

Kristian Berg

Die Graphematik der Morpheme im Deutschen und Englischen

Konvergenz und Divergenz

Sprachvergleichende Studien zum Deutschen

Herausgegeben von
Eva Breindl und Lutz Gunkel

Im Auftrag des
Leibniz-Instituts für Deutsche Sprache

Gutachterrät

Ruxandra Cosma (Bukarest), Martine Dalmas (Paris), Livio Gaeta (Turin),
Matthias Hüning (Berlin), Sebastian Kürschner (Eichstätt-Ingolstadt),
Torsten Leuschner (Gent), Marek Nekula (Regensburg), Attila Péteri (Budapest),
Christoph Schroeder (Potsdam), Björn Wiemer (Mainz)

Band 10

IDS

LEIBNIZ-INSTITUT FÜR
DEUTSCHE SPRACHE

Kristian Berg

**Die Graphematik
der Morpheme im
Deutschen und
Englischen**

DE GRUYTER

Diese Publikation wurde mit Mitteln aus dem Publikationsfonds für Open-Access-Monografien der Leibniz-Gemeinschaft gefördert.

Redaktion: Dr. Anja Steinhauer

ISBN 978-3-11-060476-4

e-ISBN (PDF) 978-3-11-060485-6

e-ISBN (EPUB) 978-3-11-060518-1



Dieses Werk ist lizenziert unter der Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 Lizenz. Weitere Informationen finden Sie unter <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

© 2019 Kristian Berg, publiziert von Walter de Gruyter GmbH, Berlin/Boston
Dieses Buch ist als Open-Access-Publikation verfügbar über www.degruyter.com.
Satz und Layout: Annett Patzschewitz
Druck und Bindung: CPI books GmbH, Leck

www.degruyter.com

Danksagung

Diese Arbeit ist die überarbeitete Fassung meiner Habilitationsschrift, die ich an der Carl von Ossietzky Universität Oldenburg im Juli 2017 eingereicht habe. Begutachtet wurde sie von Nanna Fuhrhop, Ingo Plag und Beatrice Primus. Allen dreien danke ich herzlich für ihre Mühen.

Den Ausgangspunkt nahm die Arbeit in Nanna Fuhrhops DFG-Projekt „Die Wortschreibung im Deutschen und Englischen“, in dem ich seit 2011 beschäftigt war. Zu Anfang waren die Phonem-Graphem-Korrespondenzen ein zentraler Gegenstand. Wie konsistent ist das Deutsche, wie inkonsistent ist das Englische wirklich? Mit der Zeit jedoch gewann die Idee immer mehr an Reiz, die Systeme rein graphematisch bzw. morphologisch-graphematisch zu beschreiben. Wie weit kann man mit einer solchen Beschreibung kommen? Die vorliegende Arbeit ist eine Antwort auf diese Frage.

Für fachliche Gespräche bedanke ich mich bei Mark Aronoff, Franziska Buchmann, Martin Evertz, Nanna Fuhrhop, Vera Heyer, Beatrice Primus, Niklas Reinken, Karsten Schmidt, Niklas Schreiber und Fabian Tomaschek. Nanna Fuhrhop hat die Arbeit im besten Sinne des Wortes betreut; sie hatte immer ein offenes Ohr und eine pragmatische Lösung.

Ich danke außerdem den Reihenherausgebern Lutz Gunkel und Eva Breindl für die Aufnahme der Arbeit in die Reihe „Konvergenz und Divergenz“ und Sebastian Kürschner, der in diesem Rahmen ein weiteres sehr hilfreiches Gutachten angefertigt hat. Anja Steinhauer hat die Arbeit lektoriert; Annett Patzschewitz von der Publikationsstelle des IDS hat den Satz betreut. Beide sind darüber trotz 150 Tabellen und Abbildungen, trotz spitzer Klammern und fehlender Literaturangaben nicht verzweifelt (glaube ich). Danke schön! Alle verbliebenen Fehler und Ungereimtheiten sind natürlich meine eigenen.

Oldenburg, im August 2018

Inhalt

Abbildungsverzeichnis — IX

Abkürzungsverzeichnis — XIII

1	Einleitung — 1
1.1	Zielsetzung — 1
1.2	Deutsches vs. englisches Schriftsystem — 2
1.3	Theoretischer Rahmen — 4
1.4	Terminologie — 8
1.5	Aufbau — 11
2	Datengrundlage — 13
2.1	CELEX — 13
2.2	CELEX-Teilkorpus einfacher Stämme — 14
2.3	Korpus Affixe — 18
2.4	CELEX-Teilkorpus homophoner Stämme — 19
3	Der graphemische Aufbau von Morphemen — 23
3.1	Inventar: Buchstaben und Grapheme — 24
3.2	Einfache Stämme — 33
3.2.1	Globale Graphotaktik — 33
3.2.1.1	Konsonanten und Vokale — 34
3.2.1.2	Exkurs: Skalierbarkeit der syntagmatischen Klassifizierung — 40
3.2.1.3	Die Verteilung von Bigrammen — 43
3.2.1.4	Anfangsränder von Stämmen — 62
3.2.1.5	Endränder von Stämmen — 64
3.2.1.6	Zusammenfassung — 68
3.2.2	Silbenstrukturelle Graphotaktik — 70
3.2.2.1	Der silbische Bau von Stämmen — 72
3.2.2.2	Anfangsränder — 75
3.2.2.3	Endränder — 87
3.2.2.4	Exkurs: Symmetrie der Anfangs- und Endränder — 101
3.2.2.5	Intervokalische Konsonanten(cluster) — 108
3.2.2.6	Kerne — 111
3.2.2.7	Interaktion der Silbenkonstituenten — 122
3.2.2.8	Zusammenfassung — 148
3.2.3	Minimale Stämme und prototypische Stämme — 150

3.2.4	Minimalpaare und funktionale Last	163
3.2.5	Zusammenfassung	186
3.3	Affixe	192
3.3.1	Der silbische Bau von Affixen	192
3.3.2	Anfangsränder	195
3.3.3	Endränder	196
3.3.4	Kerne	198
3.3.5	Minimale Affixe und prototypische Affixe	200
3.4	Zusammenfassung und Diskussion	204
4	Morphographische Korrespondenzen	209
4.1	Stämme: Einheitlichkeit	220
4.1.1	Flexion	221
4.1.2	Wortbildung	233
4.2	Stämme: Eindeutigkeit	247
4.3	Affixe: Einheitlichkeit	258
4.3.1	Flexion	258
4.3.2	Derivation	267
4.4	Affixe: Eindeutigkeit	276
4.4.1	Flexion	276
4.4.2	Derivation	280
4.5	Zusammenfassung	290
5	Zusammenfassung und Diskussion	295
6	Literatur	305
7	Anhang	317
	Anhang A: Liste der untersuchten Affixe	317
	Anhang B: Absolute Häufigkeiten der Buchstaben im Korpus	318
	Anhang C: Kreuztabellen der Minimalpaare	319
8	Sachregister	321

Abbildungsverzeichnis

- Abb. 1:** Relative Anteile der Buchstaben im deutschen und englischen Teilkorpus einfacher Stämme an der Summe aller Buchstaben im jeweiligen Korpus. N(de): 30.772, N(en): 37.519. — 25
- Abb. 2:** Dendrogramm der deutschen Buchstaben auf der Basis ihrer syntagmatischen Verteilung. — 36
- Abb. 3:** Dendrogramm der englischen Buchstaben auf der Basis ihrer syntagmatischen Verteilung. — 39
- Abb. 4:** Liniendiagramm der Ergebnisse der Monte-Carlo-Simulation zur Klassifikation zufällig ausgewählter Teile der deutschen und englischen Korpora. Die Verhältniswerte geben an, wie viele der 1.000 Durchläufe zur selben Klassifikation führen wie die Clusteranalyse des gesamten Teilkorpus. — 41
- Abb. 5:** Liniendiagramm der Ergebnisse der Monte-Carlo-Simulation zur Klassifikation zufällig ausgewählter Teile der deutschen und englischen Korpora. Die Verhältniswerte geben an, wie viele der 1.000 Durchläufe zur selben Klassifikation (außer für |y|) führen wie die Clusteranalyse des gesamten Teilkorpus. — 42
- Abb. 6:** Venn-Diagramme zu den Schnittmengen der belegten und nicht belegten Bigramme in den Teilkorpora einfacher deutscher Stämme und einfacher englischer Stämme. — 47
- Abb. 7:** Frequenzspektrum der Bigramme im deutschen (links) und englischen (rechts) Teilkorpus einfacher Stämme. Auf der horizontalen Achse ist logarithmisch skaliert die Frequenzklasse der Bigramme aufgetragen, auf der vertikalen Achse die Anzahl der Bigramme in der jeweiligen Frequenzklasse. — 49
- Abb. 8:** Relativer Anteil der Stämme mit 1, 2, 3, 4 und 5 graphematischen Silben an allen Stämmen. Datengrundlage: Teilkorpora einfacher Stämme deutsch und englisch. — 72
- Abb. 9:** Relativer Anteil der Stämme mit 1, 2, 3, 4 und 5 phonologischen Silben an allen Stämmen. Datengrundlage: Teilkorpora einfacher Stämme deutsch und englisch. — 73
- Abb. 10:** Der relative Anteil von Vokalgraphemen an allen Graphemen eines Wortes, geordnet nach der Zahl der graphematischen Silben. Datengrundlage: Teilkorpora einfacher Stämme deutsch und englisch. — 74
- Abb. 11:** Relativer Anteil der Stämme mit 0, 1, 2, 3 und 4 Graphemen im Anfangsrand an allen Stämmen. Datengrundlage: Teilkorpora einfacher Stämme deutsch und englisch. — 76
- Abb. 12:** Relative Anzahl der Konsonantengrapheme an allen einfach besetzten Anfangsrändern von Stämmen in den Korpora einfacher Stämme englisch und einfacher Stämme deutsch. — 77
- Abb. 13:** Vorläufiges graphotaktisches Schema für den Anfangsrand deutscher Stämme. — 80
- Abb. 14:** Graphotaktisches Schema für den Anfangsrand deutscher Stämme. — 83
- Abb. 15:** Vorläufiges graphotaktisches Schema für den Anfangsrand englischer Stämme. — 85
- Abb. 16:** Graphotaktisches Schema für den Anfangsrand englischer Stämme. — 87
- Abb. 17:** Relativer Anteil der Stämme mit 0, 1, 2, 3 und 4 Graphemen im Endrand an allen Stämmen. Datengrundlage: Teilkorpora einfacher Stämme deutsch und englisch. — 88
- Abb. 18:** Durchschnittliche Anzahl von Graphemen im Endrand deutscher Stämme nach der Anzahl graphematischer Silben dieser Stämme. — 89
- Abb. 19:** Relative Anzahl der Konsonantengrapheme an allen einfach besetzten Endrändern von Stämmen in den Korpora einfacher Stämme englisch und einfacher Stämme deutsch. — 89

- Abb. 20:** Vorläufiges graphotaktisches Schema für den Endrand deutscher Stämme. — 92
- Abb. 21:** Graphotaktisches Schema für den Endrand deutscher Stämme. — 95
- Abb. 22:** Vorläufiges graphotaktisches Schema für den Endrand englischer Stämme. — 98
- Abb. 23:** Graphotaktisches Schema für den Endrand englischer Stämme. — 101
- Abb. 24:** Graphotaktische Schemata für den Anfangs- und den Endrand deutscher Stämme. — 102
- Abb. 25:** Graphotaktische Schemata für den Anfangs- und den Endrand englischer Stämme. — 104
- Abb. 26:** Relativer Anteil der Stämme mit 0, 1, 2, 3 und 4 intervokalischen Graphemen an allen graphematisch zweisilbigen Stämmen. Datengrundlage: Teilkorpora einfacher Stämme deutsch und englisch. — 108
- Abb. 27:** Relative Anzahl der Konsonantengrapheme an allen einfachen intervokalischen Konsonantengraphemen zweisilbiger Stämme in den Korpora einfacher Stämme englisch und einfacher Stämme deutsch. — 109
- Abb. 28:** Relativer Anteil der Stämme mit 1, 2, und 3 Graphemen im Kern an allen Stämmen. Datengrundlage: Teilkorpora einfacher Stämme deutsch und englisch. — 112
- Abb. 29:** Relativer Anteil der Stämme mit 1, 2, und 3 Graphemen im Kern an allen graphematisch einsilbigen Stämmen. Datengrundlage: Teilkorpora einfacher Stämme deutsch und englisch. — 113
- Abb. 30:** Relative Anzahl der Vokalgrapheme an allen einfachen Kernen in einsilbigen Stämmen in den Korpora einfacher Stämme englisch und einfacher Stämme deutsch. — 113
- Abb. 31:** Relative Anzahl der Vokalgrapheme an allen einfachen Kernen in ersten Silben von graphematischen Zweisilbern in den Korpora einfacher Stämme englisch und einfacher Stämme deutsch. — 114
- Abb. 32:** Relative Anzahl der Vokalgrapheme an allen einfachen Kernen in zweiten Silben von graphematischen Zweisilbern in den Korpora einfacher Stämme englisch und einfacher Stämme deutsch. — 115
- Abb. 33:** Graphotaktisches Schema der Kombinatorik von zwei Vokalgraphemen in deutschen Silbenkernen. — 116
- Abb. 34:** Graphotaktisches Schema der Kombinatorik von zwei Vokalgraphemen in englischen Silbenkernen. — 117
- Abb. 35:** Graphotaktisches Schema der Kombinatorik von drei Vokalgraphemen in deutschen Silbenkernen. — 121
- Abb. 36:** Graphotaktisches Schema der Kombinatorik von drei Vokalgraphemen in englischen Silbenkernen. — 122
- Abb. 37:** Relativer Anteil der Buchstaben in Minimalpaaren im deutschen und englischen Korpus. — 167
- Abb. 38:** Anteil der Buchstaben, die an Minimalpaaren teilnehmen, relativ zu allen Vorkommen der betreffenden Buchstaben. — 169
- Abb. 39:** Ausschnitt des Netzwerks der Minimalpaare mit vier Buchstaben im deutschen Korpus einfacher Stämme, visualisiert mit Gephi; verwendeter Algorithmus: Yifan Hu. — 181
- Abb. 40:** Relativer Anteil der Affixe mit 0, 1 und 2 graphematischen Silben an allen Stämmen. Datengrundlage: Korpus deutscher bzw. englischer Affixe. — 192
- Abb. 41:** Relativer Anteil der Affixe mit 0, 1 und 2 Graphemen im Anfangsrand an allen Affixen. Datengrundlage: Korpus deutscher bzw. englischer Affixe. — 195
- Abb. 42:** Relativer Anteil der Affixe mit 0, 1, 2 und 3 Graphemen im Endrand an allen Affixen. Datengrundlage: Korpus deutscher bzw. englischer Affixe. — 196

- Abb. 43:** Relative Anzahl der Konsonantengrapheme an allen einfach besetzten Endrändern von Affixen in den Korpora deutscher bzw. englischer Affixe. — 197
- Abb. 44:** Relative Anzahl der Vokalgrapheme an allen Kernen von einsilbigen Affixen in den Korpora deutscher bzw. englischer Affixe. — 199
- Abb. 45:** Relative Anzahl der Vokalgrapheme an allen einfachen Kernen in einsilbigen Affixen in den Korpora deutscher bzw. englischer Affixe. — 199
- Abb. 46:** Die Verteilung prototypischer einsilbiger Stämme und einsilbiger Affixe im Deutschen und Englischen. — 206
- Abb. 47:** Übersicht über morphologische Schreibungen aus Berg et al. (2014: 305). — 209
- Abb. 48:** Homophone im deutschen und englischen Korpus nach Anzahl der graphematischen Silben. Datengrundlage: CELEX-Teilkorpus homophoner Stämme. — 251
- Abb. 49:** Relative Anteile der Buchstaben im deutschen und englischen Teilkorpus einfacher Stämme an der Summe aller Buchstaben im jeweiligen Korpus. N(de): 30.772, N(en): 37.519. — 295
- Abb. 50:** Verdoppelungen von Buchstaben, die signifikant häufiger sind, als stochastisch zu erwarten wäre. — 296
- Abb. 51:** Die Verteilung prototypischer einsilbiger Stämme und einsilbiger Affixe im Deutschen und Englischen. — 297
- Abb. 52:** Relativer Anteil der Schreibvarianten des Suffixes *-ic* bei Wörtern, die heute mit ⟨ic⟩ geschrieben werden. Datenbasis: Helsinki-Korpus (1350–1710) und Google Ngrams (1710–1850). — 301
- Abb. 53:** Relativer Anteil der Schreibvarianten des Suffixes *-ic* bei Wörtern, die heute mit ⟨ic⟩ geschrieben werden. Datenbasis: EEBO-Korpus (bis 1700) und ECCO-Korpus (ab 1700). — 302
- Abb. 54:** Streudiagramm von Texten aus dem EEBO- und dem ECCO-Korpus. Horizontale Achse: Jahr der Veröffentlichung. Vertikale Achse: Relativer Anteil der ⟨ick⟩-Formen im jeweiligen Text. Linie: Mittelwert der relativen Anteile pro Jahr. — 303

Abkürzungsverzeichnis

A	Adjektiv
AmE	Amerikanisches Englisch
BrE	Britisches Englisch
C	Konsonant
COM	Stoffsubstantiv
FEM	Femininum
MASK	Maskulinum
N	Substantiv
NEUT	Neutrum
Part.	Partizip
Pl.	Plural
Präs.	Präsens
Prät.	Präteritum
Ps.	Person
Sg.	Singular
V	Vokal

1 Einleitung

1.1 Zielsetzung

Diese Arbeit hat das Ziel, die Wortschreibung im Deutschen und im Englischen zu beschreiben und in ihren wesentlichen Zügen zu vergleichen. Beide Schriftsysteme sind alphabetisch – das deutsche gilt aber als ein insgesamt regelmäßiges und transparentes System, während die Unregelmäßigkeiten des englischen augenscheinlich Legion sind. Ein Vergleich der beiden Schriftsysteme ist einerseits aus typologischer Perspektive interessant: Wo unterscheiden sich die Systeme wie stark voneinander, wo nicht? Und etwas genereller: Nach welchen Kriterien und mit welchen Methoden können Schriftsysteme überhaupt sinnvoll verglichen werden? Andererseits ist eine Untersuchung auch einzelsprachlich fruchtbar, wie im Verlauf der Arbeit gezeigt wird. Nun liegen aber gerade zum englischen und (in noch größerem Maße) zum deutschen Schriftsystem bereits eine Vielzahl von Arbeiten vor. Kann eine weitere Arbeit überhaupt noch etwas Neues beitragen?

Das kann sie, denn der Schwerpunkt dieser Arbeit liegt – anders als in den meisten existierenden Arbeiten zum Englischen oder Deutschen – auf den *morphographischen Regularitäten*. Damit ist zweierlei gemeint. Erstens geht es um den graphematischen Aufbau von Morphemen: Welche graphematische Form haben Morpheme im Deutschen und Englischen? Welche Einheiten stehen zum Aufbau zur Verfügung und wie kombinieren sie? Wie kombinieren sie nicht? Zweitens stehen morphographische Korrespondenzen im Mittelpunkt: Wie einheitlich und eindeutig werden Morpheme in den beiden Sprachen graphematisch kodiert? Variieren Morpheme in verschiedenen Umgebungen oder werden sie konstant verschriftet?

Die phonographische Perspektive, die in vielen Darstellungen der Graphematik bis heute zentral ist, spielt demgegenüber hier nur am Rande eine Rolle. Das ist einerseits forschungshistorisch begründet. So hat sich die Graphematik – auch vor der Benutzung dieses Begriffs – immer zentral mit den Bezügen zwischen Schriftzeichen und Lauten beschäftigt. Zum Englischen liegen bspw. seit dem Aufkommen der elektronischen Datenverarbeitung fantastisch detaillierte Listen vor, welche Phoneme (z. T. positionsabhängig) mit welchen Graphemen korrespondieren (z. B. Hanna et al. 1966; Dewey 1970 u. v. a.). Mit der Zeit wurden die Analysen detaillierter und bezogen bspw. Silbenkonstituenten und graphematischen Kontext als Determinatoren für Korrespondenzen mit ein (so z. B. Kessler/Treiman 2001). Zum Deutschen liegen keine vergleichbaren empirisch basierten Aufstellungen der phonographischen Korrespondenzen vor. Sie sind

aber auch nicht wirklich nötig, weil die phonographischen Korrespondenzen im Deutschen schon auf den ersten Blick wesentlich eindeutiger sind als im Englischen. Zusammengenommen bedeutet das: Dieses Feld ist gut bearbeitet; es wird schwer sein, etwas grundsätzlich Neues zu phonographischen Korrespondenzen zu entwickeln.

Andererseits ist die Marginalisierung der Phonographie in dieser Arbeit auch Konsequenz des Typus von Schriftsystem, zu dem das englische und deutsche gehören. Es handelt sich bei beiden Schriftsystemen um Alphabetschriften. Für solche Systeme sind phonographische Korrespondenzen konstitutiv. Sie können enger und weiter sein, kontextfrei oder kontextsensitiv, zum Teil auch idiosynkratisch – bis jetzt ist aber kein alphabetisches Schriftsystem bekannt, das sich zu einem vollständig logographischen entwickelt hätte. Regelmäßige phonographische Bezüge sind also nicht überraschend, sie sind gewissermaßen die Nulllinie. Besonders interessant – und da setzt diese Arbeit an – sind vielmehr genau diejenigen Schreibungen, die nicht phonographisch expliziert werden können – zum Beispiel die morphographischen Schreibungen. Natürlich brauchen wir für die Feststellung der Abweichungen Informationen über die regelmäßigen phonographischen Bezüge; sie dienen in dieser Arbeit aber vor allem als Folie.

Damit soll nicht gesagt sein, dass die Beschreibung phonographischer Bezüge ohne theoretische Herausforderungen und Probleme wäre (siehe Abschn. 1.3, wo diese Probleme angesprochen werden); es bedeutet lediglich, dass der Schwerpunkt der Arbeit auf nicht-phonographischen Schreibungen liegt.

1.2 Deutsches vs. englisches Schriftsystem

Warum werden nun in dieser Arbeit ausgerechnet das deutsche und das englische Schriftsystem miteinander verglichen? Zum einen gehören die beiden Sprachen zu den am besten erforschten Sprachen der Welt. Diese Arbeit kann also auf einer starken Forschungsbasis aufbauen. Zum anderen handelt es sich um zwei eng verwandte westgermanische Sprachen, die beide ein Schriftsystem desselben Typs nutzen (Alphabetschrift); gleichzeitig unterscheiden sie sich auf den ersten Blick erheblich, was die Regularitäten und deren Konsistenz angeht. Das deutsche Schriftsystem gilt als ein recht regelmäßiges, das englische Schriftsystem hingegen als notorisch komplex mit unnötig vielen Ausnahmen. Es liegen Dutzende Vorschläge für eine Vereinfachung vor, die in regelmäßigen Abständen vorgebracht werden (vgl. für einen Überblick die Zusammenstellung in Yule/Yasuko 2016).

Die Unterschiede in der Regularität lassen sich mit dem Parameter der Tiefe erfassen. Katz/Frost (1992) haben m. W. den Namen geprägt (‚deep‘ vs. ‚shallow‘ orthographies); Meisenburg (1998) formuliert die Idee weiter und wendet den Parameter auf verschiedene romanische Schriftsysteme in Geschichte und Gegenwart an. Im Kern geht es um die Frage, welche Einheiten im Schriftsystem einer Sprache primär repräsentiert sind. In flachen Schriftsystemen wie dem Spanischen entsprechen sich Grapheme und Phoneme weitestgehend. Wenn man diese Entsprechungen kennt und wenn man weiß, wie ein Wort im Spanischen geschrieben wird, dann kann man es auch aussprechen (und andersherum). Tiefe Schriftsysteme wie das Französische sind im Kern ebenfalls alphabetisch, hier ist die segmentale Entsprechung von Graphemen und Phonemen allerdings überlagert von lexikalischen und morphologischen Informationen. Es bringt uns nicht viel weiter, zu wissen, dass die Infinitivform *regarder* ‚betrachten‘ mit finalelem /e/ realisiert wird; genauso werden (unter anderem) auch das Partizip Passiv (*regardé*) und die 2. Ps. Plural (*regardez*) realisiert. Wir müssen wissen, welchen morphologischen Kategorien die betreffende Wortform zugeordnet ist, um sie richtig zu schreiben.

Zwischen diesen beiden Extremen – flachen Schriftsystemen wie dem Spanischen auf der einen Seite und tiefen Systemen wie dem Französischen auf der anderen Seite – spannt sich ein Kontinuum von Schriftsystemen auf, die typologisch mit dem Parameter der Tiefe beschrieben werden können. Das Englische wird regelmäßig als tiefes Schriftsystem klassifiziert, und zwar als eines, das mehr auf die Wortebene als auf morphologische Informationen Bezug nimmt (vgl. Günther 2004: 1921). Mit anderen Worten: Die Tiefe des englischen Schriftsystems ergibt sich aus der relativen Idiosynkrasie vieler Wortschreibungen. Das deutsche Schriftsystem ist im Vergleich dazu flacher (vgl. Günther 2004: 1919 f.), weil regelmäßiger. Wenn die Auslautverhärtung als phonetisches Oberflächenphänomen modelliert wird (wenn also das Morphem *Hund* auch im Singular zugrundeliegend mit /d/ auslautet), können Grapheme und Phoneme recht eng aufeinander bezogen werden.

Unabhängig von der Angemessenheit dieser phonologischen Interpretation können wir festhalten: Deutsch und Englisch sind zwei eng verwandte und gut beschriebene Sprachen, die alphabetisch verschriftet werden; sie unterscheiden sich aber deutlich in der Konsistenz der Graphem-Phonem-Bezüge. Gleichzeitig unterscheiden sich auch ihre Flexionssysteme: Während das Deutsche noch über ein relativ reichhaltiges Kategoriensystem verfügt, das vor allem (aber nicht ausschließlich) segmental realisiert wird, sind im Englischen nur noch Reste eines solchen Systems zu finden.

Nachdem nun die Auswahl der beiden Schriftsysteme motiviert ist, geht es im Folgenden um die grundlegenden theoretischen Vorannahmen.

1.3 Theoretischer Rahmen

Dieser Arbeit liegen die folgenden drei Annahmen zugrunde:¹

1. Die Graphematik ist Teil des Sprachsystems; phonologische und graphematische Formen sind auf dieselben Einheiten der Inhaltsebene bezogen.

In altverschrifteten Sprachen wie dem Deutschen oder dem Englischen bestehen komplexe Wechselwirkungen zwischen geschriebener und gesprochener Sprache. Die Schrift prägt unser Bewusstsein über Sprache (vgl. Firth 1968; Lüdtko 1969; Aronoff 1992; Stetter 2005). Deswegen sollte die Schrift ein integraler Bestandteil eines Modells der Sprache sein.

Diese Annahme hat zwei Konsequenzen. Die erste lautet: Wörter können – wie andere sprachliche Ausdrücke auch – auf mehreren Ebenen beschrieben werden, und die graphematische Ebene ist eine von ihnen. Wörter haben mindestens eine graphematische Struktur (‹Katz›), eine phonologische Struktur (/kat.sə/),² eine semantisch-konzeptuelle Struktur (‚Katz‘) sowie eine morphologische Struktur ({COM, FEM; Sg}).³ Diese Trennung der Ebenen kann als unkontrovers gelten (vgl. Jacobs 2007). Sie ist in repräsentationellen Modellen wie Jackendoff (1997, 2002) konsequent umgesetzt: Hier sind Phonologie, Syntax und Semantik drei Komponenten, die über unterschiedliche und voneinander unabhängige kombinatorische Systeme verfügen. Zwischen ihnen vermitteln Schnittstellen. Das hier vorgeschlagene Vorgehen ist ohne Weiteres anschlussfähig an diese Grammatikmodelle. Mehr noch: Wie oben angedeutet, ist die Integration der Schrift in solche Modelle zumindest in altverschrifteten Sprachen ein Desiderat.

¹ Vgl. Dryer (2005), der den Begriff „theoretical framework“ synonym mit „descriptive theory“ verwendet, in Abgrenzung zu „explanatory theory“. Eine solche – beschreibende – Theorie soll die Basis dieser Arbeit sein.

² Hier und im Rest der Arbeit wird phonologisches Material zwischen Schrägstrichen repräsentiert, ohne dass damit eine bestimmte phonologische Theorie vertreten wird. Es soll damit auch nicht impliziert werden, dass die zentrale Ebene für Korrespondenzen die phonologische (und nicht die phonetische) ist. Es mag gute Gründe für diese Sichtweise geben (vgl. z. B. Bierwisch 1972 und Kohrt 1985b: 334) – die Frage nach der Bezugsebene ist in dieser Arbeit schlichtweg nicht von Belang. Nebenbei bemerkt kann sie überhaupt nur sinnvoll im Rahmen einer ausformulierten phonologischen Theorie bearbeitet werden.

³ Es werden hier die terminologischen Festlegungen von Eisenberg (2013a, 2013b) verwendet: Wortkategorien (lexikalische Kategorien) wie ‚COM‘ (Stoffsubstantiv) oder ‚FEM‘ (Substantiv im Femininum) werden in Großbuchstaben gesetzt, Einheitenkategorien (Flexionskategorien) wie ‚Sg‘ (Singular) werden nur initial großgeschrieben.

Die zweite Konsequenz dieser Annahme ist: Morphologie und Syntax sind medienneutral. Diese Annahme setzt ein strukturell einfacheres Sprachsystem an als die alternative Annahme zweier – medial differenzierter – Sprachsysteme und sollte daher bevorzugt werden. Die relevanten Einheiten und Relationen der neutralen Morphologie werden gebildet aus der Vereinigungsmenge der graphematisch und der phonologisch ermittelten morphologischen Einheiten und Relationen. Das Vorgehen lässt sich gut am Französischen demonstrieren: Hier wird bspw. Genus bei Partizipien phonologisch nicht gekennzeichnet, graphematisch allerdings schon (◀regardé◀ und ◀regardée◀ sind homophon). In der medienneutralen Morphologie wird die rein graphematische Opposition übernommen: *regardé* wird mit dem Merkmal (genauer: mit der Einheitenkategorie) {Mask} beschrieben, *regardée* mit dem Merkmal {Fem}. Ähnliches gilt für die Syntax. Das syntaktische Wort (nicht das phonologische oder graphematische) ist die Grundeinheit in dieser Arbeit: Wortformen werden im Rahmen dieser Arbeit syntaktisch definiert (vgl. Aronoff 1994; Wurzel 2000). Diese syntaktischen Wörter haben dann wie oben erläutert phonologische, graphematische, morphologische und semantische Teilstrukturen, die in ihrem Zusammenspiel untersucht werden können. Besonders für die graphematische und phonologische Struktur ist es wichtig festzuhalten, dass es sich hier nicht notwendigerweise um graphematische und phonologische Wörter handeln muss. Umgekehrt gibt es graphematische und phonologische Wörter, die keine syntaktischen Wörter sind (Gallmann 1999; Fuhrhop 2008).⁴

Der Gegenstand dieser Arbeit ist die Graphematik und ihr Verhältnis zur Morphologie. Die Orthographie hingegen wird nur am Rande behandelt. Dieser Schwerpunktsetzung liegt die Auffassung Eisenbergs (2013a) zugrunde, dass das Schriftsystem des Deutschen ‚natürlich‘ gewachsen ist und dass es die Aufgabe der Graphematik ist, dessen einschlägige Regularitäten zu ermitteln. Die Basis für diese Ermittlung ist der Schreibgebrauch (Eisenberg 2013a: 287). Die Orthographie ist demgegenüber eine Kodifizierung der graphematischen Regularitäten – und zwar eine von mehreren möglichen; sie ist, genau wie eine graphematische Theorie, eine Theorie über ein Schriftsystem (vgl. Eisenberg 1983).

Es gibt nun allerdings einen Unterschied zwischen der Graphematik einer Sprache und den übrigen linguistischen Beschreibungsebenen: Der Schriftge-

⁴ Die Annahme einer medienneutralen Syntax geht weit über die Wortgrenze hinaus. So besteht ein Satzanfang auch ohne Großschreibung. Und auch die Struktureinheiten, die durch syntaktische Interpunktionszeichen in der Schrift angezeigt werden und durch Phrasierung in der geschriebenen Sprache, sind medienneutral.

brauch ist sehr viel empfänglicher für Eingriffe in die kodifizierte Norm. Wenn die Norm geändert wird (wie das beispielsweise 1996 der Fall war), dann ändert sich auch der Gebrauch und damit die Graphematik. Das unterscheidet die Graphematik von der Phonologie: Die Orthoepie hat nicht annähernd dieselbe Wirkung auf das Gesprochene wie die Orthographie auf das Geschriebene. Das muss stets mit bedacht werden, wenn wir versuchen, das Schriftsystem zu beschreiben.

2. Die Aufgabe der Graphematik ist es, die Einheiten und Relationen der graphematischen Ebene sichtbar zu machen. Diese Einheiten und Relationen sind prinzipiell von zweierlei Art:

Graphematische Einheiten und Relationen sind solche, die ohne Rückgriff auf die übrigen linguistischen Strukturebenen – also autonom – ermittelt werden können. Die Schrift (zumindest die Druckschrift) ‚zerfällt‘ beispielsweise fast automatisch in die Einheiten Buchstaben und Wörter; sie ist gleichsam „vorsegmentiert“ (Kohrt 1985b: 430). Auch andere Einheiten lassen sich graphematisch-autonom bestimmen (siehe Kap. 3). Rein graphematische Regularitäten betreffen z. B. die minimale Wortlänge und Fragen der Graphotaktik: Welche Buchstaben treten verdoppelt auf, welche nicht? Welche kommen nur, welche nicht an bestimmten Positionen vor? Es handelt sich hier um eine bewusst naive Herangehensweise: Es wird so getan, als ob nur die Schrift gegeben sei, nicht aber das Sprachsystem, dessen Teil sie ist (vgl. die oft analoge Behandlung der Phonologie im amerikanischen Strukturalismus).

Für jede dieser Einheiten und Relationen ist weiterhin von Interesse, ob sie auf außergraphematische Einheiten und Relationen reduzierbar ist oder nicht; die nicht-reduzierbaren sind *genuin graphematische Einheiten und Relationen*. Der Buchstabe ist eine solche Einheit: Er ist durch eine rein graphematische Analyse identifizierbar, gleichzeitig aber nicht vollständig auf bspw. das Phonem reduzierbar. Die theoretische Möglichkeit solcher genuin graphematischen Einheiten macht die autonome Analyse notwendig (vgl. Eisenberg 1988): Wenn es sein könnte, dass Grapheme Einheiten ‚eigener Art‘ sind, dann darf eine Analyse nicht bereits davon ausgehen, dass sie phonologisch determiniert sind.

Phonographische und morphographische Einheiten und Relationen sind demgegenüber solche, die auf der Basis von phonologischen, morphologischen oder syntaktischen Einheiten und Regularitäten definiert sind. So ist beispielsweise von Interesse, wie Phoneme in der Schrift kodiert werden. Auch wenn die resultierende Einheit graphematisch heterogen ist – das Phonem /o/ kann graphematisch z. B. als <o>, <oo> oder <oh> realisiert werden –, ist die Zusammenfassung dieser Elemente zu einer Einheit legitim (z. B. der des „Phonographems“ wie bei

Nerius (Hg.) 2007).⁵ Die determinierende nicht-graphematische Einheit ist dabei selbstverständlich nicht auf das Phonem beschränkt: Auch die hierarchisch niedrigere Ebene der phonologischen Merkmale kann in ihrem graphematischen Reflex untersucht werden, ebenso wie die hierarchisch höheren Ebenen der Silbe und des Fußes. Und auch morphologische Einheiten jeder Komplexität sind Gegenstand möglicher Untersuchungen – Flexions- und Derivationsaffixe, freie und gebundene Stämme, komplexe Wörter.

Insgesamt ergibt sich damit ein etwas anderer Zugang, als graphematische Formen zu motivieren, wie es z. B. Nerius (Hg.) (2007) tut, oder graphematische Formen in phonologische Formen zu überführen, wie es für Neef (2005) zentral ist.

3. Maßgeblich ist die Verknüpfung der Teilstrukturen auf der Wortebene.

Das Wort ist die zentrale Ebene in repräsentationellen Modellen wie Jackendoff (1997, 2002): „A word [...] serves as part of the linkage between the multiple structures“ (Jackendoff 2002: 131). Die Wortebene ist die Ebene des sprachlichen Zeichens nach de Saussure (Aronoff 1976), das gilt für Teilstrukturen unterhalb der Wortebene nicht im gleichen Maße. Diese zentrale Einsicht der wort- oder lexembasierten Morphologie (vgl. z. B. Aronoff 1976; Anderson 1992; Blevins 2013) kann gewinnbringend auf die Graphematik übertragen werden: Wenn graphematische und phonologische Strukturen auf Wortebene verknüpft werden, dann sind Korrespondenzen auf Segmentebene – ähnlich den Morphemen – Generalisierungen über Teilstrukturen. Diese Phonem-Graphem-Korrespondenzen haben einen abgeleiteten Status.⁶

An einem Beispiel lässt sich die Konsequenz dieser Einsicht verdeutlichen. So diskutiert Venezky (2004: 147) Morphemkonstanz im Englischen und kommt zu dem Ergebnis, dass für bestimmte Mengen von Wörtern eine Konstanzschreibung aus phonologischen Gründen nicht infrage komme. Es gebe bspw. keine Stammkonstanz bei <decide> und <decision> (also genau nicht *<decidion> oder *<decise>), weil keine graphematische Einheit sowohl mit /d/ als auch mit /ʒ/ korrespondieren könne.

Verfolgt man eine wortbasierte Graphematik, ist diese Aussage tautologisch. Die Bedingung dafür, dass wir eine bestimmte Korrespondenz ansetzen, ist das

⁵ Das gilt freilich nur so lange, wie der Bezug zur Phonologie transparent gemacht wird. Rein graphematisch kann eine Einheit <o, oo, oh> nicht motiviert werden.

⁶ Auch in der Lesedidaktik, in der Phonem-Graphem-Korrespondenzen lange Zeit zentral waren, rücken zunehmend morphologische und silbische Einheiten in den Fokus (vgl. z. B. Bredel/Noack/Plag 2013).

regelmäßige Vorkommen im Schriftsystem. Ein Graphem, das /d/ und /ʒ/ entsprechen würde, würde aufgrund der prominenten Stelle an der Morphemgrenze automatisch regelmäßig vorkommen. Ein weiteres Beispiel: Die Korrespondenz ⟨t⟩ – /ʃ/ wie in *ignition* funktioniert auch nur vor ⟨ion⟩ und ⟨ious⟩. Würden die Stämme hier nicht konstant verschriftet werden, gäbe es auch die phonographische Korrespondenz nicht. Beide Aussagen sind Beschreibungen *eines* Sachverhalts, zwei Seiten einer Medaille.

Eine andere Situation liegt vor, wenn die phonographischen Bezüge eines *einzelnen* Wortes untersucht werden. Hier kann die Konsistenz der Phonem-Graphem-Korrespondenzen sehr wohl vor dem Hintergrund aller solcher Korrespondenzen (als Generalisierungen über die Menge aller Wörter) überprüft werden. Für *Ball* ist im Deutschen etwa *⟨Pall⟩ keine mögliche Schreibung, denn die Korrespondenz /b/ – ⟨p⟩ kommt im Deutschen auch ansonsten nicht vor.

Daher soll den Phonem-Graphem-Korrespondenzen nicht die Relevanz abgesprochen werden; es gibt ja gerade im Deutschen sehr stabile Korrespondenzen auf der Segmentebene. Aus der zentralen Rolle, die der Wortebene in dieser Arbeit zugesprochen wird, folgt aber deutlich, dass diese Korrespondenzen nicht unabhängig existieren.

Ähnliches gilt für morphologische Konstituenten unterhalb der Wortebene wie Stämme und Affixe. Diese Begriffe werden in der vorliegenden Arbeit verwendet, sie sind in gewisser Weise sogar zentral (vgl. den Titel der Arbeit). Das ändert epistemologisch nichts an ihrem abgeleiteten Status. Gegeben sind Wörter, und über Ähnlichkeiten und Differenzen in Form und Bedeutung können wir lexikalische und grammatische Teilstrukturen ermitteln – eben Stämme und Affixe.

Theoretisch werden die Regularitäten auf den verschiedenen Ebenen deklarativ als Beschränkungen modelliert, wie das etwa (neben vielen anderen) Culicover/Jackendoff (2005: 15) für die Syntax machen. Die Alternative – ein derivatonelles Modell, in dem etwa die ‚Regeln‘ zur morphologischen Schreibung auf phonographischen Schreibungen operieren – macht zu starke Vorannahmen und bringt potenziell (die aus der Phonologie bekannten) Probleme zur Motivation der Regelordnung mit sich.

1.4 Terminologie

Neben diesen grundsätzlichen Annahmen werden folgende terminologische Festlegungen getroffen:

1. Die Korrespondenzen zwischen graphematischer und phonologischer Ebene werden als *phonographische* Korrespondenzen bezeichnet, und zwar unabhängig von a) der Richtung der Implikation sowie b) der Hierarchieebene der

Korrespondenz. Für die Bezeichnung der Richtung bieten sich die Begriffe *Leserichtung* und *Schreibrichtung* an;⁷ die Ebene kann zusätzlich spezifiziert werden (*subsegmental*, *segmental*, *silbisch*). Die Korrespondenz [+plosiv] → ⟨+langer Kopf⟩ beschreibt also eine subsegmental-phonographische Beziehung in der Schreibrichtung (Plosive werden im Deutschen mit Buchstaben mit langem Kopf verschriftet, vgl. Primus 2004, 2006); ⟨p⟩ → /p/ beschreibt eine segmental-phonographische Beziehung in der Leserichtung; /d/ → ⟨d⟩ beschreibt eine silbisch-phonographische Beziehung in der Schreibrichtung (z. B. die zweite Silbe von *Nadel*).

2. Dabei wird zwischen zwei Arten von phonographischen Korrespondenzen unterschieden, *globalen* und *primären*. Globale Korrespondenzen sind Korrespondenzen, die über alle Wörter ermittelt werden, gewissermaßen blind für die Bezüge zu den anderen Ebenen. So ergibt sich aus Wörtern wie ⟨Kind⟩, ⟨Hand⟩, ⟨wild⟩ im Deutschen beispielsweise die Korrespondenz /t#/ → ⟨d⟩. Dass dieses Graphem nur auftritt, wenn im Paradigma Formen mit stamminaalem /d/ vorkommen, ist hier irrelevant. Diese Korrespondenzen sind in gewisser Weise naiv; es sind kaum Annahmen über das Schriftsystem und das Zusammenwirken der verschiedenen Ebenen notwendig. Es ist diese Art Korrespondenzen, die Neefs (2005) „graphematischen Lösungsraum“ konstituieren. Sie geben bereits einen ersten Hinweis darauf, wie flach oder tief das jeweilige Schriftsystem ist (siehe 1.2): Je eindeutiger die Korrespondenzen sind, desto flacher ist das Schriftsystem. Die globalen Korrespondenzen können als Basis dienen, um diejenigen Korrespondenzen zu ermitteln, die rein phonographisch sind – das sind all jene (systematisch vorkommenden) Korrespondenzen, die nicht morphologisch oder rein graphematisch explizierbar sind, etwa /p#/ → ⟨p⟩. Sie werden als primäre phonographische Korrespondenzen bezeichnet. Diese primären Korrespondenzen stehen am Ende der Analyse eines Schriftsystems. In diesem Sinne sind bspw. Eisenbergs primäre Korrespondenzen zu verstehen, vgl. Eisenberg (2013).
3. Die Korrespondenzen zwischen der graphematischen und der morphologischen Ebene werden als *morphographische Korrespondenzen* bezeichnet. Hier sind zwei Parameter von Interesse, die *Einheitlichkeit*, mit der Morpheme verschriftet werden, und die *Eindeutigkeit*, mit der Schreibungen wiederum Morpheme repräsentieren. Eine morphographische Korrespondenz ist *einheitlich*, wenn es für ein gegebenes Morphem genau eine Schrei-

⁷ Die Begriffswahl ist etwas unglücklich, weil sie suggeriert, dass Lesen und Schreiben nur im Überführen von Schrift in gesprochene Sprache (und andersherum) besteht. Schuld ist der Mangel an besseren intuitiv zugänglichen Alternativen.

bung gibt. Das ist z. B. der Fall bei *Kind*: Dieses Morphem wird – unabhängig von seiner Umgebung – immer als ‚Kind‘ verschriftet. Bei englisch *swim* oder deutsch *Apfel* ist das anders: Diese Morpheme haben abhängig vom Kontext zwei graphematische Formen (‚swim‘, ‚swimming‘; ‚Apfel‘, ‚Äpfel‘). Eine morphographische Korrespondenz ist *eindeutig*, wenn eine gegebene Schreibung auf genau ein Morphem verweist. Die englische Schreibung ‚rain‘ ist bspw. eindeutig: Sie verweist immer auf das Morphem *rain*. Das ist anders bei deutsch ‚Bank‘, das zwei distinkte Morpheme (Geldinstitut und Sitzmöbel) repräsentiert.

4. Analog werden die Bezüge zwischen phonologischer und morphologischer Ebene als *morphophonologische Korrespondenzen* bezeichnet, und die Korrespondenzen können ebenfalls auf ihre Einheitlichkeit und Eindeutigkeit geprüft werden.
5. Grapheme und graphematisches Material allgemein werden in spitzen Klammern gesetzt (also z. B. ‚a‘, ‚ß‘, ‚Bild‘). Buchstaben und Buchstabenkombinationen werden zwischen senkrechten Strichen gesetzt (also z. B. |a|, |q|, |sch|, vgl. Fuhrhop/Buchmann 2009). Auf diese Weise ist es möglich, den konzeptuellen Unterschied auszudrücken, der demjenigen zwischen phonologischen und phonetischen Einheiten in der Phonologie entspricht. Wenn keine Verwechslungsgefahr besteht bzw. es nicht um die Unterscheidung zwischen emischen und etischen Einheiten geht, werden schriftliche Einheiten mit spitzen Klammern ausgezeichnet.
6. Die rein graphematischen Einheiten und Relationen werden als *graphemisch* bezeichnet. Damit wird ein Vorschlag von Eisenberg (1983) aufgenommen. Notwendig ist der Begriff, weil die verbreiteten Begriffe ‚Graphematik‘ und ‚graphematisch‘ sowohl die Ebene der Schrift an sich als auch ihre Schnittstellen bezeichnen. Der Begriff ist synonym zur teilweise verwendeten Alternative ‚innergraphematisch‘ (Primus 2003, 2006; Fuhrhop/Barghorn 2012) zu lesen.
7. Die zentrale morphologische Ebene ist das Lexem. Lexeme sind abstrakte Verknüpfungen von graphematischen, phonologischen, (morpho-)syntaktischen und semantischen (Teil-)Strukturen (siehe oben). Mit Aronoff (1994: 39 ff.) bezeichnet der Begriff *Stamm* die graphematische oder phonologische Formseite des Lexems. Beides ist aber nicht miteinander gleichzusetzen, vor allem, weil viele Lexeme über mehrere Stämme verfügen, die man mit Fuhrhop (1998: 22 ff.) als Flexions-, Derivations- und Kompositionsstammformen bezeichnen kann (vgl. z. B. *Tag*, *täglich*, *Tagesform*, *Tagelohn*).

1.5 Aufbau

Im folgenden zweiten Kapitel wird die Datengrundlage motiviert und diskutiert. Das dritte Kapitel widmet sich dem graphematischen Aufbau von Morphemen und bildet den Schwerpunkt der Arbeit. Im vierten Kapitel stehen die morphographischen Korrespondenzen im Mittelpunkt. Die Arbeit schließt im fünften Kapitel mit einer Zusammenfassung und Diskussion der Ergebnisse.

2 Datengrundlage

2.1 CELEX

In dieser Arbeit soll die Wortschreibung im Englischen und im Deutschen untersucht und verglichen werden. Für die Untersuchung bietet sich die lexikalische Datenbank CELEX (Baayen/Piepenbrock/Gulikers 1995) als Datengrundlage an, und zwar aus zwei Gründen. Zum einen gibt es in CELEX Daten zum Englischen und zum Deutschen, und zwar in vergleichbarem Umfang (das niederländische Teilkorpus, das ebenfalls enthalten ist, ist für die Zwecke dieser Arbeit irrelevant und wird daher vernachlässigt). Zum anderen verknüpft CELEX graphematische Informationen zu den Lexemen mit phonologisch-segmentalen, phonologisch-silbenstrukturellen, morphologischen und morphosyntaktischen Informationen sowie mit Frequenzangaben. Hypothesen über Abhängigkeiten im Schriftsystem lassen sich auf dieser Grundlage relativ leicht prüfen: Wie häufig werden bestimmte Morpheme einheitlich geschrieben, wie häufig gibt es Variation? Wie häufig verweisen Schreibungen andersherum eindeutig auf bestimmte Morpheme?

Diese Vorteile überwiegen klar die Nachteile von CELEX: Neben einer z. T. uneinheitlichen morphologischen Analyse (siehe 2.2) ist das vor allem die Tatsache, dass das Korpus zwischen dem Kern- und dem Fremdwortschatz keinen Unterschied macht. Eine solche Differenzierung wäre wünschenswert, denn bei den Phonem-Graphem-Korrespondenzen und in der Derivationsmorphologie verhalten sich (zumindest im Deutschen) der native und der nicht-native Teil des Wortschatzes sehr unterschiedlich (vgl. z. B. Eisenberg 2011). Dass CELEX auf eine Auszeichnung hinsichtlich lexikalischer Strata verzichtet, ist bedauerlich, hat aber gute Gründe: Eine konsistente Unterscheidung mag einfach zu skizzieren sein (vgl. z. B. Albrow 1972), sie ist aber schwer zu operationalisieren. Es gibt m. W. keine vergleichbare lexikalische Datenbank, in der solche Informationen kodiert wären.

Die pragmatische Herangehensweise an dieses Problem ist, bei den einzelnen Untersuchungen ein offenes Auge für die Struktur der Daten zu haben. Was heißt das? Nehmen wir als Beispiel die Graphemdefinition. Hier ist es relevant, wie häufig ein potenzielles Graphem allein den Anfangsrand von Wörtern besetzen kann (so wird zumindest weiter unten argumentiert). Für |c| gibt es tatsächlich vier Wörter, in denen es alleine den Anfangsrand besetzt, also direkt von einem Vokalbuchstaben gefolgt wird: *Code*, *Cello*, *Campus*, *Camp* (siehe 3.2.2.2). Alle vier Wörter sind Fremdwörter (und zwar relativ unabhängig davon, ob man einen synchronen oder einen diachronen Fremdwortbegriff anlegt). Das bedeu-

tet: Im nativen Bereich kommt |c| nicht alleine im Anfangsrand vor, im Fremdwortbereich schon. Differenzierungen dieser Art sind möglich und werden an den entsprechenden Stellen vorgenommen.

Da sich Kompilierung und Annotierung der englischen und der deutschen Datenbank leicht unterscheiden, werden beide Datenbanken im Folgenden kurz vorgestellt. Die englische Datenbank enthält Informationen zu 52.447 Lemmata. Sie beruht vor allem auf zwei Wörterbüchern, dem Oxford Advanced Learner's Dictionary (Hornby 1974) mit 41.000 Lemmata und dem Longman Dictionary of Contemporary English (Procter 1978) mit 53.000 Lemmata. Zusätzlich wurden weitere Lemmata mit aufgenommen: „Other lemmata were added to enable morphological decomposition of the basic set of lemmata“ (Baayen/Piepenbrock/Gulikers 1995: Readme.txt-Datei). Um welche Lemmata es sich genau handelt und aufgrund welcher Kriterien sie ausgewählt wurden, bleibt unklar.

Die deutsche Datenbank enthält Informationen zu 51.728 Lemmata. Sie beruht im Gegensatz zur englischen (und zur niederländischen) Datenbank nicht auf einem gedruckten Wörterbuch, sondern auf den lexikalischen Datenbanken Bonnlex (Institut für Kommunikationswissenschaft und Phonetik, Bonn; vgl. Lenders 2013: 992), Molex (Institut für Deutsche Sprache, Mannheim; vgl. Kolvenbach 1980) und dem Noetic Circle Services German spelling lexicon (MIT). Wie im englischen Teil wurden auch hier nicht näher spezifizierte Lemmata ergänzt, um die morphologische Dekomposition zu automatisieren. Im zweiten Release von CELEX (das hier verwendet wird) sind darüber hinaus „about 1.000“ (Baayen/Piepenbrock/Gulikers 1995: Readme.txt-Datei) weitere, ebenfalls nicht näher spezifizierte Lemmata ergänzt worden.

CELEX ist 1995 erschienen und verwendet die traditionelle Rechtschreibung. Die Unterschiede betreffen größtenteils die s-Schreibungen – in CELEX erscheinen *Kuss* und *Ross* bspw. als ⟨Kuß⟩ und ⟨Roß⟩. Hinzu kommen einige mehr oder weniger isolierte Änderungen; diese betreffen die Wörter *Föhn*, *Känguru*, *Karamell*, *Mopp*, *rau*, *Stepp*, *Tollpatsch*, *behände*, *Gämse*, *Gräuel* und *Stängel*. Die entsprechenden Wörter wurden in der Datenbank geändert, sodass sie orthographisch auf dem neusten Stand ist.

2.2 CELEX-Teilkorpus einfacher Stämme

Neben den vollständigen CELEX-Korpora werden für bestimmte Fragestellungen auch Teilkorpora verwendet. Das betrifft zum einen die Untersuchung des graphematischen Aufbaus von Morphemen (Kap. 3). Hier wird eine Liste der Stämme und Affixe in beiden Sprachen benötigt. Die Liste der Affixe wird aus jeweils einschlägigen Grammatiken kompiliert (siehe 2.3); die Stämme werden aus CELEX

extrahiert. Wir beschränken uns zunächst auf einfache Flexionsstammformen – also auf morphologisch einfache Wörter wie deutsch *Haus* oder englisch *see* sowie deutsche Verbstämme wie *stell[en]*. Derivationsstammformen (gebundene einfache Stämme, die nur mit einem Derivationsuffix auftreten) wie deutsch *akzept-* oder englisch *electr-* werden ausgeklammert.

In CELEX sind morphologisch einfache Wörter als solche ausgezeichnet („MorphStatus“ = M). Problematisch ist allerdings, dass sich die Kategorisierung in beiden Sprachen unterscheidet. Offiziell gilt in beiden Sprachen zunächst der Grundsatz „If a stem contains at least one stem plus at least one other stem or affix, then it is said to be morphologically complex“ (Baayen/Piepenbrock/Gulikers 1995; German Linguistic Guide: 54; English Linguistic Guide: 47). Doch was heißt das für gebundene Stämme? Enthalten Wörter wie deutsch *demonstrieren* oder englisch *electrify* einen Stamm und ein Affix im Sinne von CELEX? Diese Frage wird im deutschen Korpus anders beantwortet als im englischen. Im englischen Korpus werden vergleichbare Fälle wie *durable* oder *social* als morphologisch ‚obskur‘ annotiert („MorphStatus“ = O: „[...] an analysis seems possible, but cannot be fully explained“, English Linguistic Guide: 49). Diese Kategorie ist auch für Konfixkomposita wie *neurology* einschlägig („neo-classical compounds“, Baayen/Piepenbrock/Gulikers 1995; English Linguistic Guide: 39).

Im deutschen Teilkorpus existiert diese Kategorie nicht. Die entsprechenden Fälle werden stattdessen meist als morphologisch einfache Wörter klassifiziert.⁸ Um die Korpora der beiden Sprachen vergleichen zu können, müssen Einträge wie *artikulieren* oder *Fotograf* daher im deutschen Teilkorpus identifiziert werden und in einem zweiten Schritt entfernt werden.

Die Identifikation geschieht mit der folgenden Prozedur:

- (1) Morphologisch komplex ist ein Wort, wenn es aus mindestens zwei Teilen besteht, von denen jedes auch frei oder in anderen Kombinationen als Morphem identifizierbar ist.

⁸ Interessanterweise stünde mit „MorphStatus“ = U („undetermined“) eine Kategorie für solche Fälle zur Verfügung – zumindest wird sie laut Handbuch für die strukturell ähnlichen Konfixkomposita verwendet: „Some stems use classical affixes, which don’t behave like normal German affixes (*Aerogramm* for example) [...] In all such cases the morphological status code is U[...]“ (Baayen/Piepenbrock/Gulikers 1995; German Linguistic Guide: 55)

Die Beispiele in (2) zeigen, wie die Prozedur zu verstehen ist:

- (2a) *Kongress* ist komplex wegen Konvent und Regress.
- (2b) *monochrom* ist komplex wegen monoton und polychrom
- (2c) *intelligent* ist komplex wegen Intelligenz und solvent

Diese Methode der Morphemidentifikation ist seit dem amerikanischen Strukturalismus gut etabliert (vgl. z. B. Nidas ‚Principle 6‘, 1949: 58 f.). Der einzige Unterschied betrifft die sog. unikalenen Morpheme wie *Schorn* in *Schornstein*, die Nida (1949) und andere als Morphem klassifizieren. Diese Morpheme können mit der skizzierten Methode in (1) nicht identifiziert werden – eben weil sie nur einmal vorkommen, und zwar in genau der Kombination, die gerade untersucht wird. Das ist intendiert. Wenn wir Nida (1949: 58 f.) folgen und *Schorn* als Morphem identifizieren würden, weil *Stein* auch frei oder in anderen Kombinationen vorkommt, dann müsste dieselbe Analyse auch für Wörter wie *fertig* oder *Hammer* gelten. Sowohl *-ig* als auch *-er* kommen in anderen Kombinationen vor; *fert* und *Hamm* wären damit Morpheme.⁹

Das ist erst einmal ein interessantes Ergebnis, das in manchen Fällen wohl auch die Intuition von Sprechern abbildet: Psycholinguistisch kann gezeigt werden, dass auch morphologisch einfache Wörter wie *corner* in einem gewissen Sinne als kompositionell wahrgenommen werden (für einen Überblick über Untersuchungen dieses Phänomens des ‚affix stripping‘ vgl. Aronoff/Berg/Heyer 2016). Ideal wäre eine Klassifikation, die auf dem Output von Wortbildungsregeln beruht: Ein gegebenes Wort enthält ein Affix, wenn es sich so verhält wie Wortbildungsprodukte dieses Affixes (so z. B. Aronoff 1976). *Hammer* enthält das Suffix *-er*, weil es wie viele *-er*-Bildungen ein Nomen instrumenti ist; *magic* enthält das Suffix *-ic*, weil es sich so verhält wie *-ic*-Bildungen: Es ist ein Adjektiv.

Das geschieht aus zwei Gründen nicht. Zum einen ist für viele Affixe unklar, wie genau diese Wohlgeformtheitsbeschränkungen formuliert werden sollten. Was ist die spezifische Leistung von *-ig*? Die Klärung dieser Fragen würde zu weit vom Ziel dieser Arbeit wegführen. Zum anderen sind die entsprechenden Lexeme im englischen Teilkorpus ebenfalls als morphologisch einfach annotiert. Diese

⁹ Nida (1949: 60, Fn. 53) schließt Fälle wie *hammer*, *ladder* und *otter* aus: „This *-er* is not semantically relatable to the agentive *-er* in *dancer*, *player*, *runner*, and *worker*.“ Zumindest im Deutschen ist die Situation etwas komplexer: Das Derivationsuffix *-er* kann durchaus Nomina instrumenti bilden, vgl. z. B. *Bohrer*. Das *-er* in *Hammer* kann semantisch auf das Derivationsuffix *-er* bezogen werden, und *Hamm* wäre ein unikales Morphem.

Parallelbehandlung ist der Hauptgrund für den Ausschluss unikalener Morpheme aus der Identifikationsprozedur in (1).

Als morphologisch einfach annotiert sind im deutschen Teilkorpus 6.531 Wörter, im englischen 7.398 Wörter. Die Anwendung der Prozedur in (1) führt im deutschen Teilkorpus zur Entfernung von 552 morphologisch komplexen Einträgen (z. B. *Anglophilie, Entstalinisierung, Harmonie, intelligent, kurieren, kostenaufwändig, Angelsachse, Bauernfang, Karfreitag, Kohlensäure, Reinemachefrau, Zitrusfrucht, Mikrofon, monochrom, metamorph, homophil*).

Morphologisch komplex sind im deutschen Teilkorpus außerdem Partikelverben wie *abflauen*. Das Kriterium ist hier nicht das Vorkommen von *flauen* in anderen Kombinationen, sondern die syntaktische Trennbarkeit in Verbzweitstellung (z. B. in *Der Wind flaute endlich ab*). Diese Trennbarkeit spricht dagegen, dass wir es mit einem einfachen Stamm zu tun haben. Partikelverben sind in CELEX annotiert („separable stems“, Baayen/Piepenbrock/Gulikers 1995; German Linguistic Guide: 56); auf diese Weise werden 58 Wörter ausgeschlossen (z. B. *aufrauen, ausweiden, einschieben*).

Ebenfalls aus dem deutschen Teilkorpus entfernt werden 16 Fälle von syntaktischer Konversion (i. S. v. Erben 2006: 31) wie *Angedenken, ausgefuchst, durchtrieben*. In all diesen Fällen wechselt nicht nur der Stamm, sondern auch das Flexionsmorphem des Ausgangswortes die Wortart.

Entfernt werden außerdem die Einträge für einzelne Buchstaben (*A, B, C ...*), die in beiden Teilkorpora enthalten sind, sowie Abkürzungen (das betrifft zwei Einträge im englischen Teilkorpus, *ABC* und *AA*). Sowohl Buchstaben als auch Abkürzungen sind keine einfachen Stämme im klassischen Sinne – sie sind bspw. nicht im selben Maße Basis von Ableitungen. Ein letzter Ausschluss betrifft die Einträge mit Akzent im deutschen (drei Wörter: *Café, Negligé, Piqué*) und Apostroph im englischen Teilkorpus (neun Wörter wie *o’er, ma’am*). Die Wörter mit Akzent im Deutschen sind randständig und beeinflussen die Ergebnisse besonders der syntagmatischen Analysen in Abschnitt 3.1 unverhältnismäßig. Die Wörter mit Apostroph im Englischen sind (mit Ausnahme von *Hallowe’en*) graphematische Varianten anderer, bereits im Korpus enthaltener Wörter (*over, Madam*).

Die Verben im deutschen Teilkorpus enthalten neben dem Stamm noch das Infinitivsuffix. Da es in Abschnitt 3.1 um den graphemischen Aufbau von einfachen Stämmen geht, wird das Suffix getilgt. Außerdem werden Groß- in Kleinbuchstaben umgewandelt, weil es zunächst um Buchstaben als abstrakte sprachliche Einheiten geht.

Im deutschen Teilkorpus verbleiben damit 5.862 Einträge, im englischen Teilkorpus sind es 7.361 Einträge. Als letzter Schritt werden die Duplikate getilgt, sodass nur distinkte Einträge übrig bleiben. Das Teilkorpus *einfacher deutscher*

Stämme enthält damit 5.485 Einträge. Das Teilkorpus *einfacher englischer Stämme* enthält 7.004 Einträge.

2.3 Korpus Affixe

Für die Untersuchung des graphemischen Aufbaus von Affixen wurden für das Deutsche die 119 Affixe aus Duden (2016: 702, 719, 734 f., 762 f., 773) zusammengetragen.¹⁰ Für das Englische wurden die 123 Affixe aus Huddleston/Pullum (Hg.) (2002: 1677 ff.) in eine Liste überführt. Zum Teil haben diese Affixe Varianten (z. B. *-heit/-keit/-igkeit*, *-ance/-ence*); diese Allomorphe haben jeweils einen eigenen Eintrag in der Liste.

Wenn wir – wie bei den Stämmen – diese Affixe als Types ansetzen, ergibt sich ein Problem: Einige Affixe kommen nur in wenigen Umgebungen vor. Unter den deutschen Lemmata in CELEX kommt das Suffix *-erich* etwa nur dreimal vor (*Wegerich*, *Weiderich*, *Wüterich*). Es kombiniert also nur sehr eingeschränkt. Wenn wir es mit in die Liste aufnahmen, wäre es ein Type und hätte damit denselben Stellenwert wie bspw. die hochfrequenten Affixe *ver-*, *-ung* oder *-lich*. Um das zu verhindern, wird für jedes Affix in beiden Listen ermittelt, wie häufig es in der Analyse der unmittelbaren Konstituenten in der CELEX-Lemmadatenbank vorkommt; Affixe, die weniger als zehnmal auftreten, werden aus der Liste ausgeschlossen. Das betrifft für das Deutsche 54 Affixe, für das Englische 45 Affixe.

Die Grenze von zehn Vorkommen ist natürlich willkürlich, und mit ihr ist das Problem nicht gelöst – es gibt neben den genannten hochfrequenten immer noch seltene Affixe, die nur in zehn, 20 oder 30 Wörtern vorkommen. Das Problem ist aber zumindest etwas entschärft.

Darüber hinaus sind in beiden Listen Affixe (genauer: Suffixe), die rein formal als Folgen von zwei Suffixen analysiert werden können – auch wenn sich diese Analyse im Fall des konkreten Wortbildungsprodukts verbietet, weil es keine freie Form mit nur einem Affix gibt. Es geht um Fälle wie deutsch *-igkeit* oder englisch *-ency*. Es gibt weder **nettig* (zu *Nettigkeit*) noch **sufficiency* (zu *sufficiency*), und doch verhalten sich *-igkeit* und *-ency* wie Folgen von Suffixen (*-ig + -keit* bzw. *-ence + -y*): Beide Teile existieren auch unabhängig von der Sequenz (vgl. *durstig*, *existence*), und der erste Teil passt kategorial zu den Wortbildungs-

¹⁰ Verbparkeln wie *auf* in *aufmachen* sind in der Liste nicht enthalten. Die Partikelverbbildung unterscheidet sich von den übrigen Wortbildungsmustern dadurch, dass ihre Produkte syntaktisch und morphologisch trennbar sind (*ich mache auf*, *aufgemacht*), vgl. z. B. Fleischer/Barz (2012: 91 f.).

regeln des zweiten Teils: *-ig* bildet Adjektive, und *-keit* operiert auf Adjektiven; *-ence* bildet Substantive, und *-y* operiert (unter anderem) auf Substantiven. Ausgehend von diesen Überlegungen werden im deutschen Korpus sieben Affixe ausgeschlossen, im englischen acht.

Eine letzte Modifikation betrifft die Flexionssuffixe, die noch ergänzt werden. Damit umfasst die Liste der deutschen Affixe 95 Einträge, die der englischen 93 Einträge (die vollständigen Listen finden sich in Anhang A). Die Listen werden mit Informationen zur Länge, zur graphematischen CV-Struktur sowie zur Besetzung der Silbenkonstituenten angereichert (basierend auf den Ergebnissen der Untersuchungen in 3.1.2).

2.4 CELEX-Teilkorpus homophoner Stämme

Für die Untersuchung der Eindeutigkeit von Stämmen wird ein Korpus von homophonen Stämmen benötigt, um festzustellen, wie häufig diese Stämme graphematisch differenziert werden und wie häufig nicht. Wie viele Paare vom Typ <Saite>/<Seite> gibt es, und wie viele vom Typ <Ton> (Klang)/<Ton> (Sediment)? Grundlage sind jeweils die CELEX-Korpora einfacher Stämme (siehe oben 2.2). Es ergeben sich hier zwei Probleme aus der Struktur von CELEX:

1. Homonyme im engeren Sinne haben in CELEX nur einen Eintrag. Damit eine phonologische Form zwei Einträge erhält, muss eine der folgenden Bedingungen erfüllt sein (vgl. Burnage 1995: 15 ff., 20 ff.):

Die phonologische Form ...

- a. ... hat zwei graphematische Formen (*peek/peak*);
- b. ... gehört zu zwei Wortarten (*arm.A/Arm.N*);
- c. ... hat zwei Flexionsparadigmen (*Bank – Bänke/Banken*);
- d. ... gehört zu zwei Genera (*Kiefer.MASK/Kiefer.FEM*)

Das bedeutet im Umkehrschluss, dass Paare wie deutsch *Hahn* oder englisch *date* (,Dattel‘/,Datum‘) – ,echte‘ Homonyme also – jeweils nur einen Eintrag in den Korpora haben. Es handelt sich hier aber um Fälle, die unbedingt ins Korpus homophoner Stämme gehören: Potenziell könnten *Hahn* und *date* eben heterographisch verschriftet werden (z. B. *⟨Haan⟩, *⟨dait⟩), und wenn sie nicht im Korpus enthalten sind, verzerrt das z. B. die Aussagen zu den Anteilen graphematisch differenzierter Stämme. Das ist das erste Problem.

2. Das zweite Problem ergibt sich ebenfalls aus den oben angegebenen Kriterien: Konversionen von Stämmen erhalten in CELEX regelmäßig zwei Einträge (*water.N/water.V*; *asterisk.N/asterisk.V*). Besonders im Englischen sind Konversionen praktisch allgegenwärtig, und wenn man ihre Produkte als Homonyme klassifiziert, verwischt man gerade die entscheidende Charakte-

ristik von ‚echten‘ Homonymen wie *Schimmel*. Bei syntaktischen Konversionen handelt es sich ja im Prinzip um dasselbe Lexem, das lediglich die Wortart wechselt. Beide Stämme sind semantisch eng verwandt – auch wenn dieses Kriterium schwer zu operationalisieren ist (vgl. Lyons 1968). Die Tatsache, dass *water* als Substantiv so geschrieben wird wie als Verb, kann in diesem Sinne als Instanz von Stammkonstanz interpretiert werden.

Im Korpus fehlen also die Homonyme im engeren Sinne; gleichzeitig sind viele Konversionen enthalten, die nicht als zwei Lexeme gewertet werden sollen. Beide Probleme wurden mithilfe von Wörterbüchern behoben:

1. Um die fehlenden ‚echten‘ Homonyme nachzutragen, wurde ein zweisprachiges Wörterbuch verwendet, das ungefähr den Umfang von CELEX hat (Langenscheidt Taschenwörterbuch Deutsch–Englisch/Englisch–Deutsch, 60.000 Einträge, Langenscheidt 2002). Die Homonyme, die nicht in CELEX enthalten sind (also die Homonyme im engeren Sinne) wurden manuell gesucht und ins Teilkorpus übernommen. Warum wurde kein größeres Wörterbuch verwendet? In größeren Wörterbüchern sind wahrscheinlich mehr Homonyme im engeren Sinn enthalten, aber die Menge der graphematisch differenzierten Homophone ändert sich nicht – die können nicht manuell gesucht werden, sondern nur automatisch. Ein größeres Wörterbuch verzerrt also unter Umständen die Datenbasis: Die Grundgesamtheit, an der die Anzahl der graphematisch differenzierten Homophone gemessen wird, wächst. Die Datenbasis ist aber CELEX, und wir erweitern diese Datenbank behutsam um ‚echte‘ Homonyme.
2. Um Konversionen auszuschließen, wird ebenfalls ein Wörterbuch verwendet. Wenn die zwei CELEX-Einträge Teil *eines* Lexikoneintrags sind, handelt es sich nicht um Homonymie, sondern um Polysemie – zumindest nach dem verwendeten Wörterbuch. Hier wird ein wesentlich umfangreicheres Wörterbuch herangezogen (Langenscheidt Handwörterbuch Deutsch–Englisch/Englisch–Deutsch, 120.000 Einträge, Langenscheidt 2005), weil es wichtig ist, dass die CELEX-Einträge auch tatsächlich im Wörterbuch enthalten sind. Ein umfangreiches Lexikon ist hier (im Gegensatz zum ersten Punkt) nicht problematisch, weil nur CELEX-Einträge überprüft werden.

Auf diese Weise wird mit den oben genannten Problemen umgegangen. Sie werden so selbstverständlich nicht gelöst; die notorisch schwierige Frage der Abgrenzung von Homonymie und Polysemie (vgl. z. B. Lipka 1986) wird lediglich auf die verwendeten Wörterbücher abgewälzt. Da es sich aber um zweisprachige Wörterbücher handelt, ist zumindest zu vermuten, dass bei der Kompilierung in beiden Sprachen ähnliche Kriterien ähnlich konsistent angewendet werden.

Das Teilkorpus homophoner Stämme umfasst also a) die automatisch ermittelten homophonen Einträge im Teilkorpus morphologisch einfacher Stämme, b) abzüglich der Konversionen, c) erweitert um die Homonyme im engeren Sinne. Nachdem im deutschen Teilkorpus außerdem vier offensichtliche Konversionen (wie *Kapsel* – *kapsel[n]*), zwei veraltete Einträge (*Gesell* und *Zapf*) sowie eine graphematische Variante (‹Ski›/‹Schi›) manuell entfernt wurden, umfasst das *deutsche Teilkorpus homophoner Stämme 248 Formen*, das *englische 618 Formen*. Dieses Korpus homophoner Stämme wurde abschließend danach annotiert, ob die Homophonie graphematisch aufgelöst wird, ob es sich also um Heterographen handelt (wie *Saite/Seite*) oder um Homonyme (wie *Mutter₁/Mutter₂*). Auf diese Weise sind Aussagen möglich, welcher Anteil der Homophone graphematisch differenziert wird. Problematisch ist in diesem Zusammenhang, dass die homophonen Formen teilweise mehr als zwei Lexeme repräsentieren, z. B. englisch *pair/pear/pare*, und teilweise werden nicht alle dieser Lexeme graphematisch unterschieden, z. B. *mal/Mal/mal[en]/mahl[en]/Mahl*. Hier wird wie folgt verfahren: Sobald Homophonie aufgelöst wird, wird der entsprechende Eintrag als heterographisch annotiert. Das gilt auch, wenn daneben Homographien bestehen bleiben wie im Fall von *mal/Mal/mal[en]/mahl[en]/Mahl*.

3 Der graphemische Aufbau von Morphemen

In diesem Kapitel soll der graphemische Aufbau von Morphemen untersucht werden. Es werden diejenigen Beschränkungen entwickelt, die für die graphemische Form von Morphemen gelten. Wie sind Stämme und Affixe graphemisch aufgebaut? Welche Buchstaben und Grapheme sind häufig, welche selten? Wie kombinieren sie? Welche Kombinationen sind häufig, welche sind selten? Diese Fragen werden im Folgenden bearbeitet.

Warum Morpheme und nicht Wörter? Weil Morpheme die relevantere Einheit für graphemische Strukturbeschränkungen sind. Wörter bestehen potenziell aus mehreren Morphemen, und das ist problematisch für die Beschreibung der Graphotaktik: Über Morphemgrenzen hinweg gelten praktisch keine Beschränkungen für die Abfolge von Graphemen. Wir finden im Prinzip alle Graphempaare, auch solche, die innerhalb von Morphemen nicht vorkommen, etwa <zm> (wie in englisch <quizmaster>) oder <hf> (wie in deutsch <Fischfutter>). Diese Paare sind aber zufällig und nicht systematisch; sie beruhen auf der Verkettung von graphemischen Morphemformen. Behandelt man beide gleich, wird der Blick verstellt auf die eigentlichen, systematischen Beschränkungen für den graphematischen Aufbau von Morphemen – etwa die Beschränkung, dass <zm> und <hf> keine möglichen Kombinationen innerhalb von Morphemen sind. Diese Argumentation beruht auf der Annahme, dass die Schrift im Deutschen und Englischen konkatenativ ist. Dass es hier tatsächlich nur relativ wenige (und vor allem systematisch beschreibbare) Ausnahmen gibt, wird in Kapitel 4 gezeigt, wenn es um die graphematische Variation von Stämmen und Affixen geht.

Zum Aufbau: Zunächst wird festgestellt, welche Buchstaben und Grapheme als Inventar zur Verfügung stehen (3.1). Anschließend wird der graphemische Aufbau einfacher Stämme untersucht (3.2). Hier geht es um die Graphotaktik – und zwar zunächst global, also ohne Bezug zu Strukturpositionen (3.2.1) und dann an Strukturpositionen wie Silbenkern und -rand gebunden (3.2.2). Danach geht es um Bedingungen für minimale Stämme (3.2.3). Mithilfe der Minimalpaaranalyse wird dann die funktionale Last ermittelt, der die Grapheme an den verschiedenen Strukturpositionen unterliegen (3.2.4). Für die Affixe (3.3) wird aufgrund der schmalen Datenbasis nur die silbenstrukturelle Graphotaktik analysiert. Das Kapitel schließt mit einer Zusammenfassung der wichtigsten Beobachtungen (3.4).

3.1 Inventar: Buchstaben und Grapheme

Es soll hier um den *graphemischen* Aufbau von englischen und deutschen Morphemen gehen. Damit rücken Fragen der (typo-)graphischen Variation in den Hintergrund. Buchstaben im Rahmen dieser Arbeit sind daher, Primus (2004) und anderen folgend, bereits abstrakte linguistische Einheiten. Mehr noch: Untersucht werden, ebenfalls Primus (2004, 2006) und anderen folgend, nur die Kleinbuchstaben des Alphabets.

Nach diesen Vorbemerkungen ist das Inventar der Buchstaben einigermaßen trivial: Englisch benutzt das lateinische Alphabet ohne Sonderzeichen (3a), Deutsch erweitert die Menge um die vier Buchstaben |ß|, |ä|, |ö| und |ü| (3b).¹¹

(3a) Buchstabeninventar Englisch:

a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z

(3b) Buchstabeninventar Deutsch:¹²

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

Interessanter ist die Frage nach der relativen Häufigkeit der Buchstaben in beiden Sprachen. Dazu werden in den beiden monomorphematischen Teilkorpora der beiden Sprachen einfach die Buchstaben gezählt. Die Verteilungen sowie die Unterschiede zwischen den beiden Sprachen lassen sich mithilfe von Balkendiagrammen gut darstellen. In der folgenden Abbildung sind die Buchstaben nach ihren relativen Häufigkeiten im Deutschen angeordnet (für eine Auswertung nach Tokens im Deutschen vgl. z. B. Best 2005). Die Tabelle mit den absoluten Zahlen findet sich in Anhang B.

¹¹ Zur Abgrenzung der Buchstaben von den Ziffern, Sonderzeichen und Wortzeichen siehe Berg/Primus/Wagner (2016: 347 f.).

¹² Anders als in Primus (2004, 2006) und Berg/Primus/Wagner (2016) werden |y|, |ä|, |ö| und |ü| hier als Buchstaben gewertet. Es stimmt, dass |y| nicht im Kernwortschatz auftritt (vgl. z. B. Eisenberg 2013a: 290) – dieser Kernwortschatz ist aber im CELEX-Korpus nicht ohne Weiteres identifizierbar. |ä|, |ö| und |ü| bestehen jeweils aus einem Buchstaben und Trema, und das Trema lässt sich phonographisch motivieren; graphemisch ist ein Ausschluss schwerer motivierbar.

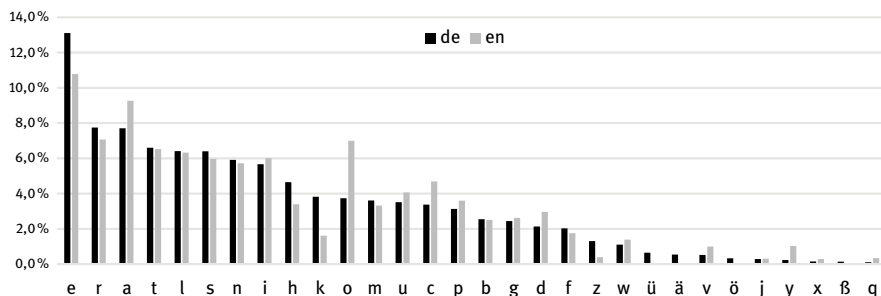


Abb. 1: Relative Anteile der Buchstaben im deutschen und englischen Teilkorpus einfacher Stämme an der Summe aller Buchstaben im jeweiligen Korpus. N(de): 30.772, N(en): 37.519.

Auf den ersten Blick verhält sich die Frequenz der Buchstaben in beiden Sprachen bemerkenswert ähnlich. Der relative Anteil der meisten Buchstaben weicht in beiden Sprachen nur minimal voneinander ab. Das spiegelt sich statistisch in einem sehr hohen Korrelationskoeffizienten wider (Pearsons $r = 0,93$, $p < 0,001$). Das ist überraschend, weil es sich um Sprachen handelt, die zwar eng miteinander verwandt sind, die aber über unterschiedliche segmentalphonologische Inventare und prosodische Muster verfügen sowie unterschiedliche Traditionen der Verschriftung haben. Warum sollten bspw. die relativen Anteile von |t|, |l|, |s|, |n|, |i| usw. in beiden Sprachen fast identisch sein?

Die Abweichungen der relativen Anteile der einzelnen Buchstaben sind demgegenüber vergleichsweise gering, aber zum Teil charakteristisch. Hier lassen sich bereits (im Vorgriff auf folgende Abschnitte) einige interessante Beobachtungen machen:

- Bei den Vokalen sind |a| und besonders |o| im Englischen häufiger als im Deutschen. Das könnte bei |o| damit zusammenhängen, dass der Buchstabe im Englischen – anders als im Deutschen – als Teil von Schreibdiphthongen auftreten kann (vgl. Abschn. 3.2.2.6). Andersherum ist |e| im Deutschen deutlich häufiger als im Englischen.
- Im Deutschen ist |h| häufiger als im Englischen. Das lässt sich auf die wesentlich höhere Frequenz von |sch| im Deutschen zurückführen – diese Verbindung kommt im monomorphematischen Korpus des Deutschen 518-mal vor, im Englischen nur elfmal. Im englischen Korpus sind dafür die Verbindungen *th*, *sh*, *wh* etc. frequenter als im deutschen.
- Im Englischen ist |c| häufiger als im Deutschen, dafür ist im Deutschen |k| häufiger. Beide sind phonographisch miteinander verwandt, und die Verteilung von |c| und |k| lässt sich unter Umständen als typologischer Parameter im Vergleich von Schriftsprachen verwenden (vgl. Fuhrhop/Buchmann/Berg 2011).

- Im Deutschen ist |z| häufiger als im Englischen, und auch hier bestehen phonographische Überschneidungen zu |c| (vgl. Fuhrhop/Buchmann/Berg 2011). Der höhere Anteil von |c| im Englischen ist damit eine Konsequenz des niedrigeren Anteils von |k| und |z|.
- Im Englischen ist |y| frequenter als im Deutschen; das lässt sich mit der Funktionalisierung als Marker von Wortenden erklären (vgl. Berg 2013).
- Für die spezifisch deutschen Buchstaben |ß|, |ä|, |ö| und |ü| ergeben sich keine starken Abweichungen, weil sie im Deutschen zu den seltensten Buchstaben gehören.

Eine interessante Frage ist, wie sich das Buchstabeninventar anderer Sprachen verhält, wenn es auf der Grundlage vergleichbarer lexikalischer Datenbanken erhoben wird. Gibt es Sprachen, die sich wesentlich vom Deutschen und Englischen unterscheiden, oder lassen sich typologische Konstanten von Alphabetschriften aufstellen?

So einfach die Frage nach dem Buchstabeninventar ist, so kompliziert ist diejenige nach dem Grapheminventar. Gegenstand ausgiebiger Diskussion ist, welche Kriterien zur Graphemdefinition herangezogen werden können oder sollen. Die Diskussion ist äußerst umfangreich und kann an dieser Stelle nicht einmal ansatzweise vollständig referiert werden; der interessierte Leser sei besonders auf Kohrt (1985b) verwiesen. Sie bewegt sich – etwas vereinfachend zusammengefasst – zwischen einer Definition, die auf das Phonemsystem Bezug nimmt (Grapheme sind Verschriftungen von Phonemen) und einer autonomen Definition, die allerdings methodisch auf die etablierten Methoden der Phonologie zurückgreift.

Teilweise lässt sich die kontrovers geführte Diskussion darauf zurückführen, dass die Kriterien, die zur jeweiligen Bestimmung des Grapheminventars führen, entweder nicht explizit gemacht werden (was seltener geschieht) oder nicht motiviert werden (was durchaus häufig passiert). Die Definition von Graphemen ist kein Selbstzweck. Man kann durchaus weit kommen ohne sie und stattdessen bspw. mit Buchstaben und Buchstabenverbindungen operieren, wie Neef (2005) das tut. Sobald aber Regularitäten einfacher erfassbar sind, bietet es sich an, gleichsam als terminologische Abkürzung von Graphemen zu sprechen. Solch einfachere Regeln können sich auf Phonem-Graphem-Beziehungen beziehen; dann kann es sinnvoll sein, |sch| und |ng| als Graphem festzulegen, damit die phonographischen Bezüge möglichst einfach sind. Es kann sich aber auch – und das ist in der vorliegenden Arbeit der Fall – darum handeln, dass im Sprachvergleich die funktionalen Einheiten der rein graphematischen Ebene verglichen werden sollen. Auch hier ist es sinnvoll, Grapheme zu definieren.

Verbreitet ist in jüngerer Zeit eine Definition wie die folgende, die auf die Parallelität zur Definition des Phonems abhebt, ohne direkt auf die Phonologie Bezug zu nehmen (vgl. so oder ähnlich z. B. Eisenberg 1988; Günther 1988: 77 u. a.):

- (4) Grapheme sind die kleinsten distinktiven Einheiten der Schrift.

Diese Definition ist auf den ersten Blick einleuchtend; sie macht wie in der Phonologie die Minimalpaaranalyse zum entscheidenden Kriterium. Unterscheiden sich zwei Wörter einer gegebenen Sprache nur hinsichtlich eines Segments, so ist dieses ein Graphem – es hat distinktive Kraft in der betreffenden Sprache.

Es werden allerdings bei näherer Betrachtung zwei Probleme sichtbar. Zum einen ist der Term ‚Einheit‘ (teilweise auch ‚Segment‘) nicht explizit genug: Als Einheit (bzw. Segment) können in der Schrift auch Teile von Buchstaben angenommen werden, da sie – anders als Merkmale in der Phonologie – durchaus voneinander trennbar sind. Legt man die obige Definition auf diese Weise aus, führt das Minimalpaar *lös[en]/los[en]* dazu, dass das Trema als kleinste distinktive Einheit angenommen werden muss. Analog dazu belegen Minimalpaare wie *lall[en]/hall[en]*, dass der ‚Spazierstock‘ des |h| ebenfalls distinktiv ist (ganz ähnlich z. B. Harweg 1971: 79 ff.).¹³ Soll dieses Problem vermieden werden, so muss die Definition in (4) entsprechend modifiziert werden, beispielsweise hinsichtlich einer Mindestgröße des Terms ‚Einheit‘:

- (4′) Grapheme sind die kleinsten distinktiven Einheiten der Schrift, die mindestens Buchstabengröße besitzen.

Das andere Problem ist, dass das paradigmatische Kriterium in bestimmten Fällen zu unintuitiven Ergebnissen führt. So gibt es zwei Belege für Minimalpaare, bei denen |q| durch einen anderen Buchstaben ersetzt wird: *quell/Duell* und *euer/quer*. Auf dieser Grundlage ist |q| ein Graphem. Dass auf |q| im nativen Wortschatz immer |u| folgt (vgl. Günther 1988: 82) und dass dies ein Argument gegen den Graphemstatus von |q| sein könnte, ist so nicht zu erfassen – dafür müssen andere Kriterien als das paradigmatische herangezogen werden. Sehr konsequent in dieser Hinsicht sind Zifonun/Hoffmann/Strecker (1997: 257), die tatsächlich nur <q> als Graphem ansetzen.

¹³ Dieses Argument funktioniert natürlich nur, wenn die Alternation mit Null-Elementen als Teil der Minimalpaarmethode akzeptiert wird (zur Diskussion darüber in der Phonologie vgl. Wagner 1982: 80 ff.).

Viele andere Autoren hingegen führen weitere graphematische Kriterien ein. Es geht hier – wie angedeutet – vor allem um die Erfassung der Tatsache, dass |q| nie ohne folgendes |u| auftritt – also um syntagmatische Information.¹⁴ Im Sinne von Shannon (1948) hat |u| hier keinerlei Informationsgehalt, der Buchstabe ist vorhersagbar. Beide Buchstaben hängen offenbar enger zusammen als andere Kombinationen und werden von den betreffenden Autoren auf dieser Grundlage als ein Graphem behandelt (vgl. z. B. Günther 1988; Fuhrhop/Peters 2013). Es ergibt sich als weitere Modifikation das folgende Kriterium (das allerdings selten explizit gemacht wird):

- (4“) Grapheme sind die kleinsten distinktiven Einheiten der Schrift, die mindestens Buchstabengröße besitzen und frei kombinieren.

Der Buchstabe |q| ist in seiner Kombinierbarkeit extrem eingeschränkt, daher wird ihm – trotz Vorhandenseins von Minimalpaaren – der Graphemstatus abgesprochen. Während dieser Fall relativ eindeutig ist, gibt es in vergleichbaren Fällen Abstufungen:

1. |c| tritt (zumindest im nativen deutschen Wortschatz) nur vor |h| und |k| auf. Auch |c| kombiniert also nicht frei; im Unterschied zu |q| ist der folgende Buchstabe allerdings nicht vollständig determiniert, sondern einer von zwei möglichen.
2. |j| tritt im Deutschen nur vor |a|, |e|, |o|, |u|, |ä|, |ö| und |ü| auf. Hier ist die Kombinierbarkeit nochmals etwas weniger eingeschränkt – aber von einer freien Kombination dennoch weit entfernt.

Möchte man |ch| und |ck| als Grapheme annehmen, so ist allein aufgrund der syntagmatischen Verteilung nicht ersichtlich, warum |ja|, |je|, |jo| etc. nicht auch Grapheme sein sollten – es sei denn, man setzt einen willkürlichen Schnitt und definiert, dass Grapheme das nachfolgende Segment nicht vollständig (|q|) oder auf eine von zwei Möglichkeiten (|c|) einschränken. Das ist natürlich eine unbefriedigende Ad-hoc-Modifikation.

Für einen umfassenderen Ansatz müssen offenbar auch Informationen über die silbenstrukturelle Position berücksichtigt werden: |c| kann nicht ohne folgendes |h| oder |k| alleine im Silbenanfangs- oder Endrand auftreten (zumindest

14 Warum diese Art von Information einen anderen Status als die paradigmatische haben sollte, wie es Zifonun/Hoffmann/Strecker (1997: 258) und Enderle (2005: 216 f.) fordern, ist nicht ersichtlich. Mit syntagmatischer Information argumentieren Strukturalisten schon lange, zumindest, was die Phonologie betrifft (vgl. z. B. Harris 1951: 59 ff.).

nicht im nativen Bereich; siehe auch Abschn. 3.2.2). Der Buchstabe |j| hingegen kann das durchaus, auch wenn die Verteilung der ihm folgenden Vokalbuchstaben leicht eingeschränkt ist (|i| kommt nur sehr selten vor). Das geschieht unter der Prämisse, dass die Silbe in der Schrift auch ohne Bezug zur Lautung ein sinnvolles Konzept ist. Diese Prämisse ist mit Primus (2003) gut motiviert. Primus argumentiert, dass die Silbe als Struktur verstanden werden kann, die zunächst unabhängig von der Phonologie ist und in den drei sprachlichen Realisationsformen Lautsprache, Schriftsprache und Gebärdensprache analog aufgebaut ist. Konstitutiv für diese abstrakte Einheit sind nach Primus a) eine bestimmte Alternationsstruktur zwischen b) unterschiedlich salienten Einheiten. Setzt man hier zunächst einen einzigen Unterschied in der Salienz an, ergeben sich zwei Gruppen von Elementen, die sich – abkürzend auch in der Schrift – Vokale und Konsonanten nennen lassen.

Wird die Silbe rein graphemisch betrachtet, so sind Konsonanten- und Vokalbuchstaben nötig, die ohne Bezug zur Phonologie definiert werden. Diese Definition wird in Abschnitt 3.2.1.1 rein graphemisch durchgeführt. Wenn es im Folgenden um Konsonanten und Vokale geht, dann geschieht das im Vorgriff auf die dort erzielte Definition.

Es ergibt sich also folgende Graphemdefinition (mit der sich nebenbei auch der Bezug auf eine segmentale Mindestgröße wie in (4'') erübrigt):

- (5) Grapheme sind die kleinsten distinktiven, silbenstrukturell autonomen Einheiten der geschriebenen Sprache.

Das bedeutet: Konstitutiv für den Graphembegriff nach (5) ist die Eigenschaft, autonom die silbenstrukturellen Positionen Anfangsrand und Endrand (für Konsonanten) bzw. Kern (für Vokale) besetzen zu können. Eine solche Definition ist anschlussfähig an nicht-lineare Ansätze (z. B. Primus 2010; Evertz 2014): Hier wird – in Anlehnung an die nicht-lineare Phonologie seit Nespor/Vogel (1986) – neben der Ebene der Buchstaben eine silbenstrukturell determinierte Graphemebene angenommen. Die Beziehungen zwischen Buchstaben- und Graphemebene können einfach oder komplex sein; einem Buchstaben können ein oder mehrere Grapheme entsprechen, und umgekehrt können einem Graphem ein oder mehrere Buchstaben entsprechen. Für die Zwecke dieser Arbeit – die ja dezidiert ‚unphonographisch‘ sind – kann auf die aufwendigere nicht-lineare Modellierung verzichtet werden.

Mit der Definition in (5) lassen sich minimale Einheiten (Buchstaben) auf ihren Graphemstatus hin überprüfen. Buchstabenverbindungen können nur dann ebenfalls geprüft werden, wenn einer der beteiligten Buchstaben nach (5) kein Graphem ist. In diesem Fall gibt es – wenn der betreffende Buchstabe die

Distribution seiner Nachbarn einschränkt – einen oder mehrere Kandidaten für komplexe Grapheme. Im nativen Wortschatz kommt |c| nicht allein im Anfangs- oder Endrand vor – es handelt sich also nicht um ein Graphem.

Unter dieser Definition kann ein komplexes Graphem nicht ausschließlich aus Graphemen bestehen. Auch wenn die Verbindung |ng| beispielsweise recht häufig ist und außerdem mit einem Phonem korrespondiert, kann sie nach (5) kein komplexes Graphem sein, da sowohl ⟨n⟩ als auch ⟨g⟩ bereits die kleinsten silbenstrukturell autonomen Einheiten sind.

Die freie Kombinierbarkeit, die Bestandteil der Definition in (4'') war, ist in (5) nicht enthalten. Das hat den Vorteil, dass (5) sehr gut operationalisierbar ist; freie Kombinierbarkeit ist, wie oben angedeutet wurde, graduell und nicht binär. Als Ausgangspunkt der Untersuchung eignet sich (5) also gut. Die Inventare, die sich ergeben, werden dann im Laufe dieses Kapitels auf der Basis von weiteren Verfahren ergänzt, die sich (in der einen oder anderen Form) alle auf die freie Kombinierbarkeit beziehen.¹⁵ So kann bspw. der Graphemstatus auch silbenstrukturell variieren – was im Anfangsrand ein Graphem ist, muss noch lange keins im Endrand sein. Außerdem kann es vorkommen, dass bestimmte Verbindungen von existierenden Graphemen sich kombinatorisch verhalten wie ein einfaches Graphem. Und schließlich ist es möglich, dass zwei Grapheme, die sich ansonsten silbenstrukturell autonom verhalten, eine feste Verbindung eingehen, die z. B. über die Minimalpaarmethode nicht auflösbar ist. All diese weiteren Verfahren – die fast alle klassisch-strukturalistische Ursprünge haben – werden im Verlauf des Kapitels an entsprechender Stelle aufgegriffen und diskutiert. Eine Zusammenfassung der Ergebnisse findet sich in Abschnitt 3.2.5.

Untersucht man nun – zunächst für das Deutsche – systematisch, ob die jeweils distinktiven Elemente auch silbenstrukturell autonom sind, so ergibt sich: Die meisten Buchstaben sind auch Grapheme. Das gilt nicht uneingeschränkt für |c|, |y| und |q|.

- |c| tritt nur in 20 Stämmen ohne folgendes |h| oder |k| auf, und alle diese Stämme sind Fremdwörter (z. B. *chic*, *Campus*, *Cello*, *Code*). Im nativen Bereich ist |c| eindeutig kein Graphem. Im nicht-nativen Bereich ist das nicht ganz so eindeutig; es gibt vereinzelt Minimalpaare, etwa *Camp/Tamp*. Der Graphemstatus von |c| steht also im nicht-nativen Bereich auf empirisch wackeligen Füßen.

¹⁵ Im Rahmen der globalen Graphotaktik (3.2.1) werden zunächst Buchstaben untersucht; erst ab der silbenstrukturellen Graphematik (3.2.2) spielen Grapheme eine Rolle. Damit wird der Definition in (5) Rechnung getragen, die auf silbenstrukturellen Informationen beruht.

- |y| tritt zwar häufiger auf (71-mal), aber auch hier sind es ausschließlich Fremdwörter (z. B. *Idyll, Lyra, Hobby*), vgl. auch Eisenberg (2013a: 290).
- |q| tritt ausschließlich vor |u| auf und ist daher kein Graphem.

Für |c| und |q| können häufige Buchstabenverbindungen geprüft werden. Bei |c| sind das die Verbindungen |ch| und |ck|, bei |q| die Verbindung |qu|. Es zeigt sich: Die beiden getesteten komplexen Grapheme sind tatsächlich distinktiv, und sie sind silbenstrukturell autonom (vgl. die Minimalpaare *Leiche/Leine, Stock/stoß[en], Qualle/Halle*).¹⁶

Es ergibt sich also als Grapheminventar des Deutschen:

(6) Grapheminventar deutsch:

einfach: ⟨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, c,¹⁷ y¹⁷⟩

komplex: ⟨ch,¹⁸ ck,¹⁸ qu⟩

Nach der Definition in (5) ist *sch* kein Graphem des Deutschen, weil das Minimalitätskriterium nicht erfüllt ist – ⟨s⟩ und ⟨ch⟩ sind bereits Grapheme. Damit wird ein notorisch schwieriger Klassifikationsfall umgangen, der im Folgenden dennoch kurz kommentiert werden soll. Zur Beantwortung der Frage, ob *sch* ein Graphem des Deutschen sei, wird oft andere als syntagmatische Information herangezogen, z. B. die Regeln zur Worttrennung am Zeilenende (so z. B. Augst (Hg.) 1985; Primus 2000: 33 f.). Am Zeilenende wird *sch* nicht getrennt (vgl. z. B. ⟨lö-[Zeilenende]schen, *⟨lös-[Zeilenende]chen⟩). Für Primus (2000) ist das ein Argument für den Graphemstatus von *sch*; schließlich handelt es sich hier um eine rein graphische Gliederung. Allerdings ist unklar, wie stark diese Evidenz zu gewichten ist; immerhin haben sich die präskriptiven Regeln zur Worttrennung lange auf die phonologische Form des betreffenden Wortes bezogen (vgl. Kohrt 1985a; Günther 1988: 84; Fuhrhop/Peters 2013: 206). Auf der anderen Seite findet sich in Gütherts (2005) diachronen Gebrauchsdaten aus dem 16.–18. Jahrhundert kein einziger Beleg für eine Trennung *s-ch*. Die Nicht-Trennung von *sch* ist also

¹⁶ Man könnte durchaus die Distinktivität abhängig von den Basispositionen (vgl. Basbøl/Wagner 1985) überprüfen. Dann hätte man für jeden Graphemkandidaten ein etwas genaueres Profil seiner Distinktivität. Da sich diese Untersuchung jedoch größtenteils mit der Ermittlung der syntagmatischen Verteilung deckt, die in Abschnitt 3.2.1 und 3.2.2 unternommen wird, soll an dieser Stelle darauf verzichtet werden. Es wird hier also nach dem etwas vereinfachenden Motto verfahren: „In einer silbenstrukturellen Position ein Graphem, immer ein Graphem.“

¹⁷ ⟨c⟩ und ⟨y⟩ sind nur im nicht-nativen Teil des Wortschatzes Grapheme.

¹⁸ ⟨ch⟩ und ⟨ck⟩ sind nur im nativen Teil des Wortschatzes Grapheme.

wohl natürlich gewachsen. Wie lässt sich die Situation bewerten? Das Verhalten von *sch* am Zeilenende deutet darauf hin, dass es eine graphemische Einheit ist – allerdings keine genuin graphemische Einheit im Sinne dieser Arbeit (vgl. oben 1.3), sondern eine graphemische Einheit, die phonographisch determiniert ist.

Im Englischen ergibt sich ein leicht abweichendes Grapheminventar. Vor allem ist |c| wesentlich besser im System verankert; es ist distinktiv und silbenstrukturell autonom (vgl. z. B. die Minimalpaare *cat/hat*, *cease/lease*, *cut/but*), also ein Graphem. Damit entfällt die Möglichkeit komplexer Grapheme wie *ch* und *ck* – die werden ja oben zumindest für den nativen Teil des Deutschen angesetzt.¹⁹ Lediglich |q| ist im Englischen nicht silbenstrukturell autonom. Wie im Deutschen füllt es diese Positionen nur zusammen mit |u| (vgl. die Minimalpaare *quiver/liver*, *quest/rest* etc.), damit ist <qu> das einzige komplexe Graphem.

(7) Grapheminventar englisch:

einfach: <a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, r, s, t, u, v, w, x, y, z>
 komplex: <qu>

Die potenziellen Grapheme <ch>, <gh>, <ph>, <rh>, <sh>, <th>, <wh> und <ng> bestehen jeweils aus Graphemen und kommen daher als eigenständige Grapheme nach der Definition (5) nicht in Frage; das Minimalitätskriterium schließt sie aus. Ansätze, die die genannten Graphemverbindungen als Grapheme klassifizieren, tun dies meist aus dezidiert phonographischer Perspektive und mit einem qualifizierenden Zusatz („compound graphemes“ bei Venezky 1970: 48; „single (but complex) grapheme“ bei Carney 1994: xxvii). Im Verlauf dieses Kapitels wird sich zeigen, dass viele dieser potenziellen Grapheme auch mit bestimmten graphemisch-kombinatorischen Kriterien zu erfassen sind (siehe 3.2.2.8).

Insgesamt decken sich die Inventare zu einem großen Teil mit den in der Literatur angenommenen, es gibt aber zu fast jedem Inventar kleinere Unterschiede. Das ist nicht überraschend, wenn man sich vor Augen führt, dass jedes Inventar auf zum Teil unterschiedlichen Definitionen beruht und unterschiedliche Zwecke verfolgt. Ich möchte in diesem Zusammenhang für etwas mehr Gelassenheit im

¹⁹ Diese Situation ist auf den ersten Blick unbefriedigend – ein und dieselbe Buchstabenverbindung, <ch>, ist in deutsch <Becher> ein Graphem, in englisch <chip> hingegen nicht. Sie ergibt sich aber notwendig aus der Definition in (5), die sich vor allem auf silbenstrukturelle Autonomie der einzelnen Bestandteile stützt. Und hier verhalten sich die beiden Schriftsysteme eben sehr unterschiedlich: Im Deutschen besetzt einfaches |c| nur in wenigen Fällen alleine den Silbenrand, während im Englischen Stämme wie *city* oder *face* Legion sind. |c| ist unterschiedlich autonom, deswegen haben Verbindungen mit |c| einen unterschiedlichen Status im Gesamtsystem.

Umgang mit Graphemdefinitionen plädieren. Grapheme sind zweckgebunden; es gibt sie nicht an und für sich. Es ist unsinnig, ‚das‘ Grapheminventar einer Sprache zu ermitteln (wie es etwa Rezec 2009: 14 ff. einfordert) – ein Umstand, auf den in der Phonologie Zelig Harris bereits 1951 hingewiesen hat.²⁰

3.2 Einfache Stämme

3.2.1 Globale Graphotaktik

Interessanter als das Inventar ist die Kombinatorik von Buchstaben und Graphemen. Bevor im nächsten Abschnitt etwas kleinteiliger der Aufbau von Silbenanfangs- und -endrändern sowie Silbenkernen ermittelt wird, geht es hier gewissermaßen global um die Möglichkeiten und Grenzen der Kombinatorik.

Die (autonome) Beschreibung der Graphotaktik gilt spätestens seit Eisenberg (1985: 124 f.) und Günther (1988: 77 f.) als Desiderat – auch wenn Eisenberg (1985: 126) betont, dass der Nutzen einer solchen Analyse ungewiss sei. Sie ist ein Desiderat, weil nur auf diese Weise ermittelt werden kann, in welcher Beziehung graphotaktische und phonotaktische Regularitäten stehen. Werden graphematische Einheiten und Regularitäten direkt aus phonologischen abgeleitet (so z. B. Garbe 1985), dann lassen sich genuin schriftsystematische Regularitäten (wenig überraschend) nicht finden. Aus einem ähnlichen Grund ist der Nutzen der Analyse a priori unklar: Es ist eben nicht sicher, wie viele der graphemischen Regularitäten mit phonologischen korrespondieren und wie viele rein graphemisch sind. Außerdem weist Eisenberg (1985: 127) darauf hin, dass graphotaktische Regularitäten keinen Wert an sich darstellen, sondern möglichst hinsichtlich ihrer Funktionalität interpretiert werden sollten.

In diesem Sinne liegt das Hauptinteresse im Folgenden auf denjenigen Regularitäten, die genuin graphemisch sind. Es geht in diesem Abschnitt nicht darum, die gesamte Graphotaktik abzubilden und zu erfassen; vieles ist phonographisch determiniert und daher eher uninteressant für die Zwecke dieser Arbeit. Vielmehr

²⁰ „At a time when phonemic operations were less frequently and less explicitly carried out, there was discussion as to what had to be done in order to arrive at ‘the phonemes’ and how one could discover ‘the phonemes’ of a language. [...] The phonemes resulted from a classification of complementary segmental elements; and this could be carried out in various ways. For a particular language, one particular arrangement may be more convenient, in terms of particular criteria, than other arrangements. The linguistic requisite is not that a particular arrangement be present, but that the criteria which determine the arrangement be explicit.“ (Harris 1951: 72, Fn. 28).

gehen wir etwas eklektisch vor und fokussieren vor allem potenziell graphemische Regularitäten.

3.2.1.1 Konsonanten und Vokale

Für die Beschreibung der Graphotaktik, vor allem der silbenstrukturellen Graphotaktik, sind Informationen über die Zugehörigkeit zu den Klassen der Vokal- und Konsonantenbuchstaben nötig. Anders können beispielsweise Silbenränder und -kerne nicht erkannt und analysiert werden. Diese Informationen werden auf graphemischer Grundlage gewonnen – es sollen nicht einfach die korrespondierenden phonologischen Kategorien übertragen werden.

Einen frühen Vorschlag in diese Richtung macht Augst (1985: 118 f.). Er ermittelt auf graphotaktischer Grundlage Konsonanten- und Vokalgrapheme, ohne Rückgriff auf die phonologischen Kategorien. Er nutzt dabei die charakteristische Verteilung von Geminaten wie <aa>, <ee> und <oo> bzw. <bb>, <dd>, <ff> usw.: Unmittelbar vor den Geminaten <aa>, <ee> und <oo> treten andere Buchstaben auf als vor den übrigen Geminaten; die Klassen sind komplementär verteilt. Mit diesem Ansatz kommt man sehr weit. Es ergeben sich aber zwei Probleme: Zum einen gibt es keine systematische Begründung für diese Verteilung. Warum sollten Buchstaben genau auf dieser Grundlage in Vokale und Konsonanten klassifiziert werden? Zum anderen gibt es Buchstaben, die nicht verdoppelt werden (im Deutschen z. B. |i|, |u|, |h|, |j|, |v|, |x| u. a.; im Englischen z. B. |a|, |i|, |u|, |h|, |j| u. a., weiter Abschn. 3.2.1.3). Für die Klassifikation dieser Buchstaben sind weitere Informationen nötig.

Ich möchte im Folgenden zeigen, wie diese Probleme heute (mehr als 30 Jahre nach Augsts Aufsatz) umgangen werden können, wie also a) die Klassenbildung systematisch motiviert werden kann und wie b) alle Buchstaben in einem Schritt klassifiziert werden können. Dazu werden – wie bei Augst (Hg.) (1985) – graphotaktische Informationen genutzt, allerdings nicht exemplarisch anhand *einer* Verteilung wie bei Augst, sondern exhaustiv über die monomorphematischen Korpora.

Die Grundidee ist: Konsonanten kommen vor allem neben Vokalen vor und umgekehrt, und diese Information wird für die Klassenbildung benutzt (auf derselben Idee beruht Sukhotins Algorithmus, vgl. Guy 1991). Wir bestimmen für jeden Buchstaben, wie häufig er vor und nach welchen anderen Buchstaben auftritt (erfasst werden also *Bigramme*, d. h. Paare von benachbarten Buchstaben). Das ist das charakteristische Profil dieses Buchstabens: ein Vektor mit n Dimensionen. Die Profile von Konsonanten und Vokalen, so ist zu vermuten, unterscheiden sich nun stark voneinander: Im Allgemeinen folgen auf Konsonanten eben Vokale und auf Vokale Konsonanten. Das lässt sich – wie oben bereits erwähnt –

mit sehr allgemeinen, modalitätsunabhängigen Prinzipien des Silbenaufbaus motivieren. Mit Primus (2003) sind Silben modalitätsübergreifend durch eine Alternationsstruktur gekennzeichnet: Einheiten aus mindestens zwei verschiedenen Kategorien alternieren; zusätzlich ist eine der beiden Klassen wahrnehmungspsychologisch prominenter/salienter.

Die Profile der Buchstaben müssen also auf dieser Grundlage in zwei Gruppen klassifiziert werden. Da die Profile sehr komplex sind (60 Dimensionen für das Deutsche, 52 Dimensionen für das Englische), scheidet eine manuelle Klassifizierung oder einfache Visualisierung aus. Stattdessen wird die statistische Klassifikationsmethode der hierarchischen Clusteranalyse verwendet. Ähnlichkeiten und Unähnlichkeiten zwischen den Buchstaben werden als sog. Dendrogramme dargestellt.²¹ Das entsprechende Dendrogramm für das Deutsche ist in Abbildung 2 wiedergegeben.²²

Wie ist diese Darstellung zu interpretieren? Das Dendrogramm gibt an, wie ähnlich sich die Elemente sind, die klassifiziert werden. Je dichter beisammen die (stets binären) Verzweigungen sind, desto ähnlicher sind die jeweiligen Elemente. Die Baumstruktur ist relevant, nicht die Reihenfolge, in der die Buchstaben in der Abbildung erscheinen – die zwei Äste jedes Knotens könnten auch andersherum angeordnet sein. Abbildung 2 ist also abstrakter, als sie auf den ersten Blick scheint.

Schauen wir uns einen konkreten Fall an, die Umlaute in Abbildung 2: Die Buchstaben |ä| und |ü| werden von einem gemeinsamen Knoten dominiert, das heißt, sie sind ähnlich. Die Verzweigung für |ö| befindet sich relativ nah an diesem Knoten; das bedeutet, dass |ö| den Buchstaben |ä| und |ü| ebenfalls ähnlich ist. Das gilt weniger für das Element, das am nächsthöheren Knoten hängt, |y|: Hier gibt der größere Abstand an, dass |y| nicht im selben Maß zur Gruppe |ä|, |ö|, |ü| gehört.

21 Altmann/Lehfeldt (1980: 278 ff.) analysieren das Phonemsystem des amerikanischen Englisch auf sehr ähnliche Weise. Da sie aber für jedes Phonem nur bestimmen, wie häufig es nach einem anderen Phonem auftritt, fällt das Ergebnis ihrer Analyse „kaum befriedigend“ (280) aus.

22 Datengrundlage sind die 5.485 Wörter des Teilkorpus einfacher Stämme (vgl. Abschn. 2.2). Aus den Wörtern dieser Liste wird zunächst eine Tabelle extrahiert, die für jeden Buchstaben angibt, wie häufig er vor und nach jedem Buchstaben auftritt. Die Berechnung der Dendrogramme erfolgt auf der Grundlage dieser ‚Buchstabenprofile‘ mit der Programmierumgebung R (www.r-project.org, Stand: 22. 11. 2018). Hierzu wird zunächst die Distanzmatrix für die euklidische Distanz zwischen den Buchstabenprofilen errechnet (Funktion *dist*). Dann wird mithilfe der hierarchischen Clusteranalyse (Funktion *hclust*) und der agglomerativen Methode „Ward.D2“ ein Objekt erzeugt, das sich grafisch darstellen lässt.

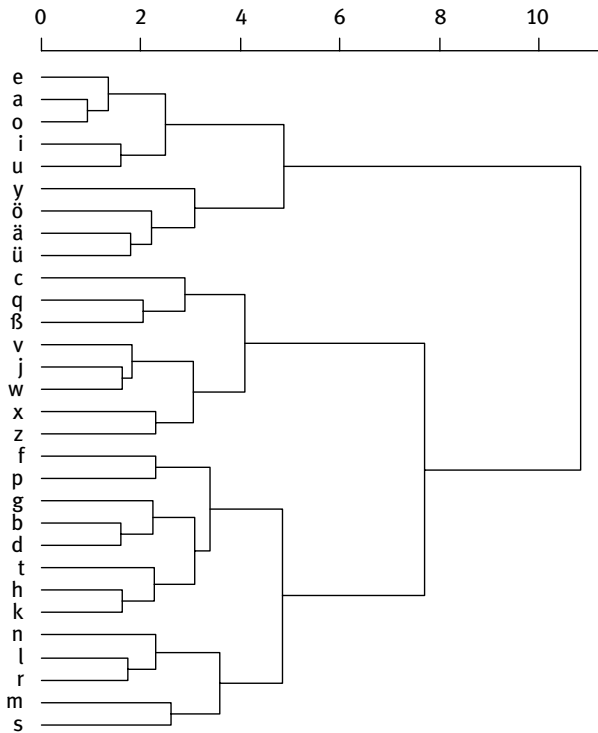


Abb. 2: Dendrogramm der deutschen Buchstaben auf der Basis ihrer syntagmatischen Verteilung.

Insgesamt zerfallen die Buchstaben in zwei Gruppen, die sich sehr deutlich voneinander unterscheiden. Nennen wir die kleinere Gruppe Vokale und die größere Konsonanten,²³ so erhalten wir für das Deutsche folgende Klassifizierung:

- (8) Vokal- und Konsonantenbuchstaben im Deutschen:
 - (8a) Vokale:
 - a, e, i, o, u, y, ä, ö, ü
 - (8b) Konsonanten:
 - b, c, d, f, g, h, j, k, l, m, n, p, q, r, s, t, v, w, x, z, ß

²³ Augst (1985: 119) weist darauf hin, dass diese phonologisch geprägten Begriffe „nicht glücklich“ sind für eine graphematische Beschreibung. Wenn sie hier dennoch verwendet werden, dann in Ermangelung besserer und gleichzeitig zugänglicher Alternativen.

Einzig die Klassifikation von |y| widerspricht der traditionellen phonographischen Klassifikation, wie sie bspw. in Schulbüchern zu finden ist. Sie ist aber sinnvoll, denn in den meisten Wörtern mit |y| verhält der Buchstabe sich wie andere Vokalbuchstaben: Er kombiniert vor allem mit Konsonantenbuchstaben. Im Korpus finden sich 52 Wörter, in denen |y| von Konsonantenbuchstaben umgeben ist oder die auf Konsonantenbuchstaben und |y| enden (z. B. die Belege in (9a), vgl. auch Berg 2012):

- (9a) Enzym, hybrid, Hymne; Baby, Hobby
 (9b) Yacht, Yoghurt; royal; Boy

Demgegenüber kommt |y| nur elfmal initial vor Vokalbuchstaben, wortmedial zwischen zwei Vokalen oder final nach einem Vokal vor (z. B. die Belege in (9b)). Die Kategorisierung als Vokalbuchstabe ist also gerechtfertigt.²⁴

Hier muss die Interpretation des Dendrogramms nicht aufhören. Die Vokalbuchstaben bestehen aus zwei Gruppen:

- (10a) a, e, i, o, u
 (10b) y, ä, ö, ü

Die Umlaute bilden – wie oben erwähnt – eine Gruppe, und zusammen mit |y| stehen sie den übrigen Vokalen gegenüber. Das ist dadurch motiviert, dass sowohl die Umlaute als auch |y| distributionell wesentlich beschränkter sind als die übrigen fünf Vokalbuchstaben; in der Mehrzahl der Fälle kombinieren sie bspw. nicht mit anderen Vokalen.

Die fünf Vokalbuchstaben wiederum sind wie folgt gegliedert: |i| und |u| bilden eine Gruppe und stehen |a|, |o| und |e| gegenüber. Diese Gliederung erfasst einige wesentliche Charakteristika der Diphthongschreibung im Deutschen. Wie Berg/Fuhrhop (2011) zeigen, sind |i| und |u| ‚gute‘ Zweitbestandteile von Schreibdiphthongen. |e| unterscheidet sich von |a| und |o| einerseits durch die wesentlich höhere Frequenz, andererseits durch die wesentlich flexiblere Distribution; in Schreibungen wie |ie| wird deutlich, „dass |e| in den Silbenkern drängt“ (Fuhrhop/Barghorn 2012: 144).

24 NB: Interessanterweise ist das korrespondierende Phonem in den Fällen (9b) ebenfalls vokalähnlich. Es wird als Halbvokal, Approximant oder Glide klassifiziert (vgl. Fuhrhop/Peters 2013: 31; Ladefoged/Maddieson 1996: 322). Auch phonographisch spricht also einiges für den Vokalstatus von |y|.

Auch bei den Konsonanten können zumindest die größeren Gruppen expliziert werden. So bestehen die Konsonanten etwa aus zwei großen Gruppen:

(11a) c, j, q, v, w, x, z, ß

(11b) b, d, f, g, h, k, l, m, n, p, r, s, t

Die erste Gruppe zeichnet sich dadurch aus, dass die Buchstaben (mit Ausnahme von |c|) die seltensten Buchstaben des Deutschen sind (vgl. Abb. 1) und dass sie alle distributionell eingeschränkt sind: Sie kombinieren nicht oder nur wenig mit anderen Konsonantenbuchstaben, wie weiter unten gezeigt wird. Die zweite Gruppe ist häufiger als die erste und kombiniert freier mit anderen Konsonantenbuchstaben. Innerhalb der Gruppe bilden |n|, |l| und |r| eine klar abgegrenzte Subgruppe; diese Buchstaben gehören (neben |t| und |s|) zu den häufigsten Buchstaben des Deutschen, und sie kombinieren zumindest im Endrand beinahe notorisch (vgl. Abschn. 3.2.2.3).

Im Englischen führt die hierarchische Clusteranalyse zu sehr vergleichbaren Ergebnissen (siehe Abb. 3).²⁵

Auch hier zerfällt die Menge der Buchstaben in zwei Gruppen; die Gliederung ist deckungsgleich mit derjenigen im Deutschen:

(12) Vokal- und Konsonantenbuchstaben im Englischen:

(12a) Vokale:

a, e, i, o, u, y

(12b) Konsonanten:

b, c, d, f, g, h, j, k, l, m, n, p, q, r, s, t, v, w, x, z

Wie im Deutschen wird |y| auch im Englischen als Vokalbuchstabe klassifiziert, und auch hier ist das gerechtfertigt, wenn man sich seine Verteilung ansieht. Medial zwischen zwei Konsonanten oder final nach einem Konsonanten kommt |y| 337-mal vor (Beispiele in (13a)). Initial vor einem Vokalbuchstaben, medial zwischen zwei Vokalbuchstaben oder final nach einem Vokalbuchstaben kommt |y| demgegenüber nur 170-mal vor; hier ist die finale Position (Typ *boy*) die frequenteste (Beispiele in (13b)).

(13a) baby, vinyl

(13b) annoy, coyote, honey, mayor, yard, yell

²⁵ Datengrundlage im Englischen sind die 7.004 Wörter des Teilkorpus einfacher englischer Stämme.

