD₁₁),
GSM/GPRS-Schnittstelle (VDI₀₉), ARD/ZDF Online-Studie (WDD₁₁),
F/A-18 Hornet-Staffel (WPD₁₁), UV/VIS-Spektroskopie (WPD₁₁),
CDU/FDP-Regierung (HAZ₀₉, BRZ₀₉, NUZ₀₉), CAD/CAM-Software
(VDI₀₉), AC/DC-T-Shirts (M₀₉), CAD/CAM-System (VDI₀₉), VfB/SC-
Reserve (BRZ₀₉), I/O-Systemen (VDI₀₉)
- c. DFB-/McDonalds-Fußballabzeichens (M₀₉), Ion-/Ion-Elektroden
(WDD₁₁)

Wie die Beispiele unter (38) zeigen, finden sich bei Komposita mit mehr als zwei Kompositionsgliedern aus der kleinsten Stichprobe vergleichsweise viele Korpusbelege mit internen Schrägstrichen, darunter jedoch nur zwei, in denen dieser unmittelbar mit einem Bindestrich kombiniert wird (vgl. (38c)). Selbst wenn man die Belege ausklammert, die nach dem Schrägstrich Teilausdrücke aufweisen, die nicht für N^o-GROSS zugänglich sind, weil es sich um Ziffern oder durchgängig großgeschriebene Abkürzungen handelt (38b), verbleibt noch eine Vielzahl von Belegen, die dafür spricht, dass auch der Schrägstrich Teilausdrücken markierter graphematischer Wörter eine graphematische Eigenständigkeit im Sinne von N^o-GROSS verleiht (vgl. (38a)). Der Schrägstrich in den Beispielen unter (38a) verhält sich dabei parallel zu dem oben besprochenen Bindestrich in mehrgliedrigen Komposita mit koordinierten Erstgliedern wie *Hals-Nasen-Ohren-Arzt* und *Hörer-Leser-Gunst*.²⁷ Anders als

²⁶ Ganz ähnlich verhalten sich auch graphisch abgesetzte, z. B. kursivierte Teilausdrücke (*Homo sapiens*-Männchen (SPK)). Auch hier verleiht das interne Merkmal der Kursivität dem Teilausdruck keine graphematische Eigenständigkeit.

²⁷ Diese Parallelität ist auch durch die Schreibnorm abgedeckt, vgl. 3.2.

der Bindestrich scheint der Schrägstrich aber auf diese Position zwischen koordinierten Teilausdrücken beschränkt zu sein, und findet sich daher, wie die Beispiele aus der gesamten Belegsammlung verdeutlichen, nur in Komposita mit mindestens drei Teilausdrücken.

Aufgrund der Beleglage ist also davon auszugehen, dass neben Bindestrich auch ein Schrägstrich Teilausdrücken markierter graphematischer Wörter durch eine graphische Abgrenzung eine Eigenständigkeit im Sinne von N^o-Gross verleiht, diese Einsatzmöglichkeit aber auf mehr als zweigliedrige Komposita beschränkt ist, und andere Interpunktionszeichen auf Wortebene Teilausdrücken keine graphematische Eigenständigkeit verleihen.

Eine Erklärung für dieses Verhalten liegt, wie oben schon angedeutet, in der Kombinatorik der Zeichen. Sowohl Klammern als auch Anführungszeichen und Abkürzungspunkte werden, wie die obigen Beispiele zeigen, regelmäßig mit einem Bindestrich kombiniert. Diese Kombinationsmöglichkeit macht deutlich, dass keines dieser Interpunktionszeichen selbständig eine graphematische Eigenständigkeit von Bestandteilen graphematischer Wörter kodieren kann. Im Gegensatz zum Schrägstrich, für den Beispiele belegen, dass er unabhängig dazu in der Lage ist, eine graphematische Eigenständigkeit zu kodieren, und nur vereinzelte Fälle eine Kombination mit dem Bindestrich belegen.

Aufgrund ihrer Formgleichheit zum Bindestrich wird auch noch das Verhalten von Ergänzungs- und Trennstrich, als weitere Divisorausprägungen, kurz diskutiert, auch wenn es sich dabei eigentlich nicht um wortinterne Phänomene handelt.

Der Trennstrich ist zwangsweise mit einem Zeilensprung kombiniert und dem Ergänzungsstrich tritt stets in Kombination mit einem Leerzeichen auf, das dem Ergänzungsstrich entweder vorausgeht (*Lebensmittelproduktion und -verarbeitung*) oder folgt (*Kinder- und Jugendfreizeit*). So ergeben sich beim Ergänzungsstrich in Vorwärtstilgungen (*Lebensmittelproduktion und -verarbeitung*) und beim Trennstrich an morphologischen Grenzen (*Haus-[Zeilensprung]tür*) auf den ersten Blick graphematisch eigenständige Wortbestandteile, die obwohl sie N^o-Positionen besetzen, kleingeschrieben werden.

Im Fall von Vorwärtstilgungen ist tatsächlich von einem graphematisch eigenständigen Wortbestandteil auszugehen. Allerdings kann hier trotz Besetzung einer N^o-Position durch den graphematisch selbständigen Teil wie im obigen Beispiel -

verarbeitung keine initiale Großschreibung realisiert werden, da das initiale Graphem der Ergänzungsstrich und kein alphabetisches Graphem ist, N^o-GROSS aber nach einem Leerzeichen ausschließlich auf die erste Position eines graphematischen Wortes zugreifen kann.²⁸ Da nur Buchstabengrapheme zweielementig sind, also in einer Minuskel- oder Majuskelrealisierung vorliegen können, Interpunktionszeichen aber nicht, kann man auch von einer koverten Großschreibung in diesen Fällen ausgehen, die theoretisch möglich ist, aber aufgrund der Einelementigkeit von Interpunktionszeichen nicht sichtbar umgesetzt werden kann. Zudem sorgt die Kleinschreibung nach dem Ergänzungsstrich in Vorwärtstilgungen auch (ähnlich wie bei wortinternen Klammern) schlichtweg dafür, dass die Tilgung ohne potentielle Binnenmajuskel aufgelöst werden kann wie in (39a). Ob darüber hinaus eine Auflösung wie in (39c) möglich ist, muss an anderer Stelle geklärt werden.

- (39) a. Germanistikvorlesungen und -seminare: Gemansitikvorlesungen und Germanistikseminare
 b. *Germanistikvorlesungen und -Seminare: Germanistikvorlesungen und GermanistikSeminare
 c. ?Germanistikvorlesungen und -Seminare: Germanistikvorlesungen und Germanistik-Seminare

Bei Trennstrichen liegt der Fall wiederum anders. Sie sind zwar ebenso wie Ergänzungsstriche wohl allein schon aufgrund der Formgleichheit nicht unmittelbar mit Bindestrichen kombinierbar, die Kombinationsmöglichkeit wird darüber hinaus aber noch durch die unveränderliche Umgebung von Ergänzungsstrich und Trennstrich (ein Leerzeichen vor dem Ergänzungsstrich in Vorwärtstilgungen, bzw. ein Zeilensprung nach dem Trennstrich) blockiert. Durch den Zeilensprung nach dem Trennstrich kommt es aber bei morphologischen Trennungen durchaus vor, dass zu Beginn einer neuer Zeile, also graphisch deutlich abgesetzt, ein Wortbestandteil steht, der eine N^o-Position besetzt, ohne dass in diesen Fällen großgeschrieben wird, obwohl das initiale Graphem eine Großschreibung erlauben würde wie *tür* in *Haus-[Zeilensprung]tür*. Damit liegt in solchen Fällen ein potentieller Verstoß gegen N^o-GROSS vor.

²⁸ Hübl & Steinbach (2011: 270) nehmen eine solche Erklärung für die Kleinschreibung von Nominalisierungen wie *das -ung* (als Substantivierung des Suffixes) an.

Verstöße in diesem Zusammenhang sind aber eindeutig funktionalisiert. Die konsequente Kleinschreibung nach Trennstrichen, auch wenn durch sie N^o-Positionen graphematisch eigenständig werden, betont die Zusammengehörigkeit der Bestandteile über die graphische Grenze des Zeilensprungs hinweg.²⁹ Außerdem verdeutlicht die Kleinschreibung wie beim Ergänzungsstrich, dass das betreffende Wort abseits der Position am Zeilenende bzw. bei Auflösung der Tilgung ohne Binnenmajuskel und Divis auskommt.³⁰

Tatsächlich ist der Bindestrich die einzig beständige Ausprägung des Divis während Ergänzungsstrich und Trennstrich jeweils als Platzhalter fungieren, die ohne Spuren zu hinterlassen durch einen Wortbestandteil ersetzt werden, sobald eine Tilgung aufgelöst oder ein graphematisches Wort gänzlich übers Zeilenende hinaus geschoben wird.

Zusammenfassung Durch drei zusätzliche Beschränkungen (MINUS, N^o-GROSS und MAJUS) konnte nach der Schreibung ohne wortinterne Spatien und Interpunktionszeichen als Standard auch die Großschreibung auf Wortebene (wortinitial und wortintern) so modelliert werden, dass die wortinterne Kleinschreibung als Standard (mit systematischen Abweichungen bei graphematisch eigenständigen Teilausdrücken) abgebildet wird. Mit dem Ranking in (40) ist die Zusammenschreibung ohne Interpunktionszeichen oder wortinterne Majuskeln so als optimaler Kandidat für die Schreibung an Kompositionsgrenzen in zweigliedrigen Determinativkomposita modelliert worden, dass auch der absteigende Optimalitätsgrad der anderen im Korpus belegten Varianten ihren absoluten Frequenzen entsprechend abgebildet werden konnte (vgl. (33)).

(40) N^o-GROSS >> MAJUS >> ZUS-MORPH >> GPKB >> GETR-AUSDR >> MINUS,
*SPAT

²⁹ Diese Betonung der Zusammengehörigkeit kann auch als zusätzliches Argument für die Kleinschreibung nach einem Ergänzungsstrich in Vorwärtstilgungen verwendet werden. Während sie beim Trennstrich der Überbrückung des Zeilensprungs dient, schlägt sie beim Ergänzungsstrich den Bogen über Leerzeichen und koordinierende Konjunktion.

³⁰ Steht ein Divis in der typischen Trennstrichumgebung an einer Kompositionsgrenze, auf die ein selbständiger, großgeschriebener Wortbestandteil in N^o-Position folgt, wie in *Bauleiter-Schulung*, ist davon auszugehen, dass der Divis nicht als Trenn- sondern als Bindestrich zu interpretieren ist und das graphematische Wort zeilenintern unverändert als *Bauleiter-Schulung* realisiert wird.

Da die Zusammenschreibung im Deutschen den Standardfall der wortinternen Schreibung darstellt, erfasst dieses Ranking die 93,5% der 1.653 zweigliedrigen Determinativkomposita als optimale Kandidaten, weil an deren Kompositionsgrenze klein und ohne Leerzeichen oder Interpunktionszeichen weitergeschrieben wird.

Es verbleiben also nur 108 Belege (6,5%) für andere Varianten wie Bindestrich, Binnenmajuskel oder Getrenntschreibung im Datenausschnitt.³¹ Die Datengrundlage für die Modellierung von Varianten abseits der unmarkierten Zusammenschreibung ist also dünner als die zur Modellierung der standardmäßigen Zusammenschreibung, wird sich aber im Folgenden als weitestgehend tragfähig erweisen. Neben den oben schon besprochenen sechs Schreibungen mit wortinternem Leerzeichen unter (41b) und drei Belegen für Binnenmajuskeln in (41c) handelt es sich – abgesehen von zwei Fällen, in denen der Bindestrich mit anderen Interpunktionszeichen kombiniert ist (41d) – in allen anderen Fällen um einfache Bindestrichschreibungen wie in (41a).

- (41) a. Mineral-Box (WDD11), FZI-Vorstand (VDI09), Doppel-Halde (WPD11)
 b. BWL Studiums (Mail), Hoffmann-von-Fallersleben Museum (BRZ09), linguisten mist, USB Verlängerung (SMS), VDI nachrichten (VDI09 (2mal))
 c. FeuchtCreme, PrivatKredit (W), LautsprecherBox (Mail)
 d. „Spiegel“-Reporter (HAZ09), (Kämpfer-)Lieder (WDD11)

6.2.2 Systematische Variation

- (42) N°-GROSS ≫ MAJUS ≫ ZUS-MORPH ≫ GPKB ≫ GETR-AUSDR ≫ MINUS,
 *SPAT

Das Ranking in (42) ist damit die optimale Basis, um nun auch systematische Variation abseits der Zusammenschreibung mit weiteren Beschränkungen als optimale

³¹ Diese Beobachtungen zum Anteil von Zusammenschreibung und Bindestrichschreibung decken sich mit denen von Borgwaldt (2013) und Kopf (2017) für N+N-Komposita. Borgwaldt (2013) kommt in ihrem Produktionsexperiment für neugebildete N+N-Komposita auf einen Anteil von 5,8% an Bindestrichschreibungen (n=828) und Kopf (2017: 181) ermittelt bei der Nachanalyse der Korpusdaten von Grube (1976) einen Wert von 7,6% (n=5.332) für lektorierte Texte (Zeitungen, Zeitschriften, wissenschaftliche und belletristische Texte).

Kandidaten zu erfassen.

Bei der ausschließlichen Betrachtung von Determinativkomposita bleibt der Einfluss der Wortbildungsart zunächst außen vor. Ebenso unberücksichtigt bleiben durch die Beschränkung auf zweigliedrige Komposita Effekte der wortinternen Hierarchie. Auch der Einfluss der Textform bzw. Einfluss ihrer Merkmale soll zunächst noch unbeachtet bleiben, um sich auf die Faktoren konzentrieren zu können, die in der Natur der Teilausdrücke begründet sind. So verbleiben, wie in den Forschungshypothesen unter 4.1 auf Basis der vorausgehenden Diskussion der einzelnen Varianten in 3.3 formuliert, der Einfluss von Eigennamen, fremdsprachlichen Ausdrücken, graphematisch markierten Wörtern oder spezifischen Buchstabenkombinationen als vermutete Einflussfaktoren zu Gunsten von Varianten abseits des Defaults der unmarkierten Zusammenschreibung. Die Verwendung von Fugenelementen und eine Lexikalisierung der komplexen Wörter lassen dagegen einen Effekt hin zur Zusammenschreibung erwarten.

Anders als bei der vorausgegangenen Modellierung der Zusammenschreibung als Standard für die Schreibung an der Kompositionsgrenze zweigliedriger Determinativkomposita stehen für die systematische Integration dieser Einflussfaktoren keine Beschränkungsvorbilder in der Literatur zu Verfügung, auf die hier aufgebaut werden könnte. Dementsprechend erarbeite und formuliere ich diese Beschränkungen im Folgenden selbst auf Basis der 1.653 Belege und der Diskussion zu diesen Einflussfaktoren in 3.3.

Markierte graphematische Wörter als Teilausdrücke

Mit Blick auf die 108 Belege, die von der Zusammenschreibung abweichen, zeigt sich, dass fast die Hälfte dieser Belege ein Kompositionsglied aufweist, das aus verschiedenen Gründen nicht der Definition eines graphematischen Wortes des Kernbereichs (vgl. Kapitel 3.1.4) entspricht, also graphisch markiert ist.

Wie die Beispiele in (43) zeigen, fallen darunter Komposita mit Ziffern (a), Einzelbuchstaben (b), Akronymen (c) oder Teilausdrücken, die graphematisch markierten Wörtern entsprechen (d). Erstere verstoßen gegen das Konzept der graphematischen Silbe (kein oder nur ein Buchstabengraphem, fehlender kompakter Kern),³² während

³² Wie oben schon erläutert, gilt ein Akronym auch dann schon als markiertes graphematisches Wort, wenn es zwar dem Konzept der graphematischen Silbe entspricht, das aber aufgrund einer durchgängigen Majuskelschreibung nicht sofort ersichtlich ist wie im Beispiel *FZI-Vorstand* (VDI09).

die Erstglieder der letzten Gruppe aufgrund der teilausdruckinternen Wortzeichen markierten graphematischen Wörtern entsprechen. So führen die Anführungszeichen um „Spiegel“ und der interne Bindestrich bei *Know-How* dazu, dass ein weiterer Teilausdruck im Kompositionsprozess mit einem Bindestrich angeschlossen wird. Auch bei *Fritz-Beske-Instituts* (VDI09) knüpft ein Bindestrich den markierten Teilausdruck *Fritz-Beske* an *Instituts* an.³³

- (43) a. 360°-Kamera (W)
 b. A-Junioren (M09)
 c. FZI-Vorstand (VDI09), Kfz-Versicherung, BWL Studiums (MAIL)
 d. Fritz-Beske-Instituts (VDI09) Experten-Know-How³⁴ (VDI09), „Spiegel“-Reporter (HAZ09)

Aufgrund ihrer graphischen Markiertheit scheinen sie alle einen Bindestrich zu brauchen, um graphematisch anschlussfähig für weitere Teilausdrücke zu sein, behalten dabei aber mit Hilfe des Bindestrich die graphematische Eigenständigkeit auch als Wortbestandteil innerhalb eines graphematischen Wortes oder verbleiben graphisch gänzlich ungebunden wie in *BWL Studiums* (Mail), wo dadurch einem morphosyntaktischen Wort gleich zwei graphematische Wörter entsprechen. Kompositionsglieder, die markierten graphematischen Wörtern entsprechen, können also als Teilausdruck nur mittels Bindestrich Teil eines komplexen, markierten graphematischen Wortes werden – in dem so die Grenze zum Teilausdruck, der einem markierten graphematischen Wort entspricht, sichtbar bleibt – oder bleiben graphematisch gesehen ein selbstständiges Wort trotz Einbindung in ein Wortbildungsmuster.

Um diesen Befund abzubilden, formuliere ich eine erste zusätzliche Beschränkung, die eine Zusammenschreibung an Grenzen zu Teilausdrücken, die graphematisch markierten Wörtern entsprechen, ausschließt:

- (44) *GRAPH

³³ Auch wenn man annimmt, dass der Eigenname als Ganzes lexikalisch und in Verbindung damit womöglich auch graphisch (mit internem Leerzeichen) im mentalen Lexikon abgespeichert ist, und der Bindestrich hier im Zuge der Wortbildung schon eine Art Reperaturmechanismus ist, hebt das die graphische Markiertheit des Teilausdrucks nicht auf.

³⁴ Trotz dreier visuell deutlich sichtbarer Bestandteile fällt der Beleg *Experten-Know-How* unter die zweigliedrigen Komposita, da *Know-How* keine Komposition sondern eine Zusammenrückung ist.

EINBINDUNG VON GRAPHEMATISCH MARKIERTEN WÖRTERN

Teilausdrücke, die graphematisch markierten Wörtern entsprechen, bleiben auch innerhalb morphologischer Bildungen graphematisch eigenständig.³⁵

Durch die Formulierung „*graphematisch eigenständig*“ im Sinne der Definition unter 6.2.1 wird also ein Bindestrich zur Anknüpfung markierter Teilausdrücke lizenziert. Ein Schrägstrich, der wie oben gezeigt ebenfalls graphematische Eigenständigkeit von Teilausdrücken kodiert, spielt hier zunächst keine Rolle. Die zweigliedrigen Determinativkomposita, die der Modellierung zugrunde liegen, bieten mangels koordinierter Teilausdrücke schlicht keine Einsatzmöglichkeit für den Schrägstrich, der erst ab drei oder mehr Kompositionsgliedern zum Einsatz kommen könnte.

In der untersuchten Stichprobe finden sich 50 Determinativkomposita, für die *GRAPH aufgrund der Markiertheit eines Teilausdruckes von Relevanz ist und unter denen sich nur ein einziger Beleg findet, der eine Zusammenschreibung aufweist (*Staats(un)wesen* (WPD₁₁)). Streng genommen müssten Belege wie dieser aber aus der Untersuchung ebenso ausgenommen werden, wie beispielsweise Koordinationen mit einem Ergänzungsstrich, da es sich in beiden Fällen nicht um eines sondern gleich zwei Lexeme (*Staatswesen* und *Staatsunwesen*) handelt. Da neben dieser einen Zusammenschreibung und der Bindestrichschreibung in 44 Fällen für diese Gruppe innerhalb der 5.000er-Stichproben auch die Getrenntschreibung in fünf Fällen belegt ist (vgl. (7)), wird *GRAPH in der Hierarchie (45) zwischen MAJUS und ZUS-MORPH eingeordnet.

(45) N°-GROSS >> MAJUS >> *GRAPH >> ZUS-MORPH >> GPKB >> GETR-AUSDR
>> MINUS, *SPAT

Durch die Positionierung von *GRAPH über GPKB und ZUS-MORPH wird die Bindestrichschreibung bei graphematisch markierten Teilausdrücken zum optimalen Kandidaten. Die Zusammenschreibung ist nur noch außerhalb des Geltungsbereiches von *GRAPH der optimale Kandidat und bei Relevanz eines Teilausdrucks für *GRAPH wird durch die Platzierung über ZUS-MORPH, wie im Tableau unter (46)

³⁵ Dass graphische Grenzen nicht nur nach, sondern auch vor graphematisch markierten Teilausdrücken liegen, zeigt ein Blick über die Stichprobe hinaus durch Belege wie *Probe-SMS* (DIV) oder *Pay-TV-Abonenten* (VDI₀₉).

ersichtlich, die Getrennschreibung, entsprechend der Beleglage, noch vor der Zusammenschreibung als gewissermaßen zweitbeste Wahl abgebildet.³⁶

(46)

	N°-GROSS	MAJUS	*GRAPH	ZUS-MORPH	GPKB	GETR-AUSDR
a. CV-Geschichte (WDD11)					*	*
b. CV Geschichte				*!		
c. CVgeschichte			*!			*
d. CVGeschichte		*!	*			*
e. CV-geschichte	*!				*	*
f. CV geschichte	*!			*		

Durch diese einfache Ergänzung werden schon 1.588 der 1.653 Schreibungen zweigliedriger Determinativkomposita, also 96%, als optimale Kandidaten abgebildet.³⁷

In diesem Tableau unter (46) und auch in den folgenden Tableaus wird der Übersichtlichkeit halber auf die Darstellung der Beschränkungen MINUS und *SPAT verzichtet. Sie sind zwar zur Modellierung der Zusammenschreibung als Default auch für Kompositionsgrenzen wesentlich und werden auch im Folgenden mitgedacht, da Verstöße gegen die beiden am niedrigsten gerankten Beschränkungen aber bei der Modellierung der systematischen Variation keine Rolle spielen, werden sie im Zuge der Modellierung nicht mehr abgebildet.

Lexikalisierung

Ebenso eindeutig wie bei Wortbestandteilen, die graphematisch markierten Wörtern entsprechen, ist die Beleglage hinsichtlich eines Einflusses des Lexikalisierungsstatus der Komposita als Ganzes. Der Effekt wirkt aber in die entgegengesetzte Richtung. Die Lexikalisierung eines komplexen Wortes ist ein sehr eindeutiger Indikator für die Zusammenschreibung, es sei denn einer der Teilausdrücke entspricht einem gra-

³⁶ Entsprechungen zu den Kandidaten (46) d.-f. finden sich nicht in der Stichprobe, so dass sie sicher weniger optimal als c. sind, ihre Hierarchie untereinander aber vorerst nicht überprüft werden kann.

³⁷ Aufgrund der geringen Belegzahl, für die *GRAPH im Vergleich zur Grundgesamtheit relevant ist, und der auch vor der Einfügung von *GRAPH schon hohen Quote optimaler Kandidaten (93,5%), liegt hier keine statistisch signifikante Verbesserung vor.

phematisch markierten Wort.

Von den 1.653 untersuchten Determinativkomposita fallen nach dem Abgleich mit der DeReWo-Wortliste, welche die 100.000 häufigsten Wortformen auf Basis des Deutschen Referenzkorpus (Stand 2014) umfasst, 757 in die Kategorie lexikalisiert,³⁸ während sich 896 Komposita nicht in identischer Schreibung auf der Liste wiederfinden.³⁹ Unter den 757 lexikalisierten Belegen finden sich nur in 16 Abweichungen von der Zusammenschreibung. In allen 16 Wortformen wird, wie in (47) zu sehen, ein durch *GRAPH legitimierter Bindestrich verwendet. Andere von der Zusammenschreibung abweichende Varianten sind unter den lexikalisierten Wortformen nicht belegt.

- (47) Willy-Brandt-Platz (BRZ09), CO₂-Emissionen (PBR, W (8mal)), EU-Kommission (WPD11), NS-Verbrechen (WPD11), SED-Diktatur (BRZ09), CO₂-Ausstoß (HPM09(2mal)), A-Junioren (M09), D-Junioren (M09)

Unter den 896 nicht lexikalisierten Determinativkomposita weichen dagegen 92 Belege, also 10,3%, von einer Zusammenschreibung ab.

Diese Daten stützen auch die Ergebnisse der Blickbewegungsstudien von Pfeiffer (2002), der bei häufig belegten zweigliedrigen Komposita längere Blickzeiten und eine erhöhte Gesamtlesezeit feststellt, was sich damit deckt, dass in den hier untersuchten lexikalisierten Belegen eine Bindestrichschreibung nur dann vorkommt, wenn sie aufgrund eines graphematisch markierten Teilausdrucks absolut notwendig ist. Die verkürzte Gesamtlesezeit, die Pfeiffer (2002) bei einer Bindestrichschreibung von seltener belegten Komposita feststellt, spiegelt sich wiederum darin, dass unter den nicht lexikalisierten Belegen deutlich mehr Varianten abseits der Zusammenschreibungen, bei denen es sich meist um Bindestrichschreibungen handelt, zu finden sind. Die Zusammenschreibung bleibt aber auch für diese Fälle die mit großem Abstand prominenteste Variante.

³⁸ In 4.4.1 wird ausführlicher erläutert, welche Belege als lexikalisiert und welche als nicht lexikalisiert gelten, und welches Verfahren dieser Entscheidung zugrundeliegt.

³⁹ Um auszuschließen, dass Belege mit einer Variante abseits der Zusammenschreibung doch auch als Zusammenschreibung lexikalisiert sind, wurden die betreffenden Belege daraufhin überprüft, mit dem Ergebnis, dass keiner der 92 Belege in einer zusammengeschriebenen Variante lexikalisiert ist, auch nicht in einer anderen Variante bezüglich der Graphem-Phonem-Korrespondenzen bei fremdsprachlichen Anteilen oder der Beteiligung von Eigennamen.

Ich formuliere daher in Anlehnung an ZUS-MORPH mit ZUS-LEX eine weitere eigene Beschränkung selbst, um diese Beobachtungen in die Analyse integrieren zu können:

(48) ZUS-LEX

ZUSAMMENSCHREIBUNG BEI LEXIKALISIERUNG

Wenn X und Y Teilausdrücke eines morphologisch gebildeten Zeichens sind, das als Komplex lexikalisiert ist, sind X und Y graphematisch unselbständig.

Wie (47) zeigt, handelt es sich bei den Belegen, die gegen ZUS-LEX verstoßen, in allen Fällen um Wortformen, die aufgrund von *GRAPH einen Bindestrich aufweisen. Bindestrichschreibungen oder andere Varianten abseits der Zusammenschreibung, die nicht auf *GRAPH zurückgehen, finden sich dagegen ausschließlich unter den nicht lexikalisierten Belegen (92 von 896).

Die Beleglage spricht also eindeutig dafür, ZUS-LEX unterhalb von *GRAPH zu platzieren. Mit GPKB gibt es bereits eine Beschränkung, die eine große Schnittmenge mit ZUS-LEX hat. Während sich der Wirkungsbereich von GPKB über alle morphologischen Bildungen erstreckt, bezieht sich ZUS-LEX auf die Teilmenge der morphologischen Bildungen, die lexikalisiert sind. Hier liegt also zunächst eine Dopplung vor. Da bisher keiner der betrachteten Belege für ein anderes Verhältnis spricht, ist vorerst anzunehmen, dass GPKB und ZUS-LEX auf einer Hierarchiestufe stehen. Dass es trotzdem sinnvoll ist, an beiden Beschränkungen festzuhalten, wird sich mit der Formulierung und Platzierung weiterer Beschränkungen zeigen.

Damit gilt zunächst das Ranking in (49), ohne dass sich durch die zusätzliche Beschränkung ZUS-LEX die Passgenauigkeit der Beschränkungshierarchie verbessert.

(49) N°-GROSS » MAJUS » *GRAPH » ZUS-MORPH » GPKB, ZUS-LEX » GETR-
AUSDR » MINUS, *SPAT

Das folgende Tableau demonstriert die Anwendung dieser aktualisierten Hierarchie auf einen lexikalisierten Beleg aus dem Kriminalroman-Korpus.

(50)

	N ^o -GROSS	MAJUS	*GRAPH	ZUS-MORPH	GPKB	ZUS-LEX	GETR-AUSDR
a. Bildunterschrift (DIV)							*
b. Bild-Unterschrift					*!	*!	
c. Bild Unterschrift				*!			
d. BildUnterschrift		*!					*
e. Bild-unterschrift	*!				*		*
f. Bild unterschrift	*!			*			

Logistische Regression

Für die verbleibenden Einflussfaktoren, also Eigennamen als Teilausdrücke, fremdsprachliche Teilausdrücke, das Auftreten von Fugenelementen oder markierten Buchstabenfolgen und die Länge der Zeichenketten sind die Effekte auf die Ausprägung von Schreibvarianten nicht so eindeutig aus den Belegfrequenzen abzulesen wie bei den von *GRAPH oder ZUS-LEX betroffenen Belegen.

Um zu überprüfen, ob auch diese anderen Faktoren dennoch einen Einfluss auf Schreibvariation an der Kompositionsgrenze in zweigliedrigen Determinativkomposita haben, der nicht alleine auf einer zufälligen Verteilung beruht, wurde deshalb eine logistische Regression in R durchgeführt (R Core Team 2016).⁴⁰

Aufgrund der eindeutigen Effekte von der Lexikalisierung einer Wortform und Teilausdrücken, die keine graphematischen Wörter des Kernbereichs sind, wurden Belege, die einen solchen Teilausdruck aufweisen oder lexikalisiert sind, vor der Regression ausgeschlossen. Während ein markiertes graphematisches Wort als Teilausdruck bei 49 von 50 Belegen zu Varianten (Bindestrich- und Getrenntschreibung) führt, geht der Effekt der Lexikalisierung nicht weniger eindeutig in die entgegen-

⁴⁰ Teilausdrücke, die graphematisch markierten Wörtern entsprechen, und eine Lexikalisierung der Determinativkomposita konnten nicht in die Regressionsanalyse mit einbezogen werden, da es für diese problematisch ist, wenn ein Prädiktor wie die graphematisch markierten Teilausdrücke immer mit einer bestimmte Ausprägung der abhängigen Variable (hier der Nicht-Zusammenschreibung) zusammenfällt. Der Ausschluss der graphematisch markierten Teilausdrücke als Prädiktor bedingt dabei auf derselben Grundlage den des Lexikalisierungsstatus als Prädiktor, da abseits der Beteiligung graphematisch markierter Teilausdrücke alle lexikalisierten Verbindungen zusammengeschrieben werden, also auch hier eine 1:1 Beziehung besteht.

Tabelle 6.2.: Datenausschnitt: Regression Kompositionsgrenzen

Textform	Determinativkomposita	Ausschnitt (prozent. Anteil)
Werbeanzeigen	318	216 (68%)
Mail	81	44 (54%)
SMS	48	30 (63%)
Zeitung	238	105 (44%)
Fachzeitschriften	226	132 (58%)
Kriminalromane	118	63 (53%)
Plenarprotokolle	293	78 (27%)
Lexikoneinträge	206	116 (56%)
Diskussionsbeiträge	125	78 (62%)
Stichprobe 5000er	1.653	862 (52%)

gesetzte Richtung: unter den 757 lexikalisierten Belegen finden sich neben 741 Zusammenschreibungen nur 16 Bindestrichschreibungen, die allesamt auf *GRAPH zurückzuführen sind.

Dadurch verbleiben als Grundlage der logistischen Regression 862 substantivische, zweigliedrige Determinativkomposita, also etwa die Hälfte der Belege, die der Modellierung des Schreibverhaltens an Kompositionsgrenzen zugrunde liegen. Unter diesen 862 Wortformen finden sich neben 803 Zusammenschreibungen auch 59 andere Schreibvarianten (55 Bindestrichschreibungen, drei Binnemajuskeln und eine Getrenntschreibung), was einer Zusammenschreibungsquote von etwa 94% entspricht und sich damit gut mit der Ausgangsgruppe deckt.

