

Produktivität in der deutschen Derivationsmorphologie

- DISSERTATION -
zur Erlangung des akademischen Grades
doctor philosophiae

vorgelegt beim
Fachbereich Germanistik der
Fakultät für Linguistik und Literaturwissenschaft
der Universität Bielefeld
in Bielefeld

von

Karina Schneider-Wiejowski

Erstgutachterin : Prof. Dr. Barbara Job
Zweitgutachter : Prof. Dr. Wolfgang Klein
externer Gutachter : Prof. Dr. Stefan Evert

Juli 2011

*»Ihrer wahren Wesensbestimmung nach ist die
Wissenschaft das Studium der Schönheit der Welt...«*

Simone Weil (1909-43)

Danksagung



„Die Dankbarkeit ist am besten und effektivsten, wenn sie nicht in leeren Phrasen verdampft“¹, sagte schon Isaac Asimov und dem schließe ich mich an.

An erster Stelle danke ich natürlich Barbara Job, Wolfgang Klein und Stefan Evert für die zahlreichen wertvollen Tipps, die sie mir im Laufe der zwei Jahre gegeben haben.

Ein weiterer Dank muss den vielen Studentinnen und Studenten der Germanistik in Bielefeld gelten. Zusammen haben wir empirisch geforscht und diskutiert, aber auch gelacht und viele sprachwissenschaftliche Fragen beantwortet.

Und als letztes danke ich meiner Familie, meinen Freunden und dem wundervollsten Ehemann dieser Welt.

¹ Foundation and Empire, Granada Publishing Limited, S. 86.

Inhaltsverzeichnis

1	Einleitung	1
2	Wortbildung im Deutschen	5
2.1	Funktion der Wortbildung	7
2.2	Einheiten der Wortbildung	10
2.3	Triebkräfte der Wortschatzerweiterung	11
2.4	Regeln und Einschränkungen der Wortschatzerweiterung	16
2.4.1	Phonologische Beschränkungen	18
2.4.2	Morphologische Beschränkungen	18
2.4.3	Syntaktische Beschränkungen	19
2.4.4	Semantische Beschränkungen	19
2.4.5	Lexikalische Beschränkungen	20
2.4.6	Pragmatische Produktivitätsbeschränkungen	21
2.4.7	Fazit	22
2.5	Konkurrenz zwischen Derivationsaffixen	23
3	Grammatikalisierung und Wortbildungswandel	26
3.1	Grammatikalisierungstheorie	27
3.2	Morphologischer Wandel	32
3.3	Stand der Forschung in der (germanistischen) Linguistik	38
3.4	Arbeitshypothesen	40
4	Produktivität in der Morphologie	43
4.1	Geschichte der morphologischen Produktivität	43
4.2	Die jüngere Produktivitätsforschung	45

Inhaltsverzeichnis

4.3	Wahrscheinlichkeitsmodelle	48
4.3.1	Baayens Produktivitätsmaß P	49
4.3.2	Kritik an P	51
4.3.3	Globale Produktivität	53
4.3.4	Kritik und Diskussion von P*	55
4.4	Psycholinguistische Aspekte im Bezug auf morphologische Produktivität	57
4.5	Fazit	60
5	Die empirischen Studien	61
5.1	Der Untersuchungsgegenstand	61
5.1.1	Das Suffix -ung	62
5.1.2	Das Suffix -nis	64
5.1.3	Die Allomorphe -heit und -keit	65
5.1.4	Das Suffix -sal	68
5.1.5	Das Suffix -tum	69
5.1.6	Das Suffix -bar	70
5.1.7	Das Suffix -sam	72
5.1.8	Das Suffix -haft	73
5.1.9	Das Suffix -lich	74
5.1.10	Das Suffix -ig	75
5.1.11	Das Suffix -mäßig	76
5.1.12	Das Suffix -isch	77
5.1.13	Das Suffix -los	79
5.1.14	Zusammenfassung	79
5.2	Datengrundlage	81
5.3	Exkurs: Deutsche Korpora	81
5.3.1	Das Durchführen einer Korpusstudie	83
5.3.2	Verfügbare freie Korpora für das Deutsche und ihre Probleme	84
5.4	Untersuchungszeitraum	88
5.5	Methode der Auswertung	89
6	Das Vorexperiment	91
6.1	Versuchsbeschreibung	91
6.1.1	Untersuchungsgegenstand	92
6.1.2	Versuchspersonen	92