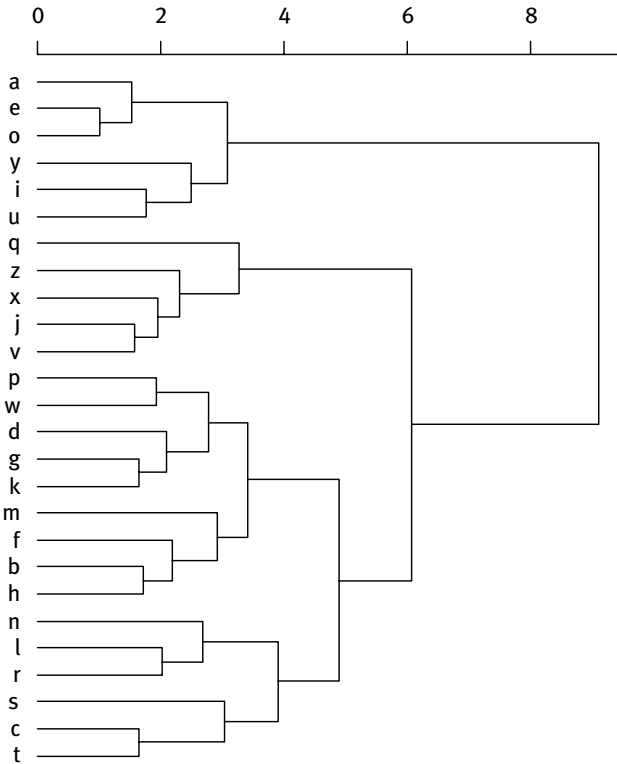


Abb. 3: Dendrogramm der englischen Buchstaben auf der Basis ihrer syntagmatischen Verteilung.

Schaut man sich das Dendrogramm etwas genauer an, entdeckt man auch im Englischen einige Gruppen, die im Laufe der Arbeit wieder auftauchen werden. Die Vokale etwa sind in zwei Hauptgruppen gegliedert:

(14a) a, e, o

(14b) i, u, y

Die Vokalbuchstaben in (14a) zeichnen sich dadurch aus, dass sie in Verbindungen mit anderen Vokalbuchstaben überwiegend an erster Stelle vorkommen. Diejenigen in (14b) treten praktisch ausschließlich an zweiter Position auf (z. B. |ai| in *pain*, |eu| in *feud*, |oy| in *boy*; weiter 3.2.2.6).

Und auch die Konsonanten bestehen aus zwei großen Gruppen. Die Gruppe in (15a) zeichnet sich – wie die entsprechende Gruppe im Deutschen – dadurch aus, dass sie die seltensten Buchstaben im englischen Korpus umfasst und dass

diese Buchstaben distributionell wesentlich eingeschränkter sind als diejenigen in der Gruppe (15b).

(15a) j, q, v, x, z

(15b) b, c, d, f, g, h, k, l, m, n, p, r, s, t, w

Die Kategorisierung mithilfe der hierarchischen Clusteranalyse liefert also für beide Sprachen stabile Ergebnisse. Diese Ergebnisse decken sich mit den Ergebnissen in Berg (2012).²⁶ Die dort verwendete Methode ist mit der hier verwendeten durchaus vergleichbar. Sie beruht aber auf der paradigmatischen Verteilung der Buchstaben, während hier die syntagmatische Verteilung genutzt wird. Grundlage in Berg (2012) ist die systematische Erfassung der Minimalpaare. Davon ausgehend werden ebenfalls Profile für die Buchstaben gebildet, die mithilfe der multidimensionalen Skalierung visualisiert werden.

Insgesamt kommen beide Methoden fast zu denselben Ergebnissen. Die hier verwendete Methode hat allerdings mindestens drei praktisch-methodische Vorteile: Zum einen ist sie einfacher. Anstatt Minimalpaare zu ermitteln – also Wörter, die sich in nur einem Element unterscheiden –, werden für jeden Buchstaben lediglich die Frequenzen der linken und rechten Kontexte erfasst. Daraus ergibt sich zweitens, dass wesentlich mehr Daten vorliegen, die als Basis der hierarchischen Clusteranalyse dienen. Während es im Deutschen bspw. auf der Basis der monomorphematischen CELEX-Lexeme 13.984 Minimalpaare gibt (Berg 2012: 44), sind es 140.562 linke und rechte Kontexte im Deutschen, die in der vorliegenden Arbeit ausgewertet wurden. Das wiederum führt drittens dazu, dass für die Klassifizierung aufgrund der syntagmatischen Verteilung weit weniger Daten notwendig sind. Das soll im folgenden Exkurs kurz gezeigt werden.

3.2.1.2 Exkurs: Skalierbarkeit der syntagmatischen Klassifizierung

Die vorgestellte Methode funktioniert (mit einer Einschränkung, siehe unten) auch mit einem Bruchteil der Datengrundlage. Um das zu zeigen, werden aus der englischen und deutschen Liste, die die Grundlage für die Berechnung der hierarchischen Clusteranalyse ist (siehe oben), jeweils 50 %, 25 % und 10 % der Einträge zufällig ausgewählt. Auf der Grundlage dieser Daten wird erneut eine Clusteranalyse durchgeführt. Das Ganze geschieht nicht einmal, sondern pro Anteil

²⁶ Im Unterschied zur vorliegenden Analyse wurde in Berg (2012) das englische |y| als Konsonantenbuchstabe klassifiziert. Das liegt an der unterschiedlichen Methode: In Berg (2012) wurde die paradigmatische Verteilung für die Klassifikation verwendet (siehe unten und 3.2.3).

(50 %, 25 %, 10 %) 1.000-mal. Wir erzeugen mit den zufällig ausgewählten Daten gewissermaßen Parallelwelten mit einem weniger umfangreichen Lexikon und überprüfen, wie die Clusteranalysen jeweils ausfallen (es handelt sich hier um eine Spielart der Monte-Carlo-Simulation vgl. z. B. Bortz/Schuster 2010: 113). Das Ergebnis lässt Rückschlüsse auf die Stabilität der Klassifikation zu: Wenn die Datengrundlage um 50 %, um 75 %, um 90 % verringert wird – wie häufig bekommen wir dann noch dasselbe Ergebnis wie bei der Analyse des kompletten Datensatzes? Tabelle 1 gibt an, bei wie vielen der jeweils 1.000 zufälligen Stichproben des Lexikons eine Klassifizierung herauskommt, die mit derjenigen übereinstimmt, die auf der Grundlage des gesamten Teilkorpus erzielt wurde (Abb. 2 und 3 oben). Abbildung 4 visualisiert die Daten als Liniendiagramm:

Tab. 1: Ergebnisse der Monte-Carlo-Simulation zur Klassifikation zufällig ausgewählter Teile der deutschen und englischen Korpora. Die Zahlen geben an, wie viele der 1.000 Durchläufe für selbige Klassifikation führen wie die Clusteranalyse des gesamten Teilkorpus.

	50 %	25 %	10 %
deutsch	925	731	251
englisch	504	444	241

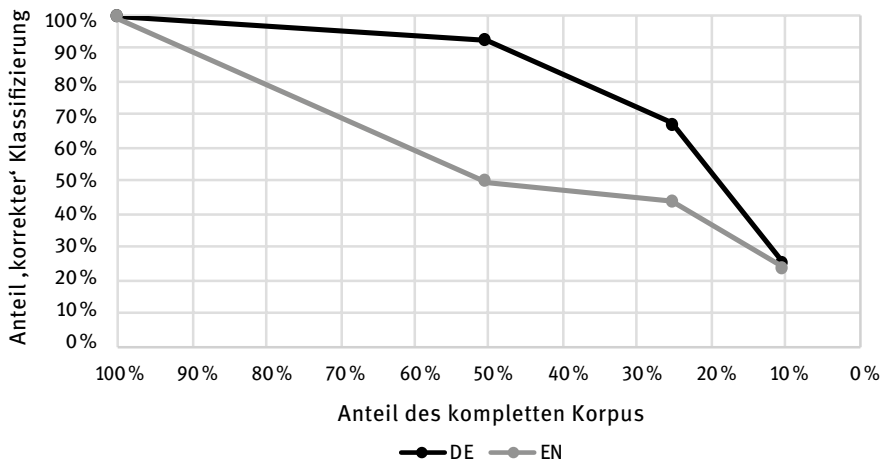


Abb. 4: Liniendiagramm der Ergebnisse der Monte-Carlo-Simulation zur Klassifikation zufällig ausgewählter Teile der deutschen und englischen Korpora. Die Verhältniszahlen geben an, wie viele der 1.000 Durchläufe zur selben Klassifikation führen wie die Clusteranalyse des gesamten Teilkorpus.

Die Klassifizierungen über den deutschen Teilkorpora sind anfangs (bei 50 % und 25 %) stabiler; bei 10 % der Korpusgröße konvergieren sie aber mit den Klassifizierungen der englischen Teilkorpora. Die Stabilität der Klassifizierung steigt deutlich, wenn wir |y| aus der Gruppe der Vokalbuchstaben herausnehmen, wie Tabelle 2 und Abbildung 5 zeigen:

Tab. 2: Ergebnisse der Monte-Carlo-Simulation zur Klassifikation zufällig ausgewählter Teile der deutschen und englischen Korpora. Die Zahlen geben an, wie viele der 1.000 Durchläufe zur selben Klassifikation der Buchstaben außer |y| führen wie die Clusteranalyse des gesamten Teilkorpus.

	50 %	25 %	10 %
deutsch ohne y	1.000	996	509
englisch ohne y	745	930	940

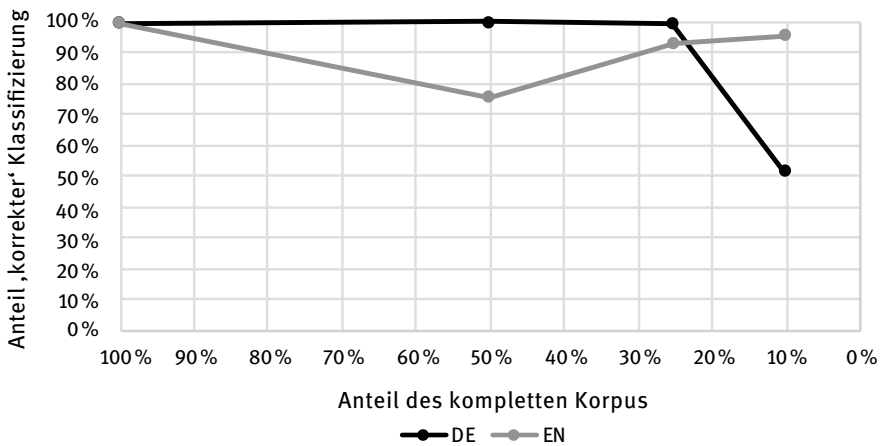


Abb. 5: Liniendiagramm der Ergebnisse der Monte-Carlo-Simulation zur Klassifikation zufällig ausgewählter Teile der deutschen und englischen Korpora. Die Verhältniszahlen geben an, wie viele der 1.000 Durchläufe zur selben Klassifikation (außer für |y|) führen wie die Clusteranalyse des gesamten Teilkorpus.

Ohne |y| werden die übrigen Vokalbuchstaben sehr viel zuverlässiger klassifiziert. Der Buchstabe |y| wird mit sinkender Datengrundlage also häufiger als Konsonantenbuchstabe klassifiziert und ist damit verantwortlich für die geringe Stabilität der Klassifizierung bei 10 % der Datenmenge in Abbildung 4 oben. Das ist

nicht überraschend, denn zum einen ist |y| in beiden Sprachen eben kein eindeutiger Vokalbuchstabe, wie oben gezeigt wurde; er wird auch in konsonantischen Kontexten verwendet, wenn auch weniger häufig. Zum anderen ist |y| in beiden Sprachen ein eher seltener Buchstabe. Im englischen Teilkorpus kommt er 389-mal vor, im deutschen sogar nur 73-mal. Wird nun das Korpus entsprechend verkleinert, ist es wahrscheinlich, dass in einigen Iterationen die konsonantische Verwendung von |y| überwiegt und der Buchstabe damit nicht als Vokalbuchstabe klassifiziert wird. Etwas Vergleichbares gilt für die Umlaute im deutschen Korpus: Sie sind zwar wesentlich häufiger als |y|, aber immer noch relativ selten, und sie sind verantwortlich für den Einbruch bei den deutschen Klassifikationen bei 10 % in Abbildung 5.

Insgesamt ist die verwendete Methode also relativ robust. Zur Erinnerung: 10 % der Daten bedeutet im englischen Teilkorpus eine Menge von 700 Wörtern, im deutschen Teilkorpus eine Menge von 549 Wörtern. Selbst auf dieser recht schmalen Datenbasis werden die wesentlichen Züge zuverlässig erkannt.

Nachdem auf diese Weise Vokale und Konsonanten ermittelt wurden, geht es im Folgenden um die globalen graphotaktischen Beschränkungen in den beiden Sprachen. Das Ziel ist es, die charakteristischen Strukturen herauszuarbeiten, und zwar explorativ mit möglichst wenigen Vorannahmen. Im besten Fall zeigen sich die Strukturen ‚von selbst‘.

3.2.1.3 Die Verteilung von Bigrammen

Dabei geht es zunächst nur um die Kombination von Buchstaben, und zwar sehr naiv unabhängig von ihrer silbenstrukturellen Position. Welche Buchstaben kombinieren mit welchen anderen Buchstaben wie häufig? Es geht hier also um Buchstabenpaare oder Bigramme, die ja bereits die Grundlage für die hierarchische Clusteranalyse bildeten.

Jedes Wort kann in eine Reihe solcher Bigramme zerlegt werden. So besteht bspw. das Wort *dünn* aus den drei Bigrammen |dü|, |ün| und |nn| und das Wort *sonic* aus den vier Bigrammen |so|, |on|, |ni| und |ic|. ²⁷ Die Bigramme liegen in Form von Kreuztabellen vor wie im folgenden Ausschnitt aus der englischen Tabelle:

²⁷ 3-Gramme und 4-Gramme sind potenziell ebenfalls von Interesse. Wenn sie im Folgenden nicht untersucht werden, so geschieht das aus zwei Gründen: Zum einen sind sie wesentlich schwerer darzustellen, zum anderen geht es an dieser Stelle um die basale Kombinationsmöglichkeit der Buchstaben; im folgenden Abschnitt (3.2.2) wird die Graphotaktik in Abhängigkeit von der Silbenstruktur dargestellt, und mit dieser Darstellung sind die (theoretisch naiveren) 3- und 4-Gramme überflüssig.

Tab. 3: Ausschnitt aus der Tabelle der Bigramme im englischen Teilkorpus ‚einfache Stämme‘.

	a	b	c	d	e	f	g	h	i	...
a	2	86	184	135	13	53	112	33	131	...
b	153	38	0	0	108	0	0	0	79	...
c	245	0	33	1	164	0	0	348	61	...
d	90	0	0	38	177	0	42	3	102	...
e	228	29	87	69	179	18	33	2	42	...
f	57	0	0	0	63	95	1	0	69	...
g	123	1	0	0	182	0	32	68	45	...
h	222	2	0	0	204	0	0	0	138	...
i	58	55	168	115	78	52	107	3	0	...
...

Hier sieht man auf den ersten Blick: nicht besonders viel. Es sind zu viele Datenpunkte, und sie sind zu unübersichtlich, als dass man Strukturen erkennen könnte. Deswegen werden im Folgenden andere Wege beschritten. Die Bigramme werden aus ganz unterschiedlichen Blickwinkeln analysiert. Konkret geht es um die folgenden Fragen:

- Welches sind die häufigsten Bigramme?
- Welche Bigramme sind nicht belegt?
- Welche Buchstaben kombinieren als Erst- oder Zweitbestandteile gut, welche nicht?
- Wie sieht das Frequenzspektrum der Bigramme aus – wie viele Bigramme kommen nur einmal vor, wie viele zweimal, dreimal usw.?
- Welche Bigramme sind häufiger, welche seltener als der Durchschnitt?
- Welche Buchstaben können verdoppelt werden, welche nicht?

Diese Fragen dienen dazu, Charakteristika der beiden Schriftsysteme sichtbar zu machen. Auf dieser Basis können dann typologische Vergleiche der Schriftsysteme angestellt werden. Die graphematische Typologie ist interessant, weil wir so gut wie nichts über typologisch-graphematische Parameter wissen – mit Ausnahme sehr grober Klassifikationen wie der in alphabetische, silbische und logographische Schriftsysteme und einen Parameter der Tiefe (vgl. oben 1.2). Wie sind bspw. komplexe Silbenkerne in verschiedenen alphabetischen Schriftsystemen aufgebaut? Welche Vokalbuchstaben können verdoppelt werden, welche nicht? Die graphematische Forschung hat gerade erst begonnen, Fragen wie diese zu

bearbeiten (vgl. z. B. Fuhrhop/Berg einger.). Eine derartige graphematische Typologie hat außerdem den praktischen Vorteil, dass sie sehr unaufwendig ist. Alles, was man braucht, ist ein morphologisch aufbereitetes, elektronisch verfügbares Lexikon einer Sprache. (Auch) vor diesem Hintergrund ist das Folgende zu lesen: als Versuch, interessante graphematisch-typologische Parameter zu identifizieren und geeignete Methoden zu entwickeln, die als Blaupause für zukünftige Untersuchungen dienen können.

Die Frage nach den häufigsten Bigrammen ist relativ leicht zu beantworten. Tabelle 4 gibt einen Überblick über die jeweils 10 häufigsten Bigramme im Deutschen und Englischen:

Tab. 4: Die 10 häufigsten Bigramme im Deutschen (links) und Englischen (rechts) mit ihrer jeweiligen Anzahl sowie dem Anteil an allen Bigrammen. Datenbasis: CELEX-Teilkorpora einfacher deutscher Stämme und einfacher englischer Stämme.

deutsch			englisch		
Bigramm	Anzahl	Anteil	Bigramm	Anzahl	Anteil
ch	827	3,3 %	le	587	1,9 %
er	756	3,0 %	in	488	1,6 %
el	537	2,1 %	ar	487	1,6 %
sc	520	2,1 %	an	419	1,4 %
te	416	1,6 %	ra	388	1,3 %
en	406	1,6 %	st	373	1,2 %
st	398	1,6 %	re	368	1,2 %
in	322	1,3 %	er	363	1,2 %
an	317	1,3 %	ch	348	1,1 %
le	300	1,2 %	on	348	1,1 %

Zwei Dinge fallen ins Auge. Erstens: Unter den zehn häufigsten Bigrammen finden sich ganz überwiegend Kombinationen aus Vokal und Konsonant (seltener Konsonant-Vokal-Kombinationen). Es gibt nur wenige Vokal-Vokal- oder Konsonant-Konsonant-Verbindungen; die wenigen Ausnahmen sind |ch|, |sc| (als Teil von |sch|) und |st| im Deutschen sowie |st| und |ch| im Englischen (allerdings sind |ch| und |sc| im Deutschen zwei der drei frequentesten Bigramme von diesem Typ). Das ist grundsätzlich erwartbar und die Basis für die Berechnung der Klassifikation – die Annahme, dass Vokale und Konsonanten grundsätzlich alternieren. Von dieser Regel scheint es nur wenige höherfrequente Ausnahmen zu geben.

Zweitens: Es gibt unter den zehn häufigsten Bigrammen im Deutschen sechs, die auch im Englischen zu den zehn häufigsten gehören; die Schnittmenge ist vergleichsweise groß. Es sind also nicht nur die Häufigkeiten der einzelnen Buch-

staben ähnlich, wie oben gezeigt, sondern auch die Häufigkeiten der frequentesten Bigramme.

Bei denjenigen Bigrammen, die überhaupt nicht vorkommen, ist das grundsätzlich ähnlich – dazu gleich mehr. Schauen wir uns zunächst die Anzahl der möglichen Bigramme sowie die Anzahl der belegten und nicht belegten Bigramme in beiden Sprachen an. Im Deutschen gibt es aufgrund des größeren Buchstabeninventars (|ä|, |ö|, |ü|, |ß|) mehr mögliche Bigramme als im Englischen – 30 x 30 gegenüber 26 x 26. Aufgrund dieses Unterschieds ist der Anteil der belegten Bigramme in beiden Sprachen unterschiedlich, wie Tabelle 5 zeigt:

Tab. 5: Mögliche Bigramme, belegte und nicht belegte Bigramme in den Teilkorpora einfacher deutscher Stämme und einfacher englischer Stämme.

	deutsch	englisch
kombinatorisch mögliche Bigramme	900	676
davon belegt	522 (58 %)	445 (66 %)
davon nicht belegt	378 (42 %)	231 (34 %)

Nimmt man |ä|, |ö|, |ü| und |ß| aus der Berechnung heraus, so sind im Deutschen 419 der insgesamt möglichen 676 Bigramme belegt (62%), 257 Bigramme sind nicht belegt (38%).

Wie groß ist nun die Ähnlichkeit zwischen den nicht belegten Bigrammen im Deutschen und Englischen – anders gefragt, wie groß ist die Schnittmenge? Betrachtet man wie im letzten Abschnitt nur diejenigen Buchstaben, die in beiden Sprachen vorkommen, so ergibt sich folgende Verteilung: Von den 419 Bigrammen, die im deutschen Teilkorpus vorkommen, sind im englischen 378 ebenfalls belegt (das sind 90%); von den 257 im Deutschen nicht belegten kommen 190 auch im Englischen nicht vor (das sind 74%). Das Verhältnis der Überschneidungen lässt sich gut mithilfe von Venn-Diagrammen visualisieren. Abbildung 6 präsentiert die Überschneidungen der belegten Bigramme sowie der unbelegten Bigramme zwischen beiden Sprachen.

Insgesamt sind die Überschneidungen in beiden Kategorien relativ groß (wenn man, wie erwähnt, im Deutschen die Umlaute und |ß| nicht berücksichtigt): Kommt eine Buchstabenverbindung also im englischen Teilkorpus vor, dann kommt sie wahrscheinlich auch im deutschen vor und umgekehrt; kommt sie im englischen Teilkorpus nicht vor, dann gilt das wahrscheinlich auch für das deutsche Teilkorpus und umgekehrt. Diese recht große Überschneidung ist überraschend, wenn man sich vor Augen führt, dass es sich hier um zwei sehr unterschiedliche Schriftsysteme handelt.

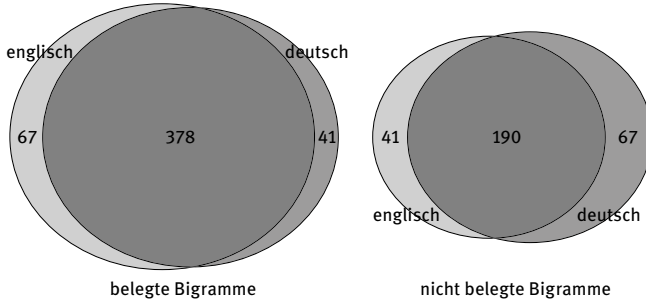


Abb. 6: Venn-Diagramme zu den Schnittmengen der belegten und nicht belegten Bigramme in den Teilkorpora einfacher deutscher Stämme und einfacher englischer Stämme.

Schaut man sich die entsprechenden nicht belegten Bigramme an, so sind Muster erkennbar. Tabelle 6 präsentiert jeweils die ersten 30 nicht belegten Bigramme (in alphabetischer Ordnung; die Umlaute und |ß| werden aus Gründen der Vergleichbarkeit ausgeklammert):

Tab. 6: Jeweils 30 nicht belegte Bigramme im Deutschen (links) und Englischen (rechts), alphabetisch sortiert. Datengrundlage: Korpora einfacher deutscher Stämme und einfacher englischer Stämme.

deutsch			englisch		
aq	bc	bf	bc	bd	bf
bg	bk	bm	bg	bh	bk
bn	bp	bq	bm	bp	bq
bv	bx	bz	bv	bw	bx
cb	cc	cd	bz	cb	cf
cf	cg	cj	cg	cj	cp
cl	cm	cn	cq	cv	cx
cp	cq	cr	db	dc	df
cs	cu	cv	dk	dp	dq
cw	cx	cy	dt	dx	fb

Bestimmte Buchstaben wie |c|, |j|, |v|, |y| scheinen in beiden Sprachen in den nicht belegten Bigrammen häufiger als andere. Für die Bigramme, die nicht vorkommen, sind offenbar vor allem bestimmte Buchstaben verantwortlich, die nicht oder nur sehr eingeschränkt kombinieren.

Welche Buchstaben sind das? Um das systematisch zu ermitteln, wird überprüft, welche Buchstaben *distributionell eingeschränkt* sind, und zwar an Erst- und Zweitposition in Bigrammen. Distributionell eingeschränkt bedeutet für die

Zwecke dieser Arbeit, dass der betreffende Buchstabe nur mit einem Drittel aller möglichen Buchstaben oder weniger kombiniert. Der Buchstabe |q| ist in beiden Sprachen ein extremes Beispiel: Er kombiniert als Erstbestandteil mit nur einem einzigen Buchstaben, nämlich |u|. Im deutschen Teilkorpus ist er auch als Zweitbestandteil in seiner Verteilung eingeschränkt, er tritt nur nach |e|, |i|, |r| und |s| auf, und zwar insgesamt nur fünfmal (vgl. die vollständige Liste *bequem, Mannequin, liquid, Marquis, Squaw*).²⁸ Im Englischen ist das anders: Hier kommt |q| als Zweitbestandteil insgesamt 75-mal vor, und zwar nach |a|, |e|, |i| und |o| sowie nach |l|, |m|, |n|, |r| und ganz besonders |s| (vgl. z. B. *square, squint, squirrel*). Tabelle 7 listet die distributionell eingeschränkten Bigramme in beiden Sprachen auf, jeweils an Erst- und Zweitposition im Bigramm, zusammen mit der Anzahl möglicher Kombinationen:

Tab. 7: Distributionell eingeschränkte Buchstaben im deutschen (links) und englischen (rechts) Teilkorpus einfacher Stämme. Distributionell eingeschränkt bedeutet, dass der Buchstabe mit maximal einem Drittel aller möglichen Buchstaben kombiniert. Angegeben ist jeweils der Kontext als Erst- und als Zweitbestandteil.

deutsch				englisch			
Erstbestandteil		Zweitbestandteil		Erstbestandteil		Zweitbestandteil	
Buchstabe	#Kombinationen	Buchstabe	#Kombinationen	Buchstabe	#Kombinationen	Buchstabe	#Kombinationen
q	1/30	q	4/30	q	1/26	x	7/26
ß	2/30	x	6/30	j	5/26	q	9/26
x	6/30	ß	7/30	v	8/26		
c	7/30	j	8/30	z	8/26		
j	7/30	v	9/30	x	9/26		
v	7/30						

Auch hier gibt es eine relativ große Schnittmenge der distributionell eingeschränkten Buchstaben in beiden Sprachen: Als Erstbestandteile sind das |q|, |j|, |v| und |x| im Englischen und Deutschen; als Zweitbestandteile |x| und |q|. Die Buchstaben |q| und |x| treten also in beiden Sprachen nur vor und nach ganz bestimmten anderen Buchstaben auf.

²⁸ Fälle wie *Aquarium* oder *Kolloquium* tauchen im Teilkorpus nicht auf, weil sie in CELEX nicht als morphologisch einfach annotiert sind, sondern als Einträge, deren morphologische Struktur ‚irrelevant‘ ist.

Außerdem gehen distributionelle Eingeschränktheit und niedrige Frequenz Hand in Hand: Mit Ausnahme von |c| im Deutschen gehören alle Buchstaben in Tabelle 7 zu den selteneren Buchstaben in beiden Sprachen (vgl. Abb. 1 oben). Das ist nicht überraschend: Je seltener ein Buchstabe vorkommt, desto geringer ist die Wahrscheinlichkeit, dass er mit einem anderen Buchstaben kombiniert. Treffen nun zwei seltene Buchstaben aufeinander, geht die Wahrscheinlichkeit ihres gemeinsamen Vorkommens gegen Null. Ein seltener Buchstabe geht also schon rein stochastisch weniger Verbindungen ein als ein häufiger Buchstabe.

Wir haben uns bis jetzt die nicht belegten und die belegten Bigramme in beiden Sprachen angesehen. In derselben Weise können nun auch die belegten Bigramme aufgegliedert werden in einmal, zweimal, dreimal belegte usw. Wir erhalten so das *Frequenzspektrum* der Bigramme (vgl. für das analoge Verteilungsmaß von Wörtern in Texten Baayen 2001). Die folgenden beiden Abbildungen zeigen die Frequenzspektren für die Bigramme im deutschen (links) und englischen Teilkorpus (rechts).

Wie sind die Diagramme zu lesen? Auf der horizontalen Achse sind (in logarithmischer Skalierung) die Frequenzklassen aufgetragen – also Bigramme, die in den Teilkorpora einmal vorkommen, zweimal, dreimal und so weiter, bis zum Höchstwert von 827 im deutschen Teilkorpus (|ch|) bzw. 587 im englischen Teilkorpus (|le|). Auf der vertikalen Achse ist für jede Frequenzklasse die Anzahl der Bigramme angegeben. Ein Beispiel: Bigramme, die nur einmal vorkommen (‘1’ auf der horizontalen Achse) sind im deutschen Teilkorpus 61-mal vertreten.

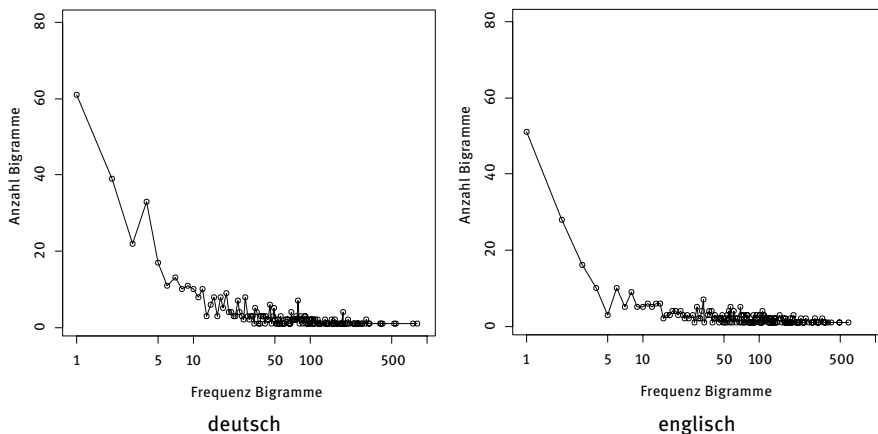


Abb. 7: Frequenzspektrum der Bigramme im deutschen (links) und englischen (rechts) Teilkorpus einfacher Stämme. Auf der horizontalen Achse ist logarithmisch skaliert die Frequenzklasse der Bigramme aufgetragen, auf der vertikalen Achse die Anzahl der Bigramme in der jeweiligen Frequenzklasse.

Auch hier gilt: Die Spektren der beiden Sprachen zeigen große Übereinstimmungen in ihrem Verlauf. Am häufigsten sind in beiden Korpora Bigramme, die einmal vorkommen; dann sinkt die Anzahl, auch wenn es Ausnahmen gibt. Bigramme mit einer Frequenz von über 100 kommen noch maximal fünfmal vor. Offenbar gibt es wie bei der Frequenzverteilung von Wörtern in einer Sprache auch bei den Bigrammen nur wenige sehr häufige und viele sehr seltene. Dieses Phänomen ist auf Wortebene als Zipf'sches Gesetz bekannt (vgl. z. B. Baayen 2001; Baroni 2009). Es besagt, dass die Frequenz von Wörtern in einem Korpus mit ihrem Rang korreliert. Das dritthäufigste Wort tritt – etwas vereinfacht – mit einem Drittel der Frequenz des häufigsten Wortes auf; das zehnthäufigste Wort mit einer Frequenz von einem Zehntel usw. Es sieht so aus, als seien die Verteilungen in Abbildung 7 ebenfalls Instanzen des Zipf'schen Gesetzes. Ob das wirklich der Fall ist, ist eine interessante Frage, die in einer weiterführenden Untersuchung verfolgt werden sollte, die auch die Frequenzspektren anderer Sprachen umfasst, die alphabetisch verschriftet werden.

Problematisch an den Aufstellungen der häufigsten, der belegten und der nicht belegten Bigramme oben ist vor allem, dass wir keinen vernünftigen Vergleichswert haben, um zu beurteilen, ob ein Vorkommen (oder Nichtvorkommen) so zu erwarten ist oder ob es besonders hoch oder niedrig ausfällt. Das Bigramm |en| zum Beispiel besteht aus den beiden hochfrequenten Buchstaben |e| und |n|. Dass auch |en| entsprechend häufig ist, ist zu erwarten. Aufschlussreicher wäre eine Aufstellung, die Informationen über die Frequenz der Bestandteile miteinbezieht und beides – reale und erwartete Werte – zueinander in Beziehung setzt.

Ein solches Verfahren haben Altmann/Lehfeldt (1980: 295 ff.) für die Phonemdistribution entwickelt. Es lässt sich ohne Modifikationen auf die Graphemdistribution übertragen. Der erwartete Wert berechnet sich aus den sog. Randsummen in Kreuztabellen wie Tabelle 3 oben. Er ist definiert als das Produkt der Spalten- und der Reihensumme der beteiligten Buchstaben, geteilt durch die Anzahl aller Bigramme. An einem Beispiel wird schnell deutlich, wie das funktioniert. Wie hoch ist die erwartete Häufigkeit von |kü| im deutschen Korpus? Dafür errechnen wir zunächst die sog. marginale Wahrscheinlichkeit, dass |k| als Erstbestandteil auftritt. |k| ist als Erstbestandteil 984-mal belegt; das ist die Reihenrandsumme n_k . Insgesamt gibt es im deutschen Teilkorpus 25.286 Bigramme (N); die Wahrscheinlichkeit, dass |k| in einem beliebig ausgewählten Bigramm der Erstbestandteil ist, liegt also bei knapp 3,9 %:

$$P_k = \frac{n_k}{N} = \frac{984}{25.286} = 0,0389.$$

Auf dieselbe Weise (nur mit der Spalten- statt der Reihenrandsumme) wird die Wahrscheinlichkeit errechnet, dass |ü| in einem beliebig ausgewählten Bigramm

der Zweitbestandteil ist. $P_{.ii}$ liegt bei 0,8 %. Die Wahrscheinlichkeit, dass beides zusammenkommt – |k| als erster und |ü| als zweiter Bestandteil –, ergibt sich aus dem Produkt der beiden marginalen Wahrscheinlichkeiten:

$$P_{k\ddot{u}} = P_{k.} \cdot P_{.ii} = 0,0389 \cdot 0,0077 = 0,0003$$

Die Wahrscheinlichkeit, dass es sich bei einem beliebig ausgewählten Bigramm genau um |kü| handelt, beträgt also 0,03 %. Wie viele Bigramme |kü| können wir auf dieser Basis erwarten? Dazu wird die Wahrscheinlichkeit $P_{k\ddot{u}}$ mit der Gesamtsumme aller Bigramme N multipliziert. Daraus ergibt sich der Erwartungswert $E_{k\ddot{u}}$:

$$E_{k\ddot{u}} = P_{k\ddot{u}} \cdot N = 0,0003 \cdot 25.286 = 7,627$$

Ausgehend allein von den Häufigkeiten der Buchstaben an Erst- und Zweitposition ergibt sich für das Bigramm |kü| also eine erwartete Häufigkeit von knapp 8 Vorkommen. Real kommt das Bigramm 15-mal vor. Wie wahrscheinlich ist es, dass eine solche (oder eine extremere) Abweichung rein zufällig vorkommt, quasi als Rauschen in den Daten?

Um das zu bestimmen, schlagen Altmann/Lehfeldt (1980) vor, die Binomialverteilung zu nutzen.²⁹ Mit ihrer Hilfe lässt sich die Wahrscheinlichkeit $P(X \leq n)$ berechnen, dass bei einem Bigramm mit der Auftretenswahrscheinlichkeit p_{ij} und einer Gesamtsumme N aller Bigramme der real vorkommende Wert dieses Bigramms n_{ij} (oder ein extremerer) auftritt – und zwar wie folgt:

$$P(X \leq n_{ij}) = \sum_{k=0}^{n_{ij}} \binom{N}{k} p_{ij}^k (1 - p_{ij})^{N-k}$$

Unterschreitet P nun bestimmte Grenzen, bewerten wir die Abweichung von erwartetem und realem Wert als nicht zufällig, sondern als signifikant. Altmann/Lehfeldt (1980) benutzen in Einklang mit dem wissenschaftlichen Usus den Wert von 0,05 als Grenze. Wir können aber durchaus weiter differenzieren und zusätzlich auch solche Verteilungen auszeichnen, die die Grenzen von 0,01 und 0,001 unterschreiten (siehe unten).

Lehfeldt (2005: 189) fasst die Grundidee anschaulich zusammen:

Wenn die Wahrscheinlichkeit p_{ij} dieser Verbindung gleich dem Produkt der Wahrscheinlichkeiten ihrer Komponenten i und j ist, d. h., wenn $p_{ij} = p_i p_j$, dann gelten die Phoneme i und j als stochastisch voneinander unabhängig. In linguistischer Interpretation bedeutet

29 Der Vorbehalt von Altmann/Lehfeldt (1980: 299 f.), dass sich die Berechnung bei großen N „wegen des unbequemen Binomialkoeffizienten als sehr mühselig“ erweise und deswegen auf Verteilungen ausgewichen werden sollte, die die Binomialverteilung approximieren, gilt heute nur noch für die Berechnung per Hand.

dies, dass die Phoneme i und j in der Anordnung ij weder eine Tendenz zur Assoziation, noch eine solche zur Dissoziation aufweisen, d. h., sie ziehen sich nicht besonders stark an, stoßen sich aber auch nicht besonders stark ab, sondern verhalten sich zueinander neutral.

In der entsprechenden Kreuztabelle (Tab. 9 unten) werden Kombinationen von Buchstaben, die sich signifikant abstoßen, mit Minuszeichen bezeichnet. Zusätzlich dazu wird – anders als bei Altmann/Lehfeldt (1980) – auch das Signifikanzniveau angegeben: ‚-‘ bedeutet, dass die Wahrscheinlichkeit für eine solche Anzahl bei unter 5 % liegt; ‚--‘ bedeutet, dass sie unter 1 %, und ‚---‘ bedeutet, dass sie unter 0,1 % liegt. Die Anziehung wird ganz analog mit Pluszeichen bezeichnet (die Signifikanzniveaus sind dieselben). Alle anderen Beziehungen sind neutral und werden mit dem Buchstaben N bezeichnet.

Der besondere Reiz des Modells von Altmann/Lehfeldt (1980) steckt aber vor allem in der Behandlung derjenigen Bigramme, die *nicht* belegt sind. Für die gibt es ja prinzipiell zwei Erklärungen: Entweder sie kommen zufällig nicht vor – wenn es mehr Wörter im Korpus gäbe, wäre das vielleicht anders. Oder ihr Nichtvorkommen ist systematisch. Altmann/Lehfeldt (1980) benutzen hier ebenfalls die Binomialverteilung, um zu entscheiden, ob ein nicht belegtes Bigramm so unwahrscheinlich ist, dass es „unzulässig“ ist (Altmann/Lehfeldt 1980: 301); das wird in Tabelle 9 mit U bezeichnet. Das Kriterium hierfür ist eine Wahrscheinlichkeit von 5 % oder weniger. Das ist vor allem der Fall, wenn der erwartete Wert vergleichsweise hoch ist; dann ist ein reales Vorkommen von 0 extrem unwahrscheinlich. Ist ein Bigramm hingegen zwar real nicht belegt, der erwartete Wert allerdings relativ niedrig, dann ist das Vorkommen von 0 nicht signifikant unwahrscheinlich, und das nicht belegte Bigramm wird als ‚virtuell‘ bezeichnet (V in den Tabellen). Tabelle 8 gibt einen Überblick über die verwendeten Bezeichnungen und Entscheidungskriterien in den folgenden Tabellen (nach Altmann/Lehfeldt 1980: 301):

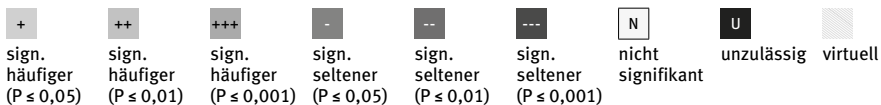
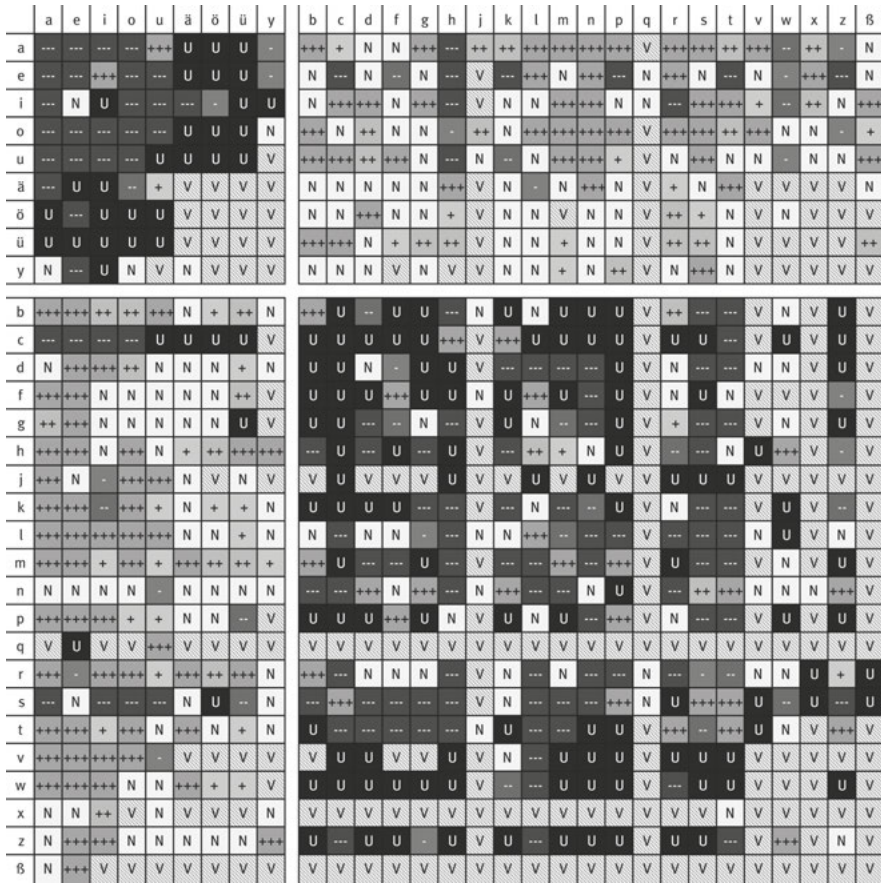
Tab. 8: Klassifikationskriterien für die Bezeichnung des Verhältnisses zwischen realem Wert (n_{ij}) und erwartetem Wert (E_{ij}) in den folgenden Tabellen (modifiziert nach Altmann/Lehfeldt 1980).

	$n_{ij} > E_{ij}$ $P \leq 0,001$	$n_{ij} > E_{ij}$ $P \leq 0,01$	$n_{ij} > E_{ij}$ $P \leq 0,05$	$n_{ij} > E_{ij}$ nicht-signifikant	$n_{ij} > E_{ij}$ $P \leq 0,05$	$n_{ij} > E_{ij}$ $P \leq 0,01$	$n_{ij} > E_{ij}$ $P \leq 0,001$
$n_{ij} > 0$	+++	++	+	N	-	--	---
$n_{ij} = 0$				V		U	

Tabelle 9 zeigt auf dieser Grundlage für das deutsche Teilkorpus, welche Verbindungen im deutschen Teilkorpus signifikant ‚anziehend‘ und ‚abstoßend‘ sind und welche nicht belegten Bigramme möglich, welche unmöglich sind. Wie in den folgenden Tabellen werden signifikant anziehende Bigramme in (je

nach Signifikanzniveau) unterschiedlichen Hellgrautönen eingefärbt, abstoßende Bigramme hingegen in Dunkelgrautönen. Neutrale Bigramme sind weiß; virtuelle sind schraffiert, und unmögliche sind schwarz eingefärbt.

Tab. 9: Kreuztabelle der Buchstaben im deutschen Teilkorpus, klassifiziert nach der Wahrscheinlichkeit, nach der der reale Wert zufällig zustande kommt. Vertikal: jeweils erste Bestandteile von Bigrammen; horizontal: jeweils zweite Bestandteile.



Auf den ersten Blick fällt ins Auge, dass Vokale mit Konsonanten kombinieren und umgekehrt. Hier sind die meisten hellgrauen und weißen Bereiche. Vokal-

Vokal- sowie Konsonant-Konsonant-Verbindungen sind signifikant seltener, als man es stochastisch erwarten sollte – oder erst gar nicht belegt: Hier finden sich die meisten schwarzen und schraffierten Zellen.

Diese Ergebnisse sind nicht überraschend. Dass Vokale und Konsonanten grundsätzlich alternieren, war ja unter anderem die theoretisch motivierte Vorannahme für die Klassifizierung der Buchstaben in Vokale und Konsonanten. In Tabelle 9 wird die Gültigkeit dieser theoretischen Vorannahme nun für das deutsche Teilkorpus demonstriert. Das Modell ist allerdings bewusst naiv: Wenn zwei Buchstaben dieselbe Frequenz als Erst- oder Zweitbestandteil haben, dann ist die erwartete Häufigkeit der Verbindung mit allen anderen Buchstaben gleich, und davon wiederum hängt die Einschätzung der Wahrscheinlichkeit in Tabelle 9 ab. Ein Beispiel: |i| kommt als Zweitbestandteil im deutschen Korpus 1.664-mal vor, |n| 1.685-mal. Die erwarteten Häufigkeiten für bspw. |oi| und |on| sind daher sehr ähnlich (71 und 72). Das Bigramm |oi| kommt aber nur zehnmal vor und ist deswegen in Tabelle 9 hochsignifikant selten; das Bigramm |on| hingegen kommt 172-mal vor und ist damit hochsignifikant häufig.

Nun wissen wir aber ja (z. B. durch die hierarchische Clusteranalyse oder Tab. 9), dass Vokale eher von Konsonanten gefolgt werden als von anderen Vokalen. Das Bigramm |oi| sollte daher einen wesentlich niedrigeren Erwartungswert haben als das Bigramm |on|. Damit würde sich für beide Verbindungen auch die Schätzung der Wahrscheinlichkeit in Tabelle 9 ändern: Wir vermuten von vornherein, dass |on| häufiger ist; ein Wert wie 172 ist damit kein krasser Ausreißer nach oben. Und |oi| wird bereits im Vorfeld als seltener eingestuft, sodass die vorgefundenen 10 Belege nicht signifikant seltener ausfallen als erwartet.

Das Modell von Altmann/Lehfeldt (1980) muss daher modifiziert werden. Doch was sind geeignete Erwartungswerte für Vokale und Konsonanten, und wie können sie als Parameter ins Modell eingearbeitet werden? Um diese nicht-triviale Frage einfach zu umgehen, folgen wir einer Randbemerkung von Altmann/Lehfeldt (1980: 303), die vorschlagen, statt der gesamten Kontingenztafel nur Teile auszuwerten – z. B. nur Vokal-Vokal- oder nur Konsonant-Konsonant-Verbindungen. So treten wir diesen Verbindungen angemessener gegenüber. Die Frage, die wir stellen, ist nur noch: Wenn wir alle Bigramme eines Typs (Vokal-Vokal, Konsonant-Konsonant etc.) jeweils als Grundgesamtheit nehmen, welche sind dann signifikant häufiger, als wir erwarten würden? Damit wird schon anerkannt, dass es wesentlich weniger Vokal-Vokal-Verbindungen gibt als bspw. Konsonant-Vokal-Verbindungen.

Auf diese Weise ergibt sich eine Tabelle, die vom Aufbau her Tabelle 9 gleicht – deren Zellen aber jeweils in jedem der vier großen Blöcke (Vokal-Vokal, Konsonant-Konsonant, Vokal-Konsonant, Konsonant-Vokal) separat berechnet wurde. Für das deutsche Teilkorpus ist das die modifizierte Tabelle 10 (nächste Seite).

Tab. 10: Kreuztabelle der Buchstaben im deutschen Teilkorpus, klassifiziert nach der Wahrscheinlichkeit, nach der der reale Wert zufällig zustande kommt. Einzelnen berechnet für jede der vier großen Gruppen von Bigrammen. Vertikal: jeweils erste Bestandteile von Bigrammen; horizontal: jeweils zweite Bestandteile.

	a	e	i	o	u	ä	ö	ü	y	b	c	d	f	g	h	j	k	l	m	n	p	q	w	s	t	v	w	x	z	ß		
a	N	---	---	---	+++	V	V	V	N	++	N	N	N	N	N	N	+++	N	+	N	++	V	---	N	N	N	N	N	N	N	-	
e	---	---	+++	---	---	V	V	V	N	---	---	-	---	-	N	U	---	+++	---	N	---	N	+++	---	---	N	N	N	N	-	N	
i	N	+++	U	+++	---	N	N	V	V	-	+++	+++	N	+++	---	V	N	---	N	+++	N	N	---	+	+++	N	N	N	++	+		
o	N	-	N	+	N	V	V	V	+	+	N	N	N	-	N	N	N	-	N	N	+++	V	N	N	N	N	N	N	N	N		
u	+++	++	N	N	U	V	V	V	V	N	+++	N	++	N	N	N	-	---	---	N	V	---	+++	N	N	N	N	N	N	N		
ä	N	U	U	N	+++	V	V	V	V	N	N	N	N	N	+++	V	N	---	N	N	N	V	N	N	+++	V	V	V	V	N		
ö	V	N	V	V	V	V	V	V	V	N	N	++	N	N	+++	V	N	U	-	N	V	N	N	N	V	+	V	V	V	N		
ü	V	V	V	V	V	V	V	V	V	+	+++	N	N	N	+++	V	N	-	N	-	N	V	N	N	N	V	V	V	V	N		
y	+++	N	U	++	U	+	V	V	V	N	N	N	V	N	V	V	N	N	-	++	V	N	+	N	V	V	V	V	V	V	N	
b	N	+	N	N	N	N	N	N	N	+++	U	N	U	U	---	N	U	+++	U	U	U	V	+++	N	---	V	N	V	U	V		
c	N	N	N	N	N	V	V	V	V	U	U	U	U	+++	V	+++	U	U	U	U	V	U	U	---	V	U	V	U	V	U		
d	---	+++	N	-	-	N	N	N	N	V	U	+++	N	V	U	V	N	N	N	U	U	V	+++	N	---	N	V	V	V	V		
f	N	N	N	-	N	N	N	N	V	U	U	U	+++	U	U	N	U	+++	U	-	U	V	+++	U	N	V	U	V	-	V		
g	---	+++	---	---	-	N	N	U	V	V	U	-	N	+	---	V	U	+++	N	+	U	V	+++	N	---	V	N	V	U	V		
h	N	N	-	N	N	N	N	N	+++	---	U	---	U	---	U	V	---	+++	+++	+++	U	V	+++	N	N	V	+++	V	-	V		
j	N	-	---	++	+++	N	V	N	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V		
k	+++	-	---	+++	N	-	N	N	N	U	U	U	U	-	---	V	---	+++	-	+++	U	V	+++	N	N	N	V	U	V	-	V	
l	N	-	+++	N	N	N	N	N	N	+++	---	++	N	N	---	N	N	+++	N	---	-	V	---	N	-	+	U	V	N	V		
m	+++	-	-	+	N	N	N	N	N	+++	U	-	-	U	---	V	---	---	+++	---	+++	U	N	---	V	N	N	V	U	V	U	
n	N	N	++	N	N	N	N	N	N	---	---	+++	N	+++	---	N	+++	---	+++	U	V	---	+++	+	-	+	++	+++	V	U	V	
p	N	N	N	N	N	N	N	---	U	U	U	U	+++	U	---	V	U	+++	U	---	+++	V	+++	N	---	V	U	V	U	V		
q	U	U	U	U	+++	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	
r	+++	---	+++	++	N	++	N	+++	N	+++	---	+++	N	+++	---	V	N	---	++++	---	N	N	++	N	++	N	V	+	V	U	V	
s	---	+++	N	N	-	N	U	N	N	---	+++	---	---	---	V	---	---	---	+++	N	U	+++	+++	V	---	V	---	V	---	V		
t	N	++	N	N	+	N	N	N	N	U	---	---	---	---	N	U	---	---	---	U	U	---	+	+++	V	N	V	+++	V	U	V	
v	N	++	N	N	---	V	V	U	V	V	V	V	V	V	V	V	N	N	V	V	V	V	V	V	V	V	V	V	V	V	V	
w	+	-	+++	---	-	++	N	N	V	V	V	V	V	V	V	N	N	V	V	V	V	++	V	V	V	V	V	V	V	V	V	
x	N	N	N	V	N	V	V	V	N	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	++	V	V	V	V	V	V	V
z	---	++	+	-	N	N	N	N	+++	V	N	V	V	N	U	V	V	N	V	V	V	V	U	V	N	V	+++	V	+	V	U	V
ß	N	+	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V

Ein Vergleich mit Tabelle 9 zeigt, dass es nun wesentlich mehr neutrale und mehr virtuelle Bigramme gibt. Die vorgefundene Verteilung entspricht also insgesamt mehr unseren Erwartungen. Das bedeutet, dass die Modifikation des Modells erfolgreich war. Das wird am Beispiel von Bigrammen mit |q| als Erstbestandteil und Vokalen als Zweitbestandteil deutlich (d. h. der erste Teil der Zeile ‚q‘ in den Tabellen oben). In Tabelle 9 ist genau eine Verbindung hochsignifikant häufiger als erwartet (|qu|); eine ist unmöglich (|qe|); die anderen sind virtuell. In Tabelle 10 hat sich das geändert. |qu| bleibt hochsignifikant häufiger, aber nun sind neben |qe| auch |qa|, |qi| und |qo| unmöglich. Wie kommt das? Das liegt daran, dass sich

die erwarteten Häufigkeiten in Tabelle 10 nur aus den Konsonant-Vokal-Verbindungen ergeben und nicht (wie in Tab. 9) aus allen Verbindungen. Das heißt, dass die 34 Vorkommen von |q| als Erstbestandteil in Tabelle 9 oben auf alle 30 Buchstaben als Zweitbestandteil aufgeteilt werden, in Tabelle 10 hingegen nur noch auf alle Vokalbuchstaben. Der erwartete Wert für die Verbindungen mit |q...| liegt in Tabelle 10 also höher, und deswegen werden mehr Verbindungen als ‚unmöglich‘ bewertet. Hier ist der Abstand zwischen tatsächlichem Wert (0) und erwartetem Wert so groß, dass er wahrscheinlich nicht zufällig ist. Und warum verhalten sich nicht alle Vokalbuchstaben gleich, warum sind |qä|, |qö|, |qü| und |qy| also nach wie vor ‚virtuell‘? Weil sich der erwartete Wert auch aus der Gesamtsumme der Zweitbestandteile ergibt, und die Umlaute und |y| sind eben wesentlich seltener als die übrigen Vokale.

Deutlich zu sehen sind einige Buchstaben, die sich als graue/schwarze Linien durch die Tabelle ziehen, horizontal etwa |c|, |j|, |q|, |v|, |w|, |x| und |ß|, vertikal |c|, |j|, |p|, |q|, |v|, |x| und |ß|. Diese Buchstaben sind distributionell eingeschränkt – auf sie können (im Fall der ersten Gruppe) nur wenige Buchstaben folgen, oder sie folgen (im Fall der zweiten Gruppe) selbst nur auf wenige Buchstaben. Es handelt sich hier um die Buchstaben, die in Tabelle 7 oben als distributionell eingeschränkt charakterisiert wurden.

Da die meisten Fragen der Verteilung im folgenden Kapitel behandelt werden, beschränkt sich dieser Teil auf einige Beobachtungen, die dort keinen Platz haben. Das betrifft vor allem Konsonant-Vokal- und Vokal-Konsonant-Verbindungen. Interessant sind außerdem die Verdoppelungen von Buchstaben.

Zunächst zu den Verbindungen aus Vokal und Konsonant (das ist das Feld rechts oben in Tab. 10). Es fällt auf, dass einige Vokalbuchstaben ein wesentlich ausgewogeneres Profil haben als andere. |a| als Erstbestandteil verbindet sich häufiger als erwartet mit vier anderen Buchstaben (in den Bigrammen |ab|, |ak|, |an|, |ap|), und signifikant seltener mit zwei Buchstaben (in den Bigrammen |ar|, |aß|). Die übrigen Verbindungen sind aber neutral (mit Ausnahme des nicht belegten |aq|). Das ist bei |e| ganz anders: Hier stehen zwei häufigere Verbindungen (|er| und |el|) elf selteneren Verbindungen gegenüber, und nur sieben Verbindungen sind neutral. Beides bedingt sich: Dadurch, dass |el| und |er| weitaus häufiger sind als erwartet (|el|: 537 statt 201, |er|: 756 statt 250), sind die meisten anderen Verbindungen weitaus seltener als erwartet. Beide Bigramme kommen ganz überwiegend stammfinal vor, |el| zu 79 % und |er| zu 68 % (das dritte zu erwartende Bigramm in der Reihe, |en|, ist nicht signifikant häufiger als erwartet).

Das gilt auch für |ig|: in 63 von 95 Stämmen (66 %) kommt das Bigramm stammfinal oder vor finalem |t| vor (z. B. *billig*, *fertig*, *Pfennig*, *berüchtigt*). Da es ganz überwiegend bei Adjektiven vorkommt, kann man es als Suffix klassifizieren – auch wenn es keine freien Stämme zu den Bildungen gibt (vgl. oben 2.2).

Entscheidend ist, dass das frequente Bigramm |ig| nicht zufällig frequent ist – es hat eine Funktion.

Auch bei der umgekehrten Abfolge Konsonant+Vokal (das ist das Feld links unten in Tab. 10) gibt es Vokalbuchstaben, die ein ausgewogeneres Profil haben als andere. Die Umlautbuchstaben und |y| verhalten sich bspw. den meisten anderen Buchstaben gegenüber neutral.³⁰ Das Gegenbeispiel ist auch hier |e|: Verbindungen mit |e| als zweitem Teil sind in nur sechs Fällen neutral, ansonsten entweder signifikant seltener als erwartet (|je|, |ke|, |le|, |me|, |re|, |we|) oder signifikant häufiger als erwartet (|de|, |ge|, |se|, |te|, |ve|, |ze|, |ße|). Bestimmte Vokale haben also – wie oben bei der umgekehrten Abfolge – gewissermaßen Favoriten, und diese frequenten Verbindungen führen dann dazu, dass andere Verbindungen seltener als erwartet ausfallen.

Es ist mit der Verbindung von realer und erwarteter Frequenz also möglich, die ‚rohen‘ Vorkommenshäufigkeiten der Bigramme angemessener zu bewerten und die charakteristischen Verbindungen für das deutsche Korpus herauszuarbeiten. Hier können dann qualitative Analysen ansetzen, um die Verteilung funktional zu interpretieren.

Verdoppelungen von Buchstaben können ein typologischer Parameter sein; nicht alle Verdoppelungen sind in allen Sprachen möglich, und die existierenden Verdoppelungen unterscheiden sich in ihrer Häufigkeit. Signifikant häufiger als erwartet (und damit ‚anziehend‘ im Sinne Lehfeldts 2005) sind im deutschen Teilkorpus die Verdoppelungen der folgenden Buchstaben:

(16a) o

(16b) b, d, f, g, l, m, n, p, s, t, z

Bei den Vokalen ist |oo| das einzige Bigramm, das signifikant häufiger als erwartet auftritt. |aa| ist in dieser Hinsicht neutral, |ee| signifikant seltener als erwartet und |ii|, |uu| sogar unmöglich. Die übrigen Doppelungen kommen nicht vor, sind aber virtuell. Das ist ein weiteres Argument dafür, dass Vokalverdoppelung im Deutschen systematisch nicht von Bedeutung ist.

Bei den Konsonanten ist das Verhältnis der Ausnutzung wesentlich höher. Neben den signifikant häufigeren Doppelungen der Konsonanten in (16b) kommen auch noch |rr| und |kk| vor. |rr| ist neutral hinsichtlich der Erwartung, |kk|

³⁰ Die Ausnahmen bei |y| sind |hy| und |zy|, die 16- bzw. 8-mal vorkommen (z. B. *Hyazinthe*, *Zyklon*). Diese Werte sind im Vergleich mit den anderen Bigrammen relativ niedrig. Weil aber jeweils beide Buchstaben global selten sind, liegen die Werte eben dennoch weit über den stochastisch erwarteten Werten.

hochsignifikant seltener – es kommt nur in drei (fremden) Stämmen vor (*Akkordeon, Mokka, Okkasion*). Damit sind – je nachdem, ob man |kk| zulässt – 12 oder 13 von insgesamt 21 Buchstaben verdoppelbar, und die meisten davon kommen wesentlich häufiger vor, als man das stochastisch erwarten würde. Die Verdoppelung von Konsonanten ist also deutlich besser im System verankert. Verdreifachungen von Buchstaben – egal ob Konsonanten- oder Vokalbuchstaben – finden sich im Korpus einfacher Stämme nicht.

Interessant ist schließlich noch die Liste derjenigen Buchstaben, die nicht verdoppelt auftreten (|c|, |h|, |j|, |q|, |v|, |w|, |x|, |ß|). Diese Liste enthält die mittlerweile üblichen Verdächtigen, also seltene Buchstaben mit eingeschränkter Kombinatorik – und |h|. Dessen Nicht-Verdoppelung hängt mit seiner phonologischen Korrespondenz zusammen. Er kommt nicht im Silbengelenk vor.

Wir können mit den Daten aus Tabelle 10 oben noch einmal zu den häufigsten Bigrammen im deutschen Korpus zurückkehren und sie besser einordnen, indem wir uns bspw. nur diejenigen häufigen Bigramme ansehen, die auch signifikant häufiger sind, als stochastisch zu erwarten wäre. Tabelle 11 zeigt die 20 häufigsten Bigramme, die häufiger sind als erwartet:

Tab. 11: Die 20 häufigsten Bigramme im deutschen Teilkorpus, die signifikant häufiger sind als erwartet.

Bigramm	Frequenz	relativ	sign.+
ch	827	3,3 %	+++
er	756	3,0 %	+++
el	537	2,1 %	+++
sc	520	2,1 %	+++
te	416	1,6 %	++
st	398	1,6 %	+++
in	322	1,3 %	+++
ra	300	1,2 %	+++
ei	295	1,2 %	+++
ge	276	1,1 %	+++
ie	256	1,0 %	+++
se	254	1,0 %	+++
de	250	1,0 %	+++
be	234	0,9 %	+
ka	211	0,8 %	+++
au	207	0,8 %	+++
ri	197	0,8 %	+++
ck	192	0,8 %	+++
ll	190	0,8 %	+++
ma	190	0,8 %	+++

Im Gegensatz zu Tabelle 4 oben (die nur die häufigsten zehn Bigramme enthält) sind hier Bigramme der Struktur Konsonant-Vokal (z. B. *te*, *ra* etc.) am häufigsten vertreten (9 von 20); unter diesen Bigrammen wiederum sind auffallend viele mit |e| als zweitem Bestandteil. Schauen wir uns für |de| beispielhaft an, wo diese Bigramme in den Stämmen des Korpus vorkommen. In 79 % der Stämme tritt das Bigramm stammfinal auf (z. B. *Bande*) oder vor stammfinalen |l|, |m|, |n| oder |r| auf (z. B. *Gondel*, *Odem*, *Frieden*, *Bruder*). |e| ist also – wenig überraschend – oft der Silbenkern der zweiten Silbe in graphematisch mehrsilbigen Wörtern. Das gilt ganz ähnlich auch für |te| und |se|. Bei |ge| und |be| liegt der entsprechende Anteil niedriger, weil beide häufig stamminitial vorkommen (z. B. *Besteck*, *geheim*).

Auch für das englische Korpus lässt sich eine Aufstellung der Bigramme aufstellen, die Abweichungen von der erwarteten Häufigkeit wiedergibt. Das geschieht in Tabelle 12 (nächste Seite). Die erwarteten Häufigkeiten sind (wie in Tab. 10 für das Deutsche) bereits für die einzelnen Klassen (Vokal-Vokal, Vokal-Konsonant usw.) angepasst.

Wie bei der Verteilung im deutschen Korpus ist auch hier zu beobachten, dass es Vokalbuchstaben gibt, die ausgewogener kombinieren – bei denen also bestimmte Verbindungen weder ‚angezogen‘ noch ‚abgestoßen‘ werden –, und daneben solche, die nur über wenige neutral bewertete Verbindungen verfügen. Wie im deutschen Korpus ist |e| sowohl als Erst- wie auch als Zweitbestandteil nur selten neutral. Die meisten Verbindungen mit |e| sind entweder signifikant häufiger oder seltener als erwartet. Das Gegenbeispiel ist |y|, bei dem der Großteil der Kombinationen mit Konsonantenbuchstaben (wieder als Erst- wie auch als Zweitbestandteil) neutral ist.

Bei den Verbindungen von Vokal- und Konsonantenbuchstaben gibt es einige interessante Bigramme: |ah| z. B. ist auf den Lehnwortbereich beschränkt und kommt in mehr als der Hälfte der 32 Belege stammfinal vor (z. B. *cheetah*, *pariah*). Die signifikant häufigeren Verbindungen |ig|, |ow| und |ug| sind aus phonographischer Perspektive interessant, weil sie – trotz ihres konsonantischen Bestands – oft nicht (im Fall von |ow|: nie) mit konsonantischen Phonemen korrespondieren (z. B. *night*, *blow*, *plough*).

Ebenfalls aus phonographischer Perspektive interessant ist die Verteilung von |c| und |k| als ersten Bestandteilen. Während sich beide mit |y| als zweitem Bestandteil neutral verhalten, sind sie vor |a|, |e|, |i|, |o| und |u| beinahe komplementär verteilt: |ca| und |co| sind signifikant häufiger als erwartet, |ce| und |ci| signifikant seltener, |cu| ist neutral; |ka|, |ko| und |ku| sind signifikant seltener als erwartet, |ke| und |ki| signifikant häufiger. Der Kontext vor |e| und |i| ist nun genau derjenige, in dem |c| nicht mit /k/, sondern mit /s/ korrespondiert (vgl. z. B. *cell*,

city vs. cat, cot, cut, vgl. Fuhrhop/Buchmann/Berg 2011: 283 f.). Und in genau diesem Kontext ist |k| häufiger als erwartet (z. B. kettle, kite).³¹

Tab. 12: Kreuztabelle der Buchstaben im englischen Teilkorpus, klassifiziert nach der Wahrscheinlichkeit, nach der der reale Wert zufällig zustande kommt. Einzeln berechnet für jede der vier großen Gruppen von Bigrammen. Vertikal: jeweils erste Bestandteile von Bigrammen; horizontal: jeweils zweite Bestandteile.

	a	e	i	o	u	y	b	c	d	f	g	h	j	k	l	m	n	p	q	r	s	t	v	w	x	z	
a	---	---	+++	---	+++	+++	N	N	N	N	N	+++	N	+	N	N	---	N	-	+++	N	N	++	N	N	+	
e	+++	+++	---	---	---	N	---	---	---	---	---	N	-	+++	---	N	---	N	+++	N	+++	N	N	N	++	-	
i	+	+++	U	++	---	U	N	+++	+	+	+++	-	N	N	N	N	+++	N	+	---	N	N	N	---	N	+	
o	---	---	---	+++	+++	N	N	-	N	N	N	N	+	-	N	N	+++	N	N	---	N	N	+++	N	+	-	
u	N	+++	+++	---	U	---	++	N	N	+	+++	-	N	-	-	+++	---	N	U	N	+++	N	---	U	-	N	
y	N	N	-	++	-	U	N	N	N	V	N	N	N	N	N	N	N	+	V	N	N	N	N	N	++	V	
b	+	---	N	+	+++	N	+++	U	U	U	U	U	N	U	+++	U	---	U	V	+++	-	---	V	U	V	V	
c	++	---	---	+++	N	N	U	-	---	U	U	+++	V	+++	-	---	---	U	U	+	---	---	U	---	V	N	
d	---	N	N	N	N	N	U	U	+++	U	+++	---	+	U	+++	N	-	U	V	+++	---	U	N	N	V	N	
f	N	-	+++	N	N	-	U	U	U	+++	---	U	N	U	+++	U	U	U	V	+++	U	N	V	U	V	V	
g	N	+	---	N	++	N	---	U	U	U	+++	+++	V	---	+++	N	+++	U	V	+++	---	U	V	N	V	V	
h	+	-	N	N	N	++	N	U	U	V	U	U	N	U	-	+++	N	-	V	+++	U	+++	V	N	V	V	
j	N	-	N	N	++	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
k	---	+++	+++	---	-	N	N	U	N	V	N	N	V	N	+++	V	+++	U	V	N	N	---	N	N	V	V	
l	N	+++	N	---	---	N	N	N	+++	N	---	V	N	+++	+	---	N	N	---	N	N	+++	N	+++	-	V	V
m	+++	---	N	N	N	N	+++	---	---	N	U	V	U	---	+++	N	+++	N	---	N	---	N	---	V	U	V	V
n	-	+++	N	N	---	N	U	+++	+++	N	+++	---	N	+++	---	U	+++	U	N	---	N	+++	++	U	+++	N	
p	++	-	N	N	N	N	U	U	U	U	++	V	---	+++	---	---	+++	V	+++	N	---	V	U	V	V	V	
q	U	U	U	U	+++	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	
r	N	---	+++	+	-	+	+++	N	+++	N	+++	---	N	N	---	+++	+++	N	N	-	+++	N	+++	-	V	N	
s	---	+++	N	N	N	N	---	+++	U	U	---	N	V	N	---	N	N	+++	+++	U	+++	+++	U	+++	V	U	
t	N	N	N	N	N	N	U	+	U	---	U	+++	V	U	---	---	---	U	U	+++	-	N	U	+++	V	N	
v	---	+++	+	---	---	N	V	V	V	V	V	V	V	V	V	V	V	V	V	V	N	V	V	+	V	V	V
w	+++	-	+++	-	U	-	U	U	N	V	-	+++	V	N	N	N	+++	-	V	N	N	---	V	V	V	V	
x	N	N	++	N	N	N	V	V	V	V	V	V	V	N	V	V	V	V	V	V	N	++	V	V	V	V	
z	-	+++	N	N	-	N	V	V	V	V	U	V	V	++	V	V	V	V	U	V	U	V	V	V	V	+++	

³¹ Das gilt vor allem für |ki|. Dass |ke| häufiger ist als erwartet, liegt vor allem daran, dass es häufig stammfinal oder vor stammfinalem |l|, |n|, |r| oder |t| vorkommt (z. B. like, bracket, poker).

Was die Verdoppelung von Buchstaben angeht, so sind im englischen Teilkorpus signifikant häufiger als erwartet:

(17a) e, o

(17b) b, d, f, g, l, m, n, p, s, v, z

Neben diesen Verdoppelungen kommen auch |rr| (signifikant seltener als erwartet) und |tt| (neutral hinsichtlich der Erwartung) vor. Diese Liste entspricht mit zwei Ausnahmen derjenigen im deutschen Korpus (vgl. (16) oben): Zusätzlich zu den Buchstaben, die im deutschen Korpus häufiger als erwartet verdoppelt vorkommen, sind im englischen Korpus auch |ee| und |vv| häufiger als erwartet. |vv| erklärt sich durch die niedrige globale Frequenz von |v|; sie führt zu einem sehr niedrigen erwarteten Wert von |vv| (0,0021). Der reale Wert liegt bei genau 1 (*savvy*), und dieser Unterschied ist stochastisch hochsignifikant. Auf dieser Grundlage können wir |vv| ausschließen; |ee| ist aber tatsächlich im englischen Korpus wesentlich stärker vertreten als im deutschen. Die Verdoppelungen |ii|, |uu| und |yy| sind unmöglich, |aa| hochsignifikant seltener als erwartet. |aa| kommt im Korpus genau zweimal vor, im Onomatopoetikon *baa* sowie im Lehnwort *bazaar*. Im Kernwortschatz ist es – genau wie |ii|, |uu| und |yy| – unmöglich. Verdreifachungen sind in einfachen Stämmen – wie im deutschen Korpus – nicht belegt.

Wir können nun auch im englischen Korpus die absolute Frequenz der Bigramme verknüpfen mit der Information über die Signifikanz dieser Frequenz und uns bspw. die 20 häufigsten Bigramme anzeigen lassen, die signifikant häufiger sind, als stochastisch zu erwarten wäre:

Tab. 13: Die 20 häufigsten Bigramme im englischen Teilkorpus, die signifikant häufiger sind als erwartet.

Bigramm	Frequenz	relativ	sign.+
le	587	1,9 %	+++
in	488	1,6 %	+++
ar	487	1,6 %	+++
st	373	1,2 %	+++
er	363	1,2 %	+++
ch	348	1,1 %	+++
ri	325	1,1 %	+++
ro	323	1,1 %	+
co	291	1,0 %	+++
se	281	0,9 %	+++
el	271	0,9 %	+++

Bigramm	Frequenz	relativ	sign.+
ne	251	0,8 %	+++
ca	245	0,8 %	++
et	241	0,8 %	+++
ma	239	0,8 %	+++
ea	228	0,7 %	+++
ll	223	0,7 %	+++
ha	222	0,7 %	+
pa	216	0,7 %	++
nt	211	0,7 %	+++

Auch hier sind Bigramme der Struktur Konsonant-Vokal (z. B. |le|, |ri|) häufiger vertreten als die übrigen Strukturen (10 von 20). Anders als im deutschen Korpus ist |e| hier aber kein auffälliger Zweitbestandteil; er kommt nur in drei Bigrammen vor (|le|, |se| und |ne|). Darunter ist mit |le| allerdings auch das absolut frequenteste Bigramm (vgl. oben Tab. 4); in zwei Dritteln der Belege ist |le| wortfinal (z. B. *apple, twinkle*). Es handelt sich hier um ein Bigramm, das sich – wie in geringerem Maße auch |er| – distributionell ähnlich verhält wie ein Suffix (es ist eine relativ häufige Verbindung, die wortfinal auftritt; vgl. unten 3.2.5 und Berg 2016).

Interessant wird es, wenn wir die Listen im deutschen und englischen Korpus vergleichen. Die Schnittmenge beträgt neun Bigramme (18a). Nur im deutschen Korpus kommen die elf Bigramme in (18b) vor, nur im englischen Korpus diejenigen in (18c):

- (18a) ch, el, er, in, ll, ma, ri, se, st
 (18b) au, be, ck, de, ei, ge, ie, ka, ra, sc, te
 (18c) ar, ca, co, ea, et, ha, le, ne, nt, pa, ro

Mit der Ausnahme von |sc| im deutschen Korpus, das nur als Teil von |sch| auftritt, sehen viele der Bigramme in (18b) tatsächlich deutsch aus und nicht englisch; die in (18c) sehen wiederum eher englisch und nicht deutsch aus. Es wäre interessant zu untersuchen, ob sich die jeweilig ‚typischen‘ Bigramme in Pseudowortuntersuchungen als besonders salient für die Zuschreibung zu einer Sprache erweisen.

3.2.1.4 Anfangsränder von Stämmen

Interessant und relevant für die Beschreibung der Graphotaktik sind neben der Verteilung der Bigramme in Stämmen auch deren Ränder. Mit welchen Buchstaben können Stämme anfangen und enden, mit welchen nicht? Tabelle 14 gibt an,

wie häufig die einzelnen Buchstaben im deutschen Teilkorpus an initialer Position auftreten, in Tabelle 15 sind die entsprechenden Daten für das englische Teilkorpus aufgelistet.

Tab. 14: Anzahl der Buchstaben in stamminitialer Position im deutschen Teilkorpus in aufsteigender Reihenfolge.

#Stämme	#Stämme	#Stämme			
ß	0	j	70	l	239
x	1	i	80	h	258
ü	5	v	93	g	267
y	6	n	134	a	272
ö	8	e	135	t	296
c	13	z	135	p	346
ä	16	d	143	m	380
q	29	w	205	b	384
u	33	r	228	k	595
o	66	f	238	s	811

Tab. 15: Anzahl der Buchstaben in stamminitialer Position im englischen Teilkorpus in aufsteigender Reihenfolge.

#Stämme	#Stämme	#Stämme			
x	2	o	126	g	313
z	23	v	127	f	339
u	26	n	144	m	404
y	49	w	224	t	482
q	53	r	251	b	487
i	91	h	270	p	610
j	93	l	271	c	759
k	98	a	298	s	1.048
e	109	d	307		

Wählen wir als binäres Kriterium belegte vs. unbelegte Anfangsränder, so ist im deutschen Korpus lediglich |ß| unmöglich; im englischen Korpus sind alle Buchstaben in initialer Position belegt. Es gibt aber in beiden Korpora Buchstaben, die selten im Anfangsrand auftreten. Sehr selten ist in beiden Korpora |x|; im deutschen Korpus gibt es einen Beleg (*Xenon*), im englischen zwei (*xenon*, *xerox*). Beide Wörter sind rezente Neubildungen, die an altgriechische Wurzeln angelehnt sind (*xénos* ‚fremd‘ und *xeros* ‚trocken‘). Es spricht also einiges dafür, in

beiden Sprachen |x| zu den unmöglichen Anfangsrändern zu zählen. Für den nativen deutschen Wortschatz gehört auch |y| in diese Gruppe: Alle Belege mit initialem |y| sind Fremdwörter (z. B. *Yak, Yen, Yoga*). Für die übrigen Buchstaben gilt das nicht – auch nicht für die Umlaute im deutschen Korpus (z. B. *Ähre, Öhr, übel*). Sie sind selten belegt, aber nicht auf einen Teil des Wortschatzes beschränkt.

3.2.1.5 Endränder von Stämmen

Für die Endränder gelten offenbar mehr Beschränkungen. Tabelle 16 listet die Anzahl der deutschen Stämme auf, die mit dem jeweiligen Buchstaben enden, in Tabelle 17 geschieht dasselbe für das englische Korpus:

Tab. 16: Anzahl der Buchstaben in stammfinaler Position im deutschen Teilkorpus in aufsteigender Reihenfolge.

#Stämme		#Stämme		#Stämme	
j	0	ß	36	g	182
q	0	u	51	m	186
ä	1	p	59	k	192
ö	1	i	65	h	261
ü	1	o	66	s	266
c	2	b	73	n	504
w	3	f	119	t	566
v	9	z	128	l	608
y	10	a	142	r	672
x	25	d	159	e	1.099

Tab. 17: Anzahl der Buchstaben in stammfinaler Position im englischen Teilkorpus in aufsteigender Reihenfolge.

#Stämme		#Stämme		#Stämme	
j	0	b	102	a	313
q	0	w	125	r	325
v	7	g	144	h	342
z	16	y	194	d	343
u	32	m	213	l	451
c	42	o	213	n	599
i	65	p	226	t	854
x	71	s	252	e	1.676
f	90	k	310		

Nicht belegt sind im deutschen und englischen Teilkorpus |j| und |q|. Selten sind im deutschen Teilkorpus die Umlaute. Bei den entsprechenden Belegen für finales |ö| und |ü| handelt es sich um Fremdwörter (*Menorrhö*, *Menü*). Für finales |ä| gilt das nicht (*sä[en]*) – der Beleg ist aber durch die Abwesenheit des silbeninitialen |h| isoliert (vgl. die analogen Stämme *bläh[en]*, *kräh[en]*, *näh[en]*, *mäh[en]* und *späh[en]*). Mit dieser Einschränkung können die Umlaute zu den unmöglichen Stammenden im deutschen Korpus gezählt werden. Die Buchstaben |c|, |w| und |y| sind im deutschen Korpus in finaler Position auf den nicht-nativen Bereich beschränkt. Die wenigen Belege sind allesamt Fremdwörter (|c|: *Tic*, *chic*; |w|: *Show*, *Squaw*, *Bungalow*; |y|: *Boy*, *Baby*, *Hobby*, *Hockey*, *Jockey*, *Jury*, *Party*, *Pony*, *Spray*, *Story*).

Bei |x| ist das ähnlich; etymologisch handelt es sich bei allen Wörtern um Fremdwörter (z. B. *Box*, *Index*, *Kruzifix*). Synchron im Sinne von Eisenberg (2011) ist das aber längst nicht so klar: Was macht *Box* und *fix* zu Fremdwörtern? Phonologisch und flexionsmorphologisch sind sie unauffällig, und graphematisch ist das Vorkommen von |x| kein zwingendes Kriterium – der Buchstabe kommt schließlich auch in nativen Wörtern wie *Hexe* vor. Da viele der Wörter (wie eben *Box* und *fix*, aber auch *Jux* und *lax*) weitgehend integriert sind, wird |x| nicht als dispräferierter Endrand im nativen Bereich angesetzt. Vergleichbares gilt auch für |v|: Neben Fremdwörtern wie *naiv*, *lukrativ* sind auch native Wörter wie *brav* und *hiev[en]* im Korpus enthalten.

Im englischen Teilkorpus ist |v| sehr selten. Ein Blick auf die Liste der entsprechenden Belege zeigt, dass es sich in allen Fällen um markierte Stämme handelt: *Slav* und *Yugoslav* sind Ethnonyme; *spiv* (‚Gauner‘) und *luv* (Schreibvariante zu *love*) sind umgangssprachlich; *lav* (‚Toilette‘) und *rev* (‚Drehzahl‘) sind Kurzwörter (zu *lavatory* und *revolution*); *derv* (‚Diesel(kraftstoff)‘) ist ein Akronym (zu *diesel-engined road vehicle*). Auf dieser Grundlage kann angenommen werden, dass stammfinales |v| im englischen Korpus für Stämme im Kernbereich unmöglich ist.

|c| ist im englischen Korpus ebenfalls relativ selten. Von den 42 Wörtern enden 25 auf *-ic* (z. B. *magic*, *music*). Hier haben wir es wohl mit dem Suffix zu tun, auch wenn die entsprechenden Wörter im Sinne von 2.2 morphologisch einfach sind (vgl. auch Berg/Aronoff 2017). Die übrigen 17 Wörter sind nicht im selben Maße markiert wie bei |v| oben. Zwar gibt es hier mit *sec* auch eine Kurzwortbildung zu *second*; der Rest ist aber unmarkiert wie z. B. *arc* ‚Bogen‘, *disc* ‚Scheibe‘, *zinc* ‚Zink‘, *shellac* ‚Schellack‘, *lilac* ‚lila‘ oder *havoc* ‚Verwüstung‘. Finales |c| ist durchaus selten (außer als Teil von *-ic*) und damit ein dispräferierter Endrand (i. S. v. Berg et al. 2014: 292). Ein unmögliches Stammende ist es nicht.

Finales |i| ist im Englischen ebenfalls einer der selteneren Buchstaben. Die 65 Belege lassen sich in zwei Gruppen gliedern:

- (19a) chichi, confetti, grafitti, hadji, kepi, lathi, octroi, patchouli, scampi, vermicelli
 (19b) Farsi, Hindi, Maori, Panjabi, Siswati, Somali, Swahili, Swazi

Die Gruppe in (19a) beinhaltet 57 der 65 Wörter. Diese Wörter sind Fremdwörter im Englischen, vor allem aus dem Italienischen (*scampi, vermicelli, grafitti, confetti*), dem Französischen (*octroi, kepi, chichi, patchouli*) sowie dem Hindi (*lathi*), aber auch aus anderen Sprachen wie dem Arabischen (*hadji*) oder dem Spanischen (*chilli*). Sie alle sind synchron recht gut als Fremdwörter identifizierbar – vor allem wohl, weil sie kulturell fremde Konzepte bezeichnen. Die Belege in (19b) enthalten mit dem Ethnonym *-i* ein Suffix (vgl. Marchand 1969: 354 f.) – auch wenn der ‚Stamm‘ nicht segmentierbar ist, weil er nicht in anderen Kontexten vorkommt. Insgesamt scheinen die 65 Wörter recht deutlich als Fremdwörter identifizierbar zu sein. Daher kann finales *|i|* zumindest im englischen Kernwortschatz als unmöglich angenommen werden. Wie auch immer dieser Kernwortschatz definiert wird – die Wörter in (19) gehören wohl nicht dazu.

Mit Ausnahme von *|y|* im nativen Teil des deutschen Korpus und *|i|* im Englischen handelt es sich bislang bei allen Beschränkungen um Konsonantenbuchstaben. Vokalbuchstaben scheinen an finaler Position weniger restringiert zu sein, sowohl im Englischen als auch im Deutschen. Das liegt aber zumindest teilweise daran, dass in den Aufstellungen in Tabelle 16 und 17 nicht differenziert wird, ob ein Vokalbuchstabe alleine Silbenkern ist (wie z. B. *Alibi*) oder ob er Teil einer Verbindung (wie z. B. *drei*).

Aus dieser Perspektive sind mehr Vokalbuchstaben stammfinal restringiert. Das betrifft zum einen finales *|u|* im englischen Teilkorpus. Es ist insgesamt 32-mal belegt, und in 21 Fällen enden die Stämme mit einfachem finalem *|u|*:

- (20a) ecru, fichu, impromptu, menu, ormolu, tutu, zebu
 (20b) babu, coypu, emu, gnu, guru, haiku, Hindu, jujitsu, juju, mu, vertu
 (20c) Bantu, Urdu
 (20d) thru

Bei all diesen Belegen handelt es sich um markierte Schreibungen. Die Wörter in (20a, b) sind Fremdwörter; die meisten stammen aus dem Französischen (20a), die in (20b) aus dem Japanischen (*mu, haiku, jujitsu*), aus dem Hindi (*babu, guru, Hindu*) – und aus verschiedenen anderen Gebersprachen wie dem Italienischen (*virtu*) und dem Portugiesischen (*emu*). Außerdem gibt es mit (20c) zwei Belege für Sprachbezeichnungen und mit (20d) eine informelle Schreibvariante zu *through*. Gemein ist den Belegen in (20), dass sie nicht zum Kernwortschatz des Englischen gehören.

Ganz ähnlich verhält es sich im deutschen Teilkorpus. Hier ist finales |u| insgesamt 65-mal belegt; finales |u| nach einem Konsonantenbuchstaben aber nur elfmal.

- (21a) Guru, Iglu, Kakadu, Känguru, Kanu, Manitu, Marabu, Uhu
 (21b) du, Nu, zu

Bei den Belegen in (21a) handelt es sich (auch synchron) um Fremdwörter – bzw. bei *Uhu* um ein Onomatopoeikon. Im nativen Wortschatz findet sich einfaches finales |u| hingegen nur in der geschlossenen Klasse der Funktionswörter (21b) *du*, *zu* sowie im Substantiv *Nu*, das allerdings nur direkt nach *im* auftritt.

Neben |u| ist auch finales |i| im deutschen Korpus restringiert. Von den 65 Wörtern mit finalem |i| im deutschen Teilkorpus enthalten nur 33 ein einfaches |i|. Die Belege in (22) zeigen eine Auswahl dieser 33 Wörter:

- (22a) Alkali, Hindi, Kanari, Kolibri, Maharani, mini, Pi, Sari, Taxi, Yogi
 (22b) Ätti, Kroki, Müsli, Radi
 (22c) Gummi, Kohlrabi

Bei den Belegen in (22a) handelt es sich um Fremdwörter, vor allem aus dem Französischen (*Kanari*, *Kolibri*, *Alkali*, *Taxi*) und dem Hindi (*Sari*, *Maharani*, *Hindi*), aber auch aus anderen Sprachen. Sie sind auch synchron recht gut als Fremdwörter identifizierbar – vor allem wohl, weil sie größtenteils kulturell fremde Entitäten bezeichnen. Bei den Belegen in (22b) handelt es sich um regionalsprachliche Lexeme schweizerdeutschen und bayerischen Ursprungs – also ebenfalls nicht um Wörter des Kernwortschatzes. Das ist bei (22c) unter Umständen anders. Auch hier handelt es sich etymologisch um Fremdwörter – *Gummi* kommt aus dem Arabischen, *Kohlrabi* aus dem Lateinischen –, aber beide Wörter sind schon geraume Zeit im Deutschen, *Gummi* seit dem Mittelhochdeutschen und *Kohlrabi* seit dem Frühneuhochdeutschen, und sie bezeichnen Objekte des Nahbereichs. Das Problem löst sich, wenn man mit Eisenberg (2011) u. a. den nativen Kernbereich prosodisch definiert als Einsilber und trochäische Zweisilber mit Reduktionssilbe; dann gehören beide Wörter nicht zum nativen Wortschatz (und auch flexionsmorphologisch verhalten sie sich ja markiert). Allerdings darf die Frage gestellt werden, wie realistisch diese Modellierung ist. Immerhin geht es Eisenberg explizit darum, die Intuition nativer Sprecher über den Fremdheitsstatus von Wörtern abzubilden. Wie Sprecher *Gummi*, *Kohlrabi* und andere Wörter tatsächlich beurteilen, ist eine spannende Frage, die empirisch untersucht werden müsste. Wir halten hier vorsichtig fest: Für den Kernwortschatz ist einfaches finales |i| ein dispräferiertes Stammende.

Dann müsste allerdings auch einfaches finales |o| im Kernwortschatz ausgeschlossen werden: Wörter, die so enden (wie etwa *Konto*, *Fresko*, *Torso*), mögen trochäisch sein – über eine Reduktionssilbe verfügen sie aber nicht. Bei |a| ist das wesentlich unsicherer: Der Unterschied in der Vokalqualität der zweiten Silben von *Laster* und *Pasta* (vgl. Fuhrhop/Peters 2013: 244 f.) ist jedenfalls wesentlich geringer. Auf phonologischer bzw. prosodischer Grundlage ist daher eine Klassifikation als Fremdwort ungewiss, und |a| kann nicht ohne Weiteres zu den unmöglichen Stammenden gezählt werden. Unabhängig davon gilt: Die Graphematik markiert den Unterschied zwischen Suffix *-er* und der nah-homophonen Endung sehr eindeutig (weiter 4.4).

3.2.1.6 Zusammenfassung

Dieses Teilkapitel hat sich mit der globalen Graphotaktik beschäftigt. Dabei ging es um drei Themen:

1. Zunächst stand die Frage im Mittelpunkt, wie die Buchstaben in Vokal- und Konsonantenklassen unterteilt werden können, ohne auf die Phonologie Bezug zu nehmen. Die Lösung, die hier vorgeschlagen wurde, beruht auf der syntagmatischen Verteilung der Buchstaben. Auf der Basis der Information über jeweils direkte Nachbarn wurde eine hierarchische Clusteranalyse durchgeführt (3.2.1.1). Die Ergebnisse der Analyse wurden auch für randomisiert verkleinerte Korpora überprüft (3.2.1.2).
2. Danach ging es um die Verteilung von Bigrammen, also Paaren von Buchstaben. Hier wurde eine Reihe Verfahren angewendet, um die wesentlichen Züge der beiden Sprachen herauszuarbeiten. Zentral ist die Idee, dass nicht die absoluten Vorkommen von Bigrammen bewertet werden, sondern dass diese Häufigkeiten mit den erwarteten Häufigkeiten verglichen werden und der Unterschied zwischen beiden auf Signifikanz geprüft wird (3.2.1.3).
3. Und schließlich ging es um Stammanfangs- und Stammendründer und die Frage, welche Buchstaben hier ausgeschlossen bzw. dispräferiert sind (3.2.1.4, 3.2.1.5).

Im Folgenden werden einige wesentliche Beobachtungen dieses Teilkapitels noch einmal knapp zusammengefasst. Wenn hier vom ‚Deutschen‘ und ‚Englischen‘ die Rede ist, dann ist das natürlich immer noch nur auf das verwendete Korpus bezogen.

Die *Klassifikation von Konsonanten- und Vokalbuchstaben* funktioniert auf der Basis der syntagmatischen Verteilung sehr gut und kommt zu erwartbaren Ergebnissen. |y| wird in beiden Sprachen als Vokalbuchstabe klassifiziert, aber das entspricht, wie gezeigt wurde, auch seiner hauptsächlichen Funktion. Das Verfahren ist auch für kleinere Korpora relativ stabil.

Verdoppelungen von Buchstaben: Insgesamt zeigt sich eine weitgehende Übereinstimmung zwischen den beiden Sprachen, was die häufigen Verdoppelungen angeht. Die Liste der Konsonanten, die häufig verdoppelt werden, ist fast deckungsgleich. Die Schnittmenge umfasst |b|, |d|, |f|, |g|, |l|, |m|, |n|, |p|, |s|, |z|; im Deutschen kommt noch |t| hinzu. Bei den Vokalen zeigen sich Unterschiede. Im Englischen sind nur |e| und |o| verdoppelbar, im Deutschen auch noch |a|; die Möglichkeiten gehen im Deutschen also weiter als im Englischen. Dafür sind die beiden vokalischen Verdoppelungen im Englischen wesentlich häufiger, als das zu erwarten wäre. Im Deutschen ist nur |o| häufiger, und diese Abweichung ist nicht hochsignifikant. Die Verdoppelung von Vokalbuchstaben ist also im Englischen wesentlich zentraler als im Deutschen.

Man könnte hier auch noch einen Schritt weiter gehen und nicht nur die Verdoppelung von Buchstaben untersuchen, sondern den Skopus auf Bigramme (oder Trigramme) ausweiten. Finden wir Verdoppelungen von Bigrammen außerhalb von Wörtern wie *Mama* und *Papa*? Wenn ja, welche Bigramme kommen bevorzugt vor? Wir befinden uns dann im Bereich der graphematischen Haplogie.

Verdreifachungen von Buchstaben sind in beiden Sprachen nicht zu finden, weder bei den Vokalen noch bei den Konsonanten.

Was die *Stammanfänge und Stammenden* angeht, so ergeben sich folgende Beschränkungen im deutschen (23) und englischen Teilkorpus (24):

(23a) Unmögliche Anfänge deutscher Stämme: $\beta, (x, y)$ ³²

(23b) Unmögliche Enden deutscher Stämme: $j, q, (\ddot{a}, \ddot{o}, \ddot{u}, c, w, y, Ci, Co, Cu)$ ³²

(24a) Unmögliche Anfänge englischer Stämme: x

(24b) Unmögliche Enden englischer Stämme: $j, q, (v, i, Cu)$ ³²

Diese Aufstellung ist aus zwei Gründen interessant. Zum einen, weil die Schnittmenge zwischen beiden Sprachen recht groß ist: Unmöglicher Anfangsrand in beiden Sprachen ist |x|; unmögliche Enden sind |j|, |q|, |i|, |u|. Zum anderen sind die unmöglichen Ränder keine Funktion der Buchstabenhäufigkeit. Es ist ja stochastisch erwartbar, dass besonders seltene Buchstaben auch selten an bestimmten Positionen vorkommen. Dass auch die Vokale |i| und |u| von der Beschränkung betroffen sind, spricht gegen diese Hypothese; sie sind in beiden Sprachen häufiger als andere Buchstaben wie z. B. |b|, |g| oder |f|.

³² Im Kernwortschatz unmöglich.

3.2.2 Silbenstrukturelle Graphotaktik

Die globale Graphotaktik im vorigen Abschnitt ist in ihrem Skopus begrenzt: Es wird jeweils die lineare Kombination von zwei Elementen untersucht. Diese Sicht hat ihre Grenzen einerseits dort, wo es um größere Struktureinheiten wie bspw. die Silbe geht: Wie viele Silben haben Stämme in den beiden Teilkorpora und wie sind sie aufgebaut? Andererseits kombinieren Buchstaben und Grapheme abhängig von ihrer silbenstrukturellen Position, und das kann die globale Graphotaktik nicht erfassen. Das Bigramm |kn| tritt nur im Anfangsrand von deutschen und englischen Schreibsilben auf, das Bigramm |nk| nur im Endrand – diese Information ist in Darstellungen wie Tabelle 12 oben nicht rekonstruierbar.

Der Begriff der Silbe wird – wie die Begriffe ‚Vokal‘ und ‚Konsonant‘ oben – modalitätsübergreifend als Alternation von Vokal- und Konsonantengraphemen verstanden (siehe Primus 2003). Folgen von Vokalgraphemen werden *einem* Silbenkern zugeordnet; ‹Bauer› und ‹royal› sind also graphematisch einsilbig. Damit ergibt sich die Silbenzahl pro Stamm recht einfach; wir zählen die Silbenkerne. Die Stämme ‹Frage› und ‹make› sind zweisilbig, ‹Abenteuer› und ‹village› sind dreisilbig und so fort. Die weitaus kompliziertere Frage nach der Lage der Silbengrenzen wird hier weitgehend ausgeklammert (für verschiedene Möglichkeiten der graphematischen Silbifizierung vgl. Schmidt 2014). Für die Untersuchung der Silbenanfangs- und endränder wird stattdessen auf die Stammanfangs- und -endränder zurückgegriffen, die immer auch Silbenränder sind. In 3.2.2.5 wird dann für intervokalische Konsonantencluster in Zweisilbern geschaut, inwiefern diese Cluster mit den bisher ermittelten Regularitäten für Silbenränder beschrieben werden können. Verhalten sich intervokalische Cluster graphotaktisch eher wie Anfangs- oder wie Endränder von Stämmen?

Im letzten Abschnitt wurde stillschweigend der Begriff der Konsonanten- und Vokalgrapheme benutzt, im Abschnitt zur globalen Graphotaktik hingegen waren *Buchstaben* zentral. Die Verwendung von Graphemen statt Buchstaben ist in diesem Abschnitt sinnvoll; das soll an einem Beispiel demonstriert werden. Wenn es um die Beschreibung der Kombinatorik von Silbenkernen geht, so gibt es, wenn wir Buchstaben als relevante Einheiten verwenden, unter anderem den möglichen komplexen Kern |ua|, der immerhin 20-mal vorkommt. Schaut man sich die Belege an, so enthalten 15 von ihnen |qu|, z. B. *Qual*, *Quark*. So häufig kommt unter anderem auch die Folge |aa| vor. Die Prominenz von |ua| hängt aber natürlich am |q|; ohne es ist die Folge mit fünf Vorkommen marginal und auf den Fremdwortbereich beschränkt (z. B. *Jaguar*, *Nuance*, *Guano*). Arbeitet man hingegen mit Graphemen, ist das kein Problem – ‹qu› ist ein Konsonantengraphem, und *Qual* und *Quark* haben einen einfachen Silbenkern. In beiden Sprachen ist

⟨qu⟩ ein komplexes Graphem; ⟨qu⟩ wird in den folgenden Analysen daher als *ein* ⟨C⟩-Graphem gezählt. Im Deutschen sind darüber hinaus ⟨ch⟩ und ⟨ck⟩ im nativen Bereich Grapheme (vgl. 3.1). Im nicht-nativen Bereich kommt ⟨c⟩ auch ohne folgendes ⟨h⟩ oder ⟨k⟩ vor, deswegen wurde der Graphemstatus oben auf den nativen Bereich beschränkt. Methodisch optimal wäre es nun, wenn wir die Stämme im deutschen Korpus hinsichtlich ihres Fremdwortstatus annotieren würden, um abhängig davon ⟨ch⟩ und ⟨ck⟩ als Grapheme bzw. als Graphemfolgen zu klassifizieren. Das ist natürlich nicht machbar – es ist aber auch nicht wirklich schlimm, und zwar aus zwei Gründen: Zum einen gibt es im nicht-nativen Bereich nur wenige Vorkommen von autonomem ⟨c⟩ im Korpus (20 Belege). Und zum anderen gehören Wörter mit ⟨ch⟩ und ⟨ck⟩ ganz überwiegend zum Kernwortschatz (Ausnahmen sind z. B. *chic*, *Chlor*, *Bronchie*, *Attacke*, *Jockey*). Aus diesem Grund werden auch ⟨ch⟩ und ⟨ck⟩ im Deutschen als *ein* Graphem gewertet, unabhängig vom Fremdwortstatus der jeweiligen Stämme.

Zum phonologischen Aufbau von Wörtern unter silbischer Perspektive liegt eine ganze Reihe an Arbeiten vor (für das Deutsche z. B. Augst 1971; Meinhold/Stock 1982; Duanmu 2008; für das Englische z. B. Fudge 1969; Treiman 1986; Duanmu 2008). Aus graphematischer Perspektive ist die Forschungslage deutlich eingeschränkter. Für das Englische sind mir keine Arbeiten zur silbenstrukturellen Graphotaktik bekannt; für das Deutsche gibt es einschlägige Arbeiten von Eisenberg (1988) sowie Butt/Eisenberg (1990), die sich mit dem Aufbau von Schriftsilben beschäftigen und dabei (wenn auch eher cursorisch) auf die Graphotaktik eingehen. Fuhrhop/Buchmann (2009) untersuchen die Schriftsilbe aus formaler Perspektive und bringen die Grapheme in eine Ordnung, die das graphematische Äquivalent zur Sonoritätshierarchie ist. Indirekt relevant ist schließlich auch Augst (1986), der die graphematische Struktur morphologisch komplexer Schriftwörter analysiert. Er setzt eine maximale Struktur für einsilbige lexikalische Stämme an, die unten in 3.2.2.7 diskutiert wird.

Dieser Abschnitt ist wie folgt aufgebaut: In einer ersten Übersicht geht es um die Frage, aus wie vielen Silben Stämme aufgebaut sind und wie diese Silben (typischerweise) strukturiert sind (3.2.2.1). Dann wird untersucht, wie Anfangsrand und Endrand aufgebaut sind – welche Grapheme häufig, welche selten vorkommen und welchen Beschränkungen die Kombinatorik unterliegt (3.2.2.2–3.2.2.3). Es folgt ein Exkurs, der sich mit dem symmetrischen Aufbau dieser beiden Konstituenten beschäftigt (3.2.2.4). Die oben angesprochenen intervokalen Cluster werden aus den genannten Gründen separat behandelt (3.2.2.5), bevor der Aufbau von Silbenkernen untersucht wird (3.2.2.6). Im darauf folgenden Abschnitt werden verschiedene Interaktionen der Silbenkonstituenten geprüft (3.2.2.7) und die Ergebnisse werden zusammengefasst (3.2.2.8).

3.2.2.1 Der silbische Bau von Stämmen

Aus wie vielen Silben bestehen Stämme im deutschen und englischen Teilkorpus und wie sind diese Silben aufgebaut? Die erste Frage ist relativ leicht zu beantworten. Wie oben skizziert, werden dafür nur die Silbenkerne identifiziert; Silbenkerne sind einzelne Vokalgrapheme oder Cluster von Vokalgraphemen. Daraus ergibt sich folgende Aufstellung für die beiden Sprachen:

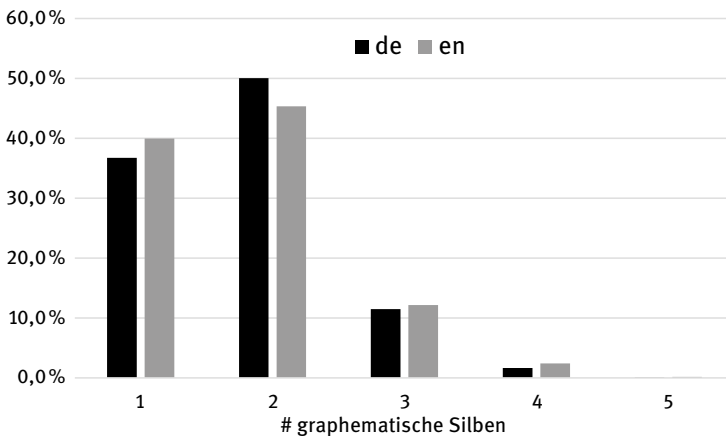


Abb. 8: Relativer Anteil der Stämme mit 1, 2, 3, 4 und 5 graphematischen Silben an allen Stämmen. Datengrundlage: Teilkorpora einfacher Stämme deutsch und englisch.

In beiden Sprachen sind zweisilbige Stämme am häufigsten; danach kommen einsilbige, dreisilbige und (marginal) viersilbige. Während die Anteile der drei- und viersilbigen Stämme bemerkenswert ähnlich sind, gibt es bei den ein- und zweisilbigen Stämmen Unterschiede. Diese sind durchaus erwartbar: Im deutschen Teilkorpus sind Zweisilber (vor allem gegenüber Einsilbern) häufiger, während der Unterschied zwischen Ein- und Zweisilbern im Englischen deutlich geringer ausfällt. Das ist mit der wichtigeren strukturellen Rolle des (graphematischen und phonologischen) Trochäus im Deutschen zu erklären, der für eine Vielzahl von Phänomenen zentral ist. Der geringere Unterschied im Englischen, so ist zu vermuten, kommt überhaupt nur zustande, weil finales <e> zur Kennzeichnung der Vokalqualität benutzt wird (z. B. in *like*, *cape*), sonst wären Einsilber die häufigste Struktur im Englischen. Da es sich bei *like*, *cape* usw. um phonologische Einsilber handelt, lässt sich diese These an den phonologischen Strukturen der einfachen Stämme überprüfen.

Wie Abbildung 9 zeigt, kehrt sich in den phonologischen Strukturen das Verhältnis um: Über 50 % aller einfachen Stämme im englischen Teilkorpus sind phonologische Einsilber.

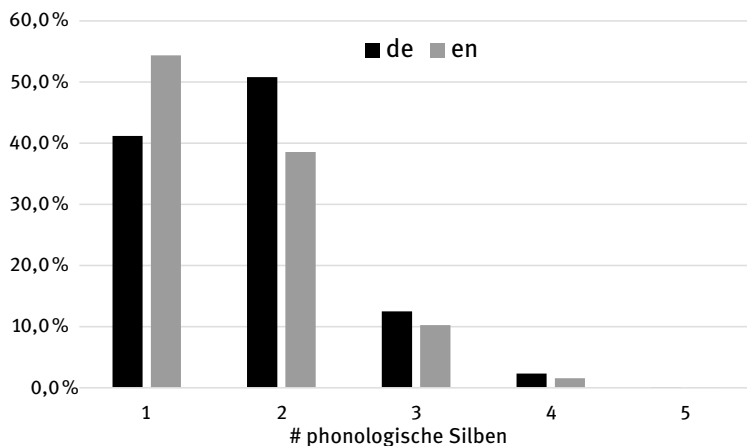


Abb. 9: Relativer Anteil der Stämme mit 1, 2, 3, 4 und 5 phonologischen Silben an allen Stämmen. Datengrundlage: Teilkorpora einfacher Stämme deutsch und englisch.

Dieser Unterschied zwischen der geschriebenen und der gesprochenen Form von Stämmen spricht für die wichtige Rolle des finalen <e> im Englischen: Obwohl phonologisch Einsilber die häufigste Struktur von Stämmen darstellen, sind es graphematisch Zweisilber.

Als Nächstes soll in beiden Korpora betrachtet werden, in welchem Verhältnis Vokalgrapheme zu allen Graphemen eines Stamms stehen, und zwar wiederum abhängig von der Silbenzahl. Dabei handelt es sich um eine sehr allgemeine Maßzahl, um das Verhältnis von Vokal- und Konsonantenbuchstaben in Stämmen zu bestimmen. Das ist interessant, weil es bereits grobe Aussagen über die Art der Silben in Stämmen erlaubt. Hat ein Stamm mehrere Silben, die einen mehrfach besetzten Anfangs- und Endrand haben, und gleichzeitig einfache Silbenkerne (wie bspw. deutsch <schüchtern>), dann ist der Wert niedrig (für *Pfingsten* liegt er bei 0,22 – zwei Vokalgrapheme auf neun Grapheme insgesamt). Hat ein Wort dagegen komplexe Silbenkerne und nur einfache Silbenränder (wie z. B. englisch *aureola*), dann ist der Wert hoch (für *aureola* liegt er bei 0,71 – fünf Vokalgrapheme auf sieben Grapheme insgesamt). Ein Wort, das nur aus ‚idealen‘ Silben im Sinne von Primus (2003) besteht (nämlich <CV>-Silben), hätte einen Wert von 0,5.

Abbildung 10 zeigt den Anteil von Vokalgraphemen an allen Graphemen eines Wortes in Abhängigkeit von der Silbenzahl in beiden Sprachen. Der Verlauf des relativen Anteils von $\langle V \rangle$ -Graphemen ist in beiden Sprachen sehr ähnlich. Der Unterschied bei den fünfsilbigen Stämmen beruht auf sehr wenigen Daten (ein Stamm im deutschen Korpus, elf im englischen) und soll deswegen nicht weiter interpretiert werden. Insgesamt ist in beiden Sprachen ein stetiger Anstieg des Anteils von $\langle V \rangle$ -Graphemen zu beobachten, von etwa 30 % bei Einsilbern auf knapp 45 % bei Viersilbern. Das ist im Großen und Ganzen erwartbar: Einsilber haben häufig besetzte Anfangs- und Endränder (vgl. 3.2.2.2 und 3.2.2.3), und häufig auch mehrfach besetzte. Bei Viersilbern hingegen kommen intervokalisch häufig einfache Konsonantengrapheme vor, und allein das führt zu höheren relativen Anteilen der $\langle V \rangle$ -Grapheme. Zwischen Ein- und Viersilbern spannt sich das Feld auf. Überraschend ist trotz der Erwartbarkeit der Ergebnisse zweierlei: einerseits die hohe Übereinstimmung zwischen den englischen und den deutschen Daten, andererseits der hohe Grad der Korrelation zwischen Silbenzahl und Anteil der $\langle V \rangle$ -Grapheme (deutsch: Pearsons $r = 0,97$, $p < 0,05$; englisch: Pearsons $r = 0,89$, $p < 0,05$).

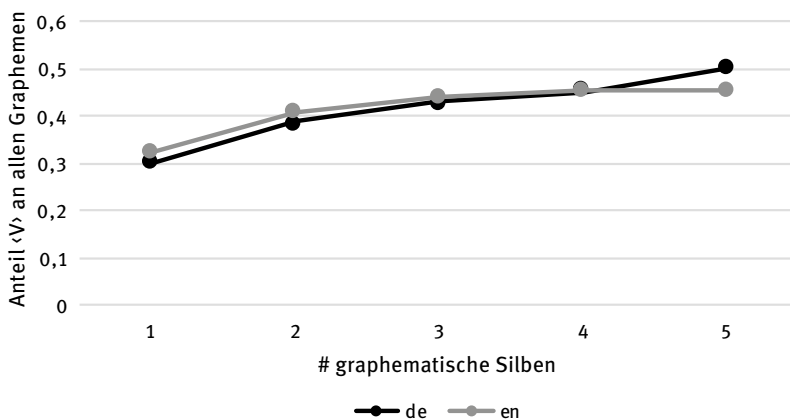


Abb. 10: Der relative Anteil von Vokalgraphemen an allen Graphemen eines Wortes, geordnet nach der Zahl der graphematischen Silben. Datengrundlage: Teilkorpora einfacher Stämme deutsch und englisch.

Der letzte Blick in diesem Abschnitt gilt den präferierten CV-Strukturen in beiden Korpora. Tabelle 18 präsentiert zu diesem Zweck für die Ein-, Zwei- und Dreisilber in beiden Sprachen die jeweils fünf häufigsten CV-Strukturen. Die Linien zwischen den Teiltabellen zeigen an, wo die jeweilige Struktur in der anderen Sprache zu finden ist.

Tab. 18: Die jeweils fünf häufigsten CV-Strukturen von Ein-, Zwei- und Dreisilbern im deutschen und englischen Teilkorpus einfacher Stämme. Die Linien zwischen den Teiltabellen zeigen gleiche Strukturen.

#Silben	deutsch		englisch	
1	CVCC	577	CVCC	613
	CCVCC	232	CCVCC	370
	CVC	173	CVC	349
	CVVC	171	CVVC	291
	CCVC	117	CCVC	261
2	CVCCVC	462	CVCCVC	426
	CVCCV	339	CVCV	375
	CVCVC	267	CVCVC	265
	CVCV	157	CVCCV	253
	CCVCCV	97	CCVCCV	205
3	CVCVCV	52	CVCCVCV	102
	CVCVCVC	50	CVCVCV	89
	CVCVCCV	43	CVCVCCV	57
	CVCCVCV	41	CVCVCVC	55
	CVCCVCVC	38	CVCCVCVC	52

Auch hier finden wir einen bemerkenswerten Grad an Übereinstimmung in den Rangfolgen. Im Ein- und Dreisilber sind die fünf häufigsten Strukturen in beiden Sprachen dieselben, im Einsilber sogar in exakt der gleichen Reihenfolge. Im Zweisilber sind vier der fünf häufigsten Strukturen in beiden Sprachen häufig. Unter Vorgriff auf die Ergebnisse in den kommenden Abschnitten lassen sich bereits folgende Beobachtungen machen: Komplexe Silbenkerne sind (mit Ausnahme von <CVVC> wie <Baum> oder <soul>) nicht unter den häufigsten Strukturen vertreten. Und komplexe Ränder und intervokalische Cluster sind ebenfalls insgesamt in der Unterzahl; Cluster aus drei Konsonanten kommen unter den häufigsten CV-Strukturen gar nicht vor.

3.2.2.2 Anfangsränder

Wie eingangs erwähnt, geht es in diesem Abschnitt nicht um Silbenanfangsränder im Allgemeinen, sondern um Stammanfangsränder im Besonderen. Das hat rein methodische Gründe; es ist eben schwierig, die Silbengrenzen in graphematischen Mehrsilbern zu bestimmen.

Betrachten wir zunächst die Komplexität der Anfangsränder in beiden Korpora. Abbildung 11 zeigt die relative Häufigkeit von leeren (0) sowie einfach bis vierfach geschlossenen Anfangsrändern:

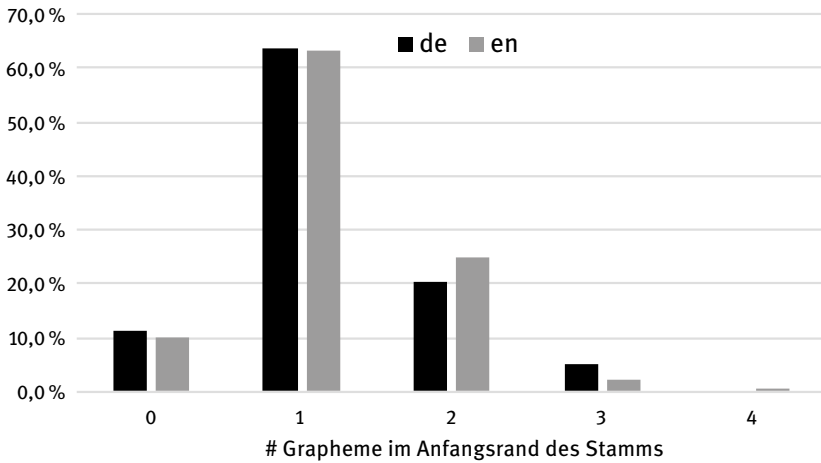


Abb. 11: Relativer Anteil der Stämme mit 0, 1, 2, 3 und 4 Graphemen im Anfangsrand an allen Stämmen. Datengrundlage: Teilkorpora einfacher Stämme deutsch und englisch.

Die Häufigkeiten sind in beiden Sprachen sehr ähnlich verteilt. Der mit Abstand frequenteste Typ ist der einfach besetzte Anfangsrand, vor dem zweifach geschlossenen und dem unbesetzten Anfangsrand. Die zweifach geschlossenen sind im englischen Korpus häufiger als im deutschen; dafür sind im deutschen Korpus die dreifach geschlossenen Anfangsränder häufiger (z. B. *Schnee*, *Schwefel*, *schrill*; zur Erinnerung: <ch> wird als ein Graphem gezählt). Es gibt insgesamt nur zwei Belege für vierfach besetzte Ränder, beide im englischen Korpus: *schnorkel* und *schwa*. Beim ersteren handelt es sich um eine graphematische Variante zu <snorkel>; zu zweiterem gibt es die Variante <shwa>. Beide Wörter kommen aus dem Deutschen (*schwa* ursprünglich aus dem Hebräischen). Vierfach besetzte Anfangsränder, so lässt sich zusammenfassen, gibt es in beiden Korpora im Prinzip nicht (mit Ausnahme extrem weniger deutscher Fremdwörter im englischen Korpus). Der maximale Anfangsrand ist in beiden Korpora also <CCC>.

Zu den Anfangsrändern im Einzelnen. Wir beginnen bei den einfach besetzten. Hier ist eher der Vergleich der Häufigkeiten interessant, mit denen die Grapheme in beiden Sprachen auftreten, als die Analyse der Verteilung in den einzelnen Sprachen. Abbildung 12 zeigt, wie hoch die relativen Häufigkeiten der Konsonantengrapheme sind, die einzeln im Anfangsrand auftreten:

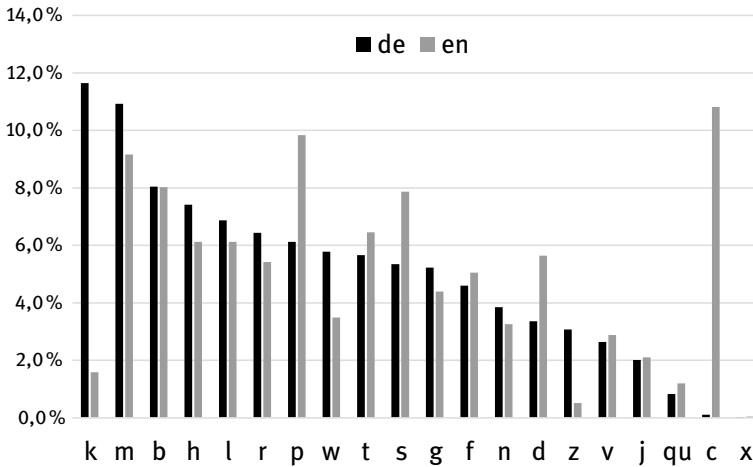


Abb. 12: Relative Anzahl der Konsonantengrapheme an allen einfach besetzten Anfangsrändern von Stämmen in den Korpora einfacher Stämme englisch und einfacher Stämme deutsch.

Besonders zwei Unterschiede fallen auf. Das betrifft einerseits ⟨k⟩: Dieses Graphem ist im deutschen Korpus der häufigste einfache Stammanfangsrand, im englischen hingegen einer der seltensten. Genau umgekehrt liegt die Situation andererseits bei ⟨c⟩: Im englischen Korpus ist es das häufigste Graphem als einfacher Stammanfangsrand, im deutschen eines der seltensten, und es kann überhaupt nur im Fremdwortbereich Graphemstatus beanspruchen (die vier Belege für einfaches initiales ⟨c⟩ sind allesamt Fremdwörter: *Code, Cello, Campus, Camp*). Beide Grapheme, ⟨c⟩ und ⟨k⟩, hängen in beiden Sprachen auch phonographisch zusammen (siehe oben 3.2.1.3); es ist also kein Zufall, dass die größten Unterschiede zwischen den Sprachen genau hier liegen. Die übrigen Unterschiede fallen demgegenüber weniger ins Gewicht. Erwähnenswert ist lediglich initiales ⟨z⟩, das im Deutschen systematisch genutzt wird, im Englischen aber stamminitial marginal ist; hier gibt es nur 23 Wörter, die mit ⟨z⟩ als einfachem Anfangsrand beginnen (z. B. *zeal, zebra, zone*).

Komplexe Anfangsränder sind theoretisch interessanter, weil hier Kombinationsbeschränkungen beobachtet werden können. Mit den 22 Konsonantengraphemen im Deutschen sind ja theoretisch 484 verschiedene ⟨CC⟩-Anfangsränder bildbar. Davon sind aber lediglich 37 tatsächlich belegt. Im Englischen kommen nur 46 der 441 theoretisch möglichen Kombinationen vor. In beiden Sprachen ist der Grad der Ausschöpfung also relativ gering. Das kann an starken Restriktionen liegen – oder an zu wenig Daten. Immerhin gibt es nur 961 Stämme mit ⟨CC⟩-Anfangsrand im deutschen Teilkorpus und 1.747 solcher Stämme im englischen. Die Wahrscheinlichkeit, dass bestimmte Kombinationen also rein zufällig nicht

vorkommen, ist durchaus real. Duanmu (2008: 68 f.) nennt das das „spotty data problem“ (übersetzbar am ehesten mit ‚Problem der punktuellen Daten‘) und demonstriert es mit folgender Rechnung: Von allen kombinatorisch möglichen (phonologischen) Zweisilbern im Englischen (etwa 6.000.000) sind nur knapp 6.000 tatsächlich belegt – also 0,1 %. Ein Teil kann auf phonotaktische Beschränkungen zurückgeführt werden – aber eben bei Weitem nicht 99,9 % der kombinatorisch möglichen Zweisilber. Duanmu bietet als Erklärung für diese Verteilung an, dass natürliche Sprachen wie Englisch oder Chinesisch jeweils knapp 10.000 Morpheme verwenden. Das bedeutet, dass nur ein Bruchteil der möglichen Formen auch verwendet wird. Das Problem ist nun, dass wir nicht wissen, ob die Auswahl dieser Formen zufällig ist oder ob sie strukturellen Prinzipien folgt (Duanmu 2008: 69). Ist das Lexikon einer Sprache zufällig oder systematisch, was die Form der Einträge anbelangt? Dieses Problem muss bei der Beschreibung der Graphotaktik mitbedacht werden; bestimmte Generalisierungen sind auf der Grundlage der Daten einfach nicht besonders stabil.³³

Im Folgenden werden zunächst komplexe Anfangsränder im deutschen Teilkorpus untersucht; den Anfang machen zweielementige Ränder. Die Kombinationen von zwei Konsonantengraphemen werden mithilfe von Kreuztabellen dargestellt (vgl. für eine ähnliche Herangehensweise Butt/Eisenberg 1990: 52 ff.). Die Kreuztabelle für das deutsche Teilkorpus hat folgende Form (siehe Tab. 19 folgende Seite).