712 der 862 Wortformen sind nur einmal belegt, nur 17 Wortformen finden sich mehr als zweimal im Datenausschnitt und unter den Belegen, die eine Variante abseits der Zusammenschreibung zeigen, ist einzig *Neckar-Nachrichten* (DIV) mehr als einmal, genau genommen zweimal, belegt. Gerade in der kleinen Gruppe der Belege, die nicht zusammengeschrieben werden, wären häufigere Dopplungen von Belegen oder ein einzelner Beleg in einer Vielzahl von Instanzen problematisch. So ist aber trotz des relativ geringen Anteils von abweichenden Varianten im Vergleich zur Zusammenschreibung auch ohne Lemmatisierung nicht davon auszugehen, dass das Ergebnis der Regression durch mehrere Instanzen derselben Wortform verfälscht wird.

Die Verteilung der Belege über die verschiedenen Textformen ist im Vergleich zur Gesamtheit der 1.635 zweigliedrigen Determinativkomposita weitgehend gleich geblieben (vgl. Tabelle 6.2). Nur die Gruppe der Belege, die dem Teilkorpus aus den Plenarprotokollen entstammen, ist vergleichsweise stark dezimiert worden von 293 auf 78 Belege, was weniger als 30% des Ausgangswertes entspricht, während diese Quote bei allen anderen Teilkorpora zwischen 44% und 68% schwankt. Dieses Phänomen lässt sich aber über den Ausschluss der lexikalisierten Belege erklären. Die Plenarprotokolle als Textform mit sehr stark ausgeprägten Mustern und formelhaften Wendungen bringen es mit sich, dass bestimmte Wortformen besonders häufig sind, und dadurch auch eher in der Wortliste erscheinen, die zur Bestimmung des Lexikalisierungsstatus genutzt wurde. Dass besonders viele Belege aus dem Teilkorpus Plenarprotokolle als lexikalisiert ausgeschlossen werden, zeigt damit die gute Passung von Korpus und Wortliste bei der Entscheidung über den Lexikalisierungsstatus. Der für die Regression genutzte Datenausschnitt kann also dennoch als repräsentativ gelten.

Auf der Basis dieses Datenausschnitts wurde eine logistische Regression (glm (R Core Team 2016)) mit der Schreibvariante (binär kodiert als Zusammenschreibung und davon abweichende Varianten)⁴¹ als abhängiger Variable durchgeführt. Das volle Modell umfasst neben den sechs unabhängigen Variablen in (51) auch alle möglichen Zwei-Wege-Interaktionen.

- (51)
- Eigennamen als erster Teilausdruck⁴²
 - fremder erster Teilausdruck
 - fremder zweiter Teilausdruck
 - Fugenelement
 - markierte Buchstabenfolge
 - Länge der Zeichenkette

Die Verteilung der beiden Ausprägungen der fünf sumgecodeten Prädiktoren (-.5,

⁴¹ D. h. also die 55 Bindestrichschreibungen, drei Binnenmajuskelschreibungen und die einzelne Getrenntschreibung wurden zu einer Kategorie von 59 Nicht-Zusammenschreibungen zusammengefasst, die den 803 zusammengeschriebenen Belegen gegenübergestellt wird.

⁴² Im Datenausschnitt sind Eigennamen nur als erster Teilausdruck belegt, weshalb diese Kategorie nur als ein Prädiktor in die logistische Regression mit eingeht.

Tabelle 6.3.: Datenausschnitt: Verteilung der binären Prädiktoren

Prädiktor (Kürzel)	betroffen	nicht betroffen
Eigenname als erster Teilausdruck (NameK1)	29	833
Fremder erster Teilausdruck (FremdK1)	219	643
Fremder zweiter Teilausdruck (FremdK2)	203	659
Fugenelement	304	558
Markierte Buchstabenfolge	49	813

.5) im Datenausschnitt zeigt sich in Tabelle 6.3 (Informationen zur Annotation der Prädiktoren finden sich unter Kapitel 4.4 Belegsammlung und Annotation). Die Länge der Zeichenketten bewegt sich bei einem Median von 13 zwischen neun und 29 Zeichen pro Beleg.⁴³ Da davon auszugehen ist, dass eine Verlängerung um ein einzelnes Zeichen bei kürzeren Belegen eher einen Effekt hat als bei Belegen, die schon 20 oder mehr Zeichen umfassen, wurden die Werte dieses Faktors logarithmisiert (10er Basis).

Um das finale Modell zu finden, wurde eine *backward elimination* mit likelihood-ratio tests durchgeführt.⁴⁴ Das heißt, aus dem vollen Modell mit allen Haupteffekten und Zwei-Wege-Interaktionen wurden Schritt für Schritt nicht signifikante Effekte herausgenommen, sofern ein Vergleich von Ausgangsmodell und vereinfachtem Modell mit anova (R Core Team 2016), nicht dafür gesprochen hat, dass das komplexere Modell die Daten signifikant besser als das vereinfachte erklärt hat. Auf diese

⁴³ Der einzelne Beleg, der in dieser Kategorie mit neun Zeichen geführt wird, obwohl aus den 5.000er-Stichproben erst Belege ab einem Umfang von 10 Zeichen erfasst wurden, ist *Opel-Krise* (WDD11). Dass ihm nur neun Zeichen zugeschlagen werden, obwohl er offensichtlich zehn Zeichen umfasst, erklärt sich dadurch, dass bei den Belegen mit Bindestrich oder anderen Interpunktions- und Sonderzeichen ebendiese nicht mitgezählt wurden, um die Ergebnisse hinsichtlich der Länge der Zeichenkette nicht dadurch zu verzerren, dass diese zusätzlichen Zeichen sich potentiell selbst begünstigen.

⁴⁴ Im Zuge der *backward elimination* hat sich die Interaktion von Namen als ersten Teilausdruck und markierter Buchstabenfolge als problematisch erwiesen. Die dünne Datenlage für die Kombination der beiden Prädiktoren (es gibt nur zwei Belege, die eine Kombination aufweisen, nämlich zwei Instanzen von *Mercedesstraße* (W)) führt zu extrem hohen und dadurch uninformativen Werten für den Standardfehler der Interaktion (Standardfehler = 1029,12 bei einem Schätzwert von 17,5) und der beteiligten Haupteffekte. Da das Modell hier offensichtlich keine zuverlässigen Schätzungen vornehmen kann, wird die Interaktion in der weiteren Analyse nicht berücksichtigt.

Tabelle 6.4.: Finales Modell der logistischen Regression

Effektname	Schätzwert	Standardfehler	χ^2	p-Wert
Intercept	8,21	2,24	3,67	p<0,001
NameK1	-3,39	0,53	40,22	p<0,001
FremdK1	-0,86	0,35	6,26	p<0,05
FremdK2	0,15	0,51	0,08	0,7730
Fugenelement	1,44	0,45	12,12	p<0,001
Länge Zeichenkette	-5,86	1,82	10,43	p<0,01
NameK1:FremdK2	2,82	1,02	7,02	p<0,01

Weise wurde das einfachste Modell ermittelt, ohne aussagekräftige Effekte zu eliminieren.

Dieses vereinfachte, finale Modell (vgl. Tabelle 6.4) beinhaltet signifikante Haupteffekte für die Prädiktoren Eigenname als erster Teilausdruck, fremder erster Teilausdruck, Fugenelement und die Länge der Zeichenkette, sowie eine signifikante Zwei-Wege-Interaktion für einen fremden zweiten Teilausdruck und Eigennamen als ersten Teilausdruck. Die χ^2 - und p-Werte für die einzelnen Effekte wurden ebenfalls durch Modellvergleiche (finales Modell vs. finales Modell, ohne den betreffenden Effekt) mittels anova (R Core Team 2016) ermittelt.

Demnach hat ein Eigenname als erster Teilausdruck einen hoch signifikanten Effekt ($\chi^2=40,22$; $p < 0,001$), die Länge der Zeichenkette einen sehr signifikanten Effekt ($\chi^2=10,43$; $p < 0,01$) und ein fremdes Erstglied hat einen signifikanten Effekt ($\chi^2=6,26$; $p < 0,05$) entgegen der Zusammenschreibung (erkennbar an den negativen Schätzwerten), der Einsatz eines Fugenelementes dagegen hat einen hoch signifikanten Effekt ($\chi^2=12,12$; $p < 0,001$) hin zur Zusammenschreibung (positiver Schätzwert).

Für ein fremdes Zweitglied und markierte Buchstabenfolgen konnten keine signifikanten Effekte festgestellt werden. Während also das Auftreten markierter Buchstabenfolgen zumindest für diese Stichprobe als Prädiktor für die Ausprägung von Schreibvarianten zu verwerfen ist,⁴⁵ und damit auch keine entsprechende Beschränkung formuliert wird, sind für fremde Teilausdrücke oder Eigennamen als Erstglied

⁴⁵ Fuhrhop (2008: 204) weist bereits darauf hin, dass eine von ihr durchgeführte Analyse des TIGER-Korpus (2005) ergeben hat, dass es sich bei Bindestrichschreibungen bei drei aufeinanderfolgenden gleichen Buchstaben um ein marginales Phänomen handelt.

sowie für das Auftreten von Fugenelementen wirksame Beschränkungen anzunehmen.

Der sehr signifikante Effekt der Länge der Zeichenkette ($\chi^2=10,43$; $p < 0,01$) weg von der Zusammenschreibung ist so zu deuten, dass entgegen der in den Hypothesen formulierten Erwartung mit zunehmender Länge des graphematischen Wortes als Zeichenkette tatsächlich die Wahrscheinlichkeit steigt, dass eine Grenzmarkierung an der Kompositionsgrenze erfolgt. Dieser Effekt beruht wahrscheinlich darauf, dass durch eine entsprechende Grenzmarkierung die Verarbeitung längerer Einheiten erleichtert werden soll, kann innerhalb dieses Rahmens aber nicht im theoretischen Modell abgebildet werden.

Erklärungsbedürftig bleibt der sehr signifikante Effekt der Interaktion von Eigennamen als ersten Teilausdruck und fremdem zweiten Teilausdruck ($\chi^2=7,02$; $p < 0,01$) in Richtung Zusammenschreibung. Den Hypothesen zufolge sollten beide Prädiktoren eigentlich einer Zusammenschreibung entgegen wirken. Das hat sich allerdings nur für den Eigennamenanteil bestätigt, während ein fremder zweiter Teilausdruck keinen signifikanten Effekt auf die Ausprägung von Schreibvarianten hat. Der signifikante Effekt der Interaktion geht nun aber in die entgegengesetzte Richtung, also hin zur Zusammenschreibung. Klarheit verschafft hier ein Blick in die Daten. So lässt sich feststellen, dass der sehr signifikante Effekt der Interaktion ($\chi^2=7,02$; $p < 0,01$) auf sieben Belegen (lemmatisiert nur sechs) beruht, die einen Eigennamen als ersten Teilausdruck und einen fremden zweiten Teilausdruck aufweisen:

- (52) Afghanistan-Konferenz (HAZ09), Whakatane-Distrikts (WPD11), Peco-Trainerinnen (VDI09), Treblinka-Prozess (WPD11), Parkinsonpatienten ((SPK) zweimal), Brundtlandbüro (M09)

Da drei dieser sieben Belege in (52) zusammengeschrieben werden, ist die Tendenz zur Zusammenschreibung im Vergleich zu anderen Determinativkomposita mit Eigennamen tatsächlich höher. Berücksichtigt man allerdings die Dopplung von *Parkinsonpatienten* (SPK) und den Umstand, dass *-büro* als fremdsprachlicher zweiter Ausdruck schon deutliche Integrationsmerkmale aufweist, bei der geringen Zahl der zugrunde liegenden Belege, relativiert sich der Effekt für die einzige im Modell verbliebene Interaktion.

Zwar belegt Tabelle 6.3, dass die Datengrundlage an sich gut geeignet für eine

Regressionsanalyse ist. Um wirklich valide Aussagen über die Effekte von Zwei-Wege-Interaktionen machen zu können, wäre aber eine breitere Datenbasis wünschenswert.

Eigennamen als Teilausdrücke

Kompositionen mit Eigennamen sind im Vergleich zu denen mit graphisch markierten Teilausdrücken mit 37 Fällen etwas seltener belegt, verhalten sich aber ähnlich.⁴⁶ Dementsprechend formuliere ich auf der Basis der Korpusfrequenzen und dem Ergebnis der logistischen Regression eine weitere Beschränkung zur Einbindung von Eigennamen in morphologischen Bildungen, die der Beschränkung zu den markierten graphematischen Wörtern ähnelt:

(53) NAME

EINBINDUNG VON EIGENNAMEN

Teilausdrücke, die Eigennamen entsprechen, bleiben auch innerhalb morphologischer Bildungen graphematisch eigenständig.

Obwohl hier nur Belege mit einem Eigennamen als ersten Teilausdruck vorliegen, ist auf der Basis der orthographischen Norm (§46), die einen Bindestrich bei Namen als zweiten Bestandteil (*Frau Müller-Weber, Annaberg-Buchholz*) obligatorisch nennt, davon auszugehen, dass es einen positionsunabhängigen Effekt abseits der prototypischen Determinativkomposita gibt.

Zur Positionierung der Beschränkung innerhalb der bisherigen Ordnung ist ein genauerer Blick in die Belege nötig. In der untersuchten Stichprobe ist diese Beschränkung für 37 Belege relevant. Die Mehrheit dieser Fälle, nämlich 27, wird mit einem Bindestrich zwischen den Teilausdrücken verschriftlicht und neben 10 Zusammenschreibungen ist auch eine einzelne Getrennschreibung belegt (vgl. (54)).

(54) a. Luther-Schau (Mog), Günter-Brand-Halle (NUZ09), Warrego-Tal

⁴⁶ Komposita mit Eigennamen als Zweitglied und damit Kopf sind selten und spielen in dieser Untersuchung keine Rolle, da monoreferentielle Eigennamen (und damit auch potentielle Kompositionsbildungen mit dementsprechenden Köpfen) als Belege ausgeschlossen wurden. In Kompositionen mit mehr als zwei Gliedern, sind dennoch wie bei *GRAPH die folgende und vorausgehende Grenze relevant, was daher auch schon hier in der Formulierung der Beschränkung NAME berücksichtigt wird.

(WPD₁₁)

- b. Guisanplatz (DIV), Parkinsonpatienten (SPK)
- c. Hoffmann-von-Fallersleben Museum (BRZ₀₉)

Für sieben der 37 Belege mit Eigennamenanteil ist nicht nur NAME relevant, sondern auch *GRAPH (z. B. für *Hoffmann-von-Fallersleben Museum* (BRZ₀₉) und *Günter-Brand-Halle* (NUZ₀₉)), so dass die Abweichung von der Zusammenschreibung auch durch die Einhaltung von *GRAPH begründet werden kann. Lexikalisiert sind zwei der Determinativkomposita mit Eigennamenanteil (*Willy-Brandt-Platz* (BRZ₀₉) und *Christuskirche* (WPD₁₁)).

Diese Beleglage spricht damit für ein Ranking von NAME über GPKB, da auch nach der Ausklammerung der sieben von *GRAPH betroffenen Belege mehr Belege mit Bindestrich vorliegen als ohne (20:10), aber unter ZUS-LEX (vgl. (55)), da sich der Bindestrich bei einem der beiden davon betroffenen Belege durch *GRAPH erklären lässt. Die Einordnung von NAME unter ZUS-LEX basiert damit auf nur einem Einzelbeleg, ist aber die konsequente Abbildung der Datengrundlage.

(55) N°-GROSS» MAJUS» *GRAPH» ZUS-MORPH» ZUS-LEX » NAME » GPKB
 » GETR-AUSDR» MINUS,*SPAT

Durch diese Positionierung löst sich außerdem die Doppelung von ZUS-LEX und GPKB auf einer gemeinsamen Stufe auf. Durch den Einschub von NAME zwischen die beiden Beschränkungen kann nämlich abgebildet werden, dass ein Eigenname als Teilausdruck mit einer erhöhten Wahrscheinlichkeit zur Grenzmarkierung führt, die auch in den meisten Fällen realisiert wird (deshalb über GPKB), es sei denn die Verbindung ist lexikalisiert (deshalb unter ZUS-LEX). Hier zeigt sich also, wie vorteilhaft es für die Modellierung ist, dass ZUS-LEX bei gleicher Wirkung wie GPKB nur die Teilmenge der lexikalisierten Verbindungen erfasst. Das so angepasste Ranking wird unter (56) am Beispiel von *Luther-Schau* (M₀₉) demonstriert.

(56)

	N ^o -GROSS	MAJUS	*GRAPH	ZUS-MORPH	ZUS-LEX	NAME	GPKB	GETR-AUSDR
a. ☞ Luther-Schau (Mo9)							*	*
b. Lutherschau						*!		*
c. Luther Schau				*!				
d. LutherSchau		*!				*		*
e. Luther-schau	*!						*	*
f. Luther schau	*!			*				

Dadurch verlieren zwar auch neun Belege (57), die trotz Namenanteil zusammengeschieden werden, ihren Status als optimal, unterm Strich verbessert sich aber die Deckung von Modell und Stichprobe, da durch diese Einordnung (55) nun 12 weitere Schreibungen an Kompositionsgrenzen, also in Summe 1.600 von 1.653, als optimal abgebildet werden, was einer Deckung 96,8% entspricht.

(57) Bonnetweiher, Guisanplatz (DIV), Hollerplatz (BRZO9), valentinstag (SMS), Mercedesstraße (W (2mal)), Parkinsonpatienten (SPK (2mal)), Brundtlandbüro (Mo9)

Auch wenn *GRAPH und NAME unterschiedliche Plätze innerhalb der Hierarchie der Beschränkungen einnehmen, sind die Parallelen in Formulierung und Wirkung doch deutlich. Ob es sinnvoll ist, sie auf einer abstrakteren Ebene zu einer Beschränkung – zur Anbindung markierter Einheiten (sei es graphisch oder semantisch) – zusammenzufassen, soll daher im Zuge der Anwendung auf weitere Wortbildungsmuster und der Spezifizierung auf die einzelnen Textformen geprüft werden. Für die hier zugrundeliegende Beleggruppe muss aber zunächst von zwei Beschränkungen ausgegangen werden.

Fremde Teilausdrücke

Bei der nächsten Gruppe, Determinativkomposita mit fremden Teilausdrücken, ergibt sich ein differenzierteres Bild. Ein wesentlicher Faktor dabei ist sicherlich, dass nun auch zweigliedrige Komposita betrachtet werden, deren beide Teilausdrücke fremd(sprachlich) sind wie in (58c), während Determinativkomposita aus zwei Ei-

genamen oder zwei nicht-graphematischen Wörtern nicht in der analysierten Stichprobe vorkommen.⁴⁷

- (58) a. Bakterien-Gemeinschaft (SPK), Businessziele (W), Browser-Fenster (Mail)
b. Abstellungstheater (NUZ09), Anlagekapitals (WPD11), Wind-Stadion (HMP09)
c. Actionprämie (W), Sozialpolitik (PBR), Burnout-Syndrom (VDI09)

Von den 1.653 untersuchten zweigliedrigen Determinativkomposita weisen 555 einen fremdsprachlichen ersten (58a) oder zweiten Teilausdruck auf (58b) oder eine Kombination fremder Teilausdrücke (58c).⁴⁸ Betrachtet man sich die Verteilung innerhalb der Belege und die jeweilige Quote der Zusammenschreibung in Tabelle 6.5, scheint besonders ein fremdes Zweitglied einen Effekt auf die Ausprägung von Schreibvarianten zu haben. Die Zusammenschreibungsquote bei einem ersten fremdsprachlichen Teilausdruck liegt bei 95%, während sie bei einem zweiten fremdsprachlichen Teilausdruck bei 84% liegt und damit auf demselben Niveau wie für fremdsprachlichen ersten und zweiten Teilausdruck.

Tatsächlich zeigt aber das Ergebnis der logistischen Regression, dass ausschließlich die Fremdsprachlichkeit des ersten Teilausdrucks einen signifikanten Effekt ($\chi^2=6,26$; $p < 0,05$) auf die Ausprägung von Varianten abseits der Zusammenschreibung hat. Für die Interaktion von einem Namen als Erstglied und einem fremden Zweitglied konnte sogar ein sehr signifikanter Effekt ($\chi^2=7,02$; $p < 0,01$) hin zur Zusammenschreibung festgestellt werden, obwohl es einen hoch signifikanten Haupteffekt für den Eigennamen ($\chi^2=40,22$; $p < 0,001$) in Richtung Grenzmarkierung gibt. Ein fremdes Zweitglied hemmt also den Effekt von einem Namen als Erstglied.

⁴⁷ Determinativkomposita mit zwei Teilausdrücken, die markierten graphematischen Wörtern entsprechen, kommen in der Gesamtbelegsammlung vor, beispielsweise in Form komplexer Abkürzungen wie *LSV-NOG* (PBR) (ein Neuordnungsgesetz für Landwirtschaftliche Sozialversicherungen). Für die theoretische Modellierung sind sie aber unproblematisch, da sie graphisch so umgesetzt werden wie Komposita mit einem betroffenen Teilausdruck. Zweigliedrige Komposita mit einem Eigennamen als Kopf tauchen in der Belegsammlung gar nicht auf, da sie als Eigennamen mit weitgehend arbiträrer Schreibung ausgeschlossen wurden.

⁴⁸ Davon in der logistischen Regression berücksichtigt wurden durch den Ausschluss lexikalisierter und von *GRAPH betroffener Belege nur 219 fremde Erst- und 203 fremde Zweitglieder.

Tabelle 6.5.: Verteilung fremder Teilausdrücke mit anteiliger Zusammenschreibung

	nicht fremder zweiter	fremder zweiter
nicht fremder erster	1098 (96%)	243 (84%)
fremder erster	207 (95%)	105 (86%)

Diese Diskrepanz zwischen dem Ergebnis der Regression und dem ersten Blick auf die Belege erklärt sich vor allem durch den Einfluss von *GRAPH und NAME, die (auch in Kombination mit einem fremden zweiten Teilausdruck) starke Variantentripper sind, und die Position des ersten Teilausdrucks besetzen. Die niedrige Quote an Zusammenschreibungen bei einem fremden zweiten Teilausdruck, für den im Zuge der logistischen Regression kein signifikanter Effekt ermittelt werden konnten ($\chi^2=0,08$; $p = 0,77$) ist also nicht auf das fremde Zweitglied zurückzuführen, sondern auf graphematische markierte Teilausdrücke (27 Fälle wie *CO₂-Emissionen* (PBR)) oder Eigennamen als Erstglieder (10 Fälle wie *Afghanistan-Konferenz* (HAZ09)).

Das Ergebnis der logistischen Regression passt auch zur Hypothese von Bredel (2008), dass der Bindestrich zwischen zwei grammatisch ungleichen Einheiten (hier einer fremden und einer nativen) steht, und deckt sich mit meinem Befund auf der Beleggrundlage nach Buchmann (2015: 239), dass bei Zweigliedrigkeit Belege mit einem fremden Bestandteil eher mit dem Bindestrich verschriftlicht werden und Belege mit zwei fremden Bestandteilen eher zusammengeschieden werden.

In einem anderen Rahmen zu überprüfen wäre, ob sich dieser Effekt tatsächlich auf einen fremden ersten Teilausdruck konzentriert. Falls dem so ist, stellt sich natürlich die Frage, wie diese Asymmetrie zustande kommt.

Im Rahmen dieser Untersuchung lässt sich aber aufgrund des bestätigten Effektes für fremdsprachliche erste Teilausdrücke bereits eine weitere Beschränkung in Anlehnung an *GRAPH und NAME formulieren. Sie ist aber erstmals auf die linke Seite der wortbildungsmorphologischen Grenze beschränkt:

(59) FREMD

FREMDSPRACHLICHE TEILAUSTRÜCKE

Fremde erste Teilausdrücke bleiben auch in morphologischen Bildungen graphematisch eigenständig.

Zur Positionsbestimmung innerhalb der Beschränkungsordnung erfolgt wieder eine detaillierte Analyse der für FREMD relevanten Belege. Da abseits von Belegen, die von *GRAPH betroffen sind, alle lexikalisierten Belege mit einem oder mehreren fremden Teilausdrücken (in Summe 181 Belege) zusammengeschrieben werden, ist FREMD unter ZUS-LEX zu platzieren. Nach den Ergebnissen der logistischen Regression ist auch nicht von einer Interaktion im Sinne einer Kopplung der Beschränkungen mit NAME auszugehen, so dass sich nur noch die Frage stellt, in welchem Verhältnis FREMD zu GPKB steht.

Von 1.653 Determinativkomposita haben 312 ein fremdsprachliches Erstglied und sind dadurch für FREMD relevant. Neben 286 Zusammenschreibungen finden sich darunter auch 26 Belege für andere Schreibvarianten, wie von den Beispielen unter (60) verdeutlicht.⁴⁹

- (60) a. Senioren-Initiativen (HMP09), Qualitäts-Garantie (W), Business-Klasse (W), Browser-Fenster (Mail)
 b. PrivatKredit (W)
 c. linguisten mist (SMS)

Wird FREMD über GPKB eingeordnet, werden nur noch 1.313 der 1.653 Belege, also knapp 80%, als optimal abgebildet, bleibt FREMD durch GPKB blockiert sind es weiterhin 1.600 Belege, also 96,8%, die als optimal abgebildet werden.

Aufgrund dieser Datenlage ist also zunächst nicht davon auszugehen, dass FREMD für diesen Datenausschnitt GPKB überzuordnen ist. FREMD wird zwar in die Beschränkungshierarchie aufgenommen, kann aber durch seine Position, wie in (61) zu erkennen, keine unmittelbare Wirkung entfalten.

- (61) N°-GROSS» MAJUS» *GRAPH» ZUS-MORPH» ZUS-LEX » NAME » GPKB


⁴⁹ Dieses Verhältnis zwischen Zusammenschreibung und anderen Varianten steht damit klar im Widerspruch zu Buchmann (2015: 237), die auf der Basis ihrer Sammlung von Bindestrichschreibungen (und der Suche nach bindestrichlosen Äquivalenten) davon ausgeht, dass Fremdwörter als Wortbestandteile bevorzugt mit einem Bindestrich segmentiert werden. Diese Diskrepanz beruht wahrscheinlich darauf, dass Buchmann (2015) zwar den Bindestrichschreibungen mit Fremdwortanteil ihre Zusammenschreibungsäquivalente gegenüberstellt, so aber durch den Zugang über die Bindestrichvariante die Lexeme gänzlich unberücksichtigt bleiben, die trotz Fremdwortanteil in ihrem Untersuchungskorpus ausschließlich zusammengeschrieben werden.

» FREMD » GETR-AUSDR » MINUS,*SPAT

Das Beschränkungsprofil eines optimalen, für FREMD relevanten Kandidaten wie *Browserfenster* in (62) verschlechtert sich dadurch lediglich im Vergleich zu einem optimalen Kandidaten ohne fremden Teilausdruck wie *Hausfenster*, der eben nicht gegen FREMD verstößt.

Diese unterschwellige Einordnung von FREMD ist möglicherweise dem Einfluss der orthographischen Norm geschuldet, die das Setzen eines Bindestrichs nur bei gänzlich fremdsprachlichen Verbindungen als Option einführt. Der häufigere Einsatz des Bindestrichs nach fremden Erstgliedern ist also ausdrücklich nicht lizenziert. Der signifikante Effekt von fremdsprachlichen ersten Teilausdrücken auf die Ausprägung von Schreibvarianten abseits der Zusammenschreibung ist also ein graphematischer Prozess, der ohne gegenläufige Norm noch deutlicher zu Tage träte.

(62)

	N°-GROSS	MAJUS	*GRAPH	ZUS-MORPH	ZUS-LEX	NAME	GPKB	FREMD	GETR-AUSDR
a.  Browserfenster								*	*
b. Browser-Fenster (MAIL)							*!		*
c. Browser Fenster				*!					
d. BrowserFenster		*!						*	*
e. Browser-fenster	*!						*		*
f. Browser fenster	*!			*					

Diesen Prozess bestätigt auch eine Analyse von 937 N+N-Komposita mit Bindestrichschreibung aus dem COW-Korpus durch Kopf (2017: 182-184) unter dem Gesichtspunkt der normgerechten Schreibung an der Kompositionsfuge. Unter den 29% der Schreibungen, in denen der Bindestrich nicht der Norm entspricht, machen Komposita mit fremdem Erstglied mit 71,5% wie *Acerola-Kirsche* und *Anime-Helden* den mit Abstand größten Anteil aus.

Tatsächlich stimmt der Einsatz des Bindestrichs nach fremden Erstgliedern genau mit dem Muster überein, dass schon *GRAPH und NAME zeigen: Nach markierten Strukturen (egal ob graphematisch, semantisch, morphologisch) gibt es eine Ten-

denz weg von der Zusammenschreibung. Bei Eigennamen und Teilausdrücken, die graphematisch markierten Wörtern entsprechen, ist der Effekt wohl nur deutlicher, weil er von der orthographischen Norm gedeckt ist.

Eine letzte Möglichkeit für FREMD, noch über GPKB seine Wirkung zu entfalten, könnte die Kombination mit einer Beschränkung sein, die auf dem beobachteten Effekt von Fugenelementen auf die wortinterne Schreibung basiert und im Folgenden diskutiert wird.

Fugenelemente

Anders als das Beispiel *Senioren-Initiativen* (HMP09) oben vermuten lässt, ergibt sich für Determinativkomposita, die ein Fugenelement wie das *-en-* im vorangegangenen Beispiel aufweisen, in der logistischen Regression ein hoch signifikanter Effekt ($\chi^2=12,12$; $p < 0,001$) hin zur Zusammenschreibung.

Von den hier betrachteten 1.653 zweigliedrigen Determinativkomposita weist mehr als ein Drittel ein Fugenelement auf. In nur neun von diesen genau 587 Fällen findet sich eine Abweichung von der Zusammenschreibung:

- (63) Funktionalitäts-Symbole, Qualitäts-Garantie (W), Senioren-Initiative (HMP09), Abstellungs-Theater (NUZ09), Heiducken-Bewegung, Georgs-Kirche (WPD11), Experten-Know-How, Ideen-Auswahl (VDI09), linguisten mist (SMS)

Diese Abweichungen von der Zusammenschreibung sind in drei Fällen durch *GRAPH (*Experten-Know-How* (VDI09)) und NAME (*Georgs-Kirche*, *Heiducken-Bewegung* (WPD11)) gedeckt und umfassen daneben auch drei Beispiele aus Werbetexten und SMS, die großzügiger mit Varianten umgehen.⁵⁰

Buchmann (2015: 238) weist drauf hin, dass 10% ihrer erfassten Bindestrichschreibungen ein Fugenelement aufweisen. Diesen Wert kann die vorliegende Untersuchung mit neun Belegen für eine Abweichung von der Zusammenschreibung trotz Fugenelement (63) auf 108 insgesamt abweichende Belege (8,3%) durchaus bestäti-

⁵⁰ Aufgrund der der Tatsache, dass die 108 Belege des für die logistische Regression gewählten Datenausschnitts, die nicht zusammengeschieden werden, auf neu Textformen verteilt, zum Teil sehr niedrige absolute Frequenzen aufweisen, bot sich leider nicht die Möglichkeit, die Kontexte schon in logistischen Regression mit einzubeziehen. Eine Diskussion der Textsortenspezifik erfolgt daher erst anhand des optimalitätstheoretischen Modells in 6.5.

gen.

Die besprochene Beleglage deckt sich auch mit den Ergebnissen von Kopf (2017), die die Kombinationsmöglichkeit von Bindestrich und Fugenelement in der Gegenwartssprache und dem Frühneuhochdeutschen (1500-1900) untersucht. Dabei beobachtet sie nicht nur eine Umkehrung der Verhältnisse von einer starken Kopplung im Frühneuhochdeutschen zum gegenseitigen Ausschluss im Gegenwartsdeutschen, sondern findet auch noch eine schlüssige Erklärung für die Extreme:

Zurückzuführen ist dies in beiden Fällen auf die Funktion des Bindestrichs als Segmentierungshilfe markierter Bestandteile: Das noch neue, aus Reanalyse syntaktischer Strukturen entstandene Kompositionsmuster beinhaltete auffällige Erstglieder. Für sie breitete sich zunächst die Zusammenschreibung aus, mit Aufkommen des Bindestrichs wurden diese Komposita jedoch schnell wieder graphematisch aufgebrochen und ihre ungewöhnliche Struktur markiert. Heute sind Erstglieder mit Fugenelement völlig unmarkierte Bestandteile von Komposita und entsprechend auch kaum mehr mit Bindestrich zu finden.

(Kopf 2017: 200)

Anders als Eigennamen, graphematisch markierte Wörter oder Teilausdrücke aus einer fremden Sprache, die als markierte Bestandteile einen Bindestrich begünstigen, zeugt ein Fugenelement also von Unmarkiertheit.

Dementsprechend formuliere ich die Beschränkung ZUS-FE:

(64) ZUS-FE

FUGENELEMENTE

Teilausdrücke mit Fugenelement sind graphematisch unselbständig.