6.1.3	Methode der Auswertung	92
6.2	Ergebnisse	92
6.2.1	Gesamtanzahl der Bildungen	92
6.2.2	Typenhäufigkeit	94
6.2.3	Andere Überlegungen	96
6.2.4	Lexikalisiertheit	99
6.3	Zusammenfassung	102
7	Ergebnisse der Korpusstudie mit dem DWDS	105
7.1	Die substantivbildenden Suffixe	105
7.1.1	Tokenhäufigkeit	105
7.1.2	Typenhäufigkeit	108
7.1.3	Typen-Token-Relationen	110
7.1.4	Hapax Legomena	111
7.1.5	Baayens P	113
7.1.6	P*	115
7.2	Die adjektivbildenden Suffixe	127
7.2.1	Tokenhäufigkeit	127
7.2.2	Typenhäufigkeit	129
7.2.3	Typen-Token-Relationen	130
7.2.4	Hapax Legomena	131
7.2.5	Baayens P	132
7.2.6	P*	134
7.3	Fazit I	147
7.3.1	Die substantivbildenden Suffixe aus synchroner Sicht	148
7.3.2	Die adjektivbildenden Suffixe aus synchroner Sicht	151
7.4	Fazit II	153
7.4.1	Die substantivbildenden Suffixe aus diachroner Sicht	154
7.4.2	Die adjektivbildenden Suffixe aus diachroner Sicht	155
8	Methodologische Diskussion	157
8.1	LNRE-Modelle	164
8.1.1	Vergleich des Zipf-Mandelbrot-, des finiten Zipf-Mandelbrot- und des GIGP-Modells	175
8.2	Normierung der Stichprobengröße	177

Inhaltsverzeichnis

8.3	Probleme mit dem Modell	187
8.4	Exemplarische Wachstumskurven der substantivbildenden Suffixe . . .	188
8.5	Exemplarische Wachstumskurven der adjektivbildenden Suffixe . . .	193
8.6	Exemplarische Wachstumskurven zur diachronen Betrachtung morphologischer Produktivität	197
9	Desiderata	201

Einleitung

„Denn eben wo Begriffe fehlen, da stellt ein Wort zur rechten Zeit sich ein“, heißt es in Goethes Faust¹ und es bietet sich deshalb an, diese Aussage einleitend für die vorliegende Arbeit zu verwenden, da sie sich mit dem Prozess der Wortbildung beschäftigt, der dafür zuständig ist, neue Wörter zu schaffen: Derivation. Dabei steht ein ganz bestimmtes Phänomen von derivationellen Bildungen im Fokus der wissenschaftlichen Betrachtung: ihre **Produktivität**.

Unter Produktivität wird die Eigenschaft eines Wortbildungsaffixes verstanden, neue, noch nicht lexikalisierte Formen zu produzieren. Im Gegensatz zum Phänomen der Wortbildung selbst, die sowohl in der Synchronie als auch in ihrem Wandel eingehend untersucht worden ist, blieb die Produktivität bisher noch weitgehend unberücksichtigt. Das gilt insbesondere für korpusbasierte, empirische Untersuchungen. Diese Lücke soll mit dieser Arbeit geschlossen werden.²

¹Faust 1, Vers 1995 f.

²Dass es sich dabei durchaus um eine wichtige Forschungsaufgabe handelt, verdeutlicht u.a. Erben (2000, 151):

„Es bleibt eine lohnende Aufgabe, das Vordringen produktiver Typen der Wortbildung in bestimmten Textsorten zu untersuchen, ihre syntaktische und stilistische Wirksamkeit sowie ihren Anteil beim Aus- und Aufbau bestimmter lexikalischer Paradigmen (Wortfelder), die der Sprachgemeinschaft oder bestimmten sozialen Gruppen kommunikationswichtig werden. Der jeweils verfügbare, herkömmliche Wortschatz kann ein Hemmnis der Weiterbildung und Nutzung aller Bildungsmöglichkeiten sein [...], aber auch Grund und Möglichkeit für treffende Neubildungen, welche die bisherige Ausdrucksskala genauer werden lassen.“

Die vorliegende Arbeit ist sowohl dem Bereich der **germanistischen Linguistik** wie auch der **quantitativen Morphologie** zuzuordnen. Die quantitative Morphologie befasst sich mit den morphologischen Einheiten, ihren Eigenschaften und Prozessen und verfolgt zudem das Ziel, Erklärungen für bestimmte Häufigkeiten und statistische Verteilungen zu finden.

In Kapitel 2 dieser Arbeit wird der Bereich der Wortbildung zunächst theoretisch eingeleitet. Skizziert werden die **Funktionen** und die **Einheiten** der Wortbildung, zudem sollen **Triebkräfte** (Produktivität, Kreativität, Analogie), die zur Wortschatzerweiterung beitragen, beschrieben werden. Des Weiteren ist es notwendig, **Regeln und Restriktionen** von Wortbildungsmodellen zu erläutern, welche die Produktivität einschränken oder antreiben können.

Im Anschluss daran folgt mit Kapitel 3 ein weiterer theoretischer Abschnitt, in welchem zunächst die **Grammatikalisierungstheorie** vorgestellt wird. Mit dieser Theorie wird der Weg eines sprachlichen Zeichens von der Lexik in die Grammatik beschrieben. Derivationsaffixe können aus freien Morphemen entstehen: auf ihrem Weg zum grammatischen Morphem gewinnen sie – im Zuge ihrer Grammatikalisierung – an Abstraktheit und damit auch an Produktivität. Letztendlich ist es allerdings möglich, dass die Produktivität eines Affixes im Laufe der Zeit wieder abnimmt und am Ende des unidirektionalen Weges nur ein morphologischer Rest übrig bleibt.