Welche Generalisierungen sind über diese Daten möglich? Eine erste Beobachtung ist – wenig überraschend, aber theoretisch für das Folgende sehr wichtig –, dass die Kombinationen nicht symmetrisch sind. Wir finden entweder die Kombination $\langle\alpha\beta\rangle$ oder $\langle\beta\alpha\rangle$, aber nicht beides. Diese Generalisierung hat in Tabelle 19 nur eine Ausnahme: Wir finden sowohl $\langle sp\rangle$ (*Sport*, *spar[en]*) als auch $\langle ps\rangle$ (*Psi*, *Psyche*). Die letztere Abfolge ist klar die Ausnahme: Sie ist wesentlich seltener als die umgekehrte Abfolge (2 vs. 76 Stämme), und sie kommt nur in fremden Stämmen vor.

Die beobachtete Asymmetrie hat nun Folgen für die Beschreibung der graphotaktischen Möglichkeiten und Grenzen. Nehmen wir an, dass in den Daten $\langle\alpha\beta\rangle$ und $\langle\beta\gamma\rangle$ vorkommen, nicht aber $\langle\gamma\alpha\rangle$ und $\langle\alpha\gamma\rangle$. Wir können trotzdem schließen, dass $\langle\alpha\gamma\rangle$ eine mögliche Abfolge ist und $\langle\gamma\alpha\rangle$ eine unmögliche ($\langle\alpha\rangle$ kommt nur vor $\langle\beta\rangle$ vor und $\langle\beta\rangle$ nur vor $\langle\gamma\rangle$, also kommt $\langle\alpha\rangle$ nur vor $\langle\gamma\rangle$ vor, nicht danach). Das ist ein stärkeres Argument als das Nicht-Vorkommen von $\langle\gamma\alpha\rangle$, das ja, wie oben ausgeführt, zufällige Gründe haben kann.

³³ Eine Möglichkeit, dem Problem beizukommen, ist die Erhebung von psycholinguistischen Daten, wie Blevins (2010) es in ihrer kritischen Rezension von Duanmu (2008) anmerkt.

Tab. 19: Kreuzklassifikation der zweifach besetzten Anfangsränder von Stämmen im deutschen Teilkorpus.

	c	ch	f	h	j	k	l	m	n	p	qu	r	s	t	w	z
b							38					66				
ch							2					2				
d												25				
f					1		48					29				
g							26		3			56				
k				1			73		43			73				
p			23	12			42		2			40	2			
r				4												
s	1	149		3		11	3	3	1	72	1			121	1	2
t				12								85				
v							1									
w												4				
z															28	

Auf diese Weise lassen sich Abfolgeschemata für die Silbenkonstituenten aufstellen – ähnlich den Schemata, mit denen in der Phonologie zur Beschreibung der Silbenstruktur gearbeitet wird (z. B. Vennemann 1982). Die Schemata zu Anfangs- und Endrändern werden in Abschnitt 3.2.2.4 verglichen und zusammengefasst.

Deutlich zu erkennen sind in Tabelle 19 zwei bevorzugte zweite Bestandteile, nämlich ⟨l⟩ und ⟨r⟩, und ein bevorzugter erster Bestandteil, ⟨s⟩. Etwas weniger eindeutig ist ⟨h⟩ als zweiter Bestandteil, der sich immerhin noch mit fünf unterschiedlichen Graphemen verbindet (⟨k⟩, ⟨p⟩, ⟨r⟩, ⟨s⟩, ⟨t⟩). Bei den entsprechenden Wörtern handelt es sich allesamt um Fremdwörter (wie *Khan*, *Pharao*, *Rhesus*, *Show*, *Thermostat*); ⟨h⟩ ist also nur im nicht-nativen Bereich ein frequenter Zweitbestandteil. Zwischen diesen vier Graphemen (⟨s⟩ auf der einen und ⟨h⟩, ⟨l⟩ und ⟨r⟩ auf der anderen Seite) spannt sich das Feld auf, und die übrigen Grapheme finden sich zwischen diesen Extremen wieder. Die längste Kette ist

- (25) ⟨s⟩ → ⟨p⟩ (z. B. *Sport*) → ⟨f⟩ (z. B. *Pflaume*) → ⟨r⟩ (z. B. *frei*) → ⟨h⟩ (z. B. *Rheuma*).

Auf diese Weise ergibt sich für die Anfangsränder folgendes Schema:

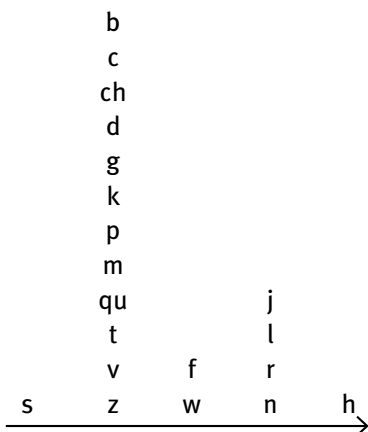


Abb. 13: Vorläufiges graphotaktisches Schema für den Anfangsrand deutscher Stämme.

Das Schema besteht aus fünf Strukturpositionen – so viele sind notwendig, das ergibt sich aus der längsten Kette. Elemente, die sich eine Strukturposition teilen, kombinieren nicht miteinander; so sind etwa die Kombinationen ⟨bc⟩ und ⟨dr⟩ im Anfangsrand nicht belegt. Zum Teil gibt es auch indirekt keine weitere Evidenz für die genaue Abfolge von zwei Graphemen (wenn die Elemente nicht Teil von Ketten sind); in diesem Fall lässt sich keine bestimmte Ordnung festlegen. Die genaue Ausgestaltung des Schemas ergibt sich also nicht zwangsläufig aus Tabelle 19. ⟨b⟩, ⟨d⟩, ⟨g⟩ und ⟨v⟩ könnten bspw. auch gemeinsam mit ⟨s⟩ an erster Stelle gruppiert werden oder mit ⟨f⟩ und ⟨w⟩ an dritter Stelle. Wir wissen nur, dass sie vor der letzten Gruppe auftreten. Abbildung 13 ist ein mögliches Schema, das nun mit den drei- und vierelementigen Anfangsrändern abgeglichen werden muss.

Bei den dreifach besetzten Anfangsrändern gibt es weniger Stämme, andererseits ist die Anzahl verschiedener Anfangsränder weit geringer. Im Deutschen kommen ⟨CCC⟩-Anfangsränder 272-mal vor, es gibt aber nur 18 Types.

Tab. 20: Dreifach besetzte Anfangsränder von Stämmen im Teilkorpus einfacher deutscher Stämme.

Anfangsrand	#Stämme	Beispiel
schl	56	Schlange, schlecht
schw	45	Schwefel, schwarz
str	41	Straße, streu[en]
schn	31	Schnee, schnell
schm	31	Schmand, schmal
schr	26	Schrift, schräg
spr	17	Sprudel, sprech[en]
pfl	8	Pflaume, pflücken
spl	6	Splitter, Spleen
pfr	2	Pfropf, Pfropfen
skr	2	Skript, Skriptum
phl	1	Phlegma
thr	1	Thron
skl	1	Sklave
sph	1	Sphinx
tsch	1	tschilp[en]
dsch	1	Dschunke
phr	1	Phrase

Die Abfolgen <thr> und <phr> kollidieren mit der Abfolge von <h> im Schema oben: <h> tritt in *Thron* nicht nach <r> und <l> auf, sondern davor. Ein Ausweg ist, <ph> bzw. <th> als *ein* Segment zu betrachten; ein anderer, dasselbe mit <rh> zu tun. Auf jede dieser Weisen ist das Problem bereits gelöst, weil es keine sich widersprechenden Abfolgen gäbe. Aber auf welcher Basis sollte die Entscheidung für die eine oder andere Lösung fallen? Entscheiden wir uns dafür, <rh> als feste Kombination anzusetzen, dann spricht dafür, dass das Inventar der Elemente kleiner ist, als wenn wir den umgekehrten Weg gehen und <ph> und <th> ansetzen. Das ist ein theorieimmanentes Argument.

Ist damit <rh> ein Graphem des Deutschen? Ja – aber eben nur, wenn man der hier dargelegten Argumentation folgt. Die Schrift behandelt <rh> als eine Einheit, und zu dieser Aussage gelangt man, wenn man a) Asymmetrie der Kombinationen, b) Konsistenz der graphotaktischen Schemata und c) ein möglichst kleines Inventar von Grundeinheiten ansetzt. Ansonsten gilt das oben unter 3.1 Dargelegte: Es ist sinnlos, nach ‚dem‘ absoluten Grapheminventar zu suchen.

Ein ähnliches Problem gibt es bei <tsch> und <psch>: Auch diese Abfolgen sind vom Schema nicht gedeckt; <s> ist im Schema das erste Element. Eine mögliche Lösung ist, <sch> ebenfalls als *ein* Segment zu klassifizieren. Dafür gibt es, wie gleich gezeigt wird, noch andere Gründe.

Mit diesen Ausnahmen sind alle Kombinationen in Tabelle 20 vom Schema oben lizenziert. Das steht im Zusammenhang mit Augsts (1986: 312) Beobachtung, dass Folgen von drei Graphemen $\langle\alpha\beta\gamma\rangle$ immer in zwei Zweierfolgen $\langle\alpha\beta\rangle$ und $\langle\beta\gamma\rangle$ zerlegbar sind, die auch als Anfangsränder attestiert sind.

Ansonsten gilt für die Cluster in Tabelle 20: a) Je seltener ein Cluster ist, desto häufiger beruht er auf Fremdwörtern. b) Wenn wir $\langle\text{sph}\rangle$, $\langle\text{tsch}\rangle$ und $\langle\text{dsch}\rangle$ ausklammern, kann Tabelle 20 zu einer Kreuztabelle vereinfacht werden:

Tab. 21: Dreifach besetzte Anfangsränder deutscher Stämme, dargestellt als Kreuztabelle (ohne $\langle\text{sph}\rangle$, $\langle\text{tsch}\rangle$ und $\langle\text{dsch}\rangle$).

	r	l	w	n	m
sch	26	56	45	31	31
st	41				
sp	17	6			
pf	2	8			
sk	2	1			
ph	1	1			
th	1				

Hier wird auf den ersten Blick deutlich, dass sich $\langle\text{sch}\rangle$ anders verhält als die übrigen ersten Elemente in Tabelle 21. Drei zweite Bestandteile ($\langle\text{w}\rangle$, $\langle\text{n}\rangle$ und $\langle\text{m}\rangle$) kommen ausschließlich nach $\langle\text{sch}\rangle$ vor. Für die zweiten Bestandteile $\langle\text{r}\rangle$ und $\langle\text{l}\rangle$ zeigt sich ein sehr hoher Grad der Ausnutzung. Obwohl viele Kombinationen nur selten vorkommen und obwohl hier native und fremde Stämme aufgenommen sind, gibt es mit $\ast\langle\text{stl}\rangle$ und $\ast\langle\text{thl}\rangle$ nur zwei Lücken in der Kombinatorik. Strukturell passt $\langle\text{sch}\rangle$ also nicht in die Reihe – die Verbindung verhält sich nicht wie ein $\langle\text{CC}\rangle$ -Cluster.

Stattdessen zeigt sich auch hier: $\langle\text{sch}\rangle$ verhält sich graphotaktisch wie *ein* Graphem. Es fügt sich problemlos in die Kreuztabelle der $\langle\text{CC}\rangle$ -Anfangsränder ein (Tab. 19 oben). Alle Letztbestandteile sind dort auch unabhängig belegt. Es ergibt sich die folgende modifizierte Kreuztabelle für $\langle\text{CC}\rangle$ -Cluster (siehe Tab. 22).

Wir haben hier also ein distributionelles Argument, $\langle\text{sch}\rangle$ als ein Graphem zu betrachten: Es verhält sich in komplexen Anfangsrändern wie ein einfaches Graphem. Das ist natürlich indirekt phonographisch motiviert: $\langle\text{sch}\rangle$ korrespondiert mit einem Phonem, und das verhält sich wie andere einfache Phoneme auch.

Tab. 22: Modifizierte Kreuzklassifikation der zweifach besetzten Anfangsränder von Stämmen im deutschen Teilkorpus.

	c	f	h	j	k	l	m	n	p	qu	r	s	t	w	z	sch
b						38					66					
ch						2					2					
d											25					1
f				1		48					29					
g						26		3			56					
k			1			73		43			73					
p		23	12			42		2			40	2				
r			4													
s	1		3		11	3	3	1	72	1			121	1	2	
t			12								85					1
v						1										
w											4					
z														28		
sch						56	31	31			26			45		

Auf der Grundlage dieser Kreuztabelle und der Überlegungen oben zum Zusammenfassen von ⟨rh⟩ und ⟨sch⟩ zu Graphemen ergibt sich das modifizierte graphotaktische Schema:

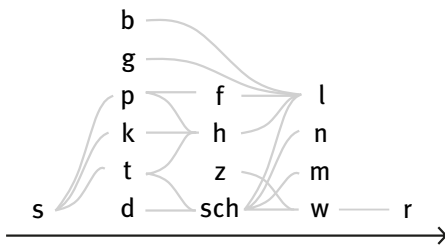


Abb. 14: Graphotaktisches Schema für den Anfangsrand deutscher Stämme.

Segmente, die durch weniger als fünf Stämme belegt sind, werden aus Gründen der Übersichtlichkeit nicht dargestellt; das betrifft im Schema oben die marginal kombinierenden Elemente ⟨rh⟩, ⟨c⟩, ⟨ch⟩, ⟨qu⟩, ⟨v⟩ und ⟨j⟩. Neben der möglichen Abfolge enthält das Schema auch Informationen über die konkreten Verbindun-

gen, die es motivieren; diese Verbindungen sind durch die grauen Linien gekennzeichnet. ⟨s⟩ kommt bspw. direkt vor ⟨p⟩, ⟨k⟩ und ⟨t⟩ vor; ⟨t⟩ wiederum vor ⟨sch⟩ usw. Das bedeutet im Umkehrschluss: Da vor ⟨b⟩, ⟨g⟩ und ⟨d⟩ kein Segment auftritt, könnten sie auch an erster Position (gemeinsam mit ⟨s⟩) angesetzt werden. Diese Information ist nützlich, wenn es weiter unten um die Frage der Vereinbarkeit der Schemata von Anfangs- und Endrand geht (siehe Abschn. 3.2.2.4).

Das Schema in Abbildung 14 lizenziert nun alle vorkommenden Kombinationen nicht-marginaler Elemente mit Ausnahme des oben diskutierten Falls ⟨ps⟩. Das Schema ist allerdings zu mächtig: Es lizenziert auch solche Kombinationen, die nicht attestiert sind, z. B. ⟨bf⟩ oder ⟨ml⟩. Zu einer vollständigen Beschreibung der Graphotaktik gehört auch eine Liste von Beschränkungen über das Schema. Hier sind aber unabhängige Daten notwendig – z. B. experimentell erhobene Wohlgeformtheitsurteile von Nicht-Wörtern. Wir wissen ansonsten nicht, wie die Lücken in den Daten zustande kommen. Sind sie zufällig oder systematisch? Die vorliegende Arbeit hört an dieser Stelle auf – das Schema des Anfangsrands wird (auch wenn es ‚übergeneriert‘) als vorläufiger Endpunkt angesetzt und nur noch mit a) dem Schema des englischen Anfangsrands und b) dem Schema des deutschen Endrands verglichen.

Im Englischen ergibt sich für die komplexen Anfangsränder – und hier zunächst für die ⟨CC⟩-Cluster – die folgende Tabelle:

Tab. 23: Kreuzklassifikation der zweifach besetzten Anfangsränder von Stämmen im englischen Teilkorpus.

	c	h	j	k	l	m	n	p	qu	r	s	t	v	w	z
b					51					82					
c		124			64					89					3
d		2	1							52				3	
f			1		69					46					
g		6			28		6			79					
k		3			1		18			3			1	1	
l					1										
p		20			63					83	6	1			
r		12													
s	65	91		28	64	22	40	79	24			123		50	
t		54								91	1			26	
w		46								24					

Auch die Daten aus dem englischen Teilkorpus lassen den Schluss zu, dass Asymmetrie ein Strukturmerkmal ist. Hier gibt es zwei Ausnahmen: Wir finden wie im deutschen Teilkorpus sowohl $\langle sp \rangle$ (*sport*) als auch $\langle ps \rangle$ (*psalm*). Darüber hinaus gibt es neben $\langle st \rangle$ (*star*) auch $\langle ts \rangle$ (*tsar*). Alle übrigen Verbindungen sind asymmetrisch.

Die längste Kette, die sich aus den einzelnen Verbindungen bilden lässt, umfasst fünf Elemente:

(26) $\langle s \rangle \rightarrow \langle t \rangle$ (*start*) $\rightarrow \langle w \rangle$ (*twig*) $\rightarrow \langle r \rangle$ (*write*) $\rightarrow \langle h \rangle$ (*rhubarb*).

Das graphotaktische Schema muss also fünf Strukturpositionen umfassen. Erfasst man so systematisch die relative Abfolge der Elemente, ergibt sich folgendes Schema:

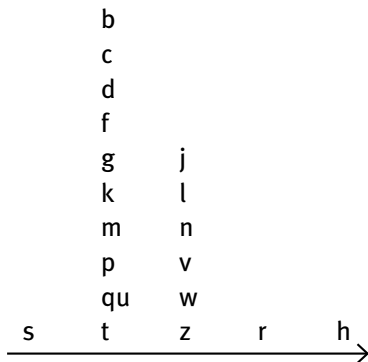


Abb. 15: Vorläufiges graphotaktisches Schema für den Anfangsrand englischer Stämme.

Es gilt wie oben bei den deutschen Daten: Dieses Schema ergibt sich nicht in jedem Punkt zwangsläufig aus Tabelle 23; es ist eine Möglichkeit. So ist z. B. die Position von $\langle b \rangle$ unklar: $\langle b \rangle$ kombiniert nicht mit $\langle s \rangle$, es könnte also auch mit $\langle s \rangle$ eine Gruppe bilden. Der Vorschlag oben trägt der Beobachtung Rechnung, dass $\langle s \rangle$ in viel stärkerem Maße als $\langle b \rangle$ als erster Bestandteil auftritt.

Nicht erfasst werden können die oben angesprochenen symmetrischen Verbindungen $\langle ps \rangle$ und $\langle ts \rangle$. Ebenfalls nicht lizenziert sind die Verbindungen $\langle pt \rangle$ (*ptarmigan*) und $\langle ll \rangle$ (*llama*), in denen jeweils Mitglieder einer Gruppe miteinander kombinieren. $\langle ps \rangle$ kommt immerhin sechsmal in den Daten vor, die übrigen Verbindungen nur einmal. Es handelt sich in allen Fällen um fremdsprachliches Material.

Passt das Schema in Abbildung 15 zu den Daten der dreifach besetzten Anfangsränder, die in Tabelle 24 aufgelistet sind? Im englischen Teilkorpus gibt es 143 Stämme, die so beginnen; sie verteilen sich auf 14 Types. Hier gibt es nun eine Reihe von Problemen. Zum einen kollidiert die Abfolge ⟨rh⟩, die wir oben im Schema angesetzt haben (aufgrund von 12 Belegen vom Typ *rhubarb*), mit den Abfolgen ⟨thr⟩ (z. B. *throw*), ⟨shr⟩ (z. B. *shrill*) und ⟨phr⟩ (z. B. *phlegm*). Zum anderen widerspricht die oben angesetzte Abfolge ⟨wh⟩ (wegen z. B. *white*) der Abfolge ⟨thw⟩ (z. B. *thwack*) (die Abfolgen ⟨tsw⟩ und ⟨khm⟩ sind ebenfalls problematisch; da es sich bei den entsprechenden Stämmen *Tswana* und *Khmer* aber um Einwohnerbezeichnungen handelt, scheint ein Ausschluss vertretbar). Und schließlich lizenziert das Schema oben die Abfolge ⟨phl⟩ nicht.

Tab. 24: Dreifach besetzte Anfangsränder von Stämmen im Teilkorpus einfacher englischer Stämme.

Anfangsrand	#Stämme	Beispiele
str	36	street, strange
scr	29	screen, scream
thr	22	throw, throne
shr	16	shrill, shrug
spr	15	spray, sprinkle
spl	10	spleen, split
sch	4	school, scheme
sph	3	sphere, sphinx
phl	2	phlegm, phlox
thw	2	thwart, thwack
phr	1	phrase
tsw	1	Tswana
chr	1	chrome
khm	1	Khmer

Wie ist zu verfahren? In allen Fällen ist die relative Anordnung von ⟨h⟩ der Ursprung des Problems. Wenn wir vermeiden wollen, dass Grapheme an mehreren Stellen im Schema auftreten, dann bleibt nur die Möglichkeit, bestimmte Kombinationen zusammenzufassen – z. B. ⟨r⟩ und ⟨h⟩ zu ⟨rh⟩. Wie oben bei der Untersuchung der Graphotaktik im deutschen Teilkorpus reicht es, entweder ⟨rh⟩ oder ⟨th⟩, ⟨sh⟩, ⟨ph⟩ anzusetzen – auf diese Weise ist das Problem bereits gelöst. Und wie oben entscheidet das Kriterium des kleineren Inventars, ⟨rh⟩ und ⟨wh⟩ als feste Kombinationen anzusetzen; hier allerdings eindeutiger als im deutschen

Teilkorpus, denn die Alternative würde zu einem wesentlich größeren Inventar führen (⟨ch⟩, ⟨dh⟩, ⟨gh⟩, ⟨kh⟩, ⟨ph⟩, ⟨sh⟩ und ⟨th⟩).

Die Überlegungen zu ⟨rh⟩ und ⟨wh⟩ führen zum modifizierten graphotaktischen Schema unten (Abb. 16). ⟨rh⟩ und ⟨wh⟩ sind nicht enthalten, weil sie nicht Teil komplexer Anfangsränder sind. Nicht alle Elemente sind in komplexen Anfangsrändern gleich stark funktionalisiert – ⟨v⟩ beispielsweise tritt überhaupt nur einmal auf, und zwar in *kvass*. Elemente mit weniger als fünf Belegen werden aus Gründen der Übersichtlichkeit nicht ins Schema aufgenommen. Im englischen Teilkorpus betrifft das ⟨j⟩, ⟨v⟩ und ⟨z⟩:

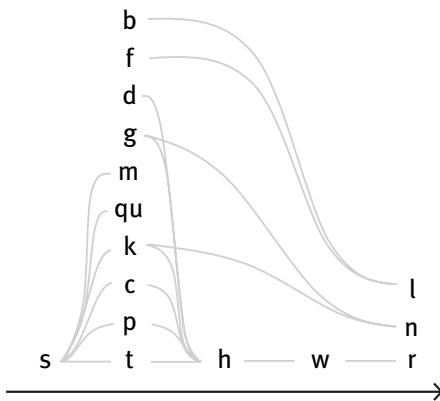


Abb. 16: Graphotaktisches Schema für den Anfangsrand englischer Stämme.

Sowohl ⟨thr⟩ als auch ⟨thw⟩ und ⟨phl⟩ sind nun vom Schema gedeckt; der Preis ist, dass ⟨rh⟩ und ⟨wh⟩ als einfache Elemente klassifiziert sind. Die grauen Linien zeigen, wie fest das jeweilige Element positional gebunden ist, indem die Verbindungen zum direkt vorangehenden und direkt folgenden Element dargestellt sind. ⟨t⟩ zum Beispiel tritt direkt nach ⟨s⟩ und vor ⟨h⟩ auf, ist also fest eingebunden. ⟨b⟩ hingegen kommt nicht nach ⟨s⟩ vor, und das direkt folgende Element ist ⟨h⟩. ⟨b⟩ könnte damit an jeder Position außer der letzten angesetzt werden.

3.2.2.3 Endränder

Wie beim Anfangsrand wird hier nicht der Endrand von Silben, sondern von Stämmen betrachtet. Wir können die Stämme in den Korpora ebenfalls danach sortieren, wie komplex ihr Endrand aussieht (siehe Abb. 17 folgende Seite).

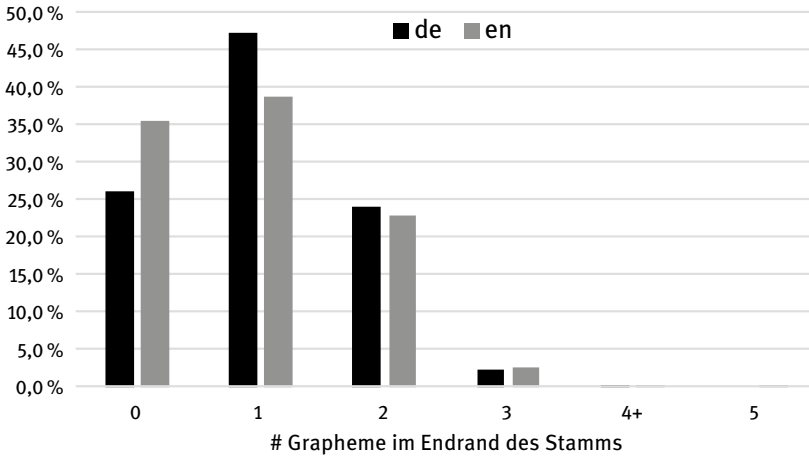


Abb. 17: Relativer Anteil der Stämme mit 0, 1, 2, 3 und 4 Graphemen im Endrand an allen Stämmen. Datengrundlage: Teilkorpora einfacher Stämme deutsch und englisch.

In beiden Teilkorpora ist ein einfacher geschlossener Endrand am häufigsten, gefolgt von einem offenen Endrand und einem zweifach geschlossenen; die übrigen Kategorien sind marginal. Ein fünffach geschlossener Endrand kommt im deutschen Korpus nicht, im englischen einmal vor (*borscht*). Vierfach geschlossene Endränder treten im deutschen Korpus achtmal auf (z. B. *Angst*, *herrsch[en]*, *selbst*, *Herbst*), im englischen viermal, und alle Belege sind Fremdwörter (*angst*, *borsch*, *dirndl*, *kirsch*) – und zwar mit einer Ausnahme (*borsch*) aus dem Deutschen. Der maximale Endrand hat im deutschen Korpus also vier Segmente, im englischen hingegen drei.

Die Anzahl der Grapheme im Endrand ist abhängig von der Silbenzahl. Generell gilt: Je weniger Silben der Stamm hat, desto mehr Grapheme sind im Endrand. Abbildung 18 zeigt die durchschnittliche Anzahl von Graphemen im Endrand für Stämme mit einer, zwei, drei und vier oder mehr graphematischen Silben (siehe Abb. 18 folgende Seite).

In einsilbigen Stämmen sind die Endränder mit durchschnittlich eineinhalb Graphemen besetzt; dieser Wert sinkt in beiden Korpora mit zunehmender Silbenzahl kontinuierlich. Im englischen Korpus liegen die Werte für die mehrsilbigen Stämme konstant unter denen im deutschen Korpus; mehrsilbige Stämme im englischen Korpus enthalten also durchschnittlich weniger Grapheme im Endrand als ihre Äquivalente im deutschen Korpus.

Wir können uns nun anschauen, wie die Endränder konkret besetzt sind. Beginnen wir bei den einfach besetzten Endrändern. Sie sind im deutschen und englischen Teilkorpus wie folgt verteilt (siehe Abb. 19 folgende Seite).

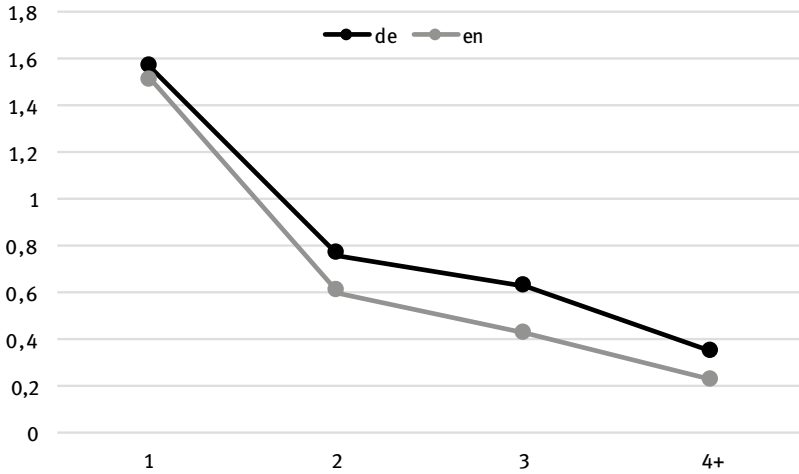


Abb. 18: Durchschnittliche Anzahl von Graphemen im Endrand deutscher Stämme nach der Anzahl graphematischer Silben dieser Stämme.

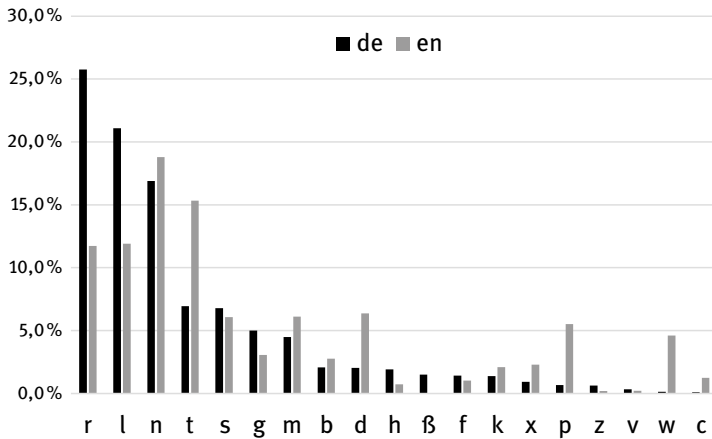


Abb. 19: Relative Anzahl der Konsonantengrapheme an allen einfach besetzten Endrändern von Stämmen in den Korpora einfacher Stämme englisch und einfacher Stämme deutsch.

Die größten Unterschiede gibt es bei ⟨l⟩ und ⟨r⟩ (diese Endränder sind im deutschen Teilkorpus wesentlich häufiger als im englischen) sowie bei ⟨t⟩, ⟨d⟩, ⟨p⟩ und ⟨w⟩ (hier ist es umgekehrt, diese Endränder sind im englischen Korpus häufiger).

Zu ⟨l⟩ und ⟨r⟩: Diese Grapheme treten im deutschen Teilkorpus häufig als Bestandteil der Stammenden ⟨el⟩ und ⟨er⟩ auf. 78% der Wörter mit einfachem

Bei den zweifach geschlossenen Endrändern fängt es an, graphotaktisch interessant zu werden – wie bei den zweielementigen Anfangsrändern oben. Wir beginnen hier mit den Daten aus dem deutschen Teilkorpus, die als Kreuztabelle dargestellt werden. Verdoppelungen von Graphemen sind grau hinterlegt (siehe Tab. 25).

Als Inventar wird hier das oben unter 3.1 aufgestellte Grapheminventar angesetzt und nicht dasjenige, das sich aus der Graphotaktik der Anfangsränder ergibt. Es ist schließlich interessant zu wissen, ob sich Anfangs- und Endränder auch in ihrem Inventar unterscheiden. Die Frage, ob und wie die Inventare vereinbar sind, wird später diskutiert (3.2.2.4, 3.2.2.8).

Die Kombinationen in Tabelle 25 sind größtenteils symmetrisch; es gibt nur neun unmittelbare Ausnahmen zu dieser Beobachtung. Da die Konsistenz des graphotaktischen Schemas wichtig ist, müssen wir uns jeweils dazu entscheiden, eine Abfolge zu ignorieren:

- Wir finden sowohl ⟨sk⟩ (*Kiosk*) als auch ⟨ks⟩ (*Keks*). Systemhafter ist sicher ⟨ks⟩, denn ⟨s⟩ kommt als zweiter Bestandteil relativ häufig vor, als erster hingegen kaum. ⟨sk⟩ wird also ignoriert.
- Neben ⟨sch⟩ (*Fisch*) kommt auch ⟨chs⟩ (*Lachs*) vor. Strukturell spricht hier einiges dafür, dass ⟨chs⟩ besser ins System passt als ⟨sch⟩: ⟨s⟩ ist ein relativ häufiger zweiter Bestandteil und kommt nach anderen Graphemen vor (siehe unten). Im Anfangsrand hat sich ⟨sch⟩ als Graphem ergeben; das wird im Endrand übernommen. ⟨sch⟩ kombiniert im deutschen Teilkorpus nur als Zweitbestandteil, und zwar nach ⟨r⟩, ⟨l⟩, ⟨m⟩, ⟨n⟩, ⟨b⟩ und ⟨t⟩.
- Neben ⟨lm⟩ (*Helm*) kommt auch einmal ⟨ml⟩ vor (*Kreml*). Diese Abfolge ist deutlich seltener, und es handelt sich um ein Fremdwort – daher wird ⟨ml⟩ im Folgenden ausgeschlossen.
- Neben ⟨cht⟩ (*Fracht*) kommt auch ⟨tch⟩ (*Match*) vor. ⟨cht⟩ passt deutlich besser ins System (⟨t⟩ tritt als zweiter Bestandteil nach fast allen Graphemen auf) und ist wesentlich häufiger. Beide Belege für ⟨tch⟩ sind Anglizismen (*Match*, *Sketch*). Die Abfolge wird daher im Folgenden ignoriert.
- Die übrigen fünf Ausnahmen ähneln einander strukturell und werden deswegen gemeinsam diskutiert. Es geht um:
 - ⟨md⟩ (*fremd*) vs. ⟨dm⟩ (*widm[en]*);
 - ⟨mt⟩ (*Samt*) vs. ⟨tm⟩ (*atm[en]*);
 - ⟨nch⟩ (*Mönch*) vs. ⟨chn⟩ (*rechn[en]*);
 - ⟨nf⟩ (*fünf*) vs. ⟨fn⟩ (*äufn[en]*);
 - ⟨ng⟩ (*Ring*) vs. ⟨gn⟩ (*regn[en]*).

In allen Fällen sind ⟨m⟩ oder ⟨n⟩ involviert, und alle Fälle beinhalten einen Verbstamm. Offenbar verhalten sich die Endränder von Verbstämmen nicht wie die Endränder anderer Stämme; hier sind z. T. andere Kombinationen

möglich. Diese Kombinationen sind aber in gewisser Weise markiert. Das sieht man z. B. daran, dass die Stämme in Derivationen mit konsonant-initialen Suffixen (wie *-bar* oder *-lich*) ein ⟨e⟩ zwischen den beiden betreffenden Graphemen aufweisen (⟨*rechenbar*, *Rechenschaft*⟩, *⟨*rechnbar*, *Rechnschaft*⟩). Die ⟨e⟩-lose Variante ist nicht immer unmöglich, aber demgegenüber deutlich markiert (z. B. ?⟨*ordnbar*⟩ gegenüber ⟨*ordenbar*⟩). Auf dieser Grundlage – und weil ⟨*n*⟩ und ⟨*m*⟩ ansonsten kaum an zweiter Position vorkommen – schließen wir die Abfolgen ⟨*dm*⟩, ⟨*tm*⟩, ⟨*nch*⟩, ⟨*fn*⟩ und ⟨*gn*⟩ im Folgenden aus.

Mit diesen Ausschlüssen ergeben sich aus der Kreuztabelle die vier bevorzugten ersten Bestandteile ⟨*h*⟩, ⟨*l*⟩, ⟨*m*⟩, ⟨*n*⟩ und ⟨*r*⟩ (wobei ⟨*h*⟩ und ⟨*m*⟩ am eingeschränktesten kombinieren) und die drei bevorzugten zweiten Bestandteile ⟨*s*⟩, ⟨*t*⟩ und ⟨*z*⟩. Zwischen diesen beiden Gruppen spannt sich das Feld auf. Die längste Kette umfasst acht Glieder:

- (27) ⟨*h*⟩ → ⟨*r*⟩ (*wahr*) → ⟨*l*⟩ (*Kerl*) → ⟨*m*⟩ (*Röteln*) → ⟨*ch*⟩ (*Mönch*) → ⟨*s*⟩ (*Lachs*) → ⟨*t*⟩ (*List*) → ⟨*sch*⟩ (*Kitsch*)

Als (vorläufiges) graphotaktisches Schema ergibt sich damit aus der Kreuztabelle:

				b				
				ch				
				ck	d			
			m	g	f			
		l	n	k	s		z	
h	r	v	w	p	x	t	sch	→

Abb. 20: Vorläufiges graphotaktisches Schema für den Endrand deutscher Stämme.

Wie oben gilt: Das Schema ergibt sich nicht zwangsläufig, es ist eins von mehreren möglichen. Problematisch sind ⟨*ph*⟩, ⟨*sh*⟩ und ⟨*th*⟩, die nicht ins Schema passen. ⟨*h*⟩ ist in Verbindung mit ⟨*r*⟩ belegt und ⟨*p*⟩, ⟨*s*⟩ und ⟨*t*⟩ sind nach ⟨*r*⟩ belegt; ⟨*p*⟩ muss also im Schema nach ⟨*h*⟩ erscheinen. Die Verbindung ⟨*sh*⟩ ist nur mit einem rezenten Fremdwort belegt, *push[en]*, und wird daher ausgeschlossen. Auch ⟨*ph*⟩ und ⟨*th*⟩ sind jeweils nur einmal belegt (*Kalligraph*, *Monolith*); allerdings handelt es sich hier um Verbindungen, die auch im Anfangsrand auffällig sind (vgl. oben 3.2.2.2). Es spricht einiges dafür, diese Verbindungen im Endrand zu Graphemen zu erklären.

Nicht erfasst werden mit dem Schema oben die Verdoppelungen von Graphemen, die im deutschen Teilkorpus bei wesentlich mehr Graphemen zu beobachten sind als im englischen. Stamminfinale Verdoppelungen sind bis auf wenige Ausnahmen morphologische Schreibungen – im Mehrsilber erscheinen intervokalisches Doppelkonsonanten (die phonographisch zu motivieren sind), und diese Doppelkonsonanten werden auf die Stammform übertragen (bzw. sind im Fall der Verbstämme bereits Teil der Stammform).

Graphotaktisch ist allein von Interesse, welche Grapheme verdoppelt werden und welche nicht. Nicht verdoppelt werden die komplexen Grapheme <ch>, <ck> und <qu> sowie die einfachen Grapheme <d>, <h>, <j>, <k>, <v>, <w> und <x>. Bis auf <d> kommen diese Grapheme grundsätzlich nicht verdoppelt vor, also auch nicht intervokalisches (z. B. *Widder*). Graphotaktisch muss es bei dieser extensionalen Definition bleiben, die gleichsam die Möglichkeiten der Verdoppelung aufzeigt. Die konkrete Verdoppelung oder Nicht-Verdoppelung ergibt sich dann morphologisch unter Bezug auf den korrespondierenden Mehrsilber.

Im nächsten Schritt muss nun überprüft werden, ob das Schema auch die dreifach geschlossenen Endränder abbilden kann. Im deutschen Teilkorpus finden sich folgende Kombinationen:

Tab. 26: Dreifach besetzte Endränder von Stämmen im Teilkorpus einfacher deutscher Stämme.

Endrand	#Stämme	Beispiel
mpf	15	Dampf, impf[en]
nst	12	Dienst, sonst
rst	11	First, berst[en]
nft	5	Vernunft, sanft
tzt	3	jetzt, letzt
lst	3	Schwulst, Wulst
rkt	2	Infarkt, Markt
pst	2	Papst, beschwipst
rft	2	Werft
hnd	2	ahnd[en]
nkt	2	Punkt
bst	2	Obst, nebst
chts	2	angesichts, nichts
rcht	1	Furcht
rrh	1	Katarrh
sst	1	bewusst
rtz	1	Hertz
nks	1	links
rnt	1	ernt[en]
rdn	1	ordnen

Endrand	#Stämme	Beispiel
nscht	1	erwünscht
rrsch	1	herrschr[en]
ndt	1	bewandt
mpt	1	prompt
rzt	1	Arzt
rks	1	Murks
lps	1	rülps[en]
nnt	1	gesinnt
nth	1	Hyazinth

Zehn der 26 Types sind nicht vom Schema oben gedeckt. Das betrifft die folgenden Fälle:

- Mit <rrsch>, <ssst> und <nnt> liegen drei Doppelkonsonantschreibungen vor, die wie gesagt von morphologischen Beschränkungen geregelt werden. Darüber hinaus handelt es sich in zwei Fällen zumindest formal um Partizipien (*bewusst*, *gesinnt*), deren Status als morphologisch einfache Einheit fragwürdig ist. Die Fälle sind für das Schema also unproblematisch (der dritte Fall ist der Verbstamm *herrschr[en]*). Interessant ist in diesem Zusammenhang noch eine andere Beobachtung: Doppelkonsonanten treten mit Ausnahme der oben genannten drei Kombinationen nicht mit anderen Konsonanten auf. Das lässt sich phonographisch aus ihrer Funktion als Silbengelenkschreibung ableiten.
- Auch bei <nscht> handelt es sich formal um ein Partizip (*erwünscht*); dieser Fall ist für das Schema morphologisch einfacher Stämme also ebenfalls nicht problematisch.
- Bei <rdn> handelt es sich um einen Verbstamm; es gilt das oben zu Verbindungen wie <tm>, <dm> etc. Geschriebene. Auch dieser Fall ist unproblematisch.
- <rrh> und <nth> enthalten <h>. Für <ph> und <th> wurde auf der Basis der zweifach besetzten Endränder eine Klassifikation als Grapheme vorgenommen; für <rh> gilt dasselbe.
- In <tzt> und <rzt> ist die Abfolge <zt> das Problem. Im Schema oben wurde <tz> angesetzt, und zwar auf Basis von Stämmen wie *Witz*, *Blitz*, *hetz[en]*. Diese Fälle können aber als morphologische Schreibungen interpretiert werden. Distributionell verhält sich <tz> wie ein Doppelkonsonant: Es ist stamfinal immer auf eine mehrsilbige Form bezogen (*Witze*, *Blitze*, *hetzen*) und tritt dort intervokalisches auf, ohne dass davor oder danach ein weiterer Konsonant steht. Die drei Ausnahmen sind genau die Stämme mit <tzt> oben. Wenn wir aber <tz> als morphologisch motivierte Schreibung ansetzen, dann muss das Schema sie nicht lizenzieren und kann stattdessen die Abfolge <zt> abdecken.

Wenn wir allerdings ‹tz› auf dieser Grundlage ausschließen, dann muss das auch für ‹ck› gelten, das sich distributionell gleich verhält.

- Schließlich ist auch ‹chts› nicht vom Schema abgedeckt: Die Abfolge ‹ts› kollidiert mit der angesetzten Abfolge ‹st›. In beiden Fällen (*angesichts*, *nichts*) sind die Stämme auf verwandte Stämme beziehbar (*Angesicht*, *nicht*), sodass es sich nicht um uneingeschränkt morphologisch einfache Stämme handelt. Die Abfolge ‹chts› wird daher im Folgenden ignoriert.

Vierelementige Endränder sind im deutschen Teilkorpus sehr selten. Es gibt nur sechs Token, die sich auf vier Types verteilen – darunter auch das in der phonologischen Diskussion beinahe notorische *Herbst* (z. B. Wiese 2000: 48; Eisenberg 2013a: 99 etc.):

Tab. 27: Vierfach besetzte Endränder von Stämmen im Teilkorpus einfacher deutscher Stämme.

Endrand	#Stämme	Beispiel
ngst	3	Angst, Hengst
lbst	1	selbst
rnst	1	ernst
rbst	1	Herbst

Alle Belege enthalten finales ‹st›, und alle sind vom Schema oben lizenziert. Mit den Modifikationen hinsichtlich ‹ph›, ‹th›, ‹rh› und der relativen Abfolge von ‹t› und ‹z› ergibt sich folgendes Bild (Elemente, die insgesamt weniger als fünfmal belegt sind, sind der Übersichtlichkeit halber nicht mit aufgenommen; das betrifft ‹v›, ‹w›, ‹rh›, ‹th›):

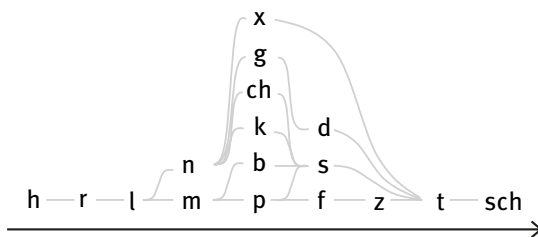


Abb. 21: Graphotaktisches Schema für den Endrand deutscher Stämme.

Im Gegensatz zum Schema des Anfangsrandes, das mit fünf Positionen auskam, ist das Schema des Endrands im deutschen Teilkorpus deutlich komplexer. Man

Zunächst fällt auf, dass stammfinale Verdoppelungen von Graphemen im englischen Teilkorpus sehr viel seltener vorkommen als im deutschen. Nur drei Grapheme werden häufig verdoppelt, ⟨b⟩, ⟨l⟩ und ⟨s⟩. Die selten verdoppelten Grapheme ⟨b⟩, ⟨d⟩, ⟨g⟩ können unabhängig über die Beschränkung der minimalen Stämme motiviert werden (vgl. unten 3.2.3): Sie helfen in den betreffenden Wörtern, die Mindestlänge von Inhaltswörtern zu erreichen (*ebb*, *add*, *odd*, *egg*). ⟨v⟩ kommt nur einmal vor – als Bestandteil von ⟨rv⟩ (*deriv*). Wir haben oben gesehen, dass ⟨v⟩ ein unmögliches Stammende im englischen Korpus ist; daher wird die Abfolge ⟨rv⟩ im Folgenden nicht weiter betrachtet.

Nicht-symmetrisch sind im englischen Teilkorpus vier Paare, für die jeweils eine Entscheidung getroffen werden muss:

- Sowohl ⟨ng⟩ (z. B. *sing*) als auch ⟨gn⟩ (z. B. *sign*) sind in Tabelle 28 zu finden. ⟨ng⟩ ist mehr als fünfmal häufiger als ⟨gn⟩ – aber als randständig kann ⟨gn⟩ mit elf Vorkommen auch nicht bezeichnet werden. Gegen ⟨gn⟩ spricht aber darüber hinaus, dass ⟨n⟩ als erster Bestandteil besser etabliert ist als ⟨g⟩ (⟨g⟩ kommt außerhalb von ⟨gn⟩ noch zweimal als erster Bestandteil vor, ⟨n⟩ hingegen fünfmal). Interessant ist in diesem Zusammenhang auch, dass ⟨gn⟩ fast ausschließlich nach ⟨i⟩ als Kern auftritt (die einzige Ausnahme ist *oppugn*); für ⟨ng⟩ gilt diese Beschränkung nicht.³⁴
- In Tabelle 28 ist neben ⟨lt⟩ (*cult*) auch ⟨tl⟩ vertreten – allerdings nur genau einmal (*axolotl*). Diese Abfolge ist gegenüber der umgekehrten Abfolge klar markiert und wird daher ausgeschlossen.
- Die übrigen drei Paare enthalten jeweils ⟨s⟩ und werden deshalb zusammen diskutiert: Es geht um ⟨sc⟩, ⟨sp⟩ und ⟨st⟩. Mit Ausnahme von ⟨sc⟩ sind die Kombinationen mit ⟨s⟩ als Erstbestandteil jeweils wesentlich häufiger. Für ⟨sc⟩ gibt es zwei Belege (*disc*, *mollusc*), für ⟨cs⟩ ebenso (*antics*, *statistics*) – Letztere sind aber auffällig: *antics* („Frassen, Grimassen“) ist ein Plural, und das -s in *statistics* ist eher ein Derivationssuffix (vgl. *statistic*, *statistical* sowie die Namen anderer Disziplinen wie *mathematics*, *genetics*; siehe auch Bauer/Lieber/Plag 2013: 124). Die drei hier diskutierten Abfolgen mit ⟨s⟩ als Zweitelement werden also ausgeschlossen.

³⁴ Bei mindestens einem Eintrag mit ⟨gn⟩, *sign*, wird oft von einer morphologischen Schreibung ausgegangen, die analog zur Morphemkonstanz im Deutschen funktioniert: Im Mehrsilber schreiben wir phonographisch *signal*, und die Graphemfolge des Stamms bleibt im Einsilber intakt (vgl. z. B. Carney 1994: 41, weiter dazu 4.1).

Die längste Kette im englischen Teilkorpus umfasst acht Elemente:

(28) ⟨r⟩ → ⟨l⟩ (*girl*) → ⟨m⟩ (*film*) → ⟨n⟩ (*autumn*) → ⟨s⟩ (*lens*) → ⟨c⟩ (*disc*) → ⟨t⟩ (*act*)
→ ⟨h⟩ (*mouth*).

In Tabelle 28 sind ⟨l⟩, ⟨m⟩, ⟨n⟩, ⟨r⟩ und ⟨s⟩ als häufige erste Bestandteile zu erkennen sowie ⟨t⟩ (und in geringerem Maße ⟨h⟩) als häufiger zweiter Bestandteil (für ⟨s⟩ gilt das nach Abzug der drei oben diskutierten Kombinationen nicht mehr).

Wir kommen auf der Grundlage dieser Überlegungen sowie der Kreuztabelle oben zu folgendem (vorläufigen) graphotaktischen Schema für Endränder im englischen Teilkorpus:

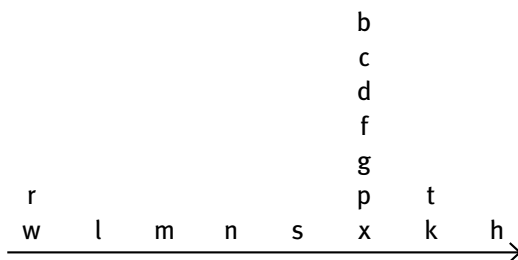


Abb. 22: Vorläufiges graphotaktisches Schema für den Endrand englischer Stämme.

Mit diesem Schema kann der Großteil der Kombinationen in Tabelle 28 erfasst werden. Für sechs Kombinationen gilt das allerdings nicht:

- Die Abfolge ⟨gm⟩ beruht auf den Einträgen *paradigm* und *phlegm*. Diese zwei Belege sind relativ isoliert, wenn man die Kette ⟨m⟩ → ⟨n⟩ (Typ *autumn*, 8x) → ⟨g⟩ (Typ *sing*, 57x) betrachtet. Die Abfolge ⟨gm⟩ wird daher ausgeschlossen.
- Für ⟨sm⟩ gilt dasselbe: ⟨m⟩ tritt vor ⟨n⟩ auf, das wiederum vor ⟨s⟩ auftritt. Dieser Abfolge widerspricht ⟨sm⟩. Das Problem ist nicht zu lösen, ohne dass eine weitere Position am rechten Rand mit ⟨m⟩ besetzt wird; diese Lösung wird weiter unten diskutiert, bis dahin wird ⟨sm⟩ ausgeklammert.
- Die Kombination ⟨bs⟩ beruht auf dem pluralisch verwendeten *plebs*³⁵ und wird daher ausgeschlossen: Es ist wahrscheinlich, dass das finale -s zumindest synchron als Pluralmarker interpretiert wird.

³⁵ Das wird an der Subjekt-Verb-Kongruenz deutlich (vgl. z. B. *But the plebs have their dialectical revenge*, Oxford English Dictionary [OED] Online, www.oed.com, Stand: 22. 11. 2018).

- Die Kombination <tz> beruht auf dem Eintrag *Blitz*, einer Kurzwortbildung zu *Blitzkrieg*. Da es sich hier offensichtlich um eine rezente Entlehnung handelt, wird die Abfolge ausgeschlossen.
- Die Abfolgen <hl>, <hm> und <hn> kollidieren mit der Position von <h> im Schema, die sich wiederum aus dem Vorkommen nach <s>, <p>, <t> und anderen Graphemen ergibt. Die entsprechenden Einträge sind entweder ursprünglich Eigennamen (*john*, *ohm*) oder Fremdwörter, die selbst im Englischen mit seinem heterogenen Wortschatz deutlich als solche identifizierbar sind (*buhl*, *kohl*). Alle drei Abfolgen werden daher nicht weiter betrachtet.

Wie verträgt sich das Schema oben nun mit dreifach besetzten Endrändern? Von diesen gibt es im englischen Teilkorpus 183 Tokens, die sich auf 30 Types verteilen:

Tab. 29: Dreifach besetzte Endränder von Stämmen im Teilkorpus einfacher englischer Stämme.

Endrand	#Stämme	Beispiel
tch	43	fetch, watch
nch	34	bench, pinch
ght	32	light, ought
rch	16	arch, church
rth	9	earth, north
lch	6	belch, squelch
nth	6	absinth, month
mph	5	lymph, triumph
rst	3	first, thirst
rsh	2	harsh, marsh
lth	2	filth, wealth
thm	2	algorithm, rhythm
mpt	2	prompt, tempt
rtz	2	hertz, quartz
rps	2	corps, turps
rrh	2	catarrh, myrrh
nct	2	sacrosanct, succinct
lph	1	syph
lsh	1	welsh
lpt	1	sculp
lst	1	whilst
ntz	1	chintz
rg	1	burgh
chm	1	drachm
rld	1	world
cht	1	yacht

Endrand	#Stämme	Beispiel
lct	1	mulct
wst	1	frowst
ltz	1	waltz
lft	1	delft

Elf der 30 Types passen nicht ins Schema. Im Einzelnen sind das:

- ⟨tch⟩ und ⟨cht⟩: Die Abfolgen sind symmetrisch. ⟨tch⟩ ist allerdings wesentlich häufiger (43x) als ⟨cht⟩, das nur einmal vorkommt (*yacht*). ⟨cht⟩ wird also ausgeschlossen. Hier ergibt sich allerdings ein neues Problem: Die Abfolge ⟨tch⟩ kollidiert mit der frequenten Abfolge ⟨ct⟩ (*act*). Die Lösung ist, ⟨ch⟩ zum Graphem zu erklären; das kann im Schema *nach* ⟨t⟩ auftreten, das einfache ⟨c⟩ hingegen davor.
- ⟨gh⟩ und ⟨rrh⟩ passen ebenfalls nicht ins Schema – ⟨gh⟩ kollidiert mit der häufigen Abfolge ⟨th⟩, und ⟨rrh⟩ (*catarrh*) enthält einen Doppelkonsonanten und ⟨h⟩. Beide Probleme lassen sich beheben, indem ⟨gh⟩ und ⟨rh⟩ als Graphem angesetzt werden.
- Drei Abfolgen enthalten ⟨tz⟩: ⟨rtz⟩, ⟨ntz⟩ und ⟨ltz⟩. Bei den entsprechenden Wörtern handelt es sich entweder um deutsche Eigennamen (*hertz*), Entlehnungen aus dem Deutschen (*quartz*, *waltz*)³⁶ oder aus dem Hindi (*chintz*). Die drei Abfolgen werden daher ausgeschlossen.
- Die Abfolge ⟨rps⟩ lässt sich nicht integrieren, weil ⟨sp⟩ relativ gut belegt ist (z. B. *crisp*, *lisp*). Und zumindest eines der Wörter, in denen ⟨rps⟩ vorkommt, ist eine Kurzwortbildung, in der -s als Derivationsmorphem interpretiert werden kann (*turps* zu *terpentine*; das andere Wort ist *corps*). Die Abfolge wird daher ausgeschlossen.
- ⟨thm⟩ und ⟨chm⟩ lassen sich nicht integrieren, ohne ⟨m⟩ zu wiederholen. Sie stehen damit in einer Reihe mit den oben diskutierten ⟨gm⟩ und ⟨sm⟩ (und mit den ähnlich aufgebauten Verbstämmen im Deutschen wie *atm[en]*, *öffn[en]*). Die Tatsache, dass es eine Reihe solcher Fälle gibt, spricht für eine gewisse Systematik. Wir tragen dem Rechnung, indem ⟨m⟩ als (einziges) extrasilbisches Element ins Schema aufgenommen wird.

Folgt man diesen Überlegungen, dann sieht das Schema für den Endrand im englischen Teilkorpus wie folgt aus (mit ⟨ch⟩ und ⟨gh⟩ als Graphemen und extrasilbi-

³⁶ *Waltz* geht interessanterweise auf deutsch *Walzer*/**Waltzer* zurück; die Schreibung mit ⟨tz⟩ scheint dort aber nie etabliert gewesen zu sein, wie eine Abfrage im Deutschen Textarchiv zeigt.

schem <m>; Elemente, die auf weniger als fünf Belegen beruhen, sind ausgeschlossen; das betrifft hier nur <rh>):

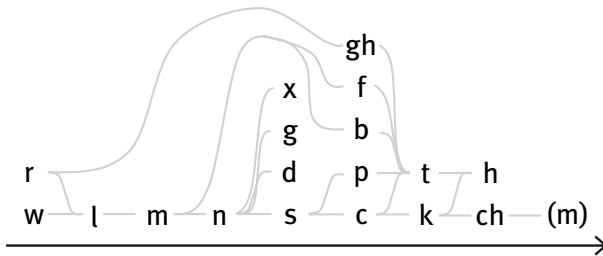


Abb. 23: Graphotaktisches Schema für den Endrand englischer Stämme.

Das Schema wird im nächsten Abschnitt mit dem Anfangsrand im englischen Teilkorpus sowie mit den Schemata des deutschen Teilkorpus verglichen. An dieser Stelle soll nur kurz das Vorgehen reflektiert werden. Das Aufstellen des Endrandschemas im englischen Korpus war schwieriger, weil es mehr (unmittelbar oder über Umwege) konfligierende Abfolgen gibt. Diese Probleme wurden relativ ad hoc gelöst, indem jeweils eine Abfolge (begründet) ignoriert wurde. Das geschah z. T. auf der Basis der morphologischen Struktur (*statistics* ist nicht uneingeschränkt morphologisch einfach), z. T. aufgrund des Status als Fremdwort (*blitz* ist als deutsche Entlehnung gut erkennbar). Dieses Vorgehen ist natürlich selektiv – wir haben eben nur über den Fremdwortstatus der problematischen Einträge entschieden. Es ist trotzdem vertretbar: zum einen, weil es sich bei den Ausnahmen allein quantitativ schon genau darum handelt – um Ausnahmen; zum anderen, weil die übrigen Fremdwörter (wie bspw. *sikh*) sich gut integrieren und unauffällig sind.

3.2.2.4 Exkurs: Symmetrie der Anfangs- und Endränder

In der Phonologie ist seit Langem die Beobachtung bekannt, dass Silbenanfangs- und -endränder symmetrisch aufgebaut sind: Im Silbenanfangsrand treten bspw. Plosive vor Liquida auf (z. B. in *treten*), im Silbenendrand danach (z. B. in *Art*). Das ist die Grundlage für die Aufstellung der Sonoritätshierarchie: Statt Anfangs- und Endrand einzeln zu spezifizieren, gibt es ein gemeinsames Strukturprinzip (das Allgemeine Silbenbaugesetz), das auf einer gemeinsamen Ordnung der Elemente aufbaut (der Sonoritätshierarchie). Es stellt sich die Frage, ob das – rein distributionell – auch in der Graphematik möglich ist. Dazu stellen wir jeweils die Stammanfangs- und Endränder im deutschen und englischen Teilkorpus gegenüber und überprüfen, inwieweit die Schemata miteinander verträglich sind. Das

Ergebnis wird mit der Längenhierarchie (Fuhrhop/Buchmann 2009, 2016) verglichen, die auf formalen Kriterien beruht.

Die Längenhierarchie baut auf den Arbeiten von Primus (2004, 2006) auf, in denen sie zeigt, a) dass die Minuskeln des lateinischen Alphabets systematisch in kleinere Bestandteile zerlegbar sind, b) dass jeweils ein Subsegment als Kopf des Buchstabens fungiert (das jeweilige Komplement nennt sie in Anlehnung an Brekle 1994 ‚Koda‘) und c) dass die Subsegmente regelmäßig mit phonologischen Merkmalen korrelieren. Fuhrhop/Buchmann (2009, 2016) gehen nun einen Schritt weiter und benutzen die Länge des Kopfs, um eine Längenhierarchie aufzustellen (analog zur Sonoritätshierarchie in der Phonologie). Sie stellen fünf Längensklassen auf: 1. Buchstaben mit langen Köpfen (|b|, |d|, |k|, |h|, |t|, |f|, |p|, |q|, |j|, |g|); 2. Buchstaben mit schrägem Kopf (|z|, |x|, |v|, |w|, |s|); 3. Buchstaben mit kurzem geradem Kopf mit oberem Koda-Anschluss (|n|, |m|, |r|, |l|); 4. Buchstaben mit kurzem geradem Kopf ohne oberen Koda-Anschluss (|i|, |u|) und 5. Buchstaben mit kurzem gebogenem Kopf (|a|, |e|, |o|). Von Klasse 1 bis Klasse 5 nimmt die Länge des Kopfs ab und die Kompaktheit nimmt zu. Für die konsonantischen Schemata des Anfangs- und Endrands sind die Klassen 1–3 relevant, also Konsonanten mit langen, mit schrägen und mit kurzen Köpfen. Ein potenzielles Problem ist, dass Fuhrhop/Buchmann Aussagen über Buchstaben treffen, nicht aber über Grapheme (wie es hier in diesem Abschnitt geschieht). Doch was ist die Länge von <ch>? Wir kommen auf dieses Problem zurück.

Beginnen wir mit den deutschen Daten. Die Schemata für den Anfangs- und Endrand sehen wie folgt aus:

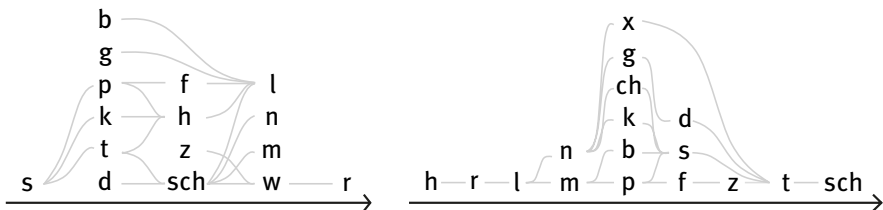


Abb. 24: Graphotaktische Schemata für den Anfangs- und den Endrand deutscher Stämme.

Was lässt sich hier generalisieren? Wir gehen von innen nach außen, also vom Kern zu den Rändern, und versuchen, Gemeinsamkeiten festzuhalten.

- <h> ist im Anfangsrand an dritter Position. Es ist hier allerdings marginal: Wenn wir wie im Endrand <ph> und <th> als Grapheme ansetzen, bleibt noch ein Beleg mit <kh> (*Khan*). Im Endrand ist <h> strukturell wesentlich stärker verankert und hier auf die erste Position beschränkt, tritt also direkt nach dem Silbenkern auf (und interagiert damit, vgl. unten 3.2.2.7).

- Abgesehen von ⟨h⟩ ist ⟨r⟩ das Graphem, das in beiden Schemata am nächsten am Kern steht.
- Danach treten in beiden Schemata ⟨l⟩, ⟨n⟩, ⟨m⟩ sowie im Anfangsrand ⟨w⟩ auf. Dort sind die Elemente ungeordnet. Im Endrand tritt ⟨l⟩ vor ⟨n⟩ und ⟨m⟩ auf.
- Ab hier werden die Gemeinsamkeiten auf den ersten Blick rarer. Dennoch ist das meiste miteinander vereinbar – mit zwei Ausnahmen. ⟨f⟩ kommt im Anfangs- und im Endrand nach ⟨p⟩ vor (*Pfand* – *Napf*). Dasselbe gilt für ⟨st⟩ (*Stahl* – *Last*). Diese Verbindungen sind also nicht symmetrisch; mehr noch, die jeweils symmetrischen Verbindungen sind nicht belegt (*⟨fp⟩, *⟨ts⟩ im Anfangs- oder Endrand). Wir können diese Asymmetrie nutzen und sie zu einem Kriterium der Graphemdefinition machen; dazu weiter unten mehr.
- Abgesehen von ⟨pf⟩ und ⟨st⟩ sind die Abfolgen größtenteils vereinbar, und zwar mit einer Modifikation: ⟨sch⟩ steht im Anfangsrand an dritter Position wegen des Vorkommens nach ⟨d⟩ und ⟨t⟩. Es handelt sich hier um je einen Beleg, und zwar ein (etymologisches) Fremdwort (*Dschungel*) und ein Onomatopoetikon (*tschilp[en]*). Ohne diese beiden Belege könnte ⟨sch⟩ auch an erster oder zweiter Position erscheinen; es ist sonst nie Zweitbestandteil.
- ⟨z⟩ tritt anders als im Endrand im Anfangsrand nicht als zweites Element auf; es könnte im Schema also auch an erster oder zweiter Stelle stehen. Damit gibt es keine Abfolge mehr, die im Anfangs- und Endrand gleich ist – außer den problematischen Verbindungen ⟨pf⟩ und ⟨st⟩. Der Endrand ist insgesamt differenzierter; die Kategorisierung des Anfangsrandes widerspricht dieser feineren Aufteilung aber nicht.

Das bedeutet: Das Schema des Endrandes gilt – mit dem Einschluss von ⟨w⟩ in die Gruppe ⟨n⟩, ⟨m⟩ und mit der Ausnahme von ⟨st⟩ und ⟨pf⟩ – auch für den Anfangsrand. Wie verhält sich nun dieses generelle distributionelle Abfolgeschema zur Längenhierarchie?

Beschränkungen über die Abfolge von Buchstaben können – analog zum Allgemeinen Silbenbaugesetz in der Phonologie – mithilfe des Allgemeinen Graphematischen Silbenbaugesetzes (im Folgenden AGS) erfasst werden. Das AGS besagt: „Die Länge der Köpfe nimmt kontinuierlich zum Silbenkern hin ab, erreicht im Silbenkern mit der Kompaktheit ihr Minimum und nimmt dann wieder zu“ (Fuhrhop/Peters 2013: 28). Problematisch für das AGS ist nun zum einen – genau wie für eine einheitliche Abfolge der Elemente – das initiale ⟨s⟩ im Anfangsrand. Fuhrhop/Buchmann (2016: 362 f.) bezeichnen es daher vorsichtig als ‚extrasyllabisch‘. Zum anderen widerspricht das silbenkernnahe ⟨h⟩ dem AGS in Stämmen wie *Sohn* oder *ahn[en]*. Fuhrhop/Buchmann (2016) bieten hier zwei Erklärungen an: Zum einen stellen sie fest, dass ⟨h⟩ vor allem im Endrand von

Verbstämmen auftritt. Diese Verbstämme kommen nun häufig als zweisilbige flektierte Wortformen vor (z. B. *ahnen*) – hier ist kein Verstoß gegen das AGS mehr zu verzeichnen. In einsilbigen Formen (wie z. B. *ahnt*) hingegen deute ⟨h⟩ „auf morphologische Komplexität hin“ (Fuhrhop/Buchmann 2016: 364); das ⟨h⟩ signalisiert, dass es „zweisilbige flektierende Formen“ (ebd.) gibt, die dem AGS entsprechen. Das ⟨h⟩ verweist in den entsprechenden Wortformen damit „auf eine flektierende Klasse“ (ebd.). Problematisch an dieser Erklärung sind die zahlreichen Ausnahmen: Wenn ⟨h⟩ in *Sohn* die Zugehörigkeit zu einer flektierenden Klasse signalisiert, warum geschieht das dann nicht in *Strom* (vgl. auch Evertz 2014: 59)? Hier ist die zweite angebotene Erklärung wesentlich überzeugender: Das ⟨h⟩ dient in Fällen wie *Sohn* als graphematische ‚Schweremarkierung‘. Fuhrhop/Buchmann (2016) greifen hier einen alten Vorschlag von Roemheld (1955) auf, der vermutet, dass Wörter, die auf einfaches ⟨l⟩, ⟨n⟩, ⟨r⟩ und ⟨m⟩ enden, aus mindestens vier Buchstaben bestehen müssen. Das tut *Strom*, nicht aber das phonographisch mögliche **Son*. Es handelt sich hier um eine rein graphematische Wohlgeformtheitsbeschränkung; sie wird weiter unten bei der Frage nach der Interaktion der Silbenkonstituenten (in diesem Fall von Anfangsrand und Endrand) wieder aufgegriffen (3.2.2.7). Insgesamt, so lässt sich schließen, ist das graphotaktische Schema des Deutschen mit dem AGS vereinbar.

Die beiden Schemata, die auf den englischen Daten beruhen, sind in Abbildung 25 gegenübergestellt:

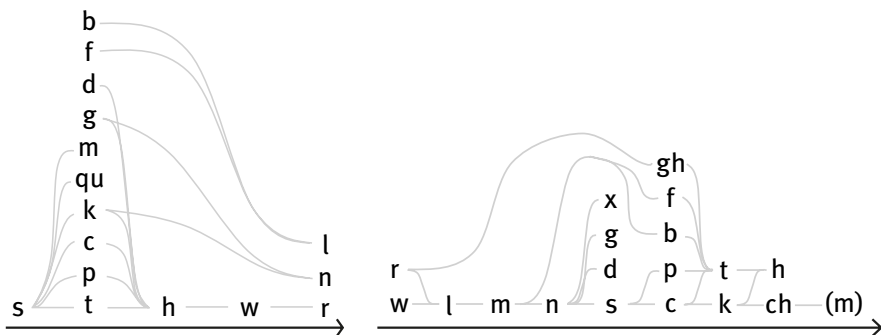


Abb. 25: Graphotaktische Schemata für den Anfangs- und den Endrand englischer Stämme.

Auch hier gehen wir von innen nach außen vor und überprüfen den Grad der Übereinstimmung; wir ignorieren dabei bewusst, dass die Grapheminventare leicht voneinander abweichen.

- ⟨r⟩ ist in beiden Schemata am kernnächsten. ⟨w⟩ tritt im Anfangsrand direkt vor ⟨r⟩ auf; das widerspricht der Abfolge im Endrand nicht.

- Im Endrand folgen dann ⟨ɫ⟩ vor ⟨m⟩ vor ⟨n⟩; auch diese Teilfolge widerspricht der Anordnung im Anfangsrand nicht. Hier sind ⟨ɫ⟩ und ⟨n⟩ in einer Gruppe mit ⟨r⟩, kombinieren aber nicht miteinander oder mit ⟨w⟩. ⟨m⟩ ist an zweiter Position angesetzt, kombiniert aber weder mit ⟨h⟩, ⟨w⟩, ⟨ɫ⟩, ⟨n⟩ oder ⟨r⟩; auch hier widersprechen sich die Schemata nicht.
- Problematisch ist die Position von ⟨s⟩. ⟨s⟩ ist sowohl im Anfangs- als auch im Endrand initial vor ⟨sc⟩ (*scar – disc*), ⟨sp⟩ (*span – crisp*) und ⟨st⟩ (*star – rest*). Wie bei den deutschen Schemata ist das Problem nicht ohne Weiteres zu lösen. Anders als im Deutschen müssten wir hier drei Grapheme ansetzen (⟨sc⟩, ⟨sp⟩ und ⟨st⟩) – oder eine von beiden Abfolgen anders motivieren, etwa als extrasyllabisches Element.
- Die übrigen Elemente sind im Anfangsrand in einer Gruppe, im Endrand in mehreren. Entscheidend ist, dass sie im Anfangsrand nicht miteinander kombinieren; die Schemata widersprechen sich also nicht.

Inwiefern ist diese Abfolge mit dem AGS vereinbar? Wie schon erwähnt und wie in den deutschen Daten ist initiales ⟨s⟩ problematisch (z. B. in *speak*, *star*); man kann auch hier dem Vorschlag von Fuhrhop/Buchmann (2016) folgen und es als extrasyllabisch klassifizieren. Darüber hinaus ist postvokalisches ⟨w⟩ (vor allem vor ⟨ɫ⟩ oder ⟨n⟩) nicht mit dem AGS vereinbar: In Stämmen wie *crawl* oder *lawn* steigt die Länge nach dem Kern an und sinkt dann wieder. Dieser Verstoß wird von Fuhrhop/Buchmann (2016) nicht aufgelistet. Phonographisch ist ⟨w⟩ hier ein Allograph von ⟨u⟩, der nur in zweiten Bestandteilen von Schreibdiphthongen auftritt (vgl. Berg/Fuhrhop 2011). Das erklärt natürlich nicht, warum wir nicht **crawl*, **lawn* schreiben – wir können aber beobachten, dass ⟨w⟩ und ⟨u⟩ vor ⟨ɫ⟩ und ⟨n⟩ annähernd komplementär verteilt sind (vgl. Fuhrhop/Buchmann/Berg 2011: 287 ff.); das ist ein graphemisches Argument für die Nähe von ⟨w⟩ zu ⟨u⟩. Wir können an dieser Stelle Folgendes festhalten: ⟨w⟩ tritt im Englischen postvokalisch vor ⟨ɫ⟩ und ⟨n⟩ auf und führt zu Verletzungen des AGS – das ist genau die Verteilung von ⟨h⟩ im Deutschen, die ebenfalls das AGS verletzt. Anders als im Deutschen ist aber postvokalisches ⟨w⟩ nicht von der Komplexität des Anfangsrandes abhängig.

Postvokalisch kommt ⟨h⟩ im Englischen nicht vor. Im Anfangsrand kombiniert es allerdings mit ⟨c⟩, ⟨g⟩, ⟨p⟩, ⟨r⟩, ⟨s⟩, ⟨t⟩ und ⟨w⟩ (z. B. in *chip*, *ghost*, *phalanx*, *rhubarb*, *ship*, *thorn*, *wheel*). Die Kombinationen ⟨rh⟩ und ⟨wh⟩ sind oben als komplexe Grapheme angesetzt worden (siehe 3.2.2.2). Die Kombination mit ⟨s⟩ ist unproblematisch, weil ⟨s⟩ eh als extrasyllabisch angesetzt werden muss. Unproblematisch für das AGS sind außerdem die Kombinationen ⟨gh⟩, ⟨ph⟩ und ⟨th⟩ – hier kombinieren zwei Buchstaben derselben Längensklasse; das ist kein Verstoß gegen das AGS. Für ⟨c⟩ gilt das nicht; ⟨c⟩ verletzt aber auch in Kombination mit

anderen Elementen das AGS, vgl. z. B. *clear, cream, disc*. Fuhrhop/Buchmann (2016: 366 ff.) schlagen vorsichtig vor, dass es sich bei <c> um einen kopflosen Buchstaben handeln könnte, der – anders als die übrigen Buchstaben – nicht weiter segmentierbar ist. Ein Argument dafür ist, dass <k> als Kombination eines langen Kopfes und <c> analysiert werden kann (Primus 2006: 19); ein weiteres Argument ist die phonographische Beobachtung, dass <c> und <k> in einer komplexen allographischen Beziehung stehen (Fuhrhop/Buchmann/Berg 2011: 283 ff.; siehe auch oben 3.2.1.3). Auch in anderen Sprache steht <c> in einer solchen Beziehung zu anderen Graphemen – im Französischen z. B. mit <qu> (vgl. Fuhrhop/Berg einge.).

Und schließlich verletzen auch einige der in den letzten beiden Abschnitten ausgeklammerten Sequenzen das AGS: Neben den marginalen Verbindungen sind das vor allem <sm> und <gn> im Endrand (etwa in *sarcasm, prism* bzw. *sign, align*). In beiden Fällen können die Verletzungen phonographisch motiviert werden: In *sarcasm* ist das finale <m> phonologisch silbisch, und damit korrespondiert die Abnahme der Länge. In *sign* modifiziert <g> die Vokalqualität und hat selbst keinen segmentalphonologischen Korrespondenten. Man könnte hier den Gedanken von Fuhrhop/Buchmann (2016) fortsetzen, dass Verstöße gegen das AGS morphologisch motiviert sind: Verstöße gegen das AGS sind morphologisch oder silbisch-phonographisch motiviert; sie zeigen entweder morphologische Komplexität oder eine Abweichung von den ‚üblichen‘ Phonem-Graphem-Korrespondenzen.

Fassen wir zusammen: Die meisten Kombinationen in Anfang- und Endrändern im deutschen und englischen Korpus sind symmetrisch. Schwächer formuliert: Die meisten Kombinationen sind zumindest nicht dezidiert asymmetrisch. Wir haben also in beiden Sprachen eine sehr vergleichbare Situation, und auch die Verletzungen (ob wir sie nun als Verletzungen der Symmetrie der Schemata oder eines Allgemeinen Graphematischen Silbenbaugesetzes fassen) weisen starke Übereinstimmungen auf.

Methodisch haben wir für die Anfangs- und Endränder jeweils beobachtet, dass die allermeisten Verbindungen nicht symmetrisch sind: Es tritt bspw. nur <rk> im deutschen Endrand auf, nicht auch <kr>. Dieses Prinzip hat nur wenige Ausnahmen (etwa im Englischen <gn>, <ng> in *sing* und *sign*). Man könnte nun einen Schritt weiter gehen und das Prinzip auch auf den Vergleich von Anfangs- und Endrändern anwenden. Die Frage ist dann: Welche Kombinationen von Graphemen kommen im Anfangs- und im Endrand *in gleicher Abfolge* vor? Ein Beispiel für eine solche Abfolge ist <st> im Deutschen: Diese Kombination kommt im Anfangsrand (z. B. *Stahl*) und im Endrand vor (z. B. *Last*). Hier verletzt <s> im Anfangsrand das AGS, wie wir oben gesehen haben – es lohnt sich aber, festzuhalten, dass das nicht zwangsläufig der Fall sein muss. Die Kombination <pf> im

Deutschen kommt ebenfalls im Anfangsrand (z. B. *Pfau*) und im Endrand vor (z. B. *Kopf*) – und dennoch können die Schemata oben (Abb. 24) das erfassen.

Ein paar der gleichen Abfolgen sind aufgrund ihrer silbenstrukturellen Graphotaktik vorläufig als komplexe Grapheme klassifiziert worden; das gilt für <sch> im Deutschen und <gh> im Englischen. Man könnte hier überlegen, Nicht-Symmetrie zu einem heuristischen Kriterium bei der Ermittlung des Graphembestands zu machen. Das ist wie folgt motiviert: Wir beobachten, dass Anfangs- und Endränder grundsätzlich symmetrisch aufgebaut sind; bestimmte Verbindungen verstoßen aber dagegen; diese Verbindungen könnten daher ‚fest‘ sein in dem Sinne, dass ihre Bestandteile nicht den allgemeinen kombinatorischen Prinzipien unterliegen (vgl. z. B. für denselben Vorschlag in der Frage, ob /pf/ eine phonologische Affrikate des Deutschen sei, Becker 1953: 253, oder Fudge 1969: 278 zu /st/ im Englischen).

Eine solche Herangehensweise führt im deutschen Korpus zu den komplexen Graphemen in (29a), im englischen Korpus zu denen in (29b):

(29a) ch, pf, ph, sch, st, th

(29b) ch, gh, ph, rh, sc, sh, sp, st, th

Die Listen enthalten zum größten Teil die üblichen Verdächtigen, wenn es um komplexe Grapheme geht; zum Teil sind die Verbindungen nur selten belegt (<ph> und <th> kommen im deutschen Korpus im Endrand nur einmal vor). <st> ist allerdings eine Überraschung: Diese Verbindung wird üblicherweise nicht als Graphem angesetzt. Hier wird im Deutschen auch sehr schön der Abstand von der Phonologie und einer phonographisch basierten Graphemdefinition deutlich: Im Anfangsrand korrespondiert <st> mit /ft/, im Endrand hingegen mit /st/. Beide Phonemkombinationen sind in der jeweils anderen Silbenkonstituente nicht belegt – phonographisch spricht also alles gegen die Annahme von <st> als Graphem. Für <pf> im Deutschen gilt das weniger – die korrespondierende Affrikate /pf/ ist ebenfalls nicht symmetrisch, und ihr Status als Phonem des Deutschen ist seit Langem Gegenstand von Diskussionen (vgl. z. B. den Überblick in Werner 1972: 50 ff.).³⁷

³⁷ Tatsächlich ist <pf> (nicht aber <st>) im Grapheminventar von Eisenberg (1988) enthalten; Eisenberg begründet die Aufnahme von <pf> damit, dass nur <f>, nicht aber <p> mithilfe der Minimalpaarmethode ersetzbar sei (*Pfahl* – *prahl* funktioniert, aber nicht *Pfahl* – *_fahl*). Das Kriterium der Minimalpaare wird in 3.2.4 auf potenzielle komplexe Grapheme angewendet.

3.2.2.5 Intervokalische Konsonanten(cluster)

Die silbenstrukturelle Graphotaktik der Konsonantengrapheme hat sich bis hierher darauf beschränkt, Stammanfangs- und -endränder zu betrachten. Sind die Ergebnisse auch auf stamminterne Konsonanten und Konsonantencluster übertragbar? Das soll hier im Folgenden untersucht werden. Wir beschränken uns dabei auf einzelne Konsonanten und Konsonantenverbindungen, die zwischen zwei Vokalgraphemen in graphematischen Zweisilbern stehen.

Das Minimum ist per Definition ein Konsonant; das Maximum sind im deutschen Teilkorpus vier Konsonanten (z. B. *Zwetschge*, mit <ch> als einem Graphem), genau wie im englischen (z. B. *minstrel*). Abbildung 26 zeigt, wie die Anzahl der intervokalischen Konsonanten in beiden Korpora verteilt ist:

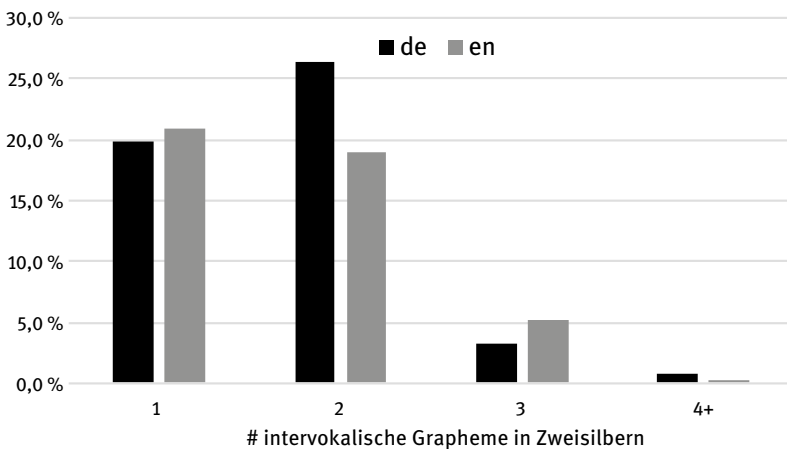


Abb. 26: Relativer Anteil der Stämme mit 0, 1, 2, 3 und 4 intervokalischen Graphemen an allen graphematisch zweisilbigen Stämmen. Datengrundlage: Teilkorpora einfacher Stämme deutsch und englisch.

Im englischen Teilkorpus sind einfache intervokalische Konsonanten der häufigste Fall, im deutschen Teilkorpus sind es Verbindungen aus zwei Graphemen; dafür sind Dreier-Cluster im englischen Korpus leicht häufiger als im deutschen. Ähnlich wie bei den Anfangs- und Endrändern sind aber Dreier-Cluster insgesamt randständig; für Verbindungen aus vier Konsonantengraphemen gilt das noch mehr. Im Folgenden werden daher nur einfache intervokalische Konsonanten sowie anschließend <CC>-Cluster betrachtet.

Bei den einfachen Konsonantengraphemen gibt es eine ganze Reihe von Abweichungen zwischen den englischen und deutschen Daten, wie in Abbil-

dung 27 zu sehen ist. In den englischen Daten sind ⟨ch⟩ und ⟨ck⟩ nicht vertreten – das hat den Grund, dass sie dort nicht als Grapheme klassifiziert wurden und entsprechend hier nicht auftauchen. ⟨ß⟩ kommt natürlich ebenfalls nur im Deutschen vor. In den englischen Daten sind ⟨r⟩, ⟨n⟩, ⟨l⟩, ⟨p⟩ sowie in besonderem Maße ⟨v⟩ und ⟨c⟩ häufiger als in den deutschen Daten. ⟨v⟩ ist der dritthäufigste einfache intervokalische Konsonant nach ⟨r⟩ und ⟨n⟩ (z. B. in *never, fever, seven*), und ⟨v⟩ ist in diese Position häufiger als im Anfangs- und Endrand. Umgekehrt sind in den deutschen Daten ⟨d⟩, ⟨t⟩, ⟨s⟩, ⟨g⟩, ⟨b⟩, ⟨f⟩ und ⟨h⟩ deutlich häufiger als in den englischen Daten. Das war in dieser Deutlichkeit eigentlich nur für ⟨h⟩ zu erwarten, das als silbeninitiales ⟨h⟩ im Deutschen, nicht aber im Englischen funktionalisiert ist (z. B. *sehen, Ruhe*). Der höhere Anteil von ⟨r⟩, ⟨n⟩ und ⟨l⟩ im englischen Korpus ist gewissermaßen die Kehrseite dieser Medaille: Intervokalisch kommen einfaches ⟨r⟩, ⟨n⟩ und ⟨l⟩ im Deutschen seltener vor, weil genau in dieser Umgebung oft (aber eben nicht immer) postvokalisches ⟨h⟩ auftritt (vgl. *Sohle, Bahre, Sahne*; zur Verteilung von ⟨h⟩ weiter 3.2.2.7).

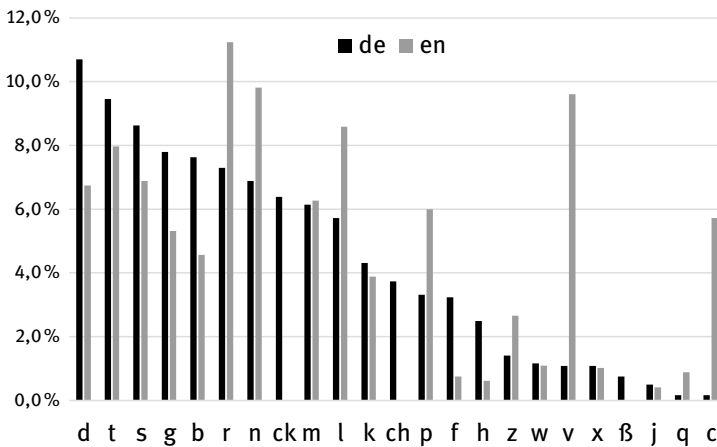


Abb. 27: Relative Anzahl der Konsonantengrapheme an allen einfachen intervokalischen Konsonantengraphemen zweisilbiger Stämme in den Korpora einfacher Stämme englisch und einfacher Stämme deutsch.

Bei den intervokalischen ⟨CC⟩-Clustern ist interessant, ob sie sich verhalten wie Anfangsränder oder wie Endränder. Wenn wir zeigen können, dass sich stamminterne ⟨CC⟩-Cluster wie Endränder verhalten, dann rechtfertigt das strukturell eine Zerlegung in eine maximale erste Silbe (mit zweifach besetztem Endrand) und eine vokal-initiale, ‚leichte‘ zweite Silbe.

Geht man nun alle <CC>-Cluster danach durch, ob sie als Anfangs- und/oder als Endrand von Stämmen vorkommen können, dann ergeben sich die folgenden Zahlen:³⁸

Tab. 30: Anzahl der intervokalischen <CC>-Cluster im deutschen und englischen Teilkorpus einfacher Stämme; Anteil der Stämme, die so auch als Anfangsrand von Stämmen vorkommen; Anteil der Stämme, die so auch als Endrand vorkommen.

	deutsch		englisch	
	#Cluster	#Types	#Cluster	#Types
<CC>-Cluster gesamt	145	1.384	143	1.352
als Anfangsrand möglich	26 (18%)	177 (13%)	30 (21%)	250 (18%)
als Endrand möglich	82 (57%)	1.251 (90%)	69 (48%)	1.036 (77%)

In beiden Teilkorpora sehen wesentlich mehr intervokalische Cluster aus wie Endränder (z. B. *hinten*, *angel*) als wie Anfangsränder (z. B. *Patron*, *bible*). Das gilt für die Anzahl der verschiedenen Cluster – das gilt aber in weit höherem Maße für die Anzahl verschiedener Types, die diese Cluster enthalten. Im deutschen Teilkorpus ist bei 90 % aller Types mit <CC>-Cluster eine Interpretation als Endrand möglich. Im englischen Korpus liegt die Zahl etwas niedriger bei 77 %.

Die meisten intervokalischen Cluster in Zweisilbern sind strukturell aufgebaut wie Endränder. Die entsprechenden Stämme ‚zerfallen‘ daher beinahe automatisch in zwei Teile, einen ‚Stamm‘ (<Falk>) und ein ‚Affix‘ (<e>). Diese Segmentierung wird uns im Laufe der Arbeit noch mehrfach begegnen. Der ‚Stamm‘ wird im Folgenden ‚Wurzel‘ genannt, das ‚Affix‘ in Anlehnung an Berg (2016) ‚Endung‘.³⁹ In graphematischen Zweisilbern, die hier relevant sind, ist die

³⁸ Wir setzen hier diejenigen komplexen Grapheme an, die sich jeweils im Anfangs- bzw. Endrand ergeben haben. Im deutschen Teilkorpus sind das im Anfangsrand <ch>, <rh> und <sch>, im Endrand <ch>, <ck>, <sch>, <ph>, <th> und <rh>; im englischen Teilkorpus sind das im Anfangsrand <rh> und <wh>; im Endrand <ch>, <gh> und <rh>.

Zum Teil kommen die intervokalischen Cluster weder als Anfangs- noch als Endrand vor (z. B. <dl> in *Adler*); zum Teil sind sie sowohl Anfangs- als auch Endrand (z. B. <st> in *Bestie*). Aus diesem Grund ergeben die Anteile in der Summe nicht 100 %.

³⁹ Der Begriff der Endung umfasst Eisenbergs (2013a: 209) ‚morphologischen Rest‘, also die durch Stammallomorphie segmentierbaren Einheiten wie *-e* in *Treppe* (wegen *Treppchen*) – er geht aber weit über Eisenbergs Begriff hinaus, indem er auch nicht segmentierbare Einheiten bezeichnet wie *-el* in *Angel* oder *-is* in *Tennis*.

Endung der Teil des Stamms, der den Reim der zweiten graphematischen Silbe umfasst – unabhängig von der Struktur der intervokalischen Konsonanten (vgl. z. B. *Falk+e*, *Ank+er*, *Hopf+en*, *Apr+il*, *Tenn+is*, *Pir+at*). Die Wurzel deckt sich größtenteils mit der von Taft (1979) vorgeschlagenen „basic orthographic syllable structure“ (BOSS); zu den Unterschieden siehe 3.2.5, vgl. auch die ähnliche Gliederung in Schmidt (2014). Der Begriff ‚Pseudosuffix‘ wird im Rahmen dieser Arbeit etwas spezifischer verwendet als bspw. bei Augst (1986). Er bezeichnet im Folgenden eine Teilmenge der Endungen, nämlich diejenigen mit ⟨e⟩ (also vor allem ⟨e⟩, ⟨er⟩, ⟨en⟩ und ⟨el⟩).

Eine theoretisch etwas informiertere Modellierung sollte darüber hinaus auch Fälle wie *Unbill*, *vergeud[en]* und *gewöhn[en]* angemessen erfassen. Diese Stämme bestehen aus zwei Teilen, und dennoch sind sie einfach im hier verwendeten Sinne – es kommen eben nicht beide Teile frei vor. Die einfachste Lösung, um Sementierungen wie *Unb+ill*, *verg+eud* und *gew+öhn* zu vermeiden, scheint es zu sein, eine Liste von Präfixen („Pseudopräfixe“ bei Augst 1986) vorzuhalten, um zu den intuitiven Segmentierungen *Un+bill*, *ver+geud* und *ge+wöhn* zu kommen.

Unabhängig davon sehen die meisten Zweisilber graphotaktisch aus wie ein einsilbiger Stamm und ein Affix. Es lohnt in diesem Zusammenhang, sich vor Augen zu führen, dass eine graphematische Gliederung wie ⟨Falk⟩+⟨e⟩ kein phonographischer Reflex ist: In der gesprochenen Sprache werden die intervokalischen Konsonanten in Stämmen wie *Falke* auf die beiden Silben aufgeteilt (/fal.kə/) – die Tatsache, dass die Sequenz der intervokalischen Konsonanten sich strukturell wie ein komplexer Endrand verhält, ist hier irrelevant. In der Schrift ist das anders.

3.2.2.6 Kerne

Auch bei den Kernen können wir zunächst global fragen: Aus wie vielen Vokalgraphemen⁴⁰ bestehen Kerne wie häufig in den beiden Teilkorpora? Abbildung 28 präsentiert die Verteilung:

⁴⁰ Da weder im englischen noch im deutschen Teilkorpus komplexe Vokalgrapheme angesetzt wurden, könnten wir hier auch von ‚Vokalbuchstaben‘ sprechen. Wenn trotzdem der Begriff ‚Graphem‘ verwendet wird, dann geschieht das, um an die Diskussion der Anfangs- und Endränder in den letzten Abschnitten anzuknüpfen.

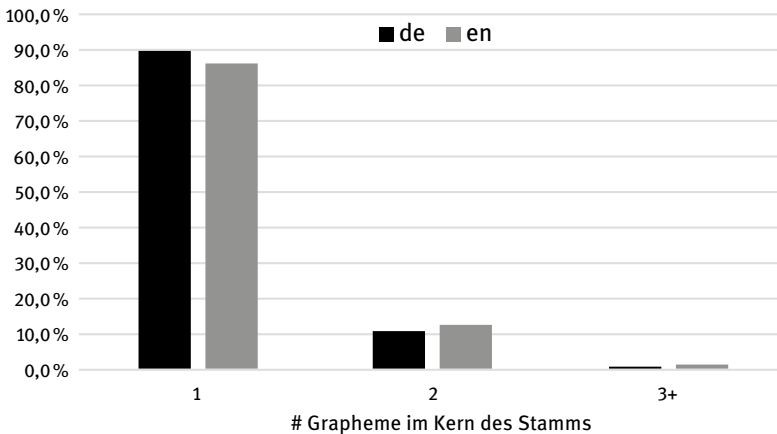


Abb. 28: Relativer Anteil der Stämme mit 1, 2, und 3 Graphemen im Kern an allen Stämmen. Datengrundlage: Teilkorpora einfacher Stämme deutsch und englisch.

Hier fällt wieder der erstaunliche Grad an Übereinstimmung zwischen den beiden Sprachen auf. Der eindeutige Standardfall ist in beiden Korpora der einfach besetzte Kern. Im Englischen sind $\langle VV \rangle$ -Verbindungen gegenüber dem Deutschen leicht häufiger; die relativen Unterschiede sind aber minimal. Auf alle Wörter – also auch auf mehrsilbige Stämme – gesehen, bilden mehrfach besetzte Kerne die Ausnahme.

Dieses Bild ändert sich etwas, wenn wir Silbenkerne in Abhängigkeit von der Silbenzahl des Stamms und der Position der Silbe betrachten. Das soll im Folgenden geschehen. Es werden nacheinander a) Kerne in graphematischen Einsilbern, b) Kerne erster Silben in graphematischen Zweisilbern und c) Kerne zweiter Silben in graphematischen Zweisilbern untersucht.

Zunächst zu den Einsilbern. Deren Kern ist – wenig überraschend – in beiden Korpora ebenfalls hauptsächlich von einem Element besetzt. Der Anteil der komplexen Kerne ist aber wesentlich höher als im Durchschnitt aller Stämme (siehe Abb. 29).

Unter den komplexen Kernen sind Verbindungen aus zwei Vokalgraphemen eindeutig der Standardfall. Dreifach besetzte Kerne (z. B. deutsch *Trauer*, *Geier*, englisch *pious*, *crayon*) sind demgegenüber sehr selten, auch in absoluten Zahlen (deutsch: 27-mal, englisch: 31-mal); vierfach besetzte Kerne kommen im deutschen Korpus gar nicht vor, im englischen nur zweimal (*miaou*, *hooey*). Bei den beiden Wörtern handelt es sich um ein Onomatopoetikon und ein Slangwort („Humbug“). Die Wörter sind also sowohl im Wortschatz als auch graphematisch markiert, und die maximale Anzahl stamminterner V-Grapheme liegt in beiden Sprachen bei drei.

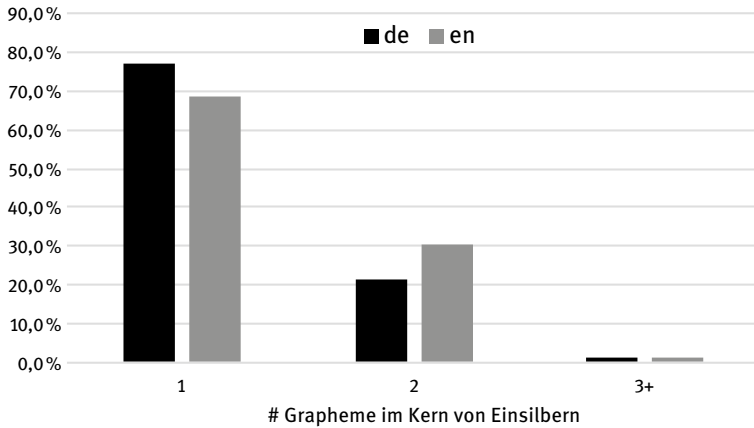


Abb. 29: Relativer Anteil der Stämme mit 1, 2, und 3 Graphemen im Kern an allen graphematisch einsilbigen Stämmen. Datengrundlage: Teilkorpora einfacher Stämme deutsch und englisch.

Unter den einfach besetzten Kernen finden wir in den beiden Korpora die folgende Verteilung von Vokalgraphemen:

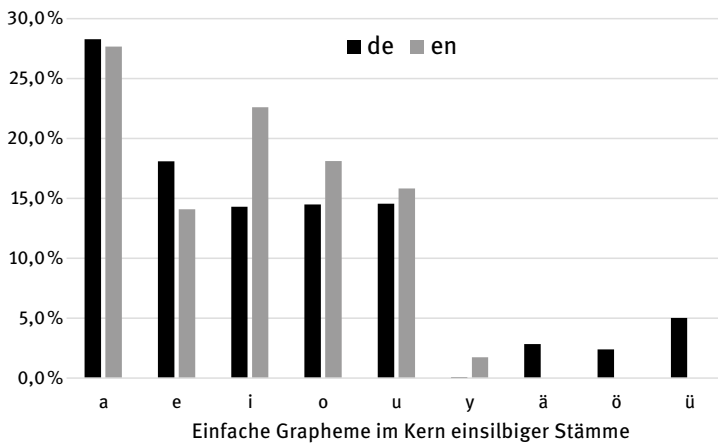


Abb. 30: Relative Anzahl der Vokalgrapheme an allen einfachen Kernen in einsilbigen Stämmen in den Korpora einfacher Stämme englisch und einfacher Stämme deutsch.

Obwohl «e» das häufigste Vokalgraphem in beiden Korpora ist (vgl. oben Abb. 1), ist es als alleiniger Silbenkern im Einsilber nicht das häufigste Element. In beiden Korpora ist «a» deutlich häufiger. Im deutschen Korpus kommt «e» ähnlich oft vor wie «i», «o» und «u»; im englischen Korpus ist es sogar seltener als diese Grapheme.

⟨y⟩ ist im englischen Korpus randständig, im deutschen extrem selten – es kommt nur einmal vor, und zwar im fremden Stamm *lynch[en]*. Bei den Umlauten ist ⟨ü⟩ etwa so häufig wie ⟨ä⟩ und ⟨ö⟩ zusammengenommen. Dass ⟨e⟩ vor allem aufgrund seiner Funktion als Silbenkern von Zweisilbern das häufigste Graphem sein würde, war klar; dass im deutschen Korpus aber ⟨a⟩ dermaßen deutlich häufiger ist als die übrigen Grapheme, ist dann doch überraschend.

Die Verteilung in Abbildung 30 ändert sich nicht wirklich für die ersten Silben von Zweisilbern, wie Abbildung 31 zeigt:

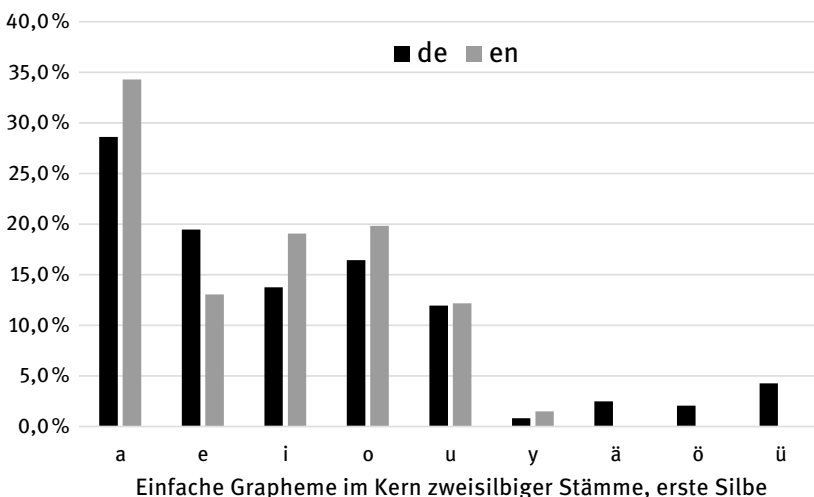


Abb. 31: Relative Anzahl der Vokalgrapheme an allen einfachen Kernen in ersten Silben von graphematischen Zweisilbern in den Korpora einfacher Stämme englisch und einfacher Stämme deutsch.

Die relativen Häufigkeiten der Grapheme sind sehr ähnlich verteilt. Das kann als Hinweis darauf verstanden werden, dass die ersten Silben von Zweisilbern strukturell mit Einsilbern vergleichbar sind.

Bei den zweiten Silben in graphematischen Zweisilbern ergibt sich eine ganz andere Verteilung:

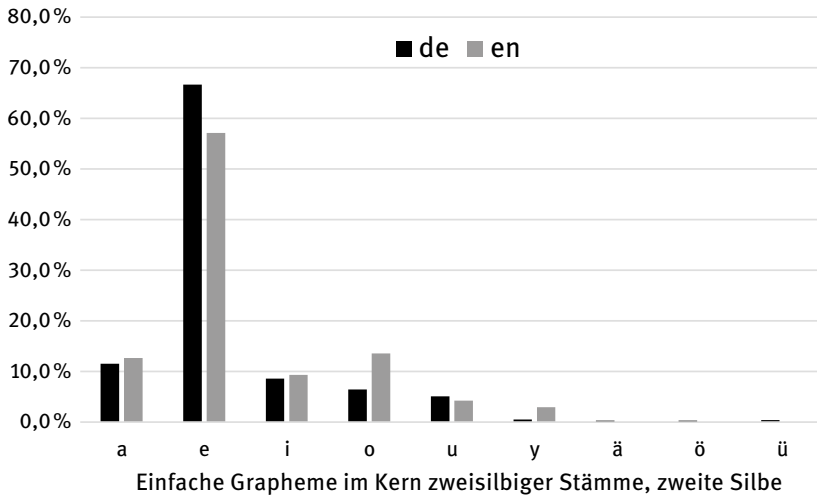


Abb. 32: Relative Anzahl der Vokalgrapheme an allen einfachen Kernen in zweiten Silben von graphematischen Zweisilbern in den Korpora einfacher Stämme englisch und einfacher Stämme deutsch.

Hier ist ⟨e⟩ in beiden Korpora eindeutig der bevorzugte einfache Silbenkern. Im deutschen Korpus fällt der Wert noch etwas höher aus als im englischen; dort ist ⟨o⟩ etwas häufiger belegt als im deutschen Korpus. Den größten Anteil daran haben Wörter, die auf ⟨on⟩ enden (z. B. *icon*, *melon*) – sie machen etwas mehr als 25 % aller Zweisilber mit ⟨o⟩ in der zweiten Silbe aus (zum Zusammenhang zwischen der graphematischen Endung ⟨on⟩ und Konsonantenverdoppelung im Englischen vgl. Berg 2016). Fast genauso häufig ist stamminales ⟨o⟩ (z. B. *hero*, *veto*).

Betrachten wir nun die zweifach besetzten Kerne. Wir beschränken uns auch hier zunächst auf graphematische Einsilber, dann auf die ersten und schließlich die zweiten Silben von graphematischen Zweisilbern. Hier findet sich die Verteilung in Tabelle 31 und 32 (siehe folgende Seite).

Im Gegensatz zu den Anfangs- und Endrändern ist Asymmetrie der Kombinationen kein Strukturprinzip der komplexen Silbenkerne. Das gilt für die deutschen Daten (links); hier finden sich z. B. die Paare ⟨ei⟩ (*Bein*) – ⟨ie⟩ (*viel*); ⟨ai⟩ (*Laib*) – ⟨ia⟩ (*Dia*); ⟨io⟩ (*Spion*) – ⟨oi⟩ (*Point*) u. a. Für die englischen Daten gilt das noch wesentlich mehr – der Grad der Ausnutzung ist erstaunlich hoch. Es fehlen lediglich die drei Verdoppelungen ⟨ii⟩, ⟨uu⟩ und ⟨yy⟩ sowie ⟨iy⟩ und ⟨yu⟩. Das macht die Beschreibung nicht unbedingt einfacher (siehe unten). Anstatt kategorial das Vorkommen bzw. Nicht-Vorkommen zu bewerten, müssen die Häufigkeitsverhältnisse mit einbezogen werden: Was sind die dominanten Verbindungen?

Tab. 31 (links) und 32 (rechts): Kreuzklassifikation der zweifach besetzten Silbenkerne von graphematisch einsilbigen Stämmen im deutschen (links) und englischen (rechts) Teilkorpus.

	a	e	i	o	u	y	ä
a	10		16	2	98	1	
e	5	15	146	1	28		
i	2	87		3			1
o	2		1	6	1	1	
u				1			
y	2	1		1			
ä				1	3		
ö		1					

	a	e	i	o	u	y
a	1	3	76	1	31	30
e	147	109	23	2	3	8
i	11	28		7	1	
o	57	15	26	108	93	8
u	2	19	9	2		2
y	12	10	1	4		

Viele Kombinationen im deutschen Korpus (linke Tabelle) sind nicht belegt. Die Verteilung kann so beschrieben werden: Wir haben eine ganze Reihe nur einmal belegter (<ay> (*Spray*), <äo> (*Äon*), <eo> (*neon*), <iä> (*Diät*), <oi> (*Point*), <ou> (*Sou*), <oy> (*Boy*), <öe> (*Böe*), <uo> (*Fluor*), <ye> (*Yen*) und <yo> (*myop*)) sowie einige zweimal belegte (<ao> (*Chaos, Tao*), <ia> (*Dia, via*), <oa> (*Boa, Toast*), <ya> (*Yacht, Yak*)). Sie alle sind Fremdwörter. Verdoppelt werden können <a>, <e> und <o> (wobei nur <oo> signifikant häufiger ist als erwartet, vgl. 3.2.1); bei den übrigen Kombinationen sind am frequentesten <ei>, <au>, <ie> und (weitaus seltener) <eu> und <ai>. Es ergibt sich das aus Eisenberg (2013a: 299) bekannte System von zwei Erstbestandteilen (<a>, <e>) und zwei Zweitbestandteilen (<i>, <u>), erweitert um <ie> (Fuhrhop/Berg inger.):

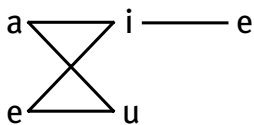


Abb. 33: Graphotaktisches Schema der Kombinatorik von zwei Vokalgraphemen in deutschen Silbenkernen.

Im englischen Korpus ist die Lage unübersichtlicher. Verdoppelt werden nur <e> und <o> (und beide sind signifikant häufiger als erwartet, vgl. 3.2.1). Der eine Beleg für <aa>, *baa*, ist eine Ausnahme; es handelt sich um ein Onomatopoeikon, das auch im Wortschatz markiert ist. Wir wissen, dass <i> und <y> in bestimmten Umge-

bungen alternieren (z. B. *happy* – *happily*, vgl. 4.1 und Berg 2013);⁴¹ wir können also die Werte für ⟨i⟩ und ⟨y⟩ zusammenfassen. Die häufigsten Kombinationen sind (abgesehen von den Verdoppelungen) in absteigender Reihenfolge: ⟨ea⟩, ⟨ai/y⟩, ⟨ou⟩, ⟨oa⟩, ⟨au⟩, ⟨ei/y⟩; sie alle haben mehr als 30 Belege. Das bedeutet: ⟨a⟩; ⟨e⟩ und ⟨o⟩ sind gute erste Bestandteile, ⟨a⟩, ⟨i/y⟩ und ⟨u⟩ sind gute zweite Bestandteile. Es ergibt sich folgendes System (Fuhrhop/Berg einger.):

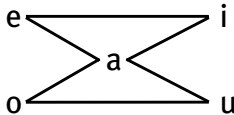


Abb. 34: Graphotaktisches Schema der Kombinatorik von zwei Vokalgraphemen in englischen Silbenkernen.

In den ersten Silben von Zweisilbern ändert sich das nicht grundlegend. Tabelle 33 und Tabelle 34 zeigen die Verteilung im deutschen (links) und englischen Teilkorpus (rechts):

Tab. 33 (links) und 34 (rechts): Kreuzklassifikation der zweifach besetzten Silbenkerne in ersten Silben von graphematisch zweisilbigen Stämmen im deutschen (links) und englischen (rechts) Teilkorpus.

	a	e	i	o	u	y	ä
a	1		5	2	65		
e	2	3	97	1	23	1	
i	4	59		3			
o	2		3		6		
u	2		1				
y	3			3			1
ä	1				6		

	a	e	i	o	u	y
a		6	20	4	32	3
e	42	21	9	8	5	
i	16	16	2	5		
o	3	3	10	25	56	3
u	3	3	11			
y	4	3	1	8	2	

⁴¹ In Tabelle 32 sind beide Grapheme als Zweitbestandteil fast gleich verteilt: Ist ⟨i⟩ möglich, dann auch ⟨y⟩ und andersherum (die Ausnahme ist ⟨yi⟩ (*yin*), obwohl ⟨yy⟩ nicht vorkommt – hier handelt es sich aber schlicht um eine Beschränkung, die die Verdoppelung von ⟨y⟩ verbietet, siehe oben 3.2.1).

Im deutschen Korpus sind wie im Einsilber $\langle ei \rangle$, $\langle au \rangle$, $\langle ie \rangle$ und $\langle eu \rangle$ die häufigsten Verbindungen. $\langle ai \rangle$ ist hingegen seltener als $\langle ou \rangle$ (z. B. *Gourmet*) oder $\langle äu \rangle$ (*räuspfern*). Die Fallzahlen sind hier allerdings schon so gering, dass es sich auch um eine zufällige Verschiebung handeln kann. Auffällig ist aber, dass die Verdoppelungen $\langle aa \rangle$, $\langle ee \rangle$ und $\langle oo \rangle$ wesentlich seltener sind: $\langle oo \rangle$ kommt nicht vor; $\langle aa \rangle$ einmal (*Waage*) und $\langle e \rangle$ dreimal (*Beere, Reede, Seele*). Verdoppelungen von Vokalgraphemen sind im Deutschen also typisch für Einsilber.

Im englischen Korpus sind (wiederum in absteigender Reihenfolge) die häufigsten Verbindungen $\langle ou \rangle$, $\langle ea \rangle$, $\langle au \rangle$, $\langle oo \rangle$, $\langle ee \rangle$ und $\langle ai/y \rangle$. Diese Reihe deckt sich größtenteils mit den oben angegebenen Verbindungen. Das gilt auch für die Verdoppelungen, die im englischen Teilkorpus – anders als im deutschen – auch in ersten Silben von Zweisilbern auftreten. In dieser Umgebung ist allerdings eine Verbindung wesentlich seltener als im Einsilber: $\langle oa \rangle$. Diese Verbindung tritt hier nur dreimal auf; zweimal in Verbindung mit $\langle r \rangle$ und der Endung $\langle se \rangle$ (*coarse, hoarse*) und einmal in einem Fremdwort (*koala*). Das lässt sich auf den ersten Blick über die phonographische Leistung von $\langle oa \rangle$ motivieren: Einsilber mit $\langle oa \rangle$ korrespondieren mit demselben Vokalphonem wie graphematische Zweisilber mit einfachem $\langle o \rangle$ in der ersten und finalelem $\langle e \rangle$ in der zweiten Silbe: *loan* und *lone* sind Homophone. Man kann $\langle a \rangle$ auf dieser Grundlage einen graphematischen ‚Silbenöffner‘ (Fuhrhop/Berg einger.) nennen, und genau diese Funktion ist ja im Zweisilber nicht notwendig, deswegen ist $\langle oa \rangle$ dort so selten. Problematisch an dieser Erklärung ist allerdings, dass sie für $\langle ea \rangle$ nicht verfängt: $\langle ea \rangle$ ist ja auch im Zweisilber prominent vertreten. Unabhängig von der Motivation bleibt festzuhalten: $\langle oa \rangle$ ist (wie die Doppelvokale im deutschen Korpus) auf graphematische Einsilber beschränkt. Abgesehen von diesen Ausnahmen unterscheidet sich die Verteilung der komplexen Silbenkerne in der ersten Silbe von Zweisilbern nicht groß von der Verteilung in Einsilbern, und zwar in beiden Sprachen.

Wie schon bei den einfachen Kernen ändert sich auch bei den doppelt besetzten Kernen das Bild, wenn wir uns stattdessen ihr Vorkommen in zweiten Silben von Zweisilbern ansehen (Tab. 35 und 36). Im deutschen Korpus kommt nur $\langle ee \rangle$ als Verdoppelung vor. Es tritt ganz überwiegend ohne Endrand auf (Typ *Kaffee*, 12 Belege). Zu dieser Regel gibt es nur vier Ausnahmen (*Kaneel, krakeel[en], Lorbeer, Paneel*). Diese Verteilung unterscheidet sich vom Einsilber, wo ein leerer Endrand nur in einem Drittel der Fälle auftritt. $\langle ee \rangle$ ist also in zweiten Silben von Zweisilbern distributionell eingeschränkter als in Einsilbern.

Tab. 35 (links) und 36 (rechts): Kreuzklassifikation der zweifach besetzten Silbenkerne in zweiten Silben von graphematisch zweisilbigen Stämmen im deutschen (links) und englischen (rechts) Teilkorpus.

	a	e	i	o	u	y	ö
a			5	2	14		
e	2	16	22	1	6	2	
i	8	61		22	9		1
o		2	3		1		
u	1	2					

	a	e	i	o	u	y
a	1	1	27	1	7	12
e	20	36	8	16	4	24
i	16	23		19	1	
o	3	5	6	36	20	7
u	2	23	8	1		
y	2			3		

Ansonsten fällt auf, dass ⟨i⟩ als erster Bestandteil präsenter ist als im Einsilber. ⟨io⟩, ⟨iu⟩ und ⟨ia⟩ sind in zweiten Silben von Zweisilbern wesentlich häufiger als in Einsilbern. Und ⟨ie⟩, das ja auch im Einsilber frequent ist, ist im Zweisilber distributionell ähnlich eingeschränkt wie ⟨ee⟩ oben: Es kommt in knapp der Hälfte der Fälle vor einfachem ⟨r⟩ vor (Typ *kapier[en]*, *Papier*, 27 Belege). Die Mehrzahl dieser Wörter sind wiederum Verbstämme. In einem Drittel der Wörter steht ⟨ie⟩ wortfinal (Typ *Aktie*, *Kopie*, 21 Belege). Es bleibt ein Rest von 13 Belegen mit anderem Endrand (z. B. *Ferien*, *Gebiet*, *Spaniel*). Besonders stark ist die Abhängigkeit vom Endrand bei ⟨iu⟩: Es kommt ausschließlich mit finalem ⟨m⟩ vor (Typ *Spatium*, *Kadmium*). Die beiden übrigen prominenten Verbindungen in Tabelle 35, ⟨ei⟩ und ⟨au⟩, verhalten sich ähnlich wie im Einsilber: Sie sind distributionell nicht eingeschränkt.

Im englischen Korpus lässt sich Vergleichbares beobachten. Hier sind die Verdoppelungen ⟨ee⟩ und ⟨oo⟩ vertreten, und beide haben eine eingeschränkte Verteilung. ⟨ee⟩ in der zweiten Silbe kommt in mehr als der Hälfte der Fälle ohne Endrand vor (Typ *coffee*, 22 Belege), und in mehr als einem Viertel der Fälle ist der Endrand einfaches ⟨n⟩ (Typ *canteen*, 10 Belege). Es gibt nur vier Belege mit anderem Endrand (*baksheesh*, *career*, *succeed*, *veneer*). Etwas Ähnliches gilt für ⟨oo⟩: Von den 36 Wörtern ist bei 11 der Endrand unbesetzt (Typ *yahoo*), und bei 17 Wörtern ist der Endrand einfaches ⟨n⟩ (Typ *baboon*). Nur acht Wörter haben einen anderen Endrand (*cahoots*, *cherook*, *chinook*, *forsooth*, *mushroom*, *nainsook*, *simoom*, *tarboosh*).

Im Einsilber finden wir diese Bindung an bestimmte Endränder nicht. Bei ⟨ee⟩ ist zwar auch im Einsilber ein leerer Endrand die häufigste Konstellation (Typ *bee*), aber ⟨ee⟩ kommt mit 11 verschiedenen Endrändern vor, von denen ⟨n⟩ erst den siebten Rang belegt. ⟨oo⟩ kommt gar mit 15 verschiedenen Endrändern vor; hier belegt ⟨n⟩ den dritten Rang. Die Verteilung der Doppelkonsonanten ist also in zweiten Silben von Zweisilbern wesentlich eingeschränkter als in Einsilbern.

Und auch sonst ähnelt die Verteilung im englischen Korpus derjenigen im deutschen Korpus: Viele der frequenten Kombinationen im Einsilber sind auch in zweiten Silben von Zweisilbern frequent. ⟨i⟩ ist als Erstbestandteil deutlich häufiger, und viele der Kombinationen sind an einen bestimmten Endrand gebunden: ⟨eo⟩ tritt häufig mit finalem ⟨n⟩ auf (Typ *surgeon*, 11 Belege), z. T. auch ohne Endrand (Typ *stereo*, 4 Belege); davon gibt es nur eine Ausnahme (*meteor*). ⟨io⟩ kommt am häufigsten stammfinal vor (Typ *audio*, 8 Belege), z. T. aber auch vor ⟨r⟩ (Typ *senior*, 3 Belege); hier gibt es 8 Ausnahmen. Am eindeutigsten ist ⟨ue⟩, das in 22 von 23 Fällen stammfinal auftritt (Typ *statue*); die Ausnahme ist *minuet*. Für die Kombinationen mit finalem ⟨y⟩ (⟨ay⟩, ⟨ey⟩, ⟨oy⟩) gilt das in noch stärkerem Maß; sie alle treten stammfinal auf.

Hier zeigt sich in beiden Korpora ein enger Zusammenhang zwischen Kern und Endrand, also denjenigen Konstituenten, die in der Phonologie oft unter der Kategorie ‚Reim‘ zusammengefasst werden. Für das Englische haben Kessler/Treiman (2001) gezeigt, dass der graphematische Reim bessere Vorhersagen über die phonographische Korrespondenz zulässt als die einzelnen Konstituenten Kern und Endrand. Die Daten, die oben präsentiert wurden, legen nahe, dass es in zweiten Silben von Zweisilbern im Englischen und Deutschen ebenfalls einen engen Zusammenhang zwischen Kern und Endrand gibt: Ist der Kern festgelegt, dann stehen nur wenige Möglichkeiten für einen Endrand zur Verfügung. Diese These wird im nächsten Abschnitt (3.2.2.7) überprüft, wenn es um die Interaktion zwischen den Silbenkonstituenten geht.

Werfen wir noch einen Blick auf Verbindungen aus drei Vokalgraphemen. Wie oben in Abbildung 28 gezeigt, sind diese Verbindungen in beiden Korpora selten. Wenn wir uns auf graphematische Einsilber beschränken, gibt es im deutschen Teilkorpus 27, im englischen 31 Wörter vom Typ *Bauer* oder *mayor*. Im deutschen Teilkorpus verteilen sich die 27 Wörter auf sechs verschiedene ⟨VVV⟩-Cluster, wie Tabelle 37 zeigt:

Tab. 37: Dreifach besetzte Kerne von graphematisch einsilbigen Stämmen im Teilkorpus einfacher deutscher Stämme.