Die Positionierung innerhalb der Hierarchie ist aber noch offen. Von einem Ranking über *GRAPH ist nicht auszugehen. Der einzige Beleg, der dafür spräche, ist *Staats(un)wesen* (WPD₁₁). Wie oben schon besprochen handelt es sich dabei aber eigentlich um eine Koordinierung von zwei Zusammenschreibungen (*Staatswesen* und *Staatsunwesen*) und keinen Einzelbeleg mit zweitem Ausdruck, der durch die Klammer markiert ist.

Nach einem Blick in die Belege scheidet auf der vorliegenden Datenbasis auch die Option, ZUS-FE über NAME einzuordnen, aus. Es gibt nur drei Belege, die einen Namen als ersten Teilausdruck aufweisen, der auch ein Fugenelement aufweist (65).

- (65) Heiducken-Bewegung (WPD₁₁), Georgs-Kirche (WPD₁₁), valentinstag (SMS)

Nur einer dieser Fälle (*valentinstag* (SMS)) wird zusammengeschrieben, während die beiden anderen Belege mit Bindestrich realisiert sind.

Offen ist damit noch das Zusammenspiel von FREMD und ZUS-FE. Hier gibt es mit 73 Belegen immerhin eine deutlich größere Schnittmenge. Die Tatsache, dass von diesen 73 Belegen mit Fugenelement und fremdem ersten Teilausdruck 67 zusammengeschrieben werden, also nur sechs Varianten mit einer Grenzmarkierung auftreten, spricht für die Dominanz von ZUS-FE über FREMD.

- (66) Funktionalitäts-Symbole (W), Senioren-Initiativen (HMP₀₉), Qualitäts-Garantie (W), Experten-Know-How (VDI₀₉), Ideen-Auswahl (VDI₀₉), linguisten mist (SMS)

Die Beispiele unter (66) zeigen außerdem, dass darunter mit *Experten-Know-How* (VDI₀₉) noch ein Beleg ist, bei dem *GRAPH ZUS-FE schlägt.

Klar ist also, dass ZUS-FE über FREMD steht. Noch offen ist, ob die Kombination beider Beschränkungen unterschwellig, also GPKB untergeordnet bleibt, oder ob ZUS-FE FREMD soweit einschränkt, dass eine gemeinsame Platzierung über GPKB geboten ist.

Im Vergleich zur Platzierung von FREMD über GPKB, die nur 1.313 Belege als optimale Varianten abbildet, ergeben sich durch die Einschränkung von FREMD durch ZUS-FE an selber Stelle immerhin 1.323 optimal abgebildete Varianten, also ein Zuwachs von 10 Belegen. Um die größtmögliche Anzahl an Belegen durch das Ranking als optimale Kandidaten zu erfassen, bleibt die Kombination von FREMD und ZUS-FE für die zweigliedrigen Determinativkomposita also unterschwellig.

Aufgrund dessen, dass ZUS-FE die Zusammenschreibung ohne Bindestrich oder Ähnliches für Komposita mit Fugenelement verlangt und damit eine Teilmenge der Beschränkung GPKB abbildet, die den Einsatz von graphematischen Segmenten ohne phonologisches Korrelat verbietet, wird diese Unterschwelligkeit auch schon dadurch erreicht, dass GPKB und ZUS-FE auf einer gemeinsamen Stufe stehen – wodurch ZUS-FE obsolet wird.

Dadurch werden durch das erweiterte Ranking in (67), das unter (68) am Beleg

Staatsexamen (WPD11) veranschaulicht wird, unverändert weiterhin 1.600 Belege also 96,8% als optimale Kandidaten abgebildet.

(67) N°-GROSS » MAJUS » *GRAPH » ZUS-MORPH » ZUS-LEX » NAME » GPKB, ZUS-FE » FREMD » GETR-AUSDR » MINUS, *SPAT

(68)

	N°-GROSS	MAJUS	*GRAPH	ZUS-MORPH	ZUS-LEX	NAME	GPKB	ZUS-FE	FREMD	GETR-AUSDR
a. Staatsexamen (WPD11)										*
b. Staats-Examen					*!		*	*		*
c. Staats Examen				*!	*			*		
d. StaatsExamen		*!								*
e. Staats-examen	*!				*		*	*		*
f. Staats examen	*!			*	*			*		

6.2.3 Modell für Kompositionsgrenzen in prototypischen Determinativkomposita

Aufbauend auf dem Basisranking in (69), das durch die Modellierung der Zusammenschreibung als optimalen Kandidaten bereits 93,5% der prototypischen Determinativkomposita als solche erfasst, konnte der Deckungsgrad des Modells durch weitere Beschränkungen, die auch Bindestrichvarianten unter bestimmten Umständen als optimale Kandidaten erfassen, auf 96,8% erhöht werden.

(69) N°-GROSS » MAJUS » ZUS-MORPH » GPKB » GETR-AUSDR » MINUS, *SPAT

Neben den Beschränkungen des Basisrankings umfasst das volle Modell in (70) für die Kompositionsgrenze in prototypischen Determinativkomposita auch die fünf im vorigen Kapitel eingeführten, zusätzlichen Beschränkungen für graphematisch markierte Wörter oder Eigennamen als Teilausdrücke, die Lexikalisierung des Kompositums, das Auftreten von Fugenelementen an der Kompositionsgrenze und für einen fremdsprachlichen ersten Teilausdruck.

- (70) N°-GROSS» MAJUS» *GRAPH» ZUS-MORPH» ZUS-LEX» NAME » GPKB,
ZUS-FE » FREMD » GETR-AUSDR» MINUS,*SPAT

Die Verbesserung des Deckungsgrades um gut 3% ist dabei allerdings ausschließlich den drei Beschränkungen *GRAPH, ZUS-LEX und NAME und ihrer Positionierung innerhalb der Hierarchie (70) geschuldet, während FREMD und ZUS-FE unerschwinglich bleiben. Die durch die Auslassung von FREMD und ZUS-FE vereinfachte Beschränkungsordnung in (71) erreicht also ebenso eine 96,8%ige Deckungsquote.

- (71) N°-GROSS» MAJUS» *GRAPH» ZUS-MORPH» ZUS-LEX» NAME » GPKB
» GETR-AUSDR» MINUS,*SPAT

Die Übertragung des Modells für determinative Kompositionsgrenzen in prototypischen, zweigliedrigen Komposita auf ähnliche Wortbildungsmuster und andere Wortbildungsarten, die textformspezifischen Anwendung und die Übertragbarkeit auf alle Belege der 5.000er-Stichproben böten noch die Möglichkeit, dass die beiden Beschränkungen im Falle einer notwendigen Anpassung des Modells doch noch entscheidend sein könnten. Im Folgenden wird sich aber zeigen, dass FREMD und ZUS-FE in den hier modellierten Kontexten durchgängig unerschwinglich bleiben. Nichtsdestotrotz sind FREMD und ZUS-FE evidenzbasiert und unabhängig motiviert, so dass ich bei dem vollen Modell bleibe.

Unabhängig davon, ob das volle Modell oder die reduzierte Variante zur Anwendung kommt, werden von den nur 53 suboptimalen Belegen (unter 1.653 Gesamtbelegen) außerdem ganze 45 Belege als Kandidaten abgebildet, die durch die Hierarchie nur eine Stufe unter dem optimalen Kandidaten stehen, also nach dem Modell die bestmögliche Alternative sind. Dazu gehören beispielsweise Getrenntschreibungen bei von *GRAPH betroffenen Belegen wie *USB Verlängerung* (SMS) oder Bindestrichschreibungen bei fremden Teilausdrücken wie *Ticket-Verkauf* (HMP09). Berücksichtigt man neben der ersten also auch die zweite Stufe der Optimalität wird eine Deckung von nahezu 100% erreicht. Unter diesem Niveau bewegen sich lediglich die folgenden acht von 1.653 Belegen:

- (72) VDI nachrichten (VDI09, 2mal), LautsprecherBox (Mail), PrivatKredit (W), FeuchtCreme (W), linguisten mist (SMS), Staats(un)wesen (WPD11),

(Kämpfer-)Lieder (WDD₁₁)

6.2.4 Übertragbarkeit auf ähnliche Wortbildungen

Neben den prototypischen zweigliedrigen Determinativkomposita, die Grundlage der Modellierung für Kompositionsgrenzen sind, finden sich unter den Belegen aus der 5.000er-Stichprobe noch weitere zweigliedrige Komposita, die ebenfalls eine determinative Struktur haben, aufgrund von Besonderheiten der (einzelnen) Teilausdrücke aber keine prototypischen Determinativkomposita sind.⁵¹

Im Folgenden soll für Phrasenkomposita, Affixoidbildungen, Inversionskomposita und Konfixkomposita diskutiert werden, ob das erarbeitete Modell auch auf diese Wortbildungen anwendbar ist.

Phrasenkomposita

Unter den Belegen aus der 5.000er-Stichproben finden sich 18 substantivische, zweigliedrige Determinativkomposita mit phrasalem Erstglied:

- (73) a. 2Hand-Laden,⁵² SecondHand-Laden (Mail), 2:2-Schützen, 2,5:1,5-Führung (BRZ09), Grand-Slam-Titel, Top-Ten-SpielerIn, Grand-Slam-Turnieren, "Top Twenty"-SpielerIn (WPD₁₁), 2-Länder-Antrag (2mal), Mehr-Länder-Antrag, Zwei-Stufen-Verfahren (PBR), Zwölf-Uhr-Sendung (DIV), Zwei-Körper-System (SPK)
- b. acht Uhr Vorstellung (SMS)
- c. Dreistufenmodell (PBR), Zweihöfesiedlung (WPD₁₁), Drei Säulentheorie (WDD₁₁)

14 davon werden mit einem Bindestrich an der Kompositionsfuge zwischen phrasa-

⁵¹ Kopulativkomposita, die aufgrund einer anderer Konstituentenstruktur auch ein anderes Verhalten hinsichtlich der Schreibvariation an der Kompositionsgrenze zeigen könnten, sind im Datenausschnitt nicht belegt.

⁵² 2Hand-Laden wurde in Analogie zu SecondHand-Laden als Phrasenkompositum klassifiziert, wodurch die Grenze zwischen 2 und Hand nicht als Wortbildungsgrenze, sondern als phrasale Grenze gilt und die Zusammenschreibung dort nicht als Verstoß gegen *GRAPH zählt. Bei einer isolierten Betrachtung wäre von einem dreigliedrigen Determinativkompositum auszugehen, das aufgrund des fehlenden Bindestrichs nach 2 dann nicht nur ein suboptimaler Kandidat wäre, sondern aufgrund seiner Dreigliedrigkeit dann hier auch noch nicht relevant wäre.

lem Erstglied und dem Kopf des Determinativkompositums verschriftlicht (73a). Die Bindestrichvariante ist dabei entsprechend *GRAPH für alle 14 Belege der optimale Kandidat, da die Erstglieder allesamt nicht zum Kernbereich der graphematischen Wörter gehören, sondern aufgrund inhärenter Interpunktionszeichen (*Mehr-Länder*, „*Top Twenty*“), Ziffern (*2Hand*, 2:2) oder einer Binnenmajuskel (*SecondHand*) markiert sind.

Auch die Getrennschreibung unter (73b) folgt *GRAPH, indem *Vorstellung* und das aufgrund des internen Leerzeichens markierte *acht Uhr* graphematisch selbstständig bleiben. Weil die graphematische Selbstständigkeit nicht durch einen Bindestrich, sondern durch ein Leerzeichen markiert wird, handelt es sich gemäß der Beschränkungshierarchie zwar nicht um einen optimalen Kandidaten, da *GRAPH noch über ZUS-MORPH steht, ist die Getrennschreibung zumindest die zweitbeste Wahl.

Auch Zusammenschreibungen von Phrasenkomposita wie unter (73c) sind mit dem Modell für Kompositionsgrenzen problemlos vereinbar, sofern der phrasale Bestandteil wie bei *Dreistufenmodell* (PBR) und *Zweihöfesiedlung* (WPD₁₁) graphematisch unmarkiert ist. Da graphematisch sowohl *Dreistufen* als auch *Zweihöfe* dem Kernbereich entsprechen, gibt es keine Beschränkung, die eine Variante abseits der Zusammenschreibung fordert. Wahrscheinlich werden die phrasalen Bestandteile dieser Komposita vom Schreiber nicht mehr als solche wahrgenommen und entsprechend auch nicht als solche wortintern gekennzeichnet durch den Durchkoppelungsbindestrich, Anführungszeichen und/oder eine Getrennschreibung. Entscheidend für die Markierung der Kompositionsgrenze, bzw. die damit einhergehende graphematische Selbstständigkeit der Teilausdrücke, ist damit nicht die phrasale Struktur des Erstgliedes selbst, sondern die Anzeige derselben über eine graphische Markiertheit.

Drei Säulentheorie (WDD₁₁) verletzt dementsprechend *GRAPH, da trotz internem Leerzeichen beim phrasalen Erstglied (*Drei Säulen*) zusammengeschrieben wird. Nach dem Ranking für Kompositionsgrenzen steht diese Verschriftlichung also gleich zwei Stufen unter dem optimalen Kandidaten.

In Summe werden also 16 der 18 Belege für Phrasenkomposita als optimale Kandidaten erfasst, weshalb davon auszugehen ist, dass das Modell für Kompositionsgrenzen in prototypischen, zweigliedrigen Determinativkomposita auch auf die Kompositionsgrenzen zwischen phrasalem Erstglied und Kopf in Phrasenkomposita

anwendbar ist.

In den Daten von Buchmann (2015: 257) zeigt sich möglicherweise ein Lexikalisierungseffekt hinsichtlich der phrasalen Bestandteile, der sich auch auf die Schreibung an der Kompositionsgrenze auswirkt. Während die selteneren bis okkasionellen Kompositionen *Rund-um-die-Uhr-Bereitschaft* und *Von-Haus-zu-Haus-Geschäft* ausschließlich durchkoppelt belegt sind, finden sich die Beispiele *Nach-Hause-Weg* (6) und *Trimm-dich-Pfad* (71), die häufiger vorkommen, auch mit Zusammenschreibung des phrasalen Bestandteils (*Trimmdich-Pfad* (3), *Nachhause-Weg* (8)) oder gänzlich zusammengeschrieben (*Trimmdichpfad* (1), *Nachhauseweg* (602)). An der in Klammern angegebenen Häufigkeit der Varianten lässt sich erkennen, dass die Zusammenschreibung *Nachhauseweg* bereits als Standard angenommen werden muss, für *Trimm-dich-Pfad* die Schreibung mit Durchkoppelungsstrich aber die beliebteste Variante bleibt. Ob sich die beiden Beispiele tatsächlich in unterschiedlichen Stadien eines Lexikalisierungsprozesses befinden, der zunächst zur Zusammenschreibung des phrasalen Bestandteils führt und anschließend zu gänzlichen Zusammenschreibung (aufgrund des durch die Zusammenschreibung der Phrase dann unmarkierten ersten Teilausdrucks), könnte nur eine (diachrone) Untersuchung einer breiteren Datenbasis überprüfen.

Ob und wie die Wortgrenzen innerhalb der phrasalen Teilausdrücke angezeigt werden, ist nicht im eigentlichen Sinne Teil des Untersuchungsgegenstandes, da hier wortbildungsmorphologische Grenzen im Wort und nicht die Grenzen zwischen Wörtern untersucht werden. Dennoch werden die syntaktisch verknüpften Wörter (und die dazwischenliegenden Grenzen) im Zuge der Komposition zu Wortbestandteilen. Wendet man das aktuelle Modell deshalb auch auf diese Grenzen an, ergibt sich auch dort die Zusammenschreibung als optimaler Kandidat, was allerdings der Beleglage unter (73) entgegensteht, die neben zwei Zusammenschreibungen, auch drei Getrenntschreibungen und acht Einsätze des Bindestrichs innerhalb des phrasalen Teilausdrucks zeigt. Vermutlich wird der Bindestrich von den Schreibern hier bevorzugt, da durch ihn trotz ununterbrochener Graphemkette die Wortgrenzen innerhalb des Syntagmas transparent bleiben und er der orthographischen Norm §44 entspricht. Die Zusammenschreibung von Phrasenkomposita an syntaktischen und wortbildungsmorphologischen Grenzen im Wort als optimaler Kandidat nach dem vorliegenden Modell stützt dagegen die Idee eines Lexikalisierungsprozesses, indem

sie den Zielpunkt einer graphematisch unauffälligen Schreibung setzt.

Affixoidbildungen

Neben den Phrasenkomposita finden sich noch 22 zweigliedrige Affixoidbildungen unter den Belegen aus den 5.000er-Teilkorpora. Wie schon bei den Phrasenkomposita unterscheidet sich auch hier das Wortbildungsmuster nicht von dem der prototypischen Determinativkomposita. Die Besonderheit liegt wie beim phrasalen Erstglied der Phrasenkomposita lediglich an den eingebundenen Konstituenten, hier also den Affixoiden, denen eine Art Übergangstellung zwischen Kompositionsgliedern und Affixen zugeschrieben wird.

Im Datenausschnitt sind, wie die Beispiele unter (74) zeigen, die Affixoide *Groß-*, *Grund-*, *Haupt-* und *Klein-* als Erstglieder sowie *-werk* und *-wesen* als Zweitglieder belegt.

- (74) Großfürstin (WDD₁₁, Grundgesetz (PBR), Hauptkommissar (DIV), Kleinarbeit (WDD₁₁), Meisterwerk (DIV), Staatswesen (PBR))

Ausnahmslos alle Belege, die als Affixoidbildungen klassifiziert sind, werden wie die Beispiele unter (74) zusammengeschrieben. Da keine der Affixoidbildungen für *GRAPH oder NAME relevant ist, fügen sich auch diese Zusammenschreibungen problemlos in das oben erarbeitete Modell für Kompositionsgrenzen in Determinativkomposita ein.

Inversionskomposita

Bei Inversionskomposita liegt der Wortakzent anders als bei prototypischen Determinativkomposita auf dem Zweitglied und spricht dafür, dass hier das Zweitglied das erste Glied näher bestimmt (vgl. Elsen 2011: 67).

Unter den Belegen dieses Datenausschnittes finden sich neun Inversionskomposita, allesamt verschiedene Wortformen der beiden Lexeme *Jahrhundert* und *Jahrzehnt*. Auch diese fügen sich problemlos in das Modell für determinative Kompositionsgrenzen.

Konfixkomposita

Bei den Konfixkomposita, für die es 23 zweigliedrige Belege im Datenausschnitt gibt (75), zeigen sich anders als bei den Affixoidbildungen und Inversionskomposita

wieder Bindestrichvarianten – allerdings nur vereinzelt und nicht so dominant wie bei den Phrasenkomposita.

- (75) a. Bürokraten, Agrarpolitik (DIV), Protokolle (Mail), Lokomotiven, Quintadena, Katastrophe (WPD₁₁), Atmosphäre, Mikrosekunde, Psychotherapie (SPK), Interessen (PBR, 3mal), Milliarden (SPK), Milliarden (HMP₀₉, 2mal), Paraffinen, Elektromobilität (W), Radiophoton (WDD₁₁), Psychotherapie (VDI₀₉)
- b. Nanobereich (Mo₉)
- c. Bio-Handel (W), Mini-Teens (BRZ₀₉), Ultra-Rechte (WPD₁₁)

Die Konfixkomposita, die ausschließlich fremdsprachliche Konfixe umfassen wie *Lokomotiven* oder durchgängig aus fremden Teilausdrücken aufgebaut sind wie *Mikrosekunde* oder *Elektromobilität*, werden offensichtlich in der Regel zusammengeschieden (vgl. (75a)). Eine Zusammenschreibung von fremdem Konfix und nativem Zweitglied ist mit *Nanobereich* dagegen nur einmal belegt, während *Bio-Handel* und *Ultra-Rechte*, die eine parallele Struktur aufweisen, mit Bindestrich verschriftlicht sind. Die dritte Bindestrichvariante *Mini-Teens* besteht dagegen durchgängig aus fremdsprachlichem Material.

Dementsprechend lassen sich auch diese Konfixkomposita gut durch das vorgelegte Modell erfassen. Alle Zusammenschreibungen werden als optimale Kandidaten erfasst und die drei Bindestrichvarianten, die unterschwellig durch FREMD getriggert werden, stehen auf der zweiten Stufe der Optimalitätsordnung. Dennoch wäre die Untersuchung einer größeren Belegmenge für den Bereich der Konfixkomposita wünschenswert, um abschätzen zu können, ob sich beispielsweise fremde Konfixe, die stets gebunden sind, als Erstglieder in zweigliedrigen Determinativkomposita tatsächlich so verhalten, wie fremde Erstglieder, die auch frei vorkommen, und ob die Nativität des Zweitgliedes eine signifikante Rolle spielt in Bezug auf die wortinterne Schreibung.

Zusammenfassung

Das Modell für die Kompositionsgrenze in prototypischen, substantivischen Determinativkomposita mit zwei Gliedern unter (70) lässt sich entsprechend der vorangegangenen Diskussion aufgrund seiner guten Passung also auch auf die zweiglied-

rigen Phrasenkomposita, Affixoidbildungen, Inversionskomposita und Konfixkomposita als Spezialfälle der Determinativkomposition übertragen. Von den 72 diskutierten Belegen werden 67 als optimale Kandidaten erfasst, was einer Quote von 93,06% entspricht, die nur geringfügig unter den 96,8% Deckungsgrad liegt, die für prototypische, zweigliedrige Determinativkomposita erreicht werden.

Da durch das entwickelte Modell, zweigliedrige Phrasenkomposita, Affixoidbildungen, Inversionskomposita und Konfixkomposita miteingeschlossen, 1.667 von 1.725 Belegen, also 96,6%, als optimale Kandidaten erfasst werden, eignet sich das Modell eindeutig auch für den nicht-prototypischen Bereich der Determinativkomposition. Dass die Zusammenschreibungsquote der diskutierten Komposita nur bei 92,7% liegt, zeigt dass neben einem Großteil der Zusammenschreibungen – nur vereinzelte Zusammenschreibungen sind nicht modellkonform – auch viele davon abweichende Bindestrichschreibungen als optimale Kandidaten erfasst werden.

Die Eignung des Modells für komplexer Komposita und den großen Bereich der Derivation und anderer Wortbildungsmuster bleibt aber noch offen.

6.3 Modellierung für Derivationsgrenzen

Auf der Basis der Beschränkungen, die bei der Modellierung der Schreibung an Kompositionsgrenzen in zweigliedrigen Determinativkomposita formuliert wurden, erfolgt nun eine Modellierung für Derivationsgrenzen.

In den 5.000er Stichproben finden sich 861 substantivische Belege, die in die Kategorie der Derivationen fallen. Darunter sind neben 790 Suffigierungen auch 15 Präfigierungen. Zirkumfigierungen sind innerhalb dieses Datenausschnittes nicht belegt. Die 56 impliziten Derivationen durch Lautwechsel wie *Zusammenhang* (PBR) bleiben zunächst unberücksichtigt, da sich dieser morphologische Prozess nicht an einer morphologischen Grenze abspielt.

Wie schon bei der Modellierung für Kompositionsgrenzen konzentriere ich mich auch hier zunächst auf die größte Gruppe, also die der Suffigierungen.

Die Verteilung der 790 Suffigierungen über die verschiedenen Textformen zeigt Tabelle 6.6. Die vergleichsweise niedrige Treffermenge für Suffigierungen erklärt sich dabei durch den Recherchemodus. Erfasst wurden Zeichenketten mit einer Mindest-

Tabelle 6.6.: Datenausschnitt: Modellierung Derivationsgrenzen

Textform	Suffigierungen
Werbeanzeigen	97
Mail	33
SMS	13
Zeitung	91
Fachzeitschriften	132
Kriminalromane	38
Plenarprotokolle	188
Lexikoneinträge	86
Diskussionsbeiträge	112
Stichprobe 5000er	790

länge von zehn Zeichen. Dieses Mindestmaß erreichen Komposita durch die Kombination wortfähiger Bestandteile sehr viel schneller und häufiger als Affigierungen, da die meisten Affixe recht kurz sind. Unter den Belegen für Suffigierungen finden sich entsprechend viele Suffigierungen bereits wortbildungsmorphologisch komplexer Basen wie in (76a) und nur wenige einfache Suffigierungen wie in (76b).

- (76) a. Krankschreibung, Informatikerin (VDI₀₉), Landwirtschaft (DIV), Öffentlichkeit (VDI₀₉), Wassersportler (W), Information (SPK), Verbundenheit (PBR), Spezialisten (W), Aktivitäten (WPD₁₁), Diskografie (WPD₁₁)
- b. Hindernissen (SMS), Leidenschaft (W), Krankheiten (VDI₀₉)

Dass alle angeführten Beispiele zusammengeschieden werden, ist dabei kein Zufall. Tatsächlich liegt bei allen 790 Suffigierungen eine Zusammenschreibung an der terminalen Derivationsgrenze (und auch an allen anderen Wortbildungsgrenzen) vor.

An den 790 Suffigierungen lassen sich leider nicht alle Prädiktoren beobachten, die für die Modellierung der Schreibung an Kompositionsgrenzen eine Rolle spielen. So gibt es keinen Beleg für die Suffigierung eines markierten graphematischen Wortes und damit auch keinen Ansatzpunkt für *GRAPH bzw. keinen Hinweis auf eine Platzierung innerhalb der Hierarchie.

Der Lexikalisierungsgrad der Suffigierungen ist sehr hoch. Während 711 Belege lexikalisiert sind, gelten nur 79 Belege als nicht lexikalisiert. Im Zusammenhang damit zeigt auch ein Blick in die Frequenzen einzelner Belege, dass im Vergleich zu den Determinativkomposita deutlich weniger nämlich nur 346 von 790 Suffigierungen (also weniger als die Hälfte) als einmalige Vorkommen belegt sind, 110 Belege kommen zwei bis viermal vor und 22 Belege fünfmal oder häufiger. Die Frequenzspitzenreiter in (77) entstammen dabei größtenteils dem Protokoll-Teilkorpus und sind symptomatisch für die dort protokollierten Standardprozesse:

- (77) Minderheit (33mal), Gesellschaft (12mal), Sicherheit (11mal), Verordnung, Informationen, Abstimmungen (jeweils 9mal)

Die hier betrachteten 790 Lexeme umfassen also nur 478 Wortformlemmata.

Da Fugenelemente für Derivationsgrenzen nicht relevant sind, bleibt nur noch der Blick auf die Prädiktoren, die an der Kompositionsgrenze einen Effekt hin zur Ausbildung von Varianten abseits der Zusammenschreibung haben. Bei den 478 Lemmata lassen sich zwar in zwei Fällen Eigennamen (78) und in vielen Fällen auch fremdsprachliche Teilausdrücke (80) beobachten, aber in allen davon betroffenen Belegen wird unverändert zusammengeschieden.

- (78) Parkinsonoid (WDD11), Amerikaner (HAZ09, 4mal)

Im Fall von Personennamen ist abseits des hier definierten Untersuchungsgegenstandes im Bereich der Flexion und der adjektivischen Wortbildung durchaus mit einer auf NAME basierenden Variante zu rechnen. So sieht die orthographische Norm nach §97 E die Möglichkeit einer Apostrophierung der Flexionsendung -s (für den Genitiv Singular wie in *Carlo's Taverne*) und des Wortbildungssuffixes -sche beispielsweise in *Einstein'sche Relativitätstheorie* zur Verdeutlichung der Grundform des Eigennamens vor. Außerhalb des hier gewählten kleinen Korpusausschnittes finden sich im mehr als drei Millionen laufende Wortformen umfassenden Gesamtkorpus dafür vereinzelte Belege:

- (79) a. Sissy's (NUZ09), Beck's, Schweighofer's (WPD11)
 b. Beck'sche, Moore'sche, Bluhm'schen (VDI09), lacan'schen (2mal), Huygen'schen, Arrow'schen (2mal), Beck'sche, Frankh'sche, Gauss'schen,

Goetz'scher, Magellan'schen, C.H.Beck'sche (3mal), Gail'scher (3mal)
Manz'schen, Späth'sches (WPD₁₁)

Die Apostrophierung des adjektivbildenden Suffixes *-sche* (79b) ist dabei deutlich häufiger belegt als eine Suffigierung der Genitivendung *-s* (79a). Allerdings konzentrieren sich die Belege stark auf die Textform der Online-Lexikoneinträge, was damit zusammenhängt, dass Ableitungen von Eigennamen in dieser Textform häufiger sind als in anderen.⁵³ Ebenfalls im Teilkorpus der Lexikoneinträge findet sich mit *Jung-Stillingschen* der einzige Beleg dieses Musters bei dem keine Apostrophierung realisiert wird.

Auch wenn diese Belege vorerst keinen Einfluss auf die Modellierung für nomenbildende Suffixe und ihre Derivationsbasen haben, liefern sie einen ersten Hinweis darauf, dass auch im Bereich der Suffigierung ein Name als Teilausdruck Einfluss auf die Schreibvariante haben kann. Obgleich NAME seine Wirkung nur bei Personennamen als Basis und *-sche* als adjektivbildendem Suffix und auch dort einzig durch die Einfügung eines Apostrophs entfalten kann, somit also gleich in dreifacher Hinsicht limitiert ist, ist er dennoch im Bereich der Suffigierung wirksam. Die Limitierung der eigentlich allgemeiner formulierten Beschränkung NAME ist womöglich ebenso der orthographischen Norm geschuldet, wie die Unterschwelligkeit von FREMD im Modell für determinative Kompositionsgrenzen.

- (80) a. Ingenieurinnen (VDI₀₉), Melancholiker (HAZ₀₉), Renovierungen (WPD₁₁), Verlinkung (WDD₁₁)
b. Kommandeurs (WDD₁₁), Tranzsparenz (RHZ₀₉), Absorption (W), Muskulatur (W), Peripherie (DIV), Stipendiat (WPD₁₁)

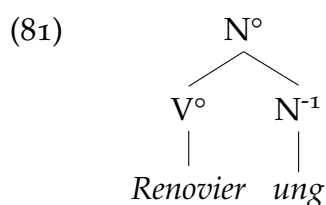
Die Belege mit fremdsprachlichen Teilausdrücken lassen sich dabei aufgliedern in solche mit (teilweise) fremder Basis aber nativem Suffix (80a) und solche, bei denen Basis und Suffix fremd sind (80b). Werden aber unabhängig davon alle zusammengeschrieben.

Für die Modellierungen anhand der oben formulierten Beschränkungen bedeutet das, dass *GRAPH mangels Datenbasis und ZUS-FE mangels Anwendbarkeit nicht

⁵³ Von 91 Belegen für suffigierte Eigennamen finden sich 50 in den beiden Teilkorpora der Online-Lexikoneinträge (31 Belege) und den dazugehörigen Diskussionsbeiträgen (19 Belege).

einzuordnen sind, während NAME und FREMD GPKB unterzuordnen sind. Da eine Schreibung mit Bindestrich genauso wenig zu beobachten ist wie eine Getrenntschreibung, ist davon auszugehen, dass GPKB auf einer Stufe mit ZUS-MORPH steht. NAME und FREMD sind auf einer gemeinsamen Stufe GPKB und ZUS-MORPH untergeordnet, so dass ZUS-LEX als Reparaturprozess über NAME irrelevant wird und in GPKB aufgeht. GPKB und ZUS-LEX verhindern beide das Einfügen von Interpunktionszeichen an wortbildungsmorphologischen Grenzen, GPKB bezieht sich dabei auf alle Grenzen, ZUS-LEX nur auf die in lexikalisierten Belegen. Sie in der Hierarchie auf eine Stufe zu stellen macht also ZUS-LEX obsolet.

N⁰-GROSS ist auf die Suffigierungen nicht anwendbar. Suffigierungen wie *-ung* sind zwar nomenbildend, können aber für sich alleine genommen keine N⁰-Position besetzen, sondern bewegen sich unterhalb dieser Ebene, wie (81) illustriert.



MAJUS blockiert auch bei Suffigierungen die Realisierung von Großbuchstaben abseits der initialen Position von graphematisch selbständigen Einheiten. Da unter den Suffigierungen auch keine Binnenmajuskeln belegt sind, kann allerdings keine valide Einstufung gegenüber ZUS-MORPH und GPKB hinsichtlich der Wirkung speziell an Wortbildungsgrenzen erfolgen, so dass auch MAJUS auf einer Stufe mit ZUS-MORPH und GPKB zu platzieren ist.

6.3.1 Modell für Suffigierungsgrenzen

Für die Suffigierungen ergibt sich auf Basis dieser Diskussion damit das Ranking in (82), das im Tableau (83) auf *Verlinkung* (VDI09) angewandt wird.