Das dritte Kapitel versteht sich zudem als Zusammenfassung der bisherigen Veröffentlichungen zum **morphologischen Wandel**. Es wird deutlich, dass der morphologische Wandel nicht nur im Zusammenhang mit dem Wandel von Derivationsaffixen auftaucht, sondern beispielsweise auch der flektivische Wandel als so betitelt wird. Aus diesem Grund bietet es sich an, den Begriff **Wortbildungswandel** einzuführen; insbesondere ist es im letzten Schritt notwendig, den Wortbildungswandel und die derivationelle Produktivität in einen Zusammenhang zu stellen. Aus diesem Zusammenhang können letztlich Arbeitshypothesen abgeleitet werden, welche mithilfe von empirischen Studien überprüft werden sollen.

Kapitel 4 stellt das zentrale Kapitel zur Produktivität und dem **Produktivitätsbegriff in der Morphologie** dar. Nach Einleitung der Thematik durch einen kurzen Überblick zur Geschichte der Produktivität werden mathematische Verfahren zur Produktivitätsmessung vorgestellt, die in den letzten zwanzig Jahren an Popularität gewonnen haben und als 'klassische Methoden' zur Messung von Produktivität bezeichnet werden dürfen.

Im fünften Kapitel werden die Rahmenbedingungen für die empirische Studie präsentiert. Zunächst erfolgt eine Beschreibung des Untersuchungsgegenstandes, eine Skizzierung der 13 Wortbildungssuffixe. Daraufhin wird der Untersuchungszeitraum angegeben, der sich aus den Möglichkeiten ergibt, die das Kernkorpus des **DWDS** (Digitale Wörterbuch der deutschen Sprache) zur Verfügung stellt. In diesem Kontext erfolgt auch ein kurzer Exkurs zur **Korpuslinguistik** des Gegenwartsdeutschen. Des Weiteren werden Aussagen der linguistischen Literatur hinsichtlich der Produktivität der zu untersuchenden Suffixe tabellarisch gegenüber gestellt. Die Auflistung zeigt, dass sich die Autoren generell einig darüber sind, welche Suffixe produktiv sind und welche Suffixe kaum zur Erweiterung des Lexikons beitragen; gleichzeitig wird an dieser Stelle jedoch deutlich, dass es sich nur um **kategoriale Etikettierungen** handelt, die gemacht werden ('produktiv', 'aktiv', 'hoch produktiv', 'beschränkt aktiv' etc.) und bisher keine empirische Überprüfung mit anschließender Bestätigung oder Validierung der Thesen stattgefunden hat.

Um einen ersten Eindruck über die Produktivität der Suffixe zu gewinnen, wurde ein **Vorexperiment** durchgeführt. Sechs Muttersprachler wurden aufgefordert, so viele Derivate wie möglich mit den entsprechenden Affixen zu bilden. Die Auswertung der Ergebnisse zeigt, dass es deutliche Unterschiede zwischen den einzelnen Suffixen gibt: so z.B. existieren Suffixe, bei denen die Sprecher mehr oder weniger unterschiedliche Bildungen (Typen) produzieren, die lexikalisiert sind, andererseits liegen auch Suffixe des Deutschen vor, mit denen vordergründig nicht-lexikalisierte Bildungen genannt werden.

Im siebten Kapitel werden die **Ergebnisse der Korpusstudie** mit dem DWDS vorgestellt. Angewandt wurden die klassischen Methoden zur Messung der Produktivität, die zuvor in Kapitel 4 erläutert worden sind. Dazu zählen die Typen-Token-Relation und die Baayensischen Maße P und P^* . Die Produktivität im engeren Sinne (P) wird rechnerisch ermittelt, indem jeweils alle einmaligen Bildungen eines Wortbildungsmodells (*Hapax Legomena*) in Relation zu der gesamten Anzahl an Token mit einem Modell gesetzt werden und P^* , die globale Produktivität, ist ein zweidimensionales Maß, bei welchem man die vorab berechneten P -Werte in Relation zu der Anzahl der Typen setzt und die Wertepaare in einem Diagramm verortet.

Bei Betrachtung der Ergebnisse fällt auf, dass die Anwendung der klassischen Maße nicht zufriedenstellend ist, um die Produktivität der Suffixe adäquat zu beschreiben. Das Hauptproblem ergibt sich aus den nicht vergleichbaren Datenmengen der

einzelnen Suffixe, denn bei zu geringen Datenmengen entstehen unter anderem hohe P -Werte, die mit Vorsicht zu betrachten sind.