⟨VVV⟩-Cluster	#Stämme	Beispiel
aue	11	Bauer, Trauer
eue	7	Feuer, teuer
eie	5	Feier, Schleier
äue	2	Gräuel, Knäuel
oya	1	royal
aie	1	Laie

Mit Ausnahme von ⟨äu⟩ und ⟨oya⟩ können wir die Verbindungen als typische ⟨VV⟩-Verbindungen + ⟨e⟩ analysieren. Wir finden also in ⟨VVV⟩-Clustern folgendes Schema:

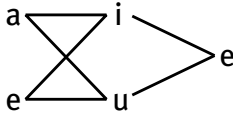


Abb. 35: Graphotaktisches Schema der Kombinatorik von drei Vokalgraphemen in deutschen Silbenkernen.

⟨äue⟩ und ⟨oya⟩ sind vom Schema nicht abgedeckt; beide sind aber randständig. Das gilt umso mehr, als *Gräuel* erst seit der Orthographiereform 1996 mit ⟨äu⟩ und nicht mit ⟨eu⟩ kodifiziert ist.

Im englischen Teilkorpus gibt es mehr Variation; hier verteilen sich die 31 Wörter auf 19 Types von ⟨VVV⟩-Clustern (Tab. 38). Wir können hier dennoch einiges an Struktur erkennen. Zum einen gibt es drei Folgen, die mit ⟨y⟩ beginnen – und dieses ⟨y⟩ ist gleichzeitig stamminitial (*year, young, yield*). Offenbar handelt es sich hier (wie beim stammfinalen ⟨y⟩) um eine Wortrandmarkierung – schließlich kommt ⟨y⟩ im Einsilber nicht nach Konsonantengraphemen vor. ⟨y⟩ kombiniert in den entsprechenden Abfolgen mit den üblichen ⟨VV⟩-Verbindungen ⟨ea⟩, ⟨ou⟩ und ⟨ie⟩.

Tab. 38: Dreifach besetzte Kerne von graphematisch einsilbigen Stämmen im Teilkorpus einfacher englischer Stämme.

⟨VVV⟩-Cluster	#Stämme	Beispiel
yea	5	year, yeast
you	4	you, young
aya	3	kayak, mayan
ayo	2	crayon, mayor
iou	2	pious, sioux
eau	2	beau, beaut
aia	1	naiad
aye	1	aye
iao	1	miaow
oya	1	soya
oue	1	roue
uoy	1	buoy
oye	1	doyen

⟨VV⟩-Cluster	#Stämme	Beispiel
eue	1	queue
aeo	1	aeon
yie	1	yield
aea	1	paeon
eye	1	eye
eyo	1	eyot

Außerdem gibt es eine Reihe von Abfolgen, in denen ⟨y⟩ medial zwischen zwei anderen Vokalgraphemen auftritt (⟨aya⟩, ⟨ayo⟩, ⟨aye⟩, ⟨oya⟩, ⟨oye⟩, ⟨eye⟩ und ⟨eyo⟩). An erster und letzter Stelle treten ⟨a⟩, ⟨e⟩ und ⟨o⟩ auf – also genau die Grapheme, die in ⟨VV⟩-Verbindungen an erster Stelle stehen können. Zu einer vollständigen Kombinatorik fehlen nur ⟨oyo⟩ und ⟨eya⟩. Nehmen wir sie als mögliche Schreibungen an, ergibt sich das folgende Schema:

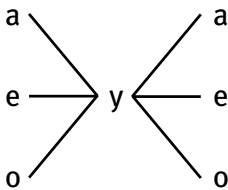


Abb. 36: Graphotaktisches Schema der Kombinatorik von drei Vokalgraphemen in englischen Silbenkernen.

Übrig bleibt ein gemischter Haufen: Wir haben viele Abfolgen, die auf (meist französischen) Fremdwörtern beruhen (*sioux*, *beau*, *beaut*, *naiad*, *roue*, *buoy*, *aeon*, *paeon*); ein Onomatopoetikon (*miaow*); ein Wort mit Adjektivsuffix *-ous* (*pious*) – und schließlich ⟨eue⟩, das auf dem auch phonographisch idiosynkratischen *queue* (mit vier adjazenten Vokalbuchstaben!) beruht. Wir begnügen uns daher mit den beiden gefundenen Mustern: ⟨y⟩ als Wortanfangsmarkierung und ⟨y⟩ als ‚Trenner‘ von Erstbestandteilen.

3.2.2.7 Interaktion der Silbenkonstituenten

Wir betrachten zunächst sehr grob die Interaktion zwischen der Besetzung der Konstituenten. Wenn der Anfangsrand zweifach besetzt ist, wie häufig ist dann ein einfacher Silbenkern? Wie häufig ein komplexer? Und etwas allgemeiner und

interessanter: Weicht die Verteilung, die wir zwischen den einzelnen Konstituenten finden, signifikant von einer zufälligen Verteilung ab? Wir beschränken uns bei der Untersuchung auf Einsilber. Als Grapheminventar wird auch hier (6) aus Abschnitt 3.1 oben verwendet. Komplexe Grapheme sind im Deutschen <qu>, <ch> und <ck>, im Englischen <qu>.

Beginnen wir mit dem Verhältnis von Anfangsrand und Kern. Tabelle 39 und Tabelle 40 zeigen den Zusammenhang zwischen Anfangsrand und Kern im deutschen (links) und englischen (rechts) Teilkorpus:

Tab. 39 (links) und Tab. 40 (rechts): Zusammenhang zwischen der Besetzung des Anfangsrandes mit 0, 1, 2 oder 3 Graphemen (Zeilen) und des Kerns mit 1, 2 oder 3 Graphemen (Spalten) in graphematischen Einsilbern im deutschen (links) und englischen (rechts) Teilkorpus einfacher Stämme.

	V	VV	VVV		V	VV	VVV
-	65	23	2	-	67	56	15
C	900	250	16	C	1.074	488	15
CC	454	121	8	CC	707	278	1
CCC	134	42	1	CCC	67	27	

Wie sind die Zahlen nun zu interpretieren? Gibt es einen Zusammenhang zwischen der Besetzung der beiden Konstituenten? Um die Frage zu beantworten, verwenden wir dieselbe Methode wie oben bei der Ermittlung signifikant häufiger/seltener Bigramme (siehe oben 3.2.1.3): Wir ermitteln zunächst die erwartete Häufigkeit für eine beliebige Kombination – zum Beispiel <CCVV> im deutschen Korpus (wie in *Stein*), das 126-mal vorkommen sollte. Real kommt diese Kombination 121-mal vor, und dieser Unterschied wird nun stochastisch auf Signifikanz geprüft. In diesem Fall ist die Abweichung nicht signifikant. Wenn wir das für alle Zellen in den Tabellen oben machen, dann ergibt sich folgendes Bild (die Kodierung folgt dem Vorgehen oben: ‚+‘, ‚++‘ und ‚+++‘ markieren ein signifikant häufigeres Vorkommen, ‚-‘, ‚--‘ und ‚---‘ ein signifikant selteneres; ‚N‘ gibt an, dass die Abweichung nicht signifikant ist. Ist eine Kombination nicht belegt, aber stochastisch möglich, wird sie mit ‚V‘ bezeichnet, ist sie hingegen sehr unwahrscheinlich, mit ‚U‘).

Tab. 41 (links) und **Tab. 42 (rechts)**: Zusammenhang zwischen der Besetzung des Anfangsrandes mit 0, 1, 2 oder 3 Graphemen (Zeilen) und des Kerns mit 1, 2 oder 3 Graphemen (Spalten) in graphematischen Einsilbern im deutschen (links) und englischen (rechts) Teilkorpus einfacher Stämme; klassifiziert nach der Wahrscheinlichkeit, dass der reale Wert zufällig ist.

	V	VV	VVV		V	VV	VVV
-	N	N	N	-	--	+	+++
C	N	N	N	C	N	N	N
CC	N	N	N	CC	N	N	---
CCC	N	N	N	CCC	N	N	V

Im deutschen Korpus (linke Tab.) gibt es offenbar keinen Zusammenhang zwischen Anfangsrand und Kern. Keine Kombination weicht signifikant von den erwarteten Werten ab. Das spiegelt sich auch in einem nicht-signifikanten Chi-Quadrat-Test wider ($p = 0,8$). Im englischen Korpus ist das etwas anders. Vor allem die leeren Anfangsränder weichen von den erwarteten Werten ab. Der Typ $\langle -V \rangle$ (z. B. *add*) ist dabei seltener, die Typen $\langle -VV \rangle$ (z. B. *earl*) und $\langle -VVV \rangle$ (z. B. *young*) sind häufiger als erwartet (entsprechend ergibt der Chi-Quadrat-Test ein signifikantes Ergebnis, $p < 0,001$). Besonders das Verhalten der letzten beiden Typen lässt sich mit der Verteilung von $\langle y \rangle$ erklären: $\langle y \rangle$ wurde oben aufgrund seiner Verteilung als Vokalbuchstabe definiert, daher ist der Anfangsrand in Wörtern wie *young* leer. 23 der 56 Wörter vom Typ $\langle -VV \rangle$ und zehn der 15 Wörter vom Typ $\langle -VVV \rangle$ beginnen mit $\langle y \rangle$. Klassifiziert man $\langle y \rangle$ in diesen Fällen anders, so ergibt sich insgesamt eine neutralere Verteilung.

Betrachten wir als Nächstes das Verhältnis von Anfangsrand und Endrand. Die entsprechenden Zahlen für das deutsche Korpus sind in Tabelle 43, die für das englische Korpus in Tabelle 44 zu finden (Stämme mit mehr als drei Konsonanten im Anfangs- oder Endrand werden im Folgenden nicht betrachtet; das betrifft im englischen Korpus sechs, im deutschen Korpus sieben Stämme):

Tab. 43 (links) und Tab. 44 (rechts): Zusammenhang zwischen der Besetzung des Anfangsrandes mit 0, 1, 2 oder 3 Graphemen (Zeilen) und des Endrands mit 0, 1, 2 oder 3 Graphemen (Spalten) in graphematischen Einsilbern im deutschen (links) und englischen (rechts) Teilkorpus einfacher Stämme.

	-	C	CC	CCC
-	3	32	44	9
C	46	443	603	69
CC	34	266	263	20
CCC	6	83	85	3

	-	C	CC	CCC
-	7	48	71	11
C	101	650	727	97
CC	72	451	405	58
CCC	8	46	35	5

Auch hier betrachten wir wieder den Unterschied zwischen erwarteter und realer Häufigkeit (Tab. 45 und Tab. 46):

Tab. 45 (links) und Tab. 46 (rechts): Zusammenhang zwischen der Besetzung des Anfangsrandes mit 0, 1, 2 oder 3 Graphemen (Zeilen) und des Endrands mit 0, 1, 2 oder 3 Graphemen (Spalten) in graphematischen Einsilbern im deutschen (links) und englischen (rechts) Teilkorpus einfacher Stämme; klassifiziert nach der Wahrscheinlichkeit, dass der reale Wert zufällig zustande kommt.

	-	C	CC	CCC
-	N	N	N	+
C	N	-	N	N
CC	N	+	N	-
CCC	N	N	N	-

	-	C	CC	CCC
-	N	N	N	N
C	N	N	N	N
CC	N	N	-	N
CCC	N	N	N	N

Die meisten Kombinationen sind in beiden Sprachen unauffällig. Im englischen Korpus (rechte Tab.) ist überhaupt nur eine Kombination signifikant, ein zweifach besetzter Anfangs- und Endrand (z. B. *start*). Diese Konstellation ist signifikant seltener – statt der erwarteten 437 Wörter gibt es nur 405 Wörter dieses Typs. Im deutschen Korpus sind vor allem die dreifachen Endränder interessant. Hier scheint sich eine These von Augst (1986: 311) zu bestätigen, der als graphematisches Schema deutscher Stämme (genauer: des ‚Kerns‘ deutscher Stämme) (30) aufstellt (ähnlich auch Enderle 2005: 223):

$$(30) \quad KKK + V + KKKK \quad K \leq 5$$

Im Anfangsrand sind drei Konsonanten möglich, im Endrand vier – <sch> und <ch> besetzen bei Augst (1986) jeweils eine Position. Die Summe der Konsonanten im Anfangs- und Endrand übersteigt fünf nicht. Tatsächlich gibt es im Korpus nur einen Stamm, der mehr als fünf Konsonanten aufweist, nämlich *Strumpf*. Selbst wenn <sch> als Verbindung aus <s> und <ch> angesetzt wird, gibt es nur zwei Ausnahmen mehr, *Schwulst* und *schrumpfen*. Und tatsächlich zeigt sich in Tabelle 45, dass dreifach besetzte Endränder seltener sind als erwartet, wenn der Anfangsrand komplex ist – und häufiger als erwartet, wenn er nicht besetzt ist (vgl. für eine äquivalente phonologische Beobachtung Maas 1992: 270).

Die weiteren Abweichungen von den erwarteten Häufigkeiten im deutschen Korpus betreffen einfach besetzte Endränder. Hier ist besonders interessant, dass Kombinationen aus einfach besetztem Anfangs- und Endrand signifikant seltener sind als erwartet. Das lässt sich damit erklären, dass stammfinales einfaches <l>, <m>, <n> und <r> selten sind – das ist genau die Umgebung, in der das sog. Dehnungs-<h> auftritt. Allerdings tritt dieses <h> (in Wörtern wie *kahl*, *fühlen* oder *Sahne*) vor allem bei unbesetzten oder einfachen Anfangsrändern auf, wie Roemheld (1955) zeigt. Da der Zusammenhang relativ komplex ist, wird er weiter unten in diesem Abschnitt separat untersucht; an dieser Stelle soll es mit dem Zusammenhang zwischen Kern und Endrand weitergehen.

In der Phonologie ist die Interaktion zwischen Kern und Endrand von Sprechsilben seit Langem bekannt, sie läuft in der germanistischen Literatur oft unter dem Schlagwort des ‚Längenausgleichs‘ (vgl. z. B. Eisenberg 2013a: 115): Nach einem ungespannten Vokal kann ein mehrfach geschlossener Endrand auftreten (wie in *Kunst*), nach einem gespannten Vokal oder Diphthong hingegen nicht (Moulton 1956). Es gibt also einen maximalen Reim im Deutschen, der ein gewisses ‚Gewicht‘ nicht überschreiten kann (das lässt sich im Rahmen der nicht-linearen Phonologie formalisieren, vgl. z. B. Wiese 2000). Es gibt eine handvoll Ausnahmen von dieser Generalisierung wie *Obst* und *Mond* – und in all diesen Ausnahmen ist das finale Segment ein koronaler Obstruent (vgl. z. B. Wiese 2000: 48). Auch im Englischen gibt es einen solchen Längenausgleich zwischen Kern und Endrand; es gibt zwar auf den ersten Blick viele Ausnahmen (z. B. *pound*, *saint*, *range*). All diesen Ausnahmen ist aber wie im Deutschen gemein, dass das finale Segment ein koronaler Obstruent ist (vgl. Giegerich 1992: 147 f.). Es kann also in beiden Sprachen nicht willkürlich Material hinzugefügt werden.

Die kurzen Ausführungen zur Phonologie setzen den Rahmen dessen, was zu erwarten ist: Die Kombination von Kern und Endrand ist in Tabelle 47 für das deutsche Korpus und in Tabelle 48 für das englische Korpus aufgeführt. Stämme mit mehr als dreifach besetzten Kernen oder Endrändern wurden ausgeschlossen; das betrifft im englischen und im deutschen Korpus jeweils sieben Einträge.

Tab. 47 (links) und Tab. 48 (rechts): Zusammenhang zwischen der Besetzung des Kerns mit 1, 2 oder 3 Graphemen (Zeilen) und des Endrands mit 0, 1, 2 oder 3 Graphemen (Spalten) in graphematischen Einsilbern im deutschen (links) und englischen (rechts) Teilkorpus einfacher Stämme.

	-	C	CC	CCC
V	18	474	958	96
VV	66	328	37	5
VVV	5	22	-	-

	-	C	CC	CCC
V	54	664	1.141	52
VV	124	572	132	21
VVV	9	17	5	-

Wenn wir nun wiederum diese Verteilungen daraufhin untersuchen, ob die Kombinationen signifikant von den statistisch erwartbaren Verteilungen abweichen, ergibt sich ein ganz anderes Bild als in den vorangegangenen Abschnitten – und dieses Bild ist in beiden Sprachen sehr ähnlich.

Im deutschen und im englischen Korpus (Tab. 49 und 50) kombinieren einfache Vokalgrapheme (erste Zeile) bevorzugt mit zwei- oder dreifach geschlossenen Endrändern, komplexe Kerne (zweite und dritte Zeile) hingegen umgekehrt mit offenen oder einfach geschlossenen Endrändern. Hier liegt – wie ja phonographisch zu erwarten war – ein eindeutiger Zusammenhang zwischen der Besetzung von Kern und Endrand vor.

Tab. 49 (links) und Tab. 50 (rechts): Zusammenhang zwischen der Besetzung des Kerns mit 1, 2 oder 3 Graphemen (Zeilen) und des Endrands mit 0, 1, 2 oder 3 Graphemen (Spalten) in graphematischen Einsilbern im deutschen (links) und englischen (rechts) Teilkorpus einfacher Stämme; klassifiziert nach der Wahrscheinlichkeit, dass der reale Wert zufällig ist.

	-	C	CC	CCC
V	---	---	+++	+
VV	+++	+++	---	---
VVV	++	++	U	V

	-	C	CC	CCC
V	---	---	+++	++
VV	+++	+++	---	---
VVV	+++	N	--	V

Dieser Zusammenhang ist im deutschen Korpus noch ausgeprägter als im englischen. Wenn wir uns die Tabelle mit den absoluten Zahlen anschauen (Tab. 47 oben), dann lässt sie sich so beschreiben: Von $\langle VC \rangle$ als Ausgangspunkt kann in zwei Richtungen abgewichen werden: Entweder der Endrand wird komplexer oder der Kern. Beides zusammen ist höchst selten. Es gibt insgesamt nur 33 Belege für einen komplexen Kern, der von einem komplexen Endrand gefolgt wird, z. B. die folgenden:

- (31) Biest, dreist, Feind, Freund, Geist, Haupt, Kiosk, Klient, leugn[en],
seufz[en]

Im englischen Korpus gibt es wesentlich mehr Ausnahmen (z. B. die 132 Stämme vom Typ <VVCC> wie etwa *point*). Das ist nicht überraschend: Zum einen scheint es auch phonologisch mehr Ausnahmen vom maximalen Reim zu geben. Zum anderen korrespondieren auch ungespannte Vokalphoneme z. T. mit komplexen Silbenkernen, vgl. etwa *breast*, *wealth*. Es ist also zu erwarten, dass es in der Schrift mehr Ausnahmen zum Längenausgleich gibt als in der Phonologie.

Kehren wir noch einmal zum oben angesprochenen Zusammenhang zwischen Anfangsrand und Endrand graphematischer Silben im Deutschen zurück. Roemheld (1955) beobachtet, dass die Verteilung des sog. ‚Dehnungs-⟨h⟩‘ auch von der Besetzung des Anfangsrandes abhängt: In Stämmen von ‚Sinnwörtern‘, die auf <ɫ>, <ɱ>, <ŋ> oder <ʀ> enden und die ansonsten aus drei Buchstaben bestehen würden, ‚beschwert‘ das ⟨h⟩ den Stamm und verleiht den ansonsten gewissermaßen zu leichten Wörtern (*⟨Mel⟩, ⟨Lem⟩, ⟨Ere⟩ etc.) mehr Gewicht (32a):

- (32a) ahm[en], dehn[en], Ehre, fahr[en], fehl[en], Lehm, Mehl, Mohn
(32b) bequem, Flur, Gral, grün, spar[en], spül[en], Kram, Krone

In den Fällen, in denen die Stämme ohne ⟨h⟩ bereits aus vier oder mehr Buchstaben bestehen, tritt dementsprechend kein ⟨h⟩ auf (32b). Es gibt Ausnahmen: zum einen kurze Wörter ohne ⟨h⟩ (z. B. *Tal*), zum anderen längere Wörter mit ⟨h⟩ (z. B. *Strahl*). Insgesamt machen die Wörter vom Typ (32a, b) aber laut Roemheld etwa 70 % aller einschlägigen Wörter aus, die Ausnahmen nur rund 30 %. Interessant ist in diesem Zusammenhang ein Hinweis Roemhelds, auf den später eingegangen werden soll: Der Unterschied zwischen der Regel und den Ausnahmen werde deutlicher, „wenn man die Wörter nicht bloß zählt, sondern auch wägt“ (Roemheld 1955: 79). Viele Ausnahmen kommen selten vor oder sind regionale Varianten; sie sollten nicht so stark ins Gewicht fallen wie hochfrequente Wörter des Kernwortschatzes.

Das ⟨h⟩ ist damit in dieser Verwendung ein rein graphematisches Mittel – man könnte es als Wohlgeformtheitsbeschränkung über Stämme von lexikalischen Wörtern modellieren. Inhaltswörter (auch ‚Autosemantika‘)⁴² müssen

⁴² Die Kategorien Autosemantika/Synsemantika bzw. Inhaltswörter/Funktionswörter sind nicht trennscharf (vgl. z. B. Brauße 1994). Wir gehen hier den wohl einzigen gangbaren Weg und klassifizieren Substantive, Verben, Adjektive und Adverbien extensional als Autosemantika und die übrigen Wortarten als Synsemantika.

bestimmte Mindestanforderungen erfüllen: Sie müssen einen Silbenkern haben und außerdem aus mindestens drei Buchstaben bestehen (siehe unten) – und in diesen minimalen Wörtern sind wiederum einfaches postvokalisches ⟨l⟩, ⟨m⟩, ⟨n⟩ oder ⟨r⟩ verboten.

Roemhelds funktionale Erklärung dieser Verteilung ist allerdings in mehrererlei Hinsicht problematisch. Er vermutet, dass die fehlende Ober- und Unterlänge von ⟨m⟩, ⟨n⟩ und ⟨r⟩ durch das ⟨h⟩ kompensiert wird, da diese Buchstaben „beim Lesen leicht übersehen“ werden (Roemheld 1955: 79).

[Das ⟨h⟩, KB] stellt sich schützend vor die einstufigen Zeichen: mit seiner langen, schmalen Form hält es das lesende Auge für einen ‚Augenblick‘ fest, und das kommt seinem Nachbarn zugute. Man liest nicht über ihn hinweg, sondern hat beim kurzen Anhalten des Auges gerade Zeit genug, ihn richtig zu erfassen. (ebd.)

Diese Idee des Lesens als zeichenweises Prozessieren ist aus heutiger Sicht nicht haltbar. Wir wissen mittlerweile, dass jeweils ganze Wortteile gleichzeitig verarbeitet werden und dass das Auge sich in einer Folge von Sakkaden und Fixationen über geschriebenen Text bewegt (vgl. für einen Überblick z. B. Rayner et al. 2012). Das ⟨h⟩ muss das Auge sicher nicht ‚festhalten‘, damit als Nebeneffekt der folgende Buchstabe ebenfalls verarbeitet werden kann. Das ist das eine Problem; ein weiteres betrifft das ⟨l⟩. Mit seiner Oberlänge widerspricht es Roemhelds These; die entsprechenden Wörter sind für Roemheld aber „nur geschichtlicher Ballast, von dem ein guter Teil abgeworfen werden könnte“ (Roemheld 1955: 79, Fn. 9).⁴³

Darüber hinaus gibt es weitere Buchstaben ohne Ober- und Unterlänge, vor denen ⟨h⟩ ebenfalls zu erwarten wäre. Roemheld selbst nennt ⟨c⟩, ⟨v⟩ und ⟨w⟩ und stellt in diesem Zusammenhang fest, dass sie „nach langem Vokal kaum vorkommen“ (Roemheld 1955: 79). Das stimmt zwar – allerdings gibt es unter den vereinzelt Stämmen keinen einzigen, der mit ⟨h⟩ geschrieben wird (*ewig, Möwe, Löwe, brav, Luv*). Und das sind nur die Buchstaben, die in der Fraktur ohne Ober- und Unterlänge sind; im lateinischen Alphabet kommen ⟨s⟩ und ⟨z⟩ dazu. Für ⟨s⟩ findet sich nun eine Reihe von nativen Wörtern, bei denen ⟨h⟩ zu erwarten wäre, aber nicht auftritt (z. B. *Besen, Busen, Dose, Hase, Hose, Käse, Nase, ras[en]*).⁴⁴

⁴³ Das ist natürlich methodisch angreifbar: Die Belege mit ⟨l⟩ nutzt Roemheld genau wie die mit ⟨m⟩; ⟨n⟩ und ⟨r⟩, um bei dem globalen Wert von 30 % Ausnahmen zu landen – ohne ⟨l⟩ wäre dieser Wert unter Umständen höher, aber das wird nicht weiter diskutiert.

⁴⁴ Dieses Gegenargument ist zugegeben schwach – gedruckt und geschrieben wurde eben über Jahrhunderte Fraktur, und wenn man eine funktionale Adaption der Schrift an die geänderten

Und schließlich gibt es auch methodische Probleme: So bleibt unklar, wie Roemheld zu seinen Daten kommt. Sind sie introspektiv gewonnen oder wurden sie systematisch (z. B. in Wörterbüchern) gesucht? Potenziell problematisch ist auch die Beschränkung auf „deutsche Wörter und alteingebürgerte Lehnwörter“ (Roemheld 1955: 75) sowie der Ausschluss von „Wörter[n], die nur in Wörterbüchern ein kümmerliches Dasein führen“ (ebd.). Ohne (zumindest etwas) genauere Kriterien über den Ein- und Ausschluss besteht die Möglichkeit, dass die Ergebnisse der Untersuchung im Sinne eines gewünschten Ergebnisses verzerrt sind (das kann ja durchaus ein unbewusster Prozess sein).

Insgesamt schießt Roemheld mit seiner Erklärung also etwas über das Ziel hinaus – er identifiziert (mithilfe eines mittlerweile überholten Lesemodells) einen Teil der Daten als systemhaft und die sich daraus ergebenden Ausnahmen sollen beseitigt werden. Das ändert nichts an der Tatsache, dass die grundsätzliche Beobachtung der Interaktion richtig und interessant ist und dass es sich beim <h> in dieser Funktion um ein rein graphematisches Mittel handelt.⁴⁵

Roemhelds Beobachtungen wurden von Augst (1986) größtenteils empirisch bestätigt. Augst untersucht das Dehnungs-<h> als eines von mehreren Mitteln, um

Leseverhältnisse und -bedürfnisse vermutet, dann in dieser Schriftart und nicht in der erst (aus historischer Perspektive) seit Kurzem verwendeten Antiqua.

45 Außerdem muss man anerkennend festhalten, dass Roemheld eine Reihe von grundlegenden Überzeugungen vertritt, die sehr modern klingen und damals sicher ihrer Zeit voraus waren. So fordert er etwa, die <h>-Schreibung aus Lesersicht zu betrachten; die Bedürfnisse der Leser zu erfüllen sei „der einzige Zweck aller Rechtschreibung“ (Roemheld 1955: 79). Das wiederum bedeutet, „das Wortbild so zu gestalten, dass das Lesen so leicht und glatt wie möglich vonstatten gehen kann“ (Roemheld 1955: 81). Hier leistet das Dehnungs-<h> einen Beitrag „zur Lesbarkeit deutscher Texte“ (Roemheld 1955: 80). Die Herausbildung dieser <h>-Schreibung lief dabei – wie graphematischer Wandel laut Roemheld generell – ungesteuert ab:

Das ursprüngliche Lautzeichen h [= ein Zeichen mit einem korrespondierenden Phonem, KB] wurde nach der Auffassung der wenigen, die sich Gedanken darüber machten, zum Lesezeichen [= ein Zeichen, das die phonographische Korrespondenz eines anderen Zeichens determiniert, KB], wurde dabei aber von vornherein ohne ihr Wissen und Wollen, „unter ihren Händen“, möchte man sagen, in seiner Verwendung so eingeschränkt, dass es seine eigentliche Wirksamkeit auf einem ganz anderen Gebiet entfaltete: es kam dem zu Beginn der Neuzeit sich immer mehr ausbreitenden Lesebedürfnis und der damit verbundenen Notwendigkeit, das Lesen so viel wie möglich zu erleichtern, entgegen, indem es als rein formales Gebilde dazu beitrug, dem Wortbild sinnwichtiger Wörter die Gestalt zu verleihen, die die Gewähr bot für eine von unliebsamen Stockungen freie Aufnahme durch das lesende Auge. (Roemheld 1955: 82).

Hier klingt der Gedanke an, dass die Graphematik einer Sprache ein selbstorganisierendes System sei und dass der Wandel zwar von Schreibern verursacht, aber nicht intendiert wurde – das ist gleichsam der Kern von Kellers Idee der ‚unsichtbaren Hand‘ im Sprachwandel (Keller 1994).

die Vokallänge zu markieren. Er stellt die folgenden Regularitäten auf (Augst 1986: 312f.):

- Das Dehnungs-⟨h⟩ steht ...
- nur bei nativen Wörtern
 - nur in der Stammsilbe, nicht in Affixen und ‚Nebensilben‘
 - nur vor einfachem postvokalischem Konsonant im Abfeld der Stammsilbe
Mit ‚Abfeld‘ sind alle Konsonanten gemeint, die dem Vokal der Stammsilbe folgen. Die Ausnahmen sind hier laut Augst *ahnden*, *fahnden*, *öhmden*, *Wöhrde* sowie die Varianten *Föhrde*, *Mährte* (neben *Förde*, *Märte*).
 - nur vor einfachem ⟨b⟩, ⟨m⟩, ⟨n⟩ oder ⟨r⟩ im Abfeld
Die Ausnahmen sind hier *Fehde*, *Lehde* sowie die Variante *Lohde* (neben *Lode*).
 - nur bei synsemantischen Lexemen
Als Alternative schlägt Augst das (besser zu operationalisierende) Kriterium der Flektierbarkeit vor; Ausnahmen sind *ohne*, *mehr*, *sehr*,⁴⁶ *eh*.
 - nur bei leerem Anfangsrand („Anfeld“ bei Augst (Hg.) 1985), nicht wenn der Anfangsrand zwei- oder mehrfach besetzt ist (Augst 1986: 313).

Dieser letzte Punkt ist besonders relevant. Hier bezieht sich Augst direkt auf Roemhelds Analyse und überprüft sie empirisch. Er kommt auf der Grundlage des Morpheminventars aus Augst (1975) zu folgenden Ergebnissen für Substantive, Verben und Adjektive (Augst 1986: 314 ff.):

Tab. 51: Anzahl der Lexeme im Morpheminventar von Augst (1975), klassifiziert nach Anzahl der Konsonanten im Anfeld sowie An- und Abwesenheit von postvokalischem ⟨h⟩ vor ⟨b⟩, ⟨m⟩, ⟨n⟩ und ⟨r⟩, Zahlen aus Augst (1986: 314).

#C im Anfeld	#+⟨h⟩	#-⟨h⟩	% +⟨h⟩
0	15	8	65 %
1	153	145	51 %
2	24	72	25 %
3+	4	47	8 %

Relativ eindeutig sind die Verhältnisse für zwei oder mehr Konsonanten im Anfangsrand: Hier überwiegen klar die Wörter ohne ⟨h⟩. Umgekehrt kommt in

⁴⁶ Eisenberg (2013a: 408) argumentiert dafür, dass *mehr* und *sehr* auf *mehren/mehrere* bzw. *versehren* beziehbar sind; damit sind diese Wörter keine eindeutig Nichtflektierbaren.

etwa zwei Drittel der Wörter mit leerem Anfangsrand das ⟨h⟩ vor. Unentschieden sieht die Situation bei einfach besetztem Anfangsrand aus; hier herrscht laut Augst (1986: 315) „echte Regellosigkeit“ (zu ganz ähnlichen Zahlen kommt Evertz 2014: 158). Daran ändert sich auch nichts, wenn man sich die Vorkommenshäufigkeit der knapp 2.000 häufigsten Wörter (nach Meier 1964) anschaut – auch dann tritt die ⟨h⟩-Schreibung nur in etwa der Hälfte aller Fälle auf. Die übrigen Kategorien werden allerdings eindeutiger: 88% der Textwörter mit leerem Anfangsrand enthalten ⟨h⟩, während das für 0% der Textwörter mit zwei oder mehr Konsonanten der Fall ist. Die Verteilung in Tabelle 51 ist insofern unschön, als die meisten Wörter genau einen Konsonanten im Anfangsrand aufweisen (vgl. 3.2.2.2), und genau für diese frequente Kategorie kann Augst keine Aussagen machen. Der einzige weitere Hinweis, den Augst in diesem Zusammenhang liefern kann, ist die Abhängigkeit von der Nebensilbe: Folgt der Stammsilbe eine weitere Silbe, die nicht nur aus ⟨e⟩ besteht, dann tritt in etwa zwei Drittel der Fälle kein ⟨h⟩ auf; es geht hier um Fälle wie *Honig* gegenüber Fällen wie *Sahne*.

Im Folgenden soll versucht werden, die bekannten Regularitäten für das Dehnungs-⟨h⟩ am Korpus deutscher Stämme empirisch zu überprüfen und auch für die Stämme mit einfachem Anfangsrand zu brauchbaren Aussagen zu kommen. Dazu wird zunächst eine Liste mit Stämmen aus dem Korpus extrahiert, die ein postvokalisches ⟨h⟩ gefolgt von mindestens einem weiteren Konsonantengraphem aufweisen. Im deutschen Teilkorpus sind das 138 Stämme. Diese Wörter sind die Grundlage für die Untersuchung. Analysiert werden der folgende Konsonant, der Typ des Silbenkerns vor dem ⟨h⟩, die Wortart, die Silbenzahl sowie der Anfangsrand.

Folgende Konsonantengrapheme treten direkt nach ⟨h⟩ auf:

Tab. 52: Anzahl der Stämme, in denen postvokalisches ⟨h⟩ jeweils vor ⟨n⟩, ⟨l⟩, ⟨r⟩, ⟨m⟩, ⟨t⟩ und ⟨d⟩ vorkommt.

Konsonant	#Stämme	Beispiel
n	43	sehn[en], Fahne
l	39	fühl[en], Stuhl
r	38	fahr[en], Ohr
m	15	nehm[en], Rahm
t	2	Draht, Naht
d	1	Fehde

Wie zu erwarten kommt postvokalisches ⟨h⟩ fast ausschließlich vor ⟨n⟩, ⟨l⟩, ⟨r⟩ und ⟨m⟩ vor. Weitere Konsonanten kommen nach diesen Graphemen nur in vier Fällen

vor: *ahnd[en]*, *fahnd[en]*, *Fähnrich*, *verwahrlos[en]*. Die letzteren beiden sind in gewissem Sinne morphologisch komplex – *Fähnrich* etwa ist zumindest etymologisch auf *Fahne* zu beziehen, und *verwahrlosen* enthält sowohl ein Derivationspräfix als auch ein Derivationsuffix (auch wenn der ‚Stamm‘ selbst nicht ohne Weiteres auf einen freien Stamm zu beziehen ist). Im Normalfall folgt also, so können wir zusammenfassen, nach ⟨h⟩ noch genau ein Konsonantengraphem, und zwar ein kurzköpfiges (⟨n⟩, ⟨b⟩, ⟨r⟩, ⟨m⟩).

Was den Kern angeht, so gibt es keinen einzigen Fall, in dem postvokalisches ⟨h⟩ nach einem komplexen Silbenkern auftritt. Alle Silbenkerne sind ausnahmslos einfach – ein Fall wie *⟨Rauhm⟩ scheint kategorial ausgeschlossen zu sein. Das ist sowohl graphemisch wie auch phonographisch nicht überraschend: Graphemisch treten ⟨VV⟩-Kerne selten mit komplexem Endrand auf (siehe oben); auch postvokalisches ⟨h⟩ ist vergleichsweise selten; schon rein stochastisch sollte ⟨VhC⟩ also sehr selten sein. Und phonographisch verweist der Name ‚Dehnungs-⟨h⟩‘ bereits auf die Funktion, die Gespanntheit zu kennzeichnen. Hier gibt es bei komplexen Kernen keine Ambiguitäten, eine Kennzeichnung mit komplexem Kern *und* postvokalischem ⟨h⟩ wäre also redundant.⁴⁷ Darüber hinaus tritt ⟨h⟩ nur in einem Fall nach ⟨i⟩ auf, dem Personalpronomen *ihr*⁴⁸ (vgl. Augst 1986: 310 f.).

Hinsichtlich der Wortart der 138 Stämme mit Dehnungs-⟨h⟩ bestätigt sich auch im vorliegenden Korpus die Beobachtung, dass ⟨h⟩ vor allem in flektierenden Stämmen auftritt:

Tab. 53: Anzahl der Stämme, in denen postvokalisches ⟨h⟩ und ein weiterer Konsonant vorkommen, klassifiziert nach Wortart (substantivisch, verbal, adjektivisch, adverbial, präpositional oder pronominal).

Wortart	#Stämme	Beispiel
Substantiv	72	Stahl, Sahne
Verb	49	fehl[en], nehm[en]
Adjektiv	12	hohl, lahm
Adverb	3	sehr, wohl
Präposition	1	ohne
Pronomen	1	ihr

⁴⁷ Problematisch an diesem Erklärungsmuster ist allerdings, dass Redundanz in der Sprache im Allgemeinen und in der Schrift im Speziellen nichts Außergewöhnliches ist.

⁴⁸ Die Formen *ihn* und *ihm* enthalten natürlich ebenfalls ⟨h⟩ nach ⟨i⟩; da sie zum selben Paradigma gehören, werden sie hier nicht mitgezählt.

Interessant ist weiterhin, welche silbische Struktur die betreffenden Stämme mit ⟨h⟩ aufweisen. An welcher Position treten graphematische Silben mit ⟨h⟩ auf? Überwiegend in Einsilbern und ersten Silben von Zweisilbern, wie Tabelle 54 zeigt. Diese Verteilung – die Konzentration auf Einsilber und erste Silben von Zweisilbern – ist für eine native Schreibung nicht überraschend. Die 15 Ausnahmen (zwölf Zweisilber mit ⟨h⟩ in der zweiten Silbe und drei Dreisilber) sind alle nicht eindeutig morphologisch einfach (neben den in Tab. 54 vermerkten handelt es sich um *Argwohn*, *befehl[en]*, *begeh[r]en*, *erwäh[n]en*, *gebü[h]r[en]*, *Gefahr*, *Gemahl*, *gewöh[n]en*, *verhohlen*, *verseh[r]en*, *vornehm*). Sie sind meist einseitig segmentierbar, etwa über die Präfixe *be-*, *ge-* oder *ver-*. Wenn wir uns die 35 Zweisilber mit ⟨h⟩ in der ersten Silbe anschauen, fällt auf, dass der Großteil ⟨e⟩ in der zweiten Silbe enthält (*ähnell[n]*, *Bühne*, *Rahmen*). Hier gibt es nur fünf Ausnahmen (*Ahming*, *Brahma*, *Dahlie*, *Fähnrich* und *Kehraus*). Im Normalfall, so können wir zusammenfassen, tritt ⟨h⟩ also in Einsilbern oder Zweisilbern mit ⟨e⟩ in der zweiten Silbe auf – den ‚kanonischen Trochäen‘ nach Primus (2010).

Tab. 54: Anzahl der Stämme, in denen postvokalisches ⟨h⟩ und ein weiterer Konsonant vorkommen und die graphematisch einsilbig, zweisilbig oder dreisilbig sind; bei Zweisilbern zusätzlich, ob ⟨h⟩ in der ersten oder zweiten Silbe auftritt.

#Silben		#Stämme	Beispiel
1		88	sehn[en], Stuhl
2	⟨h⟩ 1. Silbe	35	Fahne, Rahmen
	⟨h⟩ 2. Silbe	12	verwöhn[en], empfehl[en]
3		3	allmählich, angenehm

Eine letzte Regelmäßigkeit betrifft den Anfangsrand: Hier treten im verwendeten Korpus – wie von Eisenberg (2013a: 303) beobachtet – nie einfaches ⟨p⟩ und ⟨t⟩ auf. Zumindest die Verteilung von ⟨t⟩ lässt sich historisch erklären: Im nativen Wortschatz wurde initiales ⟨t⟩ bis in die zweite Hälfte des 19. Jahrhunderts als ⟨th⟩ verschriftet, und dieses ⟨th⟩ trat nie gemeinsam mit einem postvokalischem ⟨h⟩ auf.

Wenn wir aus den bisherigen Erkenntnissen einen operationalisierbaren⁴⁹ Kernbereich für das Auftreten des Dehnungs-⟨h⟩ definieren wollen, dann sieht

⁴⁹ Diese Einschränkung schließt die notorisch schwierige Kategorisierung hinsichtlich des Fremdwortstatus erst einmal aus.

dieser Kernbereich so aus: ⟨h⟩ tritt a) in graphematisch einsilbigen Stämmen oder in zweisilbigen Stämmen mit ⟨e⟩ als zweitem Silbenkern auf (dort in der ersten Silbe); b) nach einem einfachen Vokalgraphem (allerdings nicht nach ⟨i⟩) und c) vor einfachem ⟨l⟩, ⟨m⟩; ⟨n⟩ oder ⟨r⟩ und zwar d) bei flektierenden Wortarten; e) ⟨t⟩ ist im Anfangsrand ausgeschlossen. Der Kernbereich umfasst im Korpus einfacher deutscher Stämme 111 Einträge. Es lohnt sich, noch einmal festzuhalten, dass die typische Verteilung des Dehnungs-⟨h⟩ hier graphemisch definiert wird – es ist erstmal kein Bezug zur Phonologie notwendig.

Wir können nun den Einfluss des Anfangsrandes untersuchen, indem wir die Blickrichtung umdrehen. Dafür müssen wir zunächst bestimmen, wie viele andere Wörter den Anforderungen oben genügen, die kein ⟨h⟩ enthalten. Anders gefragt, wie viele Ausnahmen gibt es für den Kernbereich generell? Wie viele Ein- und Zweisilber enthalten einfaches ⟨l⟩, ⟨m⟩, ⟨n⟩ oder ⟨r⟩ nach einfachem Vokal? Im Korpus einfacher deutscher Stämme sind das 141; ⟨h⟩ steht also im Kernbereich nur in 44 % aller Fälle. Tabelle 55 zeigt nun die Verteilung der Stämme in Abhängigkeit von der Besetzung des Anfangsrandes. Für die Stämme ohne Anfangsrand sowie für die mit drei oder mehr Konsonanten gefüllten Anfangsränder ergeben sich Verhältnisse wie in der Untersuchung von Augst (1985), die oben zitiert wurde (dort 65 % bzw. 8 %). Der Anteil der ⟨h⟩-Schreibungen bei den einfach besetzten Anfangsrändern ist aber höher, derjenige der zweifach besetzten Anfangsränder niedriger als bei Augst (dort 51 % bzw. 25 %).

Tab. 55: Anteil der ⟨h⟩-Schreibungen in Abhängigkeit von der Besetzung des Anfangsrandes mit 0, 1, 2 und 3 oder mehr Graphemen. Datenbasis: alle Ein- und Zweisilber mit einfachem Vokal und folgendem einfachem ⟨l⟩, ⟨m⟩, ⟨n⟩ oder ⟨r⟩ im Teilkorpus einfacher deutscher Stämme.

#Konsonanten im Anfangsrand	#Stämme mit ⟨h⟩	#Stämme ohne ⟨h⟩	% Stämme mit ⟨h⟩
0	13	6	68 %
1	86	58	60 %
2	10	51	16 %
3+	2	26	7 %

Wir können nun zusätzlich die Wortart miteinbeziehen. Zwar sind ⟨h⟩-Schreibungen, wie oben gezeigt, in allen flektierenden Wortarten vertreten; Eisenberg (2013a: 312) stellt aber fest, dass postvokalisches ⟨h⟩ unter den Verben häufiger auftritt als in anderen Wortarten: „Insgesamt steht es nur in jedem zweiten möglichen Fall, bei den Verben aber viel häufiger.“ In Tabelle 56 wird die Information über den Anfangsrand mit der Information über die Wortart kombiniert.

Tab. 56: Anteil der ⟨h⟩-Schreibungen in Abhängigkeit von der Besetzung des Anfangsrandes mit 0, 1, 2 und 3 oder mehr Graphemen und der Wortart (Adjektiv, Substantiv, Verb).
 Datenbasis: alle Ein- und Zweisilber mit einfachem Vokal und folgendem einfachem ⟨b, m, n⟩ oder ⟨r⟩ im Teilkorpus einfacher deutscher Stämme.

#Konsonanten im Anfangsrand	Adjektiv % ⟨h⟩	Substantiv % ⟨h⟩	Verb % ⟨h⟩
0	0 % (0/1)	64 % (9/14)	100 % (4/4)
1	58 % (7/12)	50 % (46/92)	83 % (33/40)
2	0 % (0/4)	13 % (6/46)	27 % (4/11)
3+	0 % (0/5)	13 % (2/16)	0 % (0/7)
gesamt	32 % (7/22)	38 % (63/168)	66 % (41/62)

Die Beobachtung von Eisenberg kann also bestätigt werden: Der Anteil der ⟨h⟩-Schreibung liegt bei den Verben höher als bei den übrigen Wortarten (Zeile ‚gesamt‘ in Tab. 56). Warum Verben? Eisenberg (2013a: 313) vermutet eine phonographische Funktionalisierung: Flexionsformen wie ⟨holst⟩ mit dreifach geschlossenem Endrand sehen aus, als hätte die korrespondierende phonologische Form einen ungespannten Vokal. Das ⟨h⟩ in ⟨zahlst⟩ zeigt „tatsächlich die Dehnung an“ (ebd.). Diese Erklärung hat allerdings zwei Schwachpunkte: Zum einen tritt ⟨h⟩ ja, wie wir gesehen haben, nur vor einfachem ⟨b, m, n⟩ und ⟨r⟩ auf. Nach Eisenbergs Erklärung sollten wir es auch vor anderen einfachen Konsonanten im Endrand eines Stammes erwarten. Tatsächlich gibt es im CELEX-Teilkorpus morphologisch einfacher Stämme 72 Gegenbeispiele wie ⟨hebst⟩, ⟨lobst⟩, ⟨rufst⟩ und ⟨sagst⟩; besonders ⟨g⟩ ist als einfacher Endrand sehr häufig vertreten. Demgegenüber gibt es ebenfalls 72 Verbstämme mit einfachem ⟨b, m, n⟩ oder ⟨r⟩, von denen auch nicht alle differenziert werden. Es ist also nicht so, dass ⟨b, m, n⟩ und ⟨r⟩ wesentlich häufiger in Verbstämmen vorkommen als andere Konsonanten. Zum anderen haben wir gesehen, dass die Verteilung von ⟨h⟩ auch bei Verben abhängig von der Besetzung des Anfangsrandes ist. Phonographisch-funktional macht das wenig Sinn: In Formen wie ⟨sparst⟩ ist der Endrand schließlich komplex; die Wortform müsste ebenfalls eine Dehnungsmarkierung erhalten wie in ⟨zahlst⟩.

Eisenbergs Erklärung muss dennoch nicht verworfen werden. Die Schrift kann ja auch in einem engeren Bereich funktionalisiert sein, ohne das Schema auszuweiten auf analoge Konstruktionen. Mit anderen Worten: Es spricht nichts dagegen, dass ⟨h⟩ bei Verben nur vor ⟨l⟩, ⟨m⟩, ⟨n⟩ und ⟨r⟩ funktional ist. Die Schreibung kann genutzt werden – sie ist aber nicht notwendig, wie die zahlreichen Beispiele vom Typ ⟨sagst⟩ deutlich machen. Weiter unten (3.2.3) wird die ⟨h⟩-Schreibung als ‚Gewichtsausgleich‘ motiviert. Ein Nebeneffekt ist, dass die Vokalqualität in bestimmten Umgebungen markiert wird und dass diese Markierung vom Leser genutzt werden kann.

Die Zahlen in Tabelle 54 werden noch eindeutiger, wenn man nicht nur zählt, sondern auch wägt, wie Roemheld vorschlägt. Bei den leeren Anfangsrändern entfernen wir *Ahm* (als Variante zu *Ohm*) sowie das Fremdwort *Ale* und *Elen* (veraltet für *Elch*). Die verbleibenden 15 Wörter enthalten zu 73% ⟨h⟩; die vier Ausnahmen sind *Ar*, *elend*, *Öl* und *Ur*.

Unter den einfach besetzten Anfangsrändern finden sich Wörter wie die folgenden:

- (33a) Bor, Dom, Fan, Före, Gel, Gen, Golem, Harem, Karenz, Mol, Moment, Nomen, None, rar, Ren, Rum, Run, Selen, seren, Sol, Vene, Yen, Zar, Zen
 (33b) Kamel, Komet
 (33c) bohn[en], Dämel, Feme, Föhn, gehr[en], hehl[en], kör[en], Kule, Mahr, nöl[en], nör[en], sohr

Die Wörter in (33a) sind Fremdwörter. Da das Dehnungs-⟨h⟩ nur in nativen Wörtern auftritt (vgl. Roemheld 1955), ist ein Dehnungs-⟨h⟩ bei den Wörtern in (33a) nicht zu erwarten. Die Kategorisierung als Fremdwort ist natürlich schwierig; die Liste in (33a) enthält sowohl synchron einfach zu identifizierende Fremdwörter (*Yen*, *Zen*) wie auch einige intransparente Fälle (z. B. *Före*). Man kann sich über den Status einiger Wörter streiten; mehr als die oben angegebenen Wörter finden sich aber wohl nicht. Die Wörter in (33b) sind schon länger im Deutschen, allerdings sind sie etymologisch fremd und phonologisch Jamben; auch das kann als Fremdwortmarker interpretiert werden. Die Wörter in (33c) schließlich sind selten: Sie kommen im Korpus des Mannheimer Morgen nicht vor; dieses Korpus umfasst 5,4 Millionen Textwörter. Das deckt sich gut mit der introspektiven Beobachtung, dass es für viele dieser Wörter schwierig ist, Bedeutungen anzugeben.

Bei den zweifach und mehr besetzten Anfangsrändern werden durch die beiden Kriterien folgende Stämme ausgeschlossen:

- (34a) Bram, Chlor, Chor, Chrom, Flor, Khan, Klan, Krem, Phon, Phonem, Planet, Schal, Share, Slum, Szene, Tram
 (34b) Bräme, Brom, Drohn, Gral, klon[en], Klon, Prahm, Schäre, Schram, skor[en], spä[n]en, Splen, Stär, Ster

Die Wörter in (34a) sind Fremdwörter, die Wörter in (34b) kommen im Korpus des Mannheimer Morgen nicht vor. Der Großteil der Wörter aus beiden Gruppen enthält kein ⟨h⟩, sodass sich die Verhältnisse etwas zugunsten der ⟨h⟩-Schreibung verschieben.

Tabelle 57 zeigt, wie sich die Verteilung der ⟨h⟩-Schreibung mit diesen Modifikationen ändert: Insgesamt bleiben 181 Stämme übrig. Diese verteilen sich wie folgt auf die Anfangsränder bzw. Wortarten:

Tab. 57: Modifizierte Tabelle 56 nach Ausschluss der Fremdwörter und infrequenten Wörter.

#Konsonanten im Anfangsrand	Adjektiv % ⟨h⟩	Substantiv % ⟨h⟩	Verb % ⟨h⟩	gesamt % ⟨h⟩
0	0 % (0/1)	70 % (7/10)	100 % (4/4)	73 % (11/15)
1	75 % (6/8)	73 % (44/60)	88 % (30/34)	78 % (80/102)
2	0 % (0/4)	16 % (4/25)	44 % (4/9)	21 % (8/38)
3+	0 % (0/5)	18 % (2/11)	0 % (0/7)	9 % (2/23)
gesamt	33 % (6/18)	53 % (57/106)	70 % (38/54)	

Mit diesen Modifikationen ergibt sich ein recht klares Bild: Ist der Anfangsrand leer oder einfach besetzt, tritt meist ein ⟨h⟩ auf; ist der Anfangsrand mit mehr als einem Element besetzt, tritt es meist nicht auf (Spalte ‚gesamt‘).⁵⁰ Bei den Verben

⁵⁰ Die von Evertz (2014: 158 f.) beobachtete graduelle Abnahme des Anteils von ⟨h⟩-Schreibungen lässt sich hier (wie auch oben in Tab. 55) nur eingeschränkt beobachten. Das hat theoretische Konsequenzen: Evertz (2014: 161) folgert aus dem graduellen Zusammenhang, dass der Anfangsrand graduell zum Silbengewicht beitragen kann. Aus den hier präsentierten Daten ergibt sich ein anderer Schluss: Unbesetzte und einfache Anfangsränder stehen komplexen Anfangsrändern gegenüber.

liegt der Anteil der <h>-Schreibungen insgesamt höher als bei den übrigen Wortarten. In der mit Abstand größten Gruppe, den einfachen Anfangsrändern, ist der Unterschied zwischen den Wortarten allerdings relativ gering. Das deutet darauf hin, dass der Einfluss der Wortart geringer ist als der Einfluss des Anfangsrandes. Die Generalisierung auf der Basis des Anfangsrandes trägt recht weit: Zu den 142 Fällen, auf denen sie basiert (leere oder einfach besetzte Anfangsränder mit <h>, zweifach oder mehr besetzte Anfangsränder ohne <h>), gibt es nur die 36 Ausnahmen in (35) ((35a) mit Fällen, in denen <h> ‚fehlt‘; (35b) mit Fällen, in denen <h> entgegen der Erwartung auftritt). Die Generalisierung erfasst also 80 % der Daten.

(35a) Ar, Bär, bar, Bar, Dame, elend, gar, gär[en], hol[en], hör[en], Hure, Kur, Kür, Lore, mal[en], Mal, Mär, Mole, Name, Öl, Rune, Samen, Sole, Ur, Wal, Ware

(35b) dröhn[en], Pfahl, Pfuhl, prahl[en], Stahl, stehl[en], stöhn[en], Strahl, Strähne, Stuhl

Was ist nun die Domäne des postvokalischen <h>? Wenn wir Roemheld folgen, ist die Funktion, das betreffende Wort gewissermaßen zu beschweren, ihm mehr graphische Substanz zu verleihen. Daraus folgt, dass postvokalisches <h> in Einsilbern häufiger auftreten sollte als in Zweisilbern. Das ist aber nicht so: Ob der Stamm einsilbig ist (z. B. *kahl*) oder ob er zweisilbig ist, ob er (wenn zweisilbig) mit <e> endet (z. B. *Sahne*) oder anders (z. B. *Dahlie*) – all das ändert das Verhältnis nicht.⁵¹ Wenn sich aber nun *hohl*, *Kohle* und *Fohlen* gleich verhalten, dann spricht das für eine Gliederung in ‚Stamm‘ (*hohl*, *Kohl*, *Fohl*) und ‚Affix‘ (*e*, *en*). Diese Segmentierung wurde ja schon oben bei den intervokalischen Konsonanten beobachtet (vgl. auch Augst 1986: 321). Die Domäne des postvokalischen <h> ist also der Stamm bzw. die Wurzel. Dafür spricht auch ein Teil der Daten, die in Tabelle 55 bis 57 ausgeschlossen wurden. Postvokalisches <h> tritt vor allem in ersten Silben von Zweisilbern auf, deswegen gehören Formen wie *Gefahr* oder *befehl[en]* nicht zum Kernbereich. Auch dort wird aber die Wurzel (*fahr*, *fehl*) beschwert. Dieses Ergebnis wird unten in 3.2.5 im Zusammenhang mit der Struktur minimaler Stämme diskutiert.

Die Generalisierung oben erfasst 80 % der Daten; der Anteil liegt also noch etwas höher als der von Roemheld ermittelte. Das liegt daran, dass wir die Frequenz als Kriterium ins Boot geholt haben. Oben ist das nicht weiter begründet

⁵¹ Bei Einsilbern gibt es im Korpus 22 Ausnahmen unter 125 einschlägigen Stämmen (82,5%); bei Zweisilbern mit finalem <e> sind es 9 Ausnahmen auf 42 einschlägige Stämme (79%); bei Zweisilbern mit anderen finalen Graphemen sind es 10 Ausnahmen auf 48 Stämme (79%).

worden; das soll hier kurz geschehen. Die Frequenz ist einerseits theoretisch interessant: Was passiert, wenn wir uns nur frequentere Stämme anschauen? Wie verschiebt sich das Verhältnis von <h>-Schreibungen zu allen Schreibungen? Andererseits – und das ist hier wichtiger – ist es unmittelbar relevant, wenn die Erkenntnisse der Graphematik praktisch genutzt werden sollen, wenn aus graphematischen Regularitäten didaktische Regeln gewonnen werden sollen.

Wenn wir vorgehen wie Roemheld und Augst und eine möglichst exhaustive Datenbasis anstreben und auf dieser Datenbasis didaktische Empfehlungen ableiten – dann führt das unter Umständen dazu, dass Schülerinnen und Schüler Regeln lernen, die unnötig komplex sind. Und das nur, damit die Regeln Wörter erfassen, die Schülerinnen und Schüler in ihrem Leben wahrscheinlich niemals schreiben werden und denen sie wahrscheinlich nicht einmal begegnen werden. Machen wir uns das an zwei Beispielen kurz deutlich, den Verben *nören* (norddeutsch ‚kurz zwischendurch schlafen‘) und *kören* (‚(ein Tier) zur Zucht auswählen‘). *nören* kommt im Kernkorpus des DWDS gar nicht vor (alle fünf Treffer sind falsche Lemmatisierungen); *kören* fünfmal, aber in genau einem Buch von Erwin Strittmatter. Das DWDS-Kernkorpus umfasst 100 Millionen Textwörter aus dem 20. Jahrhundert. Um einmal dem Verb *nören* zu begegnen, reichen also im Fall des DWDS die 100 Millionen Textwörter nicht, und die fünf Treffer für *kören* sind eben nicht gleichmäßig auf die Gesamtgröße verteilt, sondern stammen alle aus demselben Buch; es ist damit ähnlich selten wie *nören*. 100 Millionen Textwörter scheinen in einem Menschenleben durchaus machbar zu sein (vgl. Baayen/Milin/Ramscar 2016).⁵² In diesem Fall wäre unser fiktiver Leser aber am Ende des Lebens eben noch nicht auf *nören* (und wahrscheinlich auch nicht auf *kören*) gestoßen.

Was ist die Alternative? Es ist sinnvoller, dass didaktische Regeln möglichst genau diejenigen Wörter erfassen, die Schüler verwenden oder wahrscheinlich später in ihrem Leben verwenden werden. Extrem seltene Wörter, landschaftlich geprägte Wörter, Wörter aus Fachsprachen, veraltete Wörter – sie alle gehören nicht dazu. Es ist nicht schlimm, wenn eine Rechtschreibregel sie nicht erfasst: Im unwahrscheinlichen Fall, dass man sie schreiben muss, schaut man in einem Rechtschreibwörterbuch nach. Auf diese Weise kommt die Regel mit weniger Ausnahmen aus, und diese Ausnahmen sind als Liste lernbar.

Zurück zu den Interaktionen zwischen Silbenkonstituenten. Bislang haben wir die Interaktion von Kernen und Rändern relativ grob betrachtet, nämlich

⁵² Baayen, Milin und Ramscar (2016) gehen von einer (theoretischen) Obergrenze von knapp 1,5 Milliarden Textwörtern in 80 Lebensjahren aus – unter der Voraussetzung, dass man 80 Jahre lang alle zwei Sekunden ein Wort hört oder liest und 8 Stunden pro Tag schläft.

unter der Frage, mit wie vielen Segmenten die Konstituenten jeweils besetzt sind. Wir können aber hier einen Schritt weiter gehen und untersuchen, ob es konkrete Grapheme gibt, die nur unter bestimmten Bedingungen auftreten.

Die meisten dieser Restriktionen betreffen den Endrand. Es gibt in beiden Sprachen eine Reihe von Graphemen und Graphemfolgen, die nur nach einfach besetzten Kernen auftreten. Umgekehrt gibt es im deutschen Korpus nur drei Endränder, die nicht nach einfachem Kern vorkommen ($\langle sk \rangle$ in *Kiosk*, $\langle fn \rangle$ in *äufn[en]* und $\langle fz \rangle$ in *seufz[en]*); im englischen Korpus gibt es keinen solchen Endrand. Wir betrachten im Folgenden wiederum nur Einsilber im deutschen und englischen Korpus. Im deutschen Korpus treten die Grapheme und Graphemverbindungen in (36) nur nach einfachem Vokalgraphem auf, und zwar mehr als zehnmals:

(36a) ff, ll, mm, nn, pp, rr, ss, tt

(36b) ck, tz

(36c) hl, hn, hr

(36d) ft, ld, lm, lt, lz, mpf, ng, nk, pf, ps, rm, rn, rch, rsch, rt, rz

Die Aufstellung enthält verdoppelte Grapheme (36a); Graphemverbindungen, die sich verhalten wie Doppelkonsonanten (vgl. oben 3.2.2.3) (36b); Verbindungen aus postvokalischem $\langle h \rangle$ und $\langle b \rangle$, $\langle n \rangle$ und $\langle r \rangle$ (36c) sowie eine große Restgruppe von meist zweifach besetzten Endrändern.

Grapheme und Graphemverbindungen, die sowohl nach einfachen als auch nach komplexen Kernen auftreten, sind etwas schwieriger zu klassifizieren. Hier bietet sich das Verhältnis der Auftretenswahrscheinlichkeiten als Vergleichsmaß an (in der Statistik auch als ‚relatives Risiko‘ bekannt). Wir ermitteln für jeden Endrand die Auftretenswahrscheinlichkeit nach einfachem und nach komplexem Kern; das ist jeweils der Quotient aus der Anzahl des Auftretens und der Summe aller Fälle. Die Graphemfolge $\langle nd \rangle$ kommt bspw. nach einfachem besetztem Kern 38-mal vor; insgesamt gibt es 1.553 Einsilber mit einfach besetzten Kernen. Die Wahrscheinlichkeit für $\langle nd \rangle$ nach einfachem Kern ist also $\frac{38}{1553} = 0,024$. Die Wahrscheinlichkeit für $\langle nd \rangle$ nach komplexem Kern berechnet sich analog; sie liegt bei $\frac{2}{436} = 0,005$. Das Verhältnis der Wahrscheinlichkeiten für $\langle nd \rangle$ liegt also bei $\frac{0,025}{0,002} = 5,33$. Das bedeutet: $\langle nd \rangle$ ist nach einfachen Kernen mehr als fünfmal so wahrscheinlich wie nach komplexen Kernen. Bei Verhältnis von 1 wären beide Vorkommen gleich wahrscheinlich; bei einem Verhältnis zwischen 0 und 1 wäre das Vorkommen nach komplexen Kernen wahrscheinlicher.

Verhältnisse, die relativ ausgewogen sind, sind erst einmal uninteressant. Interessanter sind diejenigen Endränder, die wesentlich häufiger mit einfachem oder mit komplexem Kern auftreten. Wenn wir dementsprechend den Bereich zwischen 0,5 und 2 ausklammern – hier ist eine Wahrscheinlichkeit maximal

doppelt so groß wie die andere, also relativ ähnlich – und wenn wir weiterhin nur diejenigen Endränder betrachten, die insgesamt mehr als zehnmals vorkommen, dann treten die Endränder in Tabelle 58 bevorzugt nach einfachem Kern auf. Die drei komplexen Endränder sind nicht weiter überraschend; sie fügen sich gut in die Liste (36d) oben ein. Interessant ist ⟨x⟩, das sich hier wie ein komplexer Endrand verhält – es ist in einer Gruppe mit Doppelkonsonanten und mehrfach besetzten Endrändern. Es gibt nur einen Beleg mit komplexem Kern und ⟨x⟩, *feix[en]*. Dazu passt, dass ⟨x⟩ nur marginal mit anderen Graphemen kombiniert, vgl. Tabelle 28 oben (3.2.2.3).

Tab. 58: Verhältnis der Wahrscheinlichkeiten des Auftretens von Endrändern nach einfachem und nach komplexem Silbenkern; die Liste umfasst nur solche Endränder, die nach einfachem Kern mindestens doppelt so wahrscheinlich sind wie nach komplexem Kern. Datengrundlage: Liste einsilbiger einfacher deutscher Stämme.

Endrand	#nach ⟨V⟩	#nach ⟨VV⟩	Verhältnis der Wahrscheinlichkeiten
nd	38	2	5,33
nz	11	1	3,09
x	10	1	2,81
tsch	19	2	2,67

Die Endränder in Tabelle 59 sind nach komplexem Kern wahrscheinlicher. Am eindeutigsten liegen die Verhältnisse beim unbesetzten Endrand; er ist etwa 12-mal so wahrscheinlich, wenn der Kern mehrfach besetzt ist. Das ist nicht überraschend, wenn man sich die ‚Minimalanforderungen‘ für Stämme im Deutschen anschaut; dazu in 3.2.3 mehr. Es läuft – kurz zusammengefasst – darauf hinaus, dass Stämme mit einfach besetzten Kernen nur dann möglich sind, wenn a) der Stamm aus mindestens drei Graphemen besteht und b) der Reim aus mindestens zwei Graphemen. Es gibt zwar Stämme, die gegen diese Mindestanforderungen verstoßen, aber es sind – neben Kurzwörtern (*Klo*) und Fremdwörtern (*Pi, pro*) – vor allem Funktionswörter wie z. B. Interjektionen (*ja*), Adverbien (*da, so*) und Pronomen (*du, wo*). Ein leerer Endrand nach einfachem Kern verstößt nun gegen die zweite Mindestanforderung – in *pro* ist der Reim eben nur einfach besetzt.

Außer dem leeren Endrand finden sich in Tabelle 59 nur einfach besetzte Endränder. ⟨ch⟩ besteht aus zwei Buchstaben und wurde oben aus anderen Gründen als Graphem klassifiziert; es reiht sich – anders als ⟨ck⟩ – gut in die Gruppe der einfachen Endränder ein. Interessant ist außerdem, dass die einfachen Endränder und die Doppelkonsonanten fast komplementär verteilt sind: Viele Doppelkonsonanten kommen nie nach komplexen Kernen vor (siehe oben 36a), und

viele einfache Konsonanten sind nach komplexen Kernen wesentlich häufiger als nach einfachen Kernen.

Tab. 59: Verhältnis der Wahrscheinlichkeiten des Auftretens von Endrändern nach einfachem und nach komplexem Silbenkern; die Liste umfasst nur solche Endränder, die nach komplexem Kern mindestens doppelt so wahrscheinlich sind wie nach einfachem Kern. Datengrundlage: Liste einsilbiger einfacher deutscher Stämme.

Endrand	#nach ⟨V⟩	#nach ⟨VV⟩	Verhältnis der Wahrscheinlichkeiten
-	18	66	0,08
f	9	23	0,11
l	20	31	0,18
ß	13	19	0,19
p	5	7	0,20
s	24	30	0,22
m	17	20	0,24
n	32	36	0,25
d	14	15	0,26
ch	39	31	0,35
b	27	21	0,36
t	38	27	0,40
k	9	6	0,42

Im englischen Korpus treten die folgenden Graphemverbindungen nur nach einfachem Vokalgraphem auf, und zwar mehr als zehnmals:

(37a) ll, ff

(37b) g

(37c) ck, ft, lf, lk, mp, ng, rk, rm, sp, wl, wn

Wie wir schon mehrfach gesehen haben, sind stammfinale Konsonantenverdopplungen im englischen Korpus seltener als im deutschen. Daher überrascht der geringe Umfang der Liste in (37a) im Gegensatz zum deutschen Pendant in (36a) oben nicht. Überraschend ist vielmehr, dass einfaches ⟨g⟩ ausschließlich nach einfachem Vokalgraphem vorkommt. Wörter wie *⟨leeg⟩ oder *⟨soag⟩ sind graphematisch (und phonologisch!) keine wohlgeformten Wörter des Englischen.

Wenn wir uns nun diejenigen Endränder anschauen, die wesentlich häufiger (aber eben nicht exklusiv) nach einfachen Kernen vorkommen, dann zeigt sich die Verteilung in Tabelle 60 (auch hier sind nur Endränder mit aufgenommen, die insgesamt mehr als zehnmals vorkommen).

Tab. 60: Verhältnis der Wahrscheinlichkeiten des Auftretens von Endrändern nach einfachem und nach komplexem Silbenkern; die Liste umfasst nur solche Endränder, die nach einfachem Kern mindestens doppelt so wahrscheinlich sind wie nach komplexem Kern. Datengrundlage: Liste einsilbiger einfacher englischer Stämme.

Endrand	#nach ⟨V⟩	#nach ⟨VV⟩	Verhältnis der Wahrscheinlichkeiten
w	64	1	27,46
nk	56	1	24,03
b	60	2	12,87
sh	55	2	11,80
ss	44	2	9,44
tch	41	2	8,79
sk	18	1	7,72
rt	32	2	6,86
rch	13	1	5,58
mb	13	1	5,58
x	24	2	5,15
nch	30	4	3,22
lt	28	4	3,00
rl	13	2	2,79

Neben der Verdoppelung ⟨ss⟩ und den komplexen Endrändern ⟨nk⟩ etc. ist vor allem erstaunlich, dass mit ⟨w⟩, ⟨b⟩ und ⟨x⟩ gleich drei einfache Grapheme in der Liste enthalten sind. Das Verhalten von ⟨b⟩ erinnert an das von ⟨g⟩ oben: Auch ⟨b⟩ ist nach komplexen Kernen so gut wie ausgeschlossen. Es gibt hier nur zwei Ausnahmen, *daub* und *boob*, von denen Letzteres ein Slangwort ist (vgl. OED). Verhalten sich ⟨w⟩, ⟨b⟩ und ⟨x⟩ also ebenfalls wie komplexe Grapheme, wie ⟨x⟩ im Deutschen? Nein, denn alle drei Grapheme sind – anders als ⟨x⟩ im Deutschen – regelmäßig Teil von komplexen Endrändern, vgl. z. B. *lawn*, *limb*, *jinx*.

Tabelle 61 listet diejenigen Endränder auf, die nach komplexem Kern wahrscheinlicher sind. ⟨gh⟩ und ⟨ch⟩ sind die einzigen nicht-einfachen Grapheme unter den Endrändern, die nach komplexem Kern präferiert sind. Das spricht für ihren Graphemstatus – sie verhalten sich wie einfache Grapheme. Von den übrigen Endrändern sind vor allem ⟨k⟩, ⟨ɫ⟩ und ⟨f⟩ bemerkenswert, die alleine nur sehr selten nach einfachen Silbenkernen vorkommen. ⟨f⟩ und ⟨ɫ⟩ sind auch im deutschen Korpus nach einfachen Kernen seltener als nach komplexen Kernen; im englischen Korpus ist die Anzahl der Ausnahmen aber noch einmal niedriger. Es lohnt sich, die Ausnahmen näher anzusehen:

- Für ⟨k⟩ gibt es nur zwei Belege nach einfachem Kern, und zwar die Fremdwörter *trek* (aus dem Niederländischen) und *flak* (aus dem Deutschen).

- Einfaches ⟨l⟩ kommt vor in *nil* und *gel* sowie in den umgangssprachlichen Ausdrücken *gal* und *pal*, dem geologischen Fachbegriff *col* und dem Fremdwort *sol*.
- Einfaches ⟨f⟩ tritt in den Funktionswörtern *if* und *of* auf sowie in *clef* und *chef*, zwei ursprünglich französischen Wörtern.

Tab. 61: Verhältnis der Wahrscheinlichkeiten des Auftretens von Endrändern nach einfachem und nach komplexem Silbenkern; die Liste umfasst nur solche Endränder, die nach komplexem Kern mindestens doppelt so wahrscheinlich sind wie nach einfachem Kern. Datengrundlage: Liste einsilbiger einfacher englischer Stämme.

Endrand	#nach ⟨V⟩	#nach ⟨VV⟩	Verhältnis der Wahrscheinlichkeiten
k	2	42	0,02
l	6	81	0,03
f	4	22	0,08
gh	4	16	0,11
ch	8	30	0,11
r	29	71	0,18
-	54	120	0,19
d	55	65	0,36
n	68	67	0,44

⟨k⟩, ⟨l⟩ und ⟨f⟩ kommen also nur sehr selten nach einfachem Kern vor, und diese Ausnahmen sind im Wortschatz markiert; es handelt sich meist um Fremdwörter oder umgangssprachliche Ausdrücke. Wie kann diese Verteilung interpretiert werden? Wir können hier (wie oben schon) auf die Beobachtungen im Zusammenhang mit den Minimalitätsanforderungen für Stämme Bezug nehmen, die unten in 3.2.3 diskutiert werden: Englische Stämme bestehen aus mindestens drei Graphemen, und mindestens zwei davon befinden sich im Reim. Diese Bedingungen erfassen minimale Stämme wie *pit*, *hat*, *bin* etc. Einfaches ⟨k⟩, ⟨l⟩ und ⟨f⟩ ist aber offenbar zu ‚leicht‘, um alleine den Endrand nach einfachem Silbenkern zu besetzen; Formen wie **lak*, **pil* oder **stif* sind nicht wohlgeformt.⁵³ In diesem Kontext treten die drei Grapheme fast ausschließlich verdoppelt auf – bzw. als ⟨ck⟩ im Fall von ⟨k⟩. ⟨f⟩ und ⟨l⟩ sind neben ⟨s⟩ außerdem die einzigen Grapheme, die substantziell verdoppelt werden (vgl. oben 3.2.2.3); das ist die Rückseite der Medaille.

⁵³ Diese Generalisierung gilt nur für Einsilber; in Mehrsilbern kommt zumindest einfaches finales ⟨l⟩ häufig vor (vgl. *April*, *metal*, *barrel*).

Dass hier der Reim tatsächlich die entscheidende Einheit ist, um die Verteilung zu beschreiben, wird deutlich, wenn wir uns Endränder nach komplexen Silbenkernen ansehen: Hier sind die drei Grapheme unmarkiert, vgl. z. B. ⟨look⟩, ⟨beef⟩, ⟨foul⟩. Warum nun ausgerechnet ⟨k⟩, ⟨b⟩ und ⟨f⟩ nicht alleine den Endrand nach einfachem Kern besetzen können, muss unklar bleiben; eine rein formale Ursache ist unwahrscheinlich, schließlich gibt es formal sehr ähnliche Grapheme, die sich anders verhalten (⟨b⟩ wie ⟨k⟩; ⟨t⟩ wie ⟨f⟩).

Bislang ging es in diesem Abschnitt um die Interaktion von Konstituenten im Einsilber – Anfangsrand und Endrand im Deutschen, Kern und Endrand im Deutschen und Englischen. Eine weitere potenzielle Interaktion betrifft den Kern und den Endrand von Mehrsilbern. Diese Interaktion ist aber etwas anders gelagert: Wie wir oben gesehen haben, gibt es bei komplexen Silbenkernen in zweiten Silben von Zweisilbern (wie etwa *Nation*) nur wenig Variation im Endrand. Es gibt oft für einen gegebenen Silbenkern nur wenige mögliche Endränder, und zwar sowohl im englischen wie im deutschen Teilkorpus. Das scheint bei Silbenkernen in Einsilbern anders zu sein – diese These wurde zumindest oben aufgeworfen, und hier ist der Rahmen, sie zu überprüfen. Doch wie lässt sich das Ausmaß an Variation erfassen?

Dafür bietet sich ein Maß aus der Informationstheorie an, die *Entropie*. Der Begriff wurde von Shannon (1948) aus der Thermodynamik entlehnt und bezeichnet den Grad der Information, den Informationsgehalt einer Nachricht, eines Wortes, eines Zeichens (für einen guten Überblick vgl. Gleich 2011: 204 ff.). Das Konzept wird seit einiger Zeit ertragreich auf linguistische Fragestellungen angewendet, vgl. z. B. Moscoso del Prado/Kostić/Baayen (2004), Moscoso del Prado (2016); außerdem ist Shannons Aufsatz selbst sehr nah an den graphemischen Fragestellungen, die in dieser Arbeit diskutiert werden – Shannon untersucht die Informationsdichte englischer Texte.

Die Entropie für jeden Silbenkern setzt sich zusammen aus den negativen logarithmierten Wahrscheinlichkeiten jedes Endrands, der nach diesem Silbenkern auftritt. Nehmen wir als Beispiel den Silbenkern ⟨io⟩ in zweiten Silben von deutschen Zweisilbern; als Endränder kommen hier ⟨n⟩ (13-mal), leerer Endrand (dreimal), ⟨t⟩ (zweimal), ⟨m⟩ (zweimal) und ⟨s⟩ (zweimal) vor. Insgesamt gibt es 22 Zweisilber mit ⟨io⟩ in der zweiten Silbe. Die Entropie (in Bits) von ⟨n⟩ als Endrand berechnet sich dann wie folgt:

$$\begin{aligned} H_n &= -p(n) * \log(p(n)) \\ &= -\frac{13}{22} * \log\left(\frac{13}{22}\right) \\ &= 0,45 \end{aligned}$$

Die Entropie von <ɒ> ist hingegen etwas niedriger, sie liegt bei 0,31 Bits. <ɒ> ist zwar wesentlich seltener, der Informationsgehalt ist also höher – aber die Wahrscheinlichkeit von <ɒ> ist eben aufgrund des selteneren Auftretens auch niedriger. Die Entropie für alle Endränder ist nun einfach die Summe der einzelnen Entropien. Für Endränder nach <io> ergibt sich so eine Entropie von 1,78 Bits. Zum Vergleich: Die Entropie von Endrändern nach <ee> beträgt 1,01 Bits, die von Endrändern nach <ei> hingegen 3,17 Bits. Das spiegelt die ungleiche Verteilung der Endränder wider: Nach <ee> treten nur drei verschiedene Endränder auf, und einer ist dominant (der Endrand ist in drei Viertel der Fälle leer, Typ <Kaffee>). Ohne zu wissen, welcher Endrand im konkreten Fall folgt, hat man gute Chancen, richtig zu liegen, wenn man auf den leeren Endrand tippt. Der Informationsgehalt dieses Endrands ist also relativ niedrig. Nach <ei> kommen hingegen elf ähnlich häufige Endränder vor. Jeder dieser Endränder trägt daher einen höheren Informationsgehalt.

Wenn wir nun wissen wollen, wie sich die Entropie im Ein- und Zweisilber insgesamt unterscheidet, können wir die Entropien der einzelnen Kerne (wie <ei> im Zweisilber im Beispiel oben) mitteln. Der Nachteil ist, dass dann seltene Kerne (wie <ai>, das nur fünfmal vorkommt) genauso viel Einfluss auf die globale Entropie haben wie häufige Kerne (bspw. <ie>, das 61-mal vorkommt). Um das zu umgehen, werden die einzelnen Entropien gewichtet nach der Anzahl der Vorkommen des jeweiligen Kerns.

Tabelle 62 präsentiert die auf diese Weise ermittelten globalen Werte, getrennt nach Ein- und Zweisilbern und nach einfachen und komplexen Kernen. Zur Erinnerung: Hohe Werte zeigen eine hohe Informationsdichte, also viel Variation. Niedrige Werte zeigen umgekehrt eine geringe Informationsdichte, also wenig Variation.

Tab. 62: Mittlere gewichtete Entropie der Endränder für einfache und komplexe Silbenkerne im Einsilber und im Zweisilber.

	deutsch		englisch	
	Einsilber	Zweisilber	Einsilber	Zweisilber
einfacher Kern	5,69	2,85	5,39	2,61
komplexer Kern	3,51	1,81	3,19	1,52

In Einsilbern gibt es also tatsächlich mehr Variation als in Zweisilbern, was den Übergang vom Silbenkern zum Endrand des Stamms betrifft, und zwar sowohl im deutschen als auch im englischen Korpus. Die Entropie im Einsilber ist rund dop-

pelt so groß wie im Zweisilber. Das gilt für einfache wie auch für komplexe Kerne. Und es gilt auch, wenn wir bei den einfachen Kernen im Zweisilber des Deutschen «e» ausschließen: Man könnte ja vermuten, dass die frequenten Endungen «er», «en», «el» etc. für die große Differenz verantwortlich sind. Nimmt man bei den Zweisilbern einfaches «e» im zweiten Kern aus der Berechnung heraus, so ergibt sich statt der in Tabelle 61 angegebenen 2,85 ein höherer Wert von 3,90 Bits – der aber immer noch wesentlich niedriger liegt als der Vergleichswert im Einsilber (5,69 Bits).

Dieses Ergebnis bekräftigt, was vermutet wurde: In zweiten Silben von Zweisilbern kombinieren Kern und Endrand nicht so frei wie im Einsilber. Viele zweite Silben sind – zumindest was ihren Reim angeht – feste Verbindungen, die sich in dieser Hinsicht wie Suffixe verhalten.

3.2.2.8 Zusammenfassung

Dieses Kapitel hat sich mit der silbisch informierten Graphotaktik beschäftigt. Dabei ging es um drei Themen:

1. Zum einen wurde untersucht, wie Stämme in beiden Sprachen silbisch aufgebaut sind (3.2.2.1). Aus wie vielen graphematischen Silben bestehen sie, und was sind bevorzugte CV-Strukturen?
2. Zum anderen stand der Aufbau der Silbenkonstituenten im Fokus. Welche Abfolgebeschränkungen gibt es für Kerne, für Anfangs- und Endränder von Stämmen (3.2.2.2, 3.2.2.3, 3.2.2.6)? Sind Anfangs- und Endränder spiegelbildlich aufgebaut (3.2.2.4)? Wie sind intervokalische Konsonanten verteilt (3.2.2.5)?
3. Und schließlich wurden mögliche Interaktionen zwischen den Silbenkonstituenten untersucht (3.2.2.7). Hängt der Aufbau einer Konstituente vom Aufbau einer anderen ab?

Die wichtigsten Einsichten werden hier noch einmal zusammengefasst. Es gilt: Wenn vom ‚Deutschen‘ und vom ‚Englischen‘ die Rede ist, dann ist das verkürzend für das deutsche bzw. englische Teilkorpus, auf denen die Erkenntnisse beruhen.

Anzahl und Aufbau der graphematischen Silben: Etwa jeder zweite einfache Stamm im Deutschen und Englischen ist zweisilbig. Einsilbige Stämme sind etwas seltener; Stämme mit mehr als zwei Silben fallen kaum ins Gewicht. Je mehr Silben ein Stamm hat, desto mehr nähert sich das Verhältnis zwischen Konsonanten und Vokalen im Stamm 1:1 an. Hier zeigt sich jeweils eine weitgehende Übereinstimmung zwischen dem Deutschen und dem Englischen; das gilt auch für die präferierten CV-Strukturen.

Aufbau der Silbenkonstituenten: Der Anfangsrand und der Endrand graphematischer Stämme ist meist mit einem Graphem besetzt. Vor allem die Besetzung des Endrandes variiert aber in Abhängigkeit von der Silbenzahl: Je mehr Silben der Stamm hat, desto weniger Grapheme sind im Endrand. Für die mehrfach besetzten Anfangs- und Endränder können graphotaktische Schemata aufgestellt werden, die mögliche Abfolgen erfassen. Grundlegend ist hier die Einsicht, dass zwei Grapheme innerhalb einer Konstituente nur auf eine Art serialisiert werden können. Nur in Ausnahmefällen kommen beide Abfolgen vor, etwa <ng> und <gn> im englischen Endrand. Im Anfangs- und Endrand finden wir größtenteils spiegelbildliche Abfolgen. Abweichungen von dieser Symmetrie können als Kriterium für den Graphemstatus genutzt werden: <st> z. B. verhält sich anders als andere Abfolgen, indem es im Anfangs- und Endrand nicht symmetrisch auftritt, sondern seine Abfolge beibehält (*Stahl* – *Last* – **Lats*). Das spricht dafür, dass sich <st> wie eine einfache Einheit verhält, also wie ein Graphem.

Auch der Silbenkern ist ganz überwiegend einfach besetzt. In zweifach besetzten Kernen kombinieren die Vokalgrapheme nicht willkürlich, sondern es lassen sich bevorzugte erste und zweite Bestandteile ausmachen. Die Gemeinsamkeit zwischen beiden Sprachen ist hier, dass <i> und <u> jeweils bevorzugte zweite Bestandteile sind (diese Erkenntnis lässt sich auf weitere Sprachen übertragen, vgl. Fuhrhop/Berg eing.).

Interaktion der Konstituenten: Die Besetzung der Konstituenten hängt teilweise von der Besetzung anderer Konstituenten ab. Im Deutschen findet sich eine (global schwache) Interaktion zwischen Anfangs- und Endrand: Wenn der Anfangsrand unbesetzt ist, ist ein dreifach besetzter Endrand häufiger als erwartet; wenn der Anfangsrand komplex ist, ist ein solcher Endrand seltener als erwartet. Untersucht wurde auch die Abhängigkeit von postvokalischem <h> von der Besetzung des Anfangsrandes im Deutschen; die Ergebnisse fielen dabei eindeutiger aus als in der früheren Literatur. Postvokalisches <h> tritt in knapp drei Viertel der einschlägigen Stämme mit unbesetztem oder einfach besetztem Anfangsrand vor <l>, <m>, <n> oder <r> auf. Postvokalisches <h> beschwert Stämme mit diesen (ansonsten einfachen) Endrändern. Die Domäne ist allerdings nicht der Stamm, sondern die Wurzel (also der Pseudostamm): Neben *lahm* tritt postvokalisches <h> eben auch in Zweisilbern wie *Sahne* oder *Gefahr* auf, die als Kombinationen von Wurzel und Endung interpretiert werden können.

Substanzielle Interaktionen finden sich in beiden Sprachen zwischen Kern und Endrand; das ist phonographisch erwartbar. Komplexe graphematische Silbenkerne kombinieren wesentlich seltener als erwartet mit komplexen Endrändern; einfache Silbenkerne kombinieren seltener als erwartet mit unbesetzten oder einfach besetzten Endrändern. Diese Interaktion ist auch im Detail interes-

sant. So gibt es in beiden Sprachen Grapheme, die nur oder vor allem nach einfachem Kern auftreten, und solche, die vor allem nach komplexem Kern auftreten. Im Deutschen kommt ⟨x⟩ nur nach einfachem Kern vor, gleichzeitig kombiniert es im Endrand nicht; ⟨x⟩ verhält sich damit wie ein komplexes Graphem. Umgekehrt kommt ⟨ch⟩ wesentlich häufiger nach komplexem Kern vor und verhält sich damit wie ein einfaches Graphem. Im Englischen gilt dasselbe für ⟨ch⟩ und ⟨gh⟩. Hier ist außerdem interessant, dass ⟨b⟩, ⟨f⟩ und ⟨k⟩ zu ‚leicht‘ sind, um alleine den Endrand nach einfachem Kern im Einsilber zu besetzen.

Eine Interaktion etwas anderer Art ist abhängig von der Silbenzahl: In zweiten Silben von Zweisilbern hängen Kern und Endrand enger zusammen als im Einsilber. Bei gegebenem Kern gibt es dort weniger Variation; die Entropie ist niedriger, der Endrand ist vorhersagbarer.

3.2.3 Minimale Stämme und prototypische Stämme

In diesem Abschnitt geht es einerseits um graphemische Mindestbedingungen, die Stämme im Deutschen und Englischen erfüllen müssen. Wie müssen Stämme mindestens beschaffen sein? Wie lang müssen sie sein? Andererseits wird untersucht, welche Strukturen unter Stämmen prototypisch sind. So sind ⟨VCC⟩-Wörter wie *Ast* oder *ant* im Deutschen und Englischen wohlgeformt, aber relativ selten. Eine Aufstellung häufiger Strukturen (die sich teilweise mit der Untersuchung in 3.2.2.1 oben überlappt) kann dann mit den häufigen Strukturen von Affixen verglichen werden (das geschieht in 3.4).

Wie sehen nun minimale Stämme im Deutschen und Englischen aus? Es scheint zunächst in beiden Sprachen eine Mindestlänge von drei Buchstaben für Autosemantika zu geben (vgl. für das Englische Jespersen 1928: 149). Es gibt nur wenige Inhaltswörter mit zwei Buchstaben im Deutschen (38a, Liste aus Ramers 1998) – im Englischen sogar nur eins, wenn wir Abkürzungen wie *ad* und *bi* sowie Fremdwörter wie *pi* ausschließen (38b):

(38a) Ar, Bö, Ei, öd, Öl, Ur

(38b) ox

Für Nicht-Inhaltswörter gilt das in beiden Sprachen nicht, wie die folgenden, nicht exhaustiven Listen zeigen:

(39a) ab, an, da, du, er, es, in, ja, ob, oh, um, wo, zu

(39b) I, a, o; as, be, he, in, is, it, my, no, on, to, us, we

Diese Funktionswörter, Interjektionen und Partikeln können aus zwei Buchstaben bestehen, im Englischen sogar aus nur einem Buchstaben (das Pronomen *I*, der Indefinitartikel *a*, die Interjektion *o*). Diese Wörter sind nicht „highly exceptional“, wie Evertz (2014: 164) meint – im Gegenteil, von den fünf (mit ⟨y⟩ sechs) Vokalbuchstaben des Englischen kommen drei als Wörter vor. Die Mindestlänge für Nicht-Inhaltswörter ist im Englischen also ein Buchstabe, im Deutschen zwei.

Für Inhaltswörter reicht eine Beschränkung auf drei Buchstaben in beiden Sprachen nicht aus – bestimmte Strukturen kommen innerhalb dieser Wörter nicht vor. Beginnen wir mit dem Deutschen. Hier zeigen Ramers (1998) und Evertz (2014) überzeugend, dass eine mögliche Beschränkung sich nicht allein auf die Länge in Buchstaben beziehen kann:

(40a) ⟨CCV⟩: Klo, Pli, pro, Psi, Ski, zwo

(40b) ⟨CVV⟩: Bau, Fee, Gau, kau[en], lau, Sau, See, Tau, Tee

(40c) ⟨VCC⟩: Abt, Akt, alt, Alm, Amt, arg, Ast, Axt, eng, Ohr, oft, Ort, Ost, Uhr

Die Liste in (40a) stammt aus Evertz (2014: 151) und ist im CELEX-Korpus exhaustiv; die Listen in (40b, c) enthalten exemplarische Wörter des Typs. In (40a) gibt es laut Evertz nur ein natives Wort, *zwo* – und das ist eine umgangssprachliche Variante zu *zwei*. Der Typ ⟨CCV⟩ ist also gegenüber den anderen beiden Typen deutlich markiert – Ramers (1998: 33) leitet aus dieser Beobachtung das Verbot einfacher wortfinaler Vokalgrapheme in Inhaltswörtern des Deutschen ab. Ramers (1998) reformuliert diese Beschränkung im Rahmen einer Theorie des Silbengewichts: Deutsche Inhaltswörter bestehen aus mindestens drei Gewichtseinheiten (Moren), wobei der Anfangsrand von Silben „generell den Morenwert 1“ hat (Ramers 1998: 36), jedes Element im Kern und im Endrand aber als eigene Mora zählt. Das ist gleichbedeutend damit, dass der Reim der Silbe aus mindestens zwei Moren (und damit mindestens zwei Buchstaben) besteht – eine Formulierung, die Evertz (2014: 159) bevorzugt. Wir haben also im Deutschen zwei notwendige Bedingungen für minimale Wörter, a) die Mindestlänge von drei Buchstaben und b) einen zweifach besetzten Reim.

Ramers und Evertz untersuchen minimale Wörter – lassen sich die Ergebnisse auch auf Stämme übertragen? Der Unterschied zwischen Stämmen und Wörtern betrifft im Deutschen nur die Verben (siehe oben 2.2). Und hier gibt es tatsächlich nur drei Stämme, die aus weniger als drei Buchstaben bestehen oder deren Reim nur einfach besetzt ist, *sä[en]*, *tu[n]* und *üb[en]*. Die beiden notwendigen Bedingungen für minimale Wörter im Deutschen gelten also auch für minimale Stämme – eine Mindestlänge von drei Buchstaben und ein zweifach besetzter Reim. Und wenn wir die minimale Graphemdefinition (5) aus 3.1 anwenden, dann können wir die Regel ohne Verluste auch auf Grapheme beziehen. Es gibt

nur einen Stamm von Inhaltswörtern, der dann die Mindestlänge von drei Graphemen unterschreitet, *Eck*.

Die Untersuchung minimaler Wortstrukturen wird besonders interessant, wenn wir den Blick auf mögliche Schreibungen ausweiten. Wie viele Stämme könnten phonographisch mit zwei Graphemen verschriftet werden, werden aber anders geschrieben? Das gibt uns einen Hinweis darauf, ob (und wie) das Schriftsystem hier ‚ausweicht‘, um bestimmte Strukturen zu vermeiden. Dazu betrachten wir alle Stämme, die phonologisch aus zwei Segmenten bestehen und die theoretisch mit zwei Graphemen verschriftet werden könnten – wie z. B. *Aal* (/a:l/), das auch ⟨Al⟩ geschrieben werden könnte. Die Phonem-Graphem-Korrespondenzen des Deutschen geben diese Schreibung her.

In (41) sind alle zehn Stämme im Teilkorpus deutscher Stämme versammelt, die phonologisch die Struktur /VC/ haben, die aber graphematisch aus drei statt zwei Graphemen bestehen (vgl. Evertz 2014: 149, der zu sehr ähnlichen Zahlen kommt). Diese Stämme erfüllen die beiden notwendigen Minimalitätsbedingungen. Verletzt werden eine oder beide Bedingungen durch die fünf bereits oben genannten Stämme *Ar*, *öd*, *Öl*, *Ur* und *üb[en]*. Die Mehrzahl der phonographisch potenziell zweigraphemigen Wörter wird also mit drei Graphemen verschriftet.

(41) Aal, Aar, Aas, Ahm, Ahn, Ass, Ohm, Ohr, Öhr, Uhr

Dasselbe können wir für phonologische Sequenzen tun, die potenziell mit einfachem finalem Vokal verschriftet werden könnten – /ku:/ z. B. könnte phonographisch auch ⟨Ku⟩ geschrieben werden. Doch auch hier genügt die Mehrzahl der Schreibungen (die 40 Stämme in Beleg 42) den beiden notwendigen Bedingungen, und nur elf Stämme verletzen eine oder beide (*Klo*, *Nu*, *Pi*, *Pli*, *pro*, *Psi*, *Re*, *Ski*, *sä[en]*, *tu[n]*, *zwo*).⁵⁴

(42) bläh[en], blüh[en], brüh[en], buh[en], dreh[en], droh[en], Fee, fleh[en], flieh[en], Floh, froh, früh, geh[en], glüh[en], jäh, Klee, Knie, kräh[en], Kuh, Lee, mäh[en], muh[en], müh[en], nah, näh[en], Reh, roh, schmäh[en], See, seh[en], spä[h]en, steh[en], Stroh, Tee, weh, weh[en], zäh, Zeh, zieh[en], Zoo

⁵⁴ Es gelten hier folgende – z. T. bereits oben gemachte – Anmerkungen zum Status der Stämme: *Nu* kommt nur in Verbindung mit *im* vor; *Pi*, *Pli*, *pro*, *Psi* und *Ski* sind Fremdwörter; *Klo* und *Re* sind Abkürzungen; *zwo* ist eine umgangssprachliche Variante zu *zwei*. Effektiv bleiben hier also zwei Ausnahmen, *sä[en]* und *tu[n]*.

Dieser Vergleich der phonographischen Möglichkeiten zeigt: Es könnten potenziell wesentlich mehr Wörter die zwei notwendigen Bedingungen verletzen. Das geschieht aber nicht. Andersherum: Die Tatsache, dass fast alle Inhaltswörter im Deutschen mindestens drei Buchstaben und einen zweifach besetzten Reim aufweisen, hat keine phonologische Ursache, es ist eine rein graphematische Beschränkung, wie Evertz (2014: 150) feststellt.

Interessant ist in diesem Zusammenhang, dass mit den Bedingungen zur minimalen Struktur automatisch das sog. silbeninitiale ⟨h⟩ bei Verben erfasst wird. Worum geht es beim silbeninitialen ⟨h⟩? In Fällen wie *geh[en]* oder *brüh[en]* wird üblicherweise angenommen, das intervokalische ⟨h⟩ diene der besseren Gliederung des graphematischen Wortes (so z. B. Eisenberg 2013a: 301 f.): Wenn der Stamm ohne ⟨h⟩ geschrieben würde, wäre im Infinitiv die Silbenstruktur nur schwer zu erkennen (*⟨geen, brüen⟩). Das ⟨h⟩ signalisiert die Silbengrenze (und die Morphemgrenze!), und Wörter wie *gehen* und *brühen* sind von ihrem graphematischen Aufbau analog zu Wörtern wie *geben* und *brüten*. In allen Fällen sind sowohl die Zweisilbigkeit als auch die Silben- und Morphemgrenze durch den langköpfigen intervokalischen Buchstaben deutlich zu erkennen.

Genau diese Fälle erfassen die Minimalitätsbedingungen automatisch: ⟨ge⟩ und ⟨dro⟩ sind keine wohlgeformten Stämme des Deutschen, weil der Reim nicht zweifach besetzt ist. Es gibt allerdings sieben Verbstämme, deren ⟨h⟩ so nicht motiviert werden kann, *flieh[en]*, *gedeih[en]*, *leih[en]*, *reih[en]*, *weih[en]*, *zeih[en]* und *zieh[en]*. Hier ist die Funktion allein die Markierung der silbischen und morphologischen Struktur. Es fällt allerdings auf, dass ⟨h⟩ hier nur nach ⟨ei⟩ und ⟨ie⟩ auftritt. Die fünf Belege mit ⟨eih⟩ oben sind Ausnahmen; es gibt im Korpus zehn Verbstämme, die auf ⟨ei⟩ und eben nicht auf ⟨eih⟩ enden (z. B. *schnei[en]*, *schrei[en]*, *spei[en]*, *frei[en]*). Für ⟨ie⟩ gilt das nicht; hier kann vermutet werden, dass durch die Schreibung mit ⟨h⟩ Vokalcluster der Form ⟨iee⟩ in Flexionsformen vermieden werden (z. B. ⟨fliehe⟩, *⟨fliee⟩). Interessanterweise kommen bzw. kamen diese Cluster in Schreibvarianten aber durchaus vor, z. B. in ⟨geschrieen⟩ (als Variante zu ⟨geschrien⟩, vgl. „Du hast gut geschrieen, Danton“, Büchner 1879: 75) oder in ⟨Kniee⟩ (als Varianten des Plurals von ⟨Knie⟩, vgl. „die Begier, das Glück in die Kniee zu zwingen“, Mann 1901: 280; weiter zur formalen Variation 4.3.1).⁵⁵ Hier scheinen zwei konkurrierende Beschränkungen am Werk zu sein: Einerseits sollen ⟨iee⟩- und ⟨eie⟩-Cluster vermieden werden; andererseits ist das ⟨h⟩ nach komplexen Kernen als Schweremarkierer überflüssig.

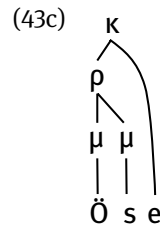
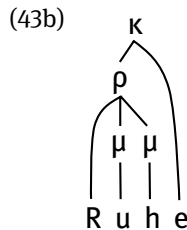
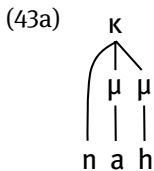
Was ist mit den nicht-verbalen ⟨h⟩-Schreibungen wie *Ruhe* oder *nahe*? Sie lassen sich auf den ersten Blick ebenfalls nur schwer mit den Minimalitätsbedingun-

⁵⁵ Für diesen Hinweis danke ich Nanna Fuhrhop.

gen fassen – schließlich umfassen sie auch ohne ⟨h⟩ drei Grapheme (vgl. **Rue*, **nae*).

Eine Möglichkeit wäre, ⟨Ruh⟩ als Wurzel von *Ruhe* anzusetzen. Das ⟨e⟩ wäre dann eine Endung. Diese Segmentierung in Wurzel und Endung hat sich im Verlauf der Arbeit bereits mehrfach ergeben – besonders relevant ist sie im Zusammenhang mit dem postvokalischen ⟨h⟩ im Deutschen (3.2.2.7). Hier wurde argumentiert, dass dessen Domäne nicht der Stamm, sondern die Wurzel sei; auf diese Art konnten auch die ⟨h⟩-Schreibungen in *Sahne* und *Gefahr* erfasst werden.

Wenn wir auch beim silbeninitialen ⟨h⟩ so vorgehen, dann ergibt sich hinsichtlich der Minimalitätsbedingungen, dass **Ru* keine gute Wurzel ist, weil es beide Bedingungen verletzt. Damit ist das ⟨h⟩ in *Ruhe* motiviert. Das hat allerdings zur Konsequenz, dass in Stämmen wie *Ode*, *öde* und *Öse* die Wurzeln *Od*, *öd* und *Ös* ebenfalls eine Bedingung verletzen, nämlich die Mindestlänge von drei Graphemen (die zweite Bedingung wird nicht verletzt – der Reim enthält zwei Gewichtseinheiten). Die Lösung ist einfach: Wir beziehen die Mindestlänge auf den gesamten Stamm, die Minimalitätsbeschränkung des Reims aber (wenn anwendbar) auf die Wurzel. Eine schematische Darstellung der Verhältnisse in Anlehnung an Hayes (1989) hilft, das zu verdeutlichen (Stämme werden mit gr. κ für *kormós* ‚Stamm‘ bezeichnet, Wurzeln mit gr. ρ für *ríza* ‚Wurzel‘):



In (43a) würde das phonographisch mögliche *⟨na⟩ beide Minimalitätsbedingungen verletzen: Der Stamm wäre zu kurz und der Reim zu leicht. ⟨h⟩ beschwert hier den Reim durch eine zusätzliche Gewichtseinheit. In (43b) hätte die phonographisch mögliche Wurzel *⟨Ru⟩ einen zu leichten Reim. Da sich die Mindestlänge von drei Graphemen auf den gesamten Stamm bezieht, besteht hier kein Problem. Das sieht man an (43c): Der Reim der Wurzel ⟨Ös⟩ ist schwer genug, und der gesamte Stamm ist lang genug; beide Bedingungen werden erfüllt.

Das ist natürlich für sich genommen nicht viel mehr als ein Taschenspielertrick: Wurzeln in Zweisilbern können ja gar nicht anders, als aus der Minimalstruktur ⟨VC⟩ zu bestehen; Wurzeln der Form ⟨V⟩ oder ⟨CV⟩, deren Reim nicht zweifach besetzt ist, sind unmöglich – zumindest, wenn man, wie in dieser Arbeit

vorgeschlagen, graphematisch syllabiert. Das lässt sich an folgendem Beispiel zeigen. Der minimale graphematische Zweisilber ist ⟨VCV⟩ – die Struktur ⟨VV⟩ würde im Rahmen dieser Arbeit als graphematischer Einsilber gewertet, weil die Vokalgrapheme nicht durch Konsonanten(cluster) getrennt werden. In ⟨VCV⟩-Stämmen ist ⟨VC⟩ die Wurzel und ⟨V⟩ die Endung. Anders ausgedrückt: Es gibt keine Wurzel, die auf Vokal endet. Wenn aber alle Wurzeln auf Konsonanten enden, dann ist die Reim-Beschränkung immer erfüllt; der Reim ist immer mit zwei Gewichtseinheiten besetzt.

Wenn wir die These vom zweifach besetzten Reim bei Wurzeln wirklich überprüfen wollen, müssen wir CV-Strukturen theoretisch erlauben. Das kann auf mindestens zwei Arten geschehen.

Zum einen können wir ⟨V⟩-finale Wurzeln wie *Ru in *Rue erlauben. Wir müssen dafür prinzipiell die Möglichkeit zulassen, dass finales ⟨e⟩ nach Vokal eine eigene graphematische Silbe konstituiert und damit als Endung klassifiziert werden kann. Das ist graphemisch durchaus zu rechtfertigen. Aus der Verteilung der Schreibung komplexer Silbenkerne in 3.2.2.6 ergibt sich, dass die Kombinationen ⟨ae⟩, ⟨oe⟩ und ⟨ue⟩ im deutschen Silbenkern marginal sind; wenn sie auftreten, handelt es sich umgekehrt wahrscheinlich um einen graphemischen Hiatus.

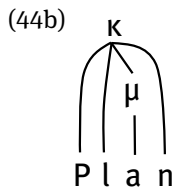
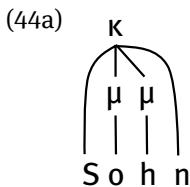
Wenn wir auf dieser Grundlage nach potenziell zu leichten Wurzeln suchen (also nach Stämmen, die auf ⟨CV+e⟩, ⟨CV+el⟩, ⟨CV+em⟩, ⟨CV+en⟩ oder ⟨CV+er⟩ enden), finden wir nur fünf drei- und viersilbige Stämme (*Aloe*, *Oboe*, *ambigue*, *Morgue*, *Statue*), keine zweisilbigen vom Typ *⟨Rue⟩. In all diesen Fällen ist die Wurzel deutlich komplexer als ⟨CV⟩. Die Tatsache, dass es keine Stämme wie *⟨nae⟩, *⟨Loe⟩ oder *⟨Rue⟩ gibt, dafür aber eine Reihe von Stämmen wie ⟨nahe⟩, ⟨Lohe⟩ und ⟨Ruhe⟩, ist ein Argument dafür, dass auch der Reim von Wurzeln mindestens zweifach besetzt sein muss.

Zum anderen können sich Wurzeln der Form ⟨V⟩ oder ⟨CV⟩ theoretisch auch stammfinal ergeben, wenn der Stamm mit einem Pseudopräfix beginnt, z. B. *gesche[en] (siehe oben 3.2.2.5). Das ist aber nicht der Fall: Es gibt keine Stämme, die mit einem der Präfixe in der Liste deutscher Affixe beginnen und auf einfachen Vokal enden. Zumindest nicht, wenn wir dem Präfix eine halbwegs konstante semantische Funktion zuschreiben wollen – ansonsten gibt es knapp zwei Dutzend Stämme vom Typ *Unke*, *Geste*, *Erbe*, *Urne*. Hier sind *un-*, *ge-*, *er-* und *ur-* allerdings keine Instanzen der Präfixe, sondern homographischer Stammanfänge.

Beide Argumente sprechen dafür, dass auch Wurzeln über zweifach besetzte Reime verfügen müssen. Ansonsten wäre mit Strukturen wie *⟨Rue⟩ oder *⟨gesche[en]⟩ zu rechnen.

Wir können nun einen Schritt weiter gehen und auch Stämme mit Dehnungs-⟨h⟩ mit den Minimalitätsbeschränkungen beschreiben. Das soll im Fol-

genden gezeigt werden. Wie oben in 3.2.2.7 gezeigt wurde, hängt postvokalisches ⟨h⟩ im Deutschen von der Besetzung des Anfangsrand ab. Es tritt nach einfachem Vokalgraphem und vor einfachem ⟨l⟩, ⟨m⟩, ⟨n⟩ oder ⟨r⟩ im Endrand des Stamms bzw. der Wurzel auf, wenn der Anfangsrand nicht oder nur einfach besetzt ist (z. B. *Rahm*, *Ahr*, *Sahne*). Das lässt sich wie folgt in die Beschreibung oben einbinden: ⟨l⟩, ⟨m⟩, ⟨n⟩ und ⟨r⟩ tragen im Endrand nicht ausreichend zum Gewicht des Reims bei; sie sind zu leicht. Das führt zu einer Verletzung des minimalen Gewichts des Reims. Diese Verletzung kann nur kompensiert werden, wenn der Stamm insgesamt umfangreicher ist und aus mindestens vier (statt drei) Graphemen besteht (und damit der Anfangsrand mit mindestens zwei Graphemen besetzt ist). Das lässt sich am besten an einigen Beispielen explizieren:



In (44a) trägt das einfache ⟨n⟩ im Endrand nicht genug zum Silbengewicht bei. Die phonographisch mögliche Form *⟨Son⟩ ist zu leicht, der Reim besteht nicht aus zwei vollständigen Gewichtseinheiten. Die Schreibung ⟨Sohn⟩ umgeht das, indem mit ⟨h⟩ eine zusätzliche Gewichtseinheit auftritt. In (44b) ist das nicht notwendig. Hier ist zwar ebenfalls das einfache ⟨n⟩ im Endrand nicht mit einer Gewichtseinheit assoziiert; das Wort erfüllt aber die ‚verschärfte‘ minimale Längbedingung, es hat insgesamt mehr Substanz. Das ⟨h⟩ hat also im Deutschen, so können wir zusammenfassen, eine Funktion – es reguliert das Silbengewicht im Reim und sichert so die minimale Struktur, die für geschriebene Stämme im Deutschen notwendig ist. Mit dieser Beschreibung kann sowohl das Dehnungs-⟨h⟩ als auch das silbeninitiale ⟨h⟩ erfasst werden. Das silbeninitiale ⟨h⟩ fungiert darüber hinaus als Silbentrenner; es macht die morphologische Struktur in Wörtern wie *sehen* sichtbar.

Im Englischen greifen ebenfalls beide Bedingungen, wie Evertz (2014: 161 ff.) zeigt – es gibt allerdings etwas mehr Ausnahmen. Für Wörter mit der Struktur ⟨VC⟩ gilt das erstmal nicht – hier sind im englischen Teilkorpus 14 Einträge zu finden, und die Synsemantika überwiegen eindeutig:

- (45) an, as, at, ex, id, if, in, it, of, on, or, ox, up, us

Diese Liste zeigt die Abhängigkeit der Minimalitätsbedingungen vom lexikalischen Status der Wörter eindrucksvoll: Nur ein Eintrag in der Liste (45) ist ein Autosemantikon (*ox*); *ex* ist ein Kurzwort, und der Rest sind Funktionswörter. Nicht im Teilkorpus aufgeführt ist *ax*, eine Schreibvariante zu *axe*.

Wörter der Struktur <CV> sind etwas häufiger, allerdings sind auch hier Auto-semantika im engeren Sinne extrem selten:

- (46) be, by, do, fa, go, he, la, ma, me, mi, mo, mu, my, no, pa, pi, po, re, si, so, ti, to, we

Eine größere Gruppe sind die Silben der sog. Solmisation, die an die Tonhöhe musikalischer Töne gebunden werden (*do, re, mi, fa, so, la, ti, do* und *si*). Außerdem finden sich Fremdwörter (*mu, pi*) und Kurzwörter in der Liste in (46) (*ma, pa, mo* [Br. E. für *moment*], *po* [für *chamber pot*]). Die verbleibenden Wörter sind alleamt Funktionswörter (*be, by, go, he, me, my, no, to, we*).

Bei den Strukturen <CCV> und <CCCV> ändert sich das Bild etwas: Hier sind Ausnahmen tatsächlich häufiger als im Deutschen – es gibt eine Reihe Inhaltswörter vom Typ *dry*.

- (47) chi, cry, dry, fly, fro, fry, gnu, phi, ply, pro, pry, she, shy, ski, sky, sly, spa, spy, sty, the, thy, thru, try, two, who, why, wry

Unter den Belegen in (47) sind Fremdwörter (*chi, gnu, phi, ski, spa*) und Synsemantika (*she, the, thru, who*); außerdem *fro* (entweder Teil der Phrase *to and fro* oder Kurzform von *Afro* [*hairstyle*]), *pro* und *two* – die letzteren beiden keine eindeutigen Inhaltswörter. Die größte Subgruppe aber ist die der Wörter auf <y> (z. B. *cry, dry, shy*). Sie umfasst 16 Einträge; sie stellen die eigentlichen Ausnahmen der notwendigen Bedingung vom zweifach besetzten Reim dar. Gleichzeitig fungiert <y> als einziges Vokalgraphem auch als Marker für Wortenden. Hier besteht ein Zusammenhang: Es macht Sinn, <y> in dieser Position zwei Gewichtseinheiten (und nicht nur eine) zuzuschreiben. Das ist weniger ad hoc, als es auf den ersten Blick aussieht: Schließlich alterniert <y> regelmäßig mit <ie> (z. B. <lady> – <ladies>, <cry> – <cries>, vgl. auch Berg et al. 2014).

Drehen wir auch hier die Richtung um und betrachten diejenigen Stämme, die aus der Abfolge Vokal + Konsonant bestehen und potenziell mit zwei Graphemen geschrieben werden könnten. Die Liste in (48) enthält diejenigen Stämme, die den Minimalitätsbedingungen genügen und aus drei Buchstaben bestehen.

- (48) add, ass, ebb, eel, eff, egg, ill, inn, odd, off, ore

Mit der Ausnahme von *off* enthält (48) nur Inhaltswörter, bei denen die Schreibung von der phonographisch ökonomischsten Variante (*⟨ad⟩, *⟨as⟩, *⟨eb⟩ etc.) abweicht, um eine Mindestlänge von drei Graphemen zu erreichen.

Noch umfangreicher ist die Liste der Wörter, die potenziell mit einfachem finalem Vokalbuchstaben geschrieben werden könnte – etwa *clue* als *⟨clu⟩. Hier gibt es 78 Stämme, die mit mindestens drei Buchstaben geschrieben werden, obwohl das phonographisch nicht notwendig wäre:

- (49) bee, blah, blow, blue, bow, buy, clue, crow, cue, die, doe, dough, due, dye, fee, flea, flee, floe, flow, flue, foe, free, ghee, glee, glow, glue, grow, hie, high, hue, key, knee, know, lea, lee, lie, low, lye, mow, nigh, pea, pee, pie, plea, queue, roe, row, rue, rye, scree, sea, see, show, sigh, sloe, slow, snow, sow, spree, stow, sue, tea, tee, thigh, though, three, throw, tie, toe, tow, tree, true, twee, vie, wee, woe, zee

Diese Stämme sind unterschiedlich frequent und unterschiedlich zentral im Wortschatz – vgl. etwa *blue* und *queue*. Insgesamt ist das Bild aber klar: Inhaltswörter, die den Minimalitätsbedingungen nicht genügen, sind die Ausnahme. Besonders interessant für die Diskussion von finalem ⟨y⟩ oben sind die Schreibungen *dye*, *lye* und *rye*: Hier zeigt sich, dass auch Stämme mit finalem ⟨y⟩ die Mindestlänge von drei Graphemen erfüllen, obwohl dazu phonographisch kein Anlass besteht. Wir haben also auch im Englischen eine Situation wie im Deutschen: Die Mindestlänge bezieht sich auf Grapheme, die Minimalitätsbeschränkung des Reims hingegen auf abstraktere Gewichtseinheiten.

Also: Im Englischen und im Deutschen gibt es dieselben zwei Minimalitätsbedingungen für lexikalische Stämme (vgl. Evertz 2014). Diese Stämme müssen aus mindestens drei Graphemen bestehen, und ihr Reim muss mindestens zwei Gewichtseinheiten enthalten. Aus diesen beiden Bedingungen ergeben sich als mögliche minimale Strukturen ⟨VVC⟩ (*Eis*, *oil*), ⟨VCC⟩ (*alt*, *urn*), ⟨CVC⟩ (*rot*, *red*), ⟨CVV⟩ (*bau[en]*, *cue*) und ⟨VCV⟩ (*öde*, *ape*). Die größere Zahl an Ausnahmen im Englischen lässt sich auf ⟨CCV⟩-Strukturen mit finalem ⟨y⟩ (z. B. in *cry*) zurückführen. ⟨y⟩ markiert hier aber auch die Wortgrenze und alterniert in flektierten Wortformen mit ⟨ie⟩ (*cries*). Es macht also Sinn, ⟨y⟩ als zweimorig anzusetzen.

Wir können auf diese Weise auch finale ⟨b⟩, ⟨f⟩ und ⟨s⟩ im Englischen erfassen: Wie oben gezeigt wurde (3.2.2.7), treten diese Grapheme kaum allein im Endrand nach einfachem Vokal auf. Das ist eine Situation, die mit ⟨b⟩, ⟨m⟩, ⟨n⟩ und ⟨r⟩ im Deutschen vergleichbar ist (wir sehen im Englischen nur keine Interaktion mit dem Anfangsrand). Daher können wir eine analoge Beschränkung im Englischen

annehmen: Einfaches ⟨l⟩, ⟨f⟩ und ⟨s⟩ im Endrand trägt nicht zum Gewicht des Stamms bei. Damit wird erfasst, dass Schreibungen wie *⟨staf⟩, *⟨glas⟩ und *⟨pil⟩ nicht vorkommen: Sie erfüllen das Mindestgewicht des Reims nicht, da nur der Vokal zum Gewicht beiträgt. Die tatsächlichen Formen verdoppeln den finalen Konsonanten (*staff*, *glass*, *pill*) – die Beschränkung gilt also nur für einfaches ⟨l⟩, ⟨f⟩ und ⟨s⟩ im Endrand.

Gilt die Minimalitätsbedingung im Englischen auch für Wurzeln? Diese Frage ist etwas schwerer zu beantworten als im Deutschen, wo die ⟨h⟩-Schreibungen einige Hinweise lieferten. Wir können nur auf den Test der präfigierten Stämme zurückgreifen. Und hier gibt es – wie im Deutschen – keine Stämme, die mit einem der Präfixe aus der Liste englischer Präfixe beginnen und mit einfachem Vokalgraphem enden. Es gibt allenfalls eine Reihe von Stämmen, die mit Präfixen homographisch beginnen, dieses Präfix aber nicht enthalten, z. B. *disco*, *extra*, *outré*, *uncle*. Auch im Englischen scheint die Beschränkung des zweifach besetzten Reims also für Wurzeln zu gelten.

Wir können die generalisierten Minimalitätsbedingungen für Stämme und Wurzeln im Deutschen und Englischen abschließend wie folgt formalisieren:

- (50) Minimalitätsbeschränkungen für Stämme und Wurzeln:
- (50a) Stämme bestehen aus mindestens drei Graphemen.
- (50b) Der Reim von Stämmen und Wurzeln ist mit mindestens zwei Graphemen besetzt.
- (50c) Jedes Graphem im Reim ist wiederum mit einer Gewichtseinheit verknüpft.
- (50d) Es gelten folgende sprachspezifische Besonderheiten:
- Im Deutschen sind ⟨l⟩, ⟨m⟩, ⟨n⟩ und ⟨r⟩ im Endrand leichter; ein daraus resultierendes zu geringes Gewicht kann durch einen komplexen Anfangsrand kompensiert werden.
- Im Englischen sind einfaches ⟨l⟩, ⟨f⟩ und ⟨s⟩ im Endrand leichter; ⟨y⟩ im Kern verhält sich wie zwei Grapheme.

Diese Beschränkungen finden sich sehr ähnlich bei Evertz (2014: 160 f.), der wiederum auf Ramers (1998: 37) aufbaut. Die entscheidende Neuerung in dieser Arbeit ist, die Beschränkung nicht auf minimale Wörter, sondern auf minimale Stämme und minimale Wurzeln zu beziehen. So können neben Wörtern wie *froh* auch Stämme wie *seh[en]* und Wurzeln wie *Nehr[ung]* erfasst werden.

Nachdem wir nun Minimalitätsbeschränkungen für beide Sprachen aufgestellt haben, sehen wir uns an, welche Strukturen häufig und welche selten sind. Daraus sollen Strukturbeschränkungen für prototypische Stämme abgeleitet wer-

den.⁵⁶ Da es hier (auch) um die Abgrenzung zu Affixen geht – also die Frage, ob wir Stämme und Affixe allein aufgrund ihrer graphematischen Gestalt und ihres Aufbaus auseinanderhalten können –, beschränken wir uns im Folgenden auf die Strukturen von Einsilbern. Das hat den zusätzlichen Vorteil, dass die Variation im Gegensatz zu Mehrsilbern überschaubar bleibt.

Variieren können wir im Einsilber die Besetzung der drei Konstituenten Anfangsrand, Endrand und Kern.⁵⁷ Wenn wir uns zunächst auf einfach besetzte Kerne beschränken, können wir die Variation im Anfangs- und Endrand zweidimensional darstellen. In Tabelle 63 ist die Anzahl der entsprechenden Strukturen im deutschen Korpus dargestellt. Die Zahl 41 in Zeile ‚-‘, Spalte ‚CC‘ besagt, dass es 41 Stämme mit leerem Anfangsrand und einfach besetztem Endrand (sowie einfachem Kern) im deutschen Korpus gibt, also den Typ ⟨VCC⟩, z. B. ⟨Ast⟩. Die Zellen sind abhängig von der Anzahl der belegten Stämme schattiert: Je mehr Stämme es für eine Struktur gibt, desto dunkler erscheint die Zelle. Die Strukturen ⟨V⟩, ⟨VC⟩ und ⟨CV⟩ genügen nicht der Mindestbeschränkung für Stämme im Deutschen und sind deswegen schraffiert.