(82) MAJUS, ZUS-MORPH, GPKB ≫ NAME, FREMD ≫ GETR-AUSDR ≫ MINUS, *SPAT

(83)

	MAJUS	ZUS-MORPH	GPKB	NAME	FREMD	GETR-AUSDR
a. Verlinkung (VDI09)					*	*
b. Verlink-ung			*!			*
Verlink ung		*!				*
VerlinkUng	*!				*	*

Die wortinterne Schreibvariation an Suffigierungsgrenzen lässt sich damit vergleichsweise einfach modellieren. Da anders als bei determinativen Kompositionsgrenzen keine Varianten abseits der Zusammenschreibung belegt sind, kann auch keine valide Aussage hinsichtlich einer abgestuften Optimalität getroffen werden. Der Datenausschnitt deutet vielmehr darauf hin, dass abseits der Zusammenschreibung, keine anderen Varianten zu erwarten sind.

Dagegen steht – womöglich aber nur im adjektivischen Bereich, also außerhalb des eigentlichen Untersuchungsgegenstandes – die orthographische Norm, die in §41 für Ableitungen von Einzelbuchstaben (*n-te Potenz*) den Bindestrich an der Suffigierungsgrenze obligatorisch nennt. Damit wäre die Beschränkung *GRAPH, die solche Fälle für den hier untersuchten substantivischen Bereich erfassen würde, bei einem größeren Datenausschnitt möglicherweise doch auch für Suffigierungen relevant. Dann stellt sich aber die Frage, ob der Bindestrich bei Ableitungen tatsächlich auf suffigierte Einzelbuchstaben beschränkt ist oder im Sinne von *GRAPH auch andere Derivationsbasen betrifft, die graphematisch markierten Wörtern entsprechen. §41 der Rechtschreibnorm verneint das aber für andere Basen mit den Beispielen *abcllich*, *ÖVPler*, *der 68er*, *ein 32stel* und *100%ig* ausdrücklich. Ein Blick ins Korpus über die kleinen 5.000er-Stichproben und die komplexen Substantive hinaus zu den adjektivischen Belegen offenbart immerhin vereinzelte Beispiele für Suffigierungen mit *-er* von Derivationsbasen, die graphematisch markierten Wörtern entsprechen und mit Bindestrich an der Suffigierungsgrenze realisiert werden (84a). Daneben gibt es aber auch vergleichbare Belege für eine Zusammenschreibung (84b). Eine Getrennschreibung ist nicht belegt.

(84) a. 36-er, 70-300-er (DIV), 90-er, PB-er, 1960-er (WDD11), 2000-er (HAZ09)

- b. 60er (M09), 123er, 129er, 142er, 1950er (BRZ09), 1670er (SPK)

Unter den Beispielen aus dem Gesamtkorpus in (84) ist mit *PB-er* (WDD11) auch zumindest ein substantivischer Beleg, der für eine Relevanz von *GRAPH für substantivische Derivationsgrenzen spricht. Eine differenzierte Betrachtung dieses Phänomens auf einer breiteren Datenbasis, die eine fundierte Einordnung von *GRAPH in das Ranking für Derivationsgrenzen ermöglicht, wäre also wünschenswert.

6.3.2 Übertragbarkeit auf ähnliche Wortbildungen

Da Zirkumfigierungen im gewählten Datenausschnitt nicht belegt sind und implizite Derivationen per se keine wortbildungsmorphologische Grenze aufweisen, bleiben nur die Präfigierungen, um eine Übertragbarkeit des Modells für Suffigierungen auf andere Derivationsmuster zu testen.

Präfigierungen

Bei den wenigen Belegen für Präfigierungen (85) zeigt sich im Vergleich zu den Suffigierungen ein differenzierteres Bild.

- (85) a. Kompromiss (3mal), Kompromisses (PBR), Missstände (DIV), Programmes (VDI09), Ressourcen (SPK, 2mal), Interpreten, Urgroßtante (WDD11), Missverstaendnis (SMS), Unmenschlichkeit (HAZ09)
- b. ANTI-OXIDANT (W, 2mal), Ex-Präsident (M09)

Unter den 15 Belegen für Präfigierungen aus den 5.000er-Stichproben finden sich neben 12 Zusammenschreibungen (85a) auch drei Bindestrichvarianten (85b). Die Bindestrichvarianten gehen dabei auf zwei Lexeme mit fremden Präfixen (*anti-* und *-ex*) zurück. Zusammenschreibungen wie *Kompromiss*, *Programm*, *Ressourcen* oder *Interpreten* sprechen aber gegen einen durchschlagenden Effekt von FREMD.

Dennoch geben die Bindestrichvarianten Anlass, das Modell für Suffigierungen für den gesamten Derivationsbereich leicht anzupassen. Der Bindestrich in drei Fällen spricht dafür, GPKB unterhalb von MAJUS und ZUS-MORPH zu platzieren wie in (86), um damit der Tatsache Rechnung zu tragen, dass an Derivationsgrenzen zwar Bindestriche aber keine Binnenmajuskel und keine Getrenntschreibung belegt sind.

- (86) MAJUS, ZUS-MORPH» GPKB» NAME, FREMD» GETR-AUSDR
 » MINUS,*SPAT

Lexikalisiert sind elf der 15 Präfigierungen, darunter auch die Bindestrichvariante *Ex-Präsident* (Mo9), die zudem wie oben unter 3.2 schon erwähnt, auch noch dank einer Einzelfallregelung als normgerechte Alternative gilt. Wieso gerade diese Präfigierung als Bindestrichvariante lexikalisiert und sogar von der Norm gedeckt ist, kann im Rahmen des Modells nicht erklärt werden. Buchmann (2015: 249), die die Diskussion unter anderem am Beispiel *Ex-Freundin* führt, das in ihrem Untersuchungskorpus deutlich häufiger in der Schreibweise mit Bindestrich als ohne belegt ist (751:69), vermutet eine Bindestrichschreibung aufgrund des unklaren morphologischen Verhaltens von *ex* (gebunden vs. wortfähig) in Kombination mit der Fremdsprachlichkeit. Aufschluss bezüglich dieses hier randständigen Phänomens könnte nur die Untersuchung einer breiteren Datenbasis bieten.

Abseits der Stichproben im Umfang von insgesamt 45.000 Textwörtern finden sich im Gesamtkorpus unter gut drei Millionen Textwörtern durch die gezielte Suche nach Bindestrichen, Getrenntschreibungen, Apostrophierungen oder Binnenmajuskeln noch 51 weitere Belege mit auffälligen Schreibungen an der Affigierungsgrenze. Darunter ist neben 50 Bindestrichschreibungen mit *VDInis* (VDIo9) auch eine Zusammenschreibung trotz graphematisch markiertem erstem Teilausdruck. Es handelt sich dabei um die Bezeichnung für kindlichen Leser der *VDI nachrichten*, *VDInis* hat also Namencharakter.

- (87) a. Ex-ASV (Mo9), Ex-BVB-Stürmer (HMPo9), Ex-SAP Manager (VDIo9), Ex-NHL-Verteidiger (HMPo9), Ex-NRW-Ministerpräsident (VDIo9)
 b. Ex-Asbacher, Ex-Schalk, Ex-Boyzone-Sänger, Ex-Stasi-Offiziere (RHZo9), Ex-Siemens-Chef (VDIo9), Ex-Siemens-Manager (NUZo9), Anti-Spengler, Anti-Antifa (WPD11)
 c. Ex-Star, Ex-Schwiegervater, Ex-Nationalspieler (HMPo9), Co-Autor (SPK (2mal)), Co-Autor, Ex-Staatssekretär (VDIo9), Anti-Imps, Ex-Ehefrau, Co-Regisseur (WPD11), Ex-Ehemann, Co-Abhängige, Ex-Tour-Sieger (Mo9), Ur-Einwohner (WDD11), Ex-Außenminister (3mal), Ex-Nationaltorwart, Ex-Nationaltorhüter (NUZo9), Ex-Europameister, Ex-Stadtdirektor, Ex-Nationalspieler (RHZo9), Ex-Atomaufsichtschef

(PBR)

- d. Re-Release (3mal), Ur-Sehnsucht (WPD₁₁), Re-Emission, Re-Artikulation (WDD₁₁), Co-Abhängigkeit, De-Industrialisierung (Mo9), Re-Industrialisierung (VDIo9 (3mal))
- e. PB-er (WDD₁₁), Taxadien-5-alpha-ol (SPK)

Einen Einblick in die Bindestrichschreibungen bietet (87). Unter den 50 belegten Bindestrichschreibungen sind nur zwei Fälle belegt (87e), die einen Bindestrich zwischen Suffix und Derivationsbasis aufweisen. Alle anderen Bindestrichschreibungen betreffen Präfigierungsgrenzen. Bei *Taxadien-5-alpha-ol* (SPK) handelt es sich um chemische Bezeichnung eines Stoffes, einer Vorstufe zum Krebsmedikament Taxol, die Schreibung mit gleich mehreren Bindestrichen dient damit wohl zur Verdeutlichung der chemischen Struktur, während sich hinter *PB-er* (WDD₁₁) eine Personenbezeichnung verbirgt – aus dem Kontext des Belegs geht hervor, dass er auf Bewohner von Berlin-Prenzlauer Berg referiert. Für *PB-er* kann analog zu der Modellierung für Kompositionsgrenzen ein Einfluss von *GRAPH angenommen werden, wie er sich auch schon bei den oben unter (84) angeführten adjektivischen Belegen angedeutet hat. Abseits beiden genannten Belege sind die Suffixe *-ol* und *-er* aber regelmäßig in Zusammenschreibungen wie *Alkohol* und *Fahrer* belegt.

Auch unter den 47 Präfigierungen mit einem Bindestrich an der Präfigierungsgrenze finden sich fünf Belege (87a), in denen sich der Einsatz des Bindestrichs nach *ex-* durch *GRAPH erklären lässt, wenn man eine Anwendbarkeit über die Kompositionsgrenze in zweigliedrigen Komposita hinaus annimmt. Unter den Präfigierungen finden sich darüber hinaus auch noch acht Belege (87b), deren Bindestrichschreibung wie bei *Ex-Siemens-Chef* (VDIo9) auf den Einfluss von NAME zurückgeführt werden kann, sofern man auch hier einen erweiterten Geltungsbereich annimmt. In dieser Gruppe ist neben dem sehr dominanten *ex-* bei Bindestrichschreibungen von Präfigierungen in 27 von 48 Belegen mit *anti-* in *Anti-Spengler* (WPD₁₁) auch ein weiteres fremdes Präfix mit nachfolgenden Bindestrich belegt.

Abseits des möglichen Geltungsbereiches von Beschränkungen, die im Zuge der Modellierung für determinative Kompositionsgrenzen formuliert wurden, findet sich eine auffällige Häufung von Bindestrichschreibungen bei einer Kombination von Präfix und folgender Personenbezeichnung wie *Ex-Star* (HMPo9). Innerhalb die-

se Musters ist neben dem auch hier dominanten *ex-* eine größere Bandbreite an Präfixen belegt, was beispielsweise *Co-Autor* (SPK), *Anti-Imps* (WPD11) und *Ur-Einwohner* (WDD11) verdeutlichen. Mit 23 Belegen machen diese Fälle fast die Hälfte der Bindestrichschreibungen nach Präfigierungen aus. Die Ausführungen im Bericht des Rechtschreibrates zur Periode von 2011-2016 bestätigen den Verdacht, dass es sich hier um eine systematische Abweichung von der Zusammenschreibung handelt (vgl. Güthert 2016: 10-11). Für *Co-/Ko-* und *Ex-* in Verbindung mit Personenbezeichnungen und nur dort wird explizit auf die Möglichkeit einer Schreibung mit dem Bindestrich in seiner Funktion zur Hervorhebung der einzelnen Bestandteile verwiesen.

Gründe für diese Einschränkung werden dabei nicht erläutert. Möglicherweise spielt die semantische Nähe von Personenbezeichnungen zu Eigennamen eine Rolle.⁵⁴ Die Bandbreite der verwendeten Präfixe in (87c) spricht meiner Meinung nach eher dafür, davon auszugehen, dass die Personenbezeichnung der Auslöser für die Abweichung von der Zusammenschreibung ist und nicht das jeweilige Präfix, da mit Ausnahme von *ur-* alle verwendeten Präfixe außerhalb dieser Kombination in Zusammenschreibung mit der Derivationsbasis belegt sind, wie die komplexen Beispiele unter (88) zeigen.

- (88) Experiment (SPK), Exportauftragsindex (VDI09), Koordinatensystem (DIV),
Oberflächen-Antigen (SPK)

Dass besonders häufig ein Bindestrich nach *ex-* zu finden ist, könnte dementsprechend damit zusammenhängen, dass es sich besonders häufig mit Personenbezeichnungen verbindet. Anders als *co-* und *anti-* und *ur-*, wird es auch selbst, bzw. eine graphematisch identische Zeichenkette als Personenbezeichnung verwendet (*Dein Ex hat sich gemeldet.*).

Mit *Co-Autor* (SPK, DIV) und *Ex-Nationalspieler* (RHZ09, HMP09) sind sogar zwei Lexeme aus dieser Gruppe als Bindestrichschreibung lexikalisiert, wogegen mit *Ureinwohner* nur zur einer der Bindestrichvarianten eine zusammengesriebene

⁵⁴ Unter den Determinativkomposita der Modellierungsgrundlage ist allerdings an einer kleinen Stichprobe nicht zu beobachten, dass Personenbezeichnungen als Teilausdrücke einen ähnlichen Effekt weg von der Zusammenschreibung haben wie Eigennamen. Unter 30 überprüften Komposita mit Personenbezeichnungen, von denen die Hälfte lexikalisiert ist, weisen nur drei eine Bindestrichschreibung auf, die allerdings in allen drei Fällen auf *GRAPH zurückzuführen ist wie in *A-Junioren* (Mo9). Alle anderen werden zusammengesrieben wie *Bürgerbüro* (Mo9) und *Bauerngewinn* (BRZ09).

Entsprechung lexikalisiert ist. Wenn die Bindestrichschreibungen bei der Präfigierung von Personenbezeichnungen als Effekt von NAME gelten können, wäre eine Überprüfung des Verhältnis von NAME und ZUS-LEX angebracht. Aufgrund des Recherchemodus in den virtuellen Korpora, der sich gezielt auf Schreibungen mit Bindestrichen und Apostrophen konzentriert und damit ein verzerrtes Bild liefert, kann aber zum einen nicht ausgeschlossen werden, dass auch für Kombinationen aus Präfix und Personennamen die Zusammenschreibung der Standardfall ist. Zum anderen ist nicht zu vergessen, dass bei fremdsprachlichen Präfixen mit FREMD schon eine Beschränkung einer Zusammenschreibung entgegensteht und bei den Präfixen *anti-* und *co-* auch die Hiäte in Belegen wie *Co-Autor* (SPK) und *Anti-Antifa* (WPD₁₁) einen Effekt haben könnten, wie er sich auch bei den Bindestrichvarianten abseits des diskutierten Musters unter (87d) andeutet.

Die elf verbleibenden Bindestrichschreibungen nach Präfixen weisen heterogene Derivationsbasen auf (87d). Das Präfix *ex-* ist dabei tatsächlich nicht vertreten, stattdessen aber *de-*, *re-*, *co-* und *ur-*. Abgesehen vom bis auf den Bindestrich unauffälligen *Ur-Sehnsucht* und *Re-Release* (WPD₁₁) wo der Bindestrich zwischen zwei identischen graphematischen Silben steht, zeigen alle Belege dieser Gruppe aber ein Aufeinandertreffen zweier Vokale an der Wortbildungsgrenze wie in *De-Industrialisierung* (Mog). Für markierte Buchstabenfolgen wie diese wurde zwar ein Effekt hin zur Grenzmarkierung erwartet, konnte aber auf Basis der untersuchten Belege nicht bestätigt werden.

Aufgrund der Ergebnisse aus den 5.000er-Stichproben mit drei Bindestrichvarianten auf 15 Präfigierungen ist trotz der im Gesamtkorpus belegten Varianten mit Bindestrich (und Binnenmajuskel) davon auszugehen, dass die Bindestrichvariante auch bei Präfigierungen deutlich seltener ist im Vergleich zur Zusammenschreibung. Darüber hinaus lässt sich die Mehrheit der Bindestrichschreibungen durch einen potentiellen Einfluss von *GRAPH und NAME bzw. die dem ähnliche Kombination von Präfix und Personenbezeichnung erklären. Dazu kommt, dass abgesehen von Einzelfällen Bindestrichschreibungen nur bei fremden Präfixen belegt sind (also eine mögliche Auswirkung von FREMD) und sich in den Belegen eine Häufung von Hiäten an der Wortbildungsgrenze zeigt, für die bei der Modellierung von Kompositionsgrenzen aber mangels breiter Datenbasis kein Effekt auf die Ausprägung von Schreibvarianten bestätigt werden konnte. Zwischen den Präfigierungen und


Suffigierungen scheint eine deutliche Assymetrie dahingehend zu bestehen, dass die Zusammenschreibung bei Suffigierungen noch dominanter ist. Ob dieser Unterschied im Kontrast Präfix und Suffix begründet ist oder womöglich auf den Anteil markierter Derivationsbasen von Präfigierungen und Suffigierungen zurückzuführen ist, kann hier nicht abschließend geklärt werden.

6.3.3 Modell für Derivationsgrenzen

(89) MAJUS, ZUS-MORPH» GPKB» NAME, FREMD» GETR-AUSDR
 » MINUS,*SPAT

Das Ranking in (89) spiegelt also entsprechend der drei belegten Bindestrichschreibungen von Präfigierungen wieder, dass ein Bindestrich an Derivationsgrenzen eher akzeptiert ist, als eine Getrennschreibung oder Binnenmajuskel. Der optimale Kandidat bleibt aber auch nach der Berücksichtigung der Präfigierungen die Zusammenschreibung. Auch wenn sich abseits der Modellierungsgrundlage Belege finden, die für einen Einfluss von *GRAPH auch an Derivationsgrenzen sprechen, muss die Beschränkung für diesen Datenausschnitt noch ausgeklammert werden, weil so keine Positionsbestimmung in Relation zu den anderen Beschränkungen möglich ist.

(90)

	MAJUS	ZUS-MORPH	GPKB	NAME	FREMD	GETR-AUSDR
a.  Urgroßtante (WDD11)					*	*
b. Ur-Großtante			*!			*
c. Ur Großtante		*!				
UrGroßtante	*!				*	*

Das Modell für Derivationsgrenzen basiert dementsprechend auf 805 Derivationsgrenzen (790 Suffigierungs- und 15 Präfigierungsgrenzen) und bildet 802 der 805 Belege als optimale Kandidaten ab, womit aufgrund der hohen Zusammenschreibungsquote ein Deckungsgrad von nahezu 100% erreicht wird.

6.4 Zusammenführung der Modelle

Da in den vorausgehenden Teilkapiteln gezeigt werden konnte, dass sich mit einem Teil der Beschränkungen, die für Kompositionsgrenzen erarbeitet wurden, auch die Schreibvariation an Derivationsgrenzen modelliert werden kann, liegt es nahe im nächsten Schritt zu prüfen, ob sich beide Grenzkategorien in einem Modell erfassen lassen.

- (91) MAJUS, ZUS-MORPH» GPKB» NAME, FREMD» GETR-AUSDR
» MINUS,*SPAT
- (92) N°-GROSS» MAJUS» *GRAPH» ZUS-MORPH» ZUS-LEX» NAME » GPKB,
ZUS-FE» FREMD» GETR-AUSDR» MINUS,*SPAT

Auf den ersten Blick stellt die Position von GPKB im Vergleich der Modelle für determinative Kompositionsgrenzen (92) und Derivationsgrenzen (91) ein Problem dar. Während GPKB für Kompositionsgrenzen NAME untergeordnet ist, blockiert die Beschränkung GPKB im Ranking (91) alle potentiell auf NAME basierenden Varianten.

Bei näherer Betrachtung wird dieses Problem allerdings innerhalb des zugrundeliegenden Datenausschnittes gut durch ZUS-LEX gelöst. Es sind mit *Parkinsonoid*⁵⁵ (WDD11) und *Amerikaner* (HAZ09) überhaupt nur zwei Lexeme für Affigierungen von Eigennamen belegt, und für alle vier Instanzen von *Amerikaner* sorgt ZUS-LEX dafür, dass eine Zusammenschreibung trotz Verstoß gegen NAME der optimale Kandidat ist (vgl. Tableau (93)). Womit nur *Parkinsonoid* (WDD11) für das Ranking (92) problematisch ist, da aufgrund des Verstoßes gegen NAME eigentlich eine Bindestrichvariante der optimale Kandidat ist (vgl. Tableau (94)). Hier offenbart sich also, wie wichtig ZUS-LEX trotz der großen Schnittmenge mit GPKB für die Modellierung der wortinternen Schreibvariation ist. Auch ganz ohne eine Veränderung der Beschränkungshierarchie sorgt ZUS-LEX aufgrund des höheren Lexikalisierungsgrades von Derivationen dafür, dass hier seltener Varianten abseits der Zusammenschrei-

⁵⁵ Das Institut für Deutsche Sprache (grammis) führt *oid* als adjektivbildendes Affix auf Basis von Nomina und Eigennamen. Beim vorliegenden Beleg *Parkinsonoid* ist aber eindeutig am Kontext zu erkennen, dass es sich um eine substantivierte Instanz handelt: (*dominierendes Parkinsonoid* + *Demenz*) (WDD11).

zung zugelassen werden als an Kompositionsgrenzen. Klar ist aber auch, dass diese Ordnung aufgrund des kleinen Datenausschnittes der zugrunde liegt noch auf wackeligen Beinen steht. Mit einer breiteren Datenbasis könnte sich beispielsweise herausstellen, dass ZUS-LEX nicht ausreicht um die Wirkung von NAME bei Derivaten einzugrenzen. Die Datenlage für die Positionierung von ZUS-LEX über NAME ist mit nur je zwei ausschlaggebenden Lexemen für Derivationsgrenzen und Kompositionsgrenzen noch unangenehm dünn und bedarf weiterer Überprüfung.

(93)

	N°-GROSS	MAJUS	*GRAPH	ZUS-MORPH	ZUS-LEX	NAME	GPKB	ZUS-FE	FREMD	GETR-AUSDR
a. ☞ Amerikaner (HAZ09)						*				*
b. Amerika-ner					*!		*			*

(94)

	N°-GROSS	MAJUS	*GRAPH	ZUS-MORPH	ZUS-LEX	NAME	GPKB	ZUS-FE	FREMD	GETR-AUSDR
a. ☞ Parkinson-oid							*			*
b. Parkinsonoid (WDD11)						*!				*

Da in der Stichprobe keine Affigierungen markierter graphematischer Wörter belegt sind, die dem Ranking gemäß einen Bindestrich als optimale Kandidaten aufweisen müssten, stellt auch die Positionierung von *GRAPH über ZUS-MORPH, ZUS-LEX und GPKB zunächst kein Problem dar. Sie bietet sogar die Möglichkeit systematische Bindestrichschreibungen wie *Ex-ASV* (Mo9), die sich bei Präfigierungen außerhalb der Modellierungsgrundlage gezeigt haben zu erklären. Problematisch hinsichtlich der abgestuften Optimalität kann aber durchaus die Positionierung über ZUS-MORPH werden. Dadurch wird nämlich nicht nur eine Bindestrichschreibung zum optimalen Kandidaten bei der Affigierung eines graphematisch markierten Wortes wie *der 11-er*, sondern auch die Getrennschreibung (*der 11 er*) müsste noch besser bewertet werden als die Zusammenschreibung (*der 11er*). Dementsprechend könnte abseits dieser Stichprobe ein Positionswechsel von *GRAPH und ZUS-MORPH zu besseren Ergebnissen führen, wenn keine oder kaum Getrennschreibungen aufgrund von *GRAPH belegt sind.

Auf der vorliegenden Datenbasis bildet die Beschränkungshierarchie für determi-

native Kompositionsgrenzen aber tatsächlich auch 801 der 805 Belege für Suffigierungen und Präfigierungen als optimale Kandidaten ab und modelliert damit die Schreibvarianten ebensogut wie die finale Beschränkungshierarchie für die Derivationsgrenzen (91), die nur einen einzelnen Beleg mehr als optimale Variante erfasst. Hier zwei verschiedene Modelle aufgrund eines Einzelbeleges anzunehmen, ist empirisch unbegründet.

Das anhand der determinativen Kompositionsgrenzen ermittelte Modell in (92) erreicht dementsprechend, die Affigierungen mit eingenommen, einen Deckungsgrad von 97,5%, da 2.468 der 2.530 bisher diskutierten Belege als optimale Kandidaten abgebildet werden.

Durch die Platzierung in der Beschränkungsordnung werden einige der Beschränkungen unwirksam wie z. B. FREMD, in ihrer Wirkung eingeschränkt wie beispielsweise NAME, oder obsolet wie ZUS-FE. Nichtsdestotrotz wird am vollen Modell festgehalten, da es für all diese Beschränkungen empirische Evidenzen gibt, und aufgrund ihres universellen Charakters davon auszugehen ist, dass es andere Kontexte gibt, in denen sie sich als wirksam erweisen.

6.5 Textformspezifik

Mit dem anhand der zweigliedrigen Determinativkomposita und Derivationen entwickelten Modell unter (95) als Rüstzeug soll im nächsten Schritt, wie bereits angekündigt, überprüft werden, wie stark der bisher unberücksichtigte Einfluss der Textform auf die Ausprägung von Schreibvarianten an wortbildungsmorphologischen Grenzen ist.

(95) N°-GROSS» MAJUS» *GRAPH» ZUS-MORPH» ZUS-LEX» NAME » GPKB,
ZUS-FE» FREMD» GETR-AUSDR» MINUS,*SPAT

Dazu wird das Modell für Kompositionsgrenzen in substantivischen, zweigliedrigen Determinativkomposita und Derivationsgrenzen der substantivischen Wortbildung, auf die entsprechenden Belege der neun verschiedenen Textform angewandt, um den individuellen Deckungsgrad des Modells für die einzelnen Textformen zu ermitteln. Dabei wird sich zeigen, dass das Modell in (95) ohne Veränderung der

Beschränkungsordnung in allen Textformen mindestens knapp 94% der Belege als optimale Kandidaten erfasst. Allein für das SMS-Teilkorpus kann eine Verbesserung der Deckungsquote um einige Prozentpunkte durch einen Positionswechsel von ZUS-MORP und GPKB erreicht werden. Anders als erwartet sind also keine weiteren pragmatischen oder textformspezifischen Beschränkungen nötig, um das Modell einzelnen Textformen anzupassen.

6.5.1 Werbeanzeigen

Im Fall des Werbeanzeigen-Teilkorpus sind in der 5.000er-Stichprobe 535 Belege für wortbildungsmorphologisch komplexe Substantive erfasst worden. 421 dieser Wortformen sind mit ihrer terminalen Wortbildungsgrenze in die Modellierung des Modells zur wortinternen Schreibvariation mit eingeflossen. Darunter befinden sich neben 322 zweigliedrigen determinativen Komposita auch 99 Derivationen.

Tabelle 6.7.: Modellierungsgrundlage: Werbeanzeigen
Optimalität

	optimal	suboptimal
Determinativkomposita	307 (292 zusammen)	15 (2 zusammen)
Affigierungen	97 (97 zusammen)	2 (0 zusammen)
Summe	404 (389 zusammen)	17 (2 zusammen)

Von diesen 421 Wortformen erfasst das Modell, wie in Tabelle 6.7 zu sehen, 404 als optimale Kandidaten (darunter 307 Komposita und 97 Derivationen) und nur 17 als suboptimal (darunter 15 Komposita und zwei Derivationen). Das entspricht einer Passungsquote von 96%, die damit nur etwas unterhalb der oben ermittelten Quote über alle Textformen hinweg von 97,5% liegt.

Während fast alle 17 suboptimalen Kandidaten an mindestens einer Grenze von der Zusammenschreibung abweichen, findet sich unter den optimalen Kandidaten nur ein geringer Anteil abweichender Varianten. Die optimalen Kandidaten, die das Ergebnis einer Derivation sind, werden alle zusammengeschrieben, von den 307 optimalen Kompositionskandidaten weichen immerhin 15 von der Zusammenschreibung ab. Neben *Strass-Steinchen*, dessen Bindestrichschreibung durch NAME ausgelöst wird, handelt es sich in allen anderen 14 Fällen dabei um eine Bindestrichschrei-

bung aufgrund von *GRAPH.

- (96) 360°-Kamera, CO₂-Ausstoß, CO₂-Emmission (3mal), CO₂-Emissionen(7mal), BEST-Tarif, SMART-Tarif

Die 17 Belege für suboptimale Kandidaten unter (97) lassen sich dabei grob in zwei Gruppen einteilen. Die Belege unter (97a) weisen zwar eine Bindestrichschreibung auf, die vom Modell nicht gedeckt ist, sind aber sonst unauffällig und damit in ihrer Optimalität nur eine Stufe unter den Zusammenschreibungen anzusiedeln. Auf derselben Stufe bewegt sich *Mercedesstraße* in (97b) als nicht lexikalisierte Zusammenschreibung trotz NAME. Die übrigen Belege sind dagegen noch weniger optimal, entweder aufgrund der Binnenmajuskel (97c) oder durch Systemverstöße hinsichtlich der Groß- und Kleinschreibung wie *PFLEGE-SYSTEM* und *direkt-granulat* in (97d).

- (97) a. Bussiness-Klasse, Apotheken-Magnesium, Funktionalitäts-Symbole, Qualitäts-Garantie, Bio-Handel, Top-Konditionen, Park-Assistenten
 b. Mercedesstraße (2mal)
 c. FeuchtCreme, PrivatKredit
 d. ZELL-RENAISSANCE, ZELL-ERNEUERUNG, PFLEGE-SYSTEM, ANTI-OXIDANT (2mal), direkt-granulat

Obwohl wir uns in der Textform der Werbeanzeigen bewegen, für die wie in 4.2.1 geschildert eine größere Bandbreite der wortinternen Varianten erwartet wurde, ist für die bisher betrachteten zweigliedrigen Determinativkomposita und die Affigierungen keine Getrennschreibung belegt. Mangels belegter Getrennschreibungen gibt es also für das Teilkorpus der Werbeanzeigen auch keine textformspezifische Evidenz für die Positionierung von *GRAPH über ZUS-MORPH. Da diese Positionierung textformübergreifend aber evidenzbasiert ist und eine Veränderung der Beschränkungsordnung den Deckungsgrad für die Werbeanzeigen nicht verbessert, wird die Hierarchie von *GRAPH über ZUS-MORPH beibehalten.

Auffällig ist außerdem, dass alle suboptimalen Kandidaten unter (97) fremdsprachliche Anteile aufweisen. Dementsprechend kann man sich die Frage stellen, ob auch FREMD innerhalb der Hierarchie verschoben werden muss, um eine optimale Passung für die Textform der Werbeanzeigen zu erzielen. Ein Blick in die Belege

zeigt allerdings, dass die Passgenauigkeit darunter vielmehr leiden würde, statt sich zu verbessern. Allein unter den Komposita gibt es 39 Belege mit fremdem Erstglied und nicht-fremdem Zweitglied. Auch ohne den Einfluss einer Lexikalisierung sind darunter nur zwei Bindestrichschreibungen im Vergleich zu 37 Zusammenschreibungen, die fast alle durch eine Positionierung von FREMD über GPKB ihren Status als optimale Kandidaten einbüßen würden, da ZUS-FE davon nur sieben Belege betrifft und damit auch nicht ausreicht, um FREMD entsprechend einzuschränken.

Auch die Position von NAME bleibt unverändert gegenüber dem textformübergreifenden Modell, da zumindest eines von beiden Lexemen mit Relevanz für NAME als optimaler Kandidaten erfasst wird. *Strass-Steinchen* als nicht lexikalisiertes Kompositum mit einem Eigennamen als Erstglied weist eine modellkonforme Bindestrichschreibung auf, während *Mercedesstraße* trotz Eigennamenanteil als nicht lexikalisierte Verbindung zusammengeschrieben wird.

Dementsprechend eignet sich für die Textform der Werbeanzeigen das unveränderte Modell aus (95). Im Tableau unter 6.5.10 auf Seite 245 wird das durch Korpusbelege unterschiedlicher Optimalitätsgrade veranschaulicht wird. Dabei werden erstmalig die Verstöße an allen wortbildungsmorphologischen Grenzen angezeigt, statt wie bisher nur die terminale Grenze zu berücksichtigen. Außerdem werden auch die unterschweligen Verstöße gegen MINUS und *SPAT wieder dokumentiert.