Dieses, aber auch andere Probleme wie Zufallsschwankungen und Lexikalisiertheit, welche sich insbesondere in hohen Token- und Typenfrequenzen einiger Suffixe erkennbar macht, sollen im achten Kapitel detailliert diskutiert werden. Ziel der Diskussion ist es, eine bisher kaum beachtete Möglichkeit zur Produktivitätsmessung vorzustellen und anschließend exemplarisch anzuwenden: das **finite Zipf-Mandelbrot-Modell** (fZM). Dieses Modell gehört zu den so genannten LNRE-Modellen (*large number of rare events*) und wurde bisher nicht mit Daten des deutschen Derivationsystems getestet, sodass die Anwendung eine Novität darstellt. Mit diesem Modell ist es z.B. möglich, Datenmengen zu vergrößern, um für alle Suffixe die gleiche Anzahl an Token zur Berechnung von P zur Verfügung zu haben. Des Weiteren lassen sich Wachstumskurven von Typen und Hapax Legomena grafisch darstellen, was eine qualitative Analyse ermöglicht, indem die Verläufe der Kurven interpretiert und auch miteinander verglichen werden können.

Kapitel 9 bildet den Abschluss der Arbeit, indem die gewonnenen Erkenntnisse zusammengefasst und diskutiert werden. Insbesondere gilt es hier aufzuzeigen, welche Wünsche seitens der Linguisten an die Mathematik herangetragen werden sollten, um das fZM-Modell zu verbessern und weiterzuentwickeln.

Demnach versteht sich diese Arbeit a) als empirische Studie zur Überprüfung der kategorialen Aussagen der linguistischen Literatur in Bezug auf die Produktivitätseigenschaften der Wortbildungssuffixe und b) als allgemeine Diskussion zum Phänomen der morphologischen Produktivität.

Wortbildung im Deutschen

In dieser Arbeit werden native Derivationsuffixe untersucht. Einleitend ist es deshalb notwendig, einige morphologische Begrifflichkeiten zu erklären, um dieser Arbeit einen theoretischen Rahmen zu geben.

Die linguistische Disziplin der Morphologie wird in die beiden großen Teildisziplinen Flexion und Wortbildung¹ eingeteilt. Durch Flexionsmorpheme werden grammatische Beziehungen zwischen Wortformen ausgedrückt. Flexionsmorpheme verändern die lexikalische Bedeutung der Basis nicht, sondern fügen Informationen hinzu, die die Basis satzfähig machen.

Der Bereich der Wortbildung lässt sich in die Teilgebiete Komposition und Derivation² aufteilen. Die Komposition ist ein Prozess, bei dem schon vorhandenes Wortmaterial zu einem neuen Wort zusammengesetzt wird. Dieses Phänomen der Wortbildung soll aber im weiteren Verlauf der Arbeit nicht untersucht werden.

¹Vgl. dazu auch Fleischer und Barz (1995, 5), die darauf aufmerksam machen, dass die Wortbildung deutlich von der Wortschöpfung abzugrenzen ist. Den Unterschied erklären sie damit, dass bei der Wortbildung bereits vorhandenes Material verwendet ist, was bei der Wortschöpfung nicht der Fall ist. Bei letzterer werden Lautkomplexe geschaffen, die noch nicht als bedeutungstragende Elemente in der jeweiligen Sprache existieren, sodass sich von einer Entstehung neuer Wortwurzeln sprechen lässt. Beispiele für Wortschöpfungen wird man vor allem in der Kindersprache (*Wauwau* für Hund, *Töfföff* für Motorrad) und im Bereich der Science Fiction finden. Ebenso können Onomatopoeica, schallnachahmende Geräusche (*Ticktack* für das Ticken einer Uhr, *peng* oder *boing* für einen Knall), in vielen Fällen mit dem Prinzip der Wortschöpfung erklärt werden.

²Zur systematischen Beschreibung der Wortbildungsart Derivation siehe Beard (1998).

Die Derivation ist dafür verantwortlich, dass neue Lexeme geschaffen werden. Von der Komposition unterscheidet sich die Derivation dadurch, dass die morphematischen Konstruktionen, aus denen ein neues Wort geschaffen wird, nicht aus mehreren freien Morphemen bestehen, die eine eigenständige (lexikalische) Bedeutung besitzen, sondern nur eine Basis existiert, an die ein (oder auch mehrere) Affixe angehängt werden. Die Hauptfunktion der Derivation besteht in der Erweiterung des Lexikons.³

Im weiteren Verlauf der Arbeit wird nun wie folgt vorgegangen: zunächst einmal soll deutlich gemacht werden, welche Funktion der Wortbildung zuzuschreiben ist; anschließend werden in Kürze die Einheiten der Wortbildung skizziert.

Danach erfolgt die Annäherung an das zentrale Thema dieser Arbeit, indem die Triebkräfte der Wortschatzerweiterung, zu denen – neben Kreativität und Analogie – auch die Produktivität zählt, erläutert werden.