Tab. 63: Anzahl der Strukturen von Einsilbern mit einfachem Kern im Korpus einfacher deutscher Stämme, sortiert nach der Besetzung des Anfangsrandes (Zeilen) und des Endrandes (Spalten).

	-	C	CC	CCC+
-			41	10
C		241	577	71
CC	7	172	256	19
CCC+		47	84	3

⁵⁶ Prototypizität wird damit nur an Vorkommensfrequenz gebunden – was häufig ist, ist prototypisch. Das ist natürlich nicht unproblematisch. Als erster Ausgangspunkt ist die Vorgehensweise dennoch vertretbar; sie kann in späteren Untersuchungen beispielsweise mit diachronen Daten angereichert werden.

⁵⁷ Diese Untersuchung wiederholt z. T. die Untersuchung in 3.2.2.7, in der es um die Interaktion der Silbenkonstituenten ging. Hier steht im Gegensatz dazu der Aufbau prototypischer Stämme im Vordergrund.

Das ‚Zentrum‘ der Stämme mit einfachem Kern ist eindeutig die Struktur <CVCC>. Von dort führt jede Veränderung (etwa die Reduktion des Endrands oder die weitere Besetzung des Anfangsrands) schrittweise zu weniger Belegen: <CVCC> hat mehr Belege als <CCVCC>, das wiederum mehr Belege hat als <CCCVCC>.

Bei den zweifach besetzten Kernen zeigt sich ein vergleichbares Bild – allerdings ist das Zentrum horizontal um eine Zelle verschoben, und es gibt insgesamt weniger Belege. Das erklärt sich aus der Interaktion aus Kern und Endrand (vgl. oben 3.2.2.7): Kombinationen aus komplexem Kern und komplexem Endrand sind dispräferiert; komplexe Kerne kombinieren meist mit einfachen Endrändern. In Tabelle 64 genügt nur die Struktur <VV> nicht den Mindestbeschränkungen; sie ist daher schraffiert.

Tab. 64: Anzahl der Strukturen von Einsilbern mit zweifach besetztem Kern im Korpus einfacher deutscher Stämme, sortiert nach der Besetzung des Anfangsrands (Zeilen) und des Endrands (Spalten).

	-	C	CC	CCC+
-		17	3	1
C	33	188	26	3
CC	25	88	7	1
CCC+	6	35	1	

Das Zentrum ist hier die Struktur <CVVC>. Auch hier führt jede Veränderung einer Konstituente zu einer Abnahme in der Anzahl der belegten Stämme.

Im englischen Korpus ergibt sich ein sehr vergleichbares Bild. Bei den Einsilbern mit einfachem Kern bildet die Struktur <CVCC> das Zentrum.

Tab. 65: Anzahl der Strukturen von Einsilbern mit einfachem Kern im Korpus einfacher englischer Stämme, sortiert nach der Besetzung des Anfangsrands (Zeilen) und des Endrands (Spalten).

	-	C	CC	CCC+
-			46	5
C		349	613	89
CC	26	261	370	50
CCC+	3	28	33	4

Wie in den deutschen Daten führt jede Modifikation des Anfangs- und Endrands der Struktur <CVCC> schrittweise zu weniger Belegen. Die Verteilung für die Einsilber mit komplexem Silbenkern entspricht ebenfalls der Verteilung im Deutschen; das kann auch hier mit der Interaktion von Kern und Endrand begründet werden.

Tab. 66: Anzahl der Strukturen von Einsilbern mit zweifach besetztem Kern im Korpus einfacher deutscher Stämme, sortiert nach der Besetzung des Anfangsrands (Zeilen) und des Endrands (Spalten).

	-	C	CC	CCC+
-		28	20	7
C	71	291	114	12
CC	46	189	35	8
CCC+	6	18	2	1

Das Zentrum liegt bei der Struktur <CVVC>; von hier nimmt die Anzahl der Belege ab, egal, in welche Richtung man sich in Tabelle 66 bewegt. Und insgesamt liegen die Zahlen für komplexe Kerne (Tab. 66) niedriger als für einfache Kerne (Tab. 65). Daher resultiert auch eine Veränderung des Kerns in einer niedrigeren Belegzahl.

Man kann die deutschen und die englischen Daten auch so interpretieren: Die Struktur <CVCC> ist der Prototyp für Stämme. Eine Abweichung von dieser Struktur – eine Änderung in der Besetzung einer ihrer Konstituenten – führt dazu, dass die Struktur weniger belegt ist als der Prototyp. Das lässt sich als Präferenzgesetz im Sinne Vennemanns (1988) formulieren:⁵⁸

- (51) Die graphematische Form eines einsilbigen Stamms ist umso mehr bevorzugt,
 (51a) je näher die Anzahl der Grapheme im Anfangsrand bei eins liegt,
 (51b) je näher die Anzahl der Grapheme im Kern bei eins liegt und
 (51c) je näher die Anzahl der Grapheme im Endrand bei zwei liegt.

⁵⁸ Vennemann (1988) basiert seine Präferenzgesetze auf einer ganzen Reihe von sehr unterschiedlichen (synchronen, diachronen, psycholinguistischen, strukturellen) sprachübergreifenden Daten; hier werden lediglich Frequenzdaten benutzt.

Diese präferierte Struktur für Stämme wird in 3.4 mit der präferierten Struktur für Affixe verglichen, die wiederum in 3.3.5 herausgearbeitet wird. Wir wollen hier abschließend noch einen Blick auf die graphemische Gestalt deutscher Wurzeln werfen. Zur Erinnerung: In graphematischen Zweisilbern wurden Endungen als der Bereich definiert, der den Reim der zweiten graphematischen Silbe umfasst – und Wurzeln als ihr Komplement.

Wenn wir die 2.746 Zweisilber im Deutschen so segmentieren und uns gleichzeitig auf die Endungen ⟨e⟩, ⟨el⟩, ⟨en⟩, ⟨er⟩ und ⟨em⟩ beschränken, erhalten wir 1.741 einsilbige Wurzeln wie *Alg* (in *Alge*) oder *Bagg* (in *Bagger*). Damit sind also bereits 63 % der Zweisilber im deutschen Korpus erfasst. Wie sind nun die drei Konstituenten dieser Silbe besetzt, vor allem im Vergleich zu den tatsächlichen einsilbigen Stämmen oben? Wir beschränken uns auf Wurzeln, in denen der Silbenkern einfach besetzt ist.

Tab. 67: Anzahl der Strukturen von einsilbigen Wurzeln mit einfachen Vokalgraphemen im Korpus einfacher deutscher Stämme, sortiert nach der Besetzung des Anfangsrandes (Zeilen) und des Endrands (Spalten).

	-	C	CC	CCC+
-		38	79	6
C	1	330	589	40
CC	5	120	206	11
CCC	1	18	57	3

Wie Tabelle 67 zeigt (und wie vor allem im Abgleich mit Tab. 63 deutlich wird), weisen Wurzeln fast dieselbe Verteilung auf wie Stämme. Prototypisch ist auch hier die Struktur ⟨CVCC⟩ (z. B. *Kaps* in *Kapsel*, *Mess* in *Messer*), und die Verteilung lässt sich mit denselben Präferenzaussagen (51) beschreiben. Das bedeutet: Stämme und Wurzeln verhalten sich sehr ähnlich, zumindest, was die fünf Endungen ⟨e⟩, ⟨el⟩, ⟨en⟩, ⟨er⟩ und ⟨em⟩ angeht (genau wie sich auch Suffixe und Endungen ähnlich verhalten).

3.2.4 Minimalpaare und funktionale Last

Um die Struktur des Systems zu ermitteln, reicht es nicht, die Distribution der Elemente syntagmatisch zu ermitteln, wie das in den vorigen Abschnitten geschehen ist. Mindestens ebenso wichtig ist die funktionale Last der Elemente: Wo

unterscheiden sich wie viele Wörter nur in einem Segment? Wir bewegen uns also entlang der paradigmatischen Achse. Sind Buchstaben oft der einzige Unterschied zwischen zwei Wörtern – wie etwa |n| und |r| in *nein* und *rein* –, so sind sie funktional höher belastet als Buchstaben, die nicht Teil von Minimalpaaren sind. Das ist aus der Leserperspektive unmittelbar einleuchtend: Bei bestimmten Buchstaben führt das ‚Verlesen‘ zu anderen, tatsächlich existierenden Wörtern, bei anderen Buchstaben nicht. Es geht hier gewissermaßen um das Ermitteln von neuralgischen Stellen in Stämmen, und von solchen, die unauffällig sind. Das ist das eine; Minimalpaare können aber auch verwendet werden, um potenzielle Grapheme auf ihre ‚Funktionalität‘ hin zu überprüfen: Ist bspw. die Verbindung |pf| ein Graphem des Deutschen? Im Folgenden wird zuerst die Frage der funktionalen Last bearbeitet, dann geht es um den Graphemstatus von Verbindungen.

Zunächst sind hier einige theoretische Festlegungen notwendig:

- Wir kümmern uns an dieser Stelle um Buchstaben, nicht um Grapheme. Das ist einerseits etwas einfacher zu operationalisieren, andererseits ist es aus Lesersicht unwahrscheinlich, dass Paare unterschiedlicher Länge wie bspw. ⟨t⟩ und ⟨sch⟩ in *treiben* und *schreiben* zu Problemen führen.
- Wir vernachlässigen außerdem die Groß- und Kleinschreibung. Dass in Paaren wie *rag[en]* und *Tag* |r| eigentlich mit |T| (und nicht mit |t|) alterniert, wird also erst einmal ausgeblendet. Stämme werden auch nicht nach Wortarten getrennt; es ist verschiedentlich argumentiert worden, nur Wörter derselben Wortart könnten ‚eigentliche‘ Minimalpaare sein, da nur sie im Satzkontext tatsächlich austauschbar sind (vgl. z. B. Lyons 1968: 132; Meinhold/Stock 1982: 83 f.). Und auch die Vorkommenshäufigkeit der einzelnen Wörter, die an Minimalpaaren teilnehmen, wird nicht weiter untersucht. Das muss berücksichtigt werden, wenn mit der funktionalen Belastung argumentiert wird: Wir untersuchen die funktionale Belastung im Lexikon einfacher Stämme.
- Alternationen mit Zero (wie etwa in *rein* – *ein*) werden nicht untersucht. Relevant sind nur Minimalpaare im engeren Sinne (vgl. für eine kritische Diskussion etwa Wagner 1982: 80 ff.).

Die Minimalpaaranalyse wurde in Berg (2012) dazu verwendet, Buchstaben in Vokale und Konsonanten zu klassifizieren (siehe oben 3.2.1.1). Wir gehen hier einen Schritt weiter und interpretieren die funktionale Last einzelner Elemente, Oppositionen und Strukturpositionen. ‚Funktionale Last‘ (bzw. ‚funktionelle Belastung‘) ist ein Begriff, den der Prager Strukturalismus bekannt gemacht hat. Sein Ursprung liegt allerdings bei Gilliéron (1918), wie King (1967: 831 f.) zeigt. Der Begriff beschreibt mit einer unmittelbar zugänglichen Metapher die Tatsache, dass bestimmte Phoneme an bestimmten Strukturpositionen beson-

ders häufig in Opposition zu anderen Segmenten stehen und damit einer gewissen Belastung unterliegen: Es ist für diese Segmente wichtiger, dass Sprecher und Hörer sie richtig äußern bzw. verstehen, weil ein Versprechen oder Verhören u. U. zu Missverständnissen führen kann. Diese Idee wird nun diachron sehr interessant: Gilliéron (1918) vermutet, dass der Abbau von Oppositionen im Sprachwandel wahrscheinlicher ist, je geringer die funktionale Last dieser Opposition ist. Andersherum: Je mehr Minimalpaare es für eine bestimmte Opposition gibt (je höher also die funktionale Last ist), desto unwahrscheinlicher ist ein Abbau dieser Opposition, da dieser Abbau zu mehr Homonymien führen würde. Trotz zahlreicher Untersuchungen im 20. Jhd. (vgl. z. B. King 1967; Kaplan 2011) konnte diese Hypothese erst in Wedel/Kaplan/Jackson (2013) statistisch-empirisch bestätigt werden.⁵⁹

Die Idee der funktionalen Last wurde in der Phonologie entwickelt; sie ist aber ohne Weiteres auf die Graphematik übertragbar. Wenn man es genau nimmt, ist sie dort sogar sinnvoller anwendbar. So sorgt z. B. die Koartikulation in der gesprochenen Sprache für eine gewisse Redundanz – /t/ wird beispielsweise vor /i/ anders artikuliert als vor /u/ (vgl. z. B. Eisenberg 2013a: 50 f.), sodass der Status von *Tisch/Tusch* als Minimalpaar im Gesprochenen unklar ist. In der geschriebenen Sprache ist das anders: Sie ist strikt und kontextunabhängig segmentiert. Der Buchstabe |t| hat vor allen anderen Buchstaben dieselbe Form. Man könnte also sagen, dass die Idee der Minimalpaare im Geschriebenen überhaupt erst ihre eigentliche Wirkung als heuristische Methode entfalten kann (vgl. auch Eisenberg 1988: 146).⁶⁰

Die Verteilung der Minimalpaare kann aus mindestens drei Perspektiven erfolgen. Zum einen können wir einfach für jeden Buchstaben bestimmen, wie häufig er Teil von Minimalpaaren ist. Zum anderen können wir feststellen, wie häufig zwei Buchstaben miteinander in Opposition stehen – hier ergeben sich Kreuzta-

⁵⁹ Wedel/Kaplan/Jackson (2013) benutzen dazu eine umfangreiche Datenbasis von 56 Phonempaaren, deren Opposition im Laufe der Zeit abgebaut wurde, sowie von 578 Phonempaaren, deren Opposition intakt ist; die Daten stammen aus insgesamt neun Sprachen.

⁶⁰ Eisenberg reagiert hier auf den Vorwurf, seine graphematischen Methoden seien „ein Rückgriff in die Mottenkiste des klassischen Strukturalismus“ (Eisenberg 1988: 146). Er entgegnet überzeugend:

Ein Grund für die Beschränktheit strukturalistischer Ansätze war ja, dass der segmental-sequentielle Aufbau, den die meisten Verfahren stillschweigend unterstellten, die Lautform natürlicher Sprachen nur zum Teil bestimmt. Im Geschriebenen ist er aber jedenfalls in einem sehr viel höheren Maße gegeben als im Gesprochenen. Es könnte daher sein, dass mancher fehlgeschlagene Versuch der Strukturalisten erst in der Graphematik seine Fruchtbarkeit erweist.

bellen wie in Berg (2012). Und schließlich können wir die Perspektive umdrehen und uns die Umgebungen anschauen, in denen Buchstaben in Opposition stehen: Bei *Baum* vs. *Raum* ist das z. B. der einfache Anfangsrand im Einsilber.

Beginnen wir mit ein paar sehr globalen Kennzahlen, mit denen wir das Verhältnis von Minimalpaaren in beiden Korpora etwas näher fassen können. Insgesamt gibt es im englischen Korpus 12.257 Minimalpaare, im deutschen Korpus hingegen 5.367 Minimalpaare. Dass es im englischen Korpus wesentlich mehr Minimalpaare sind als im deutschen, ist zumindest zum Teil auf den größeren Umfang des englischen Korpus zurückzuführen (englisches Korpus: 7.004 einfache Stämme; deutsches Korpus: 5.485 einfache Stämme, vgl. 2.2). Die Anzahl der Wörter, die an Minimalpaaren teilnehmen, unterscheidet sich hingegen weniger stark. Im deutschen Korpus sind 2.336 der 5.485 Stämme Teile von Minimalpaaren (43%); im englischen Korpus sind es 3.454 von 7.004 Stämmen (49%).

Wie kann das sein? Es gibt im englischen Korpus häufiger Minimalpaare, die sich einen Kontext ‚teilen‘ – die also eigentlich Minimaltripel, -quadrupel etc. sind. Ein Beispiel für ein Quintupel ist z. B. die Menge *cite*, *kite*, *mite*, *rite*, *site*. Diese fünf Wörter entsprechen nun nicht fünf Minimalpaaren, sondern zehn (die seien hier nur der Vollständigkeit halber aufgeführt: *cite/kite*, *cite/mite*, *cite/rite*, *cite/site*, *kite/mite*, *kite/rite*, *kite/site*, *mite/rite*, *mite/site*, *rite/site*). Wir suchten also Kontexte (z. B. im Beispiel oben *_ite*) und bestimmten die Anzahl entsprechender Wörter (im Beispiel oben fünf). Tabelle 68 zeigt für das englische und deutsche Korpus, wie viele Minimalpaare, -tripel etc. zu finden sind.

Tab. 68: Anzahl der Minimalpaare, -tripel, -quadrupel etc. im deutschen und englischen Teilkorpus einfacher Stämme.

Größe der Menge	deutsch		englisch	
	#Kontexte	% Kontexte	#Kontexte	% Kontexte
2	1.696	71,9 %	2107	63,8 %
3	414	17,5 %	586	17,8 %
4	149	6,3 %	281	8,5 %
5	57	2,4 %	117	3,5 %
6	19	0,8 %	77	2,3 %
7	11	0,5 %	46	1,4 %
8	8	0,3 %	32	1,0 %
9	5	0,2 %	24	0,7 %
10+	1	0,0 %	31	0,9 %

Die Unterschiede zwischen den beiden Korpora sind absolut nicht besonders groß (mit Ausnahme vielleicht der Minimalpaare im engeren Sinne, also der Kontexte, denen nur zwei Wörter entsprechen). Sie reichen aber trotzdem aus, um den sehr großen Unterschied in der Gesamtzahl der Minimalpaare zu motivieren. Allein die 24 Kontexte im englischen Korpus, denen neun Wörter entsprechen, sind für 864 Minimalpaare verantwortlich; die fünf entsprechenden Kontexte im deutschen Korpus nur für 144 Minimalpaare. Insgesamt ergeben die Abweichungen in Tabelle 68 die Differenz von knapp 7.000 Minimalpaaren zwischen den Korpora.

Welche Buchstaben nehmen nun häufig, welche selten an Minimalpaaren teil? Abbildung 37 zeigt den relativen Anteil der Buchstaben an allen Minimalpaaren:

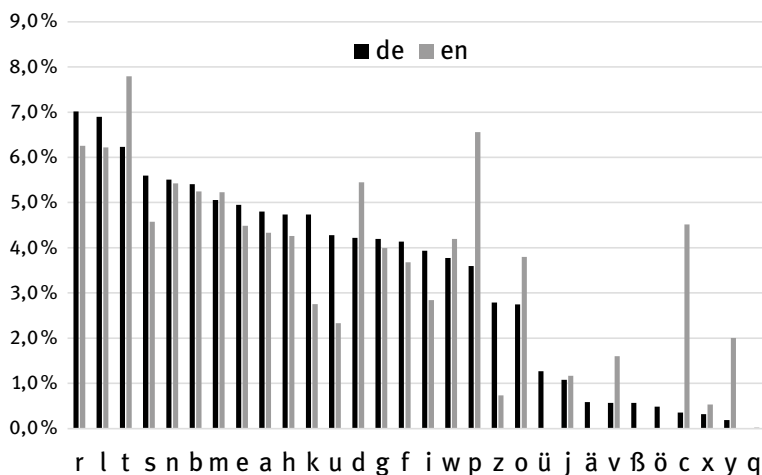


Abb. 37: Relativer Anteil der Buchstaben in Minimalpaaren im deutschen und englischen Korpus.

Dafür, dass es sich um Buchstaben in Minimalpaaren in zwei Sprachen handelt, weisen die Daten einen hohen Grad an Gemeinsamkeit auf. Nur neun Buchstaben weichen in beiden Korpora um mehr als ein Prozent voneinander ab:

- Im deutschen Korpus sind |k|, |u|, |i| und |z| in Minimalpaaren häufiger als im englischen Korpus. |k| und |z| sind im englischen Korpus absolut (also außerhalb von Minimalpaaren) wesentlich seltener als im deutschen (vgl. Abb. 1 oben). Es gibt hier also gewissermaßen weniger ‚Material‘ für Minimalpaare. Der höhere Anteil von |i| und |u| im deutschen Korpus ist interessant: die beiden Buchstaben sind die bevorzugten zweiten Bestandteile von <VV>-Silben-

- kernen (siehe oben 3.2.2.6), und in dieser Strukturposition stehen sie häufig in Opposition mit Konsonanten (z. B. *kein/Kerl*; dazu unten mehr).
- Im englischen Korpus sind |t|, |d|, |p|, |c| und |y| häufiger als im deutschen. Mit |c| und |y| verhält es sich umgekehrt wie mit |k| und |z|: |c| und |y| sind im englischen Korpus absolut (außerhalb von Minimalpaaren) wesentlich häufiger als im deutschen Korpus. Damit ist auch die Wahrscheinlichkeit größer, dass sie an Minimalpaaren teilnehmen. |d| und |p| kommen im englischen Korpus wesentlich häufiger als einfache Anfangsränder vor, und |d|, |p| und |t| kommen wesentlich häufiger als einfache Endränder vor (vgl. Abb. 12 und Abb. 19 oben). Einfache Anfangs- und Endränder sind die prädestinierten Positionen für Minimalpaare (siehe unten).

|r|, |l| und |t| sind also im deutschen Korpus, |t|, |p| und |r| im englischen Korpus am häufigsten Teil von Minimalpaaren; diese Buchstaben sind funktional am meisten belastet. Ihr Verlesen oder Verschreiben führt häufiger als das anderer Buchstaben dazu, dass ein anderes Lexem ‚aktiviert‘ wird.

Nun sagen diese relativen Zahlen noch nichts darüber aus, wie häufig die Minimalpaare relativ zu allen Vorkommen eines Buchstabens sind. |r| und |t| etwa dominieren in beiden Sprachen die Minimalpaare, aber beides sind auch häufige Buchstaben. Es ist hilfreich, auch einen Blick auf die Anzahl der Minimalpaare relativ zu allen Buchstaben zu werfen. Konkret fragen wir, wie viele Vorkommen eines bestimmten Buchstabens ersetzbar sind. Relevant ist die Anzahl der Kontexte, nicht die der Minimalpaare. Anders ausgedrückt: Da im Moment von Interesse ist, wie häufig Buchstaben generell ersetzbar sind, beschränken wir uns hier darauf, ob ein Buchstabe an einer bestimmten Position ersetzbar ist oder nicht, unabhängig davon, mit wie vielen anderen Buchstaben er in Opposition steht. Die Daten werden in Abbildung 38 (nächste Seite) präsentiert.

Die Übersicht zeigt, wie häufig die einzelnen Buchstaben funktional belastet sind. Sie zeigt nicht, wie stark sie funktional belastet sind (ob mehr als ein anderes Wort den Kontext teilt). Andersherum: Wenn wir für einen beliebigen Buchstaben ein beliebiges Wort mit diesem Buchstaben heraussuchen, dann sagt uns Abbildung 38, wie hoch die Wahrscheinlichkeit ist, dass der Buchstabe Teil eines Minimalpaares ist. Über die Hälfte aller Vorkommen von |w| im Englischen sind bspw. Teil von Minimalpaaren, in denen |w| durch einen anderen Buchstaben ersetzt werden kann.

Aus dieser Perspektive unterscheiden sich die beiden Korpora deutlicher voneinander. Insgesamt sind fast alle Buchstaben im englischen Korpus häufiger belastet als im deutschen; die Ausnahmen sind |z| und |u| (sowie trivialerweise |ä|, |ö|, |ü| und |ß|). In beiden Sprachen sind seltene Buchstaben am meisten belastet (|ß|, |w|, |j|, |ü|, |ö| im deutschen Korpus und |w|, |j|, |k| im englischen).

Das ist keine hinreichende Bedingung, wie |q| und |v| zeigen: Sie sind in beiden Sprachen selten, funktional aber nicht häufig belastet. Der größte Unterschied zwischen beiden Sprachen findet sich bei |c|, das im deutschen Korpus kaum, im englischen Korpus hingegen in einem Viertel aller Vorkommen funktional belastet ist. Dieser Unterschied erklärt sich durch die enge syntagmatische Verbindung, die |c| im Deutschen mit |h| und |k| eingeht (siehe oben 3.1). Im Englischen tritt |c| wesentlich autonomer auf und kann entsprechend an Minimalpaaren teilnehmen; im Deutschen ist das die Ausnahme.

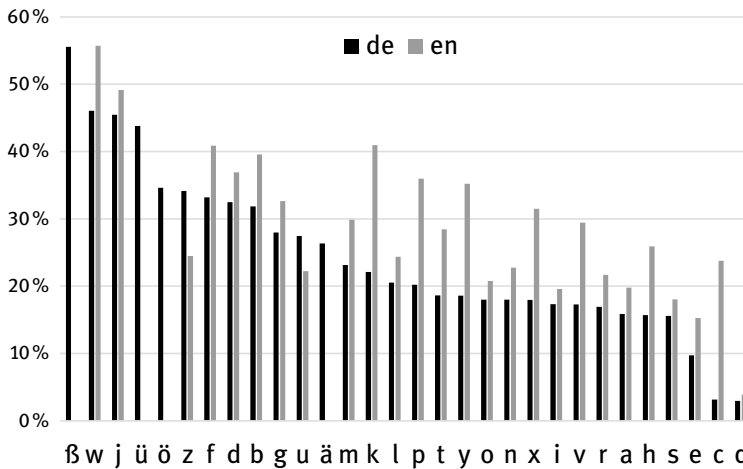


Abb. 38: Anteil der Buchstaben, die an Minimalpaaren teilnehmen, relativ zu allen Vorkommen der betreffenden Buchstaben.

Was alterniert bei diesen Minimalpaaren nun miteinander? Gehen wir auch diese Frage erst einmal sehr allgemein an und betrachten, welche Klassen von Buchstaben miteinander alternieren:

Tab. 69 (links) und **Tab. 70 (rechts):** Anzahl der Minimalpaare nach Kategorienkombination der Buchstaben im deutschen (Tab. 69) und englischen (Tab. 70) Teilkorpus.

	C	V		C	V
C	3.829 (71 %)		C	9.035 (74 %)	
V	583 (11 %)	955 (18 %)	V	1.591 (13 %)	1.631 (13 %)

In beiden Korpora überwiegen eindeutig Minimalpaare, in denen Konsonantenbuchstaben miteinander alternieren. An zweiter Stelle stehen jeweils Paare, in denen zwei Vokalbuchstaben alternieren, und erst an letzter Stelle solche, in denen ein Vokal- und ein Konsonantenbuchstabe das Minimalpaar bilden (der Unterschied zwischen diesen beiden letzten Kategorien ist allerdings im englischen Korpus gering).

Der Grund für diese Verteilung ist zumindest teilweise stochastisch. Es gibt in beiden Sprachen deutlich mehr Konsonanten- als Vokalbuchstaben, und auch in einzelnen Wörtern sind Konsonantenbuchstaben häufiger als Vokalbuchstaben (siehe oben Abb. 10). Für jedes gegebene Wort ist die Anzahl potenzieller Wörter, die sich nur in einem Konsonantenbuchstaben unterscheiden, also größer als die Menge der Wörter in den übrigen Kategorien.

Welche Buchstaben stehen nun häufig, welche selten in Opposition? Eine komplette Übersicht in Form einer Kreuztabelle (wie in Berg 2012) findet sich für beide Sprachen in Anhang C. An dieser Stelle sollen nur die 20 häufigsten Minimalpaare betrachtet werden. Tabelle 71 präsentiert die Daten aus dem deutschen Korpus.

Tab. 71: Die 20 häufigsten Minimalpaare im deutschen Korpus einfacher Stämme.

Buchstaben	#Minimalpaare	Beispiel
a, e	122	dank[en], denk[en]
l, r	120	flau, Frau
a, i	99	bald, Bild
a, u	80	Hammer, Hummer
n, r	74	Ende, Erde
e, i	71	Fenster, finster
a, o	69	Falter, Folter
s, t	68	kraus, Kraut
e, o	67	Herde, Horde
l, n	64	bald, Band
i, u	63	bitter, Butter
r, t	60	Flur, Flut
e, u	60	Kegel, Kugel
h, l	55	stehl[en], stell[en]
l, m	53	Lade, Made
l, s	52	Hals, Hass
n, t	51	gerne, Gerte
r, s	51	Kurs, Kuss
o, u	50	groß, Gruß
l, t	50	Salz, Satz

Unter den 20 häufigsten Minimalpaaren sind keine, in denen Konsonantengegen Vokalbuchstaben ausgetauscht werden. Vokal-Vokal-Paare sind mit neun Vorkommen fast so häufig in Tabelle 71 vertreten wie Konsonant-Konsonant-Paare, obwohl sie insgesamt deutlich seltener sind (vgl. Tab. 69 oben). Es gibt eben weniger mögliche Vokal-Vokal-Kombinationen als Konsonant-Konsonant-Kombinationen; unter diesen wenigen Kombinationen sind aber eine Reihe hochfrequente wie z. B. |a|/|e|, |a|/|i| oder |a|/|u|.

Bei den Vokal-Kombinationen fehlt nur |i|/|o| zu einer vollständigen Kombinatorik der nicht-umgelauteten Buchstaben; alle anderen Paare sind unter den 20 häufigsten Minimalpaaren vertreten. Bei den Konsonanten sind mit Ausnahme von |h| und |m| (|h|/|l| und |l|/|m|) nur die fünf frequentesten Buchstaben vertreten, und hier fehlt nur |n|/|s| zu einer vollständigen Kombinatorik. Das häufigste Konsonantenpaar, |l|/|r| ist aus formaler Perspektive besonders interessant: Es zeigt, dass die Opposition der beiden Buchstaben vergleichsweise stark belastet ist. Das wiederum spricht für die Form von |l|, die sich stark von |r| unterscheidet. Der Buchstabe |l| ist ja im System von Fuhrhop/Buchmann (2009) problematisch. Die Annahme, dass er einen kurzen Kopf hat, der das Mittelband nicht überschreitet, und eine Coda, die oben an diesem Kopf ansetzt, wirkt ad hoc. Auf den ersten Blick sieht |l| aus wie bspw. |b| ohne den Bogen – also ein Coda-loser Buchstabe. Nun ergibt sich aber aus der Verteilung von |l|, dass der Buchstabe sich verhält wie |m|, |n| und |r| (siehe oben 3.2.2.4) – und diese Buchstaben haben alle einen kurzen, geraden Kopf, die Coda ist oben angeschlossen. Eisenberg (2013a) zeigt in diesem Zusammenhang, dass es mit dem vorhandenen Inventar an Teilformen gar nicht möglich ist, einen anderen Buchstaben mit geradem Kopf zu konstruieren, dessen Coda oben angeschlossen ist. Dieses Argument hinkt etwas: Es ist ja durchaus möglich, sich bspw. ein vertikal gespiegeltes |r| vorzustellen, also |ɾ|. Dieser minimale Unterschied in der Orientierung wäre angesichts der relativ großen funktionalen Belastung zu gering, so lässt sich nun vermuten. Für |b| und |d| sowie |p| und |q| gilt das nicht – sie sind funktional wesentlich weniger belastet (|p|/|q| gar nicht, |b|/|d| 39-mal).

Das Bild im englischen Korpus sieht vergleichbar aus, wie Tabelle 72 zeigt (ebenfalls mit den 20 häufigsten Minimalpaaren):

Tab. 72: Die 20 häufigsten Minimalpaare im englischen Korpus einfacher Stämme.

Buchstaben	#Minimalpaare	Beispiel
a, o	248	batch, botch
a, i	231	bard, bird
l, r	198	bleak, break
a, e	183	meat, meet
a, u	177	crash, crush
d, t	162	side, site
p, t	161	hip, hit
i, o	157	find, fond
n, t	148	near, tear
e, o	143	cellar, collar
n, r	136	nude, rude
b, p	133	back, pack
l, n	130	fold, fond
i, u	129	gin, gun
l, t	128	hall, halt
s, t	123	muse, mute
r, t	121	rest, test
m, p	119	harm, harp
b, t	116	cube, cute
o, u	115	son, sun

Auch im englischen Korpus finden sich unter den 20 häufigsten Minimalpaaren keine Kombinationen aus Vokal- und Konsonantenbuchstaben, und auch hier sind Vokal-Vokal- und Konsonant-Konsonant-Kombinationen relativ ausgewogen verteilt (8x Vokal-Vokal, 12x Konsonant-Konsonant).

Bei den Vokal-Vokal-Kombinationen ist |y| überhaupt nicht vertreten. Bei den übrigen Buchstaben fehlen nur die Paare |e|/|i| und |e|/|u| zu einer vollständigen Kombinatorik. Bei den Konsonanten sind allerdings – anders als im deutschen Korpus – nicht nur die frequentesten Buchstaben Teil der frequentesten Minimalpaare. Vor allem |b|, |p| und |d| kommen mehrfach vor, obwohl sie nicht zu den häufigsten Buchstaben gehören. Die am stärksten belastete konsonantische Kombination ist auch hier ||/|r| – das Argument, dass die beiden Buchstaben formal nicht zu ähnlich sein sollten, greift also im englischen Korpus ebenso wie im deutschen.

Wie oben bereits angedeutet, sind die Minimalpaare nicht über alle silbenstrukturellen Positionen gleichermaßen verteilt. Schauen wir uns in diesem Zusammenhang zunächst an, wie viele Silben die Wörter haben, die an Minimalpaaren teilnehmen.

Tab. 73: Anzahl der einfachen Stämme im deutschen und englischen Korpus, die Teil von Minimalpaaren sind, aufgeschlüsselt nach der Anzahl der graphematischen Silben dieser Stämme.

#Silben	deutsch	englisch
1	1.763	2.595
2	1.365	1.559
3	48	68
4	5	2

Minimalpaare kommen in beiden Korpora also absolut am häufigsten bei Einsilbern vor. Hier ist der Grad der Abdeckung sehr hoch. Im deutschen Korpus gibt es 2.015 Einsilber, im englischen 2.798. Von allen Einsilbern in den Korpora nehmen also im deutschen Korpus 87 % an Minimalpaaren teil, im englischen Korpus sogar 93 %. Bei den Zweisilbern liegen die Werte in beiden Korpora wesentlich niedriger (50 % im deutschen und 49 % im englischen Korpus). Drei- und Viersilber kommen nur noch marginal in Minimalpaaren vor.

Minimalpaare gibt es also vor allem bei Ein- und Zweisilbern, und gerade bei Einsilbern nimmt der Großteil der Stämme an Minimalpaaren teil. Bemerkenswert ist wiederum der Grad der Parallelität zwischen den englischen und deutschen Daten, der die Frage aufdrängt, wie es sich in anderen Sprachen verhält.

Welche silbenstrukturellen Positionen in Ein- und Zweisilbern sind nun wie stark belastet? Betrachten wir dazu zunächst die Konsonant-Konsonant-Paare in Einsilbern. Die Strukturpositionen Anfangsrand und Endrand werden in Tabelle 74 und 75 jeweils nach der Anzahl der Buchstaben, mit denen sie besetzt sind, untersucht. Die Zahl ‚1.510‘ in Zeile ‚1‘, Spalte ‚Anfangsrand‘ in Tabelle 74 besagt, dass es 1.510 Minimalpaare im deutschen Korpus gibt, bei denen die betreffenden Buchstaben alleine den Anfangsrand besetzen (z. B. *Dorf/Torf*). Zur Erinnerung: Es geht hier um Minimalpaare, nicht um Wörter, die Teil dieser Minimalpaare sind – deswegen liegen die Zahlen in dieser Tabelle höher, als man es aus Tabelle 73 erwarten würde. Und es geht nach wie vor um Buchstaben, nicht Grapheme, deswegen spielen <ch> und <ck> im Folgenden keine Rolle.

Tab. 74 (links) und Tab. 75 (rechts): Anzahl der Konsonant-Konsonant-Minimalpaare im deutschen (Tab. 74) und englischen (Tab. 75) Teilkorpus einfacher Stämme, klassifiziert nach dem Ort der Ersetzung (Anfangsrand/Endrand) sowie der Besetzung der jeweiligen Konstituente.

#Buchstaben	Anfangsrand	Endrand	#Buchstaben	Anfangsrand	Endrand
1	1.510	502	1	3.491	1.455
2	149	493	2	711	742
3	3	17	3	21	21
4	28	6	4	0	0

Am frequentesten (und zwar mit einigem Abstand) sind im deutschen und englischen Korpus Minimalpaare im einfachen Anfangsrand, wie Tabelle 74 und Tabelle 75 zeigen. Paare wie *Butter/Kutter* oder *light/fight* sind der Standardfall. Einfache Anfangsränder sind strukturell also am meisten belastet, hier leisten die Buchstaben die größte Distinktion, hier tragen sie in einem gewissen Sinn am meisten Information (dazu unten mehr). Die zweithäufigste Position ist in beiden Korpora der einfach besetzte Endrand. Minimalpaare in drei- und vierfach besetzten Silbenkonstituenten sind sehr selten.

Warum sind einfache Anfangsränder so viel häufiger als einfache Endränder unter den Minimalpaaren zu finden (im deutschen Korpus im Verhältnis 3:1, im englischen im Verhältnis 5:2)? Sind einfache Anfangsränder vielleicht generell häufiger als einfache Endränder? Das ist in der Tat der Fall. Im deutschen Korpus gibt es 1.166 Einsilber mit einfachem Anfangsrand gegenüber 667 Einsilbern mit einfachem Endrand; im englischen Korpus sind es 1.579 Einsilber mit einfachem Anfangs- und 1.195 Einsilber mit einfachem Endrand.

Es lohnt sich außerdem, hier noch einmal die Anzahl der entsprechenden Wörter zu ermitteln, die an Minimalpaaren teilnehmen. Im deutschen Korpus gibt es 913 Einsilber mit einfachem Anfangsrand, bei denen dieser Anfangsrand ersetzt werden kann. Das heißt: In 78 % aller Einsilber mit einfachem Anfangsrand ist dieser Anfangsrand funktional belastet. Bei den einfachen Endrändern gibt es 446 Wörter; einfache Endränder sind also in 67 % der Fälle distinktiv. Im deutschen Korpus löst sich die Asymmetrie zwischen Anfangs- und Endrand so nicht vollständig auf; einfache Anfangsränder sind häufiger funktional belastet als einfache Endränder.

Im englischen Korpus ergibt sich ein etwas anderes Bild: Hier gibt es 1.360 Einsilber, deren einfacher Anfangsrand ersetzt werden kann (das sind 86 % aller Einsilber mit einfachem Anfangsrand), und 974 Einsilber mit einfachem End-

rand, der funktional belastet ist (das sind 82% aller entsprechenden Einsilber). Diese Werte unterscheiden sich kaum – wie kommt es dann zu dem relativ großen Unterschied in Tabelle 74 oben? Dort wurden Minimalpaare gezählt, hier Wörter, die an Minimalpaaren teilnehmen (anders ausgedrückt: Kontexte), und wie oben bereits ausgeführt unterscheiden sich beide Zählweisen. Der Unterschied in der Zahl der Minimalpaare deutet darauf hin, dass es bei einfachen Anfangsrändern im englischen Korpus häufig Minimaltripel, -quadrupel etc. gibt, während einfache Endränder eher Minimalpaare bilden. Wir können festhalten: Der einfache Endrand (im Einsilber) ist im Englischen stärker funktional belastet als im Deutschen.

Bei den zweifach besetzten Rändern sind jeweils nicht beide Buchstaben gleich häufig Teil von Minimalpaaren. Die jeweils randnahen Bestandteile werden häufiger ersetzt als die kernnahen Bestandteile. Minimalpaare vom Typ *Blatt/glatt* oder *mask/mast* sind also häufiger als solche vom Typ *grow/glow* oder *Samt/satt*. Tabelle 76 und Tabelle 77 präsentieren die Anzahl der Minimalpaare nach den Silbenkonstituenten und der Position (1. oder 2.).

Tab. 76 (links) und **Tab. 77 (rechts)**: Anzahl der Minimalpaare im deutschen (Tab. 76) und englischen (Tab. 77) Teilkorpus einfacher Stämme in zweifach besetzten Silbenkonstituenten nach der Konstituente (Anfangsrand/Endrand) und der Position der Ersetzung (1./2. Position in der Konstituente).

Buchstabe	Anfangsrand	Endrand	Buchstabe	Anfangsrand	Endrand
1. BS	111	180	1. BS	403	331
2. BS	38	313	2. BS	308	411

Die Tendenz, dass randnahe Buchstaben eher an Minimalpaaren teilnehmen als kernnahe, ist im deutschen Korpus ausgeprägter als im englischen. Sie ist (zumindest zum Teil) stochastisch erklärbar: Kernnah kommen in beiden Korpora vor allem |l|, |m|, |n|, |r| sowie (nach dem Kern) |h| vor; davor bzw. danach stehen u. a. |b|, |d|, |f|, |g|, |k|, |p|, |s|, |t| und |z| (das ist etwas vereinfacht – vgl. oben 3.2.2.4). Die Menge der möglichen Alternativen ist also für einen typischen Stamm wie *Blatt* stamminitial (in diesem Fall |b|) größer als an der Position danach (hier |l|).

Bis jetzt haben wir uns Konsonant-Konsonant-Minimalpaare nur im Einsilber angeschaut. Im Zweisilber ergibt sich zunächst folgende Verteilung von Minimalpaaren (hier mit der zusätzlichen Position ‚intervokalisch‘):

Tab. 78 (links) und Tab. 79 (rechts): Anzahl der Konsonant-Konsonant-Minimalpaare im deutschen (Tab. 78) und englischen (Tab. 79) Teilkorpus einfacher Stämme, klassifiziert nach dem Ort der Ersetzung (Anfangsrand/intervokalisch/Endrand) sowie der Besetzung der jeweiligen Konstituente.

#Buchstaben	Anfangsrand	intervokalisches	Endrand	#Buchstaben	Anfangsrand	intervokalisches	Endrand
1	597	216	66	1	1.250	933	31
2	23	188	5	2	176	87	5
3	3	1	0	3	3	26	0
4	5	0	0	4	0	0	0

Die wichtigste Beobachtung ist: Endränder von zweisilbigen Stämmen sind in beiden Korpora nur sehr selten funktional belastet. Die relativ wenigen Minimalpaare haben meist einen einfachen Endrand und sind vom Typ *Magen/mager*, *seven/sever*. Diese geringe funktionale Belastung spricht ebenfalls dafür, dass graphematische Zweisilber in eine Wurzel und eine Endung zerfallen. Die Endungen sind nicht wirklich kompositionell aufgebaut, sondern größere Bausteine, und das spiegelt sich hier in der geringen funktionalen Belastung der einzelnen Buchstaben wider.

Eine andere interessante Beobachtung betrifft die intervokalischen Konsonanten. Wir haben oben gesehen, dass sich <CC>-Cluster syntagmatisch (graphotaktisch) verhalten wie Endränder von Einsilbern, und diese Parallelität wiederholt sich hier. Tabelle 80 und 81 zeigen, wie häufig erste und zweite Bestandteile von intervokalischen <CC>-Clustern an Minimalpaaren teilnehmen:

Tab. 80 (links) und Tab. 81 (rechts): Anzahl der Minimalpaare im deutschen (Tab. 80) und englischen (Tab. 81) Teilkorpus einfacher Stämme in zweifach besetzten intervokalischen Clustern von Zweisilbern nach der Position der Ersetzung (1./2. Position).

#Buchstaben	intervokalisches	#Buchstaben	intervokalisches
1. BS	72	1. BS	35
2. BS	116	2. BS	52

Zweite Bestandteile von intervokalischen Clustern sind in beiden Korpora häufiger ersetzbar als erste Bestandteile. Das ist dieselbe Verteilung wie bei den End-

rändern. Auch hinsichtlich der Minimalpaare verhalten sich intervokalische Cluster also eher wie Endränder denn wie Anfangsränder. Das spiegelt die oben für die Graphotaktik gemachte Beobachtung wider; auch dort verhielten sich intervokalische Cluster eher wie Endränder von graphematischen Silben.

Wie sieht die Verteilung der Minimalpaare aus Vokalbuchstaben aus? Wir schauen uns auch hier Ein- und Zweisilber an, und bei Zweisilbern wiederum die erste und die zweite Silbe getrennt voneinander. Tabelle 82 und Tabelle 83 zeigen die Verteilung im deutschen und englischen Korpus:

Tab. 82 (links) und **Tab. 83 (rechts)**: Anzahl der Vokal-Vokal-Minimalpaare im deutschen (Tab. 82) und englischen (Tab. 83) Teilkorpus einfacher Stämme, klassifiziert nach der Anzahl der graphematischen Silben des Stamms sowie der Besetzung des Silbenkerns.

#Buch- staben	1-Silber	2-Silber		#Buch- staben	1-Silber	2-Silber	
		1. Silbe	2. Silbe			1. Silbe	2. Silbe
1	594	260	37	1	922	356	71
2	46	5	1	2	260	10	6
3	1	0	0	3	2	0	0

Wie bei den Konsonanten-Minimalpaaren gibt es deutlich mehr Minimalpaare bei Einsilbern als bei Zweisilbern, und einfache Kerne sind (wie einfache Ränder) ebenfalls deutlich präferiert. Interessant im Zusammenhang mit der Diskussion um den Status von Endungen ist der Unterschied von ersten und zweiten Silben in Zweisilbern: In Zweisilbern sind Minimalpaare in ersten Silben wesentlich häufiger als solche in zweiten Silben. Das spricht ebenfalls für die oben präsentierte Analyse einer Wurzel und einer Endung: Genau wie einfache Endränder in Zweisilbern sind auch einfache Kerne in den zweiten Silben wesentlich seltener funktional belastet, also wesentlich seltener ersetzbar. Unter der vorgeschlagenen Analyse ergibt sich diese Verteilung: Endungen sind wiederkehrende größere Einheiten, die aber nicht ohne Weiteres aufgebrochen werden können.

Werfen wir abschließend noch einen Blick auf die Minimalpaare aus Vokal und Konsonant wie etwa *kein/Kerl*. Tabelle 84 präsentiert die zehn häufigsten Strukturen, die knapp zwei Drittel der Minimalpaare dieses Typs im deutschen Korpus ausmachen:

Tab. 84: Die 10 häufigsten Vokal-Konsonant-Minimalpaare im deutschen Korpus einfacher Stämme.

Struktur	#Minimalpaare	Beispiel
CV_C	100	Heim/Helm
CVC_	67	Nase/nass
CV_	66	les[en], leg[en]
_VC	31	Aal/Wal
CCV_	28	Grab/grau
CV_CV	21	Raupe/Rampe
CCV_C	20	kraus/krass
CV_VC	14	Lauer/Lager
CV_CVC	13	heikel/Henkel
CCVC_	13	Blase/blass

Acht der zehn häufigsten Types beinhalten einen komplexen Silbenkern in einem Teil des Minimalpaars (z. B. CV_C, *Heim/Helm*). In sieben von diesen Types wird der zweite Teil des komplexen Silbenkerns ersetzt. Die Ausnahme ist die Struktur _VC (*Aal/Wal*). Dieses Muster zieht sich durch den Rest der Strukturen (nicht nur die häufigsten): 335 Minimalpaaren, in denen der zweite Teil ersetzt wird, stehen nur 40 vom Typ *Aal/Wal* gegenüber, bei denen der erste Teil ersetzt wird.

Das spricht für eine Analyse von Vokalverbindungen, die nur dem ersten Teil die (relationale) Funktion des Silbenkerns zuspricht (Fuhrhop/Berg einger.), und zwar aufgrund formaler und distributioneller Kriterien. Hier kommt ein neues, paradigmatisches Argument hinzu: Zweite Bestandteile von Vokalverbindungen können wesentlich häufiger durch Konsonanten ersetzt werden, verhalten sich also in gewisser Weise mehr wie Konsonanten, als es erste Bestandteile tun.⁶¹

Im englischen Korpus sieht die Aufstellung der zehn häufigsten Strukturen von Vokal-Konsonant-Minimalpaaren zunächst ähnlich aus (siehe Tab. 85).

Hier haben acht von zehn Strukturen einen komplexen Silbenkern. Von diesen acht Strukturen wird aber nur bei vieren der zweite Bestandteil der Vokalverbindung ersetzt. Bei den übrigen vier Strukturen wird – wie etwa bei _VC (*eat/mat*) der jeweils erste Bestandteil ersetzt. Insgesamt gibt es mehr Minimalpaare, bei denen der zweite Teil einer Vokalverbindung ersetzt wird (714 Minimalpaare vom Typ *may/man*) als solche, bei denen der erste Teil ersetzt wird (481 Minimalpaare vom Typ *eat/mat*). Der Unterschied ist aber längst nicht so deutlich wie im Deutschen.

⁶¹ Der Anstoß für diese Idee kam von Niklas Reinken.

Tab. 85: Die 10 häufigsten Vokal-Konsonant-Minimalpaare im englischen Korpus einfacher Stämme.

Struktur	#Minimalpaare	Beispiel
CV_	298	may/mat
CVC_	274	bare/barn
CV_C	172	bail/ball
_VC	140	cat/eat
CCV_	101	clue/club
_VCC	87	east/mast
C_VC	70	coat/chat
CCVC_	64	stare/start
CCV_C	34	blood/blond
C_V	30	soy/spy

Wird ein finaler Buchstabe direkt nach einem Konsonanten ersetzt (z. B. der Typ CVC_ *bare/barn* oben), dann ist in den meisten Fällen |e| involviert (in 312 von 378 Minimalpaaren, das entspricht 83 %). Das ist nicht überraschend: Finales |e| ist mit Abstand das häufigste vokalische Stammende (siehe oben 3.2.1.5).

Fassen wir zusammen: Minimalpaare treten in beiden Korpora am häufigsten in Einsilbern auf, und es werden am häufigsten Konsonanten gegeneinander ausgetauscht. Einfache Anfangsränder sind als Strukturpositionen am stärksten von Minimalpaaren betroffen – und damit funktional am stärksten belastet. Mehr als drei Viertel aller Einsilber mit einfachem Anfangsrand nehmen im deutschen Korpus an Minimalpaaren teil; im englischen Korpus sind es sogar 86 %. Einfache Anfangsränder tragen mehr dazu bei, einsilbige Stämme formal von anderen Stämmen zu unterscheiden als andere Konstituenten.

Nach einfach besetzten Anfangsrändern folgen im deutschen Korpus einfache Kerne in Einsilbern und dann einfache Endränder, im englischen Korpus ist es umgekehrt. Bei Zweisilbern ist der Endrand nur noch selten an Minimalpaaren beteiligt, und für den Silbenkern der zweiten Silbe gilt dasselbe. Die Endung von Zweisilbern – das Komplement von Tafts (1979) BOSS – verhält sich eher wie eine Einheit.

Was haben wir methodisch gemacht? Wir haben Minimalpaare nach den involvierten Buchstaben, nach Silben und nach Strukturpositionen klassifiziert und so typische Minimalpaare im englischen und deutschen Korpus herausgearbeitet, mit neuralgischen Stellen (dem stark belasteten einfachen Anfangsrand) und Stellen, die nur selten Teil von Minimalpaaren sind (komplexe Silbenränder und besonders die Bestandteile von zweiten Silben in Zweisilbern).

Hier ist noch sehr viel mehr vorstellbar. Theoretisch sind die Minimalpaare ein Netzwerk: Wir haben Einheiten (die einzelnen Wörter) und Relationen zwischen diesen Einheiten (die Relation ‚Minimalpaar‘). Sie bilden also im eigentlichen Sinn des Wortes ein System, und in diesem System ist das lexikalische Wissen des Deutschen bzw. Englischen organisiert. Dieses System kann durchaus psychologische Realität beanspruchen: Wie seit den Arbeiten von Andrews (1989, 1992) bekannt ist, hat die Anzahl der graphematischen Minimalpaare, an denen ein Wort teilnimmt, einen direkten Einfluss auf die Geschwindigkeit, mit der Wörter bzw. Nichtwörter in lexikalischen Entscheidungsaufgaben erkannt werden (das gilt zumindest für nicht hochfrequente Lexeme). Die Anzahl der Minimalpaare für ein gegebenes Wort firmiert seit Coltheart et al. (1977) sehr anschaulich unter dem Begriff der ‚orthographischen Nachbarschaft‘. Denkt man diese Metapher weiter, dann handelt es sich beim System der Minimalpaare um orthographische Straßenzüge, orthographische Städte oder orthographische Landschaften.

Es lassen sich nun mit einfachen Mitteln aus einer Liste der Einheiten und Relationen Netzwerke visualisieren (bspw. mit dem frei verfügbaren Programm Gephi, www.gephi.org, Stand: 22. 11. 2018). Es gibt verschiedene Algorithmen, die nach einem ähnlichen Prinzip verfahren: Je mehr Verbindungen ein Element zu anderen Elementen hat, desto zentraler wird es dargestellt; je weniger Verbindungen existieren, desto peripherer erscheint es. Die folgende Darstellung (Abb. 39 auf der folgenden Seite) zeigt einen Teilbereich des Minimalpaar-Netzwerks mit vier Buchstaben im deutschen Korpus (verwendeter Algorithmus: Yifan Hu).

Darstellungen wie diese lassen sich zunächst einmal explorativ verwenden. Wo gibt es Cluster und was sind die gemeinsamen Merkmale? In Abbildung 39 ist ein sehr dichter Cluster auf der linken Bildseite zu erkennen, in dem sich ausschließlich Stämme mit finalelem <ck> befinden. Auf der rechten Bildseite sind die Daten nicht so eng vernetzt und nicht so klar strukturiert. Viele Stämme enthalten finaleles <nk> (*Zink, pink, senk[en]*), zum Teil aber auch finaleles <nn> (*dünn, Tann, dann*). Etwas allgemeiner handelt es sich bei den Wörtern im Ausschnitt um solche mit einfachem Anfangsrand, einfachem Vokal und zweifach besetztem Endrand, und variiert wird jeweils (bis auf wenige Ausnahmen) der Anfangsrand oder der Kern.

Wörter, die nicht an Minimalpaaren teilnehmen, sind in der Abbildung oben nicht mit aufgenommen; sie würden am weitesten vom Zentrum dargestellt werden, weil sie maximal isoliert sind (sie sind ‚Eremiten‘, um das schöne Bild von Günther/Greese (1985) zu verwenden).

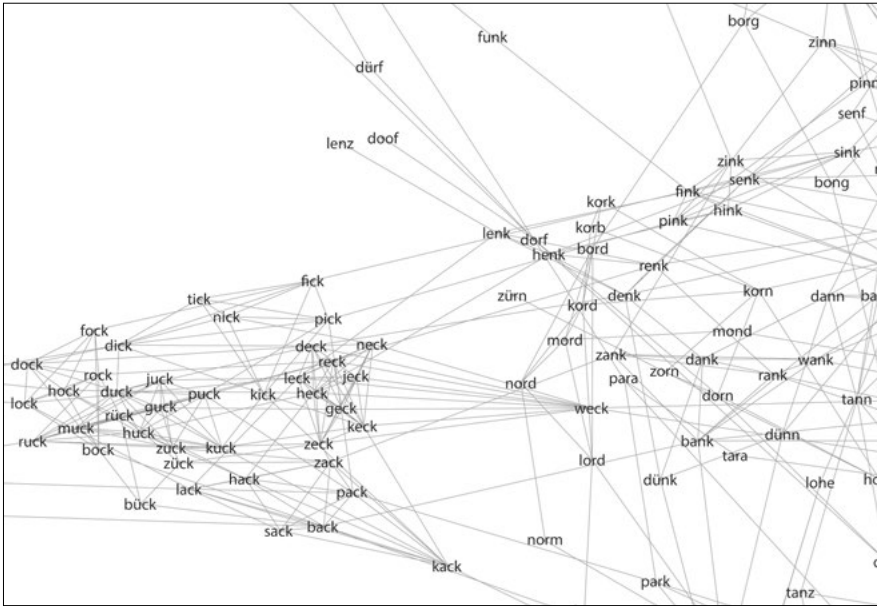


Abb. 39: Ausschnitt des Netzwerks der Minimalpaare mit vier Buchstaben im deutschen Korpus einfacher Stämme, visualisiert mit Gephi; verwendeter Algorithmus: Yifan Hu.

Man kann nun aber auch einen Schritt weiter gehen und Parameter suchen, mit denen dieses Netzwerk am besten beschrieben werden kann. Diesen Weg gehen Fushing et al. (2014) für das Englische.⁶² Sie vergleichen z. B. die Anzahl der Verbindungen für jedes Wort (den *Grad*) und stellen fest, dass Wörter mit vier Buchstaben, die mit einem Vokal beginnen, im Schnitt einen geringeren Grad haben als solche, die mit einem Konsonanten beginnen. Das ist nicht sonderlich überraschend – es gibt einfach absolut weniger vokal- als konsonant-initiale Wörter, und die Anzahl von Minimalpaaren scheint (zumindest teilweise) auf dieser absoluten Anzahl zu beruhen. In der Folge schlagen Fushing et al. (2014) aber einen Parameter vor (‘community geometry’), der geeignet scheint, automatisch linguistisch relevante Subgruppen im Netzwerk zu identifizieren. Die Anwen-

⁶² Fushing et al. (2014) bringen das Kunststück fertig, ein Netzwerk von sprachlichen Minimalpaaren zu untersuchen, ohne auch nur einmal das Konzept ‘sprachliches Minimalpaar’ zu nennen. Stattdessen beziehen sie sich auf Lewis Carrolls Spiel ‘Doublets’, bei dem der Spieler den ‘Weg’ zwischen zwei Wörtern finden muss, der aus der schrittweisen Ersetzung jeweils eines Buchstabens besteht (z. B. *love – hate: love – hove – have – hate*).

derung auf die deutschen und englischen Daten würde an dieser Stelle zu weit führen. Die systematische Analyse der Netzwerke von Minimalpaaren (graphematisch und phonologisch) ist eine eigene Untersuchung wert.

Kommen wir nun – wie einleitend versprochen – zum ‚anderen‘ Nutzen der Minimalpaarmethode, der Überprüfung des Graphemstatus von Verbindungen. Der Graphemstatus wird in den einschlägigen Definitionen oft an die Funktion gebunden, und als Funktion wird wiederum die potenzielle Bedeutungs-differenzierung angesehen (siehe oben 3.1). Auf diese Weise können die meisten Buchstaben als Grapheme etabliert werden: Aus Minimalpaaren wie *scheu/schau* ergibt sich dann, dass ⟨e⟩ und ⟨a⟩ Grapheme des Deutschen sind. Das funktioniert, wie oben gezeigt, für alle Buchstaben des Deutschen mit Ausnahme von |q| und |c| und für alle Buchstaben des Englischen mit Ausnahme von |q|. Besonders relevant wird das Verfahren nun im Zusammenhang mit den Kandidaten für komplexe Grapheme. Die Frage ist dann einerseits, ob die komplexe Einheit als ganze Teil eines Minimalpaares sein kann, ob sie sich also in gewisser Weise verhält wie einfache Grapheme. |ch| in *Teich* etwa kann durch |l| ersetzt werden (*Teil*).⁶³ Andererseits – und diese Frage ist interessanter – können wir fragen, ob alle Bestandteile des potenziellen Graphems durch Minimalpaare ersetzbar sind; ist das der Fall, verhalten sie sich eben wie kombinatorisch beschreibbare Teile und nicht wie eine Einheit (so z. B. Eisenberg 1988; Fuhrhop/Peters 2013: 204 f.). Wenn der Test positiv ausfällt, handelt es sich bei dem Kandidaten nicht um ein Graphem. Ein Beispiel ist ⟨sh⟩ im Englischen, das ja einem Phonem entspricht. Hier sind aber beide Bestandteile regelmäßig ersetzbar, vgl. z. B. die Paare *chore/shore/whore*, *shine/thine/whine*, *she/the* für den ersten Teil und *shack/slack/smack*, *share/snare/spare/stare*, *shell/smell/spell/swell* für den zweiten Teil. ⟨sh⟩ ist also nach diesem Test kein Graphem des Englischen.

Problematisch sind nun diejenigen potenziellen Grapheme, die zwar ersetzbar sind – aber nur in engen Grenzen. So ist bei z. B. |ck| im Deutschen der erste Teil ersetzbar (z. B. *leck[en]/lenk[en]*), *weck[en]/welk[en]*), der zweite Teil auch (z. B. *stech[en]/steck[en]*). Es finden sich aber nur Minimalpaare, in denen ⟨h⟩ mit ⟨k⟩ alterniert – und auch |ck| ist ja ein Kandidat für ein komplexes Graphem. Die Kombinatorik hat also hier enge Grenzen, das ist ja auch syntagmatisch gut zu sehen (siehe 3.2.2). Es liegt daher nahe, sowohl ⟨ch⟩ als auch ⟨ck⟩ im Deutschen als Grapheme zu klassifizieren. Für den Fall, dass beide Teile ersetzbar

⁶³ Hier ist allerdings etwas Vorsicht geboten, denn auf diese Weise kann auch gezeigt werden, dass ⟨badeh⟩ ein Graphem ist: Es ‚alterniert‘ mit ⟨r⟩ im Minimalpaar *Badehose/Rose*. Es handelt sich hier also nicht um eine hinreichende Bedingung für den Graphemstatus.

sind, muss also noch einmal geprüft werden, inwiefern die Ersetzbarkeit eingeschränkt ist.

Im Folgenden werden diejenigen Verbindungen zusammengetragen, die sich im Verlauf der Arbeit für das Deutsche und das Englische ergeben haben. Für jede dieser Verbindungen wird geprüft, a) wie viele Minimalpaare es für den ersten und zweiten Teil gibt (das sind die Tokens) und b) mit wie vielen unterschiedlichen einfachen Graphemen die Elemente jeweils ersetzbar sind (das sind die Types). Mit der Ausnahme von |qu| sind alle Kandidaten Verbindungen aus Konsonantenbuchstaben nach der Klassifikation in 3.2.1.1 oben (deutsch |sch| wird als Verbindung von <s> und <ch> analysiert), und es werden jeweils nur Ersetzungen mit anderen Konsonanten gewertet. Tabelle 86 zeigt die Ergebnisse für die neun deutschen Graphemkandidaten:

Tab. 86: Test auf beidseitige Ersetzbarkeit für neun Graphemkandidaten des Deutschen. Anzahl der Minimalpaare, an denen ein Teil des Graphemkandidaten teilnimmt (, #Token⁴), sowie Anzahl der Buchstaben, mit denen der Teil des potenziellen Graphems alterniert (, #Types⁴). Abschließende Bewertung des Graphemstatus aufgrund des Minimalpaarkriteriums.

Verbindung	#Token MP 1. Teil	Types MP 1. Teil	#Token MP 2. Teil	Types MP 2. Teil	Graphem?
ch	0	0	18	1	+
ck	27	4	18	1	(+)
pf	4	2	7	3	(+)
ph	1	1	0	0	+
qu	0	0	0	0	+
rh	1	1	0	0	+
sch	3	3	16	4	o
st	52	9	33	4	-
th	2	2	1	1	(+)

In der letzten Spalte wird zusammenfassend der Graphemstatus der Verbindung nach dem Minimalpaarkriterium bewertet. Diese Zuschreibung soll im Folgenden kurz begründet werden:

- |ch|: Der erste Teil dieser Verbindung ist tatsächlich nicht ersetzbar. Der Test spricht nicht gegen den Graphemstatus der Verbindung.
- |ck|: Hier sieht es etwas anders aus. Der erste Teil ist ersetzbar, der zweite auch – aber eben nur durch einen Type, nämlich <h>. Das spricht nicht uneingeschränkt gegen den Graphemstatus der Verbindung – daher die vorsichtige Bewertung ,(+)‘.

- |pf|: Es gibt im Korpus Beispiele für die Ersetzbarkeit des ersten Teils, und zwar mit ⟨r⟩ (*schlüp[en]/schlür[en]*, *Topf/Torf*) und mit ⟨f⟩ (*stopp/stopf[en]*, *tupf[en]/Tuff*). Alle Ersetzungen finden aber im Endrand des Stamms statt; im Anfangsrand ist der erste Teil also tatsächlich nicht ersetzbar, wie Eisenberg (1988: 144) feststellt. Im Endrand hingegen trifft das nicht zu, wenn es auch nur wenige Minimalpaare gibt. Daher die Bewertung mit ‚(+)'.
- |ph| und |rh|: Hier ist der zweite Teil nicht ersetzbar; die Minimalpaarmethode spricht also nicht gegen den Graphemstatus. Beide Verbindungen sind allerdings sehr selten im Korpus.
- |qu|: Hier ist kein Teil ersetzbar; *Duell* ist nicht im Teilkorpus (es hat den Status ‚I' für ‚irrelevant''), ansonsten gäbe es zumindest einen Treffer für das oben angeführte Beispiel *Quell/Duell*.
- |sch|: Hier ist der erste Teil nur bedingt ersetzbar. Es gibt drei Minimalpaare, bei denen |s| mit einem Konsonanten alterniert (*Asche/Arche*, *misch[en]/Milch*, *rasch/Ranch*), von denen das letzte ein Fremdwort beinhaltet. Das ist nur bedingt ein Argument gegen den Graphemstatus von |sch|, daher die neutrale Bewertung.
- |st|: Hier können eindeutig beide Teile häufig ersetzt werden; dieses Kriterium spricht also gegen den Graphemstatus von |st|.
- |th|: Bei dieser Verbindung sind beide Teile ersetzbar (*Thein/Rhein*, *Thon/Phon*; *Thema/Trema*), es gibt aber insgesamt nur drei Minimalpaare. Bei der Ersetzung des ersten Teils treten als Minimalpaare nur Wörter mit anderen Graphemkandidaten auf (|rh|, |ph|). Das Minimalpaarkriterium spricht also nur eingeschränkt gegen den Graphemstatus von |th|, daher erscheint in Tabelle 86 ein vorsichtiges ‚(+)'.

Insgesamt kann auf diese Weise nur eine einzige Verbindung als eben das klassifiziert werden – nämlich als Verbindung und nicht als Einheit; das ist |st|. Bei allen anderen Kombinationen ist entweder ein Teil eindeutig nicht ersetzbar oder nur bedingt ersetzbar.

Im Englischen wird neben den potenziellen komplexen Graphemen, die sich bislang ergeben haben, auch ⟨wh⟩ untersucht, das auf phonographischer Basis häufig als Einheit klassifiziert wird (vgl. z. B. Carney 1994). Die Überprüfung des Minimalpaarkriteriums bringt folgende Ergebnisse:

Tab. 87: Test auf beidseitige Ersetzbarkeit für neun Graphemkandidaten des Englischen. Anzahl der Minimalpaare, an denen ein Teil des Graphemkandidaten teilnimmt (*#Token*'), sowie Anzahl der Buchstaben, mit denen der Teil des potenziellen Graphems alterniert (*#Types*'). Abschließende Bewertung des Graphemstatus aufgrund des Minimalpaarkriteriums.

Verbindung	#Token MP	Types MP	#Token MP	Types MP	Graphem?
	1. Teil	1. Teil	2. Teil	2. Teil	
ch	45	6	41	7	-
ck	74	5	9	2	-
gh	10	4	3	2	(-)
ph	3	3	0	0	+
qu	5	3	0	0	+
rh	1	1	0	0	+
sc	0	0	73	8	+
sh	44	6	162	9	-
sp	9	4	100	9	-
st	48	9	48	9	-
th	46	6	14	4	-
wh	31	4	6	1	(-)

Wie sind die Zuschreibungen in der letzten Spalte jeweils begründet?

- |ch|: Beide Teile können mit einer Reihe von Graphemen ausgetauscht werden (z. B. *chin/shin/thin/whin*, *ache/acme/acne/acre*); das spricht gegen den Graphemstatus der Verbindung.
- |ck|: Dasselbe gilt auch für |ck|, auch wenn der zweite Teil nur gegen zwei Grapheme ausgetauscht werden kann, ⟨h⟩ (*attack/attach*) und ⟨t⟩ (*pack/pact*).
- |gh|: Der zweite Teil kann nur in drei Minimalpaaren ersetzt werden, und zwar mit ⟨n⟩ (*ghat/gnat, sigh/sign*) und mit ⟨l⟩ (*ghee/glee*). Bei ⟨n⟩ und ⟨l⟩ handelt es sich allerdings um zwei frei kombinierende Grapheme, d. h. ⟨gn⟩ und ⟨gl⟩ sind selbst keine Graphemkandidaten. Daher reichen diese wenigen Belege aus, um ⟨gh⟩ (zumindest vorsichtig) den Graphemstatus abzusprechen (,-').
- |ph| und |rh|: Wie im deutschen Korpus ist auch hier der zweite Teil nicht ersetzbar; das Minimalpaarkriterium spricht also nicht gegen den Graphemstatus der Verbindungen.
- |qu|: Hier gibt es für den ersten Bestandteil mehr Minimalpaare als im Deutschen (*quest/guest, quilt/guilt, quin/ruin, quit/suit, quite/suite*). Da der zweite Bestandteil aber nicht ersetzbar ist, spricht das Kriterium insgesamt nicht gegen den Graphemstatus.
- |sh|, |st| und |th|: Hier können jeweils beide Teile ersetzt werden, das spricht gegen den Graphemstatus der Verbindungen.

- |wh|: Der zweite Teil ist nur mit einem anderen Graphem ersetzbar, nämlich ⟨r⟩ (*white/write*). Kombinationen mit ⟨r⟩ im Anfangsrand sind allerdings unauffällig (vgl. oben 3.2.2.2), es handelt sich bei |wr| also nicht um einen Graphemkandidaten. Der Fall ist damit nicht analog zu |ch|/|ck| im Deutschen. Das Minimalpaarkriterium spricht tendenziell gegen den Graphemstatus, daher die vorsichtige Bewertung ‚(-)‘.

Anders als im Deutschen erfüllen im Englischen wesentlich mehr potenzielle komplexe Grapheme das Minimalpaarkriterium nicht. Nur |qu|, |rh| und |ph| können nach dem Kriterium als Grapheme gelten.

3.2.5 Zusammenfassung

In diesem Abschnitt ging es um den graphemischen Aufbau von Stämmen im Deutschen und Englischen. Dieser Aufbau wurde aus verschiedenen Perspektiven beleuchtet:

- Zum einen wurden die globalen graphotaktischen Regularitäten untersucht (3.2.1). Welche Buchstaben sind Konsonanten, welche Vokale? Welche Buchstaben kommen häufig miteinander vor, welche selten? Welche Buchstaben treten häufig an den Stammrändern auf, welche selten? Für eine Zusammenfassung der Ergebnisse siehe 3.2.1.6.
- Zum anderen wurde der graphotaktische Aufbau aus silbischer Perspektive betrachtet. Aus wie vielen Silben bestehen Stämme? Wie sind die Silbenkonstituenten aufgebaut, und wie interagiert ihre Besetzung? Eine Zusammenfassung der hauptsächlichen Ergebnisse findet sich in Abschnitt 3.2.2.8.
- Wir haben außerdem untersucht, wie minimale Stämme im Deutschen und Englischen aussehen müssen und wie prototypische (einsilbige) Stämme in beiden Sprachen aufgebaut sind. Beides wird weiter unten im Zusammenhang mit minimalen und prototypischen Affixen noch weiter diskutiert (Abschn. 3.4).
- Und schließlich wurde die Verteilung von Minimalpaaren und damit zusammenhängend von funktionaler Last untersucht (3.2.4). Minimalpaare sind im Englischen und Deutschen am häufigsten in einfach besetzten Konstituenten zu finden, und hier vor allem im einfachen Anfangsrand; diese Konstituente ist am stärksten funktional belastet. Die Minimalpaarmethode wurde außerdem als Kriterium zur Graphemdefinition verwendet; dazu in ein paar Absätzen mehr.

Im Folgenden sollen zwei Aspekte diskutiert werden, die sich durch das Teilkapitel 3.2 gezogen haben. Es geht um die Rolle der Segmentierung in Wurzel

und Endung sowie um eine Zusammenschau der verschiedenen Kriterien zur Graphemdefinition.

Zunächst zur Segmentierung in Wurzel und Endung. Für zweisilbige einfache Stämme im Deutschen und Englischen ergibt sich aus mehreren Perspektiven eine Gliederung in Wurzel und Endung: Die Endung umfasst den Reim der zweiten Silbe, die Wurzel ist das, was übrigbleibt. Im Deutschen geht es bspw. um Segmentierungen wie *Hamm+er*, *Ros+e*, *brutz+el[n]*, *Harn+isch*, *Atl+as*, *Fakt+or*; im Englischen z. B. um *cas+e*, *patt+ern*, *poll+en*, *dogm+a*, *tenn+is*. Die Gliederung unterscheidet sich leicht von Tafts (1979) „basic orthographic syllable structure“ (BOSS) und verwandten Konzepten: In Tafts Konzept spielt die Graphotaktik der intervokalischen Konsonanten eine Rolle, und Konsonanten, die sich aufgrund der graphotaktischen Regeln nicht mehr in den Endrand integrieren lassen, gehören nicht dazu. Im Deutschen wäre die BOSS von *Atlas* bspw. ⟨At⟩, die Wurzel, wie sie in dieser Arbeit verwendet wird, aber ⟨Atl⟩. Insgesamt führen beide Konzepte in den meisten Fällen zu isomorphen Gliederungen.

Die Formen haben in vielen Fällen denselben Aufbau wie Kombinationen aus Stamm und Suffix (z. B. ⟨lauf-en⟩, ⟨Spring-er⟩). Das gilt auch für die im fremden Bereich häufigen gebundenen Stämme (bzw. die Vergleichssegmentform bei Fuhrhop 1998: 95 ff.), vgl. z. B. ⟨demonstr-ieren⟩, ⟨Demonstr-ant⟩): „Vor Vokalen liegen potenzielle Segmentgrenzen“, wie Fuhrhop (1998: 136) feststellt.

Für eine Segmentierung von Stämmen in Wurzel und Endung spricht Folgendes:

- Zum einen gelten die Minimalitätsbedingungen nicht nur für Stämme, sondern (zumindest im Deutschen) auch für Wurzeln (3.2.3). Man kann nicht nur die minimal notwendigen Strukturen von Wörtern damit erfassen, wie Ramers (1998) und Evertz (2014) das tun, sondern auch die minimal notwendigen Strukturen von Teilen von Wörtern und Stämmen. Auch Wurzeln müssen über einen zweifach besetzten Reim verfügen, und dieser Reim muss mit zwei Gewichtseinheiten assoziiert sein. Auf diese Weise kann auch die ⟨h⟩-Schreibung im Deutschen fast komplett erfasst werden: ⟨h⟩ stellt in den meisten Fällen sicher, dass die Mindestbedingung vom zweifach besetzten Reim erfüllt ist (vgl. *Reh*, *Ruh+e*, *Sahn+e*).
- Zum anderen verhalten sich Wurzeln im Deutschen so wie tatsächliche Stämme, was die quantitative Verteilung der CV-Strukturen sowie die prototypische Struktur angeht (das gilt erst einmal nur für Wurzeln, die mit ⟨e⟩, ⟨el⟩, ⟨em⟩, ⟨en⟩ oder ⟨er⟩ kombinieren). In beiden Fällen ist die bevorzugte Struktur ⟨CVCC⟩.
- Zweifach besetzte intervokalische Konsonantencluster (z. B. in *Hunger*) verhalten sich überwiegend wie Endränder. Das gilt im Englischen, wo etwa drei Viertel aller intervokalischen Konsonantencluster so interpretiert werden

können; es gilt noch mehr im Deutschen, wo der Wert bei 90 % liegt. Das ist ein graphotaktisches Argument für die vorgeschlagene Segmentierung: ‹Hunger› sieht strukturell aus wie ‹Sänger›, und zwar nicht nur wegen der homographischen Endung (3.2.2.5).

- Ein weiteres graphotaktisches Argument ist, dass der Zusammenhang zwischen Kern und Endrand in zweiten Silben von Zweisilbern enger ist als bspw. in Einsilbern (3.2.2.6, 3.2.2.7). Zwischen Silbenkern und Endrand in zweiten Silben von Zweisilbern – also in der Endung – gibt es weit weniger Variation als im Reim von Einsilbern. Wenn der Silbenkern feststeht, ist die Menge der möglichen Endränder am Stammende stark eingeschränkt. Daran ändert sich nichts Grundsätzliches, wenn die Endungen mit ‹e› im Kern ausgeklammert werden.
- Und schließlich hat dieser engere Zusammenhang auch einen paradigmatischen Effekt: Es gibt bspw. weniger Minimalpaare im Silbenkern von zweiten Silben von Zweisilbern als im Silbenkern von ersten Silben (oder in Einsilbern). Das deutet ebenfalls darauf hin, dass Endungen (also Silbenkern und Endrand in zweiten Silben von Zweisilbern) eine engere Verbindung eingehen als andere Konstituenten.

Zusätzliche Argumente für eine Gliederung in Wurzel und Endung liefern Abkürzungen. Wie Buchmann (2015: 148 f.) zeigt, werden Wörter besonders häufig vor dem zweiten Silbenkern abgebrochen, wenn sie abgekürzt werden: ‹April› wird zu ‹Apr.›, ‹Tafel› wird zu ‹Taf.›, ‹Plural› wird zu ‹Plur.›. Die Abkürzungen zweisilbiger Wörter entsprechen also der Wurzel, die Kürzungen der Endung.

Die Gliederung in Wurzel und Endung ist also graphotaktisch gut motiviert. Endungen verhalten sich darüber hinaus im Englischen suffixähnlich: Sie haben eine charakteristische Verteilung hinsichtlich der Konsonantenverdoppelung (vgl. Berg 2016 und unten 4.1). Bestimmte Endungen wie ‹it› treten nur selten nach Doppelkonsonanten auf (vgl. *habit*, *vomit*, *limit*), andere fast ausschließlich (z. B. ‹ock› in *hammock*, *haddock*). Auch das spricht für die vorgeschlagene Gliederung.

Daraus ergibt sich eine interessante psycholinguistische Frage. Zahlreiche Studien belegen das sog. ‚affix stripping‘ (Taft/Forster 1975) bzw. die ‚morpho-orthographic decomposition‘ (z. B. Rastle/Davis/New 2004). Damit ist die Beobachtung gemeint, dass Leser auch morphologisch einfache Wörter wie *corner* in zwei Teile *corn+er* segmentieren. Die Studien gehen davon aus, dass es sich hier um einen morphologischen Effekt handelt. Es ist aber ebenso gut möglich, dass der Effekt rein graphemischer bzw. lesepsychologischer Natur ist (vgl. Aronoff/Berg/Heyer 2016).

Kommen wir abschließend zur Zusammenschau und Diskussion des Graphembegriffs. Am Anfang des Kapitels wurde in (5) eine gut zu operationalisierende Minimaldefinition vorgestellt. Im Laufe des Kapitels haben sich aber an verschiedenen Stellen Hinweise auf weitere Kriterien ergeben. Konkret handelt es sich um folgende drei Kriterien:

- Aus der Betrachtung der Graphotaktik folgt für einige Verbindungen, dass sie sich verhalten wie einfache Grapheme, für andere ist das nicht zwangsläufig der Fall. In den entsprechenden Teilkapiteln (3.2.2.2, 3.2.2.3) wurde teilweise auch nach Ökonomiegesichtspunkten argumentiert: Ein kleineres Inventar ist besser als ein umfangreicheres. Das muss bei den Bewertungen bedacht werden: In den meisten Fällen ergibt sich aus der Graphotaktik nicht unmittelbar der Graphemstatus.
- Ein weiteres Kriterium ist die Symmetrie der potenziellen Grapheme im Anfangs- und im Endrand (3.2.2.4). Wenn zwei oder mehr Buchstaben im Anfangs- und im Endrand dieselbe (und nicht die spiegelbildliche) Abfolge aufweisen, spricht das für ihren Graphemstatus. Das ist in der Graphematik ein neues Kriterium.
- Und schließlich ist die Minimalpaarbildung der einzelnen Bestandteile ein Kriterium (3.2.4). Wenn beide Bestandteile von Verbindungen einzeln in Minimalpaaren ersetzbar sind, dann handelt es sich nicht um ein Graphem, sondern um eine Verbindung frei kombinierender Grapheme (siehe oben Tab. 86 und 87).

Tabelle 88 zeigt für das Deutsche, welche Kriterien zu welchen Ergebnissen führen. Das graphotaktische Kriterium ist hinsichtlich der Konstituente (Anfangs- bzw. Endrand) spezifiziert. Ergänzt werden diese Kriterien um die minimale Graphemdefinition (5) oben und um das phonographische Kriterium, das besagt: Ein Graphem ist, was einem Phonem in der gesprochenen Sprache entspricht.

Tab. 88: Potenzielle Grapheme im Deutschen und ihre Bewertung nach sechs Kriterien. AR = Anfangsrand, ER = Endrand, n. a. = nicht anwendbar.

Verbindung	ch	ck	qu	sch	rh	ph	th	st	pf
minimale Definition	+	+	+	-	-	-	-	-	-
Phonographie	+	+	-	+	+	+	+	-	?
Graphotaktik AR				+	+	-	-	-	-
Graphotaktik ER				+	+	+	+	-	-
Symmetrie	+	n. a.	n. a.	+	+	+	+	+	+
Minimalpaare	+	(+)	+	o	+	+	(+)	-	(+)

Nach der minimalen Definition sind nur <ch>, <ck> und <qu> Grapheme im Deutschen – und <ch> und <ck> auch nur im nativen Bereich des Wortschatzes. Phonographisch ergeben sich sehr viel mehr Grapheme; <qu> und <st> sind hier keine, weil sie jeweils zwei Phonemen entsprechen, und der Status von <pf> hängt am Phonemstatus von Affrikaten (deswegen erscheint in Tab. 88 ein Fragezeichen). Bei der Untersuchung der Graphotaktik wurde vom Grapheminventar nach der minimalen Definition ausgegangen; der Status von <ch>, <ck> und <qu> wurde also vorausgesetzt und nicht geprüft. Daher sind die entsprechenden Zellen in Tabelle 88 leer. Im Anfangsrand ergeben sich die Grapheme <sch> und <rh>, im Endrand darüber hinaus auch noch <ph> und <th>. Was die Symmetrie angeht, so tritt <ck> nicht im Anfangsrand und <qu> nicht im Endrand auf; das Kriterium ist also nicht anwendbar. Alle übrigen Kandidaten erfüllen das Kriterium; es geht damit über das phonographische Kriterium hinaus. Hinsichtlich des letzten Kriteriums der Minimalpaarbildung (also der beidseitigen Ersetzbarkeit) ist <st> sicher kein Graphem; der Status von <sch> ist unklar. Der Rest der Kandidaten verhält sich (teilweise mit leichten Abstrichen) wie einfache Grapheme.

Insgesamt ergibt sich: Gegen die minimal angesetzten komplexen Grapheme <ch>, <ck> und <qu> spricht nichts. Darüber hinaus spricht einiges (in absteigender Dringlichkeit) für <rh>, <sch>, <ph> und <th>. Für <st> spricht einzig das Kriterium der Symmetrie; für <pf> darüber hinaus das Kriterium der beidseitigen Ersetzbarkeit in Minimalpaaren. Das heißt: Graphemisch gibt es Argumente für die Annahme von genau denjenigen komplexen Graphemen, die auch phonographisch motiviert werden können. Die Ausnahme ist <qu>, das phonographisch kein Graphem ist. Etwas überspitzt könnte man sagen: Je konsequenter wir eine autonome silbenstrukturell-distributionelle Graphotaktik anstreben, desto phonographischer wird sie – zumindest, was die Klassifikation als Grapheme angeht.

Die Daten zum Graphemstatus im Englischen sind in Tabelle 89 zusammengetragen; da <qu> nach der minimalen Definition als Graphem angesetzt wurde, ist es graphotaktisch nicht geprüft worden; die entsprechenden Zellen in Tabelle 89 (nächste Seite) bleiben daher leer.

Nach der minimalen Definition ergibt sich nur <qu> als komplexes Graphem; phonographisch hingegen eine ganze Reihe (<ck>, <gh>, <ph>, <rh>, <sh>, <th> und <wh>). Der Status von <ch> hängt vom Phonemstatus der korrespondierenden Affrikate [tʃ] ab. Graphotaktisch wurden im Anfangsrand <rh> und <wh> angenommen, im Endrand <ch>, <gh> und <wh>. Das Symmetriekriterium wird von allen Verbindungen erfüllt, die auch in beiden Konstituenten vorkommen. Das Minimalpaarkriterium erfüllen nur <ph>, <qu> und <rh>.

Tab. 89: Potenzielle Grapheme im Englischen und ihre Bewertung nach sechs Kriterien.
AR = Anfangsrand, ER = Endrand, n. a. = nicht anwendbar.

Verbindung	ch	ck	gh	ph	qu	rh	sh	sc	sp	st	th	wh
Minimalitätsbedingung	-	-	-	-	+	-	-	-	-	-	-	-
Phonographie	?	+	+	+	-	+	+	-	-	-	+	+
Graphotaktik AR	-	n. a.	-	-		+	-	-	-	-	-	+
Graphotaktik ER	+	-	+	-		+	-	-	-	-	-	n. a.
Symmetrie	+	n. a.	+	+	n. a.	+	+	+	+	+	+	n. a.
Minimalpaare	-	-	(-)	+	+	+	-	+	-	-	-	(-)

Damit spricht im Englischen nichts gegen $\langle qu \rangle$ und einiges für die Annahme von $\langle rh \rangle$ und (weniger) für $\langle ch \rangle$ und $\langle gh \rangle$. Im Englischen sind also phonographisch und graphemisch definierte Grapheme nicht isomorph. Das liegt vor allem daran, dass $\langle h \rangle$ als zweites Element mit einer ganzen Reihe von Graphemen auftritt. Auch wenn diese Kombinationen mit Phonemen (und nicht mit Phonemfolgen) korrespondieren, so ist in der Graphematik die (relativ) freie Kombinatorik entscheidend und ausschlaggebend für den Status als Graphemverbindung.

Im Englischen und im Deutschen führt das Kriterium der Symmetrie zum jeweils größten Inventar komplexer Grapheme. Das entwertet das Kriterium potenziell etwas: Es beruht ja gerade auf der Beobachtung, dass Kombinationen im Anfangs- und Endrand symmetrisch und nicht identisch sind. Offenbar gibt es in beiden Sprachen aber eine ganze Reihe identischer Abfolgen.

Es gibt allerdings ein Argument, das für ein solch vergrößertes Inventar spricht. Wenn wir im Deutschen neben $\langle ch \rangle$, $\langle ck \rangle$ und $\langle qu \rangle$ auch $\langle sch \rangle$, $\langle rh \rangle$, $\langle ph \rangle$, $\langle th \rangle$, $\langle st \rangle$ und $\langle pf \rangle$ als Grapheme ansetzen, lässt sich die graphotaktische Beschreibung des Anfangsrandes wesentlich vereinfachen. Der Anfangsrand ist dann fast ausschließlich zweifach besetzt. Die Ausnahmen sind $\langle spr \rangle$, $\langle spl \rangle$, $\langle skr \rangle$ und $\langle skl \rangle$ (die letzten beiden kommen ausschließlich in Fremdwörtern vor).

Im Englischen gilt dasselbe: Wenn wir $\langle ch \rangle$, $\langle gh \rangle$, $\langle ph \rangle$, $\langle rh \rangle$, $\langle sh \rangle$, $\langle sc \rangle$, $\langle sp \rangle$, $\langle st \rangle$ und $\langle th \rangle$ als Grapheme ansetzen, ist der Anfangsrand ebenfalls maximal zweifach besetzt. Hier gibt es sogar nur zwei Ausnahmen, $\langle tsw \rangle$ und $\langle khm \rangle$, die nur in jeweils einem Fremdwort auftreten (*Tswana*, *Khmer*). Das Kriterium der Symmetrie führt also in beiden Sprachen zu Inventaren, die die graphotaktische Beschreibung des Anfangsrandes deutlich vereinfachen.

3.3 Affixe

3.3.1 Der silbische Bau von Affixen

Aus wie vielen graphematischen Silben bestehen Affixe im Deutschen und Englischen und wie sind diese Silben strukturell aufgebaut? Zur Beantwortung dieser Fragen werden die Listen der Affixe benutzt, die oben (2.3) vorgestellt wurden. Wie oben schon erwähnt, sind die Ergebnisse mit mehr Vorsicht zu genießen als die Ergebnisse für die einfachen Stämme, weil einzelne (auch eingeschränkt kombinierende) Affixe die relativ kleinen Listen relativ stark beeinflussen können.

Abbildung 40 zeigt die Anzahl der Silben, aus denen deutsche und englische Affixe im Korpus bestehen:

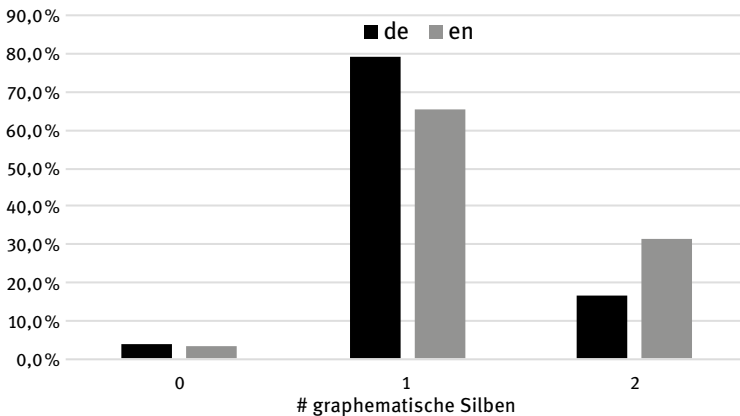


Abb. 40: Relativer Anteil der Affixe mit 0, 1 und 2 graphematischen Silben an allen Stämmen. Datengrundlage: Korpus deutscher bzw. englischer Affixe.

Affixe ohne Vokalgraphem – also nichtsilbische Affixe – sind in beiden Sprachen die Ausnahme, und in beiden Sprachen sind es ausschließlich Suffixe. Im Deutschen handelt es sich um die vier Suffixe *-n*, *-s*, *-st* und *-t*. Sie alle können als Flexionssuffixe und mit Ausnahme von *-st* auch als Derivationsuffixe auftreten. Die Flexion ist sicher der Kernbereich dieser Formen; als Derivationsuffixe sind sie wesentlich seltener. Die Form *-n* ist eine Variante des Adjektivsuffixes *-en/-ern* und tritt auf, wenn das Basissubstantiv auf *-er* endet (z. B. *kupfern*, *ledern*; Fleischer/Barz 2012: 335 f.). Die Form *-s* tritt auf als Substantivsuffix in (meist umgangssprachlichen) Bildungen wie *Klecks* und *Plumps* (Fleischer/Barz 2012: 220) und als Adverbsuffix in Bildungen wie *mittags* oder *nachts* (Fleischer/Barz

2012: 369). Beide Suffixe, *-n* und *-s*, werden von Fleischer/Barz (2012) vorsichtig als „schwach produktiv“ beschrieben. Das gilt für die Form *-t* nicht; es handelt sich um ein unproduktives Suffix (z. B. in *Fahrt*, *Naht* und weniger deutlich in *Schlacht*; vgl. Fleischer/Barz 2012: 254).

Im Englischen sind nur drei Suffixe nicht-silbisch, *-s*, *-’s* und *-th*. Von diesen tritt *-’s* tritt nur als Flexionssuffix auf und *-s* ganz überwiegend; als Adverbsuffix ist es unproduktiv (z. B. *overseas*, *sometimes*, vgl. Bauer/Lieber/Plag 2013: 328).⁶⁴ Einzig *-th* ist hier ein ‚reines‘ Derivationsuffix, das zwar reihenbildend ist (vgl. z. B. *warmth*, *length*, *depth*), aber schon lange nicht mehr produktiv (vgl. Marchand 1969: 349). Es kommt im CELEX-Korpus auf insgesamt 47 Treffer, von denen allerdings 30 Einträge Ableitungen von Ordinal- aus Kardinalzahlwörtern sind (z. B. *ninth*, *tenth*, *eleventh*, vgl. Bauer/Lieber/Plag 2013: 425 f.). Die Hauptdomäne nichtsilbischer Affixe, so können wir festhalten, ist im Deutschen wie im Englischen die Flexion. Im Gesamtsystem der Derivation sind diese Formen die Ausnahme.

Einsilbige Affixe sind in beiden Sprachen der Standardfall (siehe Abb. 40); im Deutschen sind sie noch etwas häufiger als im Englischen. Zweisilbige Affixe sind umgekehrt im Englischen etwas häufiger. Das liegt, wie wir gleich sehen werden, an der Rolle, die finales ⟨e⟩ im Englischen spielt.

Schauen wir uns die präferierten CV-Strukturen für die größte Gruppe, die Einsilber, an. Hier ist in beiden Sprachen jeweils dasselbe Schema am häufigsten:

Tab. 90: Die jeweils fünf häufigsten CV-Strukturen einisilbiger Affixe im deutschen und englischen Korpus. Die Linien zwischen den Teiltabellen zeigen gleiche Strukturen.

deutsch		englisch	
VC	24	VC	16
CVC	12	VCC	8
VCC	12	VVC	7
CV	5	CVC	7
CVCC	5	CV	5

⁶⁴ Zur marginalen Funktion als Diminutivsuffix (*moms*, *Wimblers* [für *Wimbledon*]) vgl. Bauer/Lieber/Plag (2013: 394).

Kombinationen aus einem Vokal- und einem Konsonantenbuchstaben stellen im Deutschen und im Englischen mit einigem Abstand die frequenteste Struktur (vgl. z. B. deutsch *un-*, *-er*, *-en*, *-in*; englisch *un-*, *-er*, *-al*, *-ic*). Bei den einfachen Stämmen tritt diese Struktur mit sehr wenigen Ausnahmen nur bei Funktionswörtern auf (siehe oben 3.2.3). Dasselbe gilt für die umgekehrte Struktur $\langle CV \rangle$ (wie in deutsch *be-*, *ge-* und englisch *be-*, *-ly*). Auch diese Struktur ist unter Inhaltswörtern die Ausnahme – sie verletzt dort neben der Mindestlänge von drei Graphemen auch die Bedingung, dass der Reim mit zwei Graphemen besetzt ist. Die fünf häufigsten Strukturen in einsilbigen Stämmen sind hier noch einmal wiederholt (Ausschnitt aus Tab. 18 oben):

Tab. 91: Die fünf häufigsten CV-Strukturen von Einsilbern im deutschen und englischen Teilkorpus einfacher Stämme. Die Linien zwischen den Teiltabellen zeigen gleiche Strukturen.

deutsch		englisch	
CVCC	577	CVCC	613
CCVCC	232	CCVCC	370
CVC	173	CVC	349
CVVC	171	CVVC	291
CCVC	117	CCVC	261

Was können wir feststellen? Typische einsilbige Affixe sind kürzer als typische einsilbige Stämme. Das betrifft an dieser Stelle nicht die Minimalitätsbeschränkungen. Stämme der Form $\langle VCC \rangle$ sind möglich, aber sie sind – genau wie Stämme der Form $\langle CVC \rangle$ – vergleichsweise selten, während das für Affixe nicht gilt. Diese Idee wird in Abschnitt 3.3.5 weiter diskutiert, wenn es um prototypische Affixe geht.

Zweisilbige Affixe sind im englischen Korpus deutlich häufiger als im deutschen (29 vs. 16). Hier ist eine interessante Verteilung zu beobachten: Die einfachste zweisilbige Struktur, $\langle VCV \rangle$, ist im englischen Korpus gleichzeitig die häufigste (13x), und sie wird exklusiv durch Suffixe vertreten (z. B. *-age*, *-ate*, *-ity*, *-ify*, *-ory*). Eine vergleichbare Verteilung findet sich im deutschen Korpus nicht.

3.3.2 Anfangsränder

Der Anfangsrand von Affixen ist in den meisten Fällen unbesetzt, sowohl im Deutschen wie auch im Englischen. Abbildung 41 zeigt die relative Verteilung der Grapheme im Anfangsrand von Affixen:

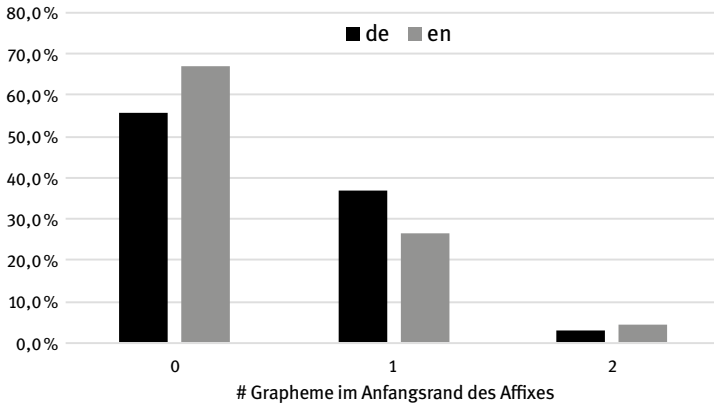


Abb. 41: Relativer Anteil der Affixe mit 0, 1 und 2 Graphemen im Anfangsrand an allen Affixen. Datengrundlage: Korpus deutscher bzw. englischer Affixe.

Unbesetzte Anfangsränder sind im Englischen etwas häufiger als im Deutschen; bei den einfach besetzten Anfangsrändern ist es umgekehrt. Zweifach besetzte Anfangsränder sind marginal.

In beiden Sprachen gibt es die Tendenz, dass unter Affixen mit unbesetzten Anfangsrändern (wie *-er*, *-ig* oder *un-*) häufiger Suffixe zu finden sind als unter Affixen mit besetztem Anfangsrand – oder andersherum, dass Suffixe häufiger unbesetzte Anfangsränder haben als Präfixe. Tabelle 92 und Tabelle 93 zeigen die Kreuzklassifikation Präfix/Suffix vs. \pm Anfangsrand für beide Sprachen:

Tab. 92 (links) und **Tab. 93 (rechts):** Anzahl von Affixen mit besetztem und unbesetztem Anfangsrand, klassifiziert nach der Art des Affixes (Präfix/Suffix) im Deutschen (Tab. 92) und Englischen (Tab. 93). Datengrundlage: Korpus deutscher bzw. englischer Affixe.

	Präfix	Suffix	% Suffix		Präfix	Suffix	% Suffix
-Anfangsrand	15	39	72 %	-Anfangsrand	14	47	77 %
+Anfangsrand	14	23	62 %	+Anfangsrand	12	16	57 %

Diese Tendenz (die in den englischen Daten etwas ausgeprägter ist) ist allerdings statistisch nicht signifikant. Ein zweiseitiger Chi-Quadrat-Test ergibt für das Deutsche $\chi^2 = 1,0234$, $p > 0,3$ und für das Englische $\chi^2 = 3,6774$, $p > 0,05$.

Da es in beiden Sprachen nur relativ wenige besetzte Anfangsränder gibt (35 im Deutschen und 24 im Englischen), wird auf eine detaillierte Untersuchung der verwendeten Grapheme hier verzichtet. Es sei nur so viel festgehalten: Der häufigste Anfangsrand ist in beiden Sprachen ⟨b⟩ (jeweils 5x im deutschen und englischen Korpus), und ⟨b⟩ kommt gleichzeitig nur als Anfangsrand von Suffixen vor.

3.3.3 Endränder

Endränder von Affixen sind in den meisten Fällen besetzt, und zwar mit einem Graphem, wie Abbildung 42 zeigt:

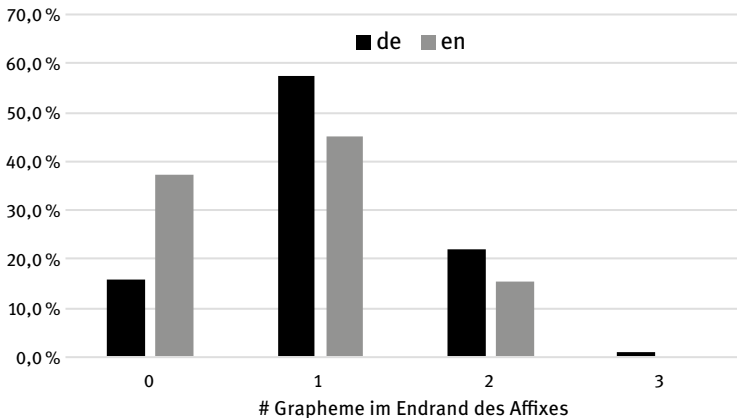


Abb. 42: Relativer Anteil der Affixe mit 0, 1, 2 und 3 Graphemen im Endrand an allen Affixen. Datengrundlage: Korpus deutscher bzw. englischer Affixe.

Hier gibt es allerdings relativ große Unterschiede zwischen dem Deutschen und dem Englischen: Unbesetzte Endränder sind im Englischen fast ebenso häufig wie einfach besetzte Endränder. Diese Verteilung ähnelt derjenigen bei den einfachen Stämmen. Und wie dort liegt es auch hier vor allem am finalen ⟨e⟩. Die unbesetzten Endränder sind vor allem Suffixe (28 Suffixe vs. 6 Präfixe), und von diesen Suffixen sind 20 zweisilbig. Von diesen 20 zweisilbigen Suffixen enden wiederum 15 auf ⟨e⟩ (z. B. *-able*, *-ise*, *-ite*, *-ence*); die übrigen fünf enden auf ⟨y⟩ (*-ory*, *-ity*, *-ify*, *-ery*, *-ety*). Dreifach besetzte Endränder tauchen nur einmal auf, und zwar im Deutschen (*-wärts*).

Bei den einfachen Endrändern haben wir – im Gegensatz zu den einfachen Anfangsrändern oben – genügend Daten, um uns die Verteilung der einzelnen Buchstaben anzuschauen:

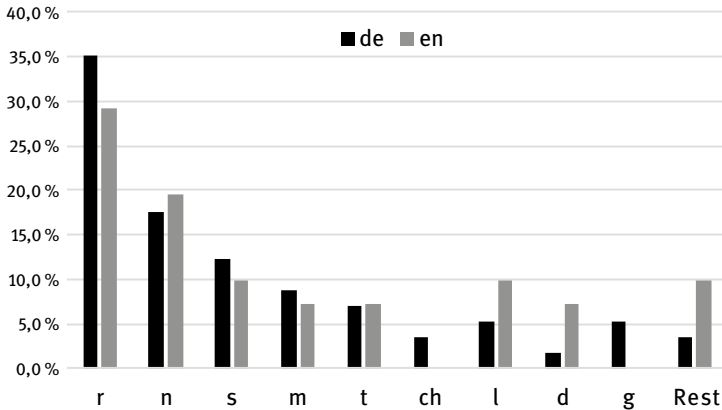


Abb. 43: Relative Anzahl der Konsonantengrapheme an allen einfach besetzten Endrändern von Affixen in den Korpora deutscher bzw. englischer Affixe.

Hier zeigt sich eine große Übereinstimmung: Die häufigsten einfachen Endränder von Affixen sind <r> (z. B. *-er*), <n> (z. B. *-en*) und <s> (z. B. *dis-*). Der Endrand <ch> und <g> kommt nur im Deutschen vor (z. B. *-lich*, *-ig*); <d> und <l> sind im Englischen frequenter (z. B. *-hood*, *-al*). Insgesamt treten nur wenige Grapheme in einfachen Endrändern auf; die Kategorie ‚Rest‘ in Abbildung 43 enthält für das Deutsche <k> und <v>, für das Englische <c>, <x>, und <p>, die alle jeweils einmal vorkommen.

Bei den zweifach besetzten Endrändern gibt es wesentlich weniger Variation als bei den Stämmen (vgl. 3.2.2.3):

Tab. 94: Zweifach besetzte Endränder in deutschen und englischen Affixen. Datengrundlage: Korpus deutscher bzw. englischer Affixe.

Endrand	deutsch	englisch
nt	3	3
st	2	3
nz	2	–
ng	2	2
ft	2	–
rz	1	–

Endrand	deutsch	englisch
ns	1	1
ss	1	3
nd	1	–
ll	1	–
rn	1	–
sm	–	1
sh	–	1

Alle Verbindungen genügen den Schemata, die für den deutschen und englischen Endrand oben aufgestellt wurden. Interessant ist, dass Doppelkonsonanten wesentlich seltener vorkommen als in Endrändern von Stämmen. Im Deutschen gibt es nur zwei Affixe mit finalen Doppelkonsonanten (*miss-* und *-ell*), im Englischen sind es drei (*-ess*, *-less* und *-ness*). Hier kann *⟨ss⟩* allerdings als morphologische Schreibung interpretiert werden: Es verhindert, dass Wörter wie *hatless* fälschlicherweise als Plurale interpretiert werden, wie es bei Schreibungen wie *⟨hatles⟩* der Fall wäre. Ebenfalls abwesend sind Verbindungen mit postvokalischem *⟨h⟩* im Deutschen und postvokalischem *⟨w⟩* im Englischen. Das postvokalisches *⟨h⟩* im Deutschen wurde oben als Möglichkeit interpretiert, das Silbengewicht zu regulieren. Das scheint bei den Affixen nicht notwendig zu sein.

3.3.4 Kerne

Betrachten wir zunächst die Anzahl der Vokalgrapheme in Kernen von einsilbigen Affixen im Deutschen und Englischen (siehe Abb. 44).

Einfache Vokalgrapheme sind in den Affixen der beiden Sprachen absolut dominant. Verbindungen von zwei (im Englischen auch drei, *-ious*, *-eous*) Graphemen sind die Ausnahme. Wenn sie überhaupt vorkommen, dann in Suffixen. Von dieser Generalisierung gibt es im Deutschen keine Ausnahmen; alle Affixe mit komplexem Silbenkern sind Suffixe (*-heit*, *-keit*, *-lein*, *-eur*, *-ei*, *-ie*, *-ier[en]*, *-ion*, *-oid*). Im Englischen gibt es genau eine Ausnahme, das Präfix *out-*.

Bei den einfach besetzten Kernen überwiegt in beiden Sprachen eindeutig *⟨e⟩*, wie Abbildung 45 zeigt. Grundsätzlich ist die Verteilung in beiden Sprachen sehr ähnlich. Die Umlaute kommen nur im Deutschen vor, *⟨y⟩* nur im Englischen – ansonsten sind die Differenzen überschaubar. Neben *⟨e⟩* kommen vor allem *⟨a⟩* und *⟨i⟩* als einfache Kerne vor; im Deutschen ist *⟨a⟩* etwas häufiger, im Englischen ist *⟨i⟩* deutlich frequenter.

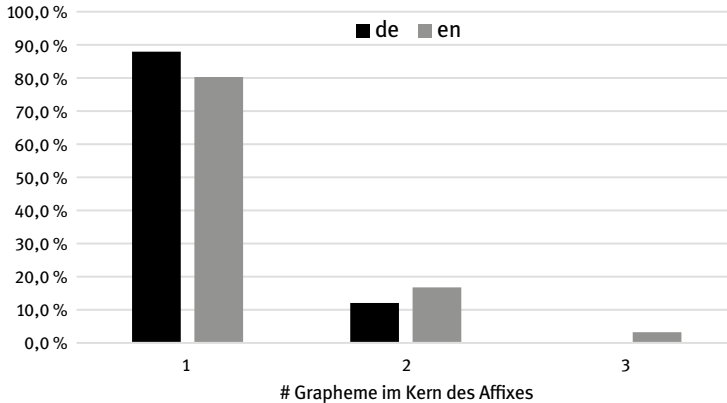


Abb. 44: Relative Anzahl der Vokalgrapheme an allen Kernen von einsilbigen Affixen in den Korpora deutscher bzw. englischer Affixe.

Zweifach besetzte Kerne verhalten sich im deutschen Korpus größtenteils wie komplexe Kerne von Stämmen, und zwar in Einsilbern. Die Kombinationen <ei> (4x), <ie> (2x) und <eu> (1x) entsprechen dem Schema für Schreibdiphthonge (Abb. 33 oben). Die beiden Ausnahmen sind hier <io> und <oi> (je 1x). Verdoppelungen von Vokalgraphemen kommen – genau wie Verdoppelungen von Konsonantengraphemen – nicht vor.

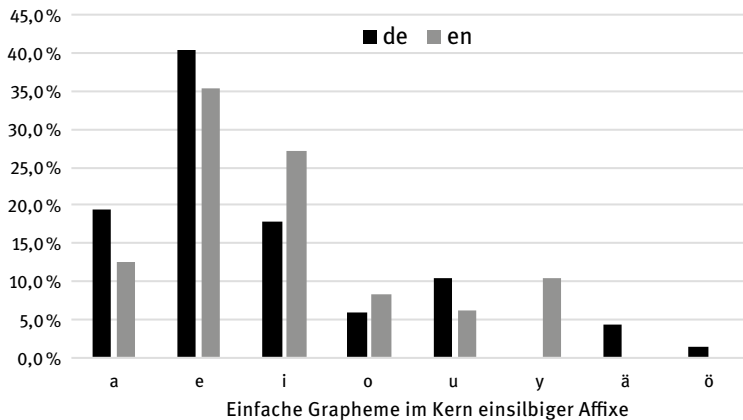


Abb. 45: Relative Anzahl der Vokalgrapheme an allen einfachen Kernen in einsilbigen Affixen in den Korpora deutscher bzw. englischer Affixe.

Im Englischen ist das anders. Zum einen sind Verdoppelungen belegt, nämlich ⟨ee⟩ (2x) und ⟨oo⟩ (1x). Zum anderen entspricht nur ⟨ou⟩ (2x) dem Schema von Schreibdiphthongen im Einsilber (Abb. 34 oben). Die übrigen Kombinationen sind in Stämmen vor allem in zweiten Silben von Zweisilbern prominent (vgl. Tab. 36 oben); es handelt sich um die Verbindungen ⟨ia⟩ (2x), ⟨ua⟩ (1x), ⟨ie⟩ (1x) und ⟨io⟩ (1x).

Komplexe Silbenkerne sind außerdem auf einsilbige Affixe beschränkt. Mit Ausnahme von englisch *counter-* und deutsch *-weise* und *-erie* sind Silbenkerne in zweisilbigen Affixen einfach. Das ist eine interessante Beobachtung: Offenbar dürfen Affixe graphematisch nicht zu ‚schwer‘ werden, genau wie lexikalische Stämme nicht zu ‚leicht‘ sein dürfen.

3.3.5 Minimale Affixe und prototypische Affixe

Minimale Affixe – das haben wir oben in 3.3.1 gesehen – bestehen aus einem Graphem, und zwar nicht einmal zwangsläufig aus einem Vokalgraphem. Bei diesen Affixen handelt es sich in beiden Sprachen um Suffixe; die einzige Ausnahme ist das Präfix *a-* im Deutschen (z. B. *asozial*) und Englischen (z. B. *acultural*). Die konsonantischen Suffixe wie *-s*, *-st* etc. sind, wie oben gezeigt, in erster Linie Flexionssuffixe. Teilweise gibt es homographische Derivationsuffixe, die sind aber seltener und/oder unproduktiv. Für Flexionssuffixe gibt es also keine Minimalitätsbeschränkung.

Affixe, die aus zwei Graphemen bestehen, sind relativ häufig, vgl. die Listen (52) für das Deutsche und (53) für das Englische:

(52a) *-al*, *-er*, *-ig*, *-iv*, *-ös*, *un-*, *ur-*

(52b) *be-*, *de-*, *ge-*, *re-*

(53a) *-al*, *-ar*, *ex-*, *-ic*, *un-*

(53b) *be-*, *-cy*, *-ly*, *-ry*, *-ty*

Die Liste in (52a) enthält Affixe (vor allem Suffixe) der Struktur ⟨VC⟩; sie ist nicht exhaustiv. Die Liste in (52b) enthält CV-Strukturen und ist vollständig. Sie enthält nur Präfixe, keine Suffixe. Offenbar ist ⟨CV⟩ keine gute Struktur für deutsche Suffixe. Das ist mit einer Minimalitätsbedingung wie der für Stämme fassbar: Der Reim muss auch bei Suffixen mit zwei Gewichtseinheiten verbunden sein. Da es eine ganze Reihe von Suffixen gibt, die diese Bedingung erfüllen und aus zwei Graphemen bestehen, gilt die zweite Bedingung von der Mindestlänge für Suffixe nicht.

Die Situation im Englischen ist auf den ersten Blick anders: Hier gibt es auch eine Reihe von Suffixen der Struktur ⟨CV⟩ in der exhaustiven Liste (53b). Aller-

dings enden alle diese Suffixe auf ⟨y⟩; sie verhalten sich damit wie Stämme, die auf einfaches ⟨y⟩ enden (z. B. *cry*, *try*). Für diese Stämme wurde die Minimalitätsbeschränkung im Englischen (Beleg (51) oben) angepasst: ⟨y⟩ im Kern verhält sich wie zwei Grapheme. Mit dieser Modifikation gilt die Beschränkung für den minimalen Reim auch für englische Derivationsuffixe: Der Reim muss mit mindestens zwei Graphemen besetzt sein. Die zweite Bedingung von der Mindestlänge gilt für Derivationsuffixe nicht.

Aus diesen Daten lassen sich die folgenden Minimalitätsbeschränkungen gewinnen: Für Flexionssuffixe gibt es keine Beschränkung. Die Beschränkungen für Derivationsaffixe unterscheiden sich je nachdem, ob es sich um Präfixe oder Suffixe handelt:

- (54) Minimalitätsbeschränkungen für Präfixe
 - (54a) Präfixe bestehen aus mindestens zwei Graphemen.
 - (54b) Der Kern von Präfixen ist mit mindestens einem Graphem besetzt.
- (55) Minimalitätsbeschränkungen für Derivationsuffixe
 - (55a) Der Kern von Derivationsuffixen ist mit mindestens einem Graphem besetzt.
 - (55b) Der Reim von Derivationsuffixen ist mit mindestens zwei Graphemen besetzt.
 - (55c) Jedes Graphem im Reim ist wiederum mit einer Gewichtseinheit verknüpft.
 - (55d) Im Englischen verhält sich ⟨y⟩ im Kern wie zwei Grapheme.

Für Präfixe gilt nur die Bedingung, dass der Kern besetzt sein muss und das Präfix insgesamt eine Mindestlänge von zwei Graphemen aufweist. Derivationspräfixe brauchen einen Reim, der mit zwei Graphemen besetzt ist – die Ausnahme ist ⟨y⟩ im Englischen, das den Reim auch alleine besetzen kann. Von diesen Minimalitätsbeschränkungen gibt es im Deutschen zwei Ausnahmen, das Präfix *a-* (z. B. *Asynergie*, *asozial*) und das Derivationsuffix *-e* (z. B. *Bremse*, *Hetze*). Im Englischen ist das Präfix *a-* ebenfalls eine Ausnahme (z. B. *apolar*), daneben das Ethnonym *-i* (z. B. *Israeli*). Das Suffix *-y*, das vor allem als Adjektivsuffix (z. B. *girly*) oder als Substantivsuffix (z. B. *burglary*) auftritt, verletzt die Beschränkung in (55) nicht, weil sich ⟨y⟩ im Kern verhält wie zwei Grapheme.

Nachdem wir auf diese Weise die minimalen Strukturen bestimmt haben, geht es im Folgenden um die prototypischen Strukturen. Da für Präfixe und Suffixe unterschiedliche Minimalitätsbeschränkungen gelten, werden beide im Folgenden getrennt betrachtet. Wir schauen uns zunächst die einsilbigen Affixe an und ordnen sie hinsichtlich der Besetzung ihrer Konstituenten. Tabelle 95 (links)

zeigt die Verteilung der einsilbigen Suffixe mit einfachem Kern im Deutschen, Tabelle 96 (rechts) die Verteilung der entsprechenden Präfixe:

Tab. 95 (links) und Tab. 96 (rechts): Anzahl der Strukturen von einsilbigen Suffixen (Tab. 95) bzw. Präfixen (Tab. 96) mit einfach besetztem Kern im Korpus deutscher Affixe, sortiert nach der Besetzung des Anfangsrandes (Zeilen) und des Endrands (Spalten).

	-	C	CC	CCC
-	1	17	11	
C		11	2	1
CC			1	

	-	C	CC
-	1	7	2
C	4	4	2
CC	1		1

Die bevorzugte Struktur ist für Suffixe und Präfixe gleichzeitig die minimale Struktur für Derivationsuffixe, $\langle VC \rangle$. Jede Abweichung davon – etwa $\langle CVC \rangle$ oder $\langle VCC \rangle$ – ist seltener. Unbesetzte Endränder gibt es (mit Ausnahme von *-e*) nur unter den Präfixen (*a-*, *be-*, *ge-*, *de-*, *re-*, *prä-*). Hier zeigt sich der Unterschied in den minimalen Strukturen, der oben angesetzt wurde. Die prototypische Struktur einsilbiger Stämme, $\langle CVCC \rangle$, wird im Deutschen nur von zwei Suffixen und zwei Präfixen geteilt (*-haft*, *-ling*, *miss-*, *durch-*).

Statt den Anfangs- und Endrand können wir auch den Silbenkern variieren. Es gibt allerdings, wie oben in 3.3.4 gezeigt wurde, nur wenige Affixe mit komplexem Kern. Diese neun Affixe sind allesamt Suffixe, und sie verteilen sich auf drei verschiedene Strukturen. Am frequentesten ist $\langle VVC \rangle$ (*-ier[en]*, *-ion*, *-eur*, *-oid*, alles Fremdsuffixe), dann $\langle CVVC \rangle$ (*-heit*, *-keit*, *-lein*) und schließlich rein vokalisches $\langle VV \rangle$ (*-ei*, *-ie*). Bei komplexen Kernen ist der Rand also maximal einfach geschlossen.

Im Englischen ist das Bild sehr ähnlich. Tabelle 97 und Tabelle 98 zeigen die Verteilung der einsilbigen Affixe mit einfachem Kern:

Tab. 97 (links) und Tab. 98 (rechts): Anzahl der Strukturen von einsilbigen Suffixen (Tab. 97) bzw. Präfixen (Tab. 98) mit einfach besetztem Kern im Korpus englischer Affixe, sortiert nach der Besetzung des Anfangsrandes (Zeilen) und des Endrands (Spalten).

	-	C	CC
-	(2)	9	8
C	(4)	3	4
CC		2	

	-	C	CC
-	1	7	
C	1	4	1
CC	1		1

Auch hier ist die bevorzugte Struktur, ⟨VC⟩, gleichzeitig die minimale Struktur für Derivationsuffixe. Bei den sechs Suffixen ohne Endrand handelt es sich um die oben besprochenen Suffixe (-cy, -ly, -ry, -ty, -i, -y), in denen (mit Ausnahme von -i) sich ⟨y⟩ im Kern verhält wie eine Verbindung aus zwei Vokalgraphemen. Im Unterschied zum Deutschen sind Suffixstrukturen vom Typ ⟨CVC⟩ (-dom, -ful, -let) im Englischen nicht häufiger als Strukturen vom Typ ⟨CVCC⟩ (-ness, -less, -ment, -ling), also der prototypischen Struktur einsilbiger Stämme.

Zweifach besetzte Kerne sind auch im Englischen seltener; hier gibt es dieselben drei Strukturen wie im Deutschen, nämlich ⟨VVC⟩ (out-, -ual, -ous, -ion, -ian, -ial, -eer), ⟨VV⟩ (-ie, -ee) und ⟨CVVC⟩ (-hood). Wie im Deutschen treten komplexe Kerne also nicht gemeinsam mit komplexen Endrändern auf.

Wir können aus diesen Beobachtungen abschließend Präferenzaussagen über die graphematische Form englischer und deutscher Derivationsaffixe machen:⁶⁵

- (56) Die graphematische Form eines Derivationsaffixes ist umso mehr bevorzugt,
- (56a) je näher die Anzahl der Silben bei eins liegt,
- (56b) je näher die Anzahl der Grapheme im Anfangsrand bei null liegt,
- (56c) je näher die Anzahl der Grapheme im Kern bei eins liegt und
- (56d) je näher die Anzahl der Grapheme im Endrand bei eins liegt.