Weil im Tableau 6.5.10 kein Ranking von Kandidaten, also mehrerer potentieller graphischer Realisierungen ein und desselben sprachlichen Zeichens, abgebildet ist, sondern lediglich Korpusbelege als Realisierung verschiedener sprachlicher Zeichen zusammengestellt wurden, um unterschiedliche Optimalitätsgrade zu illustrieren, werden auch keine fatalen Verstöße gekennzeichnet. Das Tableau dient damit lediglich dazu, zu visualisieren, dass die Beispiele in der ersten Spalte unter (a) jeweils optimale Kandidaten sind, wenngleich sie aufgrund ihrer Bestandteile gegen unterschiedliche Beschränkungen verstoßen, während das Beispiel unter (b) ein suboptimaler Kandidat ist, der im Vergleich zum optimalen Kandidaten des entsprechenden sprachlichen Zeichens (*Biohandel*) um eins abgestuft ist. Bei *FeuchtCreme* unter (c) liegen schon zwei Abstufungen vor. Besser im Ranking abschneiden würde *Feucht-Creme*, der optimale Kandidat wäre *Feuchtcreme*. Die optimalen Kandidaten zu den Beispielen unter (d) und (e), die im Ranking noch weiter abgestuft sind, wären *Antioxidant* und *Direktgranulat*.

In den Werbeanzeigen finden sich also innerhalb der Modellierungsgrundlage Belege für fünf Optimalitätsstufen nach dem erarbeiteten Modell.

Es bleibt festzuhalten, dass die Varianz der wortinternen Schreibung an wortbildungsmorphologischen Grenzen auch in den Belegen aus dem Werbeanzeigen-Korpus systematisch zu erfassen ist. Der Deckungsgrad des Modells liegt mit 96% nur knapp unter dem für die gesamte Stichprobe und selbst die suboptimalen Kandidaten weisen durch fremde Bestandteile in allen Belegen eine Regelmäßigkeit auf.

Trotz der in 4.2.1 formulierten Erwartung, dass in Werbeanzeigen gezielt Systemverstöße genutzt werden, um die Aufmerksamkeit der Rezipienten zu erregen, ist hier zwar anhand von Einzelbelegen eine größere Bandbreite der Varianten aber keine überdurchschnittliche Häufung suboptimaler Kandidaten zu beobachten.

6.5.2 Mail

Aus dem Mail-Teilkorpus sind in der 5.000er-Stichprobe 141 Belege für wortbildungsmorphologisch komplexe Substantive erfasst worden. 119 dieser Wortformen dienten mit ihrer terminalen Wortbildungsgrenze der Modellierung zur wortinternen Schreibvariation. Unter diesen 119 Belegen finden sich neben 86 zweigliedrigen, determinativen Komposita auch 33 Derivationen.

Tabelle 6.8.: Modellierungsgrundlage: Mail

	Optimalität	
	optimal	suboptimal
Determinativkomposita	83 (79 zusammen)	3 (0 zusammen)
Affigierungen	33 (33 zusammen)	0 (0 zusammen)
Summe	116 (112 zusammen)	3 (0 zusammen)

Von diesen 119 Wortformen werden, wie Tabelle 6.8 zeigt, 116 vom Modell als optimale und nur 3 als suboptimale Kandidaten abgebildet. Alle 33 Affigierungen werden zusammengesrieben und sind optimale Kandidaten. Abweichungen von der Zusammenschreibung sind nur bei den Determinativkomposita belegt. In vier Fällen handelt es sich dabei um optimale Kandidaten aufgrund von *GRAPH (98), während drei weitere eine suboptimale Variante an der Kompositionsgrenze aufweisen (99).

(98) Kfz-Versicherung, Kfz-Versicherung, 2Hand-Laden, SecondHand-Laden

Damit ergibt sich auf der Wortebene für die 119 Belege, die in die Modellierung mit eingeflossen sind, eine Deckungsquote des Modells von 97,5% die damit genau mit der textformübergreifenden Quote übereinstimmt.

Wie zu Beispiel (73a) schon erläutert, wurde *2Hand-Laden* in Analogie zu *SecondHand-Laden* als Phrasenkompositum klassifiziert. Die Zusammenschreibung trotz Ziffer, *2Hand*, und die Binnenmajuskel hier und in *SecondHand* sind also Phänomene an einer phrasalen Grenze, die nicht durch das Modell abgedeckt werden.⁵⁶ Dennoch sind diese Varianten ein erster Hinweis darauf, dass in der interaktionsorientierten Kommunikation per Mail im privaten Bereich, wie vermutet, mit einer größeren Bandbreite der Varianten gerechnet werden kann (vgl. 4.2.2).

(99) BWL Studiums, Browser-Fenster, LautsprecherBox

Von den drei suboptimalen Kandidaten unter (99) handelt es sich bei der Getrenntschreibung eines für *GRAPH relevanten Beleges (*BWL Studiums*) und eines für FREMD relevanten Beleges (*Browser-Fenster*) um Kandidaten auf der zweiten Optimalitätsstufe. Lediglich die Binnemajuskelschreibung *LautsprecherBox* ist deutlicher abgestuft. Dementsprechend ist eine Anpassung des Modells auf die Belege des Mail-Korpus nicht nötig, da die Passung in der bestehenden Form schon optimal ist.

Damit ist *LautsprecherBox* mit seiner Binnenmajuskel der einzige Beleg für eine dritte Optimalitätsstufe (vgl. Tableau 6.5.10, Seite 246), der damit auf einen größeren Variantenreichtum im Mail-Korpus hindeutet.

6.5.3 SMS

In der 5.000er-Stichprobe des SMS-Teilkorpus finden sich 76 Belege für wortbildungsmorphologisch komplexe Substantive. 63 dieser Wortformen dienen mit ihrer terminalen Wortbildungsgrenze der Modellierung zur wortinternen Schreibvariation. Unter diesen 63 Belegen sind 49 Belege für zweigliedrige, determinative Komposita und 14 Belege für Affigierungen.

Neben 59 optimalen Kandidaten finden sich unter diesen Belegen nur 4 subopti-

⁵⁶ Bei einer Einstufung von *2Hand-Laden* als dreigliedriges Determinativkompositum wäre die Zusammenschreibung *2Hand* aber ein klarer Verstoß gegen *GRAPH.

Tabelle 6.9.: Modellierungsgrundlage: SMS

	Optimalität	
	optimal	suboptimal
Determinativkomposita	45 (45 zusammen)	4 (0 zusammen)
Affigierungen	14 (14 zusammen)	0 (0 zusammen)
Summe	59 (59 zusammen)	4 (0 zusammen)

male Schreibungen, was einer Deckungsquote des Modells von 93,7% entspricht, die etwas unterhalb der Quote von 97,5% über alle Textformen hinweg liegt (vgl. 6.9).

Tabelle 6.9 veranschaulicht außerdem, dass die 59 Belege, bei denen es sich um optimale Kandidaten nach dem Modell handelt, allesamt zusammengeschrieben werden. Es ist also keine einzige Bindestrichschreibung oder sonstige von der Zusammenschreibung abweichende Variante als optimaler Kandidat belegt. Bei den vier suboptimalen Kandidaten 6.9 dominiert die Getrennschreibung. Auf Basis der geringen Belegzahl aus dem SMS-Korpus insgesamt und der vier nicht zusammengeschriebenen Belege kann es gleich zwei Gründe für das Übergewicht der Getrennschreibungen im Vergleich zur Bindestrichschreibung geben. Entweder diese Verteilung ist zufällig oder aber eine systematische Abweichung dahingehend, dass in der Kommunikationsform SMS die Getrennschreibung der Bindestrichschreibung vorgezogen wird.

(100) Abi-Prüfung, USB Verlängerung, linguisten mist, acht Uhr Vorstellung

Die Bindestriche in *Abi-Prüfung* und die Getrennschreibung von *linguisten mist* können auf FREMD zurückgeführt werden, sind aber aufgrund der Überordnung von GPKB und ZUS-MORPH suboptimale Kandidaten. Die Abweichungen von der Zusammenschreibung bei *USB Verlängerung* und *acht Uhr Vorstellung* beruhen auf *GRAPH, sind aber nach dem Modell suboptimal, weil dort der Verstoß gegen ZUS-MORPH durch die Getrennschreibung schwerer wiegt als der Verstoß gegen GPKB durch eine Bindestrichschreibung (*USB-Verlängerung*), die modellkonform wäre. Auch im Fall von *acht Uhr Vorstellung* wäre eine Bindestrichvariante (*acht Uhr-Vorstellung*) der bessere Kandidat. Der initiale Verstoß gegen N°-GROSS spielt dabei hier aufgrund der Konzentration auf das Wortinnere keine Rolle, bei einer Berück-

sichtigung aller Positionen wäre aber *Acht Uhr-Vorstellung* der bessere Kandidat.

Die Getrennschreibung bei den beiden Teilausdrücken, die graphisch markierten Wörtern entsprechen, legt also eine Anpassung der Beschränkungshierarchie im Vergleich zum allgemeinen, textformübergreifenden Modell durch einen Positionswechsel von ZUS-MORPH und GPKB wie in (101) nahe.

(101) N°-GROSS » MAJUS » *GRAPH » GPKB » ZUS-LEX » NAME » ZUS-MORPH, ZUS-FE » FREMD » GETR-AUSDR » MINUS, *SPAT

In Folge dieser Anpassung ist ZUS-FE als Teilmenge von ZUS-MORPH in direkter Nachbarschaft eigentlich obsolet. Da die Wirksamkeit von ZUS-FE als Beschränkung aber grundsätzlich belegt ist, bleibt sie hier dennoch Teil des Modells, steht aber mit ZUS-MORPH auf einer gemeinsamen Stufe. Da im SMS-Teilkorpus leider keine für NAME relevanten Kandidaten belegt sind, bleibt die Position dieser Beschränkung unverändert, bedürfte aber eigentlich einer Überprüfung auf breiterer Datenbasis.

Wie Tableau 6.5.10 auf Seite 246 verdeutlicht, werden durch diese Anpassung die Getrennschreibungen *USB Verlängerung* und *acht Uhr Vorstellung* zu optimalen Kandidaten hinsichtlich der Schreibung an den wortbildungsmorphologischen Grenzen, die Bindestrichschreibung *Abi-Prüfung* wird dagegen suboptimal. *linguisten mist* bleibt durch den Verstoß gegen N°-GROSS an der Wortbildungsgrenze suboptimal und besetzt durch die dritte, weiter abgestufte Optimalitätskategorie.

Dass Verstöße gegen die Substantivgroßschreibung keine Seltenheit im SMS-Korpus sind, zeigen auch *versicherung* und *acht Uhr Vorstellung* (Vgl. außerdem dazu Schnitzer (2012: 149-158)). Da N°-GROSS dennoch wortinitial bei 52 von 76 Belegen für wortbildungsmorphologisch komplexe Substantive und auch bei der überwiegenden Zahl der graphematisch selbständigen Teilausdrücke eingehalten wird, bleibt die Positionierung von N°-GROSS unverändert.

Für das angepasste Ranking in (101) ergibt sich durch die zwei zusätzlichen, optimalen Kandidaten eine Verbesserung der Deckungsquote auf 96,8%.

Dass im SMS-Korpus bei einer Abweichung von der Zusammenschreibung die Getrennschreibung einer Schreibung mit Bindestrich vorgezogen wird, hat wohl schreibökonomische Gründe. So werden die unmarkierten Bestandteile von *acht Uhr Vorstellung* bei der Eingabe einer SMS von einem entsprechenden Programm wahrscheinlich erkannt und zur Übernahme inklusive automatisierter Großschreibung

von *Uhr* und *Vorstellung* angeboten, bevor das Gesamtwort überhaupt fertig getippt ist. Die Leerzeichen zu entfernen oder durch Bindestriche zu ersetzen,⁵⁷ käme dann einem Überarbeitungsprozess gleich, der in interaktionsorientierten Kommunikationsformen keine hohe Priorität hat. Es gibt also einen gewissen Automatismus, der Leerzeichen in komplexen Wörtern (insbesondere zwischen Kompositionsgliedern) begünstigt, während Bindestrichschreibungen einen Mehraufwand bedeuten.

Trotz ähnlicher Kommunikationssituation zeigt sich dieser Effekt im Mail-Korpus nicht, was möglicherweise darin begründet ist, dass die Mails eher noch auf einer analogen Tastatur und ohne Worterkennungsprogramm verfasst worden sind, so dass sich keine Vorteile durch eine Getrennschreibung ergeben. Eine valide Überprüfung dieser Vermutung wäre nur möglich, wenn für jeden Beleg nachvollziehbar wäre, ob und welche Worterkennungs- und Korrekturprogramme genutzt wurden. Da auf dem heutigen Stand der Technik Mails, SMS und andere Nachrichtendienste allesamt auf den digitalen Tastaturen von Smartphones und Tablets verfasst werden können, wäre zudem zu überprüfen, ob sich Unterschiede ergeben in Abhängigkeit vom verwendeten Endgerät, bzw. der Art der Tastatur.

6.5.4 Zeitung

In der kleinsten Stichprobe im Umfang von 5.000 Textwörtern des Zeitungs-Korpus sind 427 Belege für wortbildungsmorphologisch komplexe Substantive erfasst worden. 338 davon sind in die Modellierung der Beschränkungshierarchie eingeflossen, darunter 245 Komposita und 93 Affigierungen.

Tabelle 6.10.: Modellierungsgrundlage: Zeitung

	Optimalität	
	optimal	suboptimal
Determinativkomposita	232 (213 zusammen)	13 (1 zusammen)
Affigierungen	92 (92 zusammen)	1 (0 zusammen)
Summe	324 (305 zusammen)	14 (1 zusammen)

⁵⁷ Bei Nutzung eines Smartphone kommt noch dazu, dass der Bindestrich als Zeichen häufig erst nach einem Tastaturwechsel zugänglich ist. Die verwendeten Endgeräte sind allerdings für das 2011 erfasste SMS-Korpus nicht dokumentiert.

Die Zusammenfassung in Tabelle 6.10 zeigt, dass 324 dieser Belege als optimale und 14 weitere als suboptimale Kandidaten im Modell erfasst werden.

Unter den optimalen Kandidaten finden sich 19 Abweichungen von der Zusammenschreibung aufgrund von *GRAPH (102a), NAME (102b) oder einer Kombination aus beiden durch die Zweigliedrigkeit der Personennamen in (102c). In allen Fällen handelt es sich dabei um modellkonforme Bindestrichschreibungen von Determinativkomposita, die sich über alle sechs Regionalzeitungen verteilen.

- (102)
- a. A-Junioren, D-Junioren (M09), CO₂-Ausstoß (2mal), US-Quartett (HMP09), SPD-Zeitung, „Spiegel“-Reporter (HAZ09), ETA-Ideologe, ETA-Mitbegründer (RHZ09), SED-Diktatur, 2:2-Schützen, 2,5:1,5-Führung (BRZ09)
 - b. Corsa-Klasse (HMP09), Luther-Schau (M09), Rapolder-Elf (RHZ09), Hertie-Stiftung (BRZ09), Afghanistan-Konferenz (HAZ09)
 - c. Günther-Brand-Halle (NUZ09), Willy-Brandt-Platz (BRZ09)

Bei 14 suboptimalen Kandidaten ergibt sich für das Zeitungs-Korpus eine Deckungsquote von 95,9%, die damit etwas niedriger als die textformübergreifende von 97,5% ist. Unter diesen suboptimalen Kandidaten (103) finden sich neben 12 Bindestrichschreibungen auch eine Getrenntschreibung und eine Zusammenschreibung trotz NAME, ohne dass *Brundtlandbüro* lexikalisiert ist und dementsprechend eine Zusammenschreibung auf der Basis von ZUS-LEX zu erwarten wäre. *Hoffmann-von-Fallersleben Museum* entspricht in der Struktur den Belegen unter (102c), ist aber durch das Leerzeichen anstelle des Bindestrichs kein optimaler Kandidat. Neben 13 Komposita in suboptimaler Schreibung ist auch die Bindestrichschreibung der Präfigierung (*Ex-Präsident*) belegt.

- (103)
- a. Mini-Teens, Murmel-Kuhle, Plakat-Ausstellung, Protokoll-Abteilung (BRZ09), Wind-Stadion, Einzel-Karten, Serien-Tickets, Ticket-Verkauf, VERBRAUCHER-TIPP, Senioren-Initiativen (HMP09), Abstellungs-Theater (NUZ09)
 - b. Hoffmann-von-Fallersleben Museum (BRZ09)
 - c. Brundtlandbüro (M09)⁵⁸

⁵⁸ Da Namen, insbesondere Personennamen wie *Brundtland*, einer eigenen Grammatik folgen, wird

d. Ex-Präsident (Mo9)

Die übrigen Bindestrichschreibungen unter (103a) sind wie *Ex-Präsident* zumindest teilweise relevant für FREMD, aber nicht hinreichend für eine höhere Einstufung dieser Beschränkung. Daneben gibt es aber auch hinsichtlich ihrer Bestandteile gänzlich unauffällige Belege wie *Murmel-Kuhle* oder *Einzel-Karten*. Dabei fällt auf, dass sich die Bindestrichschreibungen, die nicht modellkonform auf der Basis von *GRAPH oder NAME sind, auf die beiden Quellen *Hamburger Morgenpost* (HMP09) und *Braunschweiger Zeitung* (BRZO9) konzentrieren, obwohl die Gewichtung aller Quellen in dieser Stichprobe ausgewogen ist. Aufgrund der insgesamt geringen Anzahl von Belegen kann dies aber nur als erster Hinweis darauf gewertet werden, dass womöglich deutlichere Unterschiede zwischen den einzelnen Regionalzeitungen bestehen hinsichtlich der Neigung zu Bindestrichschreibung abseits dessen, was graphematisches System (und orthographische Norm) erwarten lassen.

Im Vergleich zu den bisher besprochenen Textformen (Werbeanzeigen, Mail, SMS) ist außerdem bemerkenswert, dass alle suboptimalen Kandidaten unter (103) durch die jeweilige Schreibung an den wortbildungsmorphologischen Grenzen jeweils nur eine Stufe unter dem optimalen Kandidaten stehen (vgl. dazu Tableau 6.5.10, Seite 247). Die Bandbreite der Variation ist also auf Zusammenschreibung, Bindestrichschreibung und Getrenntschreibung beschränkt. Optimal wäre bei der Getrenntschreibung (*Hoffmann-von-Fallersleben Museum*) eine Schreibung mit Bindestrich (*Hoffmann-von-Fallersleben-Museum*), für *Brundtlandbüro* eigentlich die Bindestrichvariante (*Brundtland-Büro*) und für die Bindestrichvarianten unter (103a) eigentlich die Zusammenschreibung.

Die unterschiedliche Optimalitätsstufe von *Amerikaner* und *Brundtlandbüro* nach dem Modell trotz identischem Profil hinsichtlich der Verstöße erklärt sich dabei durch den Einfluss von ZUS-LEX. Während *Amerika-ner* wegen seiner Lexikalisierung gegen ZUS-LEX verstieße, wäre *Brundtland-Büro* mangels Lexikalisierung nach *Name* der bessere Kandidat.

Eine Anpassung des Modells auf das Zeitungs-Korpus ist dementsprechend nicht notwendig. Auch wenn in dieser Textform überraschend viele Belege abseits von *GRAPH und NAME von der Zusammenschreibung abweichen, handelt es sich bei dieser Beleg als zweigliedriges Kompositium mit *Brundtland* als nicht weiter zu analysierendem Erstglied aufgefasst.

den suboptimalen Kandidaten zumindest um die bestmöglichen Alternativen. Damit bestätigt sich, dass Zeitungstexte standardnah und einer textorientierten Schreibhaltung entsprechend, wie erwartet (vgl. 4.2.4), weitgehend modellkonform hinsichtlich der Schreibung an Wortbildungsgrenzen sind.

6.5.5 Fachzeitschriften

Die 5.000er-Stichprobe des Fachzeitschriften-Teilkorpus hat 450 Belege für wortbildungsmorphologisch komplexe Substantive ergeben. 368 davon sind in die optimalitätstheoretische Modellierung der Schreibung an den wortinternen Grenzen eingegangen. Unter diesen 368 Belegen finden sich neben 233 zweigliedrigen, determinativen Komposita auch 135 Affigierungen.

Nach dem erarbeiteten Modell werden 358 Belege als optimale und nur 10 Belege als suboptimale Kandidaten erfasst, was einer Deckungsquote von 97,3% entspricht, die nur minimal unter der Quote von 97,5% über alle Textformen hinweg liegt.

Während alle Affigierungen zusammengeschrieben und als optimale Kandidaten erfasst werden, ist unter den Determinativkomposita, wie Tabelle 6.11 zeigt, mehr Varianz zu beobachten.

Tabelle 6.11.: Modellierungsgrundlage: Fachzeitschriften

	Optimalität	
	optimal	suboptimal
Determinativkomposita	223 (215 zusammen)	10 (2 zusammen)
Affigierungen	135 (135 zusammen)	0 (0 zusammen)
Summe	358 (350 zusammen)	10 (2 zusammen)

Die optimalen Kandidaten umfassen auch acht Bindestrichschreibungen auf der Basis von *GRAPH (104a) und NAME (104b).

- (104) a. Experten-Know-How (VDI09),⁵⁹ DLR-Institute, FZI-Vorstand, VDI-Gesellschaft, Zwei-Körper-System, CODATA-Initiative (SPK)

⁵⁹ Trotz dreier graphisch deutlich sichtbarer Bestandteile fällt der Beleg *Experten-Know-How* unter die zweigliedrigen Komposita, da *Know-How* keine Komposition sondern eine Zusammenrückung ist. Die Bindestrichschreibung macht *Know-How* dann zum graphematisch markierten Wort und sorgt für die Bindestrichschreibung nach *GRAPH an der Kompositionsgrenze.

b. Peco-Trainerinnen, Fritz-Beske-Instituts (VDI09)

Unter den suboptimalen Kandidaten sind neben Bindestrichschreibungen auch eine Getrennt- und eine Zusammenschreibung (zwei Instanzen des selben Lexems) belegt:

- (105) a. Partner-Gremium, Ideen-Auswahl, Trikot-Sponsoren (VDI09)
 b. VDI nachrichten (VDI09, 2mal)
 c. Parkinsonpatienten (SPK, 2mal)

Die Getrenntschreibung *VDI nachrichten* bestätigt die Positionierung von *GRAPH über ZUS-MORPH auch für die Publikationsform der Fachzeitschriften. Die Zusammenschreibung von *Parkinsonpatienten* trotz Eigennamen als erstem Teilausdruck, ohne dass die Verbindung lexikalisiert ist, ändert nichts an der Einstufung von NAME über GPKB, die durch die Kandidaten unter (104b) hinreichend legitimiert ist. Dementsprechend ist keine Anpassung des Modells für diese Textform nötig.

Das Tableau unter 6.5.10 auf Seite 248 zeigt zudem, dass im Fachzeitschriften-Korpus, abgesehen von *VDI nachrichten* aufgrund des Verstoßes gegen N^o-GROSS, die suboptimalen Kandidaten, wie in den Zeitungstexten, nur einfach abgestuft sind.

Bei einer isolierten Betrachtung der beiden Medien (*VDI nachrichten* und *Spektrum*), aus denen sich das Teilkorpus zusammensetzt, spräche die Beleglage aber für eine Anpassung des Modells für die Zeitschrift *Spektrum*. Durch eine Abstufung von NAME unter GPKB wäre hier ein Deckungsgrad von 100% zu erreichen. Da aus dieser Zeitschrift aber überhaupt nur zwei Abweichungen von der Zusammenschreibung (*Zwei-Körper-System*, *CODATA-Initiative*) belegt sind, bleibt es hier wie bei den Zeitungstexten bei der Andeutung, dass eine medienspezifische Vorliebe zur Grenzmarkierung möglich ist.

Für den Bereich der zweigliedrigen Komposita kann also im Fachzeitschriften-Korpus keine größere Bandbreite von Varianten oder eine besonders häufige Abweichung von der Zusammenschreibung konstatiert werden. Die in 4.2.5 formulierte Vermutung, dass eine Häufung komplexerer Kompositionen, Abweichungen von der Zusammenschreibung begünstigt, kann aber erst in Kapitel 6.6 verifiziert werden, in dem diese komplexeren Komposita erstmals berücksichtigt werden.

6.5.6 Kriminalromane

Mit nur 208 Belegen für wortbildungsmorphologisch komplexe Substantive liefert das Kriminalroman-Korpus neben den Mail- und SMS-Korpora den geringsten Anteil an Belegen aus den 5.000er-Stichproben. 164 dieser Belege (darunter 125 zweigliedrige Determinativkomposita und 39 Affigierungen) wurden in der Erarbeitung des Modells berücksichtigt.

Tabelle 6.12 zeigt, dass 161 dieser 164 Belege als optimale Kandidaten erfasst werden, was einer Deckungsquote von 98,2% entspricht, die etwas über den 97,5% über alle Textformen hinweg liegt.

Tabelle 6.12.: Modellierungsgrundlage: Kriminalromane

	Optimalität	
	optimal	suboptimal
Determinativkomposita	122 (118 zusammen)	3 (2 zusammen)
Affigierungen	39 (39 zusammen)	0 (0 zusammen)
Summe	161 (157 zusammen)	3 (2 zusammen)

Auch in den Kriminalromanen wird der Großteil der 161 optimalen Kandidaten zusammengeschrieben. Als Abweichungen davon sind unter (106) zwei Lexeme mit einer Bindestrichschreibung für nicht lexikalisierte Verbindungen aufgrund von NAME und eine Bindestrichvariante in Form des Phrasenkompositums (*Zwölf-Uhr-Sendung*) auf der Basis von *GRAPH belegt.

(106) Neckar-Nachrichten (2mal), Mersenne-Zahlen, Zwölf-Uhr-Sendung

Gleich zwei der drei suboptimalen Kandidaten unter (107) werden trotz Relevanz für NAME zusammengeschrieben, obwohl die Verbindungen nicht lexikalisiert sind.

(107) Miss-Wettbewerb, Bonnetweiher, Guisanplatz

Bezüglich einer Einordnung von NAME in der Beschränkungsordnung steht es also in Belegen zwei zu zwei. Dem würde eine Positionierung von NAME auf einer Stufe mit GPKB Rechnung tragen. Die Position von NAME bleibt aber unverändert, weil es mit *Europcar-Niederlassung* noch einen komplexeren Beleg gibt, der eine Po-

sitionierung zwischen ZUS-LEX und GPKB stützt.

Die einzige suboptimale Bindestrichvariante unter den zweigliedrigen Determinativkomposita (*Miss-Wettbewerb*) ist ein Fall für FREMD. Zusammenschreibungen wie *Ozeanwelle*, *Plastikbänder* und *Folienbeuteln* sprechen aber dagegen FREMD über GPK (b) einzuordnen.

Dementsprechend erweist sich auch für das Kriminalroman-Korpus das textformübergreifende Modell als anwendbar. Parallel zum Zeitungs-Korpus sind im Kriminalroman-Korpus, wie in 6.5.10 auf Seite 249 veranschaulicht, auch nur suboptimale Kandidaten zu beobachten, die im Vergleich zu den optimalen als zweitbeste Alternative gelten können. Dieser Befund entspricht damit den in 4.2.6 formulierten Erwartungen, dass neben der Zusammenschreibung nur motivierte Bindestrichschreibungen zu erwarten sind.

6.5.7 Plenarprotokolle

Das Plenarprotokoll-Korpus hat mit 596 Belegen mit Abstand die meisten Belege für morphologisch komplexe Substantive in der 5.000er-Stichprobe geliefert. 499 dieser Belege sind in die Modellierung mit eingeflossen. Neben 307 zweigliedrigen Determinativkomposita finden sich darunter auch 192 Affigierungen. Unabhängig von der Art der Wortbildung ist allen Wortformen gemein, dass es sich um optimale Kandidaten nach erarbeiteten Modell handelt (vgl. 6.5.10). Für die Textsorte der Plenarprotokolle wird also eine Deckungsquote von 100% erzielt, die einmalig ist unter den neun hier untersuchten Textformen.

Tabelle 6.13.: Modellierungsgrundlage: Plenarprotokolle

	Optimalität	
	optimal	suboptimal
Determinativkomposita	307 (302 zusammen)	0 (0 zusammen)
Affigierungen	192 (192 zusammen)	0 (0 zusammen)
Summe	499 (494 zusammen)	0 (0 zusammen)

Ein Blick auf Tabelle 6.13 verdeutlicht dabei, dass es unter den Belegen aus den Plenarprotokollen nur sehr wenige Instanzen gibt, die von einer Zusammenschreibung abweichen. Alle fünf abweichenden Kandidaten in (108) sind determinative

Komposita mit einem graphisch markierten ersten Teilausdruck, die deshalb als Bindestrichschreibung realisiert werden:

- (108) CO₂-Emissionen, 2-Länder-Antrag (2mal), Mehr-Länder-Antrag, Zwei-Stufen-Verfahren

Während bei *CO₂-Emissionen* der Bindestrich nach der Ziffer die nachfolgende Buchstabengraphemkette anknüpft, handelt es sich bei den übrigen Bindestrichschreibungen um Phrasenkomposita mit ebenfalls graphisch markiertem phrasalem Erstbestandteil, der eine Bindestrichschreibung nach *GRAPH zum optimalen Kandidaten macht.

Das Tableau 6.5.10 auf Seite 249 umfasst damit nur eine einzige Stufe für die allesamt optimalen Belege.

Im Vergleich zu den anderen Textformen weisen die Plenarprotokolle damit nicht nur den geringsten Anteil an Varianten abseits der Zusammenschreibung auf, sondern zeigen diese Varianten auch nur im Bereich der Komposition und ausschließlich auf der Basis von *GRAPH. Diese Beobachtungen decken sich damit mit den in 4.2.7 formulierten Erwartungen, dass nur wenige und orthographisch korrekte Bindestrichschreibungen in den Plenarprotokollen auftreten. Tatsächlich handelt es sich bei allen fünf Bindestrichschreibungen an der Kompositionsgrenze um obligatorische Bindestriche gemäß der orthographischen Norm (§ 40 und § 43). Wahrscheinlich ist diese Zurückhaltung hinsichtlich der Grenzmarkierung dem hochoffiziellen Charakter der Plenarprotokolle geschuldet.

6.5.8 Lexikoneinträge

Im 5.000er-Ausschnitt des Lexikoneinträge-Korpus sind 379 wortbildungsmorphologisch komplexe Substantive belegt. 309 davon sind mit ihrer terminalen Grenze in die Modellierung der Beschränkungshierarchie eingegangen. Darunter finden sich neben 223 determinativen Komposita auch 86 Affigierungen.

Nur vier der 309 Kandidaten gelten nach dem erarbeiteten Modell als suboptimal, so dass sich eine Deckungsquote von 98,7% ergibt, die etwas oberhalb der textformübergreifenden 97,5% liegt.

Tabelle 6.14 zeigt, dass sich auch hier die Varianten abseits der Zusammenschrei-

Tabelle 6.14.: Modellierungsgrundlage: Lexikoneinträge

	Optimalität	
	optimal	suboptimal
Determinativkomposita	219 (205 zusammen)	4 (1 zusammen)
Affigierungen	86 (86 zusammen)	0 (0 zusammen)
Summe	305 (291 zusammen)	4 (1 zusammen)

bung auf die Determinativkomposita beschränken, während alle Affigierungen zusammengeschrieben werden und in dieser Schreibung als optimale Kandidaten erfasst sind. Unter den Komposita sind dagegen auch Bindestrichschreibungen belegt. 14 optimale Kandidaten weisen einen Bindestrich auf der Basis von *GRAPH (109a) oder NAME (109b) auf.

- (109) a. EU-Kommission, NABU-Stiftung, NS-Verbrechen, Grand-Slam-Titel, Top-Ten-Spielerin, Grand-Slam-Turnieren, „Top Twenty“-Spielerin
 b. Whakatane-Distrikts, Chopin-Wettbewerb, Treblinka-Prozess, Warrego-Tal, Heiducken-Bewegung, Georgs-Kirche, Robert-Schumann-Philharmonie

Daneben finden sich auch die vier suboptimalen Varianten in (110). Neben den Belegen unter (110a) für die mangels Relevanz für *GRAPH und NAME eine Zusammenschreibung optimal wäre, findet sich mit *Staats(un)wesen* auch ein suboptimaler Beleg in Zusammenschreibung.

- (110) a. Doppel-Halde, Ultra-Rechte, Zinn-Vorkommen
 b. Staats(un)wesen

Staats(un)wesen als suboptimaler Beleg wurde in diesem Kapitel schon unter (34) zur Genüge diskutiert. Wenn man darüber hinwegsieht, dass es sich eigentlich um eine Koordination zweier Lexeme in einem graphematischen Wort und damit nicht um den eigentlichen Gegenstand der Untersuchung handelt, verstößt *Staats(un)wesen* aufgrund der internen Klammerung natürlich gegen GPKB und ist damit im Vergleich zum vermeintlich optimalen *Staatsunwesen* suboptimal. *Staatsunwesen* als graphische Realisierung bildet aber nur eines der beiden sprachlichen Zeichen ab, die

Staats(un)wesen kodiert.