Im darauf folgenden Abschnitt werden Regeln und Restriktionen von Wortbildungsmodellen beschrieben. Gerade jene Beschränkungen, denen ein Affix unterliegen kann, sind im Hinblick auf seine Produktivität von Bedeutung, da eine Vielzahl an Beschränkungen zur Abschwächung oder sogar zum Untergang der Produktivität führen kann. Ebenso ist es möglich, dass ein Affix an Produktivität verliert, weil sich ein konkurrierendes Wortbildungsmodell durchsetzen kann, was eine Verdrängung des schwächeren Modells bewirkt. Die Konkurrenz von derivationellen Affixen bildet den Abschluss dieses Kapitels.

³Neologismen müssen aber nicht zwangsweise durch eine kompositionelle oder derivationelle Struktur gebildet werden. Ebenfalls können Lehnwörter oder -elemente in den Wortschatz eingeführt werden. Auch wenn das Englische heute am auffälligsten dazu beiträgt, Wörter in den deutschen Wortschatz einzuschleusen, so existiert im Gegenwartsdeutschen auch eine Reihe an Wörtern, die aus unterschiedlichen Sprachen eingeführt worden sind: das Wort *Kajal*, das aus dem Indischen stammt, *Safari* aus der arabischen Sprache oder aber *Ketchup* aus dem Chinesischen. Auch *Lehnelemente*, Konfixe, können in den deutschen Wortschatz aufgenommen werden. Viele von ihnen stammen dabei aus der griechischen (*bio-*) oder lateinischen Sprache (*Omni-*), im Gegenwartsdeutschen jedoch findet man zahlreiche neue Lehnelemente aus dem Gegenwartsenglischen wie z.B. *-burger*. Lehnelemente werden in dieser Arbeit nicht untersucht, allerdings kann die Methodik der Produktivitätsmessung auch auf Lehnelemente angewandt werden. Auch in diesem Bereich besteht eine Forschungslücke, da keine empirischen Studien zur Produktivität von entlehnten Affixen vorliegen. Dennoch werden auch für die Konfixe kategoriale Aussagen zur Produktivität gemacht: Peter von Polenz (1994, 98) führt auf, dass die meisten Konfixe als Erbe des 17. und 18. Jahrhunderts gelten, es aber durchaus auch Konfixe gibt, die erst im 19. und 20. Jahrhundert produktiv werden. Zu diesen gehören beispielsweise *ex-*, *makro-*, *mini-*, *neo-*, *quasi-* und *super-*.

2.1 Funktion der Wortbildung

Sowohl im Deutschen als auch in allen anderen aktiv gesprochenen Sprachen gibt es den Bedarf an Neuwörtern. Diese besitzen die Funktion, bestimmte Erfindungen oder Planungen auszudrücken, für die noch kein allgemein akzeptiertes Wort (im Lexikon) existiert.

Den Bedarf an Neuwörtern erklärt Erben (1993, 22) wie folgt:

„Auch und gerade in einer Sprachgemeinschaft, die Träger einer entwickelten Sprache ist, besteht ein ungeheurer Bedarf an Neuwörtern, da die geistig-sprachliche Auseinandersetzung mit der vielseitigen und veränderlichen Wirklichkeit weitergeht, und nach wie vor die kommunikative Notwendigkeit besteht, alles, was man kennen lernt oder lehrt, auch nennen zu müssen - nicht nur das, was man vorfindet oder im Laufe der Forschung herausfindet, sondern auch das, was man empfindet, erfindet oder entwirft, was man ahnt und was man - auf neuen Tätigkeitsfeldern oder in neuen Organisationsformen - plant. So waren die Wörter Weltraumfahrt, Mondfähre oder Umweltschutz längst da, bevor das Gemeinte volle Wirklichkeit geworden ist.“

Dass dieser Drang nach der Bildung von neuen Wörtern in großem Maße besteht, kann belegt werden, indem man sich das *Wortwarte-Projekt* von Lothar Lemnitzer der Berlin-Brandenburgischen Akademie der Wissenschaften⁴ genauer anschaut. Ziel des Projektes ist es, Neubildungen des Deutschen aus online-verfügbaren Medien (z.B. *Spiegel Online* und *Die Zeit*) zu extrahieren. So findet man am 11.08.2007⁵ beispielsweise „Neuwörter“ wie *Bundesligastheniker*, *Egoliteratur*, *selbstvermarkten*, *Profibettler* oder *Topzinssparen*.⁶

Lothar Lemnitzer geht von einem weit gefassten Neologismus-Begriff aus. Neologismen sind für ihn Wörter, „die so neu sind, dass Unsicherheiten über ihre Form und auch gelegentlich über den Inhalt bestehen, also in der Phase vor ihrer Lexikalisie-

⁴<http://www.wortwarte.de/>.

⁵<http://www.sfs.uni-tuebingen.de/~lothar/nw/Archiv/Datum/d070811.html>.

⁶Bei genauer Betrachtung der Neologismen wird dem linguistischen Beobachter schnell auffallen, dass die Komposita den Großteil der Neubildungen einnehmen. Zudem handelt es sich bei den kompositionellen Strukturen in vielen Fällen um anglizistische Komposita oder zumindest anglizistische Glieder. Derivationale Bildungen sind die Seltenheit.

zung“⁷. Das Problem, das mit dem Begriff der Lexikalisierung verbunden ist, wird zu einem späteren Zeitpunkt dieser Arbeit erneut aufgegriffen und thematisiert.