Das passt gut zu den „Allgemeinen Präferenzen für Derivationsuffixe“, die Fuhrhop (1998: 92 f.) für die phonologische Form deutscher Derivationsuffixe aufgestellt hat. Einschlägig sind in diesem Zusammenhang vor allem Präferenz (1) („Derivationsuffixe mit silbischem Kern werden gegenüber solchen ohne silbischen Kern bevorzugt“) und Präferenz (4) („Derivationsuffixe, die sonor anlauten, werden gegenüber solchen bevorzugt, die stärker konsonantisch anlauten“). Präferenz (1) deckt sich zum Teil mit (56a) (Fuhrhop macht allerdings keine Aussagen zu mehrsilbigen Suffixen); Präferenz (4) korrespondiert mit (56d): Vokalisch anlautende Suffixe sind gegenüber konsonantisch anlautenden bevorzugt. Oben wurde wegen der geringen Datenbasis darauf verzichtet, die Grapheme im einfachen Anfangsrand auszuwerten, deswegen macht (56) dazu keine weitere Aussage (auch wenn die Beobachtung, dass ⟨b⟩ der häufigste Anfangsrand von

⁶⁵ Nach (56) ist ⟨ir⟩ die bessere Form für deutsch -ier[er] als das seit 1876/1901 kodifizierte ⟨ier⟩ (⟨demonstrieren⟩, *⟨demonstrieren⟩) – tatsächlich ist die heutige Schreibung problembehaftet, wie Fuhrhop/Buchmann (2016) zeigen.

Suffixen ist, ins Bild passt) – und überhaupt ist das Konzept der Sonorität ja nicht ohne Weiteres auf schriftliche Strukturen übertragbar.

Dass nun Affixe von diesen Präferenzen abweichen – bspw. *-ous* im Englischen –, kann in vielen Fällen morphologisch motiviert werden: die graphematische Form <ous> verweist eindeutig auf das Suffix *-ous* und unterscheidet sich von homophonen Endungen wie <us>, <is> oder <ice> (vgl. 4.4).

3.4 Zusammenfassung und Diskussion

In diesem Kapitel ging es um den graphemischen Aufbau von Morphemen im Deutschen und Englischen. Untersucht wurden das Buchstaben- und Grapheminventar sowie der graphemische Aufbau von Affixen und einfachen Stämmen – dieser letztere Punkt ist (auch vom Umfang her) der Schwerpunkt des Kapitels. Da die wichtigsten Erkenntnisse bereits größtenteils in den Zusammenfassungen der Teilkapitel zusammengetragen wurden (3.2.1.6, 3.2.2.8, 3.2.5), sollen hier nur noch minimale und prototypische Stämme und Affixe verglichen werden.

Für lexikalische Stämme – nicht für Funktionswörter! – gilt die Minimalitätsbeschränkung (57):

- (57) Minimalitätsbeschränkungen für Stämme und Wurzeln:
- (57a) Stämme bestehen aus mindestens drei Graphemen.
- (57b) Der Reim von Stämmen und Wurzeln ist mit mindestens zwei Graphemen besetzt.
- (57c) Jedes Graphem im Reim ist wiederum mit einer Gewichtseinheit verknüpft.
- (57d) Es gelten folgende sprachspezifische Besonderheiten:
 - Im Deutschen sind <l>, <m>, <n> und <r> im Endrand leichter; ein daraus resultierendes zu geringes Gewicht kann durch einen komplexen Anfangsrand kompensiert werden.
 - Im Englischen sind einfaches <l>, <f> und <s> im Endrand leichter; <y> im Kern verhält sich wie zwei Grapheme.

Funktionswörter sind oben zwar diskutiert worden, es ist aber keine minimale Struktur angesetzt worden. Da es im Deutschen keine Funktionswörter mit einem Buchstaben gibt und da weiterhin jedes Funktionswort über einen Silbentern verfügt, verhalten sich deutsche Funktionswörter in dieser Hinsicht wie Präfixe. Im Englischen gilt das nicht: Hier finden sich drei Wörter, die nur aus einem

Vokalgraphem bestehen, *a*, *o* und *I*. Die Tatsache, dass letztere Form großgeschrieben ist, ist in diesem Zusammenhang nicht unwesentlich – graphisch hat der Buchstabe allerdings weniger Substanz als viele Minuskeln. Und mit *a* und *I* sind zwei der drei möglichen minimalen Wörter auch noch unter den frequentesten Wörtern des Englischen. Für das Deutsche fallen also die Beschränkungen für Funktionswörter und Präfixe zusammen, im Englischen müssen Funktionswörter nur einen Silbenkern haben:

- (58) Minimalitätsbeschränkungen für Präfixe im Deutschen und Englischen und Funktionswörter im Deutschen:
- (58a) Präfixe/Funktionswörter bestehen aus mindestens zwei Graphemen.
- (58b) Der Kern von Präfixen/Funktionswörtern ist mit mindestens einem Graphem besetzt.
- (59) Minimalitätsbeschränkungen für Funktionswörter im Englischen:
Der Kern von Funktionswörtern ist mit mindestens einem Graphem besetzt.

Für Derivationsuffixe im Deutschen und Englischen gilt zusätzlich, dass der Reim zweifach besetzt sein muss (und dass ⟨y⟩ sich verhält wie zwei Grapheme):

- (60) Minimalitätsbeschränkungen für Derivationsuffixe:
- (60a) Der Kern von Derivationsuffixen ist mit mindestens einem Graphem besetzt.
- (60b) Der Reim von Derivationsuffixen ist mit mindestens zwei Graphemen besetzt.
- (60c) Jedes Graphem im Reim ist wiederum mit einer Gewichtseinheit verknüpft.
- (60d) Im Englischen verhält sich ⟨y⟩ im Kern wie zwei Grapheme.

Das sind die minimalen Beschränkungen. Anders sieht es aus, wenn wir uns die jeweils prototypischen (im Sinne von häufigen) Formen für einsilbige Stämme (61) und Affixe (62) anschauen:

- (61) Die graphematische Form eines einsilbigen Stamms ist umso mehr bevorzugt,
- (61a) je näher die Anzahl der Grapheme im Anfangsrand bei eins liegt,
- (61b) je näher die Anzahl der Grapheme im Kern bei eins liegt und
- (61c) je näher die Anzahl der Grapheme im Endrand bei zwei liegt.

- (62) Die graphematische Form eines Derivationsaffixes ist umso mehr bevorzugt,
 (62a) je näher die Anzahl der Silben bei eins liegt,
 (62b) je näher die Anzahl der Grapheme im Anfangsrand bei null liegt,
 (62c) je näher die Anzahl der Grapheme im Kern bei eins liegt und
 (62d) je näher die Anzahl der Grapheme im Endrand bei eins liegt.

Die prototypische Form von einsilbigen Stämmen ist also $\langle CVCC \rangle$, die von Affixen $\langle VC \rangle$. Abweichungen von diesen Strukturen sind seltener, und zwar umso seltener, je mehr Segmente sich unterscheiden. Wenn wir nun beide Strukturen vergleichen, stellen wir fest, dass es kaum Überlappungen gibt:

	-	C	CC	CCC
-		Affix		
C			Stamm	
CC				
CCC				

Abb. 46: Die Verteilung prototypischer einsilbiger Stämme und einsilbiger Affixe im Deutschen und Englischen.

Affixe sind selbstverständlich insgesamt seltener als Stämme; Stämme sind daher in jedem Strukturtyp häufiger. Umgekehrt kommen Affixe im Schnitt aber in mehr Wortformen vor als Stämme – und man könnte noch weiter gehen und auch die Tokenfrequenz dieser Wortformen untersuchen. Auf entsprechende Aufstellungen wird im Folgenden verzichtet. Stattdessen bleiben wir bei der groben Darstellung in Abbildung 46, die die relativen Anteile der Strukturen an allen Affixen bzw. Stämmen schematisiert.⁶⁶

Substanzielle Überlappungen gibt es nur in einer Struktur, nämlich $\langle CVC \rangle$. Auf diesen Typ entfallen im Deutschen 16 % der einsilbigen Stämme und 23 % der einsilbigen Affixe; im Englischen sind es 19 % der einsilbigen Stämme und 15 % der einsilbigen Affixe. Und genau bei diesem Strukturtyp unterscheiden sich in beiden Sprachen Affixe und Stämme hinsichtlich der möglichen Endränder. In

⁶⁶ Vgl. auch Nida (1949: 65 f.), der drei phonologische Haupttypen von Morphemen ansetzt, C, VC und CVC (wobei V und C jeweils für ein oder mehrere aufeinanderfolgende Segmente dieses Typs stehen). Nida beobachtet, dass die Strukturen C und VC im Englischen gewöhnlich unter Affixen zu finden seien, die Struktur CVC hingegen unter den Stämmen.

Stämmen sind im Englischen einfaches ⟨s⟩, ⟨f⟩ und ⟨l⟩ dispräferiert; für Affixe gilt das nicht, vgl. *-ful, dis-*. Im Deutschen sind in Stämmen einfaches ⟨l⟩, ⟨m⟩, ⟨n⟩ und ⟨r⟩ sehr selten; hier reguliert ⟨h⟩ das Silbengewicht (vgl. oben 3.2.3). In Affixen sind diese Endränder aber durchaus möglich, vgl. *-bar, -chen, kon-, -ler, -ner, -sam, -tum, ver-, zer-*. Diese Affixe sind also – obwohl es relativ viele Stämme der Form ⟨CVC⟩ gibt – ziemlich eindeutig als Affixe erkennbar.

4 Morphographische Korrespondenzen

In diesem Abschnitt stehen die Beziehungen zwischen der graphematischen Ebene und der morphologischen Ebene im Mittelpunkt. Die zentrale Frage lautet: Wo und auf welche Weise bestimmen morphologische Informationen die graphematische Form von Wörtern? Die Möglichkeiten sind prinzipiell weitreichend, wie die (leicht überarbeitete) Übersicht aus Berg et al. (2014: 305) anhand von englischen Beispielen zeigt:

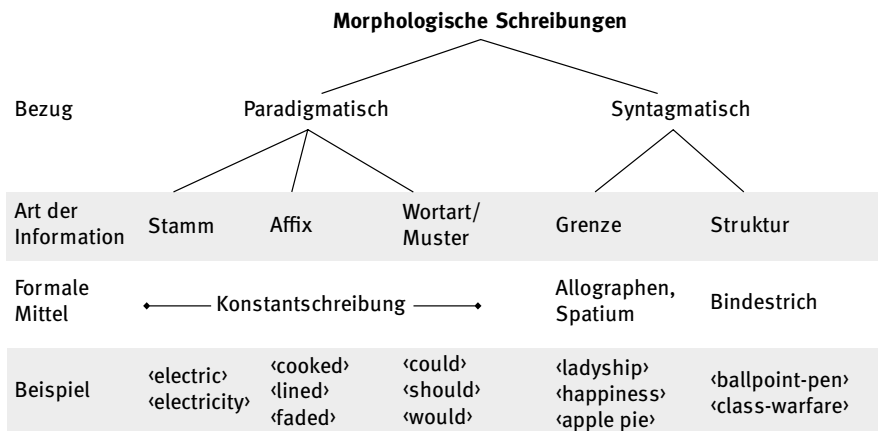


Abb. 47: Übersicht über morphologische Schreibungen aus Berg et al. (2014: 305).

Morphologische Schreibungen können zunächst sehr grob danach klassifiziert werden, ob sie paradigmatischer oder syntagmatischer Natur sind (vgl. Primus 2010). So ist bspw. die Stammkonstanz eine paradigmatische morphologische Schreibung: Sie bezieht sich auf nicht-realisierte Einheiten (also Formen *in absentia*, vgl. de Saussure 1967: 148). Der Stamm *electric* wird in allen Umgebungen konstant verschriftet, trotz Varianz in der phonologischen Form (*electrician*, *electricity*). Wenn wir das als morphologische Schreibung bezeichnen, beziehen wir die tatsächlich realisierte Form auf alle anderen möglichen Formen in anderen Umgebungen. Die Bindestrichschreibung ist hingegen eine syntagmatische morphologische Schreibung: Sie markiert die Struktur von Einheiten, die gemeinsam realisiert werden (also *in praesentia*).

Eine etwas feinere Gliederung klassifiziert die Schreibungen wiederum nach der Art der morphologischen Information, die kodiert wird: Paradigmatisch können Stämme, Affixe oder Wortarten bzw. grammatische Muster kodiert werden, syntagmatisch Grenzen und Strukturen. Die invariante (konstante) Kodie-

nung von Stämmen und Affixen ist das, was in der germanistischen Linguistik mit Begriffen wie „Morphemkonstanz“ (Günther 1988: 87 ff.; Primus 2010; Eisenberg 2013a: 310), „Stammprinzip“ (Gallmann/Sitta 1996: 35), „Schemakonstanz“ (Gallmann 1990; Dürscheid 2012: 141) oder „morphemidentifizierenden Schreibungen“ (Nerius (Hg.) 2007: 148 ff.) bezeichnet wird. Es geht im Kern darum, dass „die geschriebene morphologische Einheit [...] ihre Gestalt unter fast allen Bedingungen [behält], ihre Segmentfolge ist stabil“, wie Eisenberg (2013a: 310) es formuliert.

Im Deutschen werden beispielsweise Stämme meist konstant verschriftet – sie haben eine weitestgehend einheitliche Form. Der Stamm *Kind* wird graphematisch immer als <Kind> verschriftet, unabhängig davon, in welcher Umgebung er vorkommt (<Kind>, <Kinder>, <kindlich>).⁶⁷ Das gilt für die phonologische Form nicht; sie variiert kontextabhängig, und zwar abhängig von der silbenstrukturellen Position des Stammauslauts (/kɪnt/, /kɪndə/, /kɪntlɪç/).

Im Englischen wird ein analoges Prinzip angesetzt – allerdings nicht innerhalb von Flexionsparadigmen (hier gibt es zu viele Ausnahmen, z. B. <swim> – <swimming>), sondern in der Derivation. Es gibt es zahlreiche Beispiele, in denen die schriftliche Form stabiler ist als die phonologische, z. B. *electric* – *electricity*, *sane* – *sanity* oder *autumn* – *autumnal*. Diese Schreibungen werden meist als ‚lexical spellings‘ (Carney 1994: 18 ff.), das entsprechende Prinzip z. T. als „Identity Preservation Principle“ (Ryan 2016: 78 ff.) gefasst: „[T]he established graphic form of the base is retained as much as possible, regardless of the phonemic alternations involved“ (Venezky 1970: 120). Die interessante Frage ist, wo genau die Grenzen dieses Prinzips liegen – was heißt also ‚as much as possible‘ (vgl. auch Ryan 2016: 84 f.)?

Die Stammkonstanz ist verwandt mit der Differenzierung von Homonymen (Nerius (Hg.) 2007: 149): Bei der Stammkonstanz wird formal einheitlich gehalten, was lexikalisch dasselbe ist; bei der Differenzierung von Homonymen wird formal unterschieden, was lexikalisch unterschiedlich ist. Unmittelbar deutlich wird das, wenn homophone Lexeme in der Schrift differenziert werden. Die einschlägigen Beispiele sind <Leib>/<Laib> und <Lied>/<Lid>. Die Differenzierung von Homonymen firmiert auch unter dem Begriff des „Homonymieprinzips“ (Rahnenführer 1980: 248). Homonyme werden allerdings im Deutschen nicht systematisch differenziert: Wie Günther (1988: 88), Nerius (Hg.) (2007: 149 f.) und Eisenberg (2013a: 316 f.) zeigen, ist die Reichweite dieses Prinzips sehr begrenzt, viele Homophone werden nicht graphematisch differenziert, auch wenn es denk-

⁶⁷ Die Groß- und Kleinschreibung ist syntaktisch motiviert (vgl. Eisenberg 1981) und wird daher hier vernachlässigt.

bare Schreibvarianten gibt (z. B. ⟨Ton⟩ – *⟨Tohn⟩, ⟨Kiefer⟩ – *⟨Kifer⟩, vgl. Eisenberg 2013a: 409). Im Englischen gibt es auf den ersten Blick mehr Belege für ein entsprechendes Prinzip (vgl. z. B. die umfangreiche Liste in Carney 1994: 401 ff.), aber es wird ebenfalls häufig auf die Inkonsistenzen verwiesen (z. B. Sproat 2016: 36).⁶⁸ Wie viele Lexempaare in den beiden Sprachen differenziert werden und wie viele nicht, ist aber eine offene Forschungsfrage (siehe 4.2).

Eine Formalisierung der Idee der Stammkonstanz und des Homonymieprinzips wird in Berg (2017) entwickelt. Hier wird dafür argumentiert, Korrespondenzen zwischen der graphematischen Form und der morphologischen Struktur formal so zu behandeln wie Phonem-Graphem-Korrespondenzen (vgl. oben 1.2). Wir haben dann neben phonographischen Korrespondenzen wie ⟨d⟩/_# → /t/ (das Auftreten des Graphems ⟨d⟩ im Kontext vor einer Wortgrenze impliziert das Auftreten des Phonems /t/) auch morphographische Korrespondenzen wie ⟨ed⟩/_# → {-ed} (wortfinals ⟨ed⟩ impliziert das englische Flexionsmorphem *-ed*). Diese Bezüge können (wie bei den phonographischen Bezügen auch) in beide ‚Richtungen‘ untersucht werden; man kann dann allgemein von morphographischen Korrespondenzen reden (daher die Überschrift dieses Kapitels).

Der Bezug auf Homophone, die graphematisch differenziert werden, oder auf phonologische Alternationen, die in der Morphemkonstanz graphematisch überbrückt werden, zeigt, dass diese Phänomene nicht rein morphographisch gedacht sind, sondern oft im Kontrast zu den phonologischen Verhältnissen. Das hat einen guten Grund: Wir bemerken überhaupt nur aufgrund dieser Differenz der phonologischen Formen die Konstanz der graphematischen Formen. Aber natürlich besteht Morphemkonstanz auch in Fällen wie *Tisch – Tische – Tischler*, in denen der Stamm auch phonologisch konstant ist: Die graphematische Form ⟨Tisch⟩ variiert graphematisch nicht. Es macht Sinn, diese Fälle mit Fuhrhop/Barghorn (2012) begrifflich zu unterscheiden. Variiert die phonologische Form bei konstanter graphematischer Form (Typ *Kind, electric*), so handelt es sich um explizite Morphemkonstanz; sind beide Formen invariant (Typ *Tisch*), so handelt es sich um implizite Morphemkonstanz.

Zurück zu Abbildung 47 oben. Neben Stämmen und Affixen können auch Wortarten bzw. grammatische Funktionen konstant verschriftet werden. Das ist eine relativ neue Erkenntnis (Berg et al. 2014; Fuhrhop 2017). Die Formen ⟨would⟩, ⟨could⟩ und ⟨should⟩ repräsentieren jeweils Präteritalformen der Modalverben (‚Präteritalform‘ ist hier ausschließlich strukturell gemeint, nicht temporal, vgl. Huddleston 2002: 196–208). Außer diesen dreien gibt es nur noch eine weitere

⁶⁸ Einen Mittelweg geht Sampson (1985: 203), der das englische Schriftsystem als „partly logographic“ klassifiziert.

solche Form, ⟨might⟩. Drei der vier Modalverben enden also mit dem gleichen Muster, ⟨ould⟩.⁶⁹ Diese Form ist phonographisch auffällig: ⟨ou⟩ korrespondiert sonst nicht mit /ʊ/; außerdem hat das ⟨l⟩ keine phonologische Entsprechung. Dieses Muster ist für Modalverben reserviert: Es gibt außer den drei Modalverben nur ein Nicht-Modalverb auf ⟨ould⟩, *mould* – und auch das nur in der britischen Schreibvariante (die amerikanische Variante ist das phonographisch regelmäßige ⟨mold⟩ analog zu ⟨gold⟩, ⟨fold⟩, ⟨hold⟩). Es gibt aber, und das ist die zentrale Idee in Berg et al. (2014), vier weitere Wörter, die mit ⟨ould⟩ verschriftet werden könnten; sie enden alle phonologisch auf /ʊd/, werden aber anders geschrieben (⟨hood⟩, ⟨good⟩, ⟨pud⟩, ⟨wood⟩). Wir nutzen also die phonologische Form, um *potenzielle* graphematische Formen zu bestimmen. Dieses Argumentationsmuster wird im vorliegenden Kapitel ausgebaut. Außer den Modalverben gibt es weitere grammatische Muster bei den starken Verben (z. B. finales ⟨ought⟩, das nur dort auftritt, vgl. Fuhrhop 2017: 60 ff.).

Syntagmatisch kann die Markierung von Grenzen und Strukturen unterschieden werden, wobei die Markierung von Strukturen die von Grenzen miteinschließt – jede Markierung einer Struktur ist auch eine Markierung einer Grenze, aber nicht umgekehrt. Mit Grenzmarkierung ist gemeint, dass die Stärke der morphologischen Verbindung graphematisch kenntlich gemacht wird (Berg 2013). Das geschieht z. B. über die ⟨i⟩/⟨y⟩-Alternation im Englischen. Die beiden Grapheme alternieren stammfinal. Ist das Stammende gleichzeitig das Wortende, erscheint ⟨y⟩ (z. B. ⟨story⟩, ⟨copy⟩). Folgt ein Flexionssuffix außer -'s, tritt ⟨i(e)⟩ statt ⟨y⟩ auf (z. B. ⟨stories⟩, ⟨copied⟩). In Komposita alterniert ⟨y⟩ nicht, vgl. ⟨storybook⟩, *⟨storibook⟩. Abgeleitete Stämme mit finalem ⟨y⟩ verhalten sich nun teilweise wie Flexionsformen, indem ⟨y⟩ mit ⟨i⟩ alterniert (wie in ⟨happiness⟩); teilweise verhalten sie sich wie Komposita, indem sich die Stammform nicht ändert (z. B. ⟨storyless⟩). Interessanterweise korreliert der Grad der Alternation hochsignifikant mit dem *Type Parsing Ratio* von Hay/Baayen (2003), einem rein morphologischen Maß, das Aussagen über die Frequenz von derivierten Wörtern relativ zur Frequenz ihrer Basis macht. Wenn ein Suffix viele Wörter umfasst, deren Frequenz höher ist als die ihrer Basis (wie z. B. *government*, das frequenter ist als das Basisverb *govern*), dann ist das *Type Parsing Ratio* niedrig; das Wort wird – so ist zu vermuten – eher als Ganzes verarbeitet und nicht in seine Teile zerlegt, und die Grenze zwischen den Bestandteilen ist nicht wichtig. Umfasst ein Suffix dagegen viele Wörter, deren Basis frequenter ist (wie etwa *discernment*, das seltener ist als die Basis *discern*), dann hat das betreffende Suffix ein hohes *Type Parsing*

⁶⁹ Diachron hat sich die Schreibung mit dem Muster ⟨ould⟩ langsam herausgebildet, wie Cummins (1988: 7) zeigt.

Ratio; es wird wahrscheinlich kompositionell verarbeitet, und die Grenze zwischen den Bestandteilen ist wichtig. Es zeigt sich nun, dass ein hohes Type Parsing Ratio (also viele Wörter, in denen die Grenze wichtig ist) Hand in Hand geht mit stabilen Stammformen; hier sind die einzelnen Bestandteile transparenter, die Grenze zwischen ihnen ist mit dem stamfinalen ⟨y⟩ gut markiert. Ein niedriges Type Parsing Ratio (also viele Wörter, in denen die Grenze nicht wichtig ist) kommt hingegen häufig mit einem hohen Grad an Alternation vor.

Neben der ⟨y⟩/⟨i⟩-Alternation spielt (zumindest im Englischen) das Spatium eine Rolle als Grenzmarkierung. Es kann innerhalb morphologischer Wortformen auftreten und so Grenzen zwischen den Bestandteilen kennzeichnen (z. B. ⟨apple pie⟩). Im Deutschen ist dieser Fall markiert und stigmatisiert („Deppenleerzeichen“).

Der Bindestrich fungiert ebenfalls als Grenzmarkierung. Im Englischen kann er in Komposita – in Kombination mit Zusammenschreibung und Leerzeichen – auch morphologische Strukturen markieren: In linksverzweigenden Komposita wie ⟨pancake roll⟩ oder ⟨waste-paper basket⟩ ist die Grenze vor dem Letztglied stärker als die nach dem Erstglied (wenn wir eine unkontroverse Hierarchie Leerzeichen > Bindestrich > Zusammenschreibung annehmen, vgl. z. B. Nunberg/Briscoe/Huddleston 2002: 1761). Bei rechtsverzweigenden Komposita wie ⟨picture postcard⟩ oder ⟨brake horse-power⟩ ist es genau umgekehrt: Hier ist die Grenze nach dem Erstglied die stärkere. Auf diese Weise wird die interne hierarchische Struktur der Wortbildungen sichtbar gemacht.

Die Bandbreite morphologischer Schreibungen ist also groß – und ein Typ morphologischer Schreibungen ist in Abbildung 47 gar nicht enthalten, obwohl er in Berg et al. (2014) eine wichtige Stellung einnimmt: Die Beobachtung, dass finales ⟨s⟩ für morphologisches -s und ⟨ould⟩ für Modalverben reserviert ist, ist natürlich ebenfalls eine bestimmte Art morphologischer Schreibung (wenn auch eine indirekte). Es handelt sich bei dieser indirekten morphologischen Schreibung um das Gegenstück zur Konstanzschreibung: Die Konstanzschreibung erfasst die Tatsache, dass Morpheme eine konstante Schriftform aufweisen; wir gehen hier gewissermaßen vom Morphem zur Form. Die indirekte morphologische Schreibung erfasst die Tatsache, dass Schriftformen von Morphemen exklusiv für diese Morpheme sind; hier gehen wir von der Form zum Morphem.

Dieser Unterschied in der Betrachtungsrichtung ist im Folgenden zentral. Wir erfassen ihn mit den beiden Parametern *Einheitlichkeit* und *Eindeutigkeit*. Ein Morphem ist einheitlich, wenn es in allen Umgebungen uniform verschriftet wird.⁷⁰ Umgekehrt ist eine graphematische Form eindeutig, wenn jedes Auftre-

⁷⁰ Die Einheitlichkeit entspricht Mayerthalers (1981: 34 f.) ‚schwacher Uniformität‘.

ten der Form das entsprechende Morphem repräsentiert. Beide Parameter sind logisch unabhängig voneinander. Ein Morphem kann uneinheitlich sein wie bspw. der Stamm MANN⁷¹ mit den Formen ⟨Mann⟩, ⟨Männ⟩ – jede dieser Formen repräsentiert aber eindeutig das Lexem MANN. Umgekehrt wird der Stamm KIEFER (BAUM) einheitlich verschriftet, die entsprechende Form ⟨Kiefer⟩ ist aber nicht eindeutig, weil sie auch den Stamm KIEFER (SCHÄDELKNOCHEN) repräsentiert. Es scheint am ergiebigsten, wenn wir die Untersuchung der verschiedenen morphologischen Schreibungen in Abbildung 47 auf die Eindeutigkeit und die Einheitlichkeit graphematischer Formen beschränken und versuchen, die Bezüge zwischen Schrift und Morphologie hier möglichst präzise herauszuarbeiten.

Die Untersuchung der Eindeutigkeit und Einheitlichkeit von Lexemen ist dabei relativ unproblematisch. Bei den Affixen ist das anders – hier stellt sich die Frage nach dem Abstraktionsgrad. Um das zu zeigen, ist zunächst ein kurzer morphologischer Exkurs notwendig. Affixe sind Morpheme, und die Definition des Morphems ist mindestens so umstritten wie die des Graphems. Für viele Zwecke kommt man ganz gut ohne Morpheme zurecht (vgl. Anderson 1992), und wortbasierte Modelle (siehe oben 1.2) sind dafür besonders gut geeignet. Gerade für die Frage der Einheitlichkeit von Affixen ist das aber problematisch. Prinzipiell gibt es zwei Möglichkeiten, wenn man auf abstrakte Morpheme verzichten möchte:

1. Wir untersuchen die Einheitlichkeit von Minimalzeichen oder Morphen (s. u.) wie deutsch ⟨-s⟩ GENITIV oder ⟨-es⟩ GENITIV. Wir segmentieren komplexe sprachliche Ausdrücke in Form-Funktions-Paare – diese werden aber nicht weiter zusammengefasst. Eine Untersuchung, wie einheitlich die Formen die Minimalzeichen repräsentieren, ist überflüssig: Alle Minimalzeichen haben per Definition nur eine Form. Diese oberflächennahe Lösung scheidet also aus.
2. Die andere Möglichkeit ist, direkt bei der Funktion bzw. der Semantik anzusetzen. In der Flexionsmorphologie etwa können wir fragen, wie einheitlich das Merkmal PLURAL graphematisch kodiert wird. Das ist eine interessante Frage. Sie ist aber leider nur für die Flexionsmorphologie ohne Weiteres operationalisierbar. Was ist die Funktion von Derivationsaffixen? Das mag für Suffixe wie deutsch *-heit* oder *-ig*, die vorwiegend transponieren (Erben 2006: 24), relativ eindeutig sein; bei Suffixen wie deutsch *-ling*, *-sam* oder *-chen* ist das schon schwieriger. Es ist nicht unmöglich, hier Funktionen oder Bedeu-

⁷¹ Wir folgen hier der Terminologie von Matthews (1991: 24 ff.). Lexeme wie MANN sind abstrakte Einheiten. Sie umfassen ein Paradigma mit verschiedenen (synthetischen) Wortformen (*Mann*, *Mannes*, *Männer*, *Männern*) und werden in Kapitälchen gesetzt, um den kategorialen Unterschied deutlich zu machen.

tungen anzugeben – die Beschreibungen sind aber nicht als ‚Input‘ für die Untersuchung der Einheitlichkeit verwendbar; dafür müssten sie bspw. als Merkmalsbündel o. Ä. vorliegen. Dasselbe Problem haben wir auch bei den Stämmen – es gibt kein operationalisierbares Konzept der Bedeutungen. Außerdem führt die Semantik als Bezugsebene dazu, dass auch Synonyme untersucht würden: Die Bedeutung ‚Orange‘ kann schließlich durch die graphematischen Formen <Apfelsine> und <Orange> ausgedrückt werden. Diese Variation ist aber etwas kategorial anderes als die oben skizzierten Fälle wie <Mann>, <Männ>. Auch diese Bezugsebene kommt also für uns nicht infrage.

Es ist daher notwendig, abstrakte Morpheme anzusetzen. Im Folgenden werden – analog wie bei den Graphemen – die Definitionskriterien motiviert und transparent gemacht.

Auf und teilweise unterhalb der Wortebene gehen sprachliche Formen und Funktionen nicht weiter analysierbare Verbindungen ein, die arbiträr und konventionalisiert sind. Wir können die entsprechenden Einheiten „Minimalzeichen“ (z. B. Mugdan 1994) oder „Morphe“ (z. B. Bauer 2003) nennen. Die Wortform *trinkbare* kann bspw. in die Minimalzeichen *trink*, *bar* und *e* zerlegt werden. Die Bindung von Formen an Funktionen führt in einigen Fällen zu homonymen Minimalzeichen (vgl. z. B. Harris 1951: 199) – so z. B. in der Wortform *Erlöser*, die in die Minimalzeichen er_v , *lös* und er_N zerlegt werden kann. Beim Präfix *er-* handelt es sich um ein Verbalaffix, beim Suffix *-er* um ein Nominalaffix, das zeigen die Indizes *N* und *V*.

Die morphologische Analyse kann hier nicht aufhören, sonst wäre die Frage nach der Einheitlichkeit von Morphemen denkbar unspannend – jedes Minimalzeichen hat ja (wie oben erwähnt) genau eine Form. Wir abstrahieren also morphologisch von der Form und fassen Gruppen von Minimalzeichen zu Morphemen zusammen. Das klassische Beispiel stammt aus der englischen Substantivflexion. Abhängig vom Stammauslaut treten als Pluralmorpheme */s/* (*cats*), */z/* (*dogs*) oder */ɪz/* (*horses*) auf. Sie haben a) dieselbe Funktion und sind b) komplementär verteilt, daher werden sie zu einem abstrakten Morphem *-s PLURAL* zusammengefasst (vgl. z. B. Gleason 1961: 80). Die einzelnen Morphe */s/ PLURAL*, */z/ PLURAL* und */ɪz/ PLURAL* sind *Allomorphe* oder *Alternanten* dieses abstrakten Morphems (z. B. Bloomfield 1933: 210 ff.). Die Verteilung ist phonologisch determiniert, sie ist (in Bloomfields Begriffen) „regulär“ und „automatisch“ (Bloomfield 1933: 211). Die drei Allomorphe sind sich formal sehr ähnlich. Das gilt, so kann vermutet werden, für den Großteil der Allomorphe: „Most often, [...] allomorphs are phonemically similar“ (Gleason 1961: 80).

Man kann allerdings auch einen Schritt weiter gehen und auch seltenere Pluralformen wie *oxen* mit in die Analyse einbeziehen. Auch *-en* ist mit der Funktion

PLURAL verbunden, und dieses Minimalzeichen und die Gruppe /s ~ z ~ ɪz/ sind komplementär verteilt; daher wird das Minimalzeichen von Harris (1942: 175), Nida (1949: 44 f.) u. v. a. ebenfalls als Allomorph des abstrakten Pluralmorphems klassifiziert. Hier ist die Verteilung allerdings nicht phonologisch, sondern lexikalisch determiniert – und gleichzeitig unterscheidet sich *-en* formal deutlich von den drei anderen Allomorphen. Nida (1949: 44) stellt in diesem Zusammenhang trocken fest: „There are absolutely no limits to the degree of phonological difference between allomorphs.“

Es lohnt sich dennoch, beide Fälle zu trennen: Zum einen haben wir Allomorphe, die formseitig teildentisch sind. Das ist der Normalfall. Dann gibt es den „Extremfall“ (Bloomfield 1933: 215), dass kein Formzusammenhang zwischen den Allomorphen besteht; man kann Minimalzeichen wie *-en* PLURAL mit Bloomfield (1933: 215) als suppletiv bezeichnen. Wir beschränken uns in diesem Kapitel auf den Normalfall der Allomorphie, also auf formal ähnliche (nicht vollständig suppletive) Allomorphe – immerhin handelt es sich hier um eine graphematische Arbeit, und für eine systematische Untersuchung suppletiver Allomorphie bedürfte es u. a. erst einmal einer Aufstellung rivalisierender Affixe. Das bedeutet, dass Abweichungen von der Einheitlichkeit von Morphemen von vornherein relativ enge Grenzen gesetzt sind. Es bedeutet auch, dass vollständig suppletive Formen wie *be*, *am*, *was* oder *sein*, *bin*, *ist* jeweils nicht Allomorphe eines Morphems sind, sondern zu unterschiedlichen Morphemen gehören (die wiederum zu einem Lexem gehören).⁷²

Zu Morphemen werden also im Rahmen dieser Arbeit nur solche Minimalzeichen gruppiert, die dieselbe Funktion/Bedeutung haben und formseitig teildentisch sind.⁷³ Diese Morpheme werden auf ihre Eindeutigkeit und Einheitlichkeit

72 Nebenbei bemerkt: Für diese Beschränkung spricht, dass sprachliche Zeichen Verbindungen von Form und Funktion sind. Wenn man diesen Grundsatz ernstnimmt, dann kann nicht die Funktion allein den ‚Zuschnitt‘ des Zeichens bestimmen. Genau das passiert aber, wenn Funktionsgleichheit als definierendes Kriterium angesetzt wird. Alle Morpheme, die die morphologische Kategorie/das Merkmal PLURAL tragen, sind Allomorphe eines Morphems (Bedingung a. oben). Das hat eine Asymmetrie zur Folge: Mugdan (1994: 2546) stellt richtig fest, dass das Gegenstück von Morphemen Mengen von Minimalzeichen mit gleicher Form sind, etwa /z/ POSSESSIV, /z/ PLURAL und /z/ 3.Ps.Sg.Präs. Dafür brauchen wir aber keine Einheit ‚Morphem‘ – die jeweiligen Einheiten sind auf die Funktion reduzierbar. Die konkrete Form spielt nur insofern eine Rolle, als die einzelnen Allomorphe komplementär verteilt sein müssen (Bedingung b. oben). Möchte man sowohl die Form- als auch die Funktionsseite berücksichtigen, kommt man um eine formale Beschränkung nicht herum.

73 Ein solches Vorgehen ist in graphematisch-morphologischen Arbeiten recht verbreitet. So definiert etwa Gallmann (1990: 515) folgendermaßen: „Zwei Morpheme sind *Allomorphe* ein und desselben *Morphems*, wenn sie sich nur in bestimmten Signifiant- und/oder Signifié-Merkmalen

hin überprüft. Wir können diese beiden Parameter nun etwas präziser definieren als oben:

- (63) **Einheitlichkeit:** Die Form eines Morphems ist umso einheitlicher, je weniger Allomorphe das Morphem hat.
- (64) **Eindeutigkeit:** Die Form eines Morphems ist umso eindeutiger, je exklusiver sie das entsprechende Morphem repräsentiert. Exklusivität kann in zwei unterschiedlichen und logisch unabhängigen Beschränkungen gefasst werden:
1. Es gibt kein homographisches Morphem. Diesen Fall bezeichnen wir als *Eindeutigkeit im engeren Sinne*.
 2. Es gibt keinen homographischen Morphemteil, der als Instanz des Morphems interpretiert werden könnte. Diesen Fall bezeichnen wir mit Ryan (2016: 195) als *Distinktivität (distinctiveness)*.

Die Ausprägungen der Eindeutigkeit lassen sich am besten an einem Beispiel verdeutlichen. <-s> PLURAL verletzt im Englischen die erste Beschränkung: <-s> repräsentiert nicht nur <-s> PLURAL, sondern auch <-s> 3.Ps.SG.PRÄS; es ist also nicht eindeutig im Sinne der ersten Beschränkung. Hier sind mehrere Abstufungen vorstellbar. Der einfachste Fall ist, dass es kein homographisches Morphem gibt. In diesem Fall repräsentiert die Form das Morphem eindeutig im engeren Sinne. Das ist vor allem bei Flexionssuffixen, und hier vor allem im Deutschen, die Ausnahme; die meisten Formen repräsentieren mehrere Morpheme. Auch für <-s> im Englischen ist das der Fall, wie gerade festgestellt wurde; es repräsentiert sowohl das Morphem <-s> PLURAL als auch <-s> 3.Ps.SG.PRÄS. Das konkrete Morphem ist allerdings mithilfe der Wortart rekonstruierbar: Wenn wir aus dem syntaktischen Kontext wissen, dass *sighs* ein Verb ist, dann wissen wir, dass <-s> hier eine Instanz des Morphems <-s> 3.Ps.SG.PRÄS ist; ist *sighs* ein Substantiv, repräsentiert es <-s> PLURAL. Es ist zu überlegen, ob sich uneindeutige Morpheme in der Spezifität des Kontexts unterscheiden, der für die Desambiguierung benötigt wird: Im gerade präsentierten Fall wird nur die Wortart benötigt, im Fall von deutsch <-en> ist neben der Wortart noch mehr Kontext möglich: In *spielen* kann <-en> sowohl für das Suffix <-en> 1./3.Ps.PLURAL stehen als auch für das Suffix <-en> INFINITIV.

voneinander unterscheiden.“ Im selben Sinn versteht Nerius (Hg.) (2007: 157) unter Morphemvarianten „Formmodifikation[en] bei unveränderter Semantik“; vgl. auch die Konzeption des ‚sets‘ bei Cummings (1985: 45), die ebenfalls auf ‚similarity in their written form‘ beruht.

Das Suffix <-s> PLURAL verletzt, wie wir gesehen haben, die erste Beschränkung; es genügt aber der zweiten Beschränkung, es ist also distinktiv. Es gibt kaum einen homographischen Morphemteil, in diesem Fall ein Wortende, der als morphologisches -s (miss-)interpretiert werden könnte (siehe oben). Fast alle wortfinalen <s> sind für die Schreibung der beiden Morpheme reserviert.

Eindeutigkeit im engeren Sinne hängt bei den Flexionsaffixen davon ab, welche Suffixe konkret angesetzt werden; das wiederum hängt von der Art und Ausprägung der morphosyntakatischen Kategorien ab, die angenommen werden. Besonders im Deutschen ist unklar, wie weit z. B. das komplexe System der Nominalflexion vereinfacht werden kann (vgl. z. B. Wiese 1996; Eisenberg 2013a: 163 ff.). Aus diesem Grund wird auf eine systematische Untersuchung der Eindeutigkeit im engeren Sinne verzichtet.

Sowohl die Einheitlichkeit als auch die Eindeutigkeit sind graphemisch definiert: Wir bewerten die Bezüge zwischen Schrift und Morphologie. In beiden Fällen können wir die gewonnenen Daten aber zusätzlich mit den analogen morphophonologischen Daten vergleichen. Für die Einheitlichkeit bedeutet das: Wir vergleichen den Grad der morphographischen Einheitlichkeit mit dem der morphophonologischen Einheitlichkeit. Auf diese Weise können wir Bereiche identifizieren, in denen die Schrift z. B. phonologische paradigmatische Unterschiede überbrückt wie im Beispiel *Kind/Kinder* oben; genauso lassen sich (zumindest theoretisch) auch Bereiche identifizieren, in denen der Grad der phonologischen Einheitlichkeit höher ist als der Grad der graphematischen Einheitlichkeit. Grundsätzlich gibt es vier Möglichkeiten; sie sind in Tabelle 99 jeweils anhand eines deutschen und eines englischen Stamms illustriert:

Tab. 99: Die vier logisch möglichen Kombinationen graphematischer und phonologischer (Un-)Einheitlichkeit.

		phonologisch	
		einheitlich	uneinheitlich
graphematisch	einheitlich	bunt – bunt-e talk – talk-er	Kind – Kind-er sign – sign-al
	uneinheitlich	Kürbis – Kürbiss-e hit – hitt-ing	explod-ier[en] – Explos-ion explode – explos-ion

Phonologisch und graphematisch einheitliche Stämme sind der Standardfall (Zelle links oben); im Englischen gibt es aber wesentlich mehr Abweichungen von der graphematischen Einheitlichkeit (bei phonologischer Einheitlichkeit) als

im Deutschen (das ist die Zelle links unten). Phonologische Unterschiede können nun graphematisch überbrückt werden (Zelle rechts oben) oder graphematisch gespiegelt werden (Zelle rechts unten).

Für die Eindeutigkeit bedeutet das entsprechend, dass wir die morphographische Eindeutigkeit mit der morpho-phonologischen Eindeutigkeit vergleichen. Wir können z. B. fragen, wie häufig *ous* phonologisch mit homophonen Wortenden zusammenfällt (die Antwort ist: in vielen Fällen) oder wie häufig homophone Stämme im Englischen graphematisch differenziert werden (die Antwort: nicht systematisch, aber häufiger als im Deutschen). Erst auf diese Weise wird die ganze Leistung des Schriftsystems erkennbar. Die vier prinzipiellen Möglichkeiten phonologischer und graphematischer (In-)Variation sind in Tabelle 100 aufgeführt:

Tab. 100: Die vier logisch möglichen Kombinationen graphematischer und phonologischer (Un-)Eindeutigkeit.

		phonologisch	
		eindeutig	uneindeutig
graphematisch	eindeutig	Bett bed	Lid – Lied meat – meet
	uneindeutig	Band /bant/ – Band /bent/ lead /li:d/ – lead /led/	Bank bank

Auch hier ist phonologische und graphematische Eindeutigkeit der Standardfall: Eine phonologische Form entspricht einer graphematischen Form, und beide repräsentieren ein Lexem (Zelle links oben). Wenn nur die phonologische, nicht aber die graphematische Form zweier Lexeme differenziert wird, haben wir den (relativ seltenen) Fall homographischer Heterophone (Zelle links unten); der umgekehrte Fall, dass die Schrift mehr differenziert als die gesprochene Sprache, ist wohl häufiger – aber auch nicht systematisch (Zelle rechts oben). Schließlich gibt es auch den Fall, dass graphematisch und phonologisch nicht differenziert wird; hier handelt es sich um reine Homonyme (Zelle rechts unten).

Konstellationen wie in Tabelle 99 und 100 stehen im Mittelpunkt dieses Teilkapitels. Es ist wie folgt aufgebaut: Zuerst wird die Einheitlichkeit von Stämmen untersucht (4.1), dann ihre Eindeutigkeit (4.2). Dasselbe geschieht für die Affixe: Zuerst werden sie (jeweils getrennt nach Flexions- und Derivationsaffixen) auf Einheitlichkeit geprüft (4.3), dann auf Eindeutigkeit (4.4). Das Kapitel schließt mit einer Zusammenfassung und Diskussion der Ergebnisse (4.5).

4.1 Stämme: Einheitlichkeit

Wie einheitlich werden Stämme im Deutschen und Englischen geschrieben? Diese Frage ist zunächst rein graphematisch zu beantworten. Wir untersuchen, wie viele Fälle es gibt, in denen die graphematische Form eines Stamms variiert. Ist eine Form in allen Umgebungen stabil, wird der Stamm konstant verschriftet. Variiert er, kann gefragt werden, wie vorhersagbar diese Variation ist. Diese Art der Untersuchung ist grundsätzlich von Interesse, wie bspw. Coates (2006: 319) im Zusammenhang mit morphophonemischer Variation feststellt:

Much of interest could be said about spelling variation of parallel and equivalent type [...] 'Orthographic MP' [morphophonemics, KB] could be termed morphographics, though this term has no widespread currency, largely because the orientation of most modern linguistics leans toward speech, not writing, and tends to bypass the entire issue.

Wenn die Frage der graphemischen Einheitlichkeit geklärt ist, können phonologische Stammalternationen herangezogen werden, um zu untersuchen, wie die geschriebene Sprache mit diesen Alternanten umgeht. Wird die Alternation überbrückt, hat ein Stamm also im Geschriebenen eine konstantere Form als im Gesprochenen, liegt ein Fall von expliziter Stammkonstanz (Fuhrhop/Barghorn 2012) vor. Das Gesprochene dient hier (wie in den anderen Teilkapiteln) als Folie, vor der die spezifische Leistung der Schrift besser beurteilt werden kann. Als Datengrundlage werden vor allem CELEX sowie (für das Deutsche) das IDS-Korpus Cosmas II (<https://cosmas2.ids-mannheim.de>, Stand: 22. 11. 2018) verwendet. Die Aufstellungen haben nicht den Anspruch, exhaustiv zu sein und jede idiosynkratische Ausnahme von der Einheitlichkeit aufzulisten. Vielmehr kann es nur darum gehen, die hauptsächlichen Muster zu erfassen – die aber dafür weitestgehend vollständig.

Wir können hier zunächst verschiedene Fälle von Variation unterscheiden. Stämme variieren abhängig vom morphologischen Kontext (65a–c) oder unabhängig davon (65d).

(65) **deutsch**

- (65a) fit – fitter, *fitt, *fiter
 (65b) Tag – täglich, *taglich
 (65c) Liebe – Liebesbrief, *Liebebrief
 (65d) Foto, Photo

englisch

- rip, ripped, *ripp, *ripped
 happy – happiness, *happyness
 child – child'splay, *childplay
 color, colour

In all diesen Fällen variiert die Form des jeweiligen Stamms. Die kontextuell determinierten Varianten (65a–c) können wiederum nach dem morphologischen Prozess klassifiziert werden, der im Spiel ist: Stammvariation findet sich sowohl in der Flexion (65a) als auch in der Wortbildung, und hier sowohl in der Deriva-

tion (65b) als auch in der Komposition (65c); innerhalb von Flexionsparadigmen ist sie im Deutschen selten (vgl. z. B. Maas 2000: 330), in der Derivation gibt es – besonders bei lexikalisierten Bildungen – zahlreiche idiosynkratische Fälle, vgl. z. B. Nerius (Hg.) (2007: 162 ff.)

Es gibt auch Variation, die anders als grammatisch gebunden ist, z. B. diatopisch (z. B. BrE <colour>, AmE <color>) oder diachron (z. B. <Photo>, <Foto>). Diese Variation hat aber einen kategorial anderen Charakter als die anderen Variationstypen – und wir wissen so gut wie nichts darüber. Ist die Wahl freier Varianten bspw. intraindividuell stabil? Oder schwanken Schreiber zwischen <Kalzium> und <Calcium>, zwischen <Portemonnaie> und <Portmonee>? Aus diesem Grund wird die freie Variation als Abweichung von der Einheitlichkeit ausgeklammert. Das soll nicht heißen, dass sie theoretisch uninteressant ist – im Gegenteil; sie unterscheidet sich aber deutlich von den übrigen Variationstypen. Im Folgenden wird zuerst Variation diskutiert, die sich aus dem Kontrast flektierter Formen und freier Stämme ergibt (4.1.1), dann geht es um Variation, die sich analog bei Produkten der Wortbildung ergibt (4.1.2).

4.1.1 Flexion

Im Deutschen werden Stämme größtenteils invariant verschriftet (vgl. z. B. Eisenberg 2013a: 310). Das betrifft vor allem die Flexionsformen. Es gibt hier nur vergleichsweise wenige Ausnahmen, die in (66) exemplarisch aufgelistet sind:

- (66a) *nehm[en] – nimmt, *nehmt, *nimm[en]*
- (66b) *Apfel – Äpfel, *Apfel_{pl}, *Äpfel_{sg}*
- (66c) *fit – fitter, *fitt, *fiter*
- (66d) *liken – geliked, gelikt, geliket*
- (66e) *Google – googel[n], google[n]*
- (66f) *segel[n] – (ich) segle, segel, segele*
- (66g) *Zyklus – Zyklen, *Zyklusse, Zykluse*
- (66h) *Kürbis – Kürbisse, *Kürbise, *Kürbisse*

Die erste Gruppe von Ausnahmen hat ihren Ursprung in den unregelmäßigen und starken Verben (vgl. für das Folgende Fuhrhop 2017: 46 ff.). Neben dem charakteristischen Vokalwechsel (z. B. <gebe>/<gab>) ist es vor allem der Wechsel zwischen einfachen und Doppelkonsonanten (z. B. <greift>/<griff>), der zur Variation der Stammform führt. <h>- und <ß>-Schreibungen, die nicht in allen Stammformen vorkommen, sind demgegenüber selten (z. B. <stehe>/<stand>, <esse>/<aß>). In diesen Fällen weicht das deutsche Schriftsystem davon ab, Stämme einheitlich zu

repräsentieren. Das geschieht, wie Fuhrhop plausibel vermutet, aus phonographischen Gründen: Eine Form wie *⟨greifte⟩ ist zwar maximal einheitlich, was Stamm und Suffix angeht – sie überstrapaziert aber die Korrespondenz zur phonologischen Form /grɪf/ und ist daher nicht möglich. Interessant ist in diesem Zusammenhang Fuhrhops Beobachtung, dass dennoch die meisten Formen starker und unregelmäßiger Verben Hinweise auf ihre Wortart liefern. Wortfinale Doppelkonsonanten und wortfinale ⟨h⟩ z. B. finden sich nur in flektierenden Wortarten; ist das betreffende Wort kleingeschrieben, kann es sich nur noch um ein Verb oder ein adverbial oder prädikativ verwendetes Adjektiv handeln. Diese Informationen sind im Wortende ‚sichtbar gemacht‘ und stehen Lesern zur Verfügung (Fuhrhop 2017: 57).

Die zweite Gruppe beinhaltet Lexeme, bei denen Flexionskategorien nicht-segmental durch Umlaut markiert werden (66b). Formal variieren die Stämme im Singular und im Plural. Das ist natürlich funktional ein Vorteil, denn beide unterscheiden sich – und dieser Unterschied wird markiert. Instruktiv ist hier der Vergleich mit den korrespondierenden phonologischen Formen: In ⟨Apfel⟩/⟨Äpfel⟩, ⟨Strom⟩/⟨Ströme⟩, ⟨Hut⟩/⟨Hüte⟩ ist der Unterschied phonologisch relativ groß (/a/ – /ɛ/, /o/ – /ø/, /u/ – /y/). In der Schrift wird er überbrückt: Die Stämme sind formal intakt, die Änderung im Stammvokal wird zusätzlich durch das Trema markiert. Es markiert morphologisch die Kategorie ‚Plural‘, zum Teil gemeinsam mit einer segmentalen Markierung wie bei ⟨Hüte⟩ (*extended exponence* in Matthews’ 1991: 180 Terminologie).

Die dritte Gruppe von Ausnahmen umfasst einige einsilbige Adjektive, die aus dem Englischen entlehnt sind (66c). In der Grundform ist der finale Konsonant einfach, in allen Flexionsformen ist er verdoppelt (⟨hip⟩ – ⟨hippes⟩) – wenn die entsprechenden Wörter denn überhaupt flektiert werden, also attributiv verwendet werden. Eisenberg (2011: 238 f.) vermutet, dass das nur „gelegentlich“ geschieht, und wenn, dann im Gesprochenen: „Die Orthographie wird zum Integrationshindernis“, weil bei gegebener Grundform ⟨hip⟩ nicht sowohl der Stamm konstant bleiben kann (⟨hiper⟩) als auch das Silbengelenk verschriftet (⟨hipper⟩). Diese Aussage trifft zumindest für *fit* und *hip* nicht zu: Die flektierte Form *fitter* bspw. ist im Cosmas-Korpus W-Öffentlich häufiger als die native Form *fader* (3.852 vs. 2.030 Treffer, initiale Großschreibung wurde ausgeschlossen). Eine vollständige graphematische Integration der betreffenden Wörter ist erreicht, wenn die Doppelkonsonanten auch in der Grundform auftreten (⟨hipp⟩, ⟨fitb⟩).⁷⁴

⁷⁴ Innerhalb der flektierten Adjektive ist der Stamm natürlich konstant – die ‚Grenze‘ verläuft zwischen attributiv verwendeten Adjektiven (ein *fitter* Läufer) und adverbial sowie prädikativ verwendeten (Er ist *fit*).

Auch die vierte Gruppe besteht aus Anglizismen, hier allerdings Verben (66d). Sie haben ein stammfinales ⟨e⟩. Das unterscheidet sie von allen nativen Verbformen und ist gleichzeitig ein Grund für die Variation. Diese Variation findet hier – anders als in den übrigen Beispielen – auch innerhalb von Paradigmenzellen statt. Teilweise enthält der Stamm das finale ⟨e⟩ des englischen Verbs (⟨er fakēt⟩), teilweise nicht (⟨er fakt⟩). Das Problem scheint hier vor allem phonographisch zu sein: Für einfaches ⟨a⟩ wie in ⟨fakt⟩ ist die Korrespondenz zu /ɛɪ/ untypisch. Die Alternative, ⟨fakēt⟩, suggeriert phonologische Zweisilbigkeit. Beides ist problematisch, deswegen ist der Anteil an Variation hier relativ hoch. Schauen wir uns exemplarisch die Form des Partizips II an. Wir finden drei Schreibungen: In der Variante ⟨geliket⟩ sind Stamm und Affix einheitlich verschriftet. In der Variante ⟨geliked⟩ ist der Stamm intakt, das Affix aber uneinheitlich – als Zirkumfix für das Partizip haben wir sonst ⟨ge...t⟩ (bzw. ⟨ge...en⟩ bei starken Verben). Und in der Variante ⟨gelikt⟩ ist schließlich das Affix einheitlich, der Stamm aber nicht – das finale ⟨e⟩, das ja die Verbstämme vor allem von ihren nativen Kollegen unterscheidet, ist getilgt. Wie sind diese Varianten nun verteilt? Eine kurze Abfrage im Cosmas-Korpus W-Öffentlich ergibt die Zahlen in Tabelle 101 für die Schreibungen der Verben TUNE[N], STYLE[N], TIME[N], SHAKE[N], LIKE[N], FAKE[N] und PRIME[N]:⁷⁵

Tab. 101: Welche Partizipialformen sind wie häufig für sieben Verben aus dem Englischen?

1. Variante: ⟨ge⟩ + Stamm + ⟨t⟩; 2. Variante: ⟨ge⟩ + Stamm + ⟨d⟩; 3. Variante: ⟨ge⟩ + Stamm ohne finales ⟨e⟩ + ⟨t⟩. Datengrundlage: Cosmas-Korpus W-Öffentlich.

	ge+...e+t	ge+...e+d	ge+...+t
tun(e)	5	66	589
styl(e)	1	15	5.311
tim(e)	47	61	445
shak(e)	1	18	26
lik(e)	26	140	32
fak(e)	48	115	84
prim(e)	2	15	9

In keinem Fall ist die vollständig einheitliche Variante (Typ ⟨getunet⟩) die häufigste. Bei TUNE[N], STYLE[N] und TIME[N] ist die frequenteste Schreibung die mit einheitlichem Affix, aber Variation in der Stammform (Typ ⟨getunt⟩). Die Tilgung

⁷⁵ Zu ähnlichen Ergebnissen kommt eine Korpusrecherche des IDS von 2013, vgl. http://hypermedia.ids-mannheim.de/call/public/fragen.ansicht?v_id=140 (Stand: 22. 11. 2018).

des finalen ⟨e⟩ macht als Integrationsprozess Sinn (man kauft sich damit aber wie erwähnt phonographische Probleme ein). Bei LIKE[N], FAKE[N] und PRIME[N] ist die Variante mit intaktem Stamm und finalelem ⟨d⟩ am häufigsten (Typ ⟨gefaked⟩). Diese Schreibung erscheint auf den ersten Blick absurd: Hier wird offenbar die englische Partizipialform genommen und mit dem nativen Präfix *ge-* verbunden. Das Problem an dieser Analyse ist, dass sie für ⟨geshaked⟩ nicht aufgeht. Hier lautet die englische Partizipialform *shaken*, für ⟨geshaken⟩ findet sich im Korpus aber kein Beleg.

Damit kann ⟨ge...d⟩ als Variante des Partizipialaffixes interpretiert werden, die für englische Verbstämme reserviert ist. Die Form hat den Vorteil, dass das finale ⟨ed⟩ nicht zwangsläufig als phonographisch silbisch interpretiert werden muss – das ist es ja im Englischen meist auch nicht (vgl. ⟨tuned⟩, ⟨styled⟩, ⟨timed⟩, ⟨liked⟩, ⟨faked⟩, ⟨primed⟩). Es handelt sich hier also um einen Kompromiss: Die Formen sind einerseits durch das initiale ⟨ge⟩ integriert, andererseits markiert das finale ⟨ed⟩ sie als Anglizismen, mit allen phonographischen Konsequenzen. Sowohl der Wechsel zwischen einfachen und verdoppelten Konsonanten als auch die ‚Tilgung‘ des finalen ⟨e⟩ führen auch im Englischen zu Variation der Stammform (siehe unten). Beide Alternationen hängen zusammen und haben ihren Ursprung in der phonographischen Tatsache, dass die Vokalqualität (auch silbenstrukturell) kodiert wird. Die Integrationsprobleme, die an Formen wie *geliiked* sichtbar werden, haben ihren Ursprung im Englischen.

Auch die Wörter der fünften Gruppe (66e) sind Anglizismen. Sie enden im Englischen auf ⟨le⟩, die Form im Deutschen variiert aber: ⟨googlen⟩ und ⟨gegooglet⟩ stehen neben ⟨googeln⟩ und ⟨gegoogelt⟩. Die jeweils erste Variante entspricht der Schreibung des Stamms im Englischen; die jeweils zweite Variante mit ‚invertiertem‘ ⟨el⟩ entspricht dem nativen System. Wir schauen uns hier wiederum im Cosmas-Korpus W-Öffentlich an, wie die Formen der Verben RECYCL[EN] und GOOGL[EN] verteilt sind – die Partizipialform (die im Falle von RECYCL[EN] homonym ist mit der Form der 3.Ps.Sg.Ind.Akt.Präs) und die Infinitivform (die homonym ist mit der Form der 1./3.Ps.Plur.Akt.Präs). Tabelle 102 zeigt die Anzahl der jeweiligen Formen:

Tab. 102: Welche Partizipial- und Infinitivformen sind wie häufig für zwei Verben aus dem Englischen? Datengrundlage: Cosmas-Korpus W-Öffentlich.

	(ge)...elt	(ge)...let	...eln	...len
recyc(le)	4.188	141	2.210	611
goog(le)	2.224	208	4.397	1.553

Bei beiden Lexemen ist im verwendeten Korpus die integrierte Form (auf <el>) häufiger. Interessanterweise unterscheiden sich die Anteile aber abhängig von der Flexionskategorie: Bei den Partizipien ist der Anteil integrierter Formen deutlich höher als bei den Infinitivformen. Es bietet sich an, hier einmal alle Flexionsformen zu überprüfen, und zwar möglichst in unterschiedlichen Korpora; das würde an dieser Stelle aber zu weit führen.

Die Wörter der nächsten Gruppe (66f) zeichnen sich dadurch aus, dass sie in der 1.Ps.Sg.PRÄS. AKT. variieren: Neben dem strukturell erwartbaren (ich) <segele> (Stamm (+ Flexionssuffix)) findet sich die Form (ich) <segle>. Um zu überprüfen, wie genau die einzelnen Formen verteilt sind, werden für einige zufällig ausgewählte Verben auf <el> drei Formen erhoben: Die Form mit intaktem Stamm und intaktem Suffix (Typ <segele>), die Form mit intaktem Stamm ohne Suffix (Typ <segel>) sowie die Form mit Suffix, aber reduziertem Stamm (Typ <segle>). Die Zahlen stammen aus dem Cosmas-Korpus W-Öffentlich.⁷⁶ Die Ergebnisse sind in Tabelle 103 aufgeführt:

Tab. 103: Anzahl der Belege für drei Varianten der 1.Ps.Sg.Akt.Präs. bei 15 Verben auf -el: <el>, <ele> und <le>. Datengrundlage: Cosmas-Korpus W-Öffentlich.

	...el	...ele	...le	gesamt	% ...le
zweif(el[n])	131	316	2.595	3.042	85 %
samm(el[n])	57	199	1.569	1.825	86 %
wechs(el[n])	25	147	574	746	77 %
hand(el[n])	10	174	386	570	68 %
seg(el[n])	11	22	66	99	67 %
bett(el[n])	8	28	57	93	61 %
schmunz(el[n])	1	19	59	79	75 %
ang(el[n])	8	20	48	76	63 %
blinz(el[n])	5	14	40	59	68 %
tüft(el[n])	5	10	32	47	68 %
wack(el[n])	6	7	11	24	46 %
kritz(el[n])	2	10	3	15	20 %
schmeich(el[n])	0	1	14	15	93 %
grus(el[n])	4	8	2	14	14 %
tork(el[n])	0	0	11	11	100 %

⁷⁶ Um unerwünschte Ergebnisse möglichst auszuschließen, wurde nach Verbindungen von Personalpronomen und Verb gesucht – also *ich segle*, *ich segele*, *ich segel*.

Mit der Ausnahme von WACKEL[N], KRITZEL[N] und GRUSEL[N] sind die Formen mit reduziertem Stamm (Typ <segle>) jeweils die frequentesten Varianten; die Formen mit intaktem Stamm, aber ohne Flexionssuffix (Typ <segel>) sind bei jedem Verb die seltensten. Bei den Anteilen der Formen mit reduziertem Stamm scheint es einen Effekt der Frequenz zu geben: Je häufiger die Formen in der Summe sind (Spalte ‚gesamt‘), desto höher ist der Anteil der Formen mit reduziertem Stamm (<segle>). Von dieser Generalisierung gibt es einige Ausnahmen, besonders SCHMEICHEL[N] und TORKELE[N], die einen wesentlich höheren Anteil an reduzierten Formen aufweisen, als aufgrund ihrer Frequenz zu erwarten wäre. Die Variation in der Realisierung des Suffixes wird bei der Diskussion der Einheitlichkeit von Flexionssuffixen behandelt (4.3.1).

Zurück zur Liste in (66) oben. Die letzten beiden Fälle stehen stellvertretend für eine Reihe von Latinismen und Gräzismen. In Typ (66g) hat das Substantiv Stammflexion: Im Plural wird das finale <us> der Singularform durch andere Formative ersetzt, vor allem durch *-en*, das auch im nativen Bereich als Pluralsuffix fungiert, *Zyklus – Zyklen* (vgl. Eisenberg 2011: 224 f.). Zum Teil gibt es hier aber auch andere Alternationen, vgl. *Atlas – Atlanten*, *Korpus – Korpora*. All diese Fälle sind für die Einheitlichkeit von Stämmen unproblematisch, wenn man sie als Stammflexion analysiert: Der gebundene Stamm *Zykl-* wird einheitlich verschriftet.

Für den Typ (66h) gilt das nicht im selben Maß. Hier wird nicht-suppletiv flektiert: Der stamminale Doppelkonsonant der flektierten Form (*Kürbisse*) wird nicht auf die Grundform übertragen (**Kürbiss*), die Form des Stamms alterniert also. Das bedeutet: „Das morphologische Prinzip der Schreibung ist, was Geminatio betrifft, aufgegeben“ (Eisenberg 2011: 343). Es geht hier vor allem um solche Stämme, für die die *s*-Flexion nicht zur Verfügung steht; das sind vor allem Stämme mit finalelem <s>. Sie gehen „den Weg in die starke Flexion mit Pluralformen wie *Krokusse*, *Fidibusse*, *Omnibusse*, *Zirkusse*“ (Eisenberg 2011: 224). Das betrifft einerseits Lexeme, die nicht über Stammflexion verfügen wie *Kürbis* – es gibt hier keine Alternative wie **Kürben* (Beispiele in (67a)). Andererseits sind hier solche Lexeme einschlägig, bei denen sich neben der Stammflexion die Grundformflexion als Variante herausbildet wie *Globusse* neben *Globen* (Beispiele in (67b)):⁷⁷

⁷⁷ Neben der Stammflexion wie in *Diskus – Disken* tritt z. T. auch Nullplural auf wie in *Krokus – Krokus*. Entscheidend ist, dass in beiden Fällen die starke Flexion gewählt werden kann.

- (67a) Ananas – Ananasse, Albatros – Albatrosse, Bambus – Bambusse,
 Bus – Busse, Fidibus – Fidibusse, Iltis – Iltisse, Kirmes – Kirmesse(n),
 Kürbis – Kürbisse, Lotus – Lotusse, Penis – Penisse
- (67b) Atlas – Atlasse, Campus – Campusse, Diskus – Diskusse, Fötus – Fötusse,
 Globus – Globusse, Kaktus – Kaktusse, Korpus – Korpusse,
 Krokus – Krokusse

In all diesen Fällen sind Stämme also nicht einheitlich – zum Teil gibt es zur jeweiligen Schreibung Alternativen (67b), zum Teil nicht (67a). Die Listen beanspruchen keine Vollständigkeit; besonders die Liste in (67a) enthält aber wohl den Großteil der einschlägigen Wörter. Dieser Hinweis ist wichtig, denn so ist gut zu sehen, dass wir es hier in Bezug auf das Gesamtsystem mit relativ wenigen Ausnahmen zu tun haben.

Wir können nun generell fragen, was der Status der Wörter in den Gruppen (66a–h) ist. Von den unregelmäßigen und starken Verben in (66a) gibt es knapp 200 (vgl. z. B. Fuhrhop 2017). Die Gruppen (66c–f) sind sicher marginal – alles in allem gibt es hier wohl kaum zwei Dutzend Fälle. Die Gruppen (66g, h) sind es schon weniger, besonders die Gruppe (66g) ist unter den Latinismen weit verbreitet – die Fälle sind aber für die Einheitlichkeit unproblematisch, solange sie als Instanzen von Stammflexion aufgefasst werden. Die einzige *systematische* Varianz in der Stammschreibung geschieht in Gruppe (66b) – hier bleibt aber ja, wie oben gezeigt wurde, der Stamm intakt, und das Trema markiert (allein oder mit Pluralsuffix) die Kategorie PLURAL. Insgesamt gibt es also im Deutschen tatsächlich kaum Ausnahmen vom Prinzip, Stämme einheitlich zu verschriften.

Bislang haben wir uns Abweichungen von der Einheitlichkeit angesehen – davon gibt es im Deutschen nur wenige. Wir können nun, wie oben angekündigt, den Blick auf die Phonologie lenken. Wo gibt es morpho-phonologische Alternationen, und wie geht die Schrift damit um? Die systematischen Alternationen finden sich in (68):

- (68a) Kind (/kɪnt/) – Kinder (/kɪndɐ/)
 (68b) Apfel (/apfl/) – Äpfel (/ɛpfl/)
 (68c) Bach (/bax/) – Bäche (/bɛçə/)

Der Typ in (68a) ist das Paradebeispiel morphologischer Schreibungen: Die Schrift überbrückt die Auslautverhärtung. Graphematisch ist der Stamm stabil, phonologisch variiert er. Ähnliches gilt für den Typ in (68b), dem wir oben bereits begegnet sind. Hier variiert das entsprechende Vokalphonem relativ stark (/a/ – /ɛ/), die graphematischen Formen unterscheiden sich nur durch das Trema. Und schließlich alternieren auch die Formen in (68c) – zumindest pho-

netisch, denn phonologisch werden die Phone [x] und [ç] aufgrund ihrer komplementären Distribution und ihrer phonetischen Ähnlichkeit meist zu einem Phonem zusammengefasst.

Die Wörter in (68a, b) könnten wir auch anders verschriften (*⟨Kint⟩, *⟨Epfel⟩); die Stammkonstanz ist hier also explizit im Sinne von Fuhrhop/Barghorn (2012). Es gibt bei deutschen Stämmen also nicht nur wenige Ausnahmen von der Einheitlichkeit – es könnte potenziell wesentlich mehr Ausnahmen geben, wenn phonographischer verschriftet werden würde. Das ist ein weiteres Indiz dafür, dass die einheitliche Schreibung von Stämmen ein wichtiges Strukturprinzip im Deutschen ist.

Im Englischen gibt es demgegenüber wesentlich mehr Abweichungen, also mehr Variation in der Form von Stämmen in Flexionsformen. Die Liste in (69) zeigt exemplarische Falltypen:

- (69a) think – thought, *thanked
- (69b) knife – knives, *knifes, *knive
- (69c) lady – ladies, *ladys, *ladie
- (69d) rip – ripped, *ripp, *riped
- (69e) dine – dining, *dineing, *din

Auch im Englischen gibt es bei den unregelmäßigen Verben (69a) Variation in der Stammform (vgl. Fuhrhop 2017: 58 ff.). Meist variiert wie im Deutschen der Vokal (z. B. *ring* – *rang* – *rung*), z. T. alternieren in einigen Flexionsformen aber auch die Konsonanten (z. B. *catch* – *caught*). Viele dieser Konsonanten haben keine phonographische Korrespondenz wie bspw. das ⟨ɹ⟩ in *could*. Die Variation in der Stammform ist lexikalisch determiniert; sie ist nur in seltenen Fällen aus der graphematischen (oder der phonologischen) Struktur ableitbar.

Ebenfalls idiosynkratisch sind Fälle wie (69b), bei denen die Alternation /f/ – /v/ verschriftet wird. Bloomfield (1933: 214) gibt hier die folgende exhaustive Liste an:

- (70) knife – knives, wife – wives, life – lives, calf – calves, half – halves,
thief – thieves, leaf – leaves, sheaf – sheaves, beef – beeves,
loaf – loaves, elf – elves, shelf – shelves, hoof – hooves, scarf – scarves,
wharf – wharves⁷⁸

⁷⁸ Bloomfield bezeichnet die letzten drei Belege als Varianten, die einige Sprecher verwenden. Anders als *roof* – *rooves* und *dwarf* – *dwarves*, die er ebenfalls in diesem Zusammenhang nennt,

Interessant ist hier, dass graphematisch in der Mehrzahl der Fälle ⟨f⟩ nicht mit ⟨v⟩, sondern mit ⟨ve⟩ alterniert (z. B. ⟨half⟩ – ⟨halves⟩). ⟨v⟩ ist kein gutes Stammende im Englischen (vgl. oben 3.2.1.5), und das gilt auch für Stammformen, die nur gebunden vorkommen (in diesem Fall mit finalelem ⟨s⟩). Eine Konsequenz von Schreibungen wie ⟨leaves⟩ ist, dass Stamm und Suffix gut segmentierbar sind. In einer hypothetischen Schreibung *⟨leavs⟩ wäre das weit weniger der Fall, weil die Form ⟨leav⟩ die Wohlgeformtheitsanforderungen an englische Stämme verletzt (Berg et al. 2014: 292 f.).

Die Alternationen in (69c–e) sind, anders als die ersten beiden Fallgruppen, regelmäßig. Wenn wir die graphematische (und z. T. morphologische Struktur) der Grundform kennen, können wir die Alternationen vorhersagen. Es geht um den Wechsel zwischen ⟨y⟩ und ⟨i⟩ (69c), die Konsonantenverdoppelung (69d) und die ⟨e⟩-Tilgung (69e).

Zunächst zum Wechsel zwischen ⟨y⟩ und ⟨i⟩, der ja oben schon angesprochen wurde (vgl. für das Folgende Berg 2013; Berg et al. 2014). In Stämmen mit finalelem ⟨y⟩ in der Grundform alterniert das ⟨y⟩ in flektierten Formen mit ⟨i⟩/⟨ie⟩. Keine Alternation findet statt, wenn a) vor ⟨y⟩ ein weiteres Vokalgraphem auftritt (z. B. *boys*, *frayed*), b) das Suffix mit ⟨i⟩ beginnt (z. B. *crying*, *copying*) oder c) der Stamm ein Eigenname ist (z. B. *the two Germanys*, vgl. Bauer/Lieber/Plag 2013: 52).⁷⁹ Wenn das Suffix mit ⟨e⟩ beginnt, alterniert ⟨y⟩ mit ⟨i⟩ (z. B. *cry* – *cried*), ansonsten mit ⟨ie⟩ (z. B. *cry* – *cries*). Das führt wie oben bei der Alternation ⟨f⟩ – ⟨ve⟩ dazu, dass Formen wie ⟨cries⟩ aus einem ‚legitimen‘ Stamm und einem Suffix bestehen; würde hier ⟨y⟩ mit einfachem ⟨i⟩ alternieren, wäre die Form *⟨cris⟩ sehr viel schwe-

sind die letzten drei Belege in (70) aber heute die frequenteren Varianten, wie eine Abfrage bei Google Ngrams zeigt (<http://books.google.com/ngrams>, Stand: 22. 11. 2018).

79 Die Ausnahmen von der Alternation lassen sich graphematisch und morphologisch begründen. Würde ⟨y⟩ auch in Formen wie *frayed* alternieren, hätten wir Schreibungen wie ⟨fraied⟩. Drei oder mehr aufeinanderfolgende Vokalgrapheme sind im Englischen aber sehr selten (vgl. Berg 2013: 393, Fn. 9), und in genau solchen Umgebungen kommt auch morphemintern ⟨y⟩ vor, vgl. z. B. *mayor*, *crayon*. Offenbar hat ⟨y⟩ hier noch eine weitere Funktion, und zwar Folgen von Vokalgraphemen aufzubrechen.

Auch die zweite Ausnahme der Alternation ist graphematisch begründbar: Würde ⟨y⟩ vor *-ing* alternieren, würde das zu Schreibungen wie ⟨crieing⟩ oder ⟨criing⟩ führen. Die erste Schreibung enthält drei aufeinanderfolgende Vokalgrapheme, die zweite die Folge ⟨ii⟩, die im Englischen marginal ist (z. B. *alibiing*).

Die dritte Ausnahme hängt mit dem besonderen Status von Eigennamen zusammen. Die graphematische Stabilität von Eigennamen scheint ein wichtiges Schreibprinzip zu sein. Das ist auch an der Tatsache ablesbar, dass sich allerlei synchron längst überholte Schreibungen in Eigennamen halten (vgl. *Mecklenburg*, *Coesfeld*, *Luther*).

rer als Flexionsform zu erkennen, denn ⟨i⟩ kommt im Englischen in freien Stämmen nicht final vor.

Formen wie ⟨cried⟩ sind auf den ersten Blick Gegenbeispiele; sie lassen sich in eine Stammform ⟨cri⟩ und das Suffix ⟨ed⟩ segmentieren. Diese Verteilung lässt sich aber mit einer graphematischen Beschränkung erfassen, die ⟨ee⟩-Cluster an Morphemgrenzen verbietet (*⟨crieed⟩). Diese Beschränkung wird auch außerhalb der Beschreibung der ⟨y⟩/⟨i⟩-Alternation benötigt, vgl. z. B. ⟨die⟩/⟨died⟩/*⟨dieed⟩ (dieselbe Beschränkung wird auch gleich bei der Diskussion der ⟨e⟩-Tilgung und Konsonantenverdoppelung eine Rolle spielen, vgl. ⟨fake⟩/⟨faked⟩/*⟨fakeed⟩).⁸⁰

Warum alternieren die Stammformen hier überhaupt? ⟨y⟩ kommt im Englischen vor allem wortfinal vor, deswegen liegt es nahe, diese Verteilung als Funktion zu interpretieren: ⟨y⟩ markiert Wortenden. Wenn das Wort ein freier Stamm ist (z. B. *cry*), der flektiert wird, dann ist ⟨y⟩ in dieser Flexionsform nicht mehr final und alterniert daher (z. B. *cries*). In Komposita alterniert ⟨y⟩ nicht, bei Ableitungen teilweise (weiter 4.1.2).

Die Alternationen, die sich in der An- und Abwesenheit von finalem ⟨e⟩ sowie in einfachen und verdoppelten Konsonanten unterscheiden (69d, e), hängen eng zusammen; sie sind gewissermaßen Kehrseiten einer Medaille (vgl. z. B. Carney 1994: 129 f.; Palmer/Huddleston/Pullum 2002: 1577; Aronoff/Koch 1996: 256). Der Ursprung für beide Schreibungen ist phonographisch (vgl. für das Folgende Berg 2013: 394). Er liegt in einem Dilemma begründet: Wenn Stämme mit finalem ⟨e⟩ flektiert werden, können nicht sowohl der Stamm als auch Suffix einheitlich sein. Das würde zu Schreibungen wie *⟨hopeed⟩ führen, die sowohl segmental-phonographisch als auch suprasegmental-phonographisch irreführend wären: ⟨ee⟩ korrespondiert normalerweise mit /i/ (z. B. *feel*), und der komplexe Kern in der zweiten graphematischen Silbe (*⟨hopeed⟩) suggeriert, dass wir es mit einem korrespondierenden Jambus zu tun haben (analog zu z. B. ⟨vener⟩). Konsequenterweise ist ⟨hoped⟩ die einzige Möglichkeit.

Das hat aber wiederum zur Folge, dass Stämme mit einfachem finalem Konsonanten (wie ⟨hop⟩) anders verschriftet werden müssen, wenn sie flektieren: Eine einheitliche Schreibung von Stamm ⟨hop⟩ und Suffix ⟨ed⟩ führt hier zu systematischer Homographie mit flektierten Wörtern mit finalem ⟨e⟩ wie *hope*. Die Schreibung *⟨hoped⟩ steht daher als Flexionsform von *hop* nicht zur Verfügung. Die Lösung ist hier, den finalen Konsonanten zu verdoppeln (⟨hopped⟩).

⁸⁰ Eine alternative Analyse von *cries* ist, ⟨cri⟩ als Stammform anzusetzen und ⟨es⟩ als Form des Suffixes (z. B. Bauer/Lieber/Plag 2013: 52). Diese Analyse kann nicht die Parallelität der Alternationen vom Typ *thief* – *thieves* erfassen, wo das ⟨e⟩ ebenfalls die Stammform ‚schützt‘.

Beide Mechanismen – ‹e›-Tilgung und Konsonantenverdoppelung – führen nun zu Variation in der Stammform. Beide können graphematisch/morphologisch beschrieben werden, und bei beiden gibt es relativ wenige Ausnahmen. Im Folgenden sollen die Mechanismen und ihre Ausnahmen kurz skizziert werden. Die Verdoppelung von finalen Konsonanten beruht auf der graphematischen Form und der morphologischen Struktur der Basis. Die einfachen finalen Konsonantengrapheme ‹b, d, f, g, k, l, m, n, p, r, s, t, v, z› werden verdoppelt, wenn a) vor ihnen ein einfaches Vokalgraphem steht; b) ein Suffix mit initialem Vokalgraphem folgt; c) stammfinal eine einsilbige Wurzel wie *sit*, *drop* etc. auftritt (Berg 2016: 457). Damit wird erfasst, dass in ‹keeping›, ‹sits› und ‹hammering› die stammfinalen Konsonanten nicht verdoppelt werden: In ‹keeping›, *‹keeping› folgt der Konsonant zwei Vokalgraphemen; in ‹sits›, *‹sitts› ist das Suffix nicht vokal-initial; in ‹hammering›, *‹hammerring› ist die Wurzel nicht einsilbig. Der Verweis auf die Wurzel kann alternativ auch als Alignment-Beschränkung formuliert werden: Die Wurzel muss einsilbig sein, und ihr rechter Rand muss mit dem rechten Rand des Stamms zusammenfallen. Auf diese Weise kann erklärt werden, warum Formen wie ‹babysitting› und ‹airdropping› Doppelkonsonanten enthalten, Formen wie ‹visiting› und ‹lingering› aber nicht (*‹visitting›, *‹lingerring›). Man kann mit dem Bezug auf finale einsilbige Wurzeln auch Variation wie bspw. ‹worshipping›/‹worshipping› erklären. Diese Variation hängt an der morphologischen Analyse: Wird ‹ship› als einsilbige Wurzel analysiert, ergibt sich die Schreibung mit Doppelkonsonant; ist ‹worship› die Wurzel, wird der Konsonant nicht verdoppelt, weil sie nicht einsilbig ist.

Von dieser generellen Regel gibt es vergleichsweise wenige Ausnahmen. Die meisten dieser Ausnahmen enthalten finales ‹l› und existieren in beiden Varianten – z. B. ‹controlling›, ‹controling›, vgl. Bauer/Lieber/Plag (2013: 49 ff.). Die Varianten sind diatopisch verteilt (BrE vs. AmE).

Die Beschreibung der ‹e›-Tilgung ist einfacher; sie kann ohne Bezug auf die morphologische Struktur formuliert werden. Einfaches finales ‹e›, das auf einen Konsonanten folgt, wird in flektierten Formen getilgt, wenn ein Suffix mit initialem Vokal auftritt (z. B. ‹like› – ‹liked›). Dass hier die graphematische Form der Basis entscheidend ist, sieht man deutlich an Schreibungen wie ‹coming›. Die Basis ‹come› genügt den Bedingungen für die ‹e›-Tilgung – also wird das ‹e› getilgt. Phonographisch ist die Korrespondenz zwischen ‹coming› und /kʌmɪŋ/ schwer zu explizieren, wenn man die zahlreichen Wörter vom Typ ‹humming› – /hʌmɪŋ/ miteinbezieht, die mit Doppelkonsonant geschrieben werden.

Es scheint nur sehr wenige Ausnahmen von der ‹e›-Tilgung zu geben, und sie alle sind Stämme mit finalem ‹ge›. Ryan (2016: 91 f.) nennt in diesem Zusammenhang *ageing*, *singeing* und *bingeing*; sie verfügen über eine intakte Stammform, während bei den analog strukturierten Wortformen *lunging* und *cringing* das ‹e›

des Stamms (wie zu erwarten) meist getilgt wird. Die geringe Zahl der Ausnahmen zeigt die Stärke der Generalisierung über die ⟨e⟩-Tilgung an.⁸¹

Werfen wir abschließend auch im Englischen einen Blick auf die Phonologie, um die Möglichkeiten und Grenzen der Schreibungen besser bewerten zu können. Die regelmäßigen Stammalternationen (69c–e) sind graphemisch. Phonologisch sind die entsprechenden Flexionsformen konkatenativ aufgebaut, und der Stamm ist jeweils stabil:

- (71a) lady /leɪdi/ + s /z/ = ladies /leɪdiz/
 (71b) rip /rɪp/ + ed /t/ = ripped /rɪpt/
 (71c) dine /daɪn/ + ing /ɪŋ/ = dining /daɪnɪŋ/

Umgekehrt gibt es vereinzelte Fälle, in denen die Schrift phonologische Alternationen überbrückt. So variieren neben der /f/-/v/-Alternation auch andere stammfinale Frikative hinsichtlich ihrer Stimmhaftigkeit, und zwar /θ/-/ð/ und /s/-/z/. Bloomfield (1933: 213 f.) gibt die folgende exhaustive Liste:

- (72a) house – houses
 (72b) bath – baths, path – paths, cloth – cloths, mouth – mouths⁸²

In diesen Paaren gibt es jeweils phonologische, nicht aber graphematische Variation. Etwas Ähnliches gilt für einige wenige unregelmäßige Verben wie *hear*, deren Präsens und Präteritums-/Partizipialform graphematisch einheitlicher sind als die entsprechenden phonologischen Formen (⟨hear⟩ /hɪə/ – ⟨heard⟩ /hɜ:d/).

Es gibt also, so lässt sich zusammenfassen, im Englischen anders als im Deutschen kaum Fälle expliziter Einheitlichkeit von Stämmen – dafür gibt es eine Reihe von rein graphematischen Alternationen, die keine phonologische Entsprechung haben. Die einheitliche Schreibung von Stämmen in Flexionsparadigmen ist im Deutschen ein wichtigeres Strukturprinzip als im Englischen.

⁸¹ Der Vollständigkeit halber müssen zwei weitere Ausnahmen erwähnt werden, die mit der ⟨e⟩-Tilgung, so wie sie oben formuliert ist, nicht erfasst werden können. In *suing* und *intriguing* fällt das ⟨e⟩ des Stamms (*sue*, *intrigue*) ebenfalls weg (Ryan 2016: 97). Da das finale ⟨e⟩ nicht nach einem Vokal auftritt (so ist die Regularität oben formuliert), können die Fälle nicht erfasst werden. Sie sind aber absolut isoliert; in analogen Fällen bleibt der Stamm intakt, vgl. *rueing*, *toeing*, *shoeing*.

⁸² Daneben variiert der stammfinale Frikativ „for some speakers“ auch bei *lath* – *laths*, *oath* – *oaths*, *thruth* – *truths* und *youth* – *youths*, Bloomfield (1933: 213).

4.1.2 Wortbildung

In der Wortbildung gibt es in beiden Sprachen eine größere Bandbreite an Variationen der Stammform als bei der Flexion. Wir betrachten hier im Folgenden zwei der drei hauptsächlichen morphologischen Wortbildungsarten (i. S. v. Fleischer/Barz 2012: 83 f.), Derivation und Komposition. Die randständigeren Wortbildungsarten wie Rückbildung und Kurzwortbildung werden hier nicht behandelt. Konversion wird häufig als Zero-Ableitung begriffen (z. B. Marchand 1969: 359 ff.). In diesem Fall ist jede Variation schon definitiv ausgeschlossen. Im Deutschen werden allerdings einige Wortpaare mit Stammalternation als Instanzen von Konversion klassifiziert (auch wenn wir die implizite Derivation vom Typ *kurz/kürz[en]* ausschließen, siehe oben). Es geht um die folgenden Daten:

(73a) sing[en] – Singen

(73b) Rahmen – rahm[en]

(73c) trocken – trockn[en]

(73d) Jet – jett[en]

Der Typ in (73a) ist eine Instanz „syntaktischer Konversion“ Erben (2006: 31): Das Substantiv besteht aus dem Verbstamm und dem Infinitivsuffix. Wir können die Stammidentität ‚retten‘, wenn wir <en> beim Substantiv als Pseudosuffix klassifizieren. Unabhängig davon verhalten sich Konversionen wie (*das*) *Singen* nicht wie prototypische Substantive (vgl. z. B. Fuhrhop 2015: 43) – sie bilden bspw. keinen Plural (**die Singen*).

In (73b) haben wir dieselbe strukturelle Konstellation, aber eine andere Ableitungsrichtung: Hier wird das substantivische Pseudosuffix getilgt. Der neue Verbstamm kann wiederum per syntaktischer Konversion zum Substantiv werden – so entstehen regelmäßig homonyme (bzw. polyseme) Substantive wie <der Rahmen> – <das Rahmen>.⁸³ Das zeigt, dass es sich beim <en> in (73a) und (73b) um unterschiedliche Formen und bei den beiden Prozessen nicht einfach um unterschiedliche Konversionsrichtungen handelt. Getilgt wird neben <en> auch finales <e> (<Schule> – <schul[en]>, <Träne> – <trä[n]en[en]>), und zwar regelmäßig – das ist ein Muster, das auch bei der Derivation und der Komposition auftaucht. Stammfinales <el> und <er> bleibt hingegen erhalten, vgl. z. B. <Angel> – <angeln>, <Trauer> – <trauern> (zur Stammvariation bei Flexionsformen wie <ich angel>/<angle>/<angele> siehe 4.1.1).

⁸³ Für diesen Hinweis danke ich Nanna Fuhrhop.

In einigen wenigen Fällen wird nicht das gesamte Pseudosuffix bei der Konversion zum Verb getilgt, sondern nur das ⟨e⟩. Neben dem Beispiel in (73c) sind noch ⟨Regen⟩ – ⟨regn[en]⟩ und ⟨Atem⟩ – ⟨atm[en]⟩ einschlägig.

Typ (73d) kommt nur unter Anglizismen vor (vgl. z. B. Eisenberg 2011: 341). Hier wird der Doppelkonsonant des (desubstantivischen) Verbstamms (der sich aus der Silbengelenkschreibung von Formen wie ⟨jettet⟩ ergibt) zwar innerhalb des Flexionsparadigmas ‚übertragen‘ – nicht aber zurück auf den Substantivstamm. Der Fall ist z. T. vergleichbar mit der Alternation von einfachen und verdoppelten Konsonanten in Adjektivformen (vgl. oben 4.1.1) – nur dass es dort Variation innerhalb einzelner Lexeme gibt.

Die letzteren beiden Fälle (73c, d) sind Ausnahmen im System; die Typen in (73a, b) sind die Regel.

Bei der Derivation gibt es in beiden Sprachen substantielle Variation der Stammformen. Sie lässt sich im Deutschen in die folgenden Typen gliedern:

- (74a) Wolke – wolkig, *wolkeig
- (74b) Schatten – schattig, *schattenig
- (74c) Fieber – fiebrig, ?fieberig
- (74d) Nebel – neblig, nebelig
- (74e) trag[en] – Träger, *Trager
- (74f) Haar – Härchen, *Häärchen
- (74g) explodier[en] – Explosion, *Explodion, *explosier[en]
- (74h) kommunizier[en] – Kommunikation, *Kommunization *kommunikier[en]
- (74i) klassifizier[en] – Klassifikation, *Klassifization, *klassifikier[en]
- (74j) Muskel – muskulär, *muskelär, *Muskul
- (74k) Fiskus – fiskalisch, *fiskusalisch, *fiskusisch

Bevor wir ins Detail gehen, sind zwei Beobachtungen festzuhalten. Zum einen werden bis auf (74h) in allen Beispielen Stämme suffigiert, nicht präfigiert. Präfigierung führt nicht im selben Maß zu Variation in der Stammform wie Suffigierung. Zum anderen spielt sich die Variation in der Stammform vor allem am Stammende ab (74a, b, c, g, h, i), seltener am Stammvokal (74d, e, f, j) und nie am Stammanfang.

In den Belegen in (74a–d) wird das Stammende entweder komplett (⟨e⟩, ⟨en⟩) oder teilweise getilgt (⟨er⟩, ⟨el⟩). In allen Fällen handelt es sich um frequente Wortenden mit ⟨e⟩, und mit Ausnahme von ⟨el⟩ sind sie flexionsmorphologisch stark belastet.

Ob das Wortende getilgt wird, hängt vom Wortende und vom Derivationsuffix ab (es gibt teilweise auch Variation innerhalb von Lexemen, z. B. ⟨ekelig⟩/⟨eklig⟩; dazu gleich mehr). Zum Wortende: Hier verhält sich finales ⟨e⟩

anders als die anderen drei Elemente. Es wird so gut wie immer entweder getilgt (z. B. *freudlos*, **freudelos*) oder mit ⟨n⟩ kombiniert (z. B. *Heidentum*, **Heidetum*). Die Ausnahmen sind an zwei Händen abzählbar (*ruhelos*, *mühe-los*, *ehelich*, *ehelos*, *interesselos*, *reklamehaft*).⁸⁴ Die Frage, ob das ⟨e⟩ getilgt oder ‚gedeckt‘ wird, wird vom Suffix bestimmt. Die meisten Suffixe führen zu Stämmen mit getilgtem ⟨e⟩, z. T. mit Umlaut des Stammvokals (z. B. *Täub·chen*, *Näs·lein*), z. T. ohne (z. B. *kass·ier[en]*, *sonn·ig*, *schul·isch*, *sach·lich*, *Pat·in*). Die Suffixe *-haft*, *-weise* und *-tum* sind die Ausnahmen: Sie kommen fast ausschließlich mit ⟨en⟩ vor (z. B. *masse·n·haft*, *stufe·n·weise*, *Heide·n·tum*). Überwiegend mit ⟨en⟩ (aber nicht ausschließlich!) kommen außerdem *-los* und *-schaft* vor (z. B. *schranke·n·los* vs. *freud·los*; *Pate·n·schaft* vs. *Sipp·schaft*). Diese Suffixgruppe (*-haft*, *-weise*, *-tum*, *-los* und *-schaft*) wird uns noch mehrfach begegnen.⁸⁵

Stammfinales ⟨en⟩ wird nicht immer getilgt, und auch hier hängt die Verteilung vom Suffix ab. Stabil ist ⟨en⟩ in Ableitungen mit *-heit*, *-haft*, *-weise*, *-los* und *-schaft* (z. B. *Selten·heit*, *märchen·haft*, *haufen·weise*, *boden·los*, *Eigen·schaft*). Mit Ausnahme von *-heit* ist diese Gruppe in der Gruppe der Suffixe oben enthalten, die ⟨e⟩ mit ⟨n⟩ decken. An der Oberfläche führt das zu parallelen Formen (*masse·n·haft* vs. *märchen·haft*, *stufe·n·weise* vs. *haufen·weise* etc.). Diese formale Analogie geht noch weiter: Die Suffixe *-heit*, *-haft*, *-weise*, *-tum*, *-los* und *-schaft* (die Vereinigungsmenge der beiden oben genannten Gruppen) verbinden sich z. T. mit dem Fugenelement ⟨en⟩ (z. B. *komet·en·haft*, *Christ·en·heit*, *schicht·en·weise*, *Narr·en·tum*, *staat·en·los*, *Mach·en·schaft*). Die Suffixe verhalten sich damit ähnlich wie Zweitglieder in Komposita; hier tritt ebenfalls häufig ⟨n⟩ als Fugenelement auf, wenn der Stamm auf ⟨e⟩ endet (z. B. *Tomatensalat*, *Affenkäfig*).

Getilgt wird finales ⟨en⟩ meist ohne Umlaut in Ableitungen mit *-ig* und *-isch* (z. B. *schatt·ig*, *trop·isch*), meist mit Umlaut in Ableitungen mit *-chen*, *-el* und *-lein* (*Gärt·chen*, *fäd·el[n]*, *Tröpf·lein*). Diese Gruppe verhält sich gerade nicht wie Zweitglieder in Komposita – sondern wie man es von Derivationsuffixen erwarten würde.

Die letzten beiden Wortenden, ⟨er⟩ und ⟨el⟩, werden nicht vollständig, sondern nur teilweise getilgt (*fiebrig*, **fiebig*). Das Suffix muss mit einem Vokal

⁸⁴ Außer *reklamehaft* enthalten die Belege entweder stammfinales ⟨he⟩ oder das Suffix *-los*. Beides sind jedoch keine notwendigen Bedingungen für ⟨e⟩-Bewahrung, vgl. *mühsam*, *Mühsal* und *ehrlos*, *endlos*, *farblos* etc.

⁸⁵ Vgl. auch Fuhrhop (2000), die beobachtet, dass ⟨e⟩ an Kompositumsgrenzen möglich und stabil ist, solange es morphologisch interpretiert werden kann (z. B. in *Frischebehälter*).

beginnen; für ⟨er⟩ sind *-ig*, *-ier[en]* und *-isch* einschlägig, für ⟨el⟩ sind es *-ig*, *-isch*, *-ung* und *-er*. Bei den übrigen Suffixen wird ⟨e⟩ praktisch nicht getilgt. Bei Ableitungen auf *-ig* gibt es graphematische Variation innerhalb der Lexeme (⟨pulverig⟩ vs. ⟨pulvrig⟩). Schauen wir uns beispielhaft eine Reihe von Varianten mit ihren Belegzahlen im Cosmas-Korpus W-Öffentlich an:

Tab. 104: Verteilung der Tilgung bzw. Nicht-Tilgung von ⟨e⟩ im Pseudosuffix ⟨er⟩ bzw. ⟨el⟩ bei Ableitungen mit *-ig*. Datengrundlage: Cosmas-Korpus W-Öffentlich.

	Wort	+⟨e⟩	-⟨e⟩
a	trau(e)rig	1	98.434
b	schau(e)rig	5	3.671
c	feu(e)rig	5	3.310
d	silb(e)rig	6	2.845
e	eit(e)rig	1	141
f	wäss(e)rig	163	1.397
g	pud(e)rig	12	92
h	pulv(e)rig	40	122
i	knaus(e)rig	1.086	1.186
j	fas(e)rig	479	62
k	neb(e)lig	471	2.803
l	ad(e)lig	787	609
m	wack(e)lig	2.846	1.794
n	grus(e)lig	6.280	207
o	pumm(e)lig	637	7
p	hüg(e)lig	1.461	18
q	kug(e)lig	1.228	4
r	schläf(e)rig	3	2.407
s	kleb(e)rig	0	2.161
t	löch(e)rig	209	3.321

Die eindeutigste Verteilung gibt es bei Stämmen mit Schreibdiphthong vor ⟨er⟩ (Fälle a–c): Hier ist die Variante ohne ⟨e⟩ dominant (Typ a, ⟨traurig⟩). Es gibt keine analogen Stämme mit finalem ⟨el⟩, daher gibt es auch keine entsprechenden Ableitungen auf *-ig*. Bei den übrigen Ableitungen auf ⟨er⟩-Stämmen gibt es Variation (Fälle d–j); es finden sich vereinzelt gleichmäßig verteilte Varianten (Typ i, ⟨knauserig⟩/⟨knausrig⟩), und teilweise überwiegt auch die ⟨er⟩-Schreibung (Typ j, ⟨faserig⟩) – in den meisten Fällen wird das ⟨e⟩ aber wohl getilgt. Umgekehrtes gilt wohl für die Ableitungen von ⟨el⟩-Stämmen (Fälle k–q). Hier gibt es zwar auch Varianten, die mit getilgtem ⟨e⟩ häufiger sind (Typ k, ⟨neblig⟩), und solche, die

annähernd gleich verteilt sind (Typ 1, <adlig>/<adelig>). Bei den meisten Ableitungen überwiegt aber die <el>-Schreibung recht deutlich.

Eine interessante Gruppe bilden die Fälle r-t. Hier wird das <e> fast immer getilgt – es gibt aber keine plausible Basis auf <er>. Fleischer/Barz (2012: 338 f.) vermuten eine „alternative Derivationsstammform von Verben ohne -er im Infinitivstamm“; sie segmentieren entsprechend *klebr·ig*. Es macht aber an dieser Stelle mehr Sinn, von einem allomorphen Suffix -*rig* auszugehen – die Segmentierung wäre dann *kleb·rig*. Dafür spricht, dass die meisten Stämme mit finalelem <er> das <e> in -*ig*-Ableitungen verlieren (*holper[n]* – *holprig*). Damit ist die formale Voraussetzung für eine Reanalyse *holp·rig* gegeben – der Fall ist in dieser Hinsicht parallel zu -*er/-ler/-ner* (vgl. Fuhrhop 1998: 56 ff.). Dazu passt auch, dass es keine entsprechenden Fälle auf -*lig* gibt: Die meisten Stämme auf <el> scheinen das <e> in Ableitungen mit -*ig* zu bewahren.⁸⁶

Neben (und teilweise zusätzlich zu) diesen Tilgungen tritt der Umlaut des Stammvokals als nicht-segmentale Alternation auf (Typ 74e, *trag[en]* – *Träger*). Der Umlaut tritt regelmäßig bei den Suffixen -*chen*, -*lein*, -*ig*, -*lich*, -*er*, -*e*, -*el*, -*isch*, -*in* u. a. m. auf. Er tritt nicht auf bei (unter anderem) der Gruppe -*heit*, -*haft*, -*weise*, -*tum*, -*los* und -*schaft*, die sich oben dadurch auszeichnete, dass in Ableitungen mit diesen Suffixen keine Pseudosuffixe des Stamms getilgt werden. Auch hinsichtlich des Umlauts verhält sich die Gruppe also konsistent, indem Ableitungen rein konkatenativ verknüpft werden. In diese Gruppe gehört auch die sog. implizite Derivation vom Typ *kurz* – *kürz[en]* (vgl. Eisenberg 2013a: 280), sofern sie mit dem Umlaut und nicht mit dem Ablaut operiert (Typ *werf[en]* – *Wurf*).

Ein Sonderfall des Umlauts ist der Umlaut von Doppelvokalen (Typ 74f, *Haar* – *Härchen*). Hier erscheinen statt zwei umgelauteten Vokalgraphemen nur eines (*<Häärchen>). Auch unter den einfachen Stämmen gibt es keine Verdoppelungen der Umlautgrapheme; nur die einfachen Vokalgrapheme <a>, <e> und <o> können verdoppelt werden (siehe oben 3.2.1.3); die Verdoppelung der übrigen Vokalgrapheme (einschließlich der umgelauteten) ist graphotaktisch sanktioniert (vgl. auch Primus 2000: 61 f.).

Der Beleg (74g) steht stellvertretend für weitere Typen, in denen der stammfinale Konsonant alterniert, und zwar in gebundenen lateinischen Stämmen. Am häufigsten alternieren wohl <z> und <d> in Verbstämmen vor -*ier[en]* mit <ʈ> bzw. <s> in Ableitungen mit -(*at*)*ion* (*kommunizier[en]* – *Kommunikation*, *explor-*

⁸⁶ Analog kann man auch -*erisch* als Allomorph von -*isch* und -*erlich* als Allomorph von -*lich* ansetzen. Bildungen wie *rechmerisch* oder *lächerlich* sind direkt auf den Verbstamm *rechn[en]* und *lach[en]* beziehbar (vgl. Fleischer/Barz 2012: 344).

dier[en] – Explosion). Seltener ist die analoge Alternation ⟨t⟩/⟨s⟩ (*konvertier[en] – Konversion*).⁸⁷

Einen Sonderfall stellt die Alternation in (74i) dar. Hier steht das Verbalsuffix *-ifizier[en]* der Endung *-ifikation* gegenüber. Da es sich bei der *-ation*-Form um Ableitungen aus *-ifizier[en]*-Verben handelt, wird ⟨ifik⟩ als Teil des Verbalsuffixes angesehen. Wird ein Verb wie *identifizier[en]* mit *-ion* abgeleitet, dann alternieren also a) ein Teil des Verbalsuffixes, ⟨ier⟩, mit einem Allomorph von *-ion*, ⟨ation⟩, und b) ⟨z⟩/⟨k⟩ im Verbalsuffix. Diese Alternation betrifft also nicht die Einheitlichkeit von Stämmen, sondern von Affixen (siehe unten 4.3.2).

In lateinischen Stämmen gibt es auch Alternationen der Vokalgrapheme (74j), vor allem in adjektivischen Ableitungen mit *-är*, *-ös* und *-ell* (*Zirkel – zirkulär*, *Muskel – muskulös*, *Nomen – nominell*). Die Basis erscheint hier jeweils mit ⟨e⟩-haltigem Wortende, also mit einem Pseudosuffix in der hier verwendeten Terminologie, und in der Ableitung alterniert ⟨e⟩ mit ⟨u⟩ oder ⟨i⟩. Auch hier ist – wie oben bei den Tilgungen von ⟨e⟩, ⟨er⟩, ⟨en⟩ und ⟨el⟩ – jeweils nur das Pseudosuffix von der Alternation betroffen. Es kollidiert die nativisierte Schreibung des freien Stamms (*circulus* > ⟨Zirkel⟩, *mūsculus* > ⟨Muskel⟩) mit der nicht (vollständig) nativen Schreibung des Stamms in Ableitungen (vgl. auch Eisenberg 2011: 270 ff.).

Und schließlich gibt es besonders unter den Latinismen den Fall, dass Wörter mit Stammflexion diese Stammform auch für die Derivation ‚nutzen‘: Der Plural von *Zyklus* lautet (in den meisten Fällen) *Zyklen* (Eisenberg 2011: 224 f.), und der gebundene Stamm *Zykl-* ist gleichzeitig die Derivationsstammform (*zyklisch*).

Neben diesen mehr oder weniger regelmäßigen Alternationen gibt es eine Reihe idiosynkratischer Alternationen, die nicht mehr vollständig transparent sind (z. B. *schneid[en] – Schnitter*; *treff[en] – triftig*; *Erde – irdisch*, *korrigier[en] – Korrektur* etc.; vgl. auch Nerius (Hg.) 2007: 103 ff.). Sieht man von diesen Ausnahmen ab, dann findet graphematische Variation in deutschen Stämmen vor allem in folgenden Bereichen statt: a) beim Umlaut, b) bei Pseudosuffixen und c) (sehr viel seltener) im finalen Element gebundener lateinischer Stämme. Die Markierung mit Umlaut stellt, wie oben gezeigt wurde, einen Mittelweg zwischen einheitlicher Schreibung des Stamms und Markierung der Ableitung dar – der Stamm ist in gewissem Sinne intakt, das Trema zeigt an, dass hier morphologisch etwas passiert. Es handelt sich um einen minimalinvasiven Eingriff in die Integrität des Stamms. Die Tatsache, dass nur die Pseudosuffixe *-e* und *-en* (sowie Teile

⁸⁷ Die Alternation ⟨t⟩/⟨s⟩ kommt in Stämmen mit der lateinischen Wurzel *-mitt-* als ⟨tt⟩/⟨ss⟩ vor (*emittier[en] – Emission*). Ebenfalls an eine bestimmte Wurzel gebunden ist die Alternation ⟨n⟩/⟨s⟩ (*komponier[en] – Komposition*) – sie kommt nur in Stämmen mit der lateinischen Wurzel *-pon-* vor (*disponier[en] – Disposition*, *supponier[en] – Supposition*).

der Pseudosuffixe *-er* und *-el*) getilgt werden, nicht aber beliebige Stammenden, deutet darauf hin, dass diese Pseudosuffixe für die Integrität des Stamms weniger wichtig sind. Die einzigen ‚echten‘ Ausnahmen von der Einheitlichkeit deutscher Stämme sind damit die Alternationen in Stämmen mit gebundener lateinischer Wurzel. Einheitlichkeit von Stämmen ist also auch in der Derivation ein wichtiges Strukturprinzip.

Das zeigt auch ein Blick auf phonographisch denkbare Alternationen. Hier sind zwei der drei oben bei der Flexion diskutierten Fälle einschlägig: Die Auslautverhärtung wird nicht verschriftet (⟨Kind⟩ – ⟨kindlich⟩, *⟨kintlich⟩) und Umlaute werden mit Trema gekennzeichnet (⟨Angst⟩ – ⟨ängstlich⟩, *⟨engstlich⟩). Beides müsste nicht so sein: *⟨kintlich⟩ und *⟨engstlich⟩ sind phonographisch plausibel. Dass die Stämme auch hier stabil sind, zeigt noch einmal den hohen Stellenwert dieses Prinzips.

Im Englischen lassen sich die uneinheitlich verschrifteten Stämme wie folgt klassifizieren:

- (75a) happy – happiness, *happyness
- (75b) swim – swimmer, *swimer
- (75c) divine – divinity, *divineity
- (75d) elf – elvish, *elfish, *elve
- (75e) register – registration, *registration
- (75f) feeble – feebly, *feeblely
- (75g) explode – explosion, *explodion, *explose
- (75h) substance – substantial, *substancial, *substant
- (75i) miracle – miraculous, *miraclous
- (75j) focus – focal, *focusal

Für die Typen (75a–c) gilt zunächst, was oben im Rahmen der Flexionsformen geschrieben wurde. Es handelt sich im Kern um dieselben regelmäßigen Stammalternationen, die ⟨i⟩/⟨y⟩-Alternation (75a), die Konsonantenverdoppelung (75b) und die ⟨e⟩-Tilgung (75c). Der Unterschied zur Flexion ist, dass es in Ableitungen etwas mehr Ausnahmen von den oben formulierten Regularitäten gibt, und zwar abhängig vom jeweiligen Derivationsuffix.

Für die ⟨i⟩/⟨y⟩-Alternation wurde diese Variation in Berg (2013) genutzt, um Aussagen über die Stärke der Verbindung zu treffen. Die Suffixe in (76a) kommen ausschließlich oder fast ausschließlich mit der Alternation vor, diejenigen in (76b) ohne – diese Ableitungen sind daher mit Komposita vergleichbar, bei denen ⟨y⟩ auch bewahrt wird (z. B. ⟨ladybug⟩). Bei den Suffixen in (76c) gibt es Variation:

- (76a) -age, -al, -an, -ance, -ant, -ary, -ate, -ation, -ed, -ly, -ment, -ous, -ful(A),
-ness
(76b) -dom, -ful(N), -like, -ship, -type
(76c) -er, -able, -some, -less, -hood, -fold, -wise

In der letzten Gruppe befinden sich Suffixe, die eher alternieren (-*er* und -*able*), und solche, die das finale ⟨y⟩ eher bewahren (-*less*, -*hood*, -*fold*, -*wise*).⁸⁸ Bei den meisten Suffixen mit Variation gibt es strukturelle Faktoren, die diese Variation bestimmen. -*ness* z. B. tritt eindeutig mit Alternation auf, wenn die Basis adjektivisch ist (Typ *happiness*). Wenn die Basis hingegen substantivisch ist (der seltenere Typ *babyness*), bleibt das stammfinale ⟨y⟩ in den meisten Fällen erhalten. Formal unterscheiden sich die Gruppen deutlich: Mit Ausnahme von -*ly*, -*ment*, -*ful(A)* und -*ness* sind es vor allem vokalisches beginnende Suffixe, die mit ⟨i⟩/⟨y⟩-Alternation auftreten; umgekehrt haben alle Suffixe, die ohne diese Alternation vorkommen, einen initialen Konsonanten; etymologisch sind sie nativ. Sie verhalten sich also auch graphematisch eher wie Stämme denn wie Suffixe (vgl. oben 3.4).

Die Konsonantenverdoppelung (Typ 75b) folgt bei Derivationsuffixen grundsätzlich denselben Regularitäten wie in der Flexion (Berg 2016: 461 f.). Die Affixe verhalten sich hier einheitlich, es gibt nur wenige Ausnahmen wie die folgenden:

- (77a) *crystallize/crystalize; bimetallism/bimetalism; panelist/panelist*
(77b) *parity, gasify, scarify*
(77c) *inference, reference, preferable*

Die Belege in (77a) gehören zu einer größeren Gruppe von Wörtern, die diatopisch variieren: Im amerikanischen Englisch wird die Form mit einfachem ⟨l⟩ bevorzugt, im britischen Englisch die mit ⟨ll⟩ (vgl. auch Bauer/Lieber/Plag 2013: 50 f.). Bei den Wörtern in (77b) sollte nach der oben formulierten Regularität Konsonantenverdoppelung erwartet werden; stattdessen werden die Stämme einheitlich verschriftet. In CELEX sind dies die einzigen Fälle, in denen eine erwartbare Konsonantenverdoppelung unterbleibt. Für die Belege in (77c) schließlich wird umgekehrt mit der Verdoppelung gerechnet; diese Formen treten aber nicht auf (*⟨inference⟩, *⟨reference⟩, *⟨preferable⟩). Man könnte hier vermuten, dass die prosodische Struktur der phonologischen Form die Schreibungen mit Doppelkonsonant ‚verhindert‘: Immerhin tragen die Wörter den Hauptakzent auf der ers-

⁸⁸ Das letzte Suffix, -*some*, kommt ebenso häufig mit wie ohne Alternation vor.

ten Silbe (*'inference*), nicht – wie das die Doppelkonsonanten nahelegen – auf der zweiten (**in'ference*). Folgt man aber dieser Logik, so sollten ultimabetonte Formen wie *submittee* ohne Konsonantenverdoppelung auskommen, und in Ableitungen wie *atomic* wären Doppelkonsonanten erwartbar, um den Hauptakzent zu kennzeichnen. Das alles geschieht nicht – entscheidend ist vielmehr die graphematische und morphologische Form der Basis, und die Formen in (77c) sind Ausnahmen.

Bei der ⟨e⟩-Tilgung (75c) gibt es in zweierlei Hinsicht Variation, die von den oben beschriebenen Regularitäten abweicht. Das betrifft zum einen finales ⟨e⟩ nach ⟨c⟩ oder ⟨g⟩. Hier hat ⟨e⟩ die Funktion, die Korrespondenz zu /s/ bzw. /dʒ/ zu sichern (z. B. *sing* vs. *singe*, *music* vs. *ice*). Diese Funktion ist unabhängig von einer eventuell zusätzlichen Funktion, die Korrespondenz des Vokals in der Silbe davor als gespannt zu markieren (vgl. *ice*, wo das finale ⟨e⟩ beide Funktionen erfüllt). ⟨c⟩ und ⟨g⟩ korrespondieren regelmäßig vor ⟨e⟩ und ⟨i⟩ mit /s/ bzw. /dʒ/; in allen anderen Kontexten korrespondieren sie mit /k/ bzw. /g/. Daher bleibt das finale ⟨e⟩ in diesen Fällen erhalten, und der Stamm variiert nicht (*manage* – *manageable*, **managable*). Zum Teil gibt es vor Suffixen mit initialem Konsonanten Variation: ⟨lodgement⟩ steht neben ⟨lodgment⟩ (vgl. Ryan 2016: 65).

Gewissermaßen quer zu dieser phonographisch motivierten (Sub-)Regularität gibt es morphologisch motivierte Variation. Wie in Berg (2013) gezeigt wurde, wird finales ⟨e⟩ vor *-able* nicht konsistent getilgt; Formen wie ⟨microwaveable⟩ mit ⟨e⟩ und ⟨microwavable⟩ ohne ⟨e⟩ stehen nebeneinander.

Auch Typ (75d) ist schon bei der Flexion aufgetreten – es geht hier um die Alternation ⟨f⟩/⟨v⟩ (z. B. ⟨elf⟩ – ⟨elvish⟩). Die Derivationsstammform stimmt in dieser Hinsicht mit der Pluralstammform überein (⟨elves⟩, ⟨elvish⟩). Auch in der Derivation wird also die phonologische Alternation nicht überbrückt.

Die nächsten beiden Typen, (75e, f), haben Entsprechungen im Deutschen. Es handelt sich um die ⟨e⟩-Tilgung vor ⟨r⟩ (Typ ⟨winter⟩ – ⟨wintry⟩, **⟨wintery⟩*)⁸⁹ und um die Tilgung von ⟨le⟩ vor dem Adverbsuffix *-ly* (Typ ⟨feeble⟩ – ⟨feebly⟩, **⟨feebly⟩*). Beide erinnern an die vollständige oder teilweise Tilgung von Pseudosuffixen im Deutschen (z. B. ⟨Fieber⟩ – ⟨fiebrig⟩). Die Alternationen haben im Englischen allerdings noch größeren Ausnahmeharakter: Der Ausfall von ⟨e⟩ vor ⟨r⟩ ist auf eine kleine Gruppe von Suffixen beschränkt: *-ance* (*remembrance*), *-ant* (*ministrant*), *-ar* (*registrar*), *-ate* (*filtrate*), *-ess* (*tigress*), *-ic* (*metric*), *-ous* (*monstrous*) und *-y* (*wintry*). Und Wörter mit diesen Suffixen verhalten sich nicht immer uniform, vgl. z. B. *remembrance* vs. *ponderance*. Der Ausfall von ⟨le⟩ ist auf Ablei-

⁸⁹ Sehr viel seltener wird auch ⟨o⟩ vor ⟨r⟩ getilgt, vgl. ⟨ancestor⟩ – ⟨ancestral⟩, **⟨ancestor⟩*.

tungen mit *-ly* beschränkt; die Reichweite beider Regularitäten ist geringer als im Deutschen.

Die letzten vier Typen (75g–j) stammen alle aus dem lateinischen/französischen Teil des englischen Wortschatzes. Sie sind in ihrer Verbreitung stark begrenzt. Betrachten wir zunächst die Alternationen (75f, g). Der häufigste Fall ist wohl (75f), in dem ⟨d⟩ mit ⟨s⟩ alterniert; zum Teil mit ⟨e⟩-Tilgung wie in ⟨explode⟩ – ⟨explosion⟩, zum Teil ohne wie in ⟨defend⟩ – ⟨defensible⟩. Hier gibt es etwa drei Dutzend einschlägige Fälle, wie ein kursorischer Blick in CELEX zeigt. Wie oben erwähnt, erklärt Venezky (2004) diese Stammvariation phonographisch: Stammfinal variieren /d/ und /ʒ/, und es gibt kein Graphem, das mit beiden korrespondieren kann (dass es ein solches Graphem gäbe, wenn die Stämme einheitlich verschriftet würden, wie oben argumentiert wurde, ist im Moment nicht relevant). Die Alternation (75h) (⟨c⟩/⟨t⟩ wie in ⟨substance⟩ – ⟨substantial⟩) ist bereits wesentlich seltener. Sie tritt noch gut ein Dutzend Mal auf. Hier greift Venezkys Argument nun gerade nicht: Die Korrespondenz von stammfinalen ⟨c⟩ – /ʃ/ ist vor ⟨i⟩-initialen Suffixen im Englischen gut belegt (z. B. *suspicion*, *commercial*, *delicious*). Einer Schreibung wie *⟨substancial⟩ steht phonographisch nichts im Weg, und dennoch verzichtet das Englische hier auf die Einheitlichkeit der entsprechenden Stämme.⁹⁰

In einer Handvoll Fälle tritt in der Derivationsstammform ⟨u⟩ auf, wo in der Flexionsstammform ⟨e⟩ steht, z. B. in ⟨miracle⟩ – ⟨miraculous⟩ oder ⟨muscle⟩ – ⟨muscular⟩ (75i). Diese Paare haben direkte Entsprechungen im Deutschen – nicht nur, was das Muster angeht, sondern auch, was viele der konkreten Lexeme angeht (vgl. z. B. ⟨Mirakel⟩ – ⟨mirakulös⟩, ⟨Muskel⟩ – ⟨muskulös⟩).

Und schließlich gibt es – ebenfalls wie im Deutschen – eine Reihe von Ableitungen, in denen das Wortende des Stamms getilgt wird (Typ 75j), z. B. ⟨focus⟩ – ⟨focal⟩. Anders als im Deutschen gibt es aber oft keine Stammflexion (der Plural von *focus* ist laut OED *focuses*, die Form mit Stammflexion, *foci*, ist auf wissenschaftliche oder technische Kontexte beschränkt). Es handelt sich um Latinismen und Gräzismen, und die Überschneidung mit dem Deutschen dürfte recht hoch sein (z. B. ⟨radius⟩ – ⟨radial⟩, ⟨Radius⟩ – ⟨radial⟩; ⟨stimulus⟩ – ⟨stimulate⟩, ⟨Stimulus⟩ – ⟨stimulier[en]⟩, ⟨virus⟩ – ⟨virology⟩, ⟨Virus⟩ – ⟨Virologie⟩, ⟨velum⟩ – ⟨velar⟩, ⟨Velum⟩ – ⟨velar⟩ etc.).

⁹⁰ Neben diesen beiden Alternationen gibt es noch solche, die auf einzelne lateinische Wurzeln beschränkt sind, etwa ⟨t⟩/⟨s⟩ in Paaren wie *convert* – *conversion*, *divert* – *diversion*; ⟨t⟩/⟨ss⟩ wie in *admit* – *admission*, *emit* – *emission*; ⟨d⟩/⟨s⟩ wie in *concede* – *concession*, *recede* – *recession*; ⟨ive⟩/⟨pt⟩ wie in *receive* – *reception*, *conceive* – *conception*; ⟨c⟩/⟨k⟩ wie in *provoke* – *provocation*, *invoke* – *invocation*.

Insgesamt gibt es im lateinisch-griechischen Teil des Wortschatzes relativ große Übereinstimmungen zwischen dem Englischen und dem Deutschen. Auf den gesamten Wortschatz gesehen fallen die Ausnahmen von der Einheitlichkeit von Stämmen hier aber kaum ins Gewicht. Systematisch sind die drei Alternationen <e>-Tilgung, Konsonantenverdoppelung und <i>/<y>-Alternation. Sie führen regelmäßig zu jeweils zwei Stammformen.