Abgesehen von *Staats(un)wesen* sind also auch in den Lexikoneinträgen, wie bereits in den Zeitungstexten und Kriminalromanen, neben den optimalen Kandidaten nur einfach hinsichtlich ihrer Optimalität abgestufte Kandidaten belegt (vgl. Tableau 6.5.10, Seite 250). Damit orientieren sich auch die Lexikoneinträge entsprechend der dahinterstehenden textorientierten Schreibhaltung eng an den allgemeinen sprachlichen Standards und zeigen, wie unter 4.2.8 als Erwartung formuliert, nur Bindestrichschreibungen als Varianten neben der dominanten Zusammenschreibung.

6.5.9 Diskussionsbeiträge

Die 5.000er-Stichprobe aus dem Diskussionsbeiträge-Korpus liefert 294 Belege für wortbildungsmorphologisch komplexe Substantive. 249 dieser Belege sind mit ihrer terminalen Grenze bei der Modellierung berücksichtigt worden. Das Verhältnis zwischen Komposita (135 Belege) und Affigierungen (114 Belege) ist dabei so ausgeglichen wie in keiner anderen Textform.

Wie in Tabelle 6.15 veranschaulicht, stehen 242 optimalen Kandidaten sieben suboptimale bezüglich der Schreibung an der terminalen Wortbildungsgrenze gegenüber. Damit hat das Modell, angewandt auf die Diskussionsbeiträge, einen Deckungsgrad von 97,2%, der damit fast den textformübergreifenden 97,5% entspricht.

Tabelle 6.15.: Modellierungsgrundlage: Diskussionsbeiträge

	Optimalität	
	optimal	suboptimal
Determinativkomposita	129 (121 zusammen)	6 (1 zusammen)
Affigierungen	113 (113 zusammen)	1 (1 zusammen)
Summe	242 (234 zusammen)	7 (2 zusammen)

Unter den optimalen Kandidaten finden sich acht Belege (111), die von einer bloßen Zusammenschreibung abweichen und deren Bindestrichvarianten durch *GRAPH und NAME legitimiert sind.

- (111) a. PDF-Dateien, CV-Geschichte, POV-Ausdrücke, CAD-Systemen, BILD-Archiv

b. Maxwell-Gleichungen, Rayleigh-Zone, Opel-Krise

Die Mehrheit der sieben suboptimalen Varianten weist einen nicht modellkonformen Bindestrich auf (112a). Die Schreibung *Beob-Liste* ist dabei aufgrund der Kurzform als erstem Teilausdruck orthographisch korrekt, wird aber nicht durch *GRAPH abgedeckt, weil *Beob* der graphematischen Silbenstruktur entspricht und die Kürzung nicht durch einen Abkürzungspunkt (*Beob.*) angezeigt wird, der wiederum eine Zugriffsmöglichkeit für *GRAPH böte.

- (112) a. Beob-Liste, Mineral-Box, Salami-Taktik, Striptease-Tänzerin
 b. Drei Säulentheorie, Parkinsonoid
 c. (Kämpfer-)Lieder

Auch die beiden Zusammenschreibungen unter (112b) sind diskussionswürdig. Während beim Phrasenkompositum *Drei Säulentheorie* an der Wortbildungsgrenze zusammengeschieden wird trotz der durch ein Leerzeichen markierten phrasalen Grenze, ist *Parkinsonoid* der einzige nicht lexikalisierte Beleg für eine Affigierung eines Eigennamens in den 5.000er-Stichproben, der zusammengeschieden wird.

(Kämpfer-)Lieder verhält sich parallel zum oben unter den Lexikoneinträgen diskutierten *Staats(un)wesen*.




Selbst wenn man *(Kämpfer-)Lieder* außen vor lässt, ist mit der Zusammenschreibung von *Drei Säulentheorie* zumindest eine Wortform belegt, die nicht nur einfach abgestuft ist im Vergleich zu dem optimalen Kandidaten (*Drei Säulen-Theorie*). Das Tableau 6.5.10 auf Seite 251 umfasst also drei Optimalitätsstufen. Diese Beobachtung deckt sich, wie erwartet (vgl. 4.2.9), mit dem Befund aus den beiden anderen Kommunikationsformen (Mail und SMS), die auch einzelne Belege unterhalb der ersten beiden Optimalitätsstufen aufweisen. Da es sich jeweils nur um Einzelbelege handelt, ist der Befund aber nur ein erstes Indiz für eine größere Bandbreite wortinterner Varianten in Texten, denen eine interaktionsorientierte Schreibhaltung zugrundeliegt.

6.5.10 Textformspezifische Tableaus





Werbeanzeigen

	N°-GROSS	MAJUS	*GRAPH	ZUS-MORPH	ZUS-LEX	NAME	GPKB	ZUS-FE	FREMD	GETR-AUSDR	MINUS	*SPAT
a. ☞ Privatkunden ☞ Leistungen ☞ CO ₂ -Ausstoß							*		*	*		
b. Bio-Handel							*			*	*	
c. FeuchtCreme		*								*	*	
d. ANTI-OXIDANT		*					*			**	**	
e. direkt-granulat	*						*			***		

Mail

	N°-GROSS	MAJUS	*GRAPH	ZUS-MORPH	ZUS-LEX	NAME	GPKB	ZUS-FE	FREMD	GETR-AUSDR	MINUS	*SPAT
a.  Namensschild										*		
 Sicherheit										*		
 Kfz-Versicherung							*			**	*	
b. Browser-Fenster										**	*	
BWL Studiums				*						**	*	
c. LautsprecherBox		*								***	*	





SMS

	N°-GROSS	MAJUS	*GRAPH	GPKB	ZUS-LEX	NAME	ZUS-MORPH	ZUS-FE	FREMD	GETR-AUSDR	MINUS	*SPAT
a.  Lernstress										*		
 versicherung										*		
 USB Verlängerung							*			**	*	*
 acht Uhr Vorstellung							*			**	*	*
b. Abi-Prüfung				*					*	**	*	
c. linguisten mist	*						*			*		*

Zeitung

	N°-GROSS	MAJUS	*GRAPH	ZUS-MORPH	ZUS-LEX	NAME	GPKB	ZUS-FE	FREMD	GETR-AUSDR	MINUS	*SPAT
a. ☞ Bodenmatte (BRZ09)										*		
☞ Amerikaner (HAZ09)						*				*		
☞ SPD-Zeitung (HAZ09)							*			**	*	
☞ Luther-Schau (M09)							*			*	*	
☞ Günther-Brand-Halle (NUZ09)							*			*	*	
b. Brundtlandbüro (M09)						*				*		
Murrel-Kühle (BRZ09)							*			*	*	
Hoffmann-von-Fallerleben Museum (BRZ09)				*						**	*	
Ex-Präsident (M09)							*			***	*	

Fachzeitschriften

	N°-Gross	MAJUS	*GRAPH	ZUS-MORPH	ZUS-LEX	NAME	GPKB	ZUS-FE	FREMD	GETR-AUSDR	MINUS	*SPAT
a.  Innenleben (VDI09)										*		
 Entdeckung (SPK)										**		
 FZI-Vorstand (VDI09)							*			**	*	
 Peco-Trainerinnen (VDI09)							*			***	*	
b. Parkinsonpatienten (SPK)						*				**		
Ideen-Auswahl (VDI09)							*			**	*	
c. VDI nachrichten (VDI09)	*			*						*		*





Kriminalromane

	N°-GROSS	MAJUS	ZUS-MORPH	*GRAPH	ZUS-LEX	NAME	GPKB	ZUS-FE	FREMD	GETR-AUSDR	MINUS	*SPAT
a. Außendeich										*		
Aufklärung										**		
Zwölf-Uhr-Sendung							*			**	*	
Neckar-Nachrichten							*			**	*	
b. Miss-Wettbewerb							*			***	*	
Bonnetweiber						*				*		





Plenarprotokolle

	N°-GROSS	MAJUS	*GRAPH	ZUS-MORPH	ZUS-LEX	NAME	GPKB	ZUS-FE	FREMD	GETR-AUSDR	MINUS	*SPAT
a. Drucksache										*		
Abstimmung										**		
CO ₂ -Emissionen							*			***	*	
Mehr-Länder-Antrag							*			**	*	

Lexikoneinträge

	N°-GROSS	MAJUS	ZUS-MORPH	*GRAPH	ZUS-LEX	NAME	GPKB	ZUS-FE	FREMD	GETR-AUSDR	MINUS	*SPAT
a.  Holzfasern										*		
 Begleitung										**		
 NABU-Stiftung							*			**	*	
 Treblinka-Prozess							*			**	*	
b. Doppel-Halde										*	*	
Christuskirche						*				*		
c. Staats(un)wesen				*			**			**		

Diskussionsbeiträge

	N°-GROSS	MAJUS	*GRAPH	ZUS-MORPH	ZUS-LEX	NAME	GPKB	Zus-FE	FREMD	GETR-AUSDR	MINUS	*SPAT
a.  Landkarten										*		
 Auflistung										**		
 PDF-Dateien							*			*	*	
 Opel-Krise							*			*	*	
b. Mineral-Box							*		*	**	*	
Parkinsonoid						*				*		
c. Drei Säulentheorie			*							**	*	
(Kämpfer-)Lieder			*				**			**	*	

6.5.11 Zusammenfassung

Die vorausgehende Untersuchung zur Textformspezifik wortinterner Varianten an wortbildungsmorphologischen Grenzen in zweigliedrigen, determinativen Komposita und Affigierung hat neben kleinen Unterschieden zunächst einmal offenbart, dass das erarbeitete Modell über alle Textformen hinweg tragfähig ist.

Tabelle 6.16 macht deutlich, dass die Beschränkungshierarchie in allen neun Textformen über 90% der Schreibungen als optimale Kandidaten erfasst. Die schlechteste Quote wird für das SMS-Korpus mit nur knapp 94% erreicht, während die Belege aus dem Plenarprotokoll-Korpus zu 100% optimale Kandidaten hinsichtlich der Schreibung an den untersuchten wortbildungsmorphologischen Grenzen sind. Für alle übrigen Textformen wird eine Deckungsquote von knapp 96% bis gut 98% erreicht.

Einzig für das SMS-Korpus führt ein Positionswechsel von ZUS-MORPH und GPKB auf Basis der in den abweichenden Varianten dominierenden Getrennschreibung zu einer Verbesserung der Deckungsquote um etwa drei Prozentpunkte, da so die Getrennschreibung statt der Bindestrichschreibung für von *GRAPH betroffene Belege zum Optimalfall wird. Eine Aufwertung von FREMD hätte die Deckungsquote in allen Textformen trotz der Einschränkung durch ZUS-LEX und ZUS-FE verschlechtert, so dass dieser Varianten-Trigger durchgängig unterschwellig bleibt.

Die Gegenüberstellung der Zahlen für die optimalen und suboptimalen Belege in 6.16 inklusive der Angabe, wie viele davon jeweils zusammengeschrieben werden, zeigt außerdem, dass andere Varianten neben der Zusammenschreibung bei optimalen Belegen eher die Ausnahme sind, während unter den suboptimalen Belegen nur vereinzelte Zusammenschreibungen zu finden sind, die meistens auf eine Zusammenschreibung von Belegen zurückzuführen sind, die für NAME relevant und nicht lexikalisiert sind.

Tabelle 6.16.: Textformspezifik finales Modell

Textform	Belege	optimal	suboptimal	Quote allgemeines Modell	spezifisches Modell
Werbeanzeigen	421	404 (389 zusammen)	17 (2 zusammen)	96,4%	–
Mail	119	116 (112 zusammen)	3 (0 zusammen)	97,5%	–
SMS	63	59 (59 zusammen)	4 (0 zusammen)	93,7%	96,8%
Zeitung	338	324 (305 zusammen)	14 (1 zusammen)	95,9%	–
Fachzeitschriften	368	358 (350 zusammen)	10 (2 zusammen)	97,3%	–
Kriminalromane	164	161 (157 zusammen)	3 (2 zusammen)	98,2%	–
Plenarprotokolle	499	499 (494 zusammen)	0 (0 zusammen)	100%	–
Lexikoneinträge	309	305 (291 zusammen)	4 (1 zusammen)	98,4%	–
Diskussionsbeiträge	249	242 (234 zusammen)	7 (2 zusammen)	97,2%	–
Summe	2.530	2.468 (2.391 zusammen)	62 (10 zusammen)	97,5%	–

Dass das erarbeitete Modell dennoch nicht nur Zusammenschreibungen, sondern auch einen Großteil der davon abweichenden Varianten als optimale Kandidaten abbildet, veranschaulicht Diagramm 6.1.

Die Linien für den prozentualen Anteil an Zusammenschreibungen und der Deckungsquote des Modells ähneln sich zwar in ihrem Verlauf und stimmen für das SMS-Korpus sogar im Wert überein, da hier tatsächlich nur Zusammenschreibungen als optimale Kandidaten erfasst sind, ein identischer Verlauf liegt aber nicht vor.

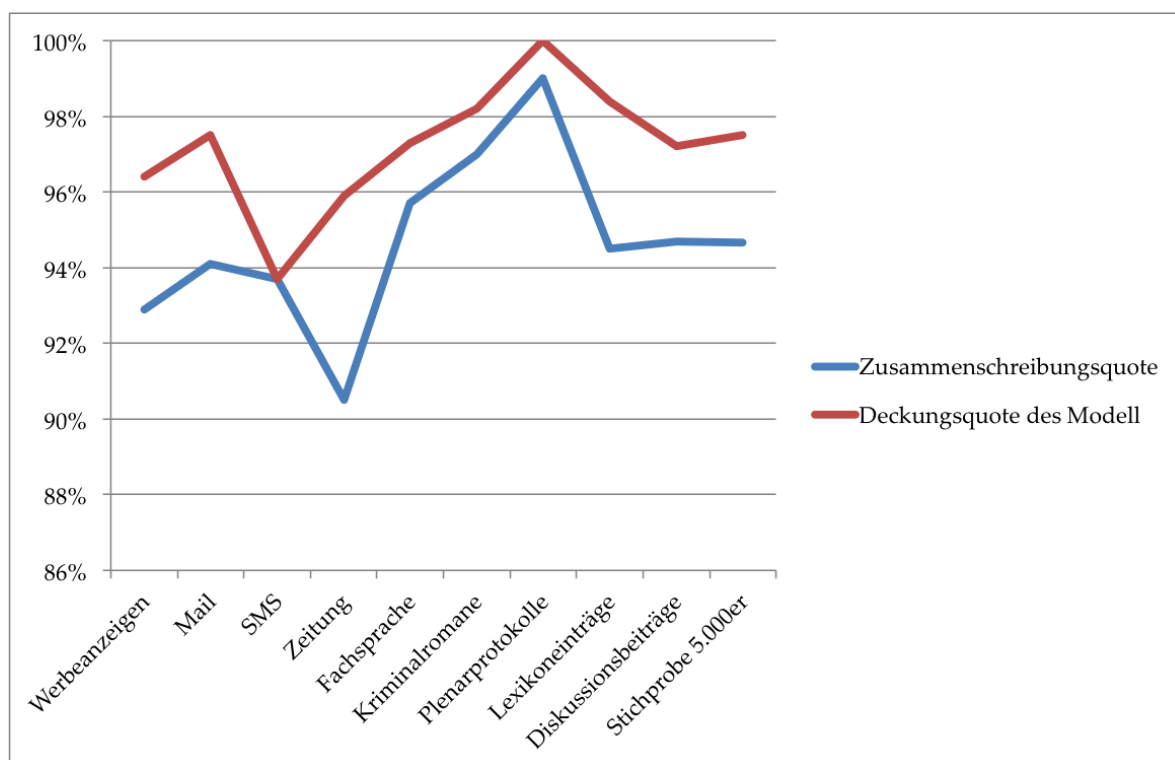


Abbildung 6.1.: Zusammenschreibung und Deckungsgrad nach Textformen

Gerade bei den Textformen mit einer vergleichsweise niedrigen Quote an Zusammenschreibungen zeigt sich, dass das Modell auch einen Großteil der davon abweichenden Varianten erfasst. Deutlich wird dies insbesondere bei den Werbeanzeigen, den Zeitungstexten, den Mails, den Lexikoneinträgen und Diskussionsbeiträgen, wo sich die Linien in Abbildung 6.1 am weitesten voneinander entfernen. So weisen im Extremfall beim Zeitungskorpus nur gut 90% der Belege eine Zusammenschreibung auf, vom Modell werden aber 96% als optimale Kandidaten erfasst. Für die Textformen Fachzeitschriften, Kriminalromane und Plenarprotokolle liegen die Linien dagegen sehr nah beieinander. Bei Zusammenschreibungsquoten über 96% wird aber

Tabelle 6.17.: Textorientierte vs. interaktionsorientierte Schreibhaltung

textorientiert	interaktionsorientiert
Plenarprotokolle	Mail
Zeitungen	SMS
Fachzeitschriften	Diskussionsbeiträge
Lexikoneinträge	
Kriminalromane	
Werbenanzeigen	

auch der Raum für Varianten als optimale Kandidaten knapp.

Was die Bandbreite der wortinternen Varianten an wortbildungsmorphologischen Grenzen angeht, lassen sich die Textformen recht klar in zwei Gruppen einteilen, und zwar anhand der Anzahl der Optimalitätsstufen nach dem erarbeiteten Modell, die durch die jeweiligen Belege besetzt sind.

Die Belege aus den Plenarprotokollen stehen als optimale Varianten hinsichtlich der Verschriftlichung wortbildungsmorphologischer Grenzen allesamt auf einer (der obersten) Optimalitätsstufe. Sie bilden zusammen mit den Belegen aus den Zeitungen, Kriminalromanen, Lexikoneinträgen und Fachzeitschriften eine Gruppe mit eng eingegrenzter Varianz, da in Zeitungen, Kriminalromanen, Lexikoneinträgen und Fachzeitschriften zwar suboptimale Belege zu finden sind, diese aber allesamt auf der nächstmöglichen Optimalitätsstufe stehen, also gewissermaßen die zweitbeste Wahl sind.⁶⁰

Dieser Gruppe gegenüber stehen die Textformen mit einem größeren Variantenreichtum: SMS, Mail, Diskussionsbeiträge und Werbeanzeigen. Für all diese Textformen sind innerhalb der jeweiligen Stichprobe, wie oben diskutiert, zumindest drei Optimalitätsstufen belegt.

Diese Gruppeneinteilung (zusammengefasst in 6.17) entspricht dabei mit Ausnahme der Werbeanzeigen den Kategorien textorientiert vs. interaktionsorientiert. Die

⁶⁰ Im Zuge dieser Gruppenbildung wird der Beleg *VDI nachrichten* (VDI09), der im Fachzeitschriften-Korpus als einziger Beleg noch eine stärker untergeordnete Optimalität hat, aufgrund des Verstoßes gegen N°-GROSS ausgeklammert, da er als Zeitschriftentitel starken Eigennamencharakter (und damit auch eine Tendenz zu graphematischen Eigenarten) hat und daher nicht repräsentativ für die Textform der Fachzeitschriften ist.

interaktionsorientierten Textformen SMS, Mail und Diskussionsbeiträge,⁶¹ deren Fokus mehr auf dem kommunikativen Prozess liegt als auf dem textlichen Produkt desselben, lassen dadurch offenbar eine etwas größere Bandbreite an Varianten zu. Die textorientierten Textformen (Plenarprotokolle, Zeitungen, Kriminalromane, Lexikoneinträge und Fachzeitschriften) sind dagegen als Produkte redaktioneller Prozesse inklusive Überarbeitung und Korrektur weniger zugänglich für suboptimale Varianten. Werbeanzeigen nehmen in diesem Zusammenhang eine Ausnahmestellung ein, da sie die größere Bandbreite an Varianten abseits der eigentlich jeweils optimalen bieten, es sich bei den Anzeigen aber um elaborierte Texte als Produkte einer textorientierten Schreibhaltung handelt. Diese Widersprüchlichkeit erklärt sich aber dadurch, dass in den Werbeanzeigen Systemverstöße auf graphematischer Ebene funktionalisiert sind und dazu dienen, die Aufmerksamkeit des Rezipienten zu erregen.

6.6 Anwendbarkeit des Modells auf komplexe Substantive im Allgemeinen

Im Rahmen der Modellierung für Kompositionsgrenze in Determinativkomposita und bei der Übertragung auf die Derivationsgrenzen hat sich der Fokus immer ausschließlich auf die Verschriftlichung der terminalen Grenze zwischen den unmittelbaren Konstituenten der substantivischen Wortbildung gerichtet. Das gilt auch für die Anwendung des Modells auf die Belege aus den einzelnen Textformen, auch wenn dabei in den Tableaus bereits Verstöße an mittelbaren Wortbildungsgrenzen angezeigt wurden.

Es ist also noch nicht abschließend geklärt, inwiefern sich das Modell auch dafür eignet, die Schreibvarianten an wortbildungsmorphologischen Grenzen – jenseits der bisher betrachteten terminalen Grenzen in zweigliedrigen Determinativkompo-

⁶¹ Obwohl drei von drei Kommunikationsformen unter den neun untersuchten Textformen in die Kategorie interaktionsorientiert fallen, ist nicht für alle Kommunikationsformen von dieser Zuteilung auszugehen. In Abhängigkeit vom situativen Rahmen gibt es auch Kommunikationsformen mit einer textorientierten Schreibhaltung, beispielsweise geschäftliche oder behördliche Korrespondenz per Brief.

sita und Affigierungen – auch in komplexeren Substantiven mit mehr als zwei Kompositionsgliedern oder anderen Wortbildungsmustern zu erfassen. Um das herauszufinden, wird zunächst die Anwendbarkeit des Modells auf alle Wortbildungsgrenzen der bisher betrachteten Substantive getestet, um anschließend auch die Passung für die 576 noch nicht untersuchten Belege aus den 5.000er-Stichproben, die mehr als zwei Kompositionsglieder aufweisen oder auf andere Wortbildungsmuster zurückgehen, zu prüfen.

Zumindest für den Bereich der Komposition gibt es nach Scherer (2012) und Geilfuß-Wolfgang (2013) Grund zur Annahme, dass bei drei oder mehr Kompositionsgliedern durch eine oder mehrere Abweichungen von der Zusammenschreibung die morphosyntaktische Struktur angezeigt werden könnte. Dieser potentielle Effekt wird aber bisher im Modell nicht abgebildet.

Die 2.530 Belege, die bereits im Zuge der optimalitätstheoretischen Modellierung berücksichtigt wurden, umfassen in Summe 5.088 Wortbildungsgrenzen (darunter 1.778 determinative Kompositionsgrenzen und 3.294 Derivationsgrenzen). Anders als bisher handelt es sich dabei nicht mehr ausschließlich um Wortbildungsgrenzen der substantivischen Wortbildung. Zusätzlich zu den bereits betrachteten Grenzen vor substantivischen Kompositionsköpfen und substantivbildenden Suffixen sowie zwischen Präfix und Substantiv finden sich darunter auch die Grenzen von Wortbildungsmustern anderer Wortarten als mittelbare Wortbildungsgrenzen. So wird beim Beleg *Reinigung* (NUZ09) nicht mehr nur die terminale Grenze vorm substantivbildenden *-ung* erfasst und bewertet, sondern auch die mittelbare Grenze vor *-ig*, die zu einem verbalen Wortbildungsmuster gehört. Im Fall des Belegs *Unangreifbarkeit* (HAZ09) kommen zur bereits erfassten substantivischen vor *-keit* gleich zwei adjektivische (nach *un-* und vor *-bar*) und eine verbale Wortbildungsgrenze nach *an-*.

Suboptimale Varianten finden sich dennoch nur an 59 Kompositionsgrenzen wie in *Doppel-Halde* (WPD11) und *Parkinsonpatienten* (SPK), vier Derivationsgrenzen wie in *Ex-Präsident* (M09) und *Parkinsonoid* (WDD11) und einer weiteren Wortbildungsgrenze in *Experten-Know-How* (VDI09). Die Bindestrichvariante *Know-How* ist dabei die einzige suboptimale an einer mittelbaren Wortbildungsgrenze. Dementsprechend werden 5.024 von 5.088 Schreibungen an Wortbildungsgrenzen vom finalen Modell als optimale Kandidaten erfasst, was einer Deckungsquote auf Grenzebene von 98,7% entspricht.

Tabelle 6.18.: Noch nicht erfasste Belege nach Wortbildungsarten

Wortbildungsart	Belege
Komposition (> 2 Kompositionsglieder)	414
Derivation (implizit)	56
Konversion	70
Sonstiges	36
Stichprobe 5000er	576

Dieser Befund legt den Schluss nahe, dass das Modell sich auch abseits der terminalen determinativen Kompositionsgrenzen und Affigierungsgrenzen an mittelbaren Grenzen der Substantivbildung als tragbar erweist. Bei diesen mittelbaren Wortbildungsgrenzen, deren Schreibung dann auch vom erarbeiteten Modell erfasst wird, handelt es sich allerdings fast ausschließlich um Derivationsgrenzen, da mehr als zweigliedrige Komposita noch nicht mit einbezogen wurden und andere Wortbildungsmuster vergleichsweise selten sind.

Die 576 noch nicht erfassten Belege der 5.000er-Stichproben sind zu gut zwei Dritteln singular. 132 Belege finden sich zwei- bis dreimal und nur zehn Belege dieser Gruppe sind in mehr als drei Instanzen belegt. Hinsichtlich der Wortbildungsart liegt der Schwerpunkt, wie Tabelle 6.18 zeigt, ganz klar auf den komplexeren Komposita mit drei oder mehr Kompositionsgliedern neben denen die belegten impliziten Derivationen und Konversionen und sonstige Wortbildungsmuster nur eine untergeordnete Rolle spielen.

Die Verteilung der 576 noch nicht analysierten Belege (und den komplexen Komposita als größter Gruppe darunter) über die Textformen, die Tabelle 6.19 zeigt, entspricht dabei in etwa den generellen Anteilen der einzelnen Textformen an der Gesamtbelegmenge auf Basis der 5.000er-Stichproben. Auch hier tragen SMS, Mail, Kriminalromane und Diskussionsbeiträge die wenigsten Belege bei, während die Werbeanzeigen und Plenarprotokolle die größten Anteile beisteuern. Die komplexen Determinativkomposita machen dabei über alle Textformen hinweg die Mehrheit der Belege aus.

Eine Anwendung des Modells auf die 576 Belege, die bisher noch nicht diskutiert wurden, ergibt auf der Wortebene 495 optimale und 77 suboptimale Kandidaten hinsichtlich der Schreibung an Wortbildungsgrenzen, was einer Deckungsquo-

Tabelle 6.19.: Noch nicht erfasste Belege nach Textform

Textform	Belege	davon komplexe Komposita
Werbeanzeigen	114	98
Mail	22	15
SMS	13	7
Zeitung	89	61
Fachzeitschriften	82	57
Kriminalromane	44	20
Plenarprotokolle	97	64
Lexikoneinträge	70	59
Diskussionsbeiträge	45	33
Stichprobe 5000er	576	414

te von knapp 85,9% entspricht. Diese Passung ist im Vergleich zur Modellierungsgrundlage, deren Belege zu 97,5% als optimale Kandidaten abgebildet werden, also deutlich schlechter, aber mit knapp 86% immer noch auf einem sehr hohen Niveau. Für alle 3.106 Belege für wortbildungsmorphologisch komplexe Substantive aus den 5.000er-Stichproben in Summe und bei Berücksichtigung aller wortbildungsmorphologischen Grenzen wird mit dem Modell aufgrund des geringen Anteils der noch nicht berücksichtigten Belege eine Quote von knapp 95,4% erreicht. Damit erweist sich das Modell auch ohne weitere Anpassung als tragfähig für die Modellierung wortinterner Schreibvarianten in morphologisch komplexen Substantiven.

Dass es unter den Derivationen, bei denen es sich hier ausschließlich um implizite handelt, und Konversionen keine Abweichungen von der Zusammenschreibung und auch keine suboptimalen Kandidaten gibt, überrascht nicht. Sowohl bei den Konversionen wie *Verschmelzen* (SPK) oder *Unterricht* (SMS), als auch bei den impliziten Derivationen wie *Unterschied* (NUZog) erfolgt der letzte Wortbildungsschritt und der damit verbundene Wechsel der Wortart nicht an einer Grenze zwischen Teilausdrücken, die markiert werden könnte, sondern innerhalb eines Teilausdrucks durch eine Ablautung (*Unterschied*), den Wegfall von Flexionsendungen (*Unterricht*) oder ohne jegliche Veränderung abseits der Substantivgroßschreibung (*Verschmelzen*). Streng genommen müssen diese Wortbildungen also bei der Ermittlung der Deckungsquote des Modells auf Wortebene also außen vor bleiben, weil es keine

Tabelle 6.20.: Noch nicht erfasste Belege nach Wortbildungsart und Optimalität

	Optimalität	
	optimal	suboptimal
Komposita	345 (307 zusammen)	69 (4 zusammen)
Derivationen	56 (56 zusammen)	0 (0 zusammen)
Konversionen	70 (70 zusammen)	0 (0 zusammen)
Sonstiges	27 (26 zusammen)	9 (0 zusammen)
Summe	499 (459 zusammen)	78 (4 zusammen)

substantivische Wortbildungsgrenze gibt, auf die das Modell anwendbar ist. Ohne implizite Derivationen und Konversionen ergibt sich eine etwas schlechtere Deckungsquote des finalen Modells von 95,3% für alle Belege der 5.000er-Stichproben und 82,7% für die komplexen Komposita und sonstigen Wortbildungsmuster im Speziellen. Auf der Grenzebene können aber die Schreibungen der Wortbildungsgrenzen vorangegangener Wortbildungsprozesse als optimale Kandidaten erfasst werden, wie die Zusammenschreibung zwischen *ver* und *schmelzen*.

Wie zu erwarten, erweisen sich die komplexen Komposita dagegen als Schwerpunkt der wortinternen Schreibvariation. An ihren Grenzen wird einerseits besonders häufig von der Zusammenschreibung abgewichen, und andererseits sind sie für den mit Abstand größten Teil suboptimaler Kandidaten unter den Schreibungen der 576 Belege verantwortlich (vgl. Tabelle 6.20).

104 der 414 komplexen Komposita werden nicht durchgängig zusammengeschrieben, was in etwa 25% dieser Gruppe entspricht. Bei den 1.653 zweigliedrigen Komposita weichen dagegen nur knapp 7% von der Zusammenschreibung ab. Nur drei Zusammenschreibungen erweisen sich als suboptimal. Was in allen vier Fällen daran liegt, dass an einer der Kompositionsgrenzen wie im Beispiel *Marientorgraben* (NUZ09) nach einem Namen keine Grenzmarkierung gemäß NAME vorgenommen wird, obwohl keine Lexikalisierung der Verbindung vorliegt. 307 zusammengeschriebene Belege als optimale Kandidaten nach dem finalen Modell sprechen demnach dafür, dass die Zusammenschreibung auch für komplexere Komposita mit drei oder mehr Kompositionsgliedern der Standardfall ist.

Zumindest abseits der Wirkungsbereiche von *GRAPH und NAME. Die beiden Beschränkungen sorgen bei immerhin 38 Komposita dafür, dass Varianten wie in (113)

mit einem oder gar mehreren Bindestrichen die optimalen Kandidaten sind, abhängig davon, welche Position der markierte Teilausdruck im Wort einnimmt und ob er dadurch an einer oder zwei Grenzen beteiligt ist.

- (113) FIFA-Hompage (HMP09), Euro-5-Niveau (W), Mersenne-Primzahl (DIV), Rhein-Mosel-Stadt (RHZ09)

Mit 65 Belegen, die an zumindest einer Wortbildungsgrenze von der Zusammenschreibung abweichen, ohne dass diese Varianten vom erarbeiteten Modell gedeckt sind, gelten aber gleich zwei Drittel der nicht durchgängig zusammengescriebenen komplexen Komposita als suboptimale Kandidaten, während bei den zweigliedrigen Komposita deutlich mehr als die Hälfte der Abweichungen von der Zusammenschreibung als optimale Kandidaten erfasst werden können. Zudem sind die suboptimalen Kandidaten sehr ungleichmäßig über die Textformen verteilt. Neben 16 komplexen Komposita mit suboptimaler Variante an einer oder mehreren Kompositionsgrenzen im Zeitungskorpus und 30 entsprechenden Belegen aus den Werbeanzeigen, die damit die Hälfte der Gruppe ausmachen, sind in allen anderen Textformen nur jeweils einige wenige Vertreter diese Kategorie belegt. Das gilt auch für Textformen wie die Plenarprotokolle und die Fachzeitschriften, die wie Tabelle 6.19 zeigt, zwar etwa ebenso viele komplexe Komposita beisteuern wie das Zeitungskorpus, bei denen sich darunter aber jeweils nur zwei suboptimale Kandidaten befinden.