In der Literatur werden folgende Gründe für Wortbildungsprozesse aufgeführt: Erschaffen von neuen Benennungseinheiten, Ersetzung von veralteten Wörtern, Innovation (Werbesprache), Bedürfnis nach motivierten Zeichen, syntaktische Forderung und Informationsverdichtung.

Als Beispiel für die Benennung eines neuen Phänomens können hier die Begriffe *Lidlisierung* und *Aldisierung* genannt werden. Der Begriff *Lidlisierung* findet sich am 05.09.2006 in der Süddeutschen Zeitung in einer Überschrift eines Artikels.⁸ Der Begriff *Aldisierung* und das damit bezeichnete Phänomen wird in einem Beitrag von Wolfgang Fritz diskutiert.⁹ In beiden Fällen werden Namen von Discountern genommen, die in Deutschland expandieren. Sie werden doppelt deriviert: zunächst werden Verben gebildet, indem das Affix *-isier*-¹⁰ an den Wortstamm gehängt werden, im Anschluss daran erfolgt eine *-ung*-Substantivierung. Die neu gebildeten Derivationen bringen die Prozesshaftigkeit des Expansionsvorgangs deutlich zum Ausdruck.

Ein anderer Grund, ein noch nicht im Wortschatz verankertes Wort zu bilden, ist, dass ein bestehender Ausdruck abgenutzt ist oder eine Wertung enthält, die in der jeweiligen Zeit nicht mehr angemessen ist. Exemplifizieren lässt sich dies durch den Begriff *Raumpflegerin*, der den Begriff *Putzfrau* ablöste.¹¹

In der Werbesprache werden stetig neue Wörter gebildet, die die Funktion besitzen, Aufmerksamkeit zu erzeugen. Ein Beispiel findet sich im Jahre 2006 bei der Automarke Toyota: „Sonderwunschlos glücklich. Nichts ist unmöglich.“ - Toyota Avensis Combi Travel, 2005.¹² Der Werbespruch, genauer gesagt, das Wort *sonderwunschlos*, lässt sich auf zwei unterschiedliche Weisen interpretieren: als Suffigierung des

⁷<http://www.wortwarte.de/Projekt/neu.html>, abgerufen am 22.11.2010 um 11:44 Uhr.

⁸„Die Lidlisierung bestimmt unser Leben“, Der Öko-Unternehmer Schweisfurth über das agroindustrielle System und was der Verbraucher tun kann, http://www.schweisfurth.de/fileadmin/dateien_news/SZ_September_06.pdf.

⁹Dazu siehe <http://www.wiwi.tu-bs.de/marketing/service/download/art/aldi.pdf>, abgerufen am 22.11.2010 um 10:55 Uhr.

¹⁰Zur Geschichte und Semantik des Affixes siehe Fleischer und Barz (1995, 311f).

¹¹Als weiteres Beispiel kann der Begriff *Negerkuss* aufgeführt werden, der durch *Schokokuss* ersetzt wurde; <http://www.esskultur.net/lm/negerkuss.html>, Abruf am 24.06.2011 um 11:46 Uhr.

¹²Quelle: <http://de.wikiquote.org/wiki/Werbesprüche>.

Substantivs *Sonderwunsch* mit dem Affix *-los*¹³ oder als Kontamination¹⁴ aus *Sonderwunsch* und *wunschlos*.

Ein weiterer Grund, der dazu führt, dass neue Wörter gebildet werden, ist das Bedürfnis nach motivierten Zeichen, die nicht isoliert im Zeichensystem stehen (vgl. Erben 2000). So findet man in alten Texten noch den Begriff *Lenz*; im Gegenwartsdeutschen wird man vorzugsweise die Wörter *Frühjahr* oder *Frühling* verwenden.

Wortbildung kann auch grammatisch motiviert sein.¹⁵ Erben (1993, 24) erklärt dieses Phänomen wie folgt:

„Daneben gibt es grammatisch bedingte Notwendigkeiten der Wortbildung. Es geht da z.T. um die Beseitigung morphologischer Mängel im Flexionssystem, wobei etwa die fehlende Pluralform durch Wortbildung 'suppliert' wird (vgl. Schnee: Schnee-massen, Regen: Regen-fälle); Trost: Tröst-ung-en, Liebe: Lieb-schaft-en) oder die mangelnde Steigerungsform geschaffen wird (vgl. tot: mause-tot, feind: spinne-feind, schade: jammer-schade; wert: wert-voll-er statt wert-er).“

Auf der Ebene der Kompositabildung im Deutschen ist vor allem der Aspekt der Informationsverdichtung zu erwähnen. Die Kompositabildung im Deutschen ist rekursiv, was bedeutet, dass sich im Prinzip immer noch ein weiteres Wort anhängen lässt. Dies kann zu hoch komplexen Wörtern führen: z.B. *Donaudampfschiffahrtselektrizitätenhauptbetriebswerkbauunterbeamtengesellschaft*.¹⁶ Doch was zunächst durchaus vorteilhaft klingt, sollte allerdings an dieser Stelle auch kritisch beleuchtet werden. Spricht man im Deutschen von einem Jägerschnitzel, so ist, so besagt es das „Kopfprinzip“¹⁷, ein Schnitzel gemeint, allerdings ein Schnitzel nach Jägerart und nicht das Schnitzel eines Jägers. Somit kann es durch die informationsverdichtende Wortbildung auch zu einer Art Unterspezifikation kommen.