Wie verhalten sich nun die korrespondierenden phonologischen Formen? In der einschlägigen Literatur wird häufig die These vertreten, dass englische Lexeme trotz phonologischer Variation relativ einheitlich verschriftet werden, dass die Schreibung des Englischen also phonologische Alternationen überbrückt (so z. B. Venezky 2004). Als Argument für diese These werden zum einen Schreibungen wie die in (78) angeführt:

- (78a) autumnal – autumn, damnation – damn, signal – sign
(z. B. Carney 1994: 245; Venezky 1999: 206)
- (78b) sane – sanity, divine – divinity, cone – conical
(z. B. Carney 1994: 14; Venezky 1999: 206 f.; Rollings 2004: 54 f.)
- (78c) electric – electricity – electrician
(z. B. Rollings 2004: 56)
- (78d) grade – gradual, revise – revision, critical – criticize
(Chomsky 1970: 292)

In (78a) korrespondiert jeweils ein Konsonantengraphem nur in bestimmten morphologischen Umgebungen – in der Grundform ist es ‚stumm‘ (oder „inert“, Carney 1994: 40). In (78b) haben Vokalgrapheme in verschiedenen Umgebungen verschiedene Korrespondenzen, etwa /eɪ/ in *sane* gegenüber /æ/ in *sanity*. In (78c) ist das finale <c> von *electric* in Ableitungen stabil, obwohl es mit drei Phonemen alterniert (/k/ vs. /s/ vs. /ʃ/). Ähnlich verhalten sich die Belege in (78d), bei denen jeweils das stamffinale Phonem alterniert.

Diese Aufstellung ist als Argument für morphologische Schreibungen im Englischen problematisch, und zwar in mehrerlei Hinsicht:

- Die Beispiele in (78a) sind idiosynkratisch; außer ihnen gibt es wahrscheinlich nur wenige andere strukturell gleiche. Für die übrigen Typen gilt das nicht, sie sind unterschiedlich regelmäßig. Am regelmäßigsten sind wohl (78b, c).
- Sampson (1985) merkt an, dass nicht alle möglichen phonologischen Alternationen graphematisch überbrückt werden. Mit anderen Worten: Es reicht nicht, die expliziten morphologischen Schreibungen wie (78) zu sammeln – wir müssen auch wissen, wie häufig genau nicht morphologisch konstant verschriftet wird.

- Es wird meist nicht unterschieden zwischen einer stabilen Schreibung von Stämmen und Affixen; unter Umständen gibt es hier aber Unterschiede, die so verwischt werden. In (78a, b) haben wir es mit phonologischer Variation im Stamm zu tun; in (78c) mit Variation in der Schreibung des Suffixes *-ic*; gerade dieser Fall ist höchst regelmäßig. Beides hängt unter Umständen zusammen, aber wenn die Kategorien analytisch nicht getrennt werden, haben wir keine Chance, das zu wissen. Problematisch sind in dieser Hinsicht die Fälle in (78d): Gehört die phonologische Variation hier zum Stamm oder zum Affix? Und ist *-ic* in *critic* ein Affix?

Es ist vor diesem Hintergrund wohl kein Zufall, dass die Frage nach dem Ausmaß und der Regelmäßigkeit der Stamm- und Affixkonstanz im Englischen bislang nicht empirisch bearbeitet worden ist.⁹¹ Dieses Desiderat kann auch im Rahmen dieser Arbeit nicht eingelöst werden; das Thema ist sicher eine eigene Arbeit wert. Statt einer systematischen Untersuchung sollen im Folgenden einige Aspekte dieses Komplexes unsystematisch kommentiert werden.

Im Kern geht es um die Frage, wo morpho-phonologische Alternationen im geschriebenen Englisch überbrückt werden und wo nicht. Schauen wir uns zwei dieser Alternationen etwas näher an, die Vokalstärkung und die trisyllabische Vokalschwächung (vgl. für das Folgende Bauer/Lieber/Plag 2013: 169 ff.); an ihnen wird noch einmal das Ausmaß der Probleme deutlich, die sich bei einer möglichen Untersuchung ergeben:

Bei der Vokalstärkung (‘vowel strengthening’) entspricht einem Reduktionsvokal in der Grundform ein Vollvokal in der abgeleiteten Form. Der reduzierte Vokal in der zweiten Silbe von *atom* oder *moral* wird in Ableitungen wie *atomic* oder *morality* zu /v/ in *atomic* bzw. /æ/ in *morality*. Diese Alternation geht normalerweise mit einer prosodischen Alternation einher: In *atom* ist der betreffende Vokal unbetont, in *atomic* trägt er den Hauptakzent. Für phonologische Beschreibungen ist das ein Dilemma: Ein und derselbe reduzierte Vokal des freien Stamms alterniert mit einer ganzen Reihe von Vollvokalen, und die konkrete Vokalqualität ist phonologisch nicht vorhersagbar. Giegerich (1999) plädiert daher dafür, das Phänomen als Leseaussprache zu verstehen: Weil wir wissen, dass *atom* mit ⟨o⟩ verschriftet wird, und weil dieses ⟨o⟩ auch in Ableitungen stabil bleibt, können wir den phonologischen Vokal über die regulären Phonem-Graphem-Korrespon-

⁹¹ Eine Ausnahme ist Sproat (2000) – ihm geht es aber um die sehr viel speziellere Frage, ob die graphematische Struktur einfacher aus einer oberflächennahen phonologischen Beschreibung oder aus einer sehr viel abstrakteren phonologischen Repräsentation i. S. v. Chomsky/Halle (1968) abgeleitet werden kann.

denzen erschließen. Die graphematische Form der Stämme ist hier also nicht nur konstanter als die phonologische – die konkrete phonologische Alternation hängt von dieser graphematischen Form entscheidend ab. Das muss bei einer Untersuchung mitbedacht werden.

Die trisyllabische Vokalschwächung (‘trisyllabic laxing’) geht auf eine Beobachtung von Chomsky/Halle (1968: 52) zurück: In gesprochenen englischen Wörtern ist die Antepenultima ungespannt. Die Generalisierung beschreibt Daten wie den Wechsel von *divine* (mit /ɹ/ in der Penultima) zu *divinity* (mit /ɪ/ in der Antepenultima). Graphematisch sind diese Fälle äußerst regelmäßig: Das finale ⟨e⟩ von ⟨divine⟩ wird getilgt, ansonsten wird konkateniert: ⟨divin⟩ + ⟨ity⟩.⁹² Außer der ⟨e⟩-Tilgung sind die Stämme einheitlich, und diese einheitliche Schreibung ist explizit, sie überbrückt phonologische Variation. Die Reichweite dieser Generalisierung ist allerdings umstritten (vgl. z. B. Bauer/Lieber/Plag 2013: 171). Das ist ein Problem für eine mögliche Untersuchung: Um zu bestimmen, ob phonologische Alternationen graphematisch überbrückt werden, ist eine Liste mit möglichst explizit formulierten Typen von phonologischen Alternationen hilfreich.⁹³

Bislang wurde lediglich die Einheitlichkeit von Stämmen in der Derivation betrachtet; es fehlt noch ein Blick auf die Komposition in beiden Sprachen. Im Englischen ist das relativ schnell erledigt: In Komposita bleiben die jeweiligen Stämme größtenteils intakt. Teilweise sind sie vom Spatium getrennt, teilweise vom Bindestrich, und teilweise sind sie verbunden (⟨apple tree⟩ vs. ⟨apple-tree⟩ vs. ⟨appletree⟩), aber die Frage der Verteilung der Varianten hat keine Auswirkung auf die Einheitlichkeit der Stämme. Die einzige Abweichung von der Einheitlichkeit ist, dass die Kompositionsglieder in einigen Fällen mit ⟨s⟩ verbunden

92 Das bedeutet nebenbei auch, dass es ein kompliziertes Unterfangen ist, die phonologische und die graphematische Form der Ableitung aufeinander zu beziehen. Die einfachste Möglichkeit dazu ist der Weg über die phonologische und graphematische Form der Basis. Auf diesen Basisformen operieren dann die regelmäßigen graphematischen und phonologischen Wortbildungsregeln (inklusive Vokalschwächung, wenn man sie in den Wortbildungsregeln formalisieren möchte). Das Ergebnis ist – wenig überraschend – eine größere Diskrepanz zwischen graphematischer und phonologischer Form als bei der Basis; jede der beiden Basisformen ist eben Gegenstand von domänenspezifischen Prozessen.

93 Diese Probleme sind sicher nicht unlösbar. So ist bspw. auch vorstellbar, datengeleitet vorzugehen: Man könnte die morphologischen Analysen in CELEX zugrundelegen (z. B. *divinity* als *divine+ity*) und den Grad der phonologischen ‚Abweichungen‘ von einer reinen Konkatenation (also einer einheitlichen Realisierung von Stämmen und Affixen) automatisch berechnen (in diesem Fall Vokalalternation /ɹ/ vs. /ɪ/). Um zu interessanten Aussagen über den Grad expliziter morphologischer Schreibungen zu kommen, ist dann aber eine weitere Klassifikation der Alternationen nötig.

werden (z. B. *sportsman*). Marchand (1969: 27) plädiert hier für eine Klassifikation als Fugenelement („linking element“) im Gegensatz zu einer Analyse des -s als Flexionssuffix. Unabhängig davon, wie dieses <s> nun eingeordnet wird, ist die nicht-konkatenative Verknüpfung im Englischen wohl marginal:

It seems safe to say [...] that neither internal inflection nor the use of linking elements plays any strong role in English compounding. (Lieber 2009: 369)

Im Deutschen sind Fugenelemente sehr viel häufiger, auch wenn sie insgesamt nur in etwa einem Drittel aller Komposita auftreten (vgl. Fuhrhop/Kürschner 2015). Außerdem werden – wie bei der Flexion und Derivation – z. T. Pseudosuffixe getilgt; dazu unten mehr. Zunächst zu den Fugenelementen. Das Inventar umfasst -s, -es, -n, -en, -ens, -er und -e (Augst 1975); diese Elemente bilden eine Teilmenge des Inventars der Flexionssuffixe. Von diesen Fugenelementen sind nur -s und -n/-en produktiv (Fuhrhop 1998). Dabei ist -s wohl vor allem phonologisch konditioniert, wie Nübling/Szczepaniak (2008) zeigen: Es kommt umso öfter vor, je mehr die prosodische Struktur des Erstglieds vom Trochäus (oder vom Einsilber) abweicht. Das Fugenelement -n hingegen ist vor allem morphologisch zu beschreiben: Es tritt bei morphologisch einfachen Erstgliedern auf -e auf (z. B. *Blumenwiese*). Sobald das finale -e des Erstglieds eine Funktion trägt (z. B. die des deadjektivischen Derivationsuffixes), werden Erst- und Zweitglied ohne Fuge verbunden (z. B. *Frischebehälter*, vgl. Fuhrhop 2000).

Die ‚Verfügung‘ eines Erstglieds auf <e> mit der Fuge <-n> (Typ *Blumenwiese*) ist allerdings nicht die einzige Möglichkeit: In einer nicht unwesentlichen Anzahl der Fälle (über 200 zweigliedrige Komposita in CELEX) wird das finale <e> des Stamms getilgt – auch hier wiederum nur, wenn es sich dabei um ein Pseudosuffix handelt (z. B. *Erdboden*, *Kirschkern*). Die übrigen Pseudosuffixe bleiben in Komposita intakt, vgl. z. B. *Angelrute*, **Angrute*; *Balkenlage*, **Balklage*; *Ackerbau*, **Ackbau*.

Fassen wir zusammen: Der Normalfall ist, dass die Stammform auch in Komposita einheitlich bleibt. Im Englischen gilt das wesentlich mehr als im Deutschen; hier tritt teilweise ein Fugenelement an die Stammform, teilweise wird das Pseudosuffix <e> im Erstglied getilgt; es gibt in diesem Bereich also mehr Abweichungen von der Einheitlichkeit als im Englischen. Eine Ausnahme von der Einheitlichkeit der Stämme hat die Überarbeitung der Rechtschreibung von 2006 beseitigt: Drei aufeinanderfolgende Konsonanten an Morphemgrenzen sind jetzt von der Norm lizenziert (Typ *Schiffahrt*); in der alten Rechtschreibung war es nicht möglich, zwischen *Bettuch* und *Betttuch* graphematisch zu differenzieren.

4.2 Stämme: Eindeutigkeit

Graphische Formen können auf mindestens zwei Arten eindeutig einen Stamm repräsentieren. Zum einen ist es vorstellbar, dass ein Stamm wie *cat* eindeutig von der Graphemfolge ⟨cat⟩ repräsentiert wird – und dass sein Vorkommen innerhalb längerer Stämme vermieden wird. Das wäre dann eine vergleichbare Situation zu Affixen wie *-ous*, das im Englischen eindeutig von der Graphemfolge ⟨ous⟩ repräsentiert wird; diese Schreibung wird als Endung von Stämmen vermieden, die wortfinale Graphemfolge ⟨ous⟩ ist für die Schreibung des Affixes reserviert (vgl. 4.4 und Berg/Aronoff 2017). Das ist für Stämme offensichtlich nicht der Fall: Beliebige gewählte einsilbige Wörter sind regelmäßig Bestandteil längerer Wörter, wie (79) und (80) zeigen:

(79a) Tal: Talg, Taler, Talar, Talk, Talent, Talmud, Talon

(79b) rat[en]: Rate, Ratte, ratter[n], Ration, Ratze

(79c) arm: Armatur, Armee, Armada

(80a) cat: catapult, catamaran, cataract, catarrh, catastrophe, catch, cater, cattle

(80b) pun: punch, pundit, punk, punnet, punt

(80c) fat: fate, fathom, fatigue

Ähnlich äußert sich Hockett (1961: 29), der die partielle Formgleichheit in den Beispielen oben mit Homophonie in Verbindung bringt. Er diskutiert das Verhältnis zwischen Phonemen und Morphemen und konstatiert, dass Morpheme aus Phonemen bestehen, dass aber umgekehrt dieselbe Phonemfolge nicht immer dasselbe Morphem repräsentiert:

It is important to note that the ‘composed of’ assumption has never been seriously taken to imply that every occurrence of a specified arrangement of specific phonemes constitutes an occurrence of one and the same morpheme. Apart from ordinary homophony (*beet* and *beat*), it is permissible for some occurrences of an arrangement of phonemes to be a morpheme, other occurrences of the same arrangement to be less or more than a morpheme – in my speech, *cat* and the beginning of *catalog*.

Psycholinguistisch gibt es zwar Hinweise darauf, dass bei der Verarbeitung von *catalog* auch *cat* aktiviert wird, wie Bowers, Davis und Hanley (2005) zeigen – in einem gewissen Sinne steckt also tatsächlich *cat* in *catalog*. Das hat aber offenbar nicht zur Vermeidung von partieller Formgleichheit in der Schrift geführt.

Zum anderen repräsentiert eine Graphemfolge einen Stamm eindeutig, wenn es keine gleichen Graphemfolgen gibt, die einen anderen Stamm repräsentie-

ren – wenn es also keine homographischen Formen gibt. So repräsentiert ‹Donner› den Stamm des Lexems DONNER eindeutig; ‹Mutter› repräsentiert hingegen zwei distinkte Lexeme, MUTTER (ELTERNTEIL) und MUTTER (SCHRAUBENTEIL). Die Frage, die sich aus dieser Perspektive auf die Eindeutigkeit von Stämmen ergibt, ist also allein: Wie viele Graphemfolgen repräsentieren ihren Stamm eindeutig und wie viele nicht? Oder anders: Wie viele homographische Formen gibt es relativ zu allen Formen?

Wenn wir hier zu einem Ergebnis gekommen sind, können wir – analog zur Eindeutigkeit der Affixe – untersuchen, wie viele der differenzierten Stämme potenziell gleich verschriftet werden *könnten*. Die Messlatte hierfür sind homophone Stämme. Homophone Stämme können trivialerweise auch gleich verschriftet werden, und wir können überprüfen, in wie vielen Fällen das tatsächlich geschieht.⁹⁴

Aus graphemischer Perspektive sind Schreibungen wie ‹Saite› und ‹Seite› nicht auffällig: Hier repräsentieren zwei distinkte Graphemfolgen zwei distinkte Lexeme – das ist erst einmal nichts anderes als die zwei Graphemfolgen ‹Dorf› und ‹Torf›, die sich ebenfalls in einem Graphem unterscheiden und zwei distinkte Stämme repräsentieren.

Wie viele Graphemfolgen repräsentieren im Deutschen und Englischen also mehrere Lexeme? Zur Beantwortung dieser Frage nutzen wir das Teilkorpus homophoner Stämme (siehe 2.4). Im deutschen Korpus gibt es 182 homographische Stämme wie *Bank* oder *Mutter* – bezogen auf die Grundgesamtheit von 5.485 Stämmen im Teilkorpus einfacher deutscher Stämme sind das 3,3 %. Die große Mehrzahl aller graphematischen Stämme im Deutschen repräsentiert also genau ein Lexem, oder anders: Die meisten Stämme sind in ihrer geschriebenen Form eindeutig. Im Englischen ist das ähnlich. Hier gibt es im entsprechenden Teilkorpus 342 homographische Stämme wie *down* oder *race*. Auf die Gesamtzahl aller 7.004 einfachen Stämme im Teilkorpus bezogen sind das 4,9 %. Der Wert liegt etwas höher als im Deutschen; dennoch repräsentieren im Englischen 95 % aller graphematischen Stammformen genau ein Lexem.

Das ist zunächst ein eindeutiges Ergebnis: Heterographie ist der Normalfall. Das ist nicht überraschend; es kann vermutet werden, dass Schriftsysteme (oder Sprachen generell) nur ein gewisses Maß an Homonymie (Homographie und/

⁹⁴ Es gibt theoretisch noch eine andere Möglichkeit, den Grad der Heterographie festzustellen, und zwar mithilfe phonologisch distinkter Wörter, die homographisch verschriftet werden könnten – z. B. englisch *bread/breed*. Der letztere Stamm könnte ebenfalls als ‹bread› geschrieben werden. Da die Erhebung potenzieller Homographien dieser Art aber relativ komplex ist, wird hier darauf verzichtet.

oder Homophonie) ‚tolerieren‘ – wie ja umgekehrt auch ‚reine‘ Synonymie vermieden wird (vgl. z. B. Aronoff 2016). Beide Phänomene sind durchaus vergleichbar: Mit ‚reiner‘ Synonymie ist gemeint, dass zwei hinreichend distinkte sprachliche Formen dieselbe Bedeutung haben – das Paar <Photo>/<Foto> fällt nicht darunter, weil sich die Formen nicht hinreichend unterscheiden. Einschlägig sind Paare wie APFELSINE/ORANGE. Mit ‚reiner‘ Homonymie ist gemeint, dass zwei hinreichend distinkte Bedeutungen mit derselben Form verknüpft sind. Paare wie BIRNE (OBST)/BIRNE (LEUCHTMITTEL) fallen nicht darunter, weil sie semantisch aufeinander beziehbar sind.

Mithilfe des Korpus homophoner Stämme (siehe 2.4) können wir die Werte von 3,3 % bzw. 4,9 % Homographen an allen Stämmen mit dem Anteil der Homophone an allen Stämmen vergleichen. Das deutsche Teilkorpus homophoner Stämme umfasst 248 Einträge (vgl. oben 2.4); bezogen auf die Gesamtzahl einfacher Stämme (5.485) sind das 4,5 %. Im englischen Teilkorpus sind es 618 Einträge, das sind relativ zu allen einfachen Stämmen (7.004) 8,9 %. In beiden Sprachen gibt es also mehr Homophone als Homographen. Das englische Korpus ist in mehrererlei Hinsicht extremer als das deutsche: Zum einen enthält es jeweils mehr Homophone und Homographen als das deutsche Korpus; zum anderen ist die Differenz zwischen Homophonen und Homographen wesentlich größer (4 % im Englischen gegenüber 1,2 % im Deutschen). Das ist ein erster Hinweis darauf, dass Homophone im Englischen häufiger graphematisch differenziert werden als im Deutschen. Insgesamt ist aber auch im Englischen der Anteil homophoner Formen an allen Stämmen relativ gering. Darauf macht bereits Bodmer (1944: 438) aufmerksam: „Though homophones are more abundant in English than in any other European languages, English homophones are few compared with the total number of words in common use.“

Wir können nun wie oben angekündigt einen Schritt weiter gehen und untersuchen, wie häufig homophone Lexeme graphematisch differenziert werden und wie häufig nicht. Auf dieser Grundlage lässt sich besser bewerten, wie wichtig Heterographie im Deutschen und Englischen ist und wie Heterographie konkret bewerkstelligt wird. Wir befinden uns hier immer mit einem Bein im Bereich der Phonographie, vor allem, wenn es um Fragen geht, wie graphematische Formen theoretisch auch aussehen könnten.

Im Deutschen umfasst das CELEX-Teilkorpus homophoner Stämme 248 Einträge, das entsprechende englische Korpus hat 618 Einträge. Ein Teil dieser homophonen Formen wird nun graphematisch differenziert, ist also heterographisch (Typ <Seite>/<Saite>, <meet>/<meat>), ein anderer Teil hingegen ist homographisch (Typ <Bank>, <date>) – diese letztere Kategorie ist unabhängig davon, ob sich die Lexeme morphosyntaktisch unterscheiden, wie etwa *Bank* in der Pluralbildung (*Bänke* vs. *Banken*). Tabelle 105 zeigt, wie die Kategorien im deutschen

und englischen Korpus verteilt sind (mit der Kategorie ‚heterographisch‘ werden auch homophone Lexeme erfasst, von denen nur ein Teil differenziert wird, z. B. WEISE₁, WAISE, WEISE₂):

Tab. 105: Anzahl der Homophone; Anzahl der heterographischen und homographischen Homophone. Datenbasis: Korpus homophoner Stämme deutsch und englisch.

		deutsch	englisch
#homophone Formen		248	618
davon	homographisch	182 (73 %)	327 (53 %)
	heterographisch	66 (27 %)	291 (47 %)

Der Anteil der heterographischen Homophone liegt – wie oben bereits angemerkt – im Englischen deutlich höher als im Deutschen. Und das Verhältnis verschiebt sich noch weiter, wenn wir im Deutschen Heterographen wie in (81) ausschließen:

- (81a) gelt[en]/Geld, gleiß[en]/Gleis, leit[en]/leid[en], Rat/Rad, walt[en]/Wald
 (81b) ess[en]/es, führ[en]/für, Reh/Re, wehr[en]/wer

Diese Paare sind jeweils heterographisch. Kann diese Differenz aber auf ein Heterographieprinzip zurückgeführt werden? Nein, denn die Schreibung ergibt sich aus anderen Prinzipien. In (81a) ist es das Prinzip der Stammkonstanz. In mehrsilbigen Flexionsformen wie *Räder* ist das stammfinale Phonem /d/, entsprechend lautet die graphematische Form ⟨Räder⟩ (*⟨Räter⟩) – und diese Schreibung wird für den Stamm in allen Umgebungen verwendet. Es gibt im Korpus deutscher homophoner Stämme 18 Einträge vom Typ in (81a). In (81b) ist es die Beschränkung über minimale Stämme, die die Form der beiden Inhaltswörter motiviert (siehe oben 3.2.3): ⟨r⟩ reicht als einfacher Endrand des Stamms (bei einfach besetztem Anfangsrand) nicht aus, hier tritt zusätzlich ⟨h⟩ auf, um die minimale Struktur zu gewährleisten. Für die Präposition *für* gilt diese Beschränkung nicht, da sie ein Funktionswort ist. Analog liegt der Fall bei ⟨ess[en]⟩: Hier verletzt der Stamm ohne Doppelkonsonant die Mindestlänge von drei Graphemen; für das Funktionswort *es* und das Fremdwort/Kurzwort *Re* gilt diese Beschränkung nicht. Im deutschen Korpus finden sich insgesamt neun analoge Fälle.

Konsequenterweise müssen dann im englischen Korpus auch Paare wie die folgenden ausgeschlossen werden, in denen Inhaltswörter die Minimalitätsbeschränkungen erfüllen:

(82) bee/be, by/buy, inn/in, no/know, we/wee

Es gibt insgesamt 10 solcher Fälle, die im Folgenden ausgeklammert werden. Hier handelt es sich zwar um heterographische Schreibungen – aber sie sind auch mit einem unabhängig motivierten Prinzip erklärbar. Außerdem werden wie im deutschen Teilkorpus fünf Paare graphematischer Varianten ausgeschlossen (z. B. <gingko>/<ginkgo>).

Damit verbleiben im deutschen Korpus 39 Heterographen von 222 Homophonen – das sind 18 %. Im Englischen verbleiben 281 Heterographen von 608 Homophonen; das ist ein Anteil von 46 %. Der Anteil im englischen Korpus liegt damit mehr als doppelt so hoch wie im deutschen Korpus; das ist eine wichtige Beobachtung. Offenbar ist die Eindeutigkeit von Stämmen weder im Deutschen noch im Englischen ein besonders wichtiges Strukturprinzip; im Englischen wird aber immerhin fast die Hälfte der Homophone graphematisch differenziert.⁹⁵

Dieses etwas kleinere Korpus wird nun weiter untersucht. Wir können uns zunächst die silbische Struktur ansehen. Abbildung 48 zeigt, dass die meisten Homophone in beiden Korpora (unabhängig davon, ob sie graphematisch differenziert werden oder nicht) graphematisch einsilbig sind:

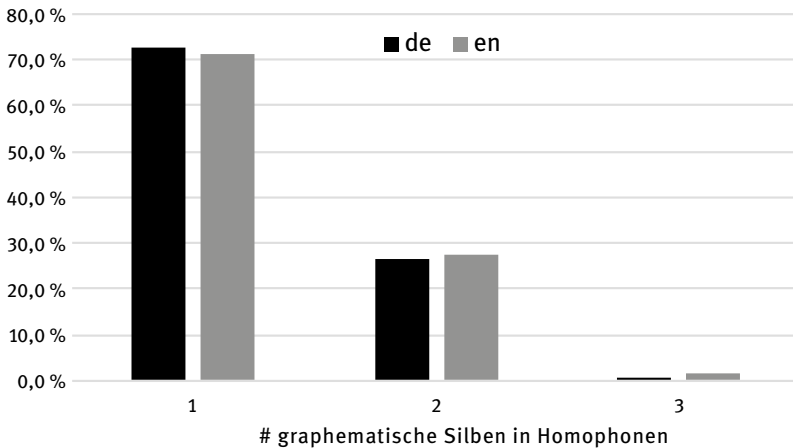


Abb. 48: Homophone im deutschen und englischen Korpus nach Anzahl der graphematischen Silben. Datengrundlage: CELEX-Teilkorpus homophoner Stämme.

⁹⁵ Wir haben hier alle nicht-heterographischen Formen als potenziell heterographisch klassifiziert. Das ist nicht immer der Fall (vgl. z. B. <Schimmel>, das kaum anders geschrieben werden könnte), aber das Bild ändert sich nicht grundsätzlich, wenn wir uns nur diejenigen Wörter ansehen, die tatsächlich alternativ verschriftet werden könnten (siehe unten).

Graphematisch differenzierte Homophone zeigen ungefähr die gleiche Verteilung – im deutschen Korpus sind 66 % der Heterographen Einsilber, im englischen sind es 69 %. Alle Zweisilber im deutschen Korpus haben ein ⟨e⟩ in der zweiten Silbe (Primus' kanonischer Trochäus, z. B. ⟨Sohle⟩/⟨Sole⟩). Im Englischen enden 60 der 80 Zweisilber mit finalem ⟨e⟩ (⟨hoarse⟩/⟨horse⟩), das sind 75 %.

Die Konzentration auf den Einsilber und den kanonischen Zweisilber ist nicht überraschend: Schließlich geht es hier um Formen, die wahrscheinlich zufällig identisch sind. Die Wahrscheinlichkeit dafür ist umso größer, je kürzer sie sind (und je einfacher die Silbenkonstituenten aufgebaut sind).

Interessanter wird es, wenn wir weiter ins Detail gehen und untersuchen, wie Heterographie ‚hergestellt‘ wird – worin sich also die heterographischen Formen unterscheiden. Diese Angaben müssen natürlich qualifiziert werden, und zwar durch die Anzahl der jeweils nicht-differenzierten Homophone. Es gibt bspw. im deutschen Korpus sieben Heterographen, die sich in der Schreibung des Diphthongs /ai/ unterscheiden (z. B. ⟨Saite⟩/⟨Seite⟩). Erweitern wir nun den Fokus und schauen uns an, wie viele Homophone auf diese Weise graphematisch differenziert werden könnten, so wird deutlich, dass die Differenzierung die Ausnahme ist – es gibt 19 Fälle wie ⟨Heide⟩, ⟨F/feige⟩ oder ⟨wein[en]⟩, in denen die Lexeme nicht unterschieden werden. /ai/ wird also nur in 27 % aller Fälle graphematisch differenziert.

Beginnen wir mit der Frage, ob in Heterographen eher Silbenkerne oder -ränder differenziert werden. Carney (1994: 403) vermutet für das Englische, Variation in der Vokalschreibung sei „the most fruitful source of homophones“. Für das Deutsche stimmt das nur eingeschränkt – es hängt vor allem an der Frage, ob postvokalisches ⟨h⟩ eine Vokalschreibung ist (siehe unten). Wir klassifizieren die Schreibung zunächst im Einklang mit den distributionellen Eigenschaften von ⟨h⟩ als Randschreibung. Ein Paar wie ⟨mahl[en]⟩/⟨mal[en]⟩ wird daher als Instanz von Variation in der Schreibung von Silbenrändern gezählt (⟨l⟩ vs. ⟨hl⟩). Tabelle 106 präsentiert die Anzahl der Heterographen für die häufigsten Muster der Differenzierung. Die Summe der Differenzierungen ist höher als die Gesamtzahl der Heterographen (39); das liegt daran, dass teilweise mehrere Muster gleichzeitig präsent sein können (z. B. ⟨lehr[en]⟩/⟨leer⟩).

Tab. 106: Übersicht über die häufigsten Muster zur Differenzierung von Homophonen im deutschen Teilkorpus homophoner Stämme.

Muster	#Heterographen	Beispiel
Silbenrand	26	
⟨h⟩	9	mal[en]/mahl[en]
⟨x⟩/⟨chs⟩	3	lax/lachs
⟨f⟩/⟨v⟩	2	fließ[en]/Vließ
Rest	12	Stadt/statt, Verb/werb[en]
Silbenkern	17	
⟨i⟩/⟨ie⟩	5	Miene/Mine
Doppelvokale	5	bet[en]/Beet
⟨ai⟩/⟨ei⟩	4	Saite/Seite
Rest	3	Ferse/Färse, biet[en]/Beat

Differenzierungen des Silbenrands überwiegen im Deutschen – zumindest, wenn man ⟨h⟩ als Silbenrandschreibung klassifiziert, ansonsten dreht sich das Verhältnis um. Die dominanten Muster sind postvokalisches ⟨h⟩ und die Variation zwischen einfachem und komplexem Silbenkern. Wir haben es hier – aus einer etwas anderen Perspektive – mit uneindeutigen Phonem-Graphem-Beziehungen zu tun: Ein Phonem kann auf mehr als eine Art verschriftet werden, und das Schriftsystem macht davon Gebrauch. Das Ergebnis sind nicht graphematische Varianten wie ⟨Foto⟩/⟨Photo⟩, sondern distinkte Lexeme. Mehr noch: Die Mittel, die bei graphematischen Varianten im Fremdwortbereich Anwendung finden (⟨f⟩/⟨ph⟩, ⟨c⟩/⟨k⟩, ⟨c⟩/⟨z⟩ etc., vgl. Eisenberg 2011: 327 ff.), sind hier gerade nicht einschlägig. Die uneindeutigen Phonem-Graphem-Korrespondenzen in Tabelle 106 gehören alle zum Kernbereich; es gibt sie aber auch im Fremdwortbereich, wenn fremde Schreibungen nicht integriert werden. Solche Korrespondenzen finden sich in der großen Restgruppe bei den Silbenrändern in Tabelle 106. Sie setzt sich vor allem aus Heterographen zusammen, die ein rezentes Fremdwort oder einen Eigennamen enthalten (⟨Bohle⟩/⟨Bowle⟩, ⟨Kot⟩/⟨Code⟩, ⟨Scheck⟩/⟨Shag⟩, ⟨lock[en]⟩/⟨Log⟩, ⟨senn[en]⟩/⟨Zen⟩, ⟨Herz⟩/⟨Hertz⟩, ⟨Kahn⟩/⟨Khan⟩). Die Variation hier ist weit weniger systematisch als die Variation im Kernbereich und wird im Folgenden ignoriert.

Drehen wir nun die Betrachtungsrichtung um und sehen uns die Fälle an, die potenziell mit den Mitteln in Tabelle 106 differenziert werden könnten. Als solche werden angesetzt:

- ⟨h⟩-Schreibung: alle Einträge mit gespanntem /a/, /e/, /o/, /u/, /ø/, /y/ vor /l/, /n/, /m/, /r/,
- ⟨x⟩/⟨chs⟩: alle Einträge mit finalelem /ks/,
- ⟨f⟩/⟨v⟩: alle Einträge mit /f/,

- ⟨i⟩/⟨ie⟩: alle Einträge mit gespanntem /i/,
- Doppelvokale: alle Einträge mit gespanntem /a/, /e/ und /o/,
- ⟨ai⟩/⟨ei⟩: alle Einträge mit /ai/.

Die Anzahl der jeweiligen Muster unter den nicht-differenzierten Homophonen ist in Tabelle 107 aufgelistet und wird mit der Anzahl der differenzierten Homophone verglichen. Im Kern geht es um die Frage, wie konsistent die Muster jeweils Homophone differenzieren: Welche Einträge außer den Heterographen könnten differenziert werden, werden aber gleich geschrieben?

Tab. 107: Graphematische Muster, die zur Differenzierung von Homophonen im Deutschen verwendet werden, nach der Anzahl, mit der sie diese Homophone differenzieren, und der Anzahl der potenziellen Differenzierungen.

Muster	#Heterographen	#Homographen	Beispiel Homograph
±⟨h⟩	9 (36 %)	16 (64 %)	Flur, scher[en]
⟨x⟩/⟨chs⟩	3 (60 %)	2 (40 %)	Wachs, Koks
⟨f⟩/⟨v⟩	2 (9 %)	21 (91 %)	Flur, fast[en]
⟨i⟩/⟨ie⟩	5 (33 %)	10 (67 %)	Krieg, Kiefer
Doppelvokale	5 (17 %)	25 (83 %)	kehr[en], les[en]
⟨ai⟩/⟨ei⟩	4 (17 %)	20 (83 %)	feige, Heide

Die meisten Muster werden nur in einer Minderheit der möglichen Fälle differenziert. Die Ausnahme ist die Silbenrandschreibung ⟨x⟩/⟨chs⟩, die häufiger heterographisch ist als homographisch. Hier liegen aber insgesamt nur fünf Fälle vor – es ist auf dieser Grundlage schwer zu beurteilen, wie regelmäßig das Muster ist. Wir können festhalten: Mit Ausnahme dieser wenigen Fälle werden die häufigsten Muster nicht einmal in der einfachen Mehrzahl der Fälle differenziert. Genau wie Lexemdifferenzierung allgemein können auch die einzelnen Muster keine Systematik beanspruchen; beides ist logisch unabhängig voneinander, es könnte durchaus auch anders sein.

Im englischen Korpus gibt es wesentlich mehr Heterographen (siehe oben). Die häufigsten Muster sind in Tabelle 108 (nächste Seite) zusammengefasst. Postvokalisches ⟨w⟩ wird hier – anders als ⟨h⟩ im deutschen Korpus – als Silbenkernschreibung klassifiziert. Dafür gibt es distributionelle Gründe: ⟨w⟩ und ⟨u⟩ sind postvokalisches komplementär verteilt (Berg/Fuhrhop 2011: 460 f.).

Silbenkerne werden im englischen Korpus häufiger differenziert als Silbenränder. Die beiden hauptsächlichen Mittel dafür sind Variation in Schreibdiphthongen (Typ ⟨meat⟩/⟨meet⟩) und finales ⟨e⟩ (Typ ⟨loan⟩/⟨done⟩). Bei den Schreibdiphthongen

sind vor allem Verschriftungen von /i/ prominent (⟨ee⟩, ⟨ea⟩ und in geringerem Maße ⟨ie⟩). Diese Verschriftungen machen die Hälfte der 48 Heterophone dieser Kategorie aus. Die Restkategorien sind sowohl bei den Kernen als auch bei den Rändern relativ umfangreich; das deutet darauf hin, dass viele Homophone eben nicht systematisch, sondern eher idiosynkratisch differenziert werden.

Tab. 108: Übersicht über die häufigsten Muster zur Differenzierung von Homophonen im englischen Teilkorpus homophoner Stämme.

Muster	#Heterographen	Beispiel
Silbenrand	144	
⟨C⟩/⟨CC⟩	22	grip/grippe
⟨w⟩/⟨wh⟩	14	wit/whit
⟨n⟩/⟨kn⟩	7	not/knot
Rest	101	bow/bough, core/corps
Silbenkern	225	
⟨VV⟩/⟨VV⟩	53	meat/meet
⟨VV⟩/⟨V_e⟩	51	loan/lone
⟨i⟩/⟨y⟩	12	die/dye
⟨V⟩ in unbetonten Silben	14	taper/tapir
⟨Vr⟩/⟨Vr⟩	11	fir/fur
Rest	85	colonel/kernel, choir/quire

Kehren wir auch hier die Blickrichtung um und fragen, wie häufig diese Muster jeweils Lexeme differenzieren könnten.

- Doppelkonsonanten: Alle Einträge, die nach ungespanntem Vokal einen einfachen Konsonanten aus der Gruppe /b/, /d/, /g/, /p/, /t/, /l/, /m/, /n/, /r/, /s/, /z/, /f/ enthalten – das ist grob die phonologische Umgebung für graphematische Doppelkonsonanten (vgl. oben 4.1.1). Wir beschränken uns hier auf phonologische Einsilber.
- ⟨w⟩/⟨wh⟩: Alle Einträge mit einfachem initialem /w/.
- ⟨n⟩/⟨kn⟩: Alle Einträge mit einfachem initialem /n/.
- ⟨VV⟩/⟨V_e⟩: Hier muss etwas ausgeholt werden. Finales ⟨e⟩ kodiert unterschiedliche phonologische Monophthonge und Diphthonge, /eɪ/ (*lane*), /i:/, (*gene*), /aɪ/ (*kite*), /əʊ/ (*lone*) und /u:/ (*cute*) sowie /ɔ:/ und /ɛ:/ (in rhotischen Varietäten vor /ɹ/, sonst stammfinal z. B. *bore*, *flare*), vgl. z. B. Venezky (1970: 102). Jeder dieser Monophthonge und Diphthonge kann auch mit einem Schreibdiphthong verschriftet werden – außer /aɪ/ (hier wird auf ⟨igh⟩ ausgewichen). /u:/ korrespondiert nur in drei Umgebungen mit ⟨u_e⟩: Nach

/j/ (*cute*), nach /l/ (*lute*) und nach /r/ (*rude*). Für /eɪ/, /i:/, /aɪ/ und /əʊ/ gibt es keine solchen Beschränkungen. Alle Heterographen mit finalem ⟨e⟩ im Korpus sind phonologische Einsilber mit geschlossenem Endrand; wir suchen also nach allen phonologischen Einsilbern mit /eɪ/, /i:/, /əʊ/, /aɪ/ oder /{j,l,r}u:/ als Kern.

- ⟨VV⟩/⟨VV⟩: Hier greifen wir die Verschriftungen von /i:/ als dem prominentesten Muster heraus. Alle Heterophone, die /i:/ differenzieren, sind phonologische Einsilber. Wir suchen also phonologisch einsilbige Formen mit /i:/.
- ⟨i⟩/⟨y⟩: Dieses Paar kommt in etwa drei Viertel der Fälle in Verschriftungen des phonologischen Diphthongs /aɪ/ vor (⟨stile⟩/⟨style⟩). Wir beschränken uns auf diese Struktur und suchen Homophone mit /aɪ/.
- ⟨V⟩ in unbetonten Silben: Wir suchen hier alle Homophone mit Monophthongen in unbetonten Silben.
- ⟨Vr⟩/⟨Vr⟩: Hier werden alle Einträge mit /ɜ:/ gesucht; dieses Muster trifft auf alle differenzierten Homophone mit Ausnahme des idiosynkratischen *bury* zu.

Legt man diese phonologischen Bedingungen an die Homophone im Korpus an, ergeben sich folgende Zahlen und Verhältnisse (siehe Tab. 109 folgende Seite).

Die Werte liegen im englischen Korpus deutlich höher als im deutschen; es werden jeweils mehr Homophone differenziert. Dennoch ist auch das Englische weit von einer konsequenten Anwendung des Prinzips entfernt, Stämme eindeutig zu verschriften, wie in Tabelle 109 zu sehen ist. Die einzige Ausnahme ist finales ⟨e⟩ in Homophonen mit /ɛ:/ als Kern (in rhotischen Varietäten vor /ɹ/, sonst stammfinal); hier wird jedes einzelne Wort im Korpus differenziert. Das kann aber auch an der relativ geringen Datenmenge liegen. Also: Kaum ein Muster wird konsequent angewendet, und es gibt einen großen idiosynkratischen Bereich – all das spricht dafür, dass die Differenzierung von Homophonen im Englischen (wie im Deutschen) nicht systematisch ist.

Einige Muster zeigen dennoch einen bemerkenswert hohen Grad der Ausnutzung. So werden etwa zwei Drittel aller Homophone, die /i:/ enthalten, im Geschriebenen differenziert. Wie ist das zu erklären? Es spricht alles dafür, dass wir es hier mit sprachhistorischen Überbleibseln zu tun haben. Die Schreibungen ⟨ee⟩ und ⟨ea⟩ korrespondierten bis ins 16. Jahrhundert mit zwei distinkten phonologischen Vokalen; die phonologische Unterscheidung ging im Laufe der Jahrhunderte verloren, und ehemals heterophonische Lexeme wie ⟨meat⟩/⟨meet⟩ wurden homophonisch; die Schrift konserviert die phonologisch überholte Unterscheidung, ⟨meat⟩/⟨meet⟩ sind heterographisch (vgl. z. B. Dobson 1968: 606 ff.; Labov 1994: 295 ff.). Es handelt sich hier also nicht um eine graphematische *Differenzierung* (im diachronen Sinne einer Herausbildung der Diffe-

renzschrift), sondern um eine graphematische *Konservierung* lexikalischer Distinktionen.

Tab. 109: Graphematische Muster, die zur Differenzierung von Homophonen im Englischen verwendet werden, nach der Anzahl, mit der sie diese Homophone differenzieren, und der Anzahl der potenziellen Differenzierungen.

Muster	#Heterographen	#Homographen	Beispiel Homographen
⟨C⟩/⟨CC⟩	6 (7 %)	83 (93 %)	lap, fan
⟨w⟩/⟨wh⟩	14 (56 %)	11 (44 %)	well, will
⟨n⟩/⟨kn⟩	7 (58 %)	5 (42 %)	nick, nip
⟨VV⟩/⟨V_e⟩	53 ⁹⁶ (50 %)	71 (50 %)	
/ei/	21 (51 %)	20 (49 %)	rail, train
/i:/	1 (10 %)	9 (90 %)	fleet, seal
/aɪ/	5 (24 %)	16 (76 %)	flight/*flite
/əʊ/	5 (31 %)	11 (69 %)	cope, toll
/{j,l,r}u:/	4 (44 %)	5 (56 %)	mule, prune
/ɔ:/	9 (47 %)	10 (53 %)	gore, shore
/eə/	8 (100 %)	–	–
⟨VV⟩/⟨VV⟩ (/i:/)	26 (72 %)	10 (28 %)	lean, seal
⟨i⟩/⟨y⟩ (/aɪ/)	8 (26 %)	23 (74 %)	like, type
⟨V⟩ in unbetonten Silben	14 (29 %)	35 (71 %)	racket, pupil
⟨Vr⟩/⟨Vr⟩	11 ⁹⁷ (50 %)	11 (50 %)	firm, stern

Inwiefern sich diese Beobachtung auch auf die anderen englischen Muster sowie auf das Deutsche übertragen lässt, muss an dieser Stelle offenbleiben. Es spricht aber einiges dafür: So korrespondierten im Deutschen viele Instanzen des postvokalischen ⟨h⟩ mit einem Frikativ (vgl. Roemheld 1955); dieser Frikativ wurde diachron getilgt, während das ⟨h⟩ in der Schrift konserviert wurde.

96 Die Heterographen ⟨none⟩/⟨nun⟩ und ⟨some⟩/⟨sum⟩ sind idiosynkratisch und werden ausgeschlossen; außerdem werden vier ⟨i_e⟩/⟨igh⟩-Varianten (*site/sight* u. a.) aufgenommen, die keine ⟨VV⟩/⟨V_e⟩-Alternationen im engeren Sinne sind. So ergibt sich die Abweichung von 53 ⟨VV⟩/⟨V_e⟩-Heterographen in Tabelle 109 und 51 solcher Heterographen in Tabelle 108.

97 Die phonologische Form des Paares ⟨bury⟩/⟨berry⟩ enthält nicht /ɜ:/ und wird daher hier nicht mitgezählt; so ergibt sich die Abweichung zu Tabelle 108.

Fassen wir zusammen: Im Englischen werden mehr Homophone graphematisch differenziert als im Deutschen. In beiden Sprachen werden Homophone aber – das ist entscheidend – nicht systematisch differenziert. Warum auch, könnte man fragen: Der Formzusammenfall bereitet im Gesprochenen keine großen Probleme, warum sollte das im Geschriebenen anders sein? Die vorhandenen heterographischen Schreibungen können wahrscheinlich als konservierte phonologische Distinktionen analysiert werden.

4.3 Affixe: Einheitlichkeit

Die Untersuchung der Einheitlichkeit von Affixen ist grob in die Bereiche Flexion (4.3.1) und Derivation (4.3.2) gegliedert; diese Kategorien haben hier eine andere Funktion als oben in 4.1. Dort spezifizieren sie den Kontext, in dem Stämme variieren, hier sind sie den Wortkategorien von Lexemen vergleichbar: Ein Affix ist entweder ein Flexions- oder ein Derivationsaffix. Der Kontext wird dann zusätzlich spezifiziert: Variiert die Form des Affixes in Verbindung mit Stämmen (wie bspw. <diskutabel> – <kompressibel>) oder – im Fall von Derivationsuffixen – mit anderen Affixen (wie bspw. <glamour> – <glamorous>)?

4.3.1 Flexion

Wir beginnen hier der Einfachheit halber mit dem Forminventar der Flexionsuffixe im Deutschen ((83), nach Heidolph/Flämig/Motsch 1984: 485) und Englischen ((84), nach Blevins 2006):⁹⁸

(83a) Substantivflexion, deutsch:

-e, -er, -(e)n, -n, -(e)s, -s

(83b) Adjektiv- und pronominale Flexion, deutsch:

-e, -(e)n, -(e)m, -(e)r, -(e)s, -(e)st

⁹⁸ Abweichend von Blevins (2006) wird hier auch das possessive -'s als Flexionsuffix angesetzt. Die Alternative ist die Klassifikation als Klitikon (so z. B. Bauer 2003) oder als phrasales Affix (z. B. Zwicky 1987), da -'s nicht auf Wörtern, sondern auf Phrasen operiert, vgl. [*the director of personell*]'s office). Die hier zu untersuchende Frage der Einheitlichkeit ist aber unabhängig vom kategorialen Status von -'s; daher wird das Element hier ebenfalls untersucht.

- (83c) Verbflexion, deutsch:
-e, -(e)st, -(e)t, -(e)n, ge-...-(e)t, (end)⁹⁹
- (84a) Substantivflexion, englisch:
-(e)s, -'s
- (84b) Adjektivflexion, englisch:
-er, -est
- (84c) Verbflexion, englisch:
-(e)s, -ed, -ing

Im Deutschen ist das Inventar wesentlich umfangreicher als im Englischen. Zusätzlich sind viele Suffixe polyfunktional. So markiert bspw. das Adjektivsuffix *-e* den Nominativ Singular der schwachen Flexion (*der schnelle Läufer, die schnelle Läuferin, das schnelle Kind*), den Akkusativ der Feminina und Neutra in der schwachen Flexion, den Nominativ und Akkusativ Singular von Feminina in der starken Flexion (*frische Milch*) sowie den Nominativ und Akkusativ Plural in der starken Flexion (*nette Kollegen*). Da die Frage der Einheitlichkeit erst einmal unabhängig von der Polyfunktionalität ist, gehen wir hier von der Form – also den Listen in (83) und (84) – aus und vermerken gegebenenfalls funktional differenzierte Variation.

Beginnen wir mit den deutschen Suffixen. Viele Flexionssuffixe haben nur eine Form; sie sind einheitlich. Im Folgenden werden nur die Ausnahmen von dieser Generalisierung diskutiert.

-(e)n PLURAL (und andere (e)-Suffixe)

Das Pluralsuffix *-(e)n* variiert (wie schon die Benennung suggeriert) hinsichtlich der An-/Abwesenheit des ⟨e⟩. Die Form lautet ⟨n⟩, wenn der Stamm bereits auf ⟨e⟩ endet (*Rose – Rosen*), ansonsten lautet sie ⟨en⟩ (*Held – Helden*). Anders ausgedrückt: Das Suffix ist nicht immer silbisch. Das ist die traditionelle Analyse (z. B. Eisenberg 2013a: 156 ff.): Das Suffix ist nicht einheitlich. Alternativ kann man für Substantive mit einfachem finalem ⟨e⟩ Stammflexion und nicht Grundformflexion ansetzen, wie es Harnisch (2001) vorschlägt: Der Stamm der Wortform *Rose* ist dann *Ros-*; er ist gebunden und tritt in dieser Form auch in Ableitungen auf (z. B. *Rös-chen*). An diesen gebundenen Stamm tritt dann das Suffix *-en*, das nun einheitlich ist: *Ros-en, Held-en* – übrigens auch im Fremdwortbereich mit seinen ebenfalls häufig gebundenen Stämmen, vgl. *Kont-en*. Die Einheitlichkeit hängt also hier von der morphologischen Interpretation ab. Das Problem an dieser Ana-

⁹⁹ Für das Partizipialsuffix *-end* ist der Status als Flexionssuffix unklar; es könnte sich auch um ein Derivationsuffix handeln (vgl. Fuhrhop/Teuber 2001).

lyse ist, dass das Ergebnis (*ein* formal einheitliches Suffix) bereits als „inter-paradigmatisches Grundprinzip“ (Harnisch 2001: 257) in Harnischs Untersuchung fungiert („gleiche Suffixe hat, was zielorientiert [...] gleichförmig ist“).

Was ist zu tun? Das Dilemma der angemesseneren Segmentierung (*Rose·n* oder *Ros·en*?) ergibt sich nur, wenn man annimmt, dass Wortformen vollständig in Morpheme gegliedert sein müssen (vgl. Anderson 2015: 18). Die Alternative ist, stattdessen wie bspw. Köpcke (1988) von Schemata für bestimmte morphosyntaktische Kategorien zu sprechen oder von ‚schematic exponence‘ wie bspw. Trommer/Zimmermann (2015). Das graphematische Schema bzw. die Wohlgeformtheitsbeschränkung für *n*-Plurale im Deutschen fordert dann, dass die betreffende Wortform auf ⟨en⟩ enden muss, und zwar unabhängig davon, ob im NOM.SG. ein finales ⟨e⟩ auftritt (wie in *Rose*) oder nicht (wie in *Held*).

Auf diese Weise bekommt man auch die Suffixe der adjektivischen und pronominalen Flexion in den Griff, die sich analog verhalten: *-(e)n* (wie in *dünn·en* vs. *müde·n*); *-(e)m* (wie in *dünn·em* vs. *müde·m*); *-(e)r* (wie in *dünn·er* vs. *müde·r*); und *-(e)s* (wie in *dünn·es* vs. *müde·s*). Sie alle haben gemein, dass die flektierte Wortform einem Schema genügt: Sie endet mit ⟨en⟩, ⟨em⟩, ⟨er⟩ oder ⟨es⟩, unabhängig von der Form der Basis. Eine weitere Wohlgeformtheitsbeschränkung kann in Fällen wie *Rosen* dafür sorgen, dass Stamm und Suffix nicht einfach konkateniert werden – indem sie bspw. zwei oder mehr aufeinanderfolgende, nicht stamminterne ⟨e⟩ verbietet (**Roseen*).

***-(e)n* INFINITIV/1.PS.PL/3.PS.PL**

Etwas anders liegt der Fall beim Infinitivsuffix (und den homonymen Suffixen der 1. und 3. Ps. Plural). Bei Verbstämmen, die konsonantisch enden, hat das Suffix die Form ⟨en⟩ (⟨bitten⟩, ⟨singen⟩). Endet der Verbstamm auf ⟨er⟩ oder ⟨el⟩, dann hat es die Form ⟨n⟩ (⟨ködern⟩, ⟨angeln⟩). Warum genau diese Endungen? Weil ⟨r⟩ und ⟨l⟩ die einzigen Grapheme sind, die im Endrand vor ⟨n⟩ auftreten können (vgl. oben 3.2.2.3). Die Wohlgeformtheitsbedingung für den Infinitiv besagt daher, dass die Form a) mit ⟨n⟩ endet und b) der Kern der Silbe, deren Endrand dieses ⟨n⟩ bildet, ein einfaches ⟨e⟩ ist. Dafür müssen wir allerdings die oben skizzierten Regeln zur Silbifizierung anpassen – nach diesen Regeln ist ja ⟨lauer[n]⟩ graphematisch einsilbig. Eine Silbe kann mit maximal zwei Vokalgraphemen im Kern besetzt sein (vgl. für eine restriktivere Beschränkung Fuhrhop/Berg einger.). Wenn wir von links nach rechts silbifizieren, ergibt sich damit automatisch die gewünschte graphematische Silbenstruktur (⟨lau.ern⟩, ⟨scheu.ern⟩, ⟨schrei.en⟩), und die doppelt besetzten Kerne sind die aus den Einsilbern bekannten Schreibdiphthonge (bei der alternativen Silbifizierung ⟨la.uern⟩, ⟨sche.uern⟩, ⟨schre.ien⟩ gilt das nur im letzten Fall; diese Analyse ist also zu verwerfen). Mit dieser Beschränkung können nun auch potenziell problematische Fälle wie ⟨atmen⟩ oder ⟨wappnen⟩ erfasst wer-

den, deren einsilbige Stämme zum graphematischen Zweisilber erweitert werden. Es ist hier nicht möglich, ⟨n⟩ zum Silbenrand hinzuzufügen – das würde die Beschränkungen über den Aufbau von Endrändern verletzen. Nicht erfasst werden die Formen ⟨sein⟩ und ⟨tun⟩. Die Formen ⟨knien⟩ und ⟨schrien⟩ (vgl. Gallmann 1990: 518) sind zwar graphematisch nicht zweisilbig – wir können aber annehmen, dass die im letzten Abschnitt vorgestellte Beschränkung über das Verbot von ⟨ee⟩ an Stammgrenzen entscheidender ist und die Alternative ⟨knieen⟩/⟨geschrieen⟩, die ja historisch belegt ist (siehe oben 3.2.3), verhindert.

-(e)s GENITIV

Das Genitivsuffix variiert – wie das Pluralsuffix *-(e)n* – hinsichtlich der An-/Abwesenheit von ⟨e⟩. Hier ist die Verteilung aber im Gegensatz zu *-(e)n* nicht ausschließlich kontextuell determiniert. Nach stamminalem ⟨s⟩, ⟨ß⟩, ⟨z⟩ oder ⟨x⟩ ist die Form ⟨es⟩ obligatorisch (des ⟨Gleises⟩, ⟨Spaßes⟩, ⟨Scherzes⟩, ⟨Juxes⟩). Diese Verteilung ist ganz offensichtlich phonologisch bedingt: Die genannten Wortformen korrespondieren alle mit phonologischen Wortformen, die stamminales /s/ enthalten. Mit der silbischen Variante des Suffixes /əs/ wird ein Zusammenfall der /s/-Segmente von Stamm und Suffix verhindert. Bei den übrigen Lexemen variiert ⟨e⟩, und es lassen sich lediglich Tendenzen der Verteilung feststellen, die miteinander interagieren (vgl. für das Folgende Bubenhofer/Hansen-Morath/Konopka 2014). Die wichtigsten Faktoren für ⟨es⟩ sind die Häufigkeitsklasse des Lexems (je seltener, desto wahrscheinlicher ist ⟨es⟩), die geographische Herkunft (in Texten aus dem Südosten ist ⟨es⟩ wahrscheinlicher) und der Stammauslaut (nach Konsonantenclustern, Affrikaten und Plosiven ist ⟨es⟩ wahrscheinlicher). Umgekehrt lässt sich für die ⟨s⟩-Formen (unter anderem) beobachten, dass sie nach Schwasilben oder Silben mit Diphthongen bevorzugt sind sowie nach Eigennamen und Fremdwörtern – und dass sich diachron eine Tendenz zugunsten von ⟨s⟩ (und zulasten von ⟨es⟩) abzeichnet.

Interessant ist außerdem ein Vergleich mit dem teilweise homonymen Pluralsuffix *-s*: Dieses Suffix ist nie silbisch (vgl. ⟨die Wracks⟩/*⟨Wrackes⟩, ⟨die Gags⟩/*⟨Gages⟩, ⟨die Profs⟩/*⟨Profes⟩).

-(e)st 2.Ps.SG

Dieses Suffix variiert in zweierlei Hinsicht. Zum einen hat es nach stamminalem ⟨d⟩ oder ⟨t⟩ in schwachen Verben die Form ⟨est⟩ ((du) ⟨betest⟩/*⟨betst⟩; (du) ⟨redest⟩/*⟨redst⟩); in starken Verben mit Vokalwechsel in der 2.Ps.SG. ist die Form hingegen *-st* ((du) ⟨lädst⟩, (du) ⟨rätst⟩). Zum anderen hat es nach stamminalem ⟨s⟩, ⟨ß⟩, ⟨z⟩ oder ⟨x⟩ die Form ⟨t⟩ (⟨du reist⟩, ⟨du reißt⟩, ⟨du reizt⟩, ⟨du hext⟩); wie oben bei Genitivsuffix ist die Verteilung also phonologisch konditioniert. Während es bei ⟨reist⟩ noch unklar ist, ob hier der Stamm oder das Suffix variiert, ist

es bei den übrigen Fällen eindeutig: Es wird eben nicht *‹hest› mit getilgtem ⟨x⟩ für diese Form von HEX[EN] geschrieben; vielmehr wird das ⟨s⟩ des Suffixes getilgt, und das, obwohl es so zu Homographien kommt: ⟨haust› ist die 2.Sg. von HAUS[EN] und HAU[EN] (vgl. Primus 2000: 21). Nach stamminalem ⟨sch⟩ gibt es Variation: Das erwartbare (du) ⟨naschst› steht neben der ⟨s⟩-losen Form (du) ⟨nascht›. Der Duden (2016: 450 f.) vermutet aufgrund nicht weiter spezifizierter Erhebungen, „dass -t allmählich -st verdrängen könnte“.

-(e)st SUPERLATIV

Auch das Superlativsuffix variiert im selben Kontext, also nach stamminalem ⟨s⟩, ⟨ß⟩, ⟨z⟩ oder ⟨x⟩. Anders als das Verbsuffix wird hier aber nicht das ⟨s⟩ getilgt, sondern ein ⟨e⟩ erweitert: ⟨der mieseste Läufer⟩/*‹mieste Läufer›, ⟨die stolzeste Teilnehmerin⟩/*‹stolzte Teilnehmerin›, ⟨das heißeste Essen⟩/*‹heißte Essen›.¹⁰⁰ Darüber hinaus tritt ⟨est› regelmäßig auch nach stamminalem ⟨t› (‹der bunteste Hund›) und ⟨d› auf (‹der fremdeste Geschmack›). Nach stamminalem ⟨p› und ⟨k› (nicht aber nach ⟨b› und ⟨g›) gibt es Variation. Tabelle 110 zeigt die Ergebnisse einiger Varianten von Superlativformen aus dem Cosmas-Korpus W-Öffentlich. Bei den Stämmen auf ⟨p› gibt es kaum Variation.¹⁰¹ Bei denen auf ⟨k› ist das anders. Hier scheint es auf den ersten Blick einen Frequenzeffekt zu geben: Je häufiger die Superlativform insgesamt ist, desto häufiger ist die ⟨e⟩-lose Variante. Das ist eine Vermutung, die an größeren Datenmengen überprüft werden müsste.

Tab. 110: Verteilung der Superlativformen ⟨ste› und ⟨este› von sieben Adjektiven im Cosmas-Korpus W-Öffentlich.

	...ste	...este
knapp...	1.345	35
schlapp...	8	0
dick...	4.410	2
schlank...	145	42
flink...	55	26
krank...	30	20
blank...	2	7

100 Es gibt tatsächlich einen Fall, in dem das ⟨s⟩ getilgt wird, und zwar den Superlativ von *groß* (‹der größte/*großeste/*größeste Kuchen›).

101 Es gibt allerdings in CELEX auch nur zwei weitere Adjektive auf ⟨p›, *plump* und *myop* – letzteres ist als Superlativform im Korpus nicht vertreten, und bei ersterem ist das Problem, dass die Variante ⟨plumpste› mit der Verbform homonym ist. Da das Korpus W-Öffentlich keine Information über die Wortart der Tokens enthält, können beide nicht automatisch differenziert werden.

-(e)t 3.Ps.Sg/2.Ps.Pl/PRÄT

Die Suffixe der 3. Ps. Sg., der 2. Ps. Pl. sowie des Präteritums sind in den meisten Fällen formgleich und lauten ⟨t⟩ ((sie/ihr) ⟨rührt⟩, ⟨spielt⟩, ⟨wohnt⟩, ⟨glaubt⟩; (ich) ⟨rührte⟩, ⟨spielte⟩, ⟨wohnte⟩, ⟨glaubte⟩). Nach basisfinalelem ⟨d⟩ oder ⟨t⟩ treten sie in der Variante ⟨et⟩ auf: (sie/ihr) ⟨leistet⟩, ⟨meldet⟩, ⟨spendet⟩; (ich) ⟨meldete⟩, ⟨spendete⟩. Das gilt auch, wenn die Numerus/Person-Suffixe an Basen mit Tempussuffix treten: (sie/ihr) ⟨leistetet⟩, ⟨meldetet⟩. Diese Beschreibung gilt für die schwachen Verben. Bei den starken Verben ist das anders: Wenn ein Vokalwechsel in der 3. Ps. Sg. auftritt, fällt das Suffix mit dem Stammende zusammen: (sie) ⟨rät⟩, ⟨hält⟩, ⟨gilt⟩. Die einzige Ausnahme ist die Form (er) ⟨lädt⟩, in der ⟨t⟩ als eigenständiges Segment identifizierbar ist (vgl. Duden 2016: 466; Neef 1996: 161 ff.). Die 2. Ps. Pl. ist von dieser Regel nicht betroffen, weil ihre Formen nicht Gegenstand von Vokalhebung sind. So kommt es zu systematischen Paaren wie (sie) ⟨hält⟩ – (ihr) ⟨haltet⟩.

-e 1.Sg.

Für das Suffix der 1. Ps. Sg. gibt es zwar keine Formvariation im engeren Sinne – aber die An-/Abwesenheit des Suffixes als Ganzes variiert. Neben regulären transparenten Formen wie (ich) ⟨gebe⟩ stehen auch (ich) ⟨geb⟩ und (ich) ⟨geb'⟩ (mit graphischer Markierung der Elision). Es handelt sich hier um ein primär phonologisches Phänomen, das auf die Schrift übertragen wird (vgl. Duden 2016: 452); im Geschriebenen, so ist zu vermuten, ist das Suffix *-e* stabiler als im Gesprochenen. Was die Verteilung der An-/Abwesenheit von *-e*/Schwa in konkreten Flexionsformen angeht, gibt es mehrere Hypothesen (vgl. z. B. Raffelsiefen 1995, die vermutet, dass Schwa der 1. Ps. Sg. nicht realisiert wird, wenn damit das Paradigma prosodisch einheitlicher wird – also *hol* statt *hole*, weil auch alle anderen Verbformen im Präsens einsilbig sind). Eine empirisch fundierte Arbeit (die geschriebene und gesprochene Sprache gleichermaßen berücksichtigt) bleibt ein Desiderat.

Bei vielen deutschen Flexionssuffixen gibt es also formale Variation; die Suffixe sind oft nicht einheitlich. Wenn man eine schemabasierte Beschreibung wählt, fallen die meisten Fälle mit schwankendem ⟨e⟩ weg; es bleiben dann aber immer noch die folgenden Suffixe:

- *-(e)s* GENITIV: (des) ⟨Haus-es⟩ vs. (des) ⟨Kraut-s⟩
- *-(e)st* 2.Ps.SG: (du) ⟨tanz-t⟩ vs. (du) ⟨film-st⟩ vs. (du) ⟨find-est⟩
- *-(e)st* SUPERLATIV: (der) ⟨leicht-est-e⟩ vs. (der) ⟨schwer-st-e⟩
- *-(e)t* 3.Ps.SG/2.Ps.PL/PRÄT: (er) ⟨find-et⟩ vs. (er) ⟨spiel-t⟩; (ich) ⟨meld-et-e⟩ vs. (ich) ⟨spiel-t-e⟩
- *-e* 1.SG: (ich) ⟨spiel⟩ vs. (ich) ⟨spiel-e⟩

Mit Ausnahme des Suffixes *-e* 1.SG ist die Variation bei allen Suffixen phonologisch konditioniert. Die korrespondierende phonologische Variation wird nicht überbrückt. Nun kann vermutet werden, dass Formen wie *(du) <filmst> statt (du) <filmst> nicht realisiert werden, weil sie mit den jeweiligen Konjunktivformen zusammenfallen, und dass wir (ich) <spielte> und nicht (ich) *<spielete> schreiben, weil das Suffix phonologisch nicht silbisch ist. Das stimmt zwar; wir können diesen Sachverhalt aber auch anders beschreiben: Im Gegensatz zum Englischen ist die Einheitlichkeit von Affixen offenbar nicht wichtig genug, um bspw. die silbisch-phonographischen Korrespondenzen zu verletzen – eine Situation, die im Englischen systematisch auftritt (vgl. bspw. die phonologisch einsilbigen, graphematisch zweisilbigen Flexionsformen <called> und <begged>). Für diese These spricht auch, dass es im Deutschen auch Variation gibt, wo die oben angebotenen Erklärungsmuster nicht greifen. So wird die phonologisch determinierte Variation von *-(e)st* nach stamminalem /s/ gerade nicht graphematisch überbrückt – wir schreiben eben nicht (du) *<reißst> oder (du) *<tanzst> mit stabiler Form des Suffixes, sondern spiegeln die phonologische Geminatenreduktion.

Überbrückt die Schrift überhaupt phonologische Alternationen der Formen? Das ist zumindest zum Teil der Fall, wie an zwei Beispielen gezeigt wird.

1. Das Suffix *-(e)n* INFINITIV/1.PS.PL/3.PS.PL ((sie kann) <singen>/(wir/sie) <singen>) kann als Schema aufgefasst werden, das eine <e>-Silbe mit finalelem <n> fordert. Die entsprechenden Formen enden (bis auf *sein* und *tun*) alle mit <en>, <eln> oder <ern>; die Markierung der entsprechenden Kategorien ist also relativ aufwendig. Phonologisch ist das anders: Hier sind die Formen nicht immer durch eine zusätzliche Silbe vom Stamm unterschieden. Besonders instruktiv sind vokalisch auslautende Stämme wie GEH[EN] mit einsilbiger Infinitivform im Gesprochenen, /ge:n/. Ähnliches gilt für Stämme mit finalen Liquiden wie FÜHL[EN] oder HÖR[EN], deren Infinitive wohl ebenfalls als einsilbig anzusetzen sind (/fy:ln/, /hœ:n/). Nach stamminalem Obstruenten hingegen wird das Suffix silbisch realisiert, z. B. in BITT[EN] oder FEG[EN] (/bitn/, /fe:gŋ/). Das Suffix variiert also im Gesprochenen; das gilt umso mehr, wenn man noch Assimilationsformen wie in KAPP[EN] (/kapn/) oder LEB[EN] (/le:bŋ/) miteinbezieht. Das Suffix ist in der Schrift wesentlich einheitlicher kodiert als im Gesprochenen.

2. Das Komparativsuffix *-er* lautet bei nicht-flektierten Adjektiven /e/ (<schöner> – /ʃøne/). In flektierten Formen hingegen tritt die Variante /ər/ auf ((eine) <schönere (Vase)> – /ʃønərə/). In der Schrift ist die Form des Suffixes hingegen einheitlich <er>.

Bei vielen anderen Suffixen ist die Lage nicht so klar. Eine phonetische Reduktion oder Assimilation der Suffixe scheint häufig möglich – wie weit sie aber gehen kann und ob sie überhaupt geschehen muss, ist eine offene Frage. Ist der Reim von *blauen* in *Das sind die blauen Reiter* einsilbig und damit identisch

mit dem Reim von *Zaun*? Ist das immer der Fall? Bevor diese Fragen phonetisch-phonologisch geklärt sind, kann die spezifische Leistung der Schrift nicht untersucht werden.

Im Englischen gibt es einerseits deutlich weniger Flexionssuffixe, und gleichzeitig gibt es nur wenige Abweichungen von der einheitlichen Schreibung. Wir betrachten auch hier nur diejenigen Suffixe, die Variation zeigen.

-(e)s PLURAL/3.PS.SG.

Das Suffix für den Plural bei Substantiven (*kites*) und das für die 3. Ps. Sg. bei Verben (*likes*) sind homonym. Anders als im Deutschen ist die Verteilung der Varianten ⟨s⟩ und ⟨es⟩ nur phonologisch zu bestimmen: Nach stammfinalen alveolaren Frikativen (/s/, /z/, /ʃ/ und /ʒ/) tritt die Form ⟨es⟩ auf (z. B. ⟨lenses⟩, (he) ⟨buzzes⟩, ⟨lashes⟩, (it) ⟨matches⟩), in allen anderen Fällen ⟨s⟩. Gegen eine graphematische Formulierung der Regel spricht, dass ⟨es⟩ nach finalelem ⟨ch⟩ auftritt (z. B. ⟨beaches⟩) – aber nur, wenn ⟨ch⟩ mit einem alveolaren Frikativ korrespondiert. Wie Ryan (2016: 102) feststellt, gibt es Fälle, in denen finaleles ⟨ch⟩ stattdessen /k/ entspricht – und hier tritt einfaches ⟨s⟩ auf (z. B. ⟨stomachs⟩, ⟨epochs⟩, ⟨eunuchs⟩, ⟨techs⟩).

Die Variante *-es* tritt außerdem nach einigen Stämmen mit finalelem ⟨o⟩ auf. Carney (1994: 174) bringt in diesem Zusammenhang die nicht-exhaustive Aufzählung ⟨dominoes⟩, ⟨heroes⟩, ⟨potatoes⟩, ⟨tomatoes⟩, ⟨torpedoes⟩ und verweist darauf, dass unsichere Schreiber aus der Pluralform ⟨potatoes⟩ oft die Singularform ⟨potatoe⟩ ableiten.¹⁰² Andere Stämme mit finalelem ⟨o⟩, besonders rezente Lehnwörter, treten nur mit der *-s*-Variante des Suffixes auf; in einer dritten Gruppe gibt es Schwankungen. Carneys Beispiele sind hier ⟨cargo(e)s⟩, ⟨commando(e)s⟩, ⟨halo(e)s⟩, ⟨tornado(e)s⟩ und ⟨volcano(e)s⟩. Ryan (2016: 102) vermutet, dass sich die Form *-s* langsam gegenüber *-es* durchsetzt.

Das ⟨e⟩ vor dem finalelen ⟨s⟩ in Formen wie *marries* wurde oben als Teil des Stamms interpretiert, nicht wie bei Bauer/Lieber/Plag (2013: 69) als Teil des Suffixes: In den einschlägigen Formen, deren Basen auf ⟨y⟩ enden, alterniert dieses ⟨y⟩ mit ⟨i⟩. Eine Schreibung wie *⟨marris⟩ sieht aber nicht aus wie ein guter englischer Plural, weil ⟨marrī⟩ nicht aussieht wie ein guter englischer Singular. Das ⟨e⟩

¹⁰² Der bekannteste Fall dürfte wohl der amerikanische Vizepräsident Dan Quayle gewesen sein, der im Wahlkampf 1992 einen Buchstabierwettbewerb an einer Schule in New Jersey (vor laufenden Fernsehkameras) moderierte und einen zwölfjährigen Schüler korrigierte: Der hatte das Wort *potato* an die Tafel geschrieben, und Quayle verlangte ein finaleles ⟨e⟩. Der Aufschrei, der daraufhin durch die Medien ging, hat wohl dazu beigetragen, dass Bush und Quayle fünf Monate später die Wahl verloren (Horobin 2013: 2f.).

sichert die Segmentierbarkeit und gehört daher zum Stamm (wie in den vergleichbaren Alternationen vom Typ <knife>/<knives> ebenfalls).

-ed PRÄT./PART.

Die Form des Präterital-/Partizipsuffixes ist unklar in Flexionsformen von Stämmen, die auf <ee> auslauten (Ryan 2016: 94 f.). Hier wird ganz offensichtlich nicht einfach konkateniert, vgl. <freed>, *<freed>. In einer schemabasierten Formulierung ist die Frage der Segmentierung weit weniger drängend – wie im Deutschen können wir annehmen, dass mit dem Suffix gebildete Wortformen bestimmten Gestaltbedingungen unterliegen, in diesem Fall, dass das Wort auf <ed> endet. Zusätzlich muss eine Abfolge von zwei <e> verboten werden, wenn sie nicht bereits in der Basis auftreten – wenn <ee> also nicht lexikalisch ist.¹⁰³ Damit können (mit einer Modifikation) auch Wortformen erfasst werden, die auf <é> enden wie *clichéd*. Hier lautet die *-ed*-Form <clichéd>, nicht *<clichéed> oder *<cliched> (Ryan 2016: 95). Offenbar wirkt die Beschränkung über das Verbot von nicht-lexikalischem <ee> auch dann, wenn eines der Grapheme mit einem Diakritikon markiert ist. Die Vermeidung von <ee> (bzw. in diesem Fall <ée>) scheint wichtiger zu sein als die vollständige Stabilität des Suffixes; die Konsequenz ist, dass <clichéd> mit <éd> und nicht mit <ed> endet. Ryan (2016: 95) stellt pointiert fest: „There is no way of recovering <ed> from <clichéd>, except with an eraser.“

Darüber hinaus gibt es diastratische Variation beim Suffix *-ing*: Ryan (2016: 93) macht darauf aufmerksam, dass in „dialect spellings“ teilweise die Form <in’> statt <ing> auftritt (z. B. <twistin’>, <feelin’>). Hier wird wohl die phonologische Variante /ɪn/ des Suffixes lautnah verschriftet – und selbst hier markiert die Schrift mithilfe des Apostrophs, dass sprachliches Material fehlt.

Im Englischen gibt es also nur an zwei Stellen Varianten unter den Flexionsuffixen: Die marginale Gruppe der Stämme auf <é> nimmt auf den ersten Blick nicht <ed>, sondern <d> als Suffix. Diese Verteilung kann mit einigen Zusatzbedingungen erfasst werden. Systematischer ist die Variation zwischen *-s* und *-es*, die auf der phonologischen Struktur des Stamms beruht. Der Fall ist vergleichbar mit dem deutschen Genitivsuffix *-s*, mit dem Unterschied, dass sich die Verteilung im Englischen nicht rein graphematisch fassen lässt. Insgesamt sind Flexionsuffixe im Englischen also einheitlich.

¹⁰³ Möchte man die Verhältnisse zwischen *free*, *-ed* und *freed* derivational beschreiben, so kann man zu Metaphern wie „dovetailing“ (Carney 1994) oder „telescoping“ (Ryan 2016) greifen.

Das gilt umso mehr, wenn man die korrespondierenden phonologischen Formen miteinbezieht. Die Suffixe *-s*, *-’s* und *-ed* haben jeweils drei phonologische Varianten, /z/, /s/, /ɪz/ bzw. /d/, /t/, /ɪd/. Die Verteilung dieser Varianten ist phonologisch determiniert: Nach stamminalem alveolarem Frikativ bzw. nach alveolarem Plosiv tritt die silbische Variante auf (⟨kisses⟩, ⟨hated⟩); ansonsten tritt nach stimmhaftem Stammauslaut die stimmhafte Variante auf (⟨begs⟩, ⟨robbed⟩), nach stimmlosem Stammauslaut die stimmlose Variante (⟨hats⟩, ⟨peaked⟩), vgl. z. B. Bauer/Lieber/Plag (2013: 68 f.). Mit Ausnahme der phonologisch determinierten Variation bei *-s* wird die gesamte phonologische Variation graphematisch überbrückt. Das gilt auch für *-’s*: Phonologisch findet sich hier dieselbe Variation wie bei *-s* (*the plan’/z/ problem, the cat’/s/ mat, the wash’/ɪz/ cycle*; Beispiele aus Ryan 2016: 104). Einzig in Kombination mit dem Pluralsuffix *-s* ist der Formaufbau nicht konkatenativ, hier erscheint als Form des Suffixes nur ⟨’⟩, vgl. *the cats’ mat, the *cats’s mat*. Phonologisch passiert etwas Vergleichbares: Der Genitiv des Plurals *horses* lautet nicht */hɔːsɪz/, sondern /hɔːsɪz/, wie die Pluralform selbst. Bauer/Lieber/Plag (2013: 145) sehen das als eine Instanz von Haplogievermeidung an (also die Vermeidung gleicher, adjazenter phonologischer Strukturen). Etwas Ähnliches kann für die Verteilung in der Schrift auch vermutet werden: Folgen zweier gleicher Suffixe (wenn wir den Apostroph vernachlässigen) sind im Englischen verboten.

4.3.2 Derivation

Das Inventar der Derivationsaffixe ist in beiden Sprachen wesentlich größer als das der Flexionssuffixe, und gleichzeitig gibt es – ebenfalls in beiden Sprachen – eine Vielzahl von Abweichungen von der Einheitlichkeit unter den lateinischen/romanischen Affixen. Wir beschränken uns im Folgenden darauf, eine Übersicht über die Verhältnisse zu geben, ohne notwendigerweise auf jeden Einzelfall einzugehen.

Den Anfang machen die Derivationsaffixe des Deutschen. Wir können hier grundsätzlich zwei Gruppen von nicht-einheitlichen Affixen unterscheiden: Zum einen gibt es Affixe, deren Form sich in weiteren Ableitungen oder unter Flexion verändert; diese Fälle sind relativ selten. Ein Beispiel ist das Movie-rungssuffix *-in*: Im Singular hat es die Form ⟨in⟩ (⟨Lehrerin⟩), im Plural die Form ⟨inn⟩ (⟨Lehrerinnen⟩). Zum anderen gibt es Affixe, deren Form abhängig ist vom Stamm, den sie modifizieren. So tritt bspw. die Variante *-keit* vor allem an Basen mit unbetonter Ultima (z. B. ⟨Fröhlichkeit⟩), die Variante *-heit* hingegen an Basen mit betonter Ultima (z. B. ⟨Schönheit⟩), vgl. Fleischer/Barz (2012:

209 ff.).¹⁰⁴ Die Verteilung kann kontextuell determiniert sein wie bei *-heit/-keit*; sie kann aber auch rein lexikalisch sein wie bei den Varianten *-abel* (⟨praktikabel⟩) und *-ibel* (⟨kompressibel⟩).

Zunächst zur Variation der ersten Gruppe. Die einschlägigen Fälle sind in (85) aufgeführt:

(85a) *-in/-inn-*; *-nis/-niss-*; *-ismus/-ismuss-*; *-abel/-abl-*

(85b) *-ös/-os-*; *-är/-ar-*; *-abel/-abil-*; *-al/-ell*

Die Fälle in (85a) variieren, wenn die betreffende Wortform flektiert wird (⟨Lehrerinnen⟩, ⟨Ergebnisse⟩, (des) ⟨Faschismusses⟩). Die Schreibung mit Doppelkonsonant wird nicht in die anderen Formen des Paradigmas übernommen, die Form der jeweiligen Derivationssuffixe ist nicht einheitlich. Für die beiden nativen Suffixe *-in* und *-nis* ist das relativ klar und auch schon beobachtet (z. B. Eisenberg 2011: 342 f.).

Warum kommt es bei *-in* und *-nis* überhaupt zur Uneinheitlichkeit – warum werden die Suffixe nicht einheitlich ⟨inn⟩ bzw. ⟨niss⟩ geschrieben? Doppelkonsonanten kommen in nativen Suffixen nicht vor (vgl. oben 3.3), nur in Fremdsuffixen wie *-ell*, *-ess*, *-esse*, *-ett*, *-ette*, die sich verhalten wie native Stämme (vgl. Eisenberg 2011: 342 f.). Bei diesen Suffixen tragen die korrespondierenden phonologischen Formen alle den Hauptakzent der jeweiligen Ableitung. Formen wie ⟨inn⟩ oder ⟨niss⟩ könnten also suggerieren, dass auch hier der Hauptakzent auf dem Suffix liegt.

Für das Fremdsuffix *-ismus* gibt es zwei Varianten nur dort, wo die Genitivform *-ismusses* verwendet wird bzw. die Pluralform *-ismusse*. Präskriptiv sind beide stigmatisiert, empfohlen ist Suffixlosigkeit im Genitiv ((des) ⟨Faschismus⟩) sowie die Variante *-ismen* im Plural ((die) ⟨Faschismen⟩) (vgl. z. B. Duden 2011: 487). Diese Varianten sind momentan eindeutig dominant, es gibt aber durchaus auch in den sog. ‚Qualitätsmedien‘ Beispiele für die seltenere Genitivform (in (86c) ist die Wortform nur ein Zitat – aber immerhin eins von Thomas Mann):

(86a) Gewiß, die Hilfe für Berlin mag eine Lokomotive sein, die mit schnaubendem Pathos und Schlagwortgerassel den schönen Zug studentischen Idealismus kräftig befördert (ZEIT 3. 9. 1965: 42)

104 Der Sachverhalt ist etwas vereinfacht dargestellt und vernachlässigt die Verteilung der Variante *-igkeit*; das Beispiel dient hier aber nur dem Zweck, die betreffende Gruppe von Affixen zu charakterisieren.

- (86b) Khol, der sich gerne als das intellektuelle Hirn des politischen Katholizismus in Österreich sieht, [...] (Süddeutsche Zeitung, 28. 1. 2003: 7)
- (86c) Thomas Mann hatte in seinem Gutachten auch noch Textproben eines gewissen Kurt Grohmann zu beurteilen. Hierzu nun heisst es: „Prosa wie Verse sind das Erzeugnis eines wohlgemeinten, aber absolut kindlichen Dilettantismus, über dessen literarische Unmöglichkeit kein Wort zu verlieren ist.“ (Neue Zürcher Zeitung, 17. 2. 2010: 55)

Nebenbei bemerkt: In (86a, c) markieren jeweils nur diese Formen eindeutig den Genitiv – ‚studentischen Idealismus‘ und ‚kindlichen Dilettantismus‘ können auch Akkusativformen sein. Die Verwendung in Nominalgruppen ohne herkömmlichen Kopf könnte eine Nische sein, in der diese Genitivformen sich einrichten – das wäre eine weitere Untersuchung wert. Unabhängig davon ist wichtig: Die *-ismusses*-Formen sind selten, aber wenn sie verwendet werden, dann immer mit Doppelkonsonant.

In allen drei Fällen, die bislang diskutiert wurden, variieren Konsonanten; sie sind entweder einfach oder verdoppelt (‹-nis› vs. ‹-nisse›). Es gibt auch einen Fall von ‹e›-Tilgung wie in Stämmen (Typ ‹segel[n]› – ‹segle›): Die Form des Fremdsuffixes *-abel* variiert, wenn das entsprechende Adjektiv attributiv verwendet wird, vgl. ‹Die Frage ist akzeptabel› – ‹eine akzeptable Frage›. Die einheitliche Form ‹akzeptabele› ist eine marginale Variante; *-abel* verhält sich wie das native Adjektiv *übel* (Eisenberg 2011: 240). Hier gilt, was oben für die fremden Adjektive vom Typ FIT festgestellt wurde: Innerhalb des Flexionsparadigmas der attributiven Adjektive ist die Form einheitlich; innerhalb eines Lexems wie AKZEPTABEL hingegen variiert die Form des Suffixes.

Die Fälle in (85b) variieren je nachdem, ob das Suffix wortfinal auftritt oder innerhalb weiterer Ableitungen. Bei den fremden Adjektivsuffixen *-ös* und *-är* tritt dann jeweils die Variante ohne Trema auf: ‹nervös› – ‹Nervosität›; ‹komplementär› – ‹Komplementarität›. Diese Fälle sind selten, es gibt (je nach morphologischer Analyse)¹⁰⁵ ein knappes Dutzend in CELEX:

- (87a) generös – Generosität, ingeniös – Ingeniosität, monströs – Monstrosität, religiös – Religiosität, seriös – Seriosität, varikös – Varikosität

¹⁰⁵ Unter den Belegen in (87b) finden sich z. B. einige unikale Basen: Kommen die Basen in *familiär* und *populär* noch in *Familie* und *populistisch* vor, gilt das für *regulär*, *utilitär* und *vulgär* nicht. Diese Adjektive könnten auch als Simplizia klassifiziert werden, deren Stämme uneinheitlich sind.

- (87b) familiär – Familiarität, populär – Popularität, regulär – Regularität,
singulär – Singularität, totalitär – Totalitarismus, utilitär – Utilitarismus,
vulgär – Vulgarität

Die Variante ohne Trema ist in einfachen Ableitungen ein Allomorph zu *-ös* und *-är* (vgl. Fleischer/Barz 2012: 349 f.): *⟨glorios⟩* vs. *⟨religiös⟩*; *⟨atomar⟩* vs. *⟨illusio-när⟩*; die Varianten mit Trema scheinen allerdings häufiger zu sein als die ohne. Prinzipiell sind von den *-ös/-os-* und den *-är/-ar-*Bildungen Ableitungen mit *-ität*, *-ismus* (und damit *-istisch*) möglich, und in allen Fällen tritt dann statt *-ös* die Variante *-os* auf. Das spiegelt die prosodischen Verhältnisse treffend wider: Die Suffixe *-ös* und *-är* tragen den Hauptakzent, wenn sie der Kopf des Wortes (i. S. v. Eisenberg 2013a: 215 f.) sind, vgl. z. B. *monströs*, *religiös*, *familiär*, *totalitär*. Der Hauptakzent ist gewissermaßen durch den Umlaut markiert, wie Schmidt (o. J.) feststellt. Werden die Wörter nun weiter abgeleitet, so ‚bewegt sich‘ der Hauptakzent auf *-ität*, *-ismus* bzw. *-istisch*, und *-ös* bzw. *-är* korrespondieren mit unbetonten Silben.

Auch *-al* und *-ell* alternieren, und zwar nicht nur in Ableitungen (*⟨kriminell⟩* – *⟨Kriminalität⟩*, *⟨eventuell⟩* – *⟨Eventualität⟩*), sondern auch in Komposita (*⟨Kriminalpolizei⟩*, *⟨Eventualfall⟩*; vgl. Fuhrhop 1998: 127 f.; Fleischer/Barz 2012: 349 f.). Es gibt keine Ableitungen und Komposita, die *-ell*-Bildungen enthalten; *-ell* ist für die weitere Wortbildung eine ‚Sackgasse‘, eine ‚schließende Form‘ nach Aronoff/Fuhrhop (2002).

Dasselbe gilt für *-abel/-abil-*. Für jedes Wort auf *-abel* scheint eine Ableitung mit *-ität* möglich, und genau hier alterniert der Vokal im Suffix: *⟨akzeptabel⟩* – *⟨Akzeptabilität⟩*, *⟨praktikabel⟩* – *⟨Praktikabilität⟩*, *⟨rentabel⟩* – *⟨Rentabilität⟩*. Es gibt keine Wörter, die auf *-abilität* enden; das Suffix ist also uneinheitlich.

Kommen wir zur zweiten Gruppe von alternierenden Affixen, denjenigen, die in Abhängigkeit vom Stamm alternieren. Die folgende Liste gruppiert die einschlägigen Fälle:

- (88a) *-ei/-erei*; *-lich/-erlich*; *-heit(/keit)/-igkeit*
 (88b) *-ion/-tion/-ation*; *-iv/-ativ*; *-or/-ator*
 (88c) *ir-/il-/im-/in-*; *kor-/kol-/kon-/co-*
 (88d) *-abel/-ibel*; *-al/-ell*; *-ant/-ent*; *-anz/-enz*; *-ante/-ente*; *-heit/-keit*

In (88a) sind Gruppen nativer Suffixe aufgeführt, von denen jeweils eines aus der Verschmelzung des anderen mit einem dritten Suffix hervorgegangen ist (vgl. Fuhrhop 1998: 44 ff.). Die Gruppen sind Allomorphe eines abstrakten Suffixes; zum Teil ist das ‚komplexere‘ Allomorph produktiv (*-erei*), zum Teil nicht (*-erlich*). Diese Fälle sind aus unserer Perspektive nicht besonders interessant, weil gra-

phematisch (und phonologisch) das eine Suffix eine Teilkette des anderen ist. Holistisch (vgl. Plank 1981) oder Output-orientiert (vgl. z. B. Plag 1999) betrachtet, gilt als formale Beschränkung für die Bildungen, dass sie auf *-ei* enden. Formen wie *Bäcker-ei*, *Pfarr-ei* und *Molk-erei* erfüllen diese Beschränkung; dass sie strukturell nicht isomorph sind, spielt hier keine Rolle.

Ganz ähnlich kann auch für die fremden Suffixe in (88b) argumentiert werden: Der formale ‚Kern‘ ist jeweils konstant, ⟨ion⟩, ⟨iv⟩ und ⟨or⟩, und zwar konstant am rechten Rand der Ableitung.¹⁰⁶ Die Suffixe sind also nicht im eigentlichen Sinn uneinheitlich.

In (88c) sind Präfixe mit phonologisch motivierter Allomorphie aufgelistet; der finale Konsonant des Suffixes ist vom intialen Element des Stamms abhängig. Solche Präfixe gibt es nur im fremden Bereich. Neben den genannten existieren noch einige weitere (*a-/an-/ar-*; *dis-/dif-/di-*). Der Status als Präfix ist synchron nicht immer eindeutig (vgl. z. B. Eisenberg 2011: 299 f.). Das muss bei der Bewertung der Nicht-Einheitlichkeit berücksichtigt werden: Wenn synchron nicht immer ein Präfix erkennbar ist, fällt die potenzielle Uneinheitlichkeit weniger ins Gewicht.

Bei (88d) handelt es sich um nicht-einheitliche Schreibungen im engeren Sinne. Wenn man berücksichtigt, dass die Paare *-ant/-ent*, *-anz/-enz* und *-ante/-ente* häufig paradigmatisch zusammenhängen (z. B. *Dominanz* – *dominant* – *Dominante*), ist das Ausmaß der Variation hier nicht besonders groß. Mit Ausnahme von *-heit/-keit* stammen alle Belege aus dem nicht-nativen Bereich und variieren hinsichtlich der Vokalgrapheme (meist ⟨a⟩ vs. ⟨e⟩). Die Verteilung von *-heit* und *-keit* ist morphologisch und prosodisch determiniert (vgl. z. B. Fleischer/Barz 2012: 209 ff.). Die Verteilung der nicht-nativen Varianten ist lexikalisch, nicht aber graphematisch oder phonologisch gesteuert; es lassen sich jeweils nur sehr grobe Domänen zuweisen. Das soll am Beispiel von *-abel/-ibel* gezeigt werden. Im Deutschen ist *-abel* wesentlich häufiger als *-ibel*; in CELEX gibt es 16 Einträge für *-ibel*, nämlich *disponibel*, *fallibel*, *flexibel*, *horribel*, *intelligibel*, *kompatibel*, *kompressibel*, *konvertibel*, *penibel*, *perzeptibel*, *plausibel*, *reversibel*, *sensibel*, *suggestibel*, *suszeptibel* und *terribel*. Während der Großteil der *-abel*-Bildungen auf *-ier[en]*-Verben beziehbar ist (vgl. Eisenberg 2011: 277), gilt das für *-ibel* nicht; nur fünf der 16 Wörter können als Ableitungen von Verben interpretiert werden (*disponier[en]*, *flektier[en]*, *kompressier[en]*, *konvertier[en]*, *perzeptier[en]*). Zum Teil sind die Basen auch unikal wie in *fallibel*, *kompatibel*, *penibel*, *plausibel*, *sensibel*, *terribel*. Die Variante *-abel* ist also regelmäßig auf *-ier[en]*-Verben (oder auf

¹⁰⁶ Außerdem alternieren die Varianten *-ion*, *-iv* und *-or* häufig, vgl. z. B. *Illustration*, *illustrativ*, *Illustrator*. Das spricht eher für *-at-* als stammbildendes Element, vgl. aber Fuhrhop (1998: 122 f.).

freie substantivische Basen, vgl. Fleischer/Barz 2012: 349) bezogen, während *-ibel*-Bildungen schwerer dekomponierbar sind.

Graphematisch gibt es also durchaus Variation unter den Derivationsaffixen – wie viel genau, das hängt z. T. von der vertretenen morphologischen Analyse ab. Gibt es phonologisch mehr Variation, wird also ein Teil von der Schrift überbückt?

Das ist in der Tat so, und zwar in mindestens drei Fällen. Zum einen gibt es – wie in der Flexion (siehe 4.3.1) – phonologische Alternationen, die ihren Ursprung in der Auslautverhärtung haben. Diese phonologischen Alternationen werden nicht verschriftet, vgl. z. B. *-ig/-ige* (⟨fettig⟩ – ⟨fettige⟩), *-and/-anden* (⟨Doktorand⟩ – ⟨Doktoranden⟩), *-iv/-ive* (⟨massiv⟩ – ⟨massive⟩).

Zum zweiten alterniert die Form der Suffixe *-er/-or*, *-är*, *-ier[en]* phonologisch, je nachdem, ob ein vokal-initiales Suffix folgt: *Bäcker* /beke/ – *Bäckerin* /bekərɪn/; *primär* /primee/ – *primäre* /primerə/; *demonstriert* /demonstriet/ – *demonstriere* /demonstrirə/. In der Schrift sind die Formen konstant.

Und drittens gibt es beim nativen Präfix *un-* z. T. Assimilationsprozesse, die in der Schrift nicht gespiegelt werden: Der Nasal des Präfixes variiert je nach Anlaut des Stamms, z. B. *undicht* /ʊndɪçt/, *unbedingt* /ʊmbədɪŋt/, *ungern* /ʊŋgeen/. Diese Assimilationsprozesse sind, anders als die auch in der Schrift sichtbaren Assimilationen bei den Fremdpräfixen wie *in-/im-/ir-* etc., nicht obligatorisch. Formen wie /ʊnbədɪŋt/ oder /ʊngeen/ sind ebenfalls unauffällig.

Wie einheitlich sind die Derivationsaffixe im Englischen? Wir können wie im Deutschen zwei Gruppen von Alternationen unterscheiden: Zum einen solche, die abhängig von weiterer Flexion oder Derivation entstehen, zum anderen Alternationen, die abhängig vom konkreten Stamm sind, auf dem das Affix operiert. Beginnen wir mit der ersten Gruppe. Die Liste in (89) enthält einschlägige Beispiele:

(89a) *-y/-i-*; *-ery/-erie-*; *-ate/-at-*

(89b) *-able/-abil-*; *-our/-or-*; *-ous/-os-*

In (89a) sind Fälle aufgeführt, in denen Derivationsuffixe unter Flexion variieren. Wir finden hier zwei der drei regulären Prozesse, die ⟨y⟩/⟨i⟩-Alternation (z. B. ⟨windy⟩, ⟨windiest⟩; ⟨bakery⟩ – ⟨bakeries⟩) und die ⟨e⟩-Tilgung (z. B. ⟨demonstrate⟩ – ⟨demonstration⟩). Konsonantenverdoppelung hingegen tritt bei Flexionsformen von affigierten Wörtern nicht auf. Das deckt sich mit der Beschreibung der Konsonantenverdoppelung oben: Nur wenn stamffinal eine einsilbige Wurzel steht (und wenn weitere graphemische Kriterien erfüllt sind), wird der stamffinale Konsonant verdoppelt.

In (89b) sind Fälle enthalten, die bei weiterer Derivation variieren. Das betrifft die Suffixe *-able* und *-ible*, die jeweils als <abil-> und <ibil-> erscheinen, wenn sie mit *-ity* (und – was seltener vorkommt – mit *-ist*) abgeleitet werden (z. B. <readable> – <readability>; <sensible> – <sensibility>, vgl. Ryan 2016: 116). Es alternieren hier <le> in der Basis mit <il> unter Derivation. Diese Alternation ist sehr häufig, denn *-ity* ist besonders auf *-able*-Ableitungen produktiv (vgl. Bauer/Lieber/Plag 2013: 246). Ebenfalls in diese Rubrik gehören *-our/-or-* und *-ous/-os-* (z. B. <glamour> – <glamorous>, <curious> – <curiosity>, vgl. z. B. Venezky 1970: 62). Die Alternation *-our/-or-* gibt es nur im britischen Englisch; im amerikanischen Englisch ist die Form des Suffixes immer <or> (vgl. Carney 1994: 428). Die Alternation *-ous/-os-* gibt es in beiden Varietäten, und sie hat eine Entsprechung im Deutschen (<generös> – <Generosität>, siehe oben).

Die zweite Gruppe uneinheitlicher Affixe umfasst solche, die in Abhängigkeit vom Stamm variieren. Wir können sie wie im Deutschen in vier Gruppen zusammenfassen:

(90a) *-ry/-ery*

(90b) *-ion/-ation; -ic/-atic; -ory/-atory*

(90c) *ir-/il-/im-/in-; cor-/col-/con-/com-*

(90d) *-able/-ible; -ce/-cy; -ant/-ent; -ise/-ize; -er/-or/-ar*

Das Suffix in (90a) leitet Substantive aus allen Wortarten ab (*devilry, cookery*). Die Verteilung der Varianten ist größtenteils prosodisch: An zweisilbige Basen tritt *-ry* (*chickenry*), an einsilbige Basen tritt *-ery* (*hennery*), vgl. Bauer/Lieber/Plag (2013: 178 f.). Zum Teil sind die Ableitungen auf *-ery* auch auf Verben auf *-er* zu beziehen (*embroider* – *embroidery*) – segmental hat das Suffix hier also die Form <y> (vgl. Bauer/Huddleston 2002: 1705). Wie im Deutschen ist diese Verteilung am besten über eine Wohlgeformtheitsbeschränkung zu erfassen, die fordert, dass die entsprechenden Bildungen auf *-ery* enden.

Die Belege in (90b) enthalten Varianten, die sich in der An-/Abwesenheit von *-at-* entscheiden. Die Interpretation dieser Form ist notorisch schwierig und umstritten: Ist sie Teil des Suffixes, ist sie Teil des Stamms (vgl. z. B. Bauer/Huddleston 2002: 1674 f.)? Die Lage wird nicht eben übersichtlicher dadurch, dass *-ate* im Englischen ein Verbsuffix ist (z. B. *demonstrate*). Ist *demonstration* also eine Ableitung von *demonstrate*? Oder ist das Suffix hier *-ation*, und ersetzt dieses Suffix *-ate*? Dann könnte man auch Fälle wie *flirtation* erfassen, in denen es kein entsprechendes *-ate*-Verb gibt und das Suffix *-ation* auf der Basis *flirt* operiert. Wir lassen die Frage hier offen und halten nur wie bei der Besprechung der deutschen Daten fest: Diese Suffixe sind nicht uneinheitlich im eigentlichen

Sinne; sie enthalten einen gemeinsamen ‚Kern‘ am rechten Rand und sind lediglich ‚erweitert‘.

In (90c) sind Präfixe aufgeführt, deren Variation phonologisch determiniert ist (Bauer/Lieber/Plag 2013: 180). Die Bedeutung ist unterschiedlich transparent: *in-* (und seine Varianten) sind bspw. wohl einfacher zu identifizieren als *con-* (vgl. z. B. *inaudible, impossible, illegal* vs. *conurbation, commixture, collateral*; Bauer/Huddleston 2002: 1672). Wenn aber der Status als Präfix synchron nicht sicher ist (wie etwa bei *con-* in *collateral*), dann muss die nicht-einheitliche Schreibung hier mit Vorsicht bewertet werden.

In (90d) schließlich sind nicht-einheitliche Affixe im engeren Sinn aufgeführt. Es handelt sich ausschließlich um Suffixe aus dem lateinisch-romanischen Stratum, die sich (mit Ausnahme von *-ise/-ize*) in den Vokalgraphemen unterscheiden. Bei *-able/-ible* und *-ance/-ence/-ancy/-ency* hängt die Zuschreibung der Nicht-Einheitlichkeit davon ab, ob die Formen überhaupt als Allomorphe abstrakter Morpheme analysiert werden oder als eigenständige Suffixe. Dafür spricht bei *-able/-ible* bspw., dass sich die Formen unterschiedlich verhalten. Die Form *-ible* ist relativ eindeutig ein ‚Class I‘-Affix i. S. v. Siegel 1979 (vgl. für das Folgende Bauer/Huddleston 2002: 1706 f.): Es tritt mit gebundenen Basen auf (z. B. *audible*), es löst z. T. morpho-phonologische Stammalternationen aus (z. B. *divide – divisible*), und es gibt einige Fälle mit Betonungsverschiebung (*ne’glect – ’negligible*); außerdem werden Bildungen mit *-ible* vom ‚Class I‘-Präfix *in-* modifiziert (z. B. *incredible*). Die Form *-able* verhält sich demgegenüber zum Teil wie ein ‚Class II‘-Affix: Morpho-phonologische Alternationen sind selten, bei Betonungsverschiebung ist die prosodische Struktur der Basis immer auch eine Variante (z. B. *’comparable* vs. *com’parable*), und die Negation wird oft mit dem nativen Präfix *un-* ausgedrückt (*unreliable, unremarkable*). Unter Umständen wird hier also die Zugehörigkeit von *-ible* zu einer bestimmten Gruppe von Affixen graphematisch markiert.¹⁰⁷

Das Substantivsuffix *-ce/-cy* operiert vor allem auf *-ant/-ent*-Basen (z. B. *brilliance, presidency*) – das ist seine ‚Nische‘ in einem Wettbewerbsmodell (Aronoff einger.). Auf diese Weise kommt es bei funktionaler Invariation zu vierfacher Formvariation (*-ance, -ancy, -ence, -ency*), die lexikalisch determiniert ist. Das Suffix ist höchstens minimal produktiv (Bauer/Lieber/Plag 2013: 196).

¹⁰⁷ Ryan (2016: 114 ff.) interpretiert die Beschreibung des Suffixes im OED so, dass sich die Schreibung *able* im Laufe der Zeit gegenüber *ible* durchgesetzt hat und im Sinne von Berg/Aronoff (2017) als Reduktion von Variation interpretiert werden kann – das folgt aber nicht zwangsläufig aus der Darstellung des OED; eine genauere Untersuchung wäre hier wünschenswert.

Die Variation zwischen *-ise* und *-ize* ist z. T. diatopisch motiviert: Im amerikanischen Englisch ist *-ize* dominant, im britischen Englisch gibt es wesentlich mehr Variation (Bauer/Huddleston 2002: 1715), die meines Wissens noch nicht systematisch untersucht ist; das Suffix ist das produktivste Verbalsuffix im heutigen Englisch (Bauer/Huddleston 2002: 1715).

Die Formen *-er/-or/-ar* werden u. a. von Bauer/Huddleston (2002: 1698) zu einem abstrakten Substantivsuffix zusammengefasst als „person/instrument nominalisation“. Produktiv ist allein die Form *-er* (z. B. *reader, golfer*); die Form *-ar* ist sehr selten (z. B. *beggar, liar*); *-or* ist häufiger, aber größtenteils auf (freie oder gebundene) lateinische Basen beschränkt (z. B. *instructor, author*). Es ist nicht ganz klar, wie weit die semantische/funktionale Übereinstimmung von *-or*- und *-er*-Bildungen reicht: Eisenberg (2011: 222) stellt für das Deutsche fest, dass „die Grundfunktion des lateinischen *or* als Suffix für Agenssubstantive [...] noch erkennbar, keineswegs aber mehr durchgängig realisiert [ist]“.

Aus der phonographischen Perspektive ergibt sich: Es gibt im Englischen eine Reihe von phonologischen Alternationen, die graphematisch überbrückt werden. Das bekannteste Beispiel ist wahrscheinlich die konstante Schreibung des Suffixes *-ic*: Während die Form in Ableitungen zwischen [ɪk] (*electrical*), /ɪs/ (*electricity*) und /ɪʃ/ (*electrician*) variiert, ist die graphematische Form einheitlich (vgl. z. B. Carney 1994: 18). Ähnliches gilt für das Suffix *-ify*, das die Variante /ɪfɪ/ (*signify*) und /ɪfi/ (*signification*) hat; auch diese Variation wird graphematisch nicht gespiegelt (vgl. Bauer/Lieber/Plag 2013: 179). Und je nachdem, ob man den Wechsel zwischen *demonstrate* und *demonstration* als Ableitung mit *-ion* wertet (und nicht als Ersetzung von *-ate* durch *-ation*), alterniert /eɪt/ mit /eɪʃ/.

In Varietäten des Englischen, in denen /ɪ/ in der Koda nicht vorkommt (nicht-rhotische Varietäten wie z. B. die Received Pronunciation), variieren auch die Suffixe *-er/-ar/-or*, *-eer*, *-ster* und *-ure* phonologisch: Werden Ableitungen mit diesen Suffixen wiederum mit einem vokal-initialen Suffix abgeleitet oder flektiert, tritt /ɪ/ auf, und es kommt z. B. zum Wechsel /ɪə/ (*engineer*) vs. /ɪ/ (*engineering*).

Und schließlich gibt es auch beim nativen Präfix *un-* phonologische Variation, die vom stamminitialen Phonem determiniert ist – ganz ähnlich wie bei den fremden Präfixen, die oben genannt wurden (/ʌm/ in *unbelievable*, /ʌn/ in *unfavourable*, /ʌŋ/ in *unkind*). Es ist allerdings unklar, wie obligatorisch diese Assimilation ist und inwiefern sie von anderen Faktoren wie der Sprechgeschwindigkeit oder dem Register abhängt (Bauer/Lieber/Plag 2013: 180).

4.4 Affixe: Eindeutigkeit

In diesem Abschnitt geht es um die Eindeutigkeit von Affixen – also die Frage, wie zuverlässig die schriftliche Form das jeweilige Affix repräsentiert. Wie oben beschrieben, können wir zwei Ausprägungen der Eindeutigkeit unterscheiden, die unabhängig voneinander sind: Zum einen kann eine Schreibung eindeutig auf *ein* Affix verweisen; das wurde oben als Eindeutigkeit im engeren Sinne bezeichnet. Eine Schreibung kann aber auch ein oder mehrere Affixe repräsentieren und sich dabei von allen nicht-morphologischen Endungen unterscheiden; wir haben das oben als Distinktivität bezeichnet. Interessanter im Rahmen dieser Arbeit ist diese zweite Dimension von Einheitlichkeit, weil sie unmittelbarer graphematisch ist: Eine Schreibung kann für die Repräsentation eines oder mehrerer Affixe ‚reserviert‘ sein; es gibt in diesem Fall keine homographischen Wortenden, die fälschlicherweise als Instanz dieses Affixes interpretiert werden können. Wie im letzten Abschnitt werden die Affixe dabei in die Bereiche Flexion (4.4.1) und Derivation (4.4.2) unterteilt und nacheinander untersucht.

4.4.1 Flexion

Die regelmäßige Flexion im Deutschen und Englischen wurde in Berg et al. (2014) und Berg (2017) auf Eindeutigkeit überprüft, die unregelmäßige Flexion in Fuhrhop (2017); vgl. für eine ähnliche Idee Augst (1986). Die Ergebnisse zur regelmäßigen Flexion im Deutschen sind in Tabelle 111 zusammengefasst (aus Berg 2017: 35). In einem CELEX-Teilkorpus¹⁰⁸ wurde für jede Form überprüft, wie viele Ausnahmen von der Eindeutigkeit es gibt – also homographische Wortenden

108 Das verwendete CELEX-Teilkorpus in Berg (2017) ist etwas umfangreicher als das hier verwendete Teilkorpus einfacher Stämme; in der vorliegenden Arbeit wurde etwas strenger darauf geachtet, dass die als morphologisch einfach annotierten Einträge auch tatsächlich morphologisch einfach sind (vgl. 2.2). In Berg (2017) umfasst das Korpus im Deutschen neben den als morphologisch einfach gekennzeichneten außerdem Einträge, deren Struktur als ‚irrelevant‘ annotiert ist (Statuscode ‚I‘). Insgesamt bilden damit 10.513 Einträge die Grundlage für die Untersuchung (zum Vergleich: in der vorliegenden Arbeit sind es 5.484 Einträge). Im Englischen wurden in Berg (2017) neben den morphologisch einfachen auch die Einträge untersucht, deren morphologische Struktur unklar ist oder die eine potenzielle Wurzel enthalten (Statuscodes ‚O‘ und ‚R‘). Das führt zu einem Korpus von 12.839 Einträgen (im Gegensatz zu den 7.004 Einträgen im Rahmen der vorliegenden Arbeit). Außerdem war der phonetische Status von [a] und [e] in Berg (2017) unklar, weswegen für die *-er*-Suffixe potenzielle Ausnahmen angesetzt wurden. Mittlerweile konnte gezeigt werden, dass die beiden Vokale nicht zusammenfallen (vgl. Tomaschek/Berg Ms.; siehe unten).

(,Anzahl Ausnahmen‘). Zum verbalen Flexionssuffix *-t* beispielsweise (5. Zeile in Tabelle 111) gibt es 68 homographische Wortenden (z. B. *dicht*), die als Instanzen eines Stammes + *-t* interpretiert werden könnten (in diesem Fall als Instanz des nicht-existierenden Stammes **dich[en]*). Darüber hinaus wurde untersucht, ob phonographisch noch weitere Ausnahmen möglich sind. Gibt es Wortenden, die phonologisch auf /t/ enden, die aber anders als mit ⟨t⟩ verschriftet werden? Wäre das der Fall, könnte man sagen, dass das Schriftsystem hier einen Bereich für das Suffix reserviert. Bei *-t* ist das in der Tat in 19 Fällen so – allerdings können die betreffenden Schreibungen (wie bspw. *fremd*) alle als Instanzen von Stammkonstanz motiviert werden; sie sind also keine eigentlichen potenziellen Ausnahmen. Und schließlich ist jeweils vermerkt, ob anderswo im System phonographische Alternativen für die Verschriftung homographischer Wortenden zur Verfügung stünden – bei *-t* kann man z. B. ⟨dt⟩ in Betracht ziehen (**dichtd*).

Bei den Suffixen wurde jeweils nach solchen der substantivischen Flexion und anderen unterschieden (±SBST in Tab. 111, vgl. Berg 2017: 25): Bei substantivischen Suffixen muss die initiale Großschreibung zum Suffix passen. Entsprechend ist bspw. für das Pluralsuffix *-e* (wie in *Tische*) ein Nicht-Substantiv wie *müde* keine Ausnahme von der Einheitlichkeit; die Kleinschreibung signalisiert bereits, dass es sich nicht um ein Substantiv handelt.

Tab. 111: Eindeutigkeit deutscher Flexionssuffixe nach Berg (2017). Für jedes der bearbeiteten Suffixe des Deutschen: die Anzahl der Ausnahmen (monomorphematische Stämme mit homographischem Wortende aus CELEX); die Anzahl potenzieller Ausnahmen (Wörter, die phonographisch so verschriftet werden könnten); phonographische Alternativen. *alles Instanzen von Morphemkonstanz; **alles Fremdwörter mit finalem ⟨x⟩.

Suffix	Anzahl Ausnahmen	Anzahl potenzieller Ausnahmen	phonographische Alternativen
-en (+SBST)	148	–	–
-en (-SBST)	34	–	–
-n (+SBST)	5	–	–
-n (-SBST)	7	–	–
-t	68	(19)*	⟨dt⟩
-st	8	–	⟨ßt⟩
-em	–	–	–
-end	6	(22)*	–
-s	57	(25)**	⟨ß⟩
-es (+SBST)	11	55 (+11)*	⟨eß⟩, ⟨is⟩
-es (-SBST)	–	–	–
-e (+SBST)	1.462	–	–
-e (-SBST)	52	–	–
-er (+SBST)	165	–	⟨a⟩

Suffix	Anzahl Ausnahmen	Anzahl potenzieller Ausnahmen	phonographische Alternativen
-er (-SBST)	46	–	⟨a⟩
ge-X-t	3	(2)*	⟨ge-X-dt⟩
ge-X-en	21	–	–

Besonders bei den Suffixen *-e*, *-er*, *-en*, *-t* und *-s* gibt es zahlreiche Ausnahmen von der eindeutigen Verschriftung. Außerdem gibt es hier kaum homophone Wortenden, die graphematisch differenziert werden. Die wenigen Fälle, die es gibt, sind alle anders erklärbar; hier wird nicht eine Schreibung für die Repräsentation von Suffixen reserviert. In den meisten Fällen handelt es sich um Instanzen von Stammkonstanz (z. B. die Tatsache, dass wir ⟨fremd⟩ und nicht *⟨fremt⟩ schreiben), und bei *-s* sind die Ausnahmen allesamt Fremdwörter auf ⟨x⟩, die nicht mit finalelem ⟨chs⟩ oder ⟨cks⟩ geschrieben werden können, weil diese Schreibungen nur im nativen Bereich vorkommen (z. B. ⟨Jux⟩ – *⟨Jucks⟩, *⟨Juchs⟩).

Es gibt allerdings in zwei Fällen potenzielle Ausnahmen, deren Schreibungen so nicht erklärt werden können. Zum einen gibt es beim substantivischen Genitivsuffix *-es* nur wenige Ausnahmen von der Einheitlichkeit (z. B. *Kirmes*), aber fünfmal so viele potenzielle Ausnahmen (z. B. *Dosis*, *Kürbis*). Hier ist noch unklar, ob die beiden Klassen phonetisch zusammenfallen: Sind die Vokale in den zweiten Silben von *Buches* (Genitiv *-es*), *Kirmes* (Wortende *-es*) und *Kürbis* (potenzielles *-es*) tatsächlich homophon? Wenn ja, dann liegt hier ein Fall von Eindeutigkeit der Suffixschreibung vor.

Der zweite Fall ist sehr ähnlich gelagert: Es geht hier um die Frage, ob die Vokale in den zweiten Silben von *Kinder*, *Laster* und *Pasta* homophon sind, wie Fuhrhop/Peters (2013: 244 f.) vermuten. Wenn das so wäre, dann würde das Deutsche in diesen Fällen in der Tat ausweichen. Phonetisch ist das allerdings in den meisten Fällen nicht so. Finales [a] und [ɐ] werden phonetisch unterschiedlich realisiert, und zwar (etwas vereinfacht) in Abhängigkeit von der Wortfrequenz (vgl. Tomaschek/Berg Ms.). Je bekannter ein Wort den Sprechern ist, desto deutlicher werden die finalen Vokale unterschieden. Der Unterschied in der Schreibung in Wörtern wie *Laster* und *Pasta* kann also auch phonographisch motiviert werden. Es handelt sich hier natürlich trotzdem um eine morphologische Schreibung – aber eben um eine implizite.

Auffallend ist, dass die beiden Ausprägungen von Eindeutigkeit im Deutschen Hand in Hand gehen: Einerseits werden homophone Wortenden nur selten graphematisch von Flexionssuffixen unterschieden, die Formen der Suffixe sind also nicht distinkt; andererseits sind viele Flexionssuffixe homonym. Formen wie *-er*, *-en/-n*, *-s*, *-t*, *-st* und *-e* repräsentieren mehr als ein Suffix, die Formen sind

polyfunktional. Eindeutig im engeren Sinne sind nur das Dativsuffix *-em*¹⁰⁹ und die Suffixe der Partizipien, *ge-X-t*, *ge-X-en* und *-end*.

Im Englischen sieht die Situation ganz anders aus, wie die Übersicht in Tabelle 112 zeigt (nach Berg 2017: 24):

Tab. 112: Eindeutigkeit englischer Flexionssuffixe nach Berg (2017). Für jedes der bearbeiteten Suffixe: die Anzahl der Ausnahmen (monomorphematische Stämme mit homographischem Wortende aus CELEX); die Anzahl potenzieller Ausnahmen (Wörter, die phonographisch so verschriftet werden könnten); phonographische Alternativen.

Suffix	Anzahl Ausnahmen	Anzahl potenzieller Ausnahmen	phonographische Alternativen
-s	59	223	⟨se⟩, ⟨ce⟩, ⟨ze⟩
-'s	0	282	⟨s⟩, ⟨se⟩, ⟨ce⟩, ⟨ze⟩
-ed	30	112	⟨id⟩, ⟨ard⟩, ⟨ead⟩
-er	505	280	⟨or⟩, ⟨ar⟩, ⟨our⟩, ⟨ure⟩
-est	24	20	⟨ist⟩
-ing	18	–	–

Zum einen ist das Inventar wesentlich geringer, und zum anderen gibt es nur eine substanzielle Ausnahme von der Eindeutigkeit, *-er*. In allen anderen Fällen ist die Zahl der Ausnahmen entweder gering, oder es gibt wesentlich mehr potenzielle Ausnahmen als tatsächliche Ausnahmen. Das ist z. B. bei *-ed* so: Hier gibt es 30 Ausnahmen, also morphologisch einfache Wörter, die aussehen wie Verbindungen aus einem Stamm + *-ed* (z. B. ⟨naked⟩, ⟨hundred⟩). Diesen 30 Ausnahmen stehen aber 112 potenzielle Ausnahmen gegenüber, also Fälle, die auch mit ⟨ed⟩ verschriftet werden könnten, die aber anders geschrieben werden (z. B. ⟨salad⟩, ⟨solid⟩). Die Ausnahme von der Eindeutigkeit, *-er*, ist außerdem polyfunktional: Es tritt auf als Komparativsuffix (z. B. *bigger*) und als Agensnominalisierer (z. B. *hitter*). Polyfunktional ist sonst nur *-s*: Es repräsentiert sowohl die 3. Ps. Sg. bei Verben (z. B. *(she) waits*) und den Plural bei Substantiven (z. B. *dogs*). Die Verteilung dieser beiden (morphosyntaktisch ja sehr verschiedenen) Suffixe ist komplementär, formal aber identisch (sie haben phonologisch dieselben Allomorphe), sodass bspw. Chomsky (1957: 29) ein abstraktes Suffix *-S* ansetzt („the morpheme which is singular for verbs and plural for nouns“, vgl. auch Aronoff 2018).

¹⁰⁹ Vgl. Bittner (2002), die zu einer ähnlichen Einschätzung aus einer ganz anderen Perspektive kommt.

Unklar ist, inwieweit die Suffixe im Gesprochenen eindeutig sind. Plag, Homann und Kunter (2017) zeigen, dass für verschiedene Instanzen von wort-finalem /s/ und /z/ sich a) die Realisierungen von morphemischem und nicht-morphemischem -s phonetisch unterscheiden und b) sich auch einige morphemische -s voneinander unterscheiden. Einige Formen repräsentieren also einige Suffixe distinkt und teilweise einheitlich; morphologische Informationen werden im phonetischen Signal kodiert (Plag/Homann/Kunter 2017: 209). Die Unterschiede sind vor allem Unterschiede in der Dauer, mit der die Elemente realisiert werden. Diese Ergebnisse werfen sehr grundsätzliche Fragen zur phonetisch-phonologischen Klassifizierung auf – wann sind zwei Segmente Instanzen eines Typs? Interessant und bislang ungeklärt ist auch, ob die Ergebnisse auf andere Affixe übertragbar sind und ob sie perzeptiv relevant sind. In jedem Fall können wir festhalten, dass bestimmte morphologische Informationen auch in der gesprochenen Sprache kodiert werden, und zwar unter Umständen in stärkerem Maße, als die traditionelle Phonologie es uns glauben macht.

Insgesamt sind die Flexionssuffixe im Englischen distinktiver und eindeutiger, und zwar in beiden Ausprägungen: Es gibt weniger homographische Wortenden (z. T. absolut, z. T. im Verhältnis zur Anzahl möglicher homographischer Wortenden), und die Suffixe sind funktional eindeutiger als im Deutschen.