Unter den 65 suboptimalen Kandidaten (114), die an zumindest einer Wortbildungsgrenze von der Zusammenschreibung abweichen, lassen sich mehrere Gruppen ausmachen.

- (114) a. Fußball-Bund, Promi-Autopilot, Tollhaus-Stimmung, Schauspielhaus-Intendant (HMP09), Prinzen-Landsitz (RHZ09), Zweitliga-Saison, Handball-Marathon (NUZ09), Vierer-Mittelfed, Kinder-Weihnachtsfeier, Weihnachtsmann-Zipfelmützen, Landesliga-Volleyballerinnen (BRZ09), Musikschul-Geschäftsstelle (M09, 3mal), Bio-Goldhirse, Bio-Fachhandel, Top-Motorsport, Bio-Granatapfel, Online-Coaching, Kopfhaut-Experten, Bio-Pflanzenauszügen, Rundum-Leistungspaket, Thymiankraut-Dickextrakt (W), Doppel-

- Rangliste, Hartplatz-Saison, Hochzeits-Malers, Offshore-Anlagen, Offshore-Windpark (3mal), Doppel-Weltrangliste (WPD₁₁), Bratwurst-Metzger (SMS), Fußball-Bundesliga, Indoor-Outdoor-Navigation (VDI₀₉), Sachverständigen-Seminar (Mail), Briefkasten-Niederlassung (DIV), Arbeitnehmer-Entsendegesetz, Körperschaftsteuer-Bemessungsgrundlage (PBR)
- b. After-Streich-Party, Gesamt-Seminar-Treffen (Mail)
 - c. CousCoussalat (Mail)
 - d. Commerzbank Filiale, Granatapfel Gesichtspflege (2mal) (W)
 - e. ForMe-Umfrage, PriTarder[®]-Technik, EcoBoost-Technologie, TipMatic[®]-Getriebe, EcoBoost-Motorentechnologie, optiway-GarantiePlus-Vertrages, Mercedes-Benz Partner (2mal), Vagisan[®] FeuchtCreme (6mal), Commerzbank PremiumKonto, Commerzbank PremiumKreditkarte, 1.4 Turbo-Benzinmotor, N^o1 LUXUS-MARKE (W), FüAK-Lehrgangsteilnehmer, DHTB Single B-Seiten Kollektion (WDD₁₁)
 - f. Austausch(Exportformat) (WDD₁₁)⁶²

Die größte Gruppe (vgl. (114a)) darunter machen Bindestrichschreibungen aus, in denen unabhängig davon, ob der Beleg drei (*Fußball-Bund* (HMP₀₉)) oder vier Kompositionsglieder (*Landesliga-Volleyballerinnen* (BRZ₀₉)) umfasst, ausschließlich an den jeweils terminalen Kompositionsgrenzen zwischen den unmittelbaren Konstituenten ein Bindestrich gesetzt wird. Zwei Bindestriche und damit eine Markierung aller Kompositionsgrenzen weisen lediglich die beiden dreigliedrigen Komposita unter (114b) auf. Daneben findet sich ein Beleg mit Binnenmajuskel (114c) und zwei weitere mit einer Getrenntschreibung (114d) an der terminalen Kompositionsgrenze.

Die Belege, die eine Kombination verschiedener Varianten abseits der Zusammenschreibung aufweisen in (114e), entstammen ausschließlich den Werbeanzeigen und Diskussionsbeiträgen. An dieser Gruppe lässt sich erstmals die Hypothese von Scherer (2012: 69) überprüfen, dass in komplexen Komposita die Grenzen zwischen unmittelbaren Konstituenten stärker markiert werden als die zwischen mittelbaren.

⁶² Dieser Beleg fällt nicht nur aufgrund der Klammerung im Vergleich zu den anderen aus dem Rahmen, es handelt sich dabei aufgrund des Originalkontextes, wie unter 6.2.1 bereits diskutiert, sehr wahrscheinlich auch um eine Koordination im Sinne von Austausch- und Exportformat.

Dabei gilt ihr das Leerzeichen als stärkste Form der Markierung aufgrund der deutlichen Unterbrechung der Zeichenkette, vor Bindestrich und Apostroph also der Einfügung eines zusätzlichen Zeichens, während die Binnemajuskel als schwächste Markierungsform ohne zusätzliches Zeichen auskommt. Tatsächlich wird die Rangfolge (Leerzeichen>Bindestrich>Binnemajuskel) zur Abbildung der morphologischen Konstituentenstruktur von fast allen Belegen eingehalten. Am häufigsten ist dabei die Kombination einer Binnemajuskel an einer mittelbaren Kompositionsgrenze mit einem Bindestrich an der unmittelbaren also terminalen Grenze wie im Beispiel *EcoBoost-Motorentechnologie* (W). Aber auch Kombinationen von Leerzeichen und Bindestrich (*Mercedes-Benz Partner* (W)) bzw. Leerzeichen und Binnemajuskeln (*Commerzbank PremiumKonto* (W)), die die Wortstruktur entsprechend anzeigen, sind belegt. Die Beispiele beschränken sich aber fast ausschließlich auf die Belege aus den Werbeanzeigen, so dass es sich womöglich um ein werbespezifisches Phänomen handelt. Auch die Markierung der unmittelbaren Kompositionsgrenzen in (114a) passt zu dieser Annahme, da ausschließlich die unmittelbaren und keine der mittelbaren Kompositionsgrenzen eine Markierung aufweist. Einzig die Schreibung von *CousCoussalat* (Mail) zeigt eine alleinige Markierung einer mittelbaren Wortbildungsgrenze auf und ist damit nicht mit der Idee vereinbar, dass die morphosyntaktische Struktur durch eine hierarchisch angezeigte Grenzmarkierung transparent gemacht wird.

Diese Regelmäßigkeit darf aber nicht darüber hinwegtäuschen, dass auch für komplexe Determinativkomposita die Zusammenschreibung der absolute Normalfall ist, was 307 optimale Kandidaten in gänzlicher Zusammenschreibung wie *Drehtüreffekt* (VDI09) und *Fußballweltmeister* (VDI09) mehr als deutlich belegen. Da nur ein Viertel der 414 komplexen Determinativkomposita lexikalisiert ist, ist auch nicht davon auszugehen, dass eine graphematische Markierung der Struktur abseits lexikalisierter Belege die bevorzugte Variante ist. Zwar konzentrieren sich die Grenzmarkierungen auf die nicht lexikalisierten Belege (90 von 104 Abweichungen von der Zusammenschreibung)⁶³, aber mit 212 Belegen werden gut zwei Drittel der Belege, die nicht lexikalisiert sind, durchgängig zusammengeschieden.

⁶³ Die Recherche nach zusammengeschiedenen Varianten der betreffenden Lexeme hat keine Treffer ergeben. Mit *Paulskirche* finden sich lediglich zwei Konstituenten des Belegs *Peter-Pauls-Kirche* (WPD11) in Zusammenschreibung.

Nimmt man also eine Beschränkung an, die für eine graphematische Markierung der Wortstruktur sorgt, ist diese wohl nicht nur ZUS-LEX, sondern auch GPKB zuzuordnen und bleibt damit unterschwellig.

Die Beleglage für eine Überprüfung der Hypothesen und Ergebnisse von Geilfuß-Wolfgang (2013) zur Schreibung dreigliedriger Determinativkomposita ist eher dünn. Nach Anschluss von Belegen, die für *GRAPH oder NAME relevant sind oder als lexikalisiert gelten, verbleiben noch 171 Belege für dreigliedrige Determinativkomposita. Wie zu erwarten, sind mehr Linksverzweigungen wie *Drehtüreffekt* (VDI09) als Rechtsverzweigungen wie *Feldhandbuch* (WDD11) belegt (102:69).⁶⁴ Von der durchgängigen Zusammenschreibung weichen nur 19 der 171 Belege ab, davon finden sich neun unter den 102 Linksverzweigungen und zehn unter den 69 Rechtsverzweigungen. Der Unterschied zwischen Links- und Rechtsverzweigungen erweist sich aber als nicht signifikant ($\chi^2=0,83$; $p > 0,05$).

Obwohl die Ergebnisse der Fragebogenstudie von Geilfuß-Wolfgang (2013) dafür sprechen, dass zumindest rechtsverzweigende, dreigliedrige Determinativkomposita mit einem Bindestrich zwischen den unmittelbaren Konstituenten besser bewertet werden als zusammengeschriebene, bleibt eine entsprechende Grenzmarkierung in den Korpusbelegen die Ausnahme. Übereinstimmend mit den Ergebnissen von Geilfuß-Wolfgang (2013) finden sich aber fast keine Belege mit einer Markierung der mittelbaren Grenze bei unmarkierter unmittelbarer Grenze, die in seiner Studie schlechter bewertet wird als eine durchgängige Zusammenschreibung. Für die linksverzweigenden, dreigliedrigen Determinativkomposita ergibt sich passend zur Datenlage im Korpus keine bessere Bewertung der Schreibung durch einen Bindestrich an der terminalen Konstituentengrenze im Vergleich zur Zusammenschreibung.

Bei den Komposita mit mehr als drei Kompositionsgliedern zeigt sich ein etwas anderes Bild. Klammert man auch hier Belege aus, die Angriffspunkte für *GRAPH, NAME oder ZUS-LEX bieten, verbleiben noch 42 Kandidaten mit mindestens vier Kompositionsgliedern. Mehr als vier Kompositionsglieder, nämlich genau fünf weisen lediglich die beiden Beispiele *Weihnachtsmann-Zipfelmützen* (BRZ09) und *Indoor-Outdoor-Navigation* (VDI09) auf.⁶⁵ Bei der Verteilung dieser Gruppe über die

⁶⁴ Ohne Ausschluss der lexikalisierten Belege wäre das Ungleichgewicht sogar noch größer (162:98).

⁶⁵ *Indoor-Outdoor-Navigation* (VDI09) ist zudem der einzige Beleg dieser Gruppe, der eine zusätzliche Bindestrichschreibung abseits der unmittelbaren Konstituentengrenze zwischen *Indoor-Outdoor* und *Navigation* aufweist. Dieser zusätzliche Bindestrich zwischen *Indoor* und *Outdoor* ist wahrscheinlich

Textformen zeigt sich, dass – anders als bei den dreigliedrigen Determinativkomposita – schon nicht mehr in allen Textformen entsprechende Belege zu finden sind. Unter den Belegen aus Diskussionsbeiträgen, Kriminalromanen, Mails und SMS-Nachrichten finden sich keine oder nur einzelne (dann aber lexikalisierte) Komposita mit mehr als drei Gliedern.⁶⁶

Nur 24 der 42 Belege mit mindestens vier Kompositionsgliedern, also etwas mehr als die Hälfte, werden durchgängig zusammengeschieden wie *Kleinmüllsammelstelle* (Mog). Die Verzweigungsrichtung scheint dabei auf den ersten Blick keine Rolle zu spielen. Sowohl bei einer beidseitigen Verzweigung wie beispielsweise *Fußballweltmeister* (W) (27 Belege, darunter 14 Zusammenschreibungen und 13 Bindestrichvarianten) als auch bei einer Rechtsverzweigung wie *Doppel-Weltrangliste* (WPD₁₁) (sechs Belege, drei davon zusammen und drei in Bindestrichschreibung) wird jeweils die Hälfte der Belege zusammengeschieden, während die andere Hälfte einen Bindestrich an der terminalen Konstituentengrenze aufweist. Lediglich bei den neun linksverzweigten Komposita dieser Gruppe wie *Mindeststeuersatzproblematik* (PBR) überwiegt die Zusammenschreibung im Verhältnis 7:2. Aufgrund der insgesamt geringen Belegzahlen für diesen Bereich können daraus aber keine tragfähigen Rückschlüsse auf einen Zusammenhang von Verzweigung und der Schreibung an den Wortbildungsgrenzen gezogen werden.

Festzuhalten bleibt, dass die Struktur von mehr als zweigliedrigen Komposita zuweilen regelmäßig und womöglich mit einer steigenden Anzahl der Kompositionsglieder zunehmend durch eine Markierung der Grenze zwischen den unmittelbaren Konstituenten der Wortbildung angezeigt wird, die Zusammenschreibung aber auch bei mehr als zwei Kompositionsgliedern der Normalfall bleibt.

Auf Basis dieser Beobachtungen formuliere ich daher eine ergänzende Beschränkung für das bisherige Modell:

- (115) STRUKTUR
TRANSPARENTE WORTSTRUKTUR

der kopulativen Struktur geschuldet, die sich im substantivischen Bereich nur selten findet, gerade bei kopulativen Adjektiven (*blau-grün*) aber regelmäßig zu Bindestrichvarianten führt.

⁶⁶ Dieses Ungleichgewicht entspricht wiederum, abgesehen von den Kriminalromanen, der Unterscheidung zwischen Textformen, denen eine textorientierte Schreibhaltung zugrundeliegt, und solchen, die interaktionsorientiert entstehen.

Unmittelbare Grenzen zwischen Teilausdrücken im Rahmen einer Wortbildung werden stärker markiert als mittelbare.

Dabei gelten die Abstufungen der Markierungsstärke nach Scherer (2012). Einzuordnen ist STRUKTUR unter GPKB. So wird der Tatsache Rechnung getragen, dass die Zusammenschreibung auch für determinative Komposita mit mehr als zwei Kompositionsgliedern, bzw. mehr als einer Kompositionsgrenze, der Normalfall ist. Durch diese Position lizenziert STRUKTUR also keinesfalls einen Bindestrich an der unmittelbaren Kompositionsgrenze als optimalen Kandidaten, sorgt aber dafür, dass ein entsprechender Kandidat die zweitbeste Wahl nach der Zusammenschreibung ist und eine Grenzmarkierung entgegen der Wortstruktur systematisch zu schlechteren Kandidaten führt. Ein weiterer Vorteil der Position und der Formulierung von STRUKTUR ist, dass auf diese Weise auch hier keine Unterscheidung zwischen Kompositions- und Derivationsgrenzen nötig ist, da sich die Bezeichnung *Teilausdruck* nach Jacobs (2005: 16) sowohl auf Kompositionsglieder als auch auf Affixe (und weitere Morphemklassen) anwenden lässt und unabhängig von der Wortbildungsart die Zusammenschreibung auch bei komplexeren Bildungen der optimale Kandidat bleibt.

(116) Kinder-Weihnachtsfeier (BRZ09), Kernkraftwerksbetreiber, Energieleitungsausbaugesetz, Energieleitungsausbaugesetzes (PBR)

Da nur vier Belege (116) der betrachteten Gruppe ein Fugenelement an der unmittelbaren Konstituentengrenze aufweisen und drei davon zusammengeschieden werden, wird STRUKTUR auch ZUS-FE untergeordnet und steht damit auf einer Stufe mit FREMD. Da aber alle drei Zusammenschreibungen den Plenarprotokollen entstammen, kann zunächst nicht ausgeschlossen werden, dass der vermutete Effekt eines Fugenelements alternativ ein Reflex der Textform ist. Der Befund zu dreigliedrigen Komposita mit einem Fugenelement an der Kompositionsgrenze zwischen den unmittelbaren Konstituenten stützt diese Einordnung der Beschränkung STRUKTUR aber zusätzlich. Unter den oben diskutierten 171 Belegen findet sich nur eine Bindestrichschreibung an der terminalen Grenze trotz Fugenelement (*Hochzeits-Malers* (WPD11)), während alle 42 weiteren Belege für dreigliedrige Determinativkomposita mit einem Fugenelement an der terminalen Wortbildungsgrenze dort auch zusammengeschieden werden wie beispielsweise *Nasenschleimhaut* (W) oder *Friedhofstor*

(WPD₁₁).

(117) N°-GROSS » MAJUS » textsc*Graph » ZUS-MORPH » ZUS-LEX » NAME
 » GPKB, ZUS-FE » STRUKTUR, FREMD » GETR-AUSDR » MINUS,*SPAT

Die konkrete Anwendbarkeit der um STRUKTUR erweiterten Beschränkungshierarchie (117) wird in den folgenden Tableaus ((118) und (119)) für Wortbildungen mit zwei Kompositionsgrenzen am Beispiel *Bergbaustadt* (VDI₀₉) und am Beispiel *Aufflistung* (WDD₁₁) für Wortbildungen mit zwei Affigierungsgrenzen demonstriert. Dabei zeigt sich, dass eine Markierung entsprechend der Wortstruktur unabhängig davon, wie die Grenzmarkierung im konkreten Fall realisiert wird (hier am Beispiel der Bindestrich- und Getrenntschreibung demonstriert), immer zu besseren Kandidaten führt als eine Grenzmarkierung entgegen der Struktur.


Diese Anpassung des finalen Modells führt dementsprechend nicht zu einer verbesserten Deckungsquote auf Wortebene insgesamt (weiterhin bei gut 95%) oder speziell für den Bereich der komplexen Determinativkomposita (weiter unter 90%), spiegelt aber wieder, dass, falls es bei komplexeren Wortbildungen zu einer Grenzmarkierung ohne direkten Einfluss der Wortbestandteile kommt, systematisch die Wortstruktur angezeigt wird. Eine Überprüfung dieser Erweiterung des Modells zum Ranking in (117) auf einer breiteren Datenbasis wäre wünschenswert, um die Effekte von Verzweigung und Anzahl der Kompositionsglieder statistisch besser fassen zu können.

Für das Modell (inklusive der Beschränkung STRUKTUR) spricht aber schon jetzt, dass abseits der optimalen Kandidaten auch der größte Teil der suboptimalen Kandidaten für alle Belege aus den 5.000er-Stichproben als bestmögliche Alternative zum optimalen Kandidaten erfasst wird. Nur 31 der insgesamt 140 als suboptimal klassifizierten Kandidaten werden schlechter bewertet. Dementsprechend gelten insgesamt 3.075 der 3.106 Belege für komplexe Substantive aus den 5.000er-Stichproben entweder als optimale Kandidaten (2.966 Belege) oder als die bestmögliche Alternative (109 Belege), was 99% der Belege entspricht.


Auch bei Betrachtung der Grenzebene wird deutlich, dass das Modell mit oder ohne Erweiterung um die Beschränkung STRUKTUR ganze 6.499 der Schreibungen an den insgesamt 6.651 Wortbildungsgrenzen der Belege aus den 5.000er-Stichproben als optimale Kandidaten erfasst und somit einen Deckungsgrad von 97,7% auf der

Grenzebene erreicht (98,7% für die 2.530 in die Modellierung eingeflossenen Belege und 94,4% für die 576 hier diskutierten verbleibenden Belege). Die bessere Deckungsquote auf der Grenzebene erklärt sich dabei dadurch, dass selbst in den komplexesten Belegen mit einer Vielzahl an Wortbildungsgrenzen fast immer nur eine einzelne Grenze eine suboptimale Schreibvariante aufweist.

(118)

	N°-GROSS	MAJUS	*GRAPH	ZUS-MORPH	ZUS-LEX	NAME	GPKB	ZUS-FE	STRUKTUR	FREMD	GETR-AUSDR	MINUS	*SPAT
a.  Bergbaustadt (VDI09)									*		**		
b. Bergbau-Stadt							*!				**	*	
c. Berg-Baustadt							*		*!		**	*	
d. Bergbau Stadt				*!							*	*	*
e. Berg Baustadt				*					*!		*	*	*

(119)

	N°-GROSS	MAJUS	*GRAPH	ZUS-MORPH	ZUS-LEX	NAME	GPKB	ZUS-FE	STRUKTUR	FREMD	GETR-AUSDR	MINUS	*SPAT
a.  Aufüstung (WDD11)									*		**		
b. Aufüst-ung					*		*!				**		
c. Auf-üstung					*		*		*		**	*!	
d. Aufüst ung				*!	*						*		*
e. Auf üstung				*	*				*		*	*!	*

7. Abschließende Diskussion

Die Grundannahme, dass wortinterne Varianten an wortbildungsmorphologischen Grenzen in komplexen Substantiven systematisch sind und sich in einem optimalitätstheoretischen Modell erfassen lassen, hat sich in dieser Untersuchung bestätigt. Unabhängig davon, ob es sich bei den Wortformen aus dem Untersuchungskorpus um Komposita, Affigierungen oder andere Wortbildungsarten und -muster handelt und in welcher Textform die die Substantive auftreten, erfasst das finale Modell in (1) im Durchschnitt gut 95% aller 3.106 Schreibungen als optimale Kandidaten hinsichtlich der Schreibvarianten an den morphologischen Grenzen im Wortinneren. Auf der Ebene der einzelnen Grenzen ergibt sich sogar ein Deckungsgrad von knapp 98%.

- (1) N°-GROSS» MAJUS» *GRAPH» ZUS-MORPH» ZUS-LEX» NAME» GPKB,
ZUS-FE» STRUKTUR, FREMD» GETR-AUSDR » MINUS,*SPAT

Wesentlich für die Passgenauigkeit des Modells ist dabei einerseits, dass zunächst die Zusammenschreibung als Default – gut 92% der 3.106 Belege werden durchgängig zusammengeschieden – modelliert wurde und zusätzlich ein Großteil der Bindestrichschreibungen – konkret 117 Belege und damit fast die Hälfte der 243 Kandidaten, die nicht ausschließlich zusammengeschieden werden – durch die Beschränkungen *GRAPH und NAME und deren Platzierung innerhalb der Beschränkungshierarchie als optimale Kandidaten erfasst wird. Andere wortinterne Varianten wie die in 31 Belegen erfasste Getrenntschreibung oder eine 17 Belege betreffende Schreibung mit Binnenmajuskel sind wesentlich seltener unter den 243 Belegen des Untersuchungskorpus, die von der durchgängigen Zusammenschreibung abweichen, und nie die optimale Schreibung für eine Wortbildungsgrenze im Sinne des Modells.

Dementsprechend ist nicht von einer willkürlichen Variation nach individueller Präferenz des Schreibers auszugehen, sondern von einem etablierten System. Diese

Erkenntnis ist auf der Basis einer starken, orthographischen Norm, wie es sie für das Deutsche gibt, natürlich allen Bedenken der Sprachkritiker zum Trotz nicht wirklich überraschend. Im Vergleich zur orthographischen Norm, die sich sehr kleinteilig mit diesem Phänomen auseinandersetzt, werden über die optimalitätstheoretische Modellierung die wesentlichen Faktoren in wenigen Beschränkungen erfasst. Neben der optimalen Schreibung, die fast immer der orthographisch korrekten entspricht, bietet die Beschränkungshierarchie aber im Sinne einer abgestuften Optimalität darüber hinaus auch die Möglichkeit, abzubilden, dass bei bestimmten Wortbildungen neben dem eigentlich optimalen Kandidaten auch regelmäßig suboptimale Schreibungen belegt sind. Bei fremdsprachlichen Teilausdrücken und graphematisch markierten Teilausdrücken als Derivationsbasis zeigen sich außerdem systematische Unterschiede zwischen orthographischer Norm und graphematischem System. Das im Modell abgebildete graphematische System begünstigt allgemein eine Grenzmarkierung nach einem ersten fremdsprachlichen Teilausdruck – wenn auch unterschwellig – und bei graphematisch markierten Teilausdrücken unabhängig vom Wortbildungsmuster. Die orthographische Norm dagegen führt einen fakultativen Bindestrich nur zwischen zwei fremdsprachlichen Kompositionsgliedern als Option auf und beschränkt den obligatorischen Bindestrich bei graphematischer Markiertheit auf Kompositionsglieder und Einzelbuchstaben.

Selbst die Belege für suboptimale Schreibvarianten an den Wortbildungsgrenzen stützen das Modell, da es sich in der Regel um Schreibungen handelt, die im Vergleich zur optimalen nur einfach abgestuft sind. So handelt es sich beispielsweise bei sechs der sieben Getrennschreibungen zweigliedriger Determinativkomposita um Belege für die aufgrund von *GRAPH optimalerweise mit einem Bindestrich verschriftlicht werden sollten, beispielsweise *BWL-Studiums* statt *BWL Studiums* (SMS). Deutlichere Optimalitätsverstöße sind dagegen selten und entweder funktionalisiert wie *VDI nachrichten* als graphisch auffälliger Zeitschriftentitel oder den Besonderheiten der Textform geschuldet wie die Getrennschreibung *linguisten mist* (SMS).

Wie erwartet, ist die Zusammenschreibung auch für komplexe Substantive im Deutschen der Standard. Alle anderen Varianten zusammengenommen machen weniger als 10% aus und betreffen zumeist nur die terminale Grenze zwischen den beiden unmittelbaren Konstituenten der Wortbildung. Abseits der Zusammenschreibung ist eine Verschriftlichtung mit Bindestrich wie *PDF-Dateien* (WDD₁₁) am häu-

figsten belegt und nach dem erarbeiteten Modell auch in der Mehrheit der Fälle als optimaler Kandidat auf Basis der Beschränkungen *GRAPH und NAME (bei nicht lexikalisierten Verbindungen) zu bewerten. Getrennschreibungen oder Schreibungen mit Binnenmajuskel sind dagegen vergleichsweise selten und nach dem finalen Modell auch nicht als optimale Kandidaten möglich. Beide Phänomene zeigen sich vereinzelt unter den zweigliedrigen Komposita wie *BWL Studiums* (SMS) und *Feucht-Creme* (W) und etwas häufiger auch miteinander bzw. mit der Zusammenschreibung und Bindestrichschreibung kombiniert in den komplexeren Komposita mit mehr als nur zwei Kompositionsgliedern aus dem Werbeanzeigen-Teilkorpus wie *EcoBoost-Motorentechologie*, *Mercedes-Benz Partner* und *Commerzbank PremiumKreditkarte* (W). In solchen Belegen entspricht die abgestufte Markierung der Wortbildungsgrenzen gemäß STRUKTUR der Wortstruktur.

Allen Grenzmarkierungen ist dabei gemein, dass sie fast ausschließlich dort auftreten, wo sie durch graphematisch markierte, namenartige oder fremdsprachliche Bestandteile begünstigt werden. Getrennschreibungen konzentrieren sich darüber hinaus auf Kandidaten, in denen eine Bindestrichschreibung die eigentlich optimale Variante ist.

In Parallelität zu den Forschungshypothesen werden im Folgenden die weiteren Ergebnisse zu den untersuchten Einflussfaktoren auf die Schreibvarianten an wortbildungsmorphologischen Grenzen zusammenfassend diskutiert.

7.1 Wortbildungsart und Lexikalisierung

Wie erwartet, ist der Variantenreichtum und die Häufigkeit von Abweichungen von der Zusammenschreibung im Bereich der Komposition deutlich größer. Unter den 805 Affigierungen weisen nur zwei Präfigierungen (*Ex-Präsident* (Mo9) und *ANTI-OXIDANT* (W, 2mal) eine Bindestrichschreibung auf, die Zusammenschreibungsquote ist damit bei fast 100%. Unter den 1.725 zweigliedrigen Determinativkompositionen, die nur zu knapp 93% zusammengeschieden werden, finden sich dagegen nicht nur regelmäßig Bindestrichschreibungen (*SPD-Zeitung* (HAZ09)), sondern auch Getrennschreibungen (*acht Uhr Vorstellung* (SMS)) und die Verwendung von Binnenmajuskeln (*FeuchtCreme* (W)). Bei den 414 Komposita mit mehr als zwei

Kompositionsgliedern liegt die Zusammenschreibungsquote nur noch bei etwa 75% und es treten Belege mit einer Kombination verschiedener Grenzmarkierungen wie *EcoBoost-Technologie* (W) auf.

Der Häufigkeitsunterschied zwischen den Komposita und Affigierungen in Bezug auf die Abweichung von einer Zusammenschreibung kann aber für die betrachteten Daten nicht auf den jeweiligen Wortbildungsprozess zurückgeführt werden, schließlich lassen sich beide Gruppen mit dem selben Modell in (1) erfassen, sondern hängt vielmehr damit zusammen, dass die Affigierungen deutlich seltener Teilausdrücke aufweisen, die für *GRAPH (keine Belege) und NAME (lediglich: *Parkinsonoid* (WDD11) und *Amerikaner* (HAZ09, 4mal)) relevant sind, und dadurch eine Abweichung von der Zusammenschreibung provozieren. Deshalb wäre eine kritische Prüfung des Modells auf einer breiteren Datenbasis, die mehr Affigierungen mit entsprechenden Bestandteilen aufweist, wünschenswert. Unter den 2.139 Komposita, die etwa zweieinhalbmal so häufig belegt sind wie Affigierungen, finden sich dagegen gleich 74 Belege mit Relevanz für NAME und 93, die von *GRAPH betroffen sind. Darüber hinaus kann die höhere Lexikalisierungsquote unter den Affigierungen (722 von 805 Affigierungen, also knapp 90% sind lexikalisiert.) bei von NAME betroffenen Belegen wiederum verstärkt für eine Zusammenschreibung sorgen.

Bei den 1.653 prototypischen zweigliedrigen Determinativkomposita sind dagegen nur 757 Varianten, also knapp 46% lexikalisiert, bei den 414 mehrgliedrigen Komposita sogar nur knapp 25%. Unter den 859 lexikalisierten Komposita finden sich, wie zu erwarten, nur 20 Abweichungen (2,3%) von der Zusammenschreibung allesamt auf der Basis von *GRAPH. Unter den 1.208 nicht lexikalisierten Determinativkomposita weichen dagegen 192, also 15,9% der Belege von einer Zusammenschreibung ab.

Obwohl unter den determinativen Komposita deutlich mehr Abweichungen von der Zusammenschreibung belegt sind als unter den Affigierungen, ist dieser Unterschied wohl nicht auf die Wortbildungsart zurückzuführen, sondern hauptsächlich darauf, dass in den Affigierungen sehr selten Teilausdrücke vorkommen, die eine Abweichung von der Zusammenschreibung erforderlich machen, und selbst diese seltenen Fälle bei Eigennamen als Teilausdrücken mit großer Wahrscheinlichkeit durch den Effekt von ZUS-LEX zusammengeschrieben werden.

Der Variantenreichtum dagegen hängt mit der Wortbildungsart bzw. dem Wortbil-

dungsmuster zusammen. Während Kompositionsgrenzen die ganze Bandbreite der beobachtbaren Variation bieten, beschränkt sich die Varianz an den Affigierungsgrenzen der Modellierungsgrundlage neben der Zusammenschreibung auf einzelne Bindestrichschreibungen, als einfach abgestufte suboptimale Kandidaten, also sozusagen die zweitbeste Wahl. Auch außerhalb der Modellierungsgrundlage finden sich abseits der Zusammenschreibung nur vereinzelte Bindestrichvarianten und eine einzige Binnenmajuskelschreibung. Die Daten deuten allerdings auf eine in einem anderen Zusammenhang noch zu klärende Asymmetrie von Präfixen und Affixen hinsichtlich der Grenzmarkierung hin.

Festzuhalten bleibt also: Der Variantenreichtum an Derivationsgrenzen ist im Vergleich zu den Kompositionsgrenzen deutlich eingeschränkter und auch die Häufigkeit von Abweichungen von der Zusammenschreibung ist an Derivationsgrenzen minimal im Vergleich zu den regelmäßigen Abweichungen an Kompositionsgrenzen. Während die unterschiedliche Bandbreite der Variation wohl tatsächlich ein Reflex des Wortbildungsmusters ist, sind die Unterschiede in der Häufigkeit wohl eher dem Einfluss der Wortbestandteile in Kombination mit dem Effekt der Lexikalisierung eines komplexen Wortes geschuldet.

7.2 Einfluss der Bestandteile

Der Einfluss der Wortbestandteile auf die Schreibung an wortbildungsmorphologischen Grenzen ist also der Entscheidende Faktor für die Ausprägung von Varianten abseits der Zusammenschreibung.

Nachgewiesen und entsprechend modelliert werden konnten die Effekte von graphematisch markierten Wörtern (*GRAPH), Eigennamen (NAME) und fremdsprachlichen Bestandteilen (FREMD) als Teilausdrücke, die einer Zusammenschreibung entgegen stehen, sowie von Fugenelementen (ZUS-FE), die für eine Zusammenschreibung sprechen.

Dabei wird sich Bredel (2008) bestätigen, die von einer graphischen Markierung der Wortbindungsrenze ausgeht, wenn dort zwei hinsichtlich ihrer grammatischen Eigenschaften heterogene Einheiten aufeinandertreffen.

7.2.1 Graphematisch markierte Wörter

Den stärksten Effekt entgegen der Zusammenschreibung haben Teilausdrücke, die graphematisch markierten Wörtern entsprechen. Innerhalb des untersuchten Datenausschnittes treten sie nur in Komposita auf und führen in den komplexen Substantiven in 90 von 93 Fällen zu einer Schreibung mit Bindestrich (*FZI-Vorstand* (VDI09)) oder gar einer Getrennschreibung (*USB Verlängerung* (SMS)). Damit sind graphematische markierte Wörter als Teilausdrücke mit Abstand am häufigsten für eine Abweichung von der Zusammenschreibung verantwortlich und betreffen etwas weniger als die Hälfte aller belegten Abweichungen.