¹³Zur semantischen Beschreibung dieses Affixes siehe Kap. 5.1.13.

¹⁴Bei der Kontamination handelt es sich um eine nicht produktive Wortbildungsart, bei welcher Wörter miteinander verschmolzen werden (vgl. Donalies 2005).

¹⁵Siehe dazu auch Fleischer und Barz (1995, 5), die Beispiele aufführen, wie Lücken im grammatischen Paradigma durch Wortbildung geschlossen werden können und in diesem Kontext die Beispiele *Getreide-arten*, *Atem-züge* und *Rat-schläge* nennen.

¹⁶Siehe dazu <http://de.wikipedia.org/wiki/Donaudampfschiffahrtselektrizitätenhauptbetriebswerkbauunterbeamtengesellschaft>.

¹⁷Das Kopfprinzip in der Linguistik sagt aus, dass jede Phrase genau einen Kopf hat. Der Kopf projiziert die Phrase und bestimmt ihre kategorialen Merkmale: [NP die hübsche [N Sängerin]]. Auch für Komposita gilt das Kopfprinzip, indem Merkmale wie Genus weitervererbt werden und der rechte Teil des Kompositums determinierend ist und damit auch die syntaktische Kategorie bestimmt: *rot*, *Wein*, *Rotwein*.

2.2 Einheiten der Wortbildung

Als Grundeinheit der Wortbildung benennen Fleischer und Barz (1995, 21) das Wortbildungsprodukt.

In Abhängigkeit der möglichen Bildungsverfahren ergeben sich daraus unterschiedliche Strukturen: komplexe Wortbildungsprodukte mit unmittelbarer Konstituentenstruktur (UK-Struktur) wie *unverantwortlich* oder *begrüßen* oder sekundäre Simplizia (*die Schau, das Band*).

Wortbildungen sind Morphemgefüge, deren Segmentierung zur Isolierung der bedeutungstragenden Einheiten, der Morpheme, führt (ebd.).

Unterscheiden lassen sich zunächst *freie* und *gebundene* Morpheme, die zudem entweder *grammatisch* oder *lexikalisch* sind. Freie lexikalische Grundmorpheme (z.B. *Baum, Tisch, Stuhl*) tragen eine begriffliche Bedeutung, freie grammatische Morpheme besitzen eine grammatische Funktion; zu ihnen gehören beispielsweise Artikel und Präpositionen.

Auch für die Kategorie der gebundenen Morpheme lässt sich die dichotome Unterscheidung 'grammatisch versus lexikalisch' heranziehen: Grammatisch gebunden sind Flektionsmorpheme; bei Affixen, die aus Derivationsbasen komplexe Wörter bilden, ist es so, dass sie zwar eine lexikalische Bedeutung besitzen, aber einen höheren „Allgemeinheitsgrad“ (Motsch 1983, 109) aufweisen als freie Morpheme. Dieser Aspekt ist für diese Arbeit insofern von Bedeutung, als dass der Allgemeinheitsgrad und die damit einhergehende Abstraktheit der Wortbildungsmorpheme statistisch untersucht werden sollen.

Nach Fleischer und Barz (1995, 28) verfügen Affixe über folgende Eigenschaften:

1. **Reihenbildung:** Damit ist gemeint, dass Affixe wiederholt in Wortbildungskonstruktionen des einen und desselben Modells auftreten können: *erklär-*, *amt-*, *bedenklich*.
2. **Abstraktheit:** Im Vergleich zu lexikalischen freien Morphemen ist die Bedeutung von Affixen abstrakter, sodass auch von einem „wesentlich höheren Allgemeinheitsgrad“¹⁸ gesprochen werden darf.
3. **Gebundenheit:** Affixe können nicht frei verwendet werden und sind zudem auf eine Position festgelegt (z.B. wortinitial bei Präfixen, wortfinal bei Suffixen).

¹⁸Motsch (1983, 109).

4. **Nicht basisfähig:** Es ist im Deutschen unmöglich, ein Affix als Basis zur weiteren Wortbildung zu verwenden: **verheit*, **entbar* etc.
5. **Einsilbigkeit:** In der Regel sind Affixe einsilbig und bestehen aus einem Vokal und einem oder mehreren Konsonanten. Eine Ausnahme jedoch bildet z.B. das Suffix *-mäßig*.
6. Affixe unterliegen bestimmten **Restriktionen:** das Suffix *-bar* lässt sich nicht mit intransitiven Verben verbinden. Dieses Phänomen ist Gegenstand des Abschnitts 2.4.