4.4.2 Derivation

Was für die Flexionssuffixe funktioniert, ist auch für die Derivationsaffixe möglich. Die einschlägigen Fragen lauten: Gibt es homographische Wortenden? Repräsentieren die Formen genau ein Affix oder mehrere? In Berg/Aronoff (2017) wurde die erste dieser Fragen für vier englische Derivationsuffixe untersucht, *-ous*, *-ic*, *-al* und *-y*. Die Ergebnisse für *-ous* sind besonders deutlich: Es gibt praktisch keine Ausnahme von der Eindeutigkeit der Graphemfolge. *⟨ous⟩* repräsentiert fast immer das Adjektivsuffix *-ous*. Es gibt aber einige hundert Wörter, die potenziell mit finalem *⟨ous⟩* geschrieben werden könnten – *service*, *tennis*, *bonus* usw. Finales *⟨ous⟩* ist also für die Schreibung des Suffixes reserviert (zu den Ergebnissen im Detail siehe unten).

Es ist kein Zufall, dass genau diese Suffixe gewählt wurden: Es handelt sich um kurze Suffixe, die (mit Ausnahme von *-y*) phonologisch aus einem Vokal und einem Konsonanten bestehen. Die Wahrscheinlichkeit, dass es homographische und homophone Wortenden gibt, ist hoch. Die Wahrscheinlichkeit sinkt, je komplexer die Affixe werden: Es ist unmittelbar einsichtig, dass es potenziell mehr Wörter im Englischen gibt, die mit *⟨ous⟩* verschriftet werden könnten, also solche,

die homographisch mit *-ify*, *-ship* oder *inter-* sind. Je komplexer ein Affix ist, desto eindeutiger ist es aus rein stochastischen Gründen.

Daher werden im Folgenden nicht alle Derivationsaffixe auf ihre Eindeutigkeit hin untersucht. Für Affixe, die eine gewisse (phonologische oder graphematische) Mindestlänge überschreiten, kann ohnehin vermutet werden, dass sie eindeutig sind. Nur bei kürzeren Affixen ist das nicht zwangsläufig schon stochastisch der Fall. Von diesen kürzeren Affixen wird ein Teil untersucht – ein Querschnitt durch die Wortarten gewissermaßen. Für das Deutsche werden folgende Affixe überprüft:

- (91a) substantivische Affixe: *-in*, *-er*, *-ant*
- (91b) adjektivische Affixe: *-ar/-är*, *-ig*, *-al/-ell*, *-iv*, *-os/-ös*, *un-*
- (91c) verbale Affixe: *be-*, *er-*

Teilweise sind die Affixe produktiv (wie *-er*, *un-* und *be-*), teilweise sind sie unproduktiv (wie *-ant* oder *-iv*); teilweise sind es Präfixe (*un-*, *be-*), teilweise sind es Suffixe (*-in*, *-er*); teilweise sind es native Affixe (*be-*, *-er*), teilweise sind sie fremd (*-al*, *-ant*). Gemeinsam haben sie, dass die Menge der abgeleiteten Wörter sich jeweils recht gut fassen lässt (siehe unten).

Die Untersuchung läuft ab wie in Berg/Aronoff (2017) skizziert: Für jedes Suffix wird im CELEX-Teilkorpus einfacher Stämme (vgl. 2.2) bestimmt, wie viele Einträge ein homographisches Ende aufweisen. Ein potenzielles Problem ist dabei, dass bspw. *billig* in CELEX (nachvollziehbarerweise) als morphologisch einfach klassifiziert wird. Damit wäre *billig* aber eine Ausnahme von der Eindeutigkeit von *-ig*: ein morphologisch einfaches Wort, das so endet wie das Suffix. Eine solche Klassifikation wäre gegenintuitiv, denn unter den kompositionell aufgebauten *-ig*-Bildungen wie *kernig*, *saftig* etc. ist *billig* formal völlig unauffällig; der einzige Unterschied ist, dass es zu *billig* keine freie Basis **Bill* gibt. Als Ausnahme von der Eindeutigkeit wird *billig* daher nicht gezählt. Etwas technischer: Es werden nur solche Einträge gezählt, die nicht dem Output der jeweiligen Wortbildungsregel entsprechen (dem ‚Wertebereich‘ in Fuhrhop 1998: 14).

Eine Einschränkung betrifft dabei die Form der ‚Stämme‘, also der Form, die von einem morphologisch einfachen Eintrag übrigbleibt, wenn die graphematische Form des Affixes ‚abgezogen‘ wird, z. B. ⟨Eim-⟩ in *Eimer*. Es wird hier die Minimalitätsbedingung aus 3.2.3 angesetzt, nach der potenzielle Stämme mindestens aus drei Graphemen bestehen müssen und einen zweifach besetzten Reim enthalten. Für das Suffix *-er* wird *Tier* also nicht als Ausnahme gewertet, weil ⟨Ti-⟩ kein wohlgeformter Stamm ist; *Zimmer* hingegen ist für *-er* eine Ausnahme von der Eindeutigkeit, denn ⟨Zimm-⟩ erfüllt die Minimalitätsbedingungen für Stämme.

Außerdem ist die (deutsche) Untersuchung (wie Berg 2017) sensitiv für Groß- und Kleinschreibung. Das bedeutet, dass jeweils nach Substantiven und Nicht-Substantiven unterschieden wird: Als Ausnahme des Adjektivsuffixes *-ig* wird bspw. *Reisig* nicht gewertet, weil hier die initiale Großschreibung bereits eine falsche Interpretation des Wortendes als Suffix verhindert. Andersherum ist *außer* keine Ausnahme zur Eindeutigkeit des Substantivsuffixes *-er*, weil hier die Kleinschreibung die Interpretation als Substantiv unwahrscheinlich macht.

Im Folgenden gehen wir die Affixe in (91) der Reihe nach durch. Zunächst werden für jedes Affix die Eigenschaften der Ableitungen spezifiziert. Dann werden Ausnahmen von der Eindeutigkeit aufgeführt und – wo anwendbar – auch potenzielle Ausnahmen, die anders verschriftet werden.

-in: Ableitungen mit dem Movierungssuffix *-in* sind trivialerweise feminine Substantive, meist Berufsbezeichnungen, aber auch Personen- oder Tierbezeichnungen (*Ärztin*, *Gattin*, *Hündin*), vgl. z. B. Fleischer/Barz (2012: 236 f.). Hier gibt es im CELEX-Teilkorpus 29 Ausnahmen wie *Medizin*, *Termin*, *Trampolin*, *Hermelin*. Wörter, die potenziell mit *-in* geschrieben werden könnten, aber anders verschriftet werden, gibt es nicht.

-er: Das Substantivsuffix *-er* bildet maskuline Nomina Agenti, Instrumenti oder Acti (vgl. z. B. Fleischer/Barz 2012: 201 ff.). Ausnahmen von der Eindeutigkeit sind also a) nicht-maskuline Substantive und b) Maskulina, die keine Person, kein Instrument und kein Abstraktum bezeichnen. *Vater* bspw. bezeichnet eine Person, *Hammer* ein Instrument und *Kummer* ein Abstraktum – diese und ähnliche Einträge verhalten sich also formal und semantisch wie *-er*-Ableitungen. Schließen wir diese aus, so verbleiben im Korpus 167 Ausnahmen wie die folgenden:

(92) Faser, Holunder, Kammer, Kloster, Schiefer, Tochter, Wunder, Zucker

All diese Fälle sehen aus wie *-er*-Ableitungen, unterscheiden sich aber von dieser Gruppe semantisch. Wir können die Eindeutigkeit auch nicht dadurch ‚retten‘, dass wir das *-er* in Fällen wie (92) als Pseudosuffix klassifizieren: Anders als finales *<en>* und *<e>* alterniert es nicht mit Derivationssuffixen (wie bei *Garten* – *Gärtchen* etc.). Ob es darüber hinaus potenzielle Ausnahmen gibt, hängt wie oben bei der Flexion davon ab, wie weit die Vokalqualität der zweiten Silben von *Laster* und *Pasta* phonetisch differenziert wird; es spricht einiges dafür, dass die Vokale nicht zusammenfallen (Tomaschek/Berg Ms.). *Pasta*, *Thema* und *Algebra* sind also keine potenziellen Ausnahmen.

-ant/-ent: Ableitungen auf *-ant* und *-ent* sind meist Nomina Agenti von *-ier[en]*-Verben (mit Ersetzung des Verbsuffixes): *konsumier[en]* – *Konsument* (vgl. z. B. Fleischer/Barz 2012: 245). Da die Kategorie Nomina Agentis auf die Basis Bezug nimmt, formulieren wir die Outputbeschränkung etwas allgemeiner: *-ant/-ent*-Ableitungen sind Personenbezeichnungen. Im CELEX-Teilkorpus gibt es 17 Ausnahmen von der Eindeutigkeit wie die folgenden:

(93) Elefant, Instrument, Kontinent, Moment, Trabant

Neben diesen 17 Ausnahmen gibt es acht potenzielle Ausnahmen wie *Zustand*, *Abstand*, *Gewand* und *Niemand*; in all diesen Fällen ergibt sich die Schreibung aber aus der Stammkonstanz (*Zustand* – *Zustände*).

-ar/-är: Ableitungen mit dem Adjektivsuffix *-ar/-är* sind Adjektive – spezifischere Beschränkungen des Outputs lassen sich nicht machen (zumindest nicht, ohne auf die Basis zu verweisen, vgl. z. B. Fleischer/Barz 2012: 349). Es gibt hier nur eine einzige Ausnahme, den Verbstamm *gebär[en]* – dessen Formen aber aufgrund des *ge*-Präfixes gut als Verben identifizierbar sind. Es gibt keine Präpositionen, Konjunktionen, Adverbien usw. auf *⟨ar⟩/⟨är⟩*. Es gibt außerdem keine potenziellen Ausnahmen, die mit *⟨ar⟩/⟨är⟩* verschriftet werden könnten.

-ig: Mit *-ig* abgeleitete Wörter sind Adjektive; viel weiter kann diese Gruppe nicht eingeschränkt werden (vgl. z. B. Fleischer/Barz 2012: 337). Wir suchen also Nicht-Adjektive (und aufgrund der Kleinschreibung Nicht-Substantive), die mit *⟨ig⟩* enden. Die gibt es nicht – es sei denn, man zählt die Kardinalzahlen der Zehner-Produkte wie *zwanzig*, *dreißig*, *vierzig* usw. dazu. Hier handelt es sich aber um Ableitungen mit *-zig* – außerdem verhalten sich die Numeralia in vielerlei Hinsicht wie Adjektive (vgl. Eisenberg 2013b: 162). Es gibt nur zwei potenzielle Ausnahmen, *antik* und *magnifik* – und bei beiden ergibt sich die Schreibung des Wortes unabhängig aus dem Prinzip der Stammkonstanz.

-al/-ell: Bei Ableitungen auf *-al/-ell* handelt es sich um Adjektive, konkreter lassen sich die adjektivischen Bildungen ohne Bezug auf die Basis nicht fassen (vgl. z. B. Fleischer/Barz 2012: 349 f.). Gesucht werden also Nicht-Adjektive (und aufgrund der Kleinschreibung Nicht-Substantive), die mit *⟨al⟩/⟨ell⟩* enden. Die einzige Ausnahme ist der Verbstamm *gesell[en]*, dessen Formen aufgrund des Präfixes gut als Verbformen identifiziert werden können. Potenzielle Ausnahmen gibt es nicht.

-iv: Auch hier ist das einzige Kriterium, das angesetzt werden kann, der Status als Adjektiv (vgl. z. B. Fleischer/Barz 2012: 350). Gesucht werden also Nicht-Adjektive (und aufgrund der Kleinschreibung Nicht-Substantive), die mit ⟨iv⟩ enden. Solche Einträge gibt es im Korpus nicht; potenzielle Ausnahmen sind ebenfalls nicht zu finden.

-os/-ös: Wie bei den vergangenen vier Suffixen ist auch hier ist das einzige Kriterium, das angesetzt werden kann, der Status als Adjektiv (vgl. z. B. Fleischer/Barz 2012: 350). Gesucht werden Nicht-Adjektive (und aufgrund der Kleinschreibung Nicht-Substantive), die mit ⟨os⟩/⟨ös⟩ enden. Es gibt hier weder tatsächliche noch potenzielle Ausnahmen von der Eindeutigkeit.

un-: Da das Präfix *un-* die Basisbedeutung negiert (vgl. z. B. Fleischer/Barz 2012: 354) und die Basis im Rahmen dieser Arbeit nicht Teil der Definition der Ableitungen sein kann, bleibt auch hier nur das Kriterium der Wortart ‚Adjektiv‘. Gesucht werden Nicht-Adjektive (und aufgrund der Kleinschreibung Nicht-Substantive), die mit ⟨un⟩ beginnen. Hier gibt es drei Ausnahmen von der Eindeutigkeit, *unser*, *unten* und *unter*; diese Einträge verhalten sich graphematisch (fast) wie die Ableitungen *ungern* oder *unwert*. Potenzielle Ausnahmen gibt es nicht.

be-: Das Verbpräfix *be-* leitet transitive Verben ab (vgl. z. B. Fleischer/Barz 2012: 383 ff.). Ausnahmen von der Eindeutigkeit sind also a) Nicht-Verben (und aufgrund der Kleinschreibung Nicht-Substantive) sowie b) intransitive Verben, die mit ⟨be⟩ beginnen. Hier gibt es folgende Ausnahmen von der Eindeutigkeit:

(94a) bewusst, berüchtigt, beschieden, beflissen, betucht

(94b) besonders, bereit, bequem, behände

Die Formen in (94a) sind deutlich als Partizipien erkennbar, auch wenn sie synchron nicht mehr auf das entsprechende Verb zurückführbar sind. Das spielt aber hier keine Rolle: Partizipien von *be-*Verben können regelmäßig adjektivisch gebraucht werden (*befahren*, *bedruckt*, *beglückwünscht*). Die eigentlichen Ausnahmen sind die vier Belege in (94b). Potenzielle Ausnahmen gibt es keine.

er-: Beim Verbpräfix *er-* steht nur die Wortart ‚Verb‘ als Definitionskriterium zur Verfügung, wenn man auf die Bedeutung der Basis verzichten möchte (vgl. z. B. Fleischer/Barz 2012: 386 ff.). Gesucht werden also Nicht-Verben (und aufgrund der Kleinschreibung Nicht-Substantive), die mit ⟨er⟩ beginnen. Solche Einträge gibt es nicht, allenfalls – wie oben bei den *be-*Verben – *er-*Partizipien (*erwünscht*,

ergeben, erlaucht, erstunken), die nicht als Ausnahmen von der Eindeutigkeit gewertet werden. Potenzielle Ausnahmen gibt es nicht.

Die Ergebnisse zu den untersuchten deutschen Affixen sind in Tabelle 113 zusammengestellt. Ernstzunehmende Ausnahmen von der Eindeutigkeit gibt es nur bei drei Affixen, und zwar den Suffixen *-er*, *-in* und *-ant/-ent*.

Tab. 113: Eindeutigkeit deutscher Derivationsaffixe. Für jedes der bearbeiteten Affixe: die Anzahl der Ausnahmen (monomorphematische Stämme mit homographischem Wortende aus CELEX) mit einem Beispiel; die Anzahl potenzieller Ausnahmen (Wörter, die phonographisch so verschrieben werden könnten) mit einem Beispiel. * alles Instanzen von Morphemkonstanz.

Affix	Anzahl Ausnahmen	Beispiel	Anzahl potenzieller Ausnahmen	Beispiel
-in	29	Termin	–	–
-er	167	Faser	–	–
-ant/-ent	17	Elefant	8*	(Zustand)
-ar/-är	1	gebär[en]	–	–
-ig	–	–	2*	(antik)
-al/-ell	1	gesell[en]	–	–
-iv	–	–	–	–
-os/-ös	–	–	–	–
un-	3	unter	–	–
be-	4	besonders	–	–
er-	–	–	–	–

Die mit Abstand meisten Ausnahmen gibt es bei *-er*; das ist auch im Englischen der Fall (siehe unten). Für *-in* und *-ant/-ent* gibt es deutlich weniger Ausnahmen; die beiden Suffixe unterscheiden sich aber in ihrer Produktivität: Das Movie-*er*-Suffix kann mit wenigen Ausnahmen jedes Substantiv ableiten, das mit dem produktiven Suffix *-er* abgeleitet wurde, während *-ant/-ent* nicht produktiv ist. Die Ausnahmen von der Eindeutigkeit sollten also bei *-in* weniger stark ins Gewicht fallen als bei *-ant/-ent*. Für eine umfassendere Beurteilung der Eindeutigkeit können auch die Tokenfrequenzen der einzelnen Bildungen erfasst werden.

Im Deutschen gibt es also – wenn wir die Ergebnisse der Stichprobe verallgemeinern – nur relativ wenige Ausnahmen von der Eindeutigkeit der Derivationsaffixe. Außerdem gibt es kaum potenzielle Ausnahmen. Das liegt vor allem daran, dass auch in Nebensilben mehrere Vokalqualitäten unterschieden werden. Das genaue Ausmaß dieser Differenzierung ist allerdings nicht klar; hier handelt es sich um ein Forschungsdesiderat.

Im Englischen sind in Berg/Aronoff (2017) bereits vier einschlägige Suffixe untersucht worden (*-ic*, *-ous*, *-al* und *-y*). Die Ergebnisse werden im Folgenden kurz vorgestellt, bevor weitere Suffixe untersucht werden.¹¹⁰

-ous: Das Suffix leitet Adjektive aus Substantiven ab (vgl. z. B. Marchand 1969: 339 f.). Gegenbeispiele zur Eindeutigkeit sind also Nicht-Adjektive, die mit <ous> enden. Die gibt es im gesamten CELEX-Korpus nicht: Wann immer ein Leser auf das Wortende <ous> trifft, verweist das Wortende auf das Suffix *-ous*, und das Wort ist ein Adjektiv. Es gibt außerdem eine Reihe potenzieller Ausnahmen: 320 Wörter enden auf /is/¹¹¹ und könnten mit <ous> verschriftet werden¹¹² – sie werden aber anders verschriftet, z. B. mit <us> (z. B. *bonus*), <is> (z. B. *tennis*), <ess> (z. B. *fortress*) oder <ice> (z. B. *service*). Und von diesen 320 Wörtern können nur sechs adjektivisch verwendet werden (z. B. *bogus*, *primus*, *emeritus*); der Status der meisten dieser sechs ist zumindest zweifelhaft (vgl. Berg/Aronoff 2017: 44 f., Fn. 8).

-ic: Auch *-ic* ist ein desubstantivisches Adjektivsuffix (vgl. z. B. Marchand 1969: 294 ff.). Gegenbeispiele sind hier daher Nicht-Adjektive, die auf <ic> enden; von diesen gibt es im CELEX-Korpus 18 (z. B. *attic*, *critic*, *fabric*). Es gibt aber darüber hinaus 38 Wörter, die phonologisch mit /ik/ enden und die mit <ic> geschrieben werden könnten. Die prominentesten ‚Ausweichschreibungen‘ sind hier <ock> (z. B. *haddock*) und <ick> (z. B. *derrick*). Von diesen Wörtern ist nur eines als Adjektiv verwendbar (*elegiac*).

-al: Zwei *al*-Suffixe können unterschieden werden, ein substantivisches (Typ *arrival*, vgl. z. B. Marchand 1969: 236 f.) und ein häufigeres adjektivisches (Typ *accidental*, vgl. z. B. Marchand 1969: 238 ff.). Gegenbeispiele zur Eindeutigkeit wären hier also Nicht-Substantive und Nicht-Adjektive mit finalelem /il/, die mit <al> ver-

110 Im Unterschied zur vorliegenden Untersuchung wurde in Berg/Aronoff (2017) das gesamte CELEX-Korpus durchsucht.

111 Flemming/Johnson (2007) zeigen empirisch, dass es unter den nicht-finalen reduzierten Vokalen im amerikanischen Englisch im Prinzip nur noch eine Vokalqualität gibt, die sie phonetisch mit [i] bezeichnen, dem Symbol für einen hohen, mittleren, ungerundeten Vokal. Die berühmte Unterscheidung der Vokalqualität in den jeweils zweiten Silben von *Rosa's* und *roses* ist von dieser Generalisierung nicht betroffen, weil die Vokale hier stamfinal auftreten; in dieser Position werden mindestens zwei Vokalqualitäten unterschieden, die Wörter sind nicht homophon.

112 Wenn [i] Teil eines längeren Suffixes ist, werden die Belege ausgeschlossen. Das betrifft die Suffixe *-less*, *-ness*, *-itis* und *-osis*: Bei *helpless* erwarten wir keine Verschriftung mit <ous>, das ergibt sich aber aus der Konstantanschreibung des Suffixes *-less*.

schriftet werden. Von diesen gibt es im Korpus 22 (z. B. *equal*, *local*, *spiral*). Nur drei dieser Wörter sind aber exklusiv verbal (z. B. *outrival*) – bei den übrigen kann es sich auch um Konversionen von Adjektiven oder Substantiven handeln. Es gibt außerdem 598 Wörter, die auf /i/ enden und nicht mit ⟨al⟩ verschriftet werden. Die häufigsten Muster sind ⟨le⟩ (z. B. *crumble*), ⟨el⟩ (z. B. *channel*), ⟨il⟩ (z. B. *devil*) oder ⟨yl⟩ (z. B. *ethyl*).

-y: Es können mindestens drei -y-Suffixe unterschieden werden (vgl. Bauer/Lieber/Plag 2013), das produktive adjektivische -y (Typ *windy*), das substantivische -y (Typ *harmony*) sowie das Diminutivsuffix -y (Typ *granny*). Ausnahmen von der Eindeutigkeit sind hier also Nicht-Adjektive und Nicht-Substantive, die phonologisch auf /i/ und graphematisch auf ⟨y⟩ enden (alle -y-Diminutive sind substantivisch, deswegen erweitert diese Kategorie die Extension nicht). Im CELEX-Korpus gibt es 133 Wörter dieses Typs (z. B. *belly*, *candy*, *lobby*)¹¹³ – aber nur elf dieser Fälle können nicht als adjektivisch oder substantivisch verwendet werden (z. B. *accompany*, *bury*, *marry*). Bei den übrigen Fällen kann es sich auch um Konversion handeln. Außerdem gibt es hier 262 Wörter, die phonologisch mit /i/ enden und also mit ⟨y⟩ verschriftet werden könnten, die aber graphematisch ‚ausweichen‘. Die häufigsten Muster sind ⟨i⟩ (z. B. *Israeli*), ⟨ey⟩ (z. B. *chimney*) oder ⟨ie⟩ (z. B. *brownie*). Das Wortende -i ist ein distinktes Suffix, ein ‚Ethnonym‘ (Marchand 1969: 354 f.). Interessanter ist aber ⟨ie⟩: Diese Schreibung ist eine Variante des Diminutivsuffixes -y (vgl. z. B. Bauer/Lieber/Plag 2013: 388 ff.). Es ist eine interessante Frage, ob die -ie-Variante im Laufe der Zeit zulasten der -y-Variante zunimmt – bei einigen Wörtern wie *cabbie/cabby* scheint das in der Tat der Fall zu sein. Hier werden eventuell gerade zwei Suffixe graphematisch differenziert, das Substantiv- bzw. Adjektivsuffix -y und das homophone Diminutivsuffix (vgl. Berg/Aronoff 2017).

Bei diesen vier Suffixen gibt es jeweils nur relativ wenige Ausnahmen von der Eindeutigkeit. Zum Teil ist die genaue Anzahl abhängig davon, was als Ausnahme gewertet wird: Wenn -al Substantive und Adjektive ableitet, ist dann ein Verb wie *spiral* eine Ausnahme? Es handelt sich hier ja offensichtlich um Konversion aus einem Substantiv. Die Alternative ist, nur diejenigen Einträge zu zählen, die nie als Adjektive oder Substantive auftreten (wie z. B. das Verb *outrival*). Wie auch immer die Daten interpretiert werden, es gibt für jedes Suffix wesentlich mehr potenzielle als tatsächliche Ausnahmen.

¹¹³ Fälle, in denen finales ⟨y⟩ Teil eines längeren Suffixes ist, werden ausgeschlossen; das betrifft die Suffixe -acy, -ary/-ery/-ory, -ancy/-ency, -ey, -ity, -ly und -ry.

Historisch zeigt sich, dass die heutige Form der Suffixe sich in den meisten Fällen in einem langen und stetigen Prozess der Selbst-Organisation herausgebildet hat, ohne dass eine Institution oder Personengruppe diesen graphematischen Wandel gesteuert hätte (Berg/Aronoff 2017).

Neben diesen Suffixen werden in dieser Arbeit die folgenden Affixe untersucht:

(95a) substantivische Affixe: *-er/-or, -ion*

(95b) adjektivische Affixe: *un-, -ive*

(95c) verbale Affixe: *-ise/-ize*

Damit wird ein Querschnitt durch den Affixbestand des Englischen analysiert, produktive (*-er*) und unproduktive (*-ion*); Präfixe (*un-*) und Suffixe (*-ise*); native (*-er*) und fremde (*-ive*).

Wir gehen hier vor wie oben bei der Untersuchung der deutschen Affixe: Zunächst werden für jedes Affix die Eigenschaften der Ableitungen spezifiziert. Dann werden Ausnahmen von der Eindeutigkeit aufgeführt und – wo anwendbar – auch potenzielle Ausnahmen, die anders verschriftet werden.

-er/-or: Dieses Suffix bildet Nomina Agenti (Typ *baker*, vgl. z. B. Marchand 1969: 273 ff.). Ausnahmen von der Eindeutigkeit sind also a) Nicht-Substantive oder b) Substantive, die nicht Personen bezeichnen. Hier gibt es 141 Ausnahmen (z. B. *never, consider, monitor*). Wenn wir Einträge entfernen, die das Verbsuffix *-er* (Typ *clatter*) enthalten könnten (nach Marchand 1969: 273 „disyllabic verbs expressive of sound or movement“ wie *flicker, stutter*), dann verbleiben noch 133 Ausnahmen von der Eindeutigkeit. Umgekehrt gibt es 83 Wörter, die mit finalelem ⟨er⟩ oder ⟨or⟩ geschrieben werden könnten, aber anders verschriftet werden. Die häufigsten Muster der Differenzierung sind ⟨ar⟩ (z. B. *cellar*), ⟨re⟩ (z. B. *centre*), ⟨ur⟩ (z. B. *murmur*) und ⟨our⟩ (z. B. *favour*). In diesem Sinne interpretiert auch Aronoff (1978) die Unterscheidung zwischen *-our* und *-or*, die das britische Englisch macht (nicht aber das amerikanische Englisch, vgl. AmE *color*, BrE *colour*): Die *-our*-Wörter sind fast ausschließlich unbelebte, nichtagentivische Substantive, während die *-or*-Wörter in der Mehrzahl agentivische Substantive sind – die Differenzierung dieser beiden Suffixe im britischen Englisch spiegelt also einen morphologischen Unterschied wider.

-ion: Bauer/Lieber/Plag (2013: 201 ff.) setzen für dieses Suffix die Formen {*-ation, -cation, -ion, -ition, -iation, -sion, -ution* und *-tion*} an. Das Suffix leitet abstrakte Substantive ab. Ausnahmen von der Eindeutigkeit dieses Affixes wären also a) Nicht Substantive auf ⟨ion⟩ (damit sind alle längeren Formen eingeschlossen)

oder b) Konkreta auf ⟨ion⟩. Solche Ausnahmen gibt es im Korpus nicht. Es gibt aber insgesamt sechs potenzielle Ausnahmen, die mit /ʃən/ (*gentian* wie *petition*) oder /jən/ (*alien*, *canyon*, *truncheon* wie *companion*) enden. Mit /zən/ (analog zu *explosion*) endet kein Beleg.

un-: Dieses Präfix negiert die Basisbedeutung (vgl. z. B. Marchand 1969: 201). Da die Basis für die Definition der Ableitungen im Rahmen dieser Arbeit nicht verwendet wird, bleibt nur die Beschränkung auf Adjektive. Ausnahmen von der Eindeutigkeit sind hier also Nicht-Adjektive, die mit ⟨un⟩ beginnen. Dafür gibt es im Korpus nur einen Beleg, *under*. Potenzielle Ausnahmen gibt es nicht.

-ive: Bei *-ive* handelt es sich um ein Adjektivsuffix (vgl. z. B. Marchand 1969: 315). Weitere Kriterien können nicht angesetzt werden, daher wird hier nach Nicht-Adjektiven gesucht, die auf ⟨ive⟩ enden. Davon gibt es im Korpus nur vier (*arrive*, *connive*, *contrive*, *motive*). Ein Beleg endet darüber hinaus mit /iv/, könnte also ebenfalls mit ⟨ive⟩ geschrieben werden, *octave*.

-ise/-ize: Das Suffix leitet Verben ab (vgl. z. B. Marchand 1969: 318 ff.). Da keine weiteren Kriterien (wie z. B. die Stelligkeit) angesetzt werden können, wird nach Nicht-Verben gesucht, die auf ⟨ise⟩/⟨ize⟩ enden. Für diesen Typ gibt es sechs Belege im CELEX-Korpus (*cerise*, *paradise*, *precise*, *premise*, *promise*, *valise*). Es gibt keine potenziellen Ausnahmen.

Die Ergebnisse zu den untersuchten englischen Affixen sind in Tabelle 114 zusammengestellt: Im englischen Schriftsystem werden bestimmte Schreibungen für bestimmte Suffixe ‚reserviert‘. Für das Präfix *un-* und die längeren Suffixe *-ion*, *-ive* und *-ise/-ize* gilt das nur eingeschränkt: Hier gibt es kaum tatsächliche Ausnahmen und potenzielle Ausnahmen. Das ist stochastisch nicht überraschend: Je aufwendiger die graphematische und die phonologische Form eines Affixes sind, desto unwahrscheinlicher sind tatsächliche und potenzielle Ausnahmen. Es gibt nur ein Affix mit einer substanziellen Anzahl von tatsächlichen Ausnahmen, *-er*; die übrigen Affixe sind eindeutig i. S. v. distinktiv. Zur Erinnerung: Wenn eine bestimmte Schreibung distinktiv ist, so bedeutet das nicht zwangsläufig, dass sie eindeutig auf ein Affix verweist – es bedeutet, dass sie auf ein oder mehrere homographische Affixe verweist und dass es keine homographischen Wortenden gibt. Finales ⟨al⟩ repräsentiert eines von zwei Suffixen, ist aber kein gutes Wortende morphologisch einfacher Wörter. Das ist ein wichtiger Punkt: Hier unterscheidet die Schrift lediglich morphologische von nicht-morphologischer Information. Die Art der morphologischen Information wird nicht weiter spezifiziert.

Tab. 114: Eindeutigkeit englischer Derivationsaffixe. Für jedes der bearbeiteten Affixe: die Anzahl der Ausnahmen (monomorphematische Stämme mit homographischem Wortende aus CELEX) mit einem Beispiel; die Anzahl potenzieller Ausnahmen (Wörter, die phonographisch so verschriftet werden könnten) mit einem Beispiel. * In Klammern: Anzahl der Belege, die als Ausnahme von der Eindeutigkeit gewertet werden können (z. B. *spiral* kann auch als Verb verwendet werden); ohne Klammern: Anzahl der Belege, die als Ausnahme von der Eindeutigkeit gewertet werden müssen. ** In Klammern: Anzahl der Ausnahmen; ohne Klammern: Ausnahmen ohne Verben mit dem Suffix *-er* (Typ *clatter*).

Affix	Anzahl Ausnahmen	Beispiel	Anzahl potenzieller Ausnahmen	Beispiel
-ous	–	–	320	bonus
-ic	18	critic	38	derrick
-al	(22) 3*	outrival	598	crumble
-y	(133) 11*	bury	262	chimney
-er/-or	(141) 133**	consider	83	cellar
-ion	–	–	6	canyon
un-	1	under	–	–
-ive	4	arrive	1	octave
-ise/-ize	6	paradise	–	–

Insgesamt scheint es weder im Deutschen noch im Englischen viele Ausnahmen von der Eindeutigkeit der Derivationsaffixe zu geben. Je länger die Affixe sind, desto unwahrscheinlicher werden solche (tatsächlichen wie auch potenziellen) Ausnahmen. Einzig *-er* ist in beiden Sprachen nicht eindeutig (z. B. *Faser*, *consider*). Das ist nicht überraschend: Das Wortende *«er»* nimmt im Englischen und noch stärker im Deutschen eine zentrale Stellung in morphologisch einfachen Wörtern ein. Hier kann offenbar nicht einmal eines der produktivsten Wortbildungsmuster (*-er*) zu einem Abbau der konkurrierenden nicht-morphologischen *«er»*-Schreibungen führen.

4.5 Zusammenfassung

Tragen wir zunächst die Beobachtungen aus den Teilkapiteln zusammen. Für die vier untersuchten Dimensionen der Morphemkonstanz sind das die folgenden:

Einheitlichkeit der Stämme:

Stämme werden im Deutschen weitestgehend einheitlich geschrieben – das war bereits vor der Untersuchung bekannt, und daran hat sich nichts geändert. Es gibt nur wenige Ausnahmen:

- Es gibt knapp 200 unregelmäßig flektierende Verben im Deutschen; hier variiert die Form innerhalb des Flexionsparadigmas (z. B. ‹nehmen›, ‹nimmst›, *‹nehmst›). Wären die Formen konkatentativ aufgebaut, würde es sich nicht um unregelmäßige Verben handeln; diese Aussage ist also beinahe tautologisch.
- Stämme mit den Pseudosuffixen ‹e› und ‹en› ‚kappen‘ das Pseudosuffix in einigen Ableitungen und Komposita (‹Wolke›, ‹wolkig›); die Pseudosuffixe ‹er› und ‹el› variieren hinsichtlich der An-/Abwesenheit des ‹e› in Flexion und Derivation (‹Fieber›, ‹fiebrig›).
- In einigen Anglizismen variieren einfache und verdoppelte Konsonanten (‹fit›, ‹fitter›, *‹fitt›); allerdings nicht innerhalb der Flexionsparadigmen. Einige wenige Anglizismen variieren hinsichtlich Tilgung bzw. Bewahrung des stamfinalen ‹e› (‹gelikt›, ‹geliket›, ‹geliked›).
- Einige Dutzend Substantive mit den Endungen ‹as›, ‹es›, ‹is›, ‹os› und ‹us› variieren innerhalb der Flexionsparadigmen hinsichtlich Konsonantenverdoppelung (z. B. ‹Kürbis›, ‹Kürbisse›, *‹Kürbiss›).
- Einige Lexeme mit umgelautetem Doppelvokal variieren, indem der umgelautete Vokal nur einfach auftritt (‹Haar›, ‹Härchen›, *‹Har›, *‹Häärchen›).
- Im lateinischen Bereich des Wortschatzes variiert in einigen Fällen der stamfinale Konsonant (‹explodier[en]›, ‹Explosion›, *‹Explodion›, *‹explosier[en]›) oder der Vokal der Endung (‹Muskel›, ‹muskulär›, *‹muskelär›).
- Ein Grenzfall sind schließlich die Umlautschreibungen wie ‹Äpfel›: Hier wurde dafür argumentiert, dass die Stammform minimalinvasiv modifiziert wird. Die Gestalt des Stamms bleibt intakt, das Trema markiert die morphologische Veränderung.

Die Ausnahmen von der Einheitlichkeit beziehen sich also neben den unregelmäßigen Verben vor allem auf fremde Elemente und auf Pseudosuffixe. Im nativen Bereich sind zumindest die Wurzeln (also der Teil der Stämme abzüglich der Endung) stabil.

Im Englischen gibt es mehr Ausnahmen; das liegt daran, dass es drei systematische Alternationen gibt, die Konsonantenverdoppelung (‹swim›, ‹swimming›, *‹swimm›), die ‹e›-Tilgung (‹dine›, ‹dining›, *‹din›) sowie die ‹i/‹y›-Alternation (‹lady›, ‹ladies›, *‹ladie›). Damit ist ein substanzialer Teil der englischen Stämme uneinheitlich. Die Alternationen sind selbstverständlich nicht idiosynkratisch; sie sind automatisch und damit vorhersagbar. Neben diesen drei Ausnahmen von der Einheitlichkeit gibt es noch einige weitere:

- Auch im Englischen sind unregelmäßige Verben ein Hort für Variation im Paradigma (z. B. ‹think›, ‹thought›, *‹thinked›).

- In einem guten Dutzend Fälle alterniert der finale Konsonant (z. B. ⟨elf⟩, ⟨elves⟩, ⟨elvish⟩, *⟨elve⟩).
- Stämme auf ⟨er⟩ und ⟨le⟩ variieren in Ableitungen hinsichtlich der An-/Abwesenheit des ⟨e⟩ (z. B. ⟨register⟩, ⟨registration⟩, *⟨registration⟩; ⟨feeble⟩, ⟨feebly⟩, *⟨feeblely⟩).
- Im lateinischen Bereich des Wortschatzes variiert in einigen Fällen der stamminale Konsonant (⟨explode⟩, ⟨explosion⟩, *⟨explodion⟩, *⟨explose⟩) oder der Vokal der Endung (⟨muscle⟩, ⟨muscular⟩, *⟨musclelar⟩). Hier gibt es große Überschneidungen zwischen dem Englischen und dem Deutschen.

Wir können hier festhalten: Im Englischen gibt es also mehr Ausnahmen von der Einheitlichkeit der Stämme als im Deutschen. Dafür spricht auch der Vergleich mit der Phonologie: Im Deutschen wird eine Reihe phonologischer Alternationen graphematisch überbrückt; im Englischen ist dieser Fall marginal.

Eindeutigkeit der Stämme:

Graphematische Formen von Stämmen repräsentieren diese Stämme in beiden Sprachen meist eindeutig – es gibt vergleichsweise wenig Homographie (nur 3,3% der Formen im Deutschen und 4,9% der englischen Formen sind homographisch). Handelt es sich bei dieser Dimension von Morphemkonstanz also um ein Strukturprinzip im Deutschen und Englischen? Dagegen spricht, dass es in beiden Sprachen eine ganze Reihe von Homophonen gibt, die potenziell graphematisch differenziert werden könnten, die aber in der Schrift zusammenfallen. So werden im deutschen Korpus nur 18% aller Homophone graphematisch differenziert. Es handelt sich hier also eindeutig nicht um ein Strukturprinzip des Deutschen. Im Englischen liegt der Anteil wesentlich höher, bei 46%. Auch hier werden Homophone aber nicht systematisch differenziert. Es könnte sich bei den heterographischen Schreibungen um fossilisierte phonologische Distinktionen handeln.

Einheitlichkeit der Affixe:

Es gibt in beiden Sprachen Ausnahmen von der Einheitlichkeit der Affixe. Im Deutschen sind das folgende Fälle:

- In der Flexion sind viele Suffixe nicht-distinkt (-(e)s, -(e)st, -e(t), -e; vgl. z. B. (des) ⟨Hauses⟩, (des) ⟨Krauts⟩; (du) ⟨spielst⟩, (du) ⟨tanzt⟩).
- Auch in der Derivation gibt es uneinheitliche Affixe, wenn auch vergleichsweise wenige: zum einen Suffixe, deren Form sich bei weiterer Ableitung oder Flexion ändert (⟨-nis⟩, ⟨-in⟩, ⟨-abel⟩, ⟨-ös⟩, ⟨-är⟩, ⟨-al⟩; vgl. z. B. ⟨Hindernis⟩, ⟨Hindernisse⟩, *⟨Hinderniss⟩; ⟨generös⟩, ⟨Generosität⟩, *⟨generos⟩).

- Zum anderen gibt es Suffixe, deren Form von der jeweiligen Basis abhängt (*-abel/-ibel*, *-al/-ell*, *-ant/-ent*, *-heit/-keit*, vgl. z. B. *⟨diskutabel⟩*, *⟨disponibel⟩*; *⟨Eitelkeit⟩*, *⟨Schönheit⟩*).
- Fälle, in denen demgegenüber phonologische Variation graphematisch überbrückt wird, sind rar.

Im Englischen ist formale Variation bei Affixen etwas seltener. Es geht hier um die folgenden Fälle:

- In der Flexion ist Einheitlichkeit die Regel; Variation gibt es nur bei *-s* (*⟨cats⟩* vs. *⟨houses⟩*).
- In vielen anderen Fällen wird phonologische Variation graphematisch überbrückt, vgl. z. B. die Form des Präterital-/Partizipsuffixes, *⟨-ed⟩*, gegenüber seinen phonologisch konditionierten Allomorphen */t/*, */d/*, */ɪd/*.
- In der Derivation gibt es einige Suffixe, die bei weiterer Ableitung oder unter Flexion variieren (*-y*, *-ery*, *-ate*, *-able*, *-our*, *-ous*, z. B. *⟨generous⟩*, *⟨generosity⟩*, **⟨generos⟩*); zwei dieser Fälle haben eine direkte Entsprechung im Deutschen.
- Außerdem gibt es Suffixe, deren Form abhängig von der jeweiligen Basis variiert (*-able/-ible*, *-ce/-cy*, *-ant/-ent*, *-ise/-ize*, *-er/-or/-ar*; vgl. z. B. *⟨readable⟩*, *⟨divisible⟩*). Hier gibt es z. T. Überschneidungen mit den Fällen im Deutschen.
- Bei den Derivationsuffixen werden ebenfalls an einigen Stellen phonologische Alternationen überbrückt, vgl. z. B. die einheitliche graphematische Form *⟨-ic⟩* im Gegensatz zu ihren phonologischen Varianten */ɪk/*, */ɪs/*, */ɪʃ/*.

Insgesamt scheint es im Englischen weniger Ausnahmen von der Einheitlichkeit zu geben als im Deutschen. Vor allem gibt es einige Fälle, in denen eindeutig phonologische Variation graphematisch überbrückt wird. Das ist eine gänzlich andere Situation als bei der Einheitlichkeit der Stämme; die Einheitlichkeit der Affixe ist im Englischen strukturell wichtiger.

Eindeutigkeit der Affixe:

Grundsätzlich gilt: Je umfangreicher die graphematische Form von Affixen ist, desto eindeutiger repräsentiert diese Form das Affix. Bei den kürzeren Suffixen zeigt sich allerdings ein Unterschied zwischen den beiden Sprachen. Im Deutschen finden wir folgende Situation vor:

- Die wenigsten deutschen Flexionsaffixe sind eindeutig, und viele sind auch nicht distinkt; *⟨-e⟩* bspw. repräsentiert eine Reihe verschiedener Suffixe, und es gibt weit über tausend morphologisch einfache Substantive, die auf *⟨e⟩* enden.

- Bei den Derivationsaffixen gibt es weit weniger Ausnahmen (vor allem *-er*, in geringerem Maße auch *-in* und *-ant/-ent*). Derivationsaffixe sind allerdings typischerweise auch länger als Flexionsaffixe, daher ist die Wahrscheinlichkeit homographischer Wortenden ohnehin niedriger.
- Die Frage, ob bestimmte Schreibungen für Affixe reserviert sind, kann nicht eindeutig beantwortet werden. Es gibt einige wenige Fälle, in denen Affixe und Wortenden graphematisch differenziert werden.

Im Englischen finden wir deutlich weniger Ausnahmen von der Eindeutigkeit:

- In der Flexion ist zwar eine Reihe von Suffixen nicht einheitlich – so repräsentiert <-s> z. B. die 3. Ps. Sg. bei Verben und den Plural bei Substantiven. Der Großteil der Suffixe ist aber distinkt: Wenn es homographische Wortenden gibt, so sind es wenige im Vergleich zur Anzahl der potenziell möglichen homographischen Wortenden. Hier werden deutlich Bereiche der Schrift für die Flexionsaffixe freigehalten.
- In der Derivation ist das ähnlich. Die einzige substanzielle Ausnahme von der Eindeutigkeit findet sich bei *-er/-or*; bei allen anderen überprüften Affixen ist die Zahl der tatsächlichen homographischen Wortenden kleiner als die Zahl der potenziell möglichen homographischen Wortenden.

Die Eindeutigkeit (bzw. im Fall der Flexionsaffixe: die Distinktivität) von Affixen ist im Englischen insgesamt höher als im Deutschen. Es gibt im Deutschen keinen eindeutigen Fall, in dem eine bestimmte Schreibung bei Wortenden vermieden wird, um ein bestimmtes Affix distinkter auszuzeichnen.

Fassen wir zusammen: Im Deutschen ist die Einheitlichkeit von Stämmen ein bestimmendes Strukturprinzip; für die übrigen drei Dimensionen finden sich jeweils mehr Ausnahmen als im Englischen. Im Englischen ist vor allem die Eindeutigkeit der Affixe prägend; ihre Einheitlichkeit hat ebenfalls nur wenige Ausnahmen. Etwas verallgemeinernd lässt sich feststellen, dass im Deutschen die Schreibung von Stämmen entscheidend ist, im Englischen die Dekodierung von Affixen – oder anders ausgedrückt: „das deutsche Schriftsystem [ist] eher lexikalisch geprägt, das englische grammatisch“ (Fuhrhop 2017: 71).

5 Zusammenfassung und Diskussion

In dieser Arbeit wurden zwei Arten von graphematischen Regularitäten untersucht, graphemische und morpho-graphematische. Im Folgenden werden wesentliche Ergebnisse dieser Untersuchung noch einmal zusammengetragen und diskutiert. Dabei soll zunächst dem Reihentitel Rechnung getragen werden: Wo konvergieren das deutsche und das englische Schriftsystem, wo divergieren sie? Was folgt, ist keine vollständige Übersicht über alle untersuchten Phänomene, sondern eine Auswahl besonders eindrücklicher Beispiele.

Überraschend ähnlich sind sich die beiden Systeme in der relativen Frequenz der Buchstaben (vgl. Abb. 1 oben, hier wiederholt als Abb. 49):

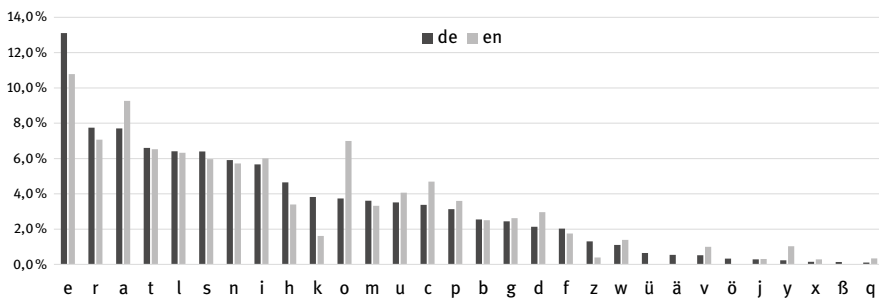


Abb. 49: Relative Anteile der Buchstaben im deutschen und englischen Teilkorpus einfacher Stämme an der Summe aller Buchstaben im jeweiligen Korpus. N(de): 30.772, N(en): 37.519.

Die relative Häufigkeit einiger Buchstaben weicht voneinander ab; so sind |e|, |h|, |k| und |z| im Deutschen etwas häufiger und |a|, |c|, |o| und |y| im Englischen. Insgesamt ist der Grad der Übereinstimmung aber bemerkenswert. Das schlägt sich in einem hohen Grad der Korrelation nieder (Pearsons $r = 0,93$, $p < 0,001$).

Auch die belegten und nicht belegten Bigramme (die Verbindungen von zwei Buchstaben) weisen große Übereinstimmungen auf. 90 % aller Bigramme, die im deutschen Korpus belegt sind, finden sich auch im englischen Korpus (siehe 3.2.1.3). Eine besonders interessante Teilklasse der Bigramme sind Verdoppelungen. Diejenigen Verdoppelungen, die signifikant häufiger sind als erwartet, stimmen größtenteils überein (Abb. 50).

englisch												
ee	oo	bb	dd	ff	gg	ll	mm	nn	pp	ss	zz	tt
deutsch												

Abb. 50: Verdoppelungen von Buchstaben, die signifikant häufiger sind, als stochastisch zu erwarten wäre.

Die Verdoppelung von Konsonantenbuchstaben ist in beiden Systemen zentral, und die Schnittmenge ist groß. |tt| ist im Englischen nicht häufiger als erwartbar – aber auch nicht seltener. Bei den Vokalen ist der Unterschied auf den ersten Blick ebenfalls klein. Im Deutschen ist nur |oo| häufiger als erwartet, im Englischen darüber hinaus auch |ee|. Diese Beschreibung verwischt aber die Tatsache, dass |oo| im Deutschen nur etwas häufiger ist, als zu erwarten wäre, während |ee| und |oo| im Englischen wesentlich häufiger sind. Die Verdoppelung von Vokalbuchstaben ist im Englischen zentraler als im Deutschen.

Typische Strukturen von Stämmen und Affixen hingegen konvergieren in den beiden Systemen weitestgehend. Die folgende Liste von Konvergenzen ist nicht abschließend:

- Die meisten Stämme in den Korpora sind zweisilbig, die meisten Affixe einsilbig.
- Die meisten einsilbigen Stämme haben einen einfach besetzten Anfangsrand, einen einfach besetzten Kern und einen zweifach besetzten Endrand (<CVCC>).
- Jede Abweichung vom Prototyp (mehr oder weniger besetzter Anfangsrand, Kern, Endrand) führt zu einer Abnahme der Belegzahlen: Es gibt bspw. mehr einsilbige Wörter vom Typ <CVCC> als vom Typ <CCVCC>, und von diesem Typ wiederum mehr als vom Typ <CCVC> etc.
- Die meisten einsilbigen Affixe haben einen leeren Anfangsrand und einen einfach besetzten Kern und Endrand (<VC>).
- Es gibt nur wenige Überschneidungen zwischen einsilbigen Stämmen und Affixen. Anders ausgedrückt: Die meisten Affixe sehen wie Affixe aus, die meisten Stämme wie Stämme. Die Struktur <CVC> (z. B. <Bus>, <-nis>, <gas>, <-ful>) ist die einzige Struktur, in der Stämme *und* Affixe quantitativ vorkommen (Abb. 46 oben, hier wiederholt als Abb. 51).

	-	C	CC	CCC
-		Affix		
C			Stamm	
CC				
CCC				

Abb. 51: Die Verteilung prototypischer einsilbiger Stämme und einsilbiger Affixe im Deutschen und Englischen.

Eine deutsche Besonderheit ist die Interaktion zwischen Anfangs- und Endrand. Bestimmte Grapheme kommen (in Autosemantika) nicht alleine im Endrand vor, wenn der Anfangsrand nicht oder nur einfach besetzt ist (*⟨Or⟩, *⟨fü[en]⟩). In diesem Kontext tritt das sog. Dehnungs-⟨h⟩ auf; es reguliert das Silbengewicht, indem es den Endrand schwerer macht (⟨Ohr⟩, ⟨fühl[en]⟩). Genauso kann auch das silbeninitiale ⟨h⟩ interpretiert werden: Es beschwert Stämme, die ansonsten die Minimalitätsbedingung des zweifach besetzten Reims nicht erfüllen würden (z. B. ⟨Schuh⟩, *⟨Schu⟩). Interessant ist in diesem Zusammenhang, dass die Domäne der ⟨h⟩-Schreibungen nicht das Wort oder der Stamm ist, sondern die Wurzel – also die Form, die nach Abzug der Endung übrigbleibt. So kann auch die ⟨h⟩-Schreibung in Fällen wie ⟨Sahn⟩ erfasst werden: Der Endrand der Wurzel *⟨San-⟩ ist zu leicht, deswegen wird sie mit ⟨h⟩ beschwert (⟨Sahn-⟩).

Eine solche Segmentierung in Wurzel und Endung (die im Deutschen häufig ein Pseudosuffix ist, also ⟨e⟩, ⟨er⟩, ⟨en⟩ und ⟨el⟩) drängt sich in beiden Systemen auf. Neben der Tatsache, dass sich die Beschränkungen über das Gewicht der Silbenkonstituenten nicht auf Stämme, sondern auf Wurzeln beziehen, spricht für eine solche Gliederung auch, dass Wurzeln häufig dieselbe Struktur haben wie einsilbige Stämme – ⟨Sahn-⟩ ist aufgebaut wie ⟨Hahn⟩. Das gilt allgemein für die typischen CV-Strukturen, es gilt aber auch konkret für die finalen Konsonantencluster in Wurzeln und Stämmen. Darüber hinaus konnte gezeigt werden, dass die Endungen (die ja definitiv Reime sind) häufig weniger kompositionell sind als Reime von Einsilbern.

Die minimalen Strukturen im Deutschen und Englischen stimmen ebenfalls weitestgehend überein:

- Stämme (etwas genauer: autosemantische Stämme) bestehen aus mindestens drei Graphemen, und ihr Reim ist mit mindestens zwei Graphemen besetzt (⟨VCC⟩, ⟨CVC⟩ oder ⟨CVV⟩).
- Sprachspezifisch sind in diesem Zusammenhang nur wenige Beschränkungen: Im Englischen verhält sich ⟨y⟩ wie zwei Grapheme (der Typ ⟨dry⟩ braucht kein weiteres Graphem im Kern oder Endrand). Im Deutschen sind einfaches ⟨l⟩, ⟨m⟩, ⟨n⟩ und ⟨r⟩ zu leicht; im Englischen sind es ⟨f⟩, ⟨l⟩ und ⟨s⟩.

- Affixe und Synsemantika sind leichter: Präfixe bestehen aus mindestens zwei Graphemen (⟨VC⟩ oder ⟨CV⟩); bei Suffixen gilt zusätzlich, dass der Reim zweifach besetzt sein muss (⟨VC⟩).
- Synsemantika unterscheiden sich im Englischen und Deutschen: Im Englischen gilt lediglich, dass der Kern einfach besetzt sein muss (⟨V⟩), im Deutschen darüber hinaus, dass die Struktur aus zwei Graphemen besteht (⟨VC⟩ oder ⟨CV⟩).

Das sind sehr weitgehende Konvergenzen, wenn man bedenkt, wie unterschiedlich konsistent die Schriftsysteme auf der segmentalen Ebene sind. Das deutsche Schriftsystem ist wesentlich näher an eineindeutigen Phonem-Graphem-Korrespondenzen als das englische, das ja gemeinhin als chaotisch oder – etwas neutraler – als stark idiosynkratisch gilt. Dass vor allem auf der suprasegmentalen Ebene so große Übereinstimmungen bestehen, ist überraschend.

Divergenzen finden sich vor allem, was die morphographische Schnittstelle anbelangt. Hier wurden mit der Einheitlichkeit und der Eindeutigkeit zwei Parameter angesetzt, die im weiteren Sinn auf das Form-Funktions-Verhältnis abheben. Ein Morphem ist einheitlich, wenn es nur eine graphematische Form gibt, die das Morphem repräsentiert. Und eine solche graphematische Form ist wiederum eindeutig, wenn jede Instanz dieser Form auf dasselbe Morphem verweist. Diese Parameter wurden für Stämme und Affixe im Englischen und Deutschen überprüft.

Es zeigt sich, dass im Deutschen die Einheitlichkeit von Stämmen wichtig ist, und zwar a) wichtiger als die Einheitlichkeit von Affixen und die Eindeutigkeit von Stämmen und Affixen im Deutschen und b) wichtiger als die Einheitlichkeit von Stämmen im Englischen. Die Einheitlichkeit von Stämmen ist im Deutschen ein entscheidendes Strukturprinzip. Das könnte einen Verarbeitungsvorteil haben – Stämme sind leichter wiedererkennbar, wenn sie formal nicht variieren. Ackermann/Zimmer (2017) zeigen empirisch, dass das zumindest für bestimmte Substantive im Deutschen gilt.

Im Englischen ist umgekehrt die Eindeutigkeit von Affixen entscheidend. Auch für diese morphographische Schreibung kann ein Verarbeitungsvorteil angenommen werden: Wenn eine bestimmte Schreibung immer ein Affix repräsentiert und nie ein zufälliges Wortende, dann liefert die Schrift den Lesern morphologische und morphosyntaktische Informationen auf dem Silbertablett: Die Leser sehen auf den ersten Blick, dass es sich bei einem bestimmten Teil eines graphematischen Wortes um grammatische und nicht um lexikalische Informationen handelt. In den meisten Fällen handelt es sich hier um Distinktivität: Eine Schreibung repräsentiert exklusiv morphologische Information und genau kein Wortende; um welches Affix es sich konkret handelt, ist aber oft nicht spezifi-

ziert. Eindeutigkeit im engeren Sinne, also der Verweis auf ein und nur ein Affix, ist die Ausnahme.

Auch für die Einheitlichkeit von Affixen gibt es im Englischen wesentlich weniger Ausnahmen als im Deutschen – und vor allem wird eine Reihe von morpho-phonologischen Alternationen graphematisch überbrückt (man denke nur an die Flexionssuffixe *-s* und *-ed* oder das Derivationsuffix *-ic*, die alle graphematisch einheitlicher sind als phonologisch).

Die Eindeutigkeit von Stämmen ist in beiden Sprachen zwar weit verbreitet; sie kann als der Normalfall gelten. Es gibt aber in beiden Sprachen einen nicht-ignorierbaren Teil an Ausnahmen – 3,3% homographischer Formen im Deutschen, 4,9% im Englischen. Wichtiger ist: Es werden im Englischen nur etwa die Hälfte aller homophonen Formen graphematisch differenziert (z. B. *meet/meat*); die andere Hälfte ist homographisch (z. B. *date₁/date₂*). Im Deutschen liegt dieser Anteil noch einmal niedriger, und zwar bei weniger als einem Fünftel aller Homophone. Es wird also längst nicht überall graphematisch differenziert, wo das lexikalisch möglich wäre.

Insgesamt zeigt sich, dass im Englischen eher Affixe verlässlich kodiert werden (sie sind oft eindeutig und einheitlich), während es im Deutschen Stämme sind, die häufig einheitlich kodiert werden. Das sind zwei grundsätzlich unterschiedliche Strategien der Leserleichterung. Im Deutschen wird der lexikalische Zugriff potenziell erleichtert, im Englischen der Aufbau grammatischer Strukturen (indem bspw. anhand eines Derivationsuffixes die Wortart ablesbar ist).

Welche Kriterien und Methoden eignen sich besonders für einen Vergleich von Schriftsystemen? Diese Frage wurde eingangs gestellt (siehe 1.1); sie soll hier kurz diskutiert werden. Die Voraussetzung ist sicherlich, dass sich für die Sprachen, die untersucht werden sollen, eine elektronische Liste der Stämme aufstellen lässt. Das ist die Minimalanforderung. Eine elektronische Datenbank, die auch phonologische und morphologische Informationen enthält, lässt weitere Untersuchungen zu; aber alleine mit einer Liste einfacher Stämme kann bereits eine Vielzahl von Informationen erhoben werden.

Das betrifft zum einen die quantitative Verteilung der Buchstaben. Welche Buchstaben sind häufig, welche selten? Das Deutsche und das Englische stimmen weitgehend überein, was die relativen Anteile der Buchstaben angeht (3.1); es wäre interessant zu wissen, ob das in anderen Sprachen ebenso ist und worauf sich die Gemeinsamkeiten oder Unterschiede zurückführen lassen. Es handelt sich hier um einen sehr groben Parameter – welche Buchstaben sind unter allen einfachen Stämmen wie häufig? – und dennoch lassen sich eventuell interessante Rückschlüsse ziehen.

Zweitens können die Buchstaben automatisch in Konsonanten und Vokale eingeteilt werden, und zwar automatisch auf der Grundlage ihrer syntagmati-

schen Verteilung. Dieses Verfahren ist etwas ausgefeilter als die reine quantitative Verteilung. Mit ihm lassen sich Aussagen über die Inventare von Vokalen und Konsonanten in der betreffenden Sprache machen, und diese Inventare können wiederum mit den etablierten (und aller Wahrscheinlichkeit nach phonographisch motivierten) Inventaren verglichen werden.

Drittens kann die Verteilung der Bigramme untersucht werden; auch dafür ist lediglich eine Liste mit der schriftlichen Form der Stämme notwendig. Wir können dann für jedes Bigramm feststellen, ob es wesentlich häufiger oder seltener auftritt, als das stochastisch zu erwarten wäre. Ein Sonderfall (allerdings ein sehr salienter) sind Verdoppelungen von Buchstaben: Welche Buchstaben können verdoppelt werden, welche nicht? Außerdem können mögliche und unmögliche Stammenden ermittelt werden.

Viertens kann die Liste der Stämme mit silbenstrukturellen Informationen angereichert werden, die wiederum auf der Einteilung in Vokale und Konsonanten beruhen. Mit diesen Informationen lassen sich nun sehr einfach Aussagen über die Anzahl der Schreibsilben sowie häufige und seltene CV-Strukturen machen. Außerdem kann die Besetzung der Silbenkonstituenten graphotaktisch beschrieben werden, und die Konstituenten können auf Interaktionseffekte untersucht werden.

Und fünftens kann schließlich überprüft werden, wie minimale Strukturen aussehen (gibt es eine Mindestlänge oder andere Mindestanforderungen für alle Stämme? Für die autosemantischen Stämme?) und wie Stämme typischerweise aussehen.

Für die Untersuchung der Einheitlichkeit und Eindeutigkeit sind dann morphologische und phonologische Informationen (und vor allem ein gewisses morphologisches Hintergrundwissen der jeweiligen Sprache) erforderlich. Dann kann aber für Stämme und Affixe festgestellt werden, wie einheitlich sie sind, und für jede graphematische Form, wie eindeutig bzw. distinktiv sie sind. Hier ist jeweils ein Vergleich mit der Phonologie instruktiv: Wo werden morpho-phonologische Alternationen überbrückt, wo nicht? Wo sind distinkte Lexeme homophon, aber heterographisch? Wo nicht?

So unterschiedlich die Untersuchungen sind, die in dieser Arbeit durchgeführt wurden – sie haben doch eines gemein: Sie beschreiben alle den synchronen Zustand des Schriftsystems. So kann expliziert werden, wie das System heute funktioniert; es bleibt aber unklar, wie die heutige Form des Systems sich herausgebildet hat.

Im Folgenden soll schlaglichtartig gezeigt werden, wie eine quantitative Untersuchung der diachronen Entwicklung aussehen könnte und was sie zu leisten vermag. Ausgangspunkt ist die Teiluntersuchung in Berg/Aronoff (2017), die sich mit der Herausbildung der Suffixschreibungen beschäftigt, und zwar für

dieselben vier Derivationsuffixe, deren Verteilung synchron ermittelt wurde (-ous, -ic, -al und -y). Die Datengrundlage bildete das relativ kleine diachrone Helsinki-Korpus (vgl. Helsinki Corpus of English Texts 1991), das nicht lemmatisiert ist und grobe Zeitangaben in 70-Jahres-Schnitten (bis 1710) enthält. Für jedes Suffix wurde mithilfe des OED die Menge der Schreibvarianten ermittelt (für -ic bspw. <ic>, <ik>, <ike>, <ique>, <icke>, <ick> etc.). Im Helsinki-Korpus wurden dann alle Wörter (i. S. v. Token, also laufende Textwörter) gesucht, die mit diesen Graphemfolgen enden (z. B. <panick>, <catholicke>), und für jedes dieser Wörter wurde überprüft, ob es heute mit <ic> endet. Daraus ergibt sich dann für jeden Zeitschnitt ein ‚Spektrum‘ der Schreibungen (z. B. 1350–1420: 65 % <ik>, 20 % <ic>, 10 % <ike> etc.).

In Abbildung 52 sind die relativen Anteile der Schreibvarianten über die Zeit aufgetragen (nach Berg/Aronoff 2017; für die Zeit ab 1710 wurde das wesentlich größere Korpus von Google Books verwendet, vgl. <https://books.google.com/ngrams>, Stand: 22. 11. 2018):

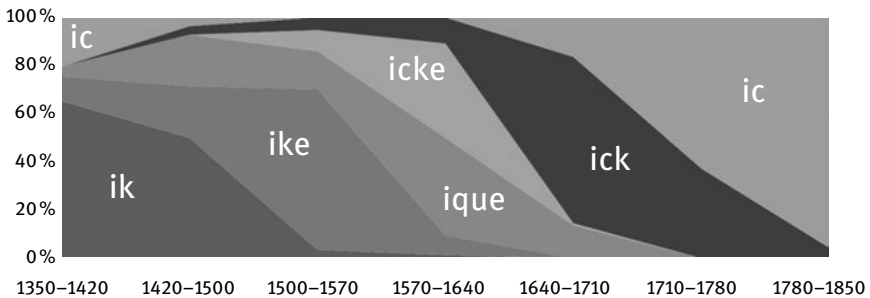


Abb. 52: Relativer Anteil der Schreibvarianten des Suffixes -ic bei Wörtern, die heute mit <ic> geschrieben werden. Datenbasis: Helsinki-Korpus (1350–1710) und Google Ngrams (1710–1850).

In beinahe jedem Zeitabschnitt ist eine andere Form dominant (<ik> > <ike> > <icke> > <ick> > <ic>). Die heutige Schreibung <ic> ist zwar im ersten Zeitschnitt bereits vertreten, verschwindet dann aber für ein paar Jahrhunderte und setzt sich erst 1710–1780 durch. Anders als bei anderen Suffixschreibungen, z. B. -ous, wird hier Variation nicht kontinuierlich abgebaut – es treten im Gegenteil zeitweise neue Varianten auf (<icke> z. B. erstmals 1500–1570).

Das Helsinki-Korpus ist relativ klein; es umfasst knapp 1,5 Millionen Token. Daher können viele interessante Fragen der Verteilung nicht mit diesem Korpus bearbeitet werden: Wie genau läuft der Übergang von einer zur anderen Variante ab? Ist er diatopisch zu motivieren, wird also eine regionale Variante in den Stan-

dard übernommen? Oder lassen sich ‚Leuchttürme‘ ausmachen – Schreiber, die eine bestimmte Variante verhältnismäßig früh verwenden und von denen wir wissen, dass sie einflussreich waren? Die Auflösung des Helsinki-Korpus ist einfach zu klein, um diese Fragen zu beantworten.

Hier bieten sich größere diachrone Korpora wie das Korpus ‚Early English Books Online‘ (EEBO, 1475–1700) und das Korpus ‚Eighteenth Century Collections Online‘ (ECCO, 1700–1799) an (www.textcreationpartnership.org, Stand: 22.11.2018); für das Deutsche ist das Deutsche Textarchiv einschlägig (www.deutschestextarchiv.de, Stand: 22.11.2018). Das EEBO-Korpus umfasst knapp 500 Millionen Token, das ECCO-Korpus 75 Millionen Token. Die zeitliche Auflösung ist viel höher, und zusätzlich liegen für jeden Text Informationen zum Autor, Verlag und Ort (sowie z. T. Drucker) vor, die für datengeleitete Analysen genutzt werden können.

Führt man die Untersuchung aus Berg/Aronoff (2017) in den EEBO- und ECCO-Korpora durch, so ergibt sich ein sehr ähnliches Bild wie im Helsinki-Korpus:

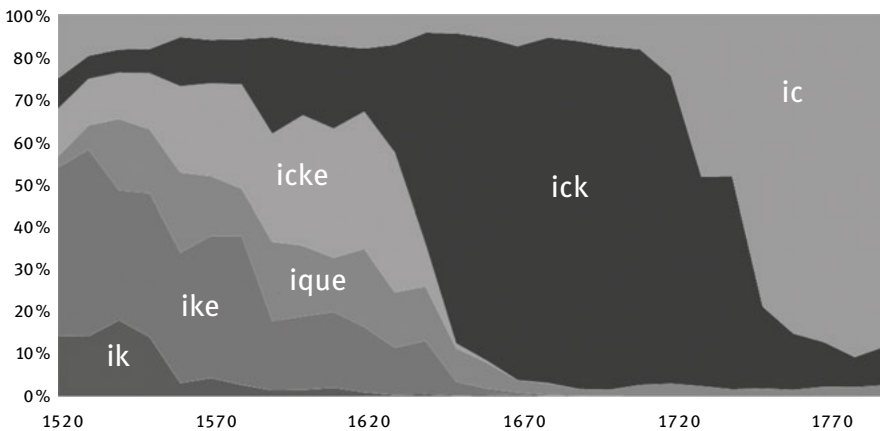


Abb. 53: Relativer Anteil der Schreibvarianten des Suffixes *-ic* bei Wörtern, die heute mit *ic* geschrieben werden. Datenbasis: EEBO-Korpus (bis 1700) und ECCO-Korpus (ab 1700).

Durch die höhere zeitliche Auflösung sind die Übergänge weniger glatt. Anders formuliert: Es können auch kleinere Schwankungen abgebildet werden. Ansonsten stimmen die Abbildung 52 und 53 überein, was die ungefähren Anteile der jeweiligen Varianten sowie die zeitliche Abfolge angeht. Das ist einerseits eine Bestätigung der Ergebnisse aus Berg/Aronoff (2017); andererseits ist es eine Ermunterung, auch mit kleineren diachronen Korpora zu arbeiten.

Wir können hier aber nun weitergehen und die Daten etwas genauer betrachten. Das geschieht in Abbildung 54, die den Übergang von <ick> zu <ic> zwischen 1650 und 1800 darstellt. Jeder Punkt im Streudiagramm ist ein Text, und die vertikale Achse zeigt an, wie hoch der Anteil der <ick>-Formen im jeweiligen Text ist.

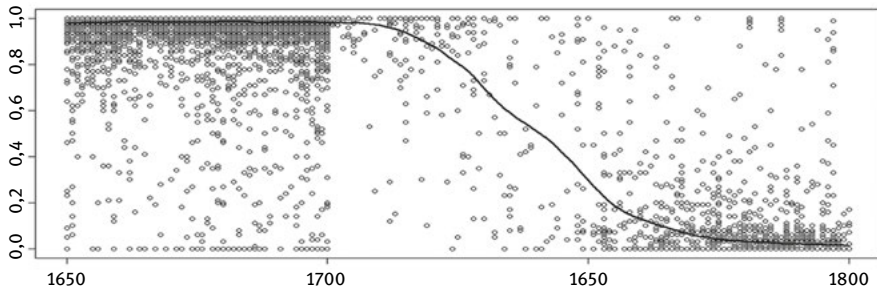


Abb. 54: Streudiagramm von Texten aus dem EEBO- und dem ECCO-Korpus. Horizontale Achse: Jahr der Veröffentlichung. Vertikale Achse: Relativer Anteil der <ick>-Formen im jeweiligen Text. Linie: Mittelwert der relativen Anteile pro Jahr.

Am Anfang ist <ick> absolut dominant, dann gibt es eine etwa 70-jährige Übergangsperiode, an dessen Ende <ic> dominiert (die Linie zeigt die Mittelwerte des Anteils von <ick> an allen Token eines Jahres an). Der gut sichtbare Bruch um 1700 ist die Grenze der beiden Korpora: Das EEBO-Korpus ist sehr viel dichter als das ECCO-Korpus, das 1700 beginnt. Auch wenn es auf den ersten Blick so aussieht, als ob die Variation schlagartig nachlässt – das ist wohl nur ein Effekt der geringeren Textdichte pro Jahr nach 1700. Wenn es mehr Text gäbe, gäbe es auch mehr Variation.

Die einzelnen Texte können nun danach klassifiziert werden, wo sie gedruckt und/oder veröffentlicht wurden oder wer sie geschrieben hat. Wir können außerdem von der Verteilung in Abbildung 54 ausgehen und die Autoren der Texte in graphematische *Innovatoren* und *Traditionalisten* einteilen. Innovatoren sind im konkreten Fall solche Schreiber, die <ic> verwenden, noch bevor die Mehrheit das tut (also vor etwa 1740). Traditionalisten sind andersherum solche Schreiber, die <ick> noch verwenden, nachdem die Mehrheit das nicht mehr tut (also nach ca. 1740). Und tatsächlich finden sich unter den Innovatoren viele einflussreiche Schreiber (z. B. der Theologe George Berkeley oder der Schriftsteller Alexander Pope, die zwischen 1710 und 1740 Texte veröffentlicht haben, in denen die neue Schreibung überwiegt). Umgekehrt stammen viele der Texte in ‚traditioneller‘ Orthographie im rechten oberen Bereich des Diagramms vom Schriftsteller Edmund Burke, einem der Vordenker des Konservatismus. Burke verwendet um

1800 noch eine Schreibung, die die meisten anderen Schreiber bereits knapp 60 Jahre vorher aufgegeben haben. Diese und andere Untersuchungen stehen noch am Anfang; es ist aber zu hoffen, dass mit einem umfangreichen diachronen Korpus und den richtigen Analysemethoden neue Erkenntnisse zum Ablauf graphematischer Wandelprozesse gewonnen werden können.

6 Literatur

- Ackermann, Tanja/Zimmer, Christian (2017): Morphologische Schemakonstanz – eine empirische Untersuchung zum funktionalen Vorteil nominalmorphologischer Wortschonung im Deutschen. In: Fuhrhop/Szczepaniak/Schmidt (Hg.). 145–174.
- Albrow, Kenneth H. (1972): *The English writing system: Notes towards a description*. London: Longman.
- Altmann, Gabriel/Lehfeldt, Werner (1980): *Einführung in die quantitative Phonologie*. (= *Quantitative Linguistics* 7). Bochum: Brockmeyer.
- Anderson, Stephen R. (1992): *A-morphous morphology*. (= *Cambridge Studies in Linguistics* 62). Cambridge, UK: Cambridge University Press.
- Anderson, Stephen R. (2015): The morpheme: Its nature and use. In: Baerman (Hg.). 11–33.
- Andrews, Sally (1989): Frequency and neighborhood size effects on lexical access: Activation or search? In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15. 802–814.
- Andrews, Sally (1992): Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18. 234–254.
- Aronoff, Mark (1976): *Word formation in generative grammar*. Cambridge, MA: MIT Press.
- Aronoff, Mark (1978): An English spelling convention. In: *Linguistic Inquiry* 9, 2. 299–303.
- Aronoff, Mark (1992): Segmentalism in linguistics: The alphabetic basis of phonological theory. In: Downing, Pamela/Lima, Susan D./Noonan, Michael (Hg.): *The linguistics of literacy*. (= *Typological Studies in Language* 21). Amsterdam u. a.: Benjamins. 71–82.
- Aronoff, Mark (1994): *Morphology by itself. Stems and inflectional classes*. Cambridge, MA: MIT Press.
- Aronoff, Mark (2016): Competition and the lexicon. In: Elia, Annibale/Iacobino, Claudio/Voghera, Miriam (Hg.): *Livelli di analisi e fenomeni di interfaccia. Atti del XLVII congresso internazionale della società di linguistica Italiana*. Rom: Bulzoni Editore. 39–52.
- Aronoff, Mark (2018): English verbs in syntactic structures. In: Hornstein, Norbert et al. (Hg.): *Syntactic structures after 60 years. The impact of the Chomskyan revolution in linguistics*. (= *Studies in Generative Grammar* 129). Berlin u. a.: De Gruyter. 381–402.
- Aronoff, Mark (eingereicht): *Competitors and alternants*. (Manuskript). Stony Brook: Stony Brook University.
- Aronoff, Mark/Fuhrhop, Nanna (2002): Restricting suffix combinations in German and English. In: *Natural Language and Linguistic Theory* 20. 451–490.
- Aronoff, Mark/Koch, Eric (1996): Context-sensitive regularities in English vowel spelling. In: *Reading and Writing* 8, 3. 251–265.
- Aronoff, Mark/Berg, Kristian/Heyer, Vera (2016): Some implications of English spelling for morphological processing. In: *The Mental Lexicon* 11. 164–185.
- Augst, Gerhard (1971): Über die Kombination von Phonemsequenzen bei Monemen. In: *Linguistische Berichte* 11. 37–47.
- Augst, Gerhard (1975): *Untersuchungen zum Morpheminventar der deutschen Gegenwortsprache*. Tübingen: Narr.
- Augst, Gerhard (1985): Dehnungs-h und Geminate in der graphematischen Struktur. In: Augst (Hg.). 112–121.

- Augst, Gerhard (Hg.) (1985): Graphematik und Orthographie. Neuere Forschungen der Linguistik, Psychologie und Didaktik in der Bundesrepublik Deutschland. (= Theorie und Vermittlung der Sprache 2). Frankfurt a. M.: Lang.
- Augst, Gerhard (1986): Zur Struktur komplexer Wörter. In: Zeitschrift für germanistische Linguistik 14, 3. 309–320.
- Baayen, R. Harald (2001): Word frequency distributions. (= Text, Speech and Language Technology 18). Dordrecht u. a.: Kluwer.
- Baayen, R. Harald/Milin, Petar/Ramscar, Michael (2016): Frequency in lexical processing. In: Aphasiology 30, 11. 1174–1220.
- Baayen, R. Harald/Piepenbrock, Richard/Gulikers, Leon (1995): The CELEX lexical database (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.
- Baerman, Matthew (Hg.) (2015): The Oxford handbook of inflection. Oxford: Oxford University Press.
- Baroni, Marco (2009): Distributions in text. In: Lüdeling, Anke/Kytö, Merja (Hg.): Corpus linguistics: An international handbook. Bd. 2. (= Handbücher zur Sprach- und Kommunikationswissenschaft 29). Berlin u. a.: De Gruyter. 803–821.
- Basbøll, Hans/Wagner, Johannes (1985): Kontrastive Phonologie des Deutschen und Dänischen: Segmentale Wortphonologie und -phonetik. (= Linguistische Arbeiten 160). Berlin/Boston: De Gruyter.
- Bauer, Laurie (2003): Introducing linguistic morphology. 2. Aufl. Edinburgh: Edinburgh University Press.
- Bauer, Laurie/Huddleston, Rodney (2002): Lexical word-formation. In: Huddleston, Rodney/Pullum, Geoffrey K. (Hg.): The Cambridge grammar of the English language. Cambridge, UK: Cambridge University Press. 1621–1722.
- Bauer, Laurie/Lieber, Rochelle/Plag, Ingo (2013): The Oxford reference guide to English morphology. Oxford: Oxford University Press.
- Becker, Henrik (1953): Zur monophonematischen Wertung. In: Zeitschrift für Phonetik und Allgemeine Sprachwissenschaft 7, 3. 253–258.
- Berg, Kristian (2012): Identifying graphematic units: Vowel and consonant letters. In: Written Language and Literacy 15, 1. 26–45.
- Berg, Kristian (2013): Graphemic alternations in English as a reflex of morphological structure. In: Morphology 23, 4. 387–408.
- Berg, Kristian (2016): Double consonants in English: Graphemic, morphological, prosodic and etymological determinants. In: Reading and Writing 29, 3. 453–474.
- Berg, Kristian (2017): Sichtbare Flexionsmorphologie im Englischen und Deutschen. In: Fuhrhop/Szczepaniak/Schmidt (Hg.). 9–40.
- Berg, Kristian/Aronoff, Mark (2017): Self-organization in the spelling of English suffixes: The emergence of culture out of anarchy. In: Language 93. 37–64.
- Berg, Kristian/Fuhrhop, Nanna (2011): Komplexe Silbenkernschreibungen im Englischen im Vergleich mit dem Deutschen. In: Linguistische Berichte 228. 443–466.
- Berg, Kristian/Primus, Beatrice/Wagner, Lutz (2016): Buchstabenmerkmal, Buchstabe, Graphem. In: Primus, Beatrice/Domahs, Ulrike (Hg.): Handbuch Laut, Gebärde, Buchstabe. (= Handbücher Sprachwissen 2). Berlin u. a.: De Gruyter. 337–355.
- Berg, Kristian et al. (2014): Morphological spellings in English. In: Written Language and Literacy 17, 2. 282–307.
- Best, Karl-Heinz (2005): Zur Häufigkeit von Buchstaben, Leerzeichen und anderen Schriftzeichen in deutschen Texten. In: Glottometrics 11. 9–31.

- Bierwisch, Manfred (1972): Schriftstruktur und Phonologie. In: *Probleme und Ergebnisse der Psychologie* 43. 21–44.
- Bittner, Dagmar (2002): Semantisches in der pronominalen Flexion des Deutschen. *Zeitschrift für Sprachwissenschaft* 21, 2. 196–233.
- Blevins, James P. (2006): English inflection and derivation. In: Aarts, Bas/McMahon, April (Hg.): *The handbook of English linguistics*. Malden, MA u. a.: Blackwell. 507–536.
- Blevins, James P. (2013): Word-based morphology from Aristotle to modern WP. In: Allan, Keith (Hg.): *The Oxford handbook of the history of linguistics*. Oxford: Oxford University Press. 375–396.
- Blevins, Juliette (2010): Syllable structure: The limits of variation, by San Duanmu. In: *Linguistic Typology* 14. 287–291.
- Bloomfield, Leonard (1933): *Language*. Chicago: University of Chicago Press.
- Bodmer, Frederick (1944): *Loom of Language*. New York: Norton.
- Bortz, Jürgen/Schuster, Christof (2010): *Statistik für Human- und Sozialwissenschaftler*. 7. Aufl. Heidelberg: Springer.
- Bowers, Jeffrey/Davis, Colin/Hanley, Derek (2005): Interfering neighbours: The impact of novel word learning on the identification of visually similar words. In: *Cognition* 97, 3. B45–B54.
- Brauß, Ursula (1994): *Lexikalische Funktionen der Synsemantika*. (= *Forschungsberichte des Instituts für deutsche Sprache* 71). Tübingen: Narr.
- Bredel, Ursula/Noack, Christina/Plag, Ingo (2013): Morphologie lesen: Stammkonstanzschreibung und Leseverstehen bei starken und schwachen Lesern. In: Neef, Martin/Scherer, Carmen (Hg.): *Die Schnittstelle von Morphologie und geschriebener Sprache*. (= *Linguistische Arbeiten* 551). Berlin u. a.: De Gruyter. 211–249.
- Brekke, Herbert E. (1994): Some thoughts on a historico-genetic theory of the lettershapes of our alphabet. In: Watt, William C. (Hg.): *Writing systems and cognition: Perspectives from psychology, physiology, linguistics, and semiotics, neuropsychology and cognition*. Dordrecht: Kluwer. 129–139.
- Bubenhof, Noah/Hansen-Morath, Sandra/Konopka, Marek (2014): Korpusbasierte Exploration der Variation der nominalen Genitivmarkierung. In: *Zeitschrift für Germanistische Linguistik* 42, 3. 379–419.
- Buchmann, Franziska (2015): *Die Wortzeichen im Deutschen*. (= *Germanistische Bibliothek* 56). Heidelberg: Winter.
- Büchner, Georg (1879): *Sämtliche Werke und handschriftlicher Nachlaß*. Frankfurt a. M.: Sauerländer. www.deutschestextarchiv.de/book/show/buechner_werke_1879 (Stand: 22. 11. 2018).
- Burnage, Gavin (1995): *CELEX. A guide for users*. Nijmegen: Center for lexical information.
- Butt, Matthias/Eisenberg, Peter (1990): Schreibsilbe und Sprechsilbe. In: Stetter, Christian (Hg.): *Zu einer Theorie der Orthographie. Interdisziplinäre Aspekte gegenwärtiger Schrift- und Orthographieforschung*. (= *Germanistische Linguistik* 99). Tübingen: Niemeyer. 33–64.
- Carney, Edward (1994): *A survey of English spelling*. New York: Routledge.
- Chomsky, Carol (1970): Reading, writing, and phonology. *Harvard educational review* 40, 2. 287–309.
- Chomsky, Noam (1957): *Syntactic structures*. (= *Janua linguarum* 4). S'-Gravenhage: Mouton.
- Chomsky, Noam/Halle, Morris (1968): *The sound patterns of English*. New York: Harper & Row.
- Coates, Richard (2006): Morphophonemics. In: Brown, Keith (Hg.): *Encyclopedia of language and linguistics*. Bd. 8: Mel-N. 2. Aufl. Amsterdam u. a.: Elsevier. 318–331.

- Coltheart, Max et al. (1977): Access to the internal lexicon. In: Dornič, Stansilav (Hg.): Attention and performance VI: Proceedings of the sixth international symposium on attention and performance, Stockholm, July 28 – August 1, 1975. London: Academic Press. 535–555.
- Cook, Vivian/Ryan, Des (Hg.) (2016): The Routledge handbook of the English writing system. New York: Routledge.
- Culicover, Peter W./Jackendoff, Ray (2005): *Simpler Syntax*. Oxford: Oxford University Press.
- Cummings, Donald Wayne (1988): *American English spelling: An informal description*. Baltimore: Johns Hopkins University Press.
- De Saussure, Ferdinand (1967): *Grundfragen der allgemeinen Sprachwissenschaft*. 2. Aufl. Berlin: De Gruyter.
- Dewey, Godfrey (1970): *Relative frequency of English spellings*. Columbia: Teachers College Press.
- Dobson, Eric J. (1968): *English Pronunciation 1500–1700*. 2. Aufl. Oxford: Clarendon Press.
- Dryer, Matthew S. (2005): Descriptive theories, explanatory theories, and basic linguistic theory. In: Ameka, Felix K./Dench, Alan/Evans, Nicholas (Hg.): *Catching language: The standing challenge of grammar writing*. (= Trends in Linguistics. Studies and Monographs 167). Berlin u. a.: De Gruyter. 207–234.
- Duanmu, San (2008): *Syllable structure: The limits of variation*. Oxford: Oxford University Press.
- Duden (2011): *Der Duden in zwölf Bänden*. Bd. 9: *Richtiges und gutes Deutsch. Das Wörterbuch der sprachlichen Zweifelsfälle*. 7., vollst. überarb. Aufl. Mannheim u. a.: Dudenverlag.
- Duden (2016): *Der Duden in zwölf Bänden*. Bd. 4: *Die Grammatik*. 9., vollst. überarb. u. aktual. Aufl. Berlin: Dudenverlag.
- Dürscheid, Christa (2012): *Einführung in die Schriftlinguistik*. 4., überarb. und aktual. Aufl. Göttingen: Vandenhoeck und Ruprecht.
- Eisenberg, Peter (1981): Substantiv oder Eigenname? Über die Prinzipien unserer Regeln zur Groß- und Kleinschreibung. In: *Linguistische Berichte* 72. 77–101.
- Eisenberg, Peter (1983): Orthografie und Schriftsystem. In: Günther, Klaus-Burkhard (Hg.): *Schrift, Schreiben, Schriftlichkeit: Arbeiten zur Struktur, Funktion und Entwicklung schriftlicher Sprache*. (= Germanistische Linguistik 49). Tübingen: Niemeyer. 41–68.
- Eisenberg, Peter (1985): Graphemtheorie und phonologisches Prinzip. Vom Sinn eines autonomen Graphembegriffs. In: Augst (Hg.). 122–128.
- Eisenberg, Peter (1988): Die Grapheme des Deutschen und ihre Beziehung zu den Phonemen. In: Baurmann, Jürgen/Günther, Klaus-B./Knoop, Ulrich (Hg.): *Aspekte von Schrift und Schriftlichkeit* (= Germanistische Linguistik 93/94). Hildesheim u. a.: Olms. 139–154.
- Eisenberg, Peter (2011): *Das Fremdwort im Deutschen*. Berlin u. a.: De Gruyter.
- Eisenberg, Peter (2013a): *Grundriss der deutschen Grammatik*. Bd. 1: *Das Wort*. 4., aktual. u. überarb. Aufl. Stuttgart: Metzler.
- Eisenberg, Peter (2013b): *Grundriss der deutschen Grammatik*. Bd. 2: *Der Satz*. 4., aktual. u. überarb. Aufl. Stuttgart: Metzler.
- Enderle, Ursula (2005): *Autonomie der geschriebenen Sprache? Zur Theorie phonographischer Beschreibungskategorien am Beispiel des Deutschen*. (= *Philologische Studien und Quellen* 188). Berlin: Schmidt.
- Erben, Johannes (2006): *Einführung in die deutsche Wortbildungslehre*. 5., durchges. u. erg. Aufl. (= *Grundlagen der Germanistik* 17). Berlin: Schmidt.
- Evertz, Martin (2014): *Visual prosody: The graphematic foot in English and German*. Dissertation an der Philosophischen Fakultät der Universität zu Köln. (Manuskript). Köln: Universität Köln.

- Firth, John Rupert (1968): *Selected papers of J. R. Firth, 1952–59*. London: Longman.
- Fleischer, Wolfgang/Barz, Irmhild (2012): *Wortbildung der deutschen Gegenwartssprache*. 4., völlig neu bearb. Aufl. Berlin u. a.: De Gruyter.
- Fleming, Edward/Johnson, Stephanie (2007): Rosa's roses: Reduced vowels in American English. In: *Journal of the International Phonetic Association* 37, 1. 83–96.
- Fudge, Erik C. (1969): Syllables. In: *Journal of Linguistics* 5. 253–286.
- Fuhrhop, Nanna (1998): *Grenzfälle morphologischer Einheiten*. (= *Studien zur deutschen Grammatik* 57). Tübingen: Stauffenburg.
- Fuhrhop, Nanna (2000): Zeigen Fugenelemente die Morphologisierung von Komposita? In: Thieroff et al. (Hg.). 201–213.
- Fuhrhop, Nanna (2008): Das graphematische Wort (im Deutschen): Eine erste Annäherung. In: *Zeitschrift für Sprachwissenschaft* 27. 189–228.
- Fuhrhop, Nanna (2015): *Orthografie*. 4., aktual. Aufl. Heidelberg: Winter.
- Fuhrhop, Nanna (2017): Sichtbare Morphologie in der Flexion der starken und unregelmäßigen Verben im Deutschen und Englischen. In: Fuhrhop/Szczepaniak/Schmidt (Hg.). 41–73.
- Fuhrhop, Nanna/Barghorn, Rebecca (2012): Prinzipien der Wortschreibung im Deutschen und Englischen am Beispiel der Schreibdiphthonge und der Doppelkonsonanten. In: Gunkel, Lutz/Zifonun, Gisela (Hg.): *Deutsch im Sprachvergleich. Grammatische Kontraste und Konvergenzen*. (= *Jahrbuch des Instituts für Deutsche Sprache* 2011). Berlin u. a.: De Gruyter. 135–160.
- Fuhrhop, Nanna/Berg, Kristian (eingereicht): *Schreibdiphthonge und graphematische Silbenkerne – modalitätsspezifisch und modalitätsübergreifend*. (Manuskript). Oldenburg: Universität Oldenburg.
- Fuhrhop, Nanna/Buchmann, Franziska (2009): Die Längenhierarchie. Zum Bau der graphematischen Silbe. In: *Linguistische Berichte* 218. 127–155.
- Fuhrhop, Nanna/Buchmann, Franziska (2016): Graphematische Silbe. In: Primus, Beatrice/Domahs, Ulrike (Hg.): *Handbuch Laut, Gebärde, Buchstabe*. (= *Handbücher Sprachwissen* 2). Berlin: De Gruyter. 356–376.
- Fuhrhop, Nanna/Kürschner, Sebastian (2015): Linking elements in Germanic. In: Müller, Peter O. et al. (Hg.): *Word-formation. An international handbook of the languages of Europe*. Bd. 1. (= *Handbücher zur Sprach- und Kommunikationswissenschaft* 40). Berlin u. a.: De Gruyter. 568–582.
- Fuhrhop, Nanna/Peters, Jörg (2013): *Einführung in die Phonologie und Graphematik*. Stuttgart: Metzler.
- Fuhrhop, Nanna/Teuber, Oliver (2000): Das Partizip 1 als adjektivischer Infinitiv. In: Bittner, Andreas/Bittner, Dagmar/Köpcke, Klaus-Michael (Hg.): *Angemessene Strukturen: Systemorganisation in Phonologie, Morphologie und Syntax*. Hildesheim u. a.: Olms. 173–190.
- Fuhrhop, Nanna/Buchmann, Franziska/Berg, Kristian (2011): The length hierarchy and the graphematic syllable. Evidence from German and English. In: *Written Language and Literacy* 14, 2. 275–292.
- Fuhrhop, Nanna/Szczepaniak, Renata/Schmidt, Karsten (Hg.) (2017): *Sichtbare und hörbare Morphologie*. (= *Linguistische Arbeiten* 565). Berlin u. a.: De Gruyter.
- Fushing, Hsieh et al. (2014): Lewis Carroll's doublets net of English words: Network heterogeneity in a complex system. In: *PLoS ONE* 9, 12. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0114177> (Stand: 22. 11. 2018).

- Gallmann, Peter (1990): Wortschreibung und Schemakonstanz. In: *Zeitschrift für Germanistik* 11, 5. 513–523.
- Gallmann, Peter (1999): Wortbegriff und Nomen-Verb-Verbindungen. In: *Zeitschrift für Sprachwissenschaft* 18, 1. 269–304.
- Gallmann, Peter/Sitta, Horst (1996): Die Neuregelung der deutschen Rechtschreibung. Regeln, Kommentar und Verzeichnis wichtiger Neuschreibungen. (= Duden Taschenbücher 26). Mannheim u. a.: Dudenverlag.
- Garbe, Burckhard (1985): Graphemtheorien und mögliche Strukturmodelle zur Beschreibung der Orthographie. In: Augst (Hg.). 1–21.
- Giegerich, Heinz (1992): *English phonology: An introduction*. Cambridge, UK: Cambridge University Press.
- Giegerich, Heinz (1999): *Lexical strata in English: Morphological causes, phonological effects*. Cambridge, UK: Cambridge University Press.
- Gilliéron, Jules (1918): *Genéalogie des mots qui designent l'abeille*. Paris: Champion.
- Gleason, Henry A. (1961): *An introduction to descriptive linguistics*. 2. Aufl. New York u. a.: Holt, Rinehart and Winston.
- Glick, James (2011): *The information: A history, a theory, a flood*. New York: Pantheon.
- Günther, Hartmut (1988): *Schriftliche Sprache. Strukturen geschriebener Wörter und ihre Verarbeitung beim Lesen*. (= Konzepte der Sprach- und Literaturwissenschaft 40). Tübingen: Niemeyer.
- Günther, Hartmut (2004): *Schriftsysteme*. In: Booij, Geert et al. (Hg.): *Morphologie. Ein internationales Handbuch zur Flexion und Wortbildung*. 2. Halbband. New York: De Gruyter. 1915–1923.
- Günther, Hartmut/Greese, Barbara (1985): *Lexical hermits and the pronunciation of visually presented words*. In: *Forschungsberichte des Instituts für Phonetik und sprachliche Kommunikation der Universität München* 21. 25–52.
- Güthert, Kerstin (2005): *Herausbildung von Norm und Usus Scribendi im Bereich der Worttrennung am Zeilenende (1500–1800)*. (= Germanistische Bibliothek 24). Heidelberg: Winter.
- Guy, Jacques (1991): *Vowel identification: An old (but good) algorithm*. In: *Cryptologia* 3. 258–262.
- Hanna, Paul R. et al. (1966): *Phoneme-grapheme correspondences as cues to spelling improvement*. Washington, DC: U. S. Department of Health, Education, and Welfare, Office of Education.
- Harris, Zelig (1942): *Morpheme alternants in linguistic analysis*. In: *Language* 18, 3. 169–180.
- Harris, Zelig (1951): *Structural linguistics*. Chicago: University of Chicago Press.
- Harnisch, Rüdiger (2001): *Grundform- und Stamm-Prinzip in der Substantivmorphologie des Deutschen. Synchronische und diachronische Untersuchungen eines typologischen Parameters*. (= Germanistische Bibliothek 10). Heidelberg: Winter.
- Harweg, Roland (1971): *Buchstabe und Graphem*. In: *Linguistische Berichte* 13. 78–80.
- Hay, Jennifer/Baayen, R. Harald (2003): *Phonotactics, parsing and productivity*. In: *Rivista di Linguistica* 15, 1. 99–130.
- Hayes, Bruce (1989): *Compensatory lengthening in moraic phonology*. In: *Linguistic inquiry* 20, 2. 253–306.
- Heidolph, Karl Erich/Flämig, Walter/Motsch, Wolfgang (Hg.) (1984): *Grundzüge einer deutschen Grammatik*. 2., unveränd. Aufl. Berlin: Akademie-Verlag.

- Helsinki Corpus of English Texts (1991): Department of Modern Languages, University of Helsinki. Compiled by Matti Rissanen (Project leader), Merja Kytö (Project secretary); Leena Kahlas-Tarkka, Matti Kilpiö (Old English); Saara Nevanlinna, Irma Taavitsainen (Middle English); Terttu Nevalainen, Helena Raumolin-Brunberg (Early Modern English). www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus (Stand: 22. 11. 2018).
- Hockett, Charles F. (1961): Linguistic elements and their relations. In: *Language* 37, 1. 29–53.
- Hornby, Albert Sydney (1974): *Oxford advanced learner's dictionary*. London: Oxford University Press.
- Horobin, Simon (2013): *Does spelling matter?* Oxford: Oxford University Press.
- Huddleston, Rodney (2002): The verb. In: Huddleston/Pullum (Hg.). 71–212.
- Huddleston, Rodney/Pullum, Geoffrey K. (Hg.) (2002): *The Cambridge grammar of the English language*. Cambridge, UK: Cambridge University Press.
- Jackendoff, Ray (1997): *The architecture of the language faculty*. Cambridge, MA: MIT Press.
- Jackendoff, Ray (2002): *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.
- Jacobs, Joachim (2007): Vom (Un-)Sinn der Schreibvarianten. In: *Zeitschrift für Sprachwissenschaft* 26. 43–80.
- Jespersen, Otto (1928): *A modern English grammar: On historical principles*. Bd. 1: Sounds and spelling. 4. Aufl. Heidelberg: Winter.
- Kaplan, Abby (2011): How much homophony is normal? In: *Journal of linguistics* 47. 631–671.
- Katz, Leonard/Frost, Ram (1992): Reading in different orthographies: The orthographic depth hypothesis. In: Frost, Ram/Katz, Leonard (Hg.): *Orthography, phonology morphology, and meaning*. Amsterdam: Elsevier. 67–84.
- Keller, Rudi (1994): *Sprachwandel. Von der unsichtbaren Hand in der Sprache*. 2., überarb. u. erw. Aufl. Tübingen u. a.: Francke.
- Kessler, Brett/Treiman, Rebecca (2001): Relationships between sounds and letters in English monosyllables. In: *Journal of Memory and Language* 44, 4. 592–617.
- King, Robert D. (1967): Functional load and sound change. In: *Language* 43, 4. 831–852.
- Kohrt, Manfred (1985a): Morphem- und Silbengrenzen in der deutschen Orthographie. In: Augst (Hg.). 64–104.
- Kohrt, Manfred (1985b): *Problemgeschichte des Graphembegriffs und des frühen Phonembegriffs*. (= Germanistische Linguistik 61). Berlin u. a.: De Gruyter.
- Kolvenbach, Monika (1980): Das morphologische Lexikon (MOLEX) des Systems Plidis. In: *Mitteilungen des Instituts für deutsche Sprache* 7. 60–66.
- Köpcke, Klaus-Michael (1988): Schemas in German Plural Formation. In: *Lingua* 74. 303–335.
- Labov, William 1994: *Principles of linguistic change. Internal factors*. Oxford: Blackwell.
- Ladefoged, Peter/Maddieson, Ian (1996): *The sounds of the world's languages*. Oxford: Blackwell.
- Langenscheidt (2002): *Langenscheidt Taschenwörterbuch Englisch*. Hrsg. v. Helmut Willmann, Gisela Türck, Heinz Messinger und Langenscheidt-Redaktion. Berlin u. a.: Langenscheidt.
- Langenscheidt (2005): *Langenscheidt Handwörterbuch Englisch*. Hrsg. v. Heinz Messinger, Martin Feller Mayer und Langenscheidt-Redaktion. Berlin u. a.: Langenscheidt.
- Lehfeldt, Werner (2005): Phonemdistribution. In: Köhler, Reinhard/Altmann, Gabriel/Piotrowski, Rajmund G. (Hg.): *Quantitative Linguistik. Ein internationales Handbuch*. (= Handbücher zur Sprach- und Kommunikationswissenschaft 27). Berlin u. a.: De Gruyter. 181–190.

- Lenders, Winfried (2013): Computational lexicography and corpus linguistics until ca. 1970/1980. In: Gouws, Rufus H. et al. (Hg.): *Dictionaries. An international encyclopedia of lexicography. Supplementary volume: Recent developments with focus on electronic and computational lexicography.* (= *Handbücher zur Sprach- und Kommunikationswissenschaft* 5.4). Berlin u. a.: De Gruyter. 982–1000.
- Lieber, Rochelle (2009): IE, Germanic: English. In: Lieber, Rochelle/Štekauer, Pavol (Hg.): *The Oxford handbook of compounding.* Oxford: Oxford University Press. 357–369.
- Lipka, Leonhard (1986): Homonymie, Polysemie oder Ableitung im heutigen Englisch. In: *Anglistik und Amerikanistik* 34. 128–138.
- Lyons, John (1968): *Introduction to theoretical linguistics.* London: Cambridge University Press.
- Lüdtke, Helmut (1969): Die Alphabetschrift und das Problem der Lautsegmentierung. In: *Phonetica* 20. 147–176.
- Maas, Utz (1992): *Grundzüge der deutschen Orthographie.* (= Reihe germanistische Linguistik 120). Tübingen: Niemeyer.
- Maas, Utz (2000): *Orthographie. Materialien zu einem erklärenden Handbuch zur Rechtschreibung des Deutschen.* Osnabrück: Buchhandlung zur Heide.
- Mann, Thomas (1901): *Buddenbrooks. Verfall einer Familie.* Berlin: Fischer.
- Marchand, Hans (1969): *The categories and types of present-day English word-formation: A synchronic-diachronic approach.* 2., vollst. überarb. u. erg. Aufl. München: Beck.
- Matthews, Peter H. (1991): *Morphology.* 2. Aufl. Cambridge, UK: Cambridge University Press.
- Mayerthaler, Willi (1981): *Morphologische Natürlichkeit.* (= *Linguistische Forschungen* 8). Frankfurt: Athenaeon.
- Meier, Helmut (1964): *Deutsche Sprachstatistik.* Hildesheim: Olms.
- Meinhold, Gottfried/Stock, Eberhard (1982): *Phonologie der deutschen Gegenwartssprache.* Leipzig: Bibliographisches Institut.
- Meisenburg, Trudel (1998): Zur Typologie von Alphabetschriftsystemen anhand des Parameters der Tiefe. In: *Linguistische Berichte* 173. 43–64.
- Moscoso del Prado Martín, Fermín (2016): Vocabulary, grammar, sex, and aging. In: *Cognitive Science* 41, 4. 950–975.
- Moscoso del Prado Martín, Fermín/Kostić, Aleksandar/Baayen, R. Harald (2004): Putting the bits together: An information-theoretical perspective on morphological processing. In: *Cognition* 94, 1. 1–18.
- Moulton, William G. (1956): Syllabic nuclei and final consonant clusters in German. In: Halle, Morris/Lunt, Horace/McLean, Hugh (Hg.): *For Roman Jakobson: Essays on the occasion of his sixtieth birthday, 11 october 1956.* Den Haag: Mouton. 372–381.
- Mugdan, Joachim (1994): Morphological units. In: Asher, Ronald E. (Hg.): *The encyclopedia of language and linguistics.* Oxford u. a.: Pergamon Press. 2543–2553.
- Neef, Martin (1996): *Worddesign. Eine deklarative Analyse der deutschen Verbflexion.* (= *Studien zur deutschen Grammatik* 52). Tübingen: Stauffenburg.
- Neef, Martin (2005): *Die Graphematik des Deutschen.* (= *Linguistische Arbeiten* 500). Tübingen: Niemeyer.
- Nerius, Dieter (Hg.) (2007): *Deutsche Orthographie.* 4., neu bearb. Aufl. Hildesheim u. a.: Olms.
- Nespor, Marina/Vogel, Irene (1986): *Prosodic phonology.* Dordrecht: Foris Publications.
- Nida, Eugene (1949): *Morphology: The descriptive analysis of words.* 2. Aufl. Ann Arbor: University of Michigan Press.

- Nübling, Damaris/Szczepaniak, Renata (2008): On the way from morphology to phonology: German linking elements and the role of the phonological word. In: *Morphology* 18, 1. 1–25.
- Nunberg, Geoffrey/Briscoe, Ted/Huddleston, Rodney (2002): Punctuation. In: Huddleston/Pullum (Hg.). 1723–1764.
- Palmer, Frank/Huddleston, Rodney/Pullum, Geoffrey K. (2002): Inflectional morphology and related matters. In: Huddleston/Pullum (Hg.). 1565–1620.
- Plag, Ingo (1999): Morphological productivity: Structural constraints in English derivation. (= *Topics in English Linguistics* 28). Berlin u. a.: De Gruyter.
- Plag, Ingo/Homann, Julia/Kunter, Gero (2017): Homophony and morphology: The acoustics of word-final S in English. In: *Journal of Linguistics* 53, 1. 181–216.
- Plank, Frans (1981): Morphologische (Ir-)Regularitäten. Aspekte der Wortstrukturtheorie. (= *Studien zur deutschen Grammatik* 13). Tübingen: Narr.
- Primus, Beatrice (2000): Das Schriftsystem des Deutschen. Teil 1: Die Graphematik. (Manuskript). Köln: Universität Köln.
- Primus, Beatrice (2003): Zum Silbenbegriff in der Schrift-, Laut- und Gebärdensprache – Versuch einer mediumübergreifenden Fundierung. In: *Zeitschrift für Sprachwissenschaft* 23, 1. 3–55.
- Primus, Beatrice (2004): A featural analysis of the Modern Roman Alphabet. In: *Written Language and Literacy* 7, 2. 235–274.
- Primus, Beatrice (2006): Buchstabenkomponenten und ihre Grammatik. In: Bredel, Ursula/Günther, Hartmut (Hg.): *Orthographietheorie und Rechtschreibunterricht*. (= *Linguistische Arbeiten* 509). Tübingen: Niemeyer. 5–43.
- Primus, Beatrice (2010): Strukturelle Grundlagen des deutschen Schriftsystems. In: Bredel, Ursula/Müller, Astrid/Hinney, Gabriele (Hg.): *Schriftsystem und Schriffterwerb: linguistisch – didaktisch – empirisch*. (= *Germanistische Linguistik* 589). Tübingen: Niemeyer. 9–45.
- Procter, Paul (1978): *Longman dictionary of contemporary English. The up-to-date learning dictionary*. Harlow u. a.: Longman.
- Raffelsiefen, Renate (1995): Conditions for stability: The case of schwa in German. (= *Theorie des Lexikons* 69). Wuppertal u. a.: *Arbeiten des Sonderforschungsbereichs* 282.
- Rahnenführer, Ilse (1980): Zu den Prinzipien der Schreibung des Deutschen. In: Nerius, Dieter/Scharnhorst, Jürgen (Hg.): *Theoretische Probleme der deutschen Orthographie*. (= *Sprache und Gesellschaft* 16). Berlin: Akademie-Verlag. 231–259.
- Ramers, Karl-Heinz (1998): Minimale Wörter: „Prosodische“ Beschränkungen graphischer Wortstrukturen. In: Kröger, Bernd J./Riek, Christine/Sachse, Georg (Hg.): *Festschrift für Georg Heike*. Frankfurt a. M.: Hector. 25–39.
- Rastle, Kathleen/Davis, Matthew H./New, Boris (2004): The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. In: *Psychonomic Bulletin and Review* 11. 1090–1098.
- Rayner, Keith et al. (2012): *Psychology of reading*. 2. Aufl. New York: Psychology Press.
- Rezec, Oliver (2009): *Zur Struktur des deutschen Schriftsystems*. Dissertation, LMU München.
- Roemheld, Friedrich (1955): Die Längenbezeichnung in der deutschen Rechtschreibung. In: *Deutscherunterricht* 7, 3. 71–82.
- Rollings, Andrew (2004): *The spelling patterns of English*. München: Lincom Europa.
- Ryan, Des (2016): *Principles of English spelling formation*. Unveröffentlichte Diss. Trinity College, Dublin.

- Sampson, Geoffrey (1985): *Writing systems: a linguistic introduction*. London u. a.: Hutchinson.
- Schmidt, Karsten (o. J.): Ist ä: in Fremdwörtern ein Ärgernis für das native System? (Manuskript). Oldenburg: Universität Oldenburg.
- Schmidt, Karsten (2014): Morphophonographic regularities in German: The graphematic syllable boundary. A non-linear graphematic approach. In: *Written Language and Literacy* 17, 2. 253–281.
- Shannon, Claude E. (1948): A mathematical theory of communication. In: *Bell System Technical Journal* 27, 3. 379–592.
- Siegel, Dorothy (1979): *Topics in English morphology*. New York: Garland.
- Sproat, Richard (2000): *A computational theory of writing systems*. Cambridge, UK: Cambridge University Press.
- Sproat, Richard (2016): English among the writing systems of the world. In: Cook/Ryan (Hg.). 27–40.
- Stetter, Christian (2005): *System und Performanz. Symboltheoretische Grundlage von Medientheorien und Sprachwissenschaft*. Weilerswist: Vellbrück.
- Taft, Marcus (1979): Lexical access via an orthographic code: The basic orthographic syllabic structure (BOSS). In: *Journal of verbal learning and verbal behavior* 18, 1. 21–39.
- Taft, Marcus/Forster, Kenneth I. (1975): Lexical storage and retrieval of prefixed words. In: *Journal of verbal learning and verbal behavior* 14. 638–647.
- Thieroff, Rolf et al. (Hg.) (2000): *Deutsche Grammatik in Theorie und Praxis*. Tübingen: Niemeyer.
- Tomaschek, Fabian/Berg, Kristian (Manuskript): Orthographic representation and phonetic characteristics of unstressed German vowels: The case of Opa vs. Oper. Tübingen: Universität Tübingen.
- Treiman, Rebecca (1986): The division between onsets and rimes in English syllables. In: *Journal of Memory and Language* 25, 4. 476–491.
- Trommer, Jochen/Zimmermann, Eva (2015): Inflectional exponence. In: Baerman (Hg.). 47–86.
- Venezky, Richard L. (1970): *The structure of English orthography*. Den Haag u. a.: Mouton.
- Venezky, Richard L. (1999): *The American way of spelling: The structure and origins of American English orthography*. New York: Guilford Press.
- Venezky, Richard L. (2004): In search of the perfect orthography. In: *Written Language and Literacy* 7. 139–163.
- Vennemann, Theo (1982): Zur Silbenstruktur der deutschen Standardsprache. In: Vennemann, Theo (Hg.): *Silben, Segmente, Akzente: Referate zur Wort-, Satz- und Versphonologie anlässlich der vierten Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft*, Köln, 2.–4. März 1982. Tübingen: Niemeyer. 261–305.
- Vennemann, Theo (1988): Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin. Berlin u. a.: De Gruyter.
- Wagner, Arthur (1982): *Das Minimalpaar*. Hamburg: Buske.
- Wedel Andrew/Kaplan, Abby/Jackson, Scott (2013): High functional load inhibits phonological contrast loss: A corpus study. In: *Cognition* 128. 179–186.
- Werner, Otmar (1972): *Phonemik des Deutschen*. Stuttgart: Wichmann.
- Wiese, Bernd (1996): Iconicity and syncretism. On pronominal inflection in Modern German. In: Sackmann, Robin/Budde, Monika (Hg.): *Theoretical linguistics and grammatical description. Papers in honour of Hans-Heinrich Lieb on the occasion of his 60th birthday*.

- (= Amsterdam Studies in the Theory and History of Linguistic Science 4). Amsterdam u. a.: Benjamins. 323–344.
- Wiese, Richard (2000): The phonology of German. Oxford: Oxford University Press.
- Wurzel, Wolfgang Ullrich (2000): Was ist ein Wort? In: Thieroff et al. (Hg.). 29–42.
- Yule, Valerie/Yasuko, Ishi (2016): Spelling reform. In: Cook/Ryan (Hg.). 413–426.
- Zifonun, Gisela/Hoffmann, Ludger/Strecker, Bruno (1997): Grammatik der deutschen Sprache. 3 Bde. (= Schriften des Instituts für Deutsche Sprache 7). Berlin u. a.: De Gruyter.
- Zwicky, Arnold M (1987): Suppressing the Zs. In: Journal of Linguistics 23. 133–148.

7 Anhang

Anhang A: Liste der untersuchten Affixe

Deutsch (nach Duden 2016: 702, 719, 734 f., 762 f., 773)

Präfixe:

a-, an-, anti-, be-, de-, des-, durch-, ent-, er-, erz-, ge-, hinter-, im-, in-, inter-, kon-, miss-, prä-, re-, trans-, über-, ultra-, um-, un-, unter-, ur-, ver-, wider-, zer-

Suffixe:

-abel, -age, -al, -ant, -anz, -ar, -är, -at, -bar, -chen, -e, -ei, -el, -ell, -em, -en, -end, -ent, -enz, -er, -erie, -ern, -es, -est, -eur, -fach, -haft, -halber, -heit, -ie, -ier, -ig, -ik, -in, -ion, -isch, -ismus, -ist, -ität, -iv, -keit, -lein, -ler, -lich, -ling, -los, -mäßig, -n, -ner, -nis, -oid, -or, -os, -ös, -s, -sam, -schaft, -st, -t, -tum, -ung, -ur, -wärts, -weg, -wegen, -weise

Englisch (nach Huddleston/Pullum (Hg.) 2002: 1677 ff.)

Präfixe:

a-, after-, anti-, be-, counter-, dis-, em-, en-, ex-, fore-, im-, in-, inter-, ir-, micro-, mid-, non-, out-, over-, post-, pre-, sub-, super-, trans-, un-, under-

Suffixe:

-able, -age, -al, -an, -ance, -ant, -ar, -ate, -atic, -cy, -dom, -ed, -ee, -eer, -en, -ence, -ent, -eous, -er, -ery, -ese, -ess, -est, -et, -ette, -ety, -ful, -hood, -i, -ial, -ian, -ible, -ic, -ie, -ify, -ine, -ing, -ion, -ious, -ish, -ism, -ist, -ite, -ity, -ive, -ize, -less, -let, -like, -ling, -ly, -ment, -ness, -or, -ory, -ous, -ry, -s, -'s, -ship, -some, -ster, -th, -ty, -ual, -ure, -y

Anhang B: Absolute Häufigkeiten der Buchstaben im Korpus

Absolute Häufigkeiten der Buchstaben in den CELEX-Teilkorpora einfacher deutscher Stämme und einfacher englischer Stämme.

Buchstabe	Frequenz de	Frequenz en
a	2.371	3.477
ä	167	
b	785	943
c	1.041	1.763
d	659	1.111
e	4.034	4.045
f	624	658
g	751	984
h	1.432	1.277
i	1.744	2.254
j	88	116
k	1.176	608
l	1.974	2.376
m	1.110	1.246
n	1.819	2.146
o	1.150	2.623
ö	104	
p	965	1.352
q	34	128
r	2.386	2.652
s	1.971	2.238
ß	45	
t	2.031	2.448
u	1.082	1.525
ü	201	
v	162	377
w	341	524
x	48	108
y	73	389
z	404	151

8 Sachregister

- Affix Stripping 188
- Allomorphe 18, 215, 216, 217, 270, 274, 279
- Anglizismus 91, 223, 224, 234, 291

- Bigramm 43
- Bindestrich 213

- Clusteranalyse 35, 38, 40, 41, 42, 43, 54, 68
- CV-Strukturen 74

- Dehnungs-⟨h⟩ 140, 155
- Dendrogramm 35, 39
- Derivationsstammform 237, 238, 241, 242
- deutsche Präfixe
 - *be-* 134, 200, 202, 281, 284
 - *er-* 155, 215, 281, 284
 - *in-* 270
 - *kon-* 270
 - *un-* 272, 284
- deutsche Suffixe
 - *-abel* 268, 269, 270, 271
 - *-al* 268, 270, 281, 283
 - *-ant* 270, 271, 281, 283, 285
 - *-ante* 270, 271
 - *-anz* 270, 271
 - *-är* 238, 268, 269, 270, 272, 281, 283
 - *-ation* 238, 270
 - *-chen* 237
 - *-e* 237
 - *-e* 1.Sg. 263
 - *-ei* 198, 202, 270, 271
 - *-el* 237
 - *-ell* 198, 238, 268, 270, 281, 283
 - *-(e)n* INFINITIV/1.Ps.PL/3.Ps.PL 260
 - *-(e)n* PLURAL 259
 - *-er* 236, 237, 272, 282, 285
 - *-(e)s* GENITIV 261
 - *-ess* 268
 - *-esse* 268
 - *-(e)st* 2.Ps.Sg 261
 - *-(e)st* SUPERLATIV 262
 - *-(e)t* 3.Ps.Sg/2.Ps.Pl/PRÄT 263
 - *-ett* 268
 - *-ette* 268
 - *-haft* 235, 237
- *-heit* 235, 237, 270
- *-ier* 236
- *-ifizier* 238
- *-ig* 16, 18, 19, 197, 200, 214, 235, 236, 237, 272, 281, 282, 283
- *-in* 237, 267, 268, 281, 282, 285
- *-isch* 235, 237
- *-ismus* 268, 270
- *-iv* 200, 270, 271, 272, 281, 284
- *-lein* 237
- *-lich* 18, 92, 197, 237, 270
- *-los* 235, 237
- *-nis* 268, 269, 296
- *-or* 270
- *-ös* 200, 238, 268, 269, 270, 281, 284
- *-schaft* 235, 237
- *-tum* 237
- *-ung* 236
- *-weise* 235, 237
- Differenzierung von Homonymen 210
- distributionell eingeschränkt 47

- Eindeutigkeit 217
- Einheitlichkeit 217
- Endung 110, 187
- englische Präfixe
 - *con-* 273
 - *in-* 273
 - *inter-* 281
 - *un-* 289
- englische Suffixe
 - *-able* 196, 240, 241, 272, 273, 274
 - *-age* 194, 240
 - *-al* 197, 200, 240, 280, 286, 287, 301
 - *-an* 240
 - *-ance* 18, 240, 241, 274
 - *-ant* 240, 241, 273, 274
 - *-ary* 240, 287
 - *-ate* 194, 240, 241, 272, 273, 275
 - *-ation* 240, 273, 275, 288
 - *-ce* 273, 274
 - *-dom* 203, 240
 - *-ed* 211, 240, 259, 266, 279, 299
 - *-ed* PRÄT./PART. 266
 - *-er* 240, 273, 275, 288

- *-ery* 196, 272, 273, 287
- *-(e)s PLURAL/3.Ps.Sg.* 265
- *-fold* 240
- *-ful(A)* 240
- *-ful(N)* 240
- *-hood* 197, 203, 240
- *-ic* 16, 65, 200, 241, 244, 273, 275, 280, 286, 293, 299, 301, 302
- *-ion* 288
- *-ise* 196, 273, 274, 275, 288, 289
- *-ive* 289
- *-less* 198, 203, 240, 286
- *-like* 240
- *-ly* 240, 242
- *-ment* 203, 240
- *-ness* 198, 203, 240, 286
- *-ory* 194, 196, 273, 287
- *-our* 272, 273, 288
- *-ous* 122, 203, 204, 219, 240, 241, 272, 273, 280, 286, 301
- *-ship* 240, 281
- *-some* 240
- *-type* 240
- *-wise* 240
- *-y* 272, 280, 287
- Entropie 146
- ⟨e⟩-Tilgung 229, 230, 231, 232, 239, 241, 242, 243, 245, 269, 272, 291

- Frequenzspektrum 49
- Fugenelement 235, 246
- funktionale Last 164

- gebundener Stamm 226, 238
- grammatische Muster 212
- graphematische Integration 222
- Graphotaktik
 - der intervokalischen Konsonanten 111
 - des Anfangsrandes 87, 195
 - des Endrands 101, 196
 - des Kerns 122
 - des Silbenkerns 198

- Heterographie 248, 249, 252
- Homographie 230, 248, 292
- Homonymieprinzip 210

- Interaktion
 - zwischen Anfangsrand und Endrand 126
 - zwischen Anfangsrand und Kern 124
 - zwischen Kern und Endrand 128
- Inventar
 - der Buchstaben 24
 - der Grapheme 33, 81, 107, 186, 191
 - ⟨i⟩/⟨y⟩-Alternation 212, 239, 240, 243, 291

- Konversion 233

- Latinismus 226, 227, 238, 242
- Lexem 10

- Minimalitätsbeschränkungen
 - für Affixe 201
 - für Stämme und Wurzeln 159
- Morphem 23
- Morphemdefinition 214
- morphologische Schreibungen 209

- Netzwerk 180
- nichtsilbische Affixe 192

- Parameter der Tiefe 3
- prototypische Strukturen
 - von Affixen 204
 - von Stämmen 163
 - von Stämmen und Affixen im Vergleich 206
- Pseudosuffix 111

- Silbe 70
- Silbengewicht 138, 156, 198, 207, 297
- Silbenzahl pro Stamm 70
- Stammflexion 226, 227, 238, 242, 259
- Stammkonstanz 7, 20, 209, 210, 211, 220, 228, 250, 277, 278, 283
- Stammvokal 235, 237
- Symmetrie 78, 107

- Umlaut 222, 235, 237, 238, 270

- Verdoppelung 57, 61, 93, 231, 237, 239

- Wurzel 110, 154, 155, 159, 187, 231