Die Belege bestätigen damit die orthographische Norm und die Beobachtungen von Buchmann (2015), dass Wortbestandteile, die graphisch auffällig sind wie Ziffern, Majuskelschreibungen, Schreibungen mit Interpunktionszeichen usw., innerhalb von Wortbildungen einen Bindestrich erforderlich machen, um an andere (graphisch unauffällige) Wortbestandteile anschließen zu können. Neu ist dagegen, dass über die hier erarbeitete Definition der markierten graphematischen Wörter alle Gruppen und Einzelfälle mit der einfachen Beschränkung *GRAPH erfasst werden und ihrem durchschlagenden Effekt durch eine entsprechende Platzierung in der Beschränkungshierarchie über GPKB und ZUS-MORPH gerecht wird, so dass bei graphematisch markierten Wörtern innerhalb von Wortbildungen immer die Schreibung mit Bindestrich der optimale Kandidat ist.

7.2.2 Eigennamen

Ebenfalls deutlich ist der Effekt eines Eigennamens als Teilausdruck. Von den 76 betroffenen Belegen werden auch einige zusammengeschieden, z. B. *Brundtlandbüro*, die Bindestrichschreibung wie in *Luther-Schau* (M09) überwiegt aber und vereinzelt liegen auch Getrennschreibungen wie *Commerzbank Filiale* (W) vor.

Wie die graphematisch markierten Wörter unterscheiden sich auch Eigennamen hinsichtlich ihres grammatischen Verhaltens vom Kernbereich und begünstigen damit eine Abweichung von der Zusammenschreibung. Das schlägt sich auch in der parallelen Formulierungen der Beschränkungen *GRAPH und NAME wieder. Eine gemeinsame Beschränkung würde jedoch der Tatsache nicht gerecht, dass der Effekt von Eigennamen auf die Schreibvariation an Wortbildungsgrenzen eingeschränkter

ist als der Effekt graphematisch markierter Wörter als Teilausdrücke. Innerhalb der Beschränkungshierarchie wird dies durch die Positionierung von NAME auf der unterhalb von ZUS-LEX modelliert.

Eine mögliche Erklärung für diesen Unterschied liegt im Einfluss der orthographischen Norm, gemäß der graphematisch markierte Teilausdrücke obligatorisch zu einer Bindestrichschreibung führen, während sie bei Eigennamen fakultativ ist. Alternativ wäre denkbar, dass sich gerade die graphematische Markiertheit stärker auf der direkt betroffenen Ebene der Schreibung auswirkt, als eine semantische Markiertheit im Fall von Eigennamen.

7.2.3 Fremdsprachlichkeit

Fremdsprachliche Teilausdrücke sind die letzte Kategorie von Wortbestandteilen, die einen Effekt weg von der Zusammenschreibung haben. Wie bei graphematisch markierten Wörtern und Eigennamen liegt auch bei fremden Teilausdrücken eine Markiertheit im Vergleich zum Kernbereich vor. Anders als bei Eigennamen und graphematisch markierten Wörtern überwiegt aber bei fremdsprachlichen Teilausdrücken trotz des durch die logistische Regression bestätigten Effektes in die entgegengesetzte Richtung die Zusammenschreibung sehr deutlich, so dass die entsprechende Beschränkung FREMD innerhalb der Hierarchie so eingeordnet wird, dass GPKB, ZUS-LEX und ZUS-MORPH eine Abweichung blockieren.

Da ein signifikanter Effekt in der logistischen Regression weg von der Zusammenschreibung nur für fremde Teilausdrücke an der ersten Position zweigliedriger, prototypischer Determinativkomposita gezeigt werden konnte, ist FREMD im Gegensatz zu *GRAPH und NAME entsprechend einseitig formuliert. Diese Einseitigkeit passt zur Hypothese von Bredel (2008), dass ein Bindestrich zwischen heterogenen Einheiten steht. Zu überprüfen bleibt, ob dieser Effekt tatsächlich einseitig ist oder sich in einem größeren Datenausschnitt auch positionsunabhängig zeigt.

Festzuhalten bleibt jedenfalls, dass nach dem finalen Modell, die Zusammenschreibung der optimale Kandidat auch für Wortbildungen mit fremdsprachlichen Teilausdrücken ist.

Auch im Fall von FREMD sorgt wahrscheinlich der Einfluss der orthographische Norm, die einen Bindestrich nur in einigen gänzlich fremdsprachlichen Verbindungen und Einzelfällen als Option bietet, dafür, dass unter den betreffenden Belegen

die überwiegende Mehrheit dennoch zusammengeschrieben wird. Somit sind alle Bindestrichvarianten auf der Basis von FREMD Belege für systematische Varianten abseits der kodifizierten Norm.

7.2.4 Fugenelemente

Anders als bei markierten graphematischen Wörtern, Eigennamen und fremdsprachlichen Teilausdrücken konnte für Fugenelemente, wie erwartet, ein Effekt in Richtung Zusammenschreibung nachgewiesen werden, der in der Beschränkung ZUS-FE zum Ausdruck kommt.

Eine Schreibung mit Bindestrich oder einer Getrennschreibung trotz Fugenelementes sind die absolute Ausnahme, da das Fugenelement anders als die bisher besprochenen Einheiten geradezu ein Beleg für die Unmarkiertheit von Teilausdrücken ist und bei einer entsprechenden Grenzmarkierung womöglich auch schwerer zu interpretieren ist. Die wenigen Belege mit Fugenelement, die von der Zusammenschreibung abweichen, zeigen allesamt einen fremdsprachlichen Teilausdruck (*linguisten mist* (SMS)), einen Eigennamen (*Georgs-Kirche* (WPD₁₁)) oder ein graphematisches markiertes Wort als Teilausdruck auf (*Experten-Know-how* (VDI₀₉)). Entsprechend des Verhältnisses von Zusammenschreibung und anderen Varianten für diese Kombinationen wurde ZUS-FE oberhalb von FREMD positioniert, aber unterhalb von *GRAPH und NAME. Letzteres beruht aber auf Einzelbelegen, so dass auf einer breiteren Datenbasis auch eine Positionierung über NAME und *GRAPH nötig sein könnte.

7.3 Textform

Bezüglich des Einflusses der Textform konnte zunächst gezeigt werden, dass es über alle Textformen hinweg eine breite gemeinsame Basis gibt, die sich darin zeigt, dass das Modell für die zweigliedrigen Komposita und Affigierungen für alle Textformen einen Deckungsgrad von mindestens 93,7% bis zu 100% erreicht. Lediglich für die Belege aus dem SMS-Korpus sorgt eine Anpassung des Modells aufgrund der häufigeren Getrennschreibungen für eine noch bessere Passung.

Unterschiede zwischen den Textformen ergeben sich hauptsächlich in Bezug auf

die Bandbreite der Schreibvarianten an wortbildungsmorphologischen Grenzen. Während in den Textformen die das Produkt einer textorientierten Schreibhaltung sind (Zeitungen, Kriminalromane, Lexikoneinträge und Fachzeitschriften) neben optimalen Kandidaten nur einfach abgestufte, suboptimale Kandidaten, also die zweitbeste Wahl nach dem finalen Modell belegt ist, zeigen sich in den interaktionsorientierten Textformen (SMS, Mail, Diskussionsbeiträge) weitere Optimalitätsstufen und damit auch schlechtere Kandidaten. Aus dem Rahmen fallen in diesem Zusammenhang lediglich die Werbeanzeigen, die klar das Ergebnis eines textorientierten Schreibprozesses sind, aber gleichzeitig die größte Bandbreite an wortinternen Varianten auf den verschiedensten Optimalitätsstufen bieten. Für die Werbeanzeigen ist aber davon auszugehen, dass es sich um bewusste Systemverstöße handelt, die im Sinne einer Aufmerksamkeitserregung funktionalisiert sind.

Die Textform hat also – abgesehen von den SMS, wo tendenziell eine Getrenntschreibung einer Bindestrichschreibung vorgezogen wird – keinen Einfluss darauf, welche Kandidaten als optimal gelten, die Schreibhaltung (textorientiert vs. interaktionsorientiert) beeinflusst aber sehr wohl die Bandbreite der wortinternen Schreibvarianten an Wortbildungsgrenzen und hat damit einen ähnlichen Effekt wie die Wortbildungsart (Derivation vs. Komposition).

7.4 Wortstruktur

Auf Basis der 414 komplexeren Komposita mit drei oder mehr Kompositionsgliedern konnte zudem der Einfluss der Konstituentenstruktur der Wortbildungen auf die Schreibvarianten an deren wortbildungsmorphologischen Grenzen modelliert werden. Die Beschränkung STRUKTUR trägt der Tatsache Rechnung, dass, falls von der Zusammenschreibung abgewichen wird, in aller Regel die terminale Wortbildungsgrenze zwischen den unmittelbaren Konstituenten einer Wortbildung markiert wird, und diese Grenzmarkierung damit die Konstituentenstruktur der Wortbildung auf graphischer Ebene sichtbar macht.

Darüber hinaus zeigt sich anhand von Belegen wie *Commerzbank PremiumKonto* und *N°1 LUXUS-MARKE (W)*, dass dort, wo unterschiedliche Varianten abseits der Zusammenschreibung in einem Beleg zusammenkommen, die Stärke der Grenzmar-

kierung die Hierarchieebenen der Konstituentenstruktur widerspiegelt. Die stärkste Grenzmarkierung findet sich also an der unmittelbaren Konstituentengrenze, während die unmittelbaren Grenzen schwächer markiert sind oder unmarkiert bleiben.

Auch dieses Phänomen erfasst STRUKTUR und verhindert damit als Bestandteil des finalen Modells, dass eine Grenzmarkierung entgegen der Wortstruktur zum optimalen Kandidaten werden kann.

7.5 Fazit

Schreibvarianten an wortbildungsmorphologischen Grenzen komplexer Substantive im Deutschen lassen sich also systematisch in einem optimalitätstheoretischen Modell erfassen. Abweichungen vom Standardfall der Zusammenschreibung sind in aller Regel auf grammatische Besonderheiten der Bestandteile zurückzuführen, während die Wortbildungsart (Affigierung vs. Komposition) und die Textform bzw. die damit verbundene Schreibhaltung (textorientiert vs. interaktionsorientiert) einen Rahmen für die jeweils mögliche Bandbreite an Varianten vorgeben, der durch die Optimalitätsstufen im Modell abgebildet wird.

Literaturverzeichnis

- Adamzik, Kirsten. 2012. Werbekommunikation textlinguistisch. In: Nina Janich (Hg), *Handbuch Werbekommunikation*, 123–142. Tübingen: Francke.
- Adamzik, Kirsten. 2016. *Textlinguistik. Grundlagen, Kontroversen, Perspektiven*. Berlin, Boston: De Gruyter, 2. Aufl.
- Anthony, Laurence. 2014. *AntConc [Computersoftware]*. Japan: Waseda University, Dissertation. <http://www.laurenceanthony.net/>.
- Barz, Irmhild. 1993. Graphische Varianten bei der substantivischen Komposition. *Deutsch als Fremdsprache* 30. 167–172.
- Berg, Kristian. 2019. *Die Graphematik der Morpheme im Deutschen und Englischen*. Berlin, Boston: De Gruyter.
- Berg, Kristian, Beatrice Primus & Lutz Wagner. 2016. Buchstabenmerkmal, Buchstabe, Graphem. In: Ulrike Domahs & Beatrice Primus (Hgg), *Handbuch Laut, Gebärde, Buchstabe*, 337–355. Berlin and Boston: De Gruyter.
- Betram, Raymon, Victor Kuperman, R. Harald Baayen & Jukka Hyöna. 2011. The Hyphen as a segmentation cue in triconstituent compound processing: It's getting better all the time. *Scandinavian Journal of Psychology* 52. 530–544.
- Beutelspacher, Albrecht. 2009. *Kryptologie*, Band 9. aktualisierte Auflage. Wiesbaden: Vieweg + Teubner.
- Boettcher, Wolfgang. 2009. *Grammatik verstehen*, Band 1. Tübingen: Niemeyer.
- Borgwaldt, Susanne. 2013. Fugenelemente und Bindestriche in neugebildeten NN-Komposita. In: *Die Schnittstelle von Morphologie und geschriebener Sprache*, 103–134. Berlin, New York: de Gruyter.
- Bredel, Ursula. 2008. *Die Interpunktion des Deutschen. Ein kompositionelles System zur Online-Steuerung des Lesens*. Tübingen: Niemeyer.
- Bredel, Ursula. 2011. *Interpunktion*. Universitätsverlag Winter.
- Buchmann, Franziska. 2015. *Die Wortzeichen im Deutschen*. Heidelberg: Winter.

- Domahs, Frank, Ria de Bleser & Peter Eisenberg. 2001. Silbische Aspekte segmentalen Schreibens – neurolinguistische Evidenz. *Linguistische Berichte* 185. 11–29.
- Donalies, Elke. 2011. *Basiswissen Deutsche Wortbildung*. Basel, Tübingen: Francke, 2. Aufl.
- Drews, Catharina. 2011. *Werden längenhaltige Wörter schneller gelesen als längenlose?* Universität Oldenburg, Magisterarbeit.
- Duden. 2013. *Duden. Die deutsche Rechtschreibung*. Berlin, Mannheim, Zürich: Dudenverlag, 26. Aufl.
- Dürscheid, Christa. 2005. Medien, Kommunikationsformen, kommunikative Gattungen. *Linguistik online* 22(1). 3–16.
- Dürscheid, Christa. 2016. *Einführung in die Schriftlinguistik*. Vandenhoeck & Ruprecht, 5. Aufl.
- Dürscheid, Christa & Sarah Brommer. 2009. Getippte Dialoge in neuen Medien. Sprachkritische Aspekte und linguistische Analysen. *Linguistik online* 37(1). 3–20.
- Eisenberg, Peter. 1989. Die Schreibsilbe im deutschen. In: Peter Eisenberg & Hartmut Günther (Hgg), *Schriftsystem und Orthographie*, Tübingen: Niemeyer.
- Eisenberg, Peter. 2012. *Das Fremdwort im Deutschen*. Berlin, Boston: De Gruyter.
- Eisenberg, Peter. 2013. *Grundriss der deutschen Grammatik*, Band 1: Wort. Stuttgart: Metzler, 4. Aufl.
- Elsen, Hilke. 2011. *Grundzüge der Morphologie des Deutschen*, 2. aktualisierte Auflage. Berlin, Boston: De Gruyter.
- Evertz, Martin. 2016. Graphematischer Fuß und graphematisches Wort. In: Beatrice Primus & Ulrike Domahs (Hgg), *Handbuch Laut, Gebärde, Buchstabe*, Kap. 20, 377–397. Berlin, Boston: De Gruyter.
- Evertz, Martin. 2018. *Visual Prosody*. Berlin, München, Boston: De Gruyter.
- Ewald, Petra. 2012. Werbekommunikation graphostilistisch. In: Nina Janich (Hg), *Handbuch Werbekommunikation*, 3–16. Tübingen: Francke.
- Fandrych, Christian & Maria Thurmair. 2011. *Textsorten im Deutschen*. Tübingen: Stauffenburg.
- Fleischer, Wolfgang & Irmhild Barz. 2012. *Wortbildung der deutschen Gegenwartssprache*. De Gruyter, 4. Aufl.
- Fuhrhop, Nanna. 2000. Zeigen Fugenelemente die Morphologisierung von Kompo-

- sita an? In: Rolf Thierhoff, Matthias Tamrat, Nanna Fuhrhop & Oliver Teuber (Hgg), *Deutsche Grammatik in Theorie und Praxis*, 201–213. Tübingen: Niemeyer.
- Fuhrhop, Nanna. 2007. *Zwischen Wort und Syntagma*. Tübingen: Niemeyer.
- Fuhrhop, Nanna. 2008. Das graphematische Wort (im Deutschen): Eine erste Annäherung. *Zeitschrift für Sprachwissenschaft* 27(2). 189–228.
- Fuhrhop, Nanna & Franziska Buchmann. 2016. Graphematische Silbe. In: Beatrice Primus & Ulrike Domahs (Hgg), *Handbuch Laut, Gebärde, Buchstabe*, Kap. 19, 357–376. Berlin, Boston: De Gruyter.
- Fuhrhop, Nanna & Jörg Peters. 2013. *Einführung in die Phonologie und Graphematik*. Stuttgart, Weimar: J.B. Metzler.
- Gaeta, Livio & Amir Zeldes. 2012. Deutsche Komposita zwischen Syntax und Morphologie: Ein Korpusbasierter Ansatz. In: Livio Gaeta & Barbara Schlücker (Hgg), *Das Deutsche als Kompositionsfreudige Sprache*, 197–218. Berlin: De Gruyter.
- Gallmann, Peter. 1985. *Graphische Elemente der geschriebenen Sprache*. Tübingen: Niemeyer.
- Gallmann, Peter. 1996. Interpunktion (Syngrapheme). In: Hartmut Günther & Otto Ludwig (Hgg), *Schrift und Schriftlichkeit. Ein interdisziplinäres Handbuch internationaler Forschung.*, Band 10 (Handbücher zur Sprach- und Kommunikationswissenschaft 2), 1456–1467. Berlin: De Gruyter.
- Gallmann, Peter. 1997. Konzepte der Nominalität. In: Gerhard Augst, Karl Blüml, Dieter Nerius & Horst Sitta (Hgg), *Die Neuregelung der deutschen Rechtschreibung. Begründung und Kritik*, 209–242. Tübingen: Niemeyer.
- Geilfuß-Wolfgang, Jochen. 2002. Optimal Hyphenation. In: Richard William Sproat, Martin Neef & Anneke Neijt (Hgg), *The Relation of Writing to Spoken Language*, 115–1130. Tübingen: Niemeyer.
- Geilfuß-Wolfgang, Jochen. 2007a. Stammkonstanz ohne Stützform. *Zeitschrift für Sprachwissenschaft* 26. 133–154.
- Geilfuß-Wolfgang, Jochen. 2007b. *Worttrennung am Zeilenende. Über die deutschen Worttrennungsregeln, ihr Erlernen in der Grundschule und das Lesen getrennter Wörter*. Tübingen: Niemeyer.
- Geilfuß-Wolfgang, Jochen. 2013. Gute und schlechte Bindestriche in dreiteiligen Komposita. In: Martin Neef & Carmen Scherer (Hgg), *Die Schnittstelle von Mor-*

- phologie und geschriebener Sprache*, 135–155. Berlin, Boston: De Gruyter.
- Glück, Helmut. 2016a. Autonomiehypothese. In: Helmut Glück & Michael Rödel (Hgg), *Metzler Lexikon Sprache*, 80. Stuttgart: Metzler.
- Glück, Helmut. 2016b. Interdependenzhypothese. In: Helmut Glück & Michael Rödel (Hgg), *Metzler Lexikon Sprache*, 301–302. Stuttgart: Metzler.
- Grube, Henner. 1976. Die Fugenelemente in neuhochdeutschen appellativischen Komposita. *Sprachwissenschaft* 1. 187–222.
- Günther, Hartmut. 1988. *Schriftliche Sprache: Strukturen geschriebener Wörter und ihre Verarbeitung beim Lesen*. Tübingen: De Gruyter.
- Güthert, Kerstin. 2016. Bericht des Rats für deutsche Rechtschreibung über die Wahrnehmung seiner Aufgaben in der Periode 2011 bis 2016. Rat für deutsche Rechtschreibung, Band. https://www.rechtschreibrat.com/DOX/rfdr_Bericht_2011-2016.pdf.
- Habscheid, Stephan (Hg). 2011. *Textsorten, Handlungsmuster, Oberflächen. Linguistische Typologien der Kommunikation*. Berlin, New York: De Gruyter.
- Heller, Klaus. 2000. Binde-Strich und Zergliederungs-Sucht. *Sprachreport. Informationen und Meinungen zur deutschen Sprache* 16(1). 26–27.
- Höhle, Tilman. 1982. Über Komposition und Derivation: zur Konstituentenstruktur von Wortbildungsprodukten. *Zeitschrift für Sprachwissenschaft* 1(1). 76–112.
- Hübl, Annika & Markus Steinbach. 2011. Wie viel Syntax steckt in der satzinternen Großschreibung? *Linguistische Berichte* 227. 255–295.
- Inhoff, Albrecht Werner, Ralph Radach & Dieter Heller. 2000. Complex Compounds in German: Interword Spaces Facilitate Segmentation but Hinder Assignment of Meaning. *Journal of Memory and Language* 24. 23–50.
- Institut für Deutsche Sprache. 2013. *Deutsches Referenzkorpus / Archiv der Korpora geschriebener Gegenwartssprache 2013 II (release vom 19.09.2013)*. Mannheim: Institut für Deutsche Sprache. www.ids-mannheim.de/DeReKo.
- Institut für Deutsche Sprache. grammis. -oid (grippoid). In: *Wörterbuch der Affixe. Grammatisches Informationssystem grammis*, <https://grammis.ids-mannheim.de/affixe/299553>.
- Jacobs, Joachim. 2005. *Spatien, Zum System der getrennt- und Zusammenschreibung im heutigen Deutsch* (Linguistik – Impulse & Tendenzen 8). Berlin, New York: De Gruyter.

- Janich, Nina. 2013. *Werbesprache. Ein Arbeitsbuch*. Tübingen: Narr.
- Klein, Wolf Peter. 2002. Der Apostroph in der Deutschen Gegenwartssprache. Logographische gebrauchserweiterung auf phonographischer Basis. *Zeitschrift für germanistische Linguistik* 30. 169–197.
- Kopf, Kristin. 2017. Fugenelement und Bindestrich in der Compositions-Fuge. In: Nanna Fuhrhop, Renata Szczepaniak & Karsten Schmidt (Hgg), *Sichtbare und hörbare Morphologie*, 177–204. Berlin, Boston: De Gruyter.
- Kupietz, Marc & Harald Lüngen. 2014. Recent Developments in DeReKo. In: Nicoletta Calzolari & al. (Hgg), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 2378–2385. Reykjavik: ELRA.
- Lesezirkel. 2014. <http://www.medien-palette.de/jetzt-sparen/geschaeftswelt.html>.
- Müller, Gereon. 2000. *Elemente der optimalitätstheoretischen Syntax*. Tübingen: Stauffenburg.
- Munske, Horst Haider. 2002. Wortbildungswandel. In: Mechthild Habermann, Horst Haider Munske & Peter O. Müller (Hgg), *Historische Wortbildung*, 23–41. Tübingen: Niemeyer.
- Neef, Martin & Beatrice Primus. 2001. Stumme Zeugen der Autonomie - Eine Replik auf Ossner. *Linguistische Berichte* 187. 353–378.
- Nübling, Damaris. 2014. Sprachverfall? : Sprachliche Evolution am Beispiel des diachronen Funktionszuwachses des Apostrophs im Deutschen. In: Albrecht Plewnia & Andreas Witt (Hgg), *Sprachverfall? : Dynamik – Wandel – Variation*, 99–123. Berlin, Boston: De Gruyter.
- Nübling, Damaris, Fabian Fahlbusch & Rita Heuser. 2012. *Namen. Eine Einführung in die Onomastik*. Tübingen: Narr.
- Ossner, Jakob. 2001a. Das <h>-Graphem im Deutschen. *Linguistische Berichte* 187. 325–351.
- Ossner, Jakob. 2001b. Worum geht es eigentlich? Replik auf die Replik von Martin Neef und Beatrice Primus. *Linguistische Berichte* 187. 379–382.
- Perkuhn, Rainer, Holger Keibel & Marc Kupietz. 2012. *Korpuslinguistik*. Paderborn: Wilhelm Fink.
- Pfeiffer, Marcus. 2002. *Lesen von Komposita*. RWTH Aachen, Magisterarbeit.
- Primus, Beatrice. 2000. Suprasegmentale Graphematik und Phonologie: Die Deh-

- nungszeichen im Deutschen. *Linguistische Berichte* 181. 5–30.
- Primus, Beatrice. 2003. Zum Silbenbegriff in der Schrift-, Laut- und Gebärdensprache – Versuch einer mediumübergreifenden Fundierung. *Zeitschrift für Sprachwissenschaft* 22(1). 3–55.
- Primus, Beatrice. 2004. A featural analysis of the Modern Roman Alphabet. *Written Language and Literacy* 7(2). 235–274.
- Primus, Beatrice. 2006. Buchstabenkomponenten und ihre Grammatik. In: Ursula Bredel & Hartmut Günther (Hgg), *Orthographietheorie und Rechtschreibunterricht*, 5–43. Tübingen: De Gruyter.
- Primus, Beatrice. 2007. The typological and historical punctuation systems Comma constraints. *Written Language and Literacy* 10(2). 103–107.
- Primus, Beatrice. 2010. Strukturelle Grundlagen des deutschen Schriftsystems. In: Ursula Bredel, Astrid Müller & Gabriele Hinney (Hgg), *Schriftsystem und Schriffterwerb*, 9–45. Tübingen: De Gruyter.
- Prince, Alan & Paul Smolensky. 1993. *Optimality Theory. Constraint Interaction in Generative Grammar*. Cambridge, Mass.: MIT Press.
- Programmbereich Korpuslinguistik. 2012. Textklassifikation in COSMAS II.
- R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rat für deutsche Rechtschreibung. 2006. *Deutsche Rechtschreibung. Regeln und Wörterverzeichnis*. Rat für deutsche Rechtschreibung, München, Mannheim.
- Rat für deutsche Rechtschreibung. 2018. *Deutsche Rechtschreibung. Regeln und Wörterverzeichnis*. Rat für deutsche Rechtschreibung, Mannheim, aktualisierte Aufl.
- Roelcke, Thorsten. 2020. *Fachsprache*. Berlin: Erich Schmidt Verlag, 4. Aufl.
- Scherer, Carmen. 2006. *Korpuslinguistik*. Heidelberg: Universitätsverlag Winter.
- Scherer, Carmen. 2010. Das Deutsche und die dräuenden Apostrophe. zur Verbreitung von 's im Gegenwartsdeutschen. *Zeitschrift für germanistische Linguistik* 38. 1–24.
- Scherer, Carmen. 2012. Vom Reisezentrum zum Reise Zentrum. Variation in der Schreibung von N+N-Komposita. In: Livio Gaeta & Barbara Schlücker (Hgg), *Das Deutsche als Kompositionsfreudige Sprache*, 57–82. De Gruyter.
- Scherer, Carmen. 2013. Kalb's Leber und Dienstag's Schnitzeltag: Zur funktionalen Ausdifferenzierung des Apostrophs im Deutschen. *Zeitschrift für Sprachwissen-*

- schaft 32(1). 75–112.
- Schmidt, Günter Dietrich. 1987. Das Affixoid. Zur Notwendigkeit und Brauchbarkeit eines beliebten Zwischenbegriffs der Wortbildung. In: Gabriele Hoppe, Günter Dietrich Schmidt & Wolfgang Rettig (Hgg), *Deutsche Lehnwortbildung. Beiträge zur Erforschung der Wortbildung mit entlehnten WB-Einheiten im Deutschen*, 53–101. Tübingen: Narr.
- Schnitzer, Caroline-Victoria. 2012. *Linguistische Aspekte der Kommunikation in den neueren elektronischen Medien, SMS – E-Mail – Facebook*. München: Ludwig Maximilian Universität, Dissertation.
- Selkirk, Elisabeth O. 1982. *The Syntax of Words*. Massachusetts: MIT Press.
- Starke, Günter. 1993. Droht uns eine Bindestrich-Inflation. *Muttersprache* 103. 50–60.
- Stein, Stephan. 1999. Majuskeln im WortInnen. Ein neuer graphostilistischer Trend für die Schreibung von Komposita in der Werbesprache. *Muttersprache* 109. 261–278.
- Stein, Stephan. 2008. Affixoide - Überlegungen aus semantischer und pragmatischer Perspektive zu einer umstrittenen Einheit der Wortbildung. In: Inge Pohl (Hg), *Semantik und Pragmatik. Schnittstellen*, 179–216. Frankfurt am Main: Peter Lang.
- Sternefeld, Wolfgang. 2000. Schreibgeminanten im Deutschen: Ein Fall für die Optimalitätstheorie. *Linguistische Berichte* 181. 35–54.
- Storrer, Angelika. 2014. Sprachverfall durch internetbasierte Kommunikation? Linguistische Erklärungsansätze - empirische Befunde. In: Albrecht Plewnia & Andreas Witt (Hgg), *Sprachverfall? : Dynamik – Wandel – Variation*, 171–196. Berlin, Boston: De Gruyter.
- Wiese, Richard. 2004. How to optimize orthography. *Written Language and Literacy* 7(2). 305–331.
- Wiese, Richard. 2011. *Phonetik und Phonologie*. Paderborn: UTB.
- Wurzel, Wolfgang. 2002. Morphologische Eigenschaften von Wörtern. In: D. Alan Cruse, Franz Hundsnurscher, Michael Job & Peter Rolf Lutzeier (Hgg), *Lexikologie/Lexicology. Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen/An international Handbook on the Nature and Structure of Words and Vocabularies*, 200–210. Berlin (u.a.): De Gruyter.
- Wurzel, Wolfgang Ullrich. 2000. Was ist ein Wort? In: Nanna Fuhrhop, Rolf Thierhoff, Oliver Teuber & Matthias Tamrat (Hgg), *Deutsche Grammatik in Theorie und*

Praxis, 29–42. Tübingen: Niemeyer.

Wurzel, Wolfgang Ulrich. 1984. *Flexionsmorphologie und Natürlichkeit. Ein Beitrag zur morphologischen Theoriebildung* (Studia Grammatica XXI). Berlin: Akademie-Verlag.

Zifonun, Gisela. 2012. Komposition (oder Halbaffigierung) zum Ausdruck von Nominalaspekt: Schmuckstück, Glücksfall und Zuckerwerk. In: Barbara Schlücker & Livio Gaeta (Hgg), *Das Deutsche als Kompositionsfreudige Sprache: strukturelle Eigenschaften und systembezogene Aspekte*, 101–133. Berlin, Boston: De Gruyter.

A. Vollständige Liste der optimalitätstheoretischen Beschränkungen

- (1) ZUS-MORPH
ZUSAMMENSCHREIBUNG VON MORPHOLOGISCHEN BILDUNGEN
Wenn X und Y Teilausdrücke eines morphologisch gebildeten Zeichens sind, gibt es zwischen <X> und <Y> kein Spatium. (Jacobs 2005: 34)
- (2) GETR-AUSDR
GETRENNTSCHREIBUNG VON TEILAUSTRÜCKEN
Wenn X und Y Teilausdrücke sind, liegt zwischen <X> und <Y> mindestens ein Spatium. (Jacobs 2005: 97)
- (3) *SPAT
SPARSAME SPATIENSETZUNG
Graphemfolgen enthalten kein Spatium. (Jacobs 2005: 127)
- (4) GPK (b)
DEPENDENZ:
Jedes Segment der graphemischen Repräsentation korreliert eindeutig mit einem Segment der phonologischen Repräsentation. (Sternefeld 2000: 36)
- (5) MINUS
GRAPHISCHE REALISIERUNG VON GRAPHEMEN
Grapheme werden als Minuskeln realisiert.

- (6) N^o-GROSS
GROSSSCHREIBUNG
Graphematisch eigenständige Wörter und Wortbestandteile, die eine N^o-Position besetzen, werden großgeschrieben.
- (7) MAJUS
POSITION VON MAJUSKELN
Eine Majuskel leitet eine graphematisch eigenständige Einheit ein.
- (8) *GRAPH
EINBINDUNG VON GRAPHEMATISCH MARKIERTEN WÖRTERN
Teilausdrücke, die graphematisch markierten Wörtern entsprechen, bleiben auch innerhalb morphologischer Bildungen graphematisch eigenständig.
- (9) ZUS-LEX
ZUSAMMENSCHREIBUNG BEI LEXIKALISIERUNG
Wenn X und Y Teilausdrücke eines morphologisch gebildeten Zeichens sind, das als Komplex lexikalisiert ist, sind X und Y graphematisch unselbständig.
- (10) NAME
EINBINDUNG VON EIGENNAMEN
Teilausdrücke, die Eigennamen entsprechen, bleiben auch innerhalb morphologischer Bildungen graphematisch eigenständig.
- (11) FREMD
FREMDSPRACHLICHE TEILAUSTRÜCKE
Fremde erste Teilausdrücke bleiben auch in morphologischen Bildungen graphematisch eigenständig.
- (12) ZUS-FE
FUGENELEMENTE
Teilausdrücke mit Fugenelement sind graphematisch unselbständig.
- (13) STRUKTUR
TRANSPARENTE WORTSTRUKTUR
Unmittelbare Grenzen zwischen Teilausdrücken im Rahmen einer Wortbildung werden stärker markiert als mittelbare