Sind Merkmale 1-3 vorhanden, so gehören diese Affixe sogar zum „Kernbereich der Affixe“ (ebd.).

Im Bezug auf die Position der Basis lassen sich Präfixe, Suffixe, Interfixe und Zirkumfixe voneinander unterscheiden. In dieser Arbeit jedoch werden nur wortfinale Wortbildungsaffixe, Suffixe, hinsichtlich ihrer Produktivität untersucht. Die Produktivitätsmessung kann allerdings auch für anderweitig positionierte Affixe angewandt werden. Dass die Notwendigkeit besteht, auch Präfixe und Zirkumfixe des Deutschen auf ihre Produktivität hin zu untersuchen, wird im Laufe der Arbeit noch deutlich werden, denn der Forschungsbericht in Kap. 3.2 und 3.3 bringt zum Vorschein, dass quantitative Studien zur Produktivität von Affixen im Allgemeinen bisher rar sind, sodass nicht nur die existierende Forschungslücke in Bezug auf Suffixe geschlossen werden sollte.

2.3 Triebkräfte der Wortschatzerweiterung

Werden Neuwörter gebildet, so geschieht dies nicht willkürlich, sondern regelbasiert; Regelbasiertheit ist ein zentrales Merkmal von produktiven Wortbildungsmodellen. Doch im Zusammenhang mit der Wortschatzerweiterung im Deutschen müssen neben der Produktivität auch zwei weitere Kräfte genannt werden: Kreativität und Analogie. Wie das Zusammenspiel dieser Triebkräfte zu verstehen ist, soll im Folgenden skizziert werden.

Produktivität ist die Möglichkeit, gemäß den Wortbildungsregeln einer Sprache neue Wörter zu bilden. Durch produktive Baumuster werden neue Lexeme geschaffen, die zur Erweiterung des Lexikons führen. Produktive Wortbildungsprozesse werden mit einem geringen Grad an Bewusstheit angewendet (Evert und Lüdelling 2001). Zudem

ist Produktivität eine graduelle Größe, die, so Fleischer und Barz (1995, 57), „in Abstufungen auftritt“.

Während Fleischer und Barz (1995) und Lohde (2006) Affixe des Deutschen in produktive, hoch/stark produktive und gering produktive und unproduktive Affixe einteilen, spricht Motsch (2004) von aktiven und inaktiven Mustern:

„Aus der Warte der **Sprachproduktion** bzw. **-rezeption** bedeutet das: es gibt Muster, die für die Bildung oder Rezeption neuer Wörter genutzt werden und solche, für die das nicht gilt. Wir unterscheiden zwischen **aktiven** und **inaktiven Wortbildungsmustern** und betrachten diese Differenzierung als Bestandteil der Beschreibung von Wortbildungsmustern. Sie reflektiert die Fähigkeit von Sprachteilnehmern, zwischen Wörtern zu unterscheiden, die zu ihrem festen Wortschatz gehören und solchen, die für sie neu sind.¹⁹

Motsch macht darauf aufmerksam, dass er den Begriff der Aktivität synonymisch zu dem Begriff der Produktivität verwendet. Ein Problem, das er selbst beleuchtet, ist die Exaktheit dieser Angaben. Er räumt ein, dass „genaue Kriterien“ zur Klassifizierung von Aktivitätsgraden fehlen und, dass es sich deshalb um rein subjektive Einschätzungen handelt. Inaktiv ist ein Muster seiner Meinung nach genau dann, wenn es entweder keine oder nur von wenigen Autoren willkürlich benutzte Textwörter gibt. Schwach aktiv bezeichnet die Eigenschaft eines Musters, nur vereinzelte Neologismen zu produzieren. Von starker Aktivität darf gesprochen werden, wenn „relativ häufig“ Neologismen mit dem Affix gebildet werde.²⁰

Fleischer und Barz (1995, 57) bezeichnen Suffixe dann als produktiv, sofern ein Wortbildungsmodell nur wenige Restriktionen aufweist und es „in hohem Maße auch für okkasionelle Neubildungen genutzt werden kann“. Ersteres stellt für sie das qualitative Kriterium dar, während die hohe Anzahl an Neologismen als quantitatives Kriterium zu bezeichnen ist.

¹⁹Motsch (2004, 19).

²⁰Des Weiteren äußert er sich wie folgt: „Die Unterscheidung zwischen aktiven und inaktiven Wortbildungsmustern kann empirisch durch die Fähigkeit von Sprechern einer Sprache begründet werden, zwischen Wörtern, die zu ihrem festen Wortschatz gehören, und solchen zu unterscheiden, die für sie fremd oder auffällig sind. Ein für einen Sprecher neues Wort muss aber nicht für alle Sprecher neu sein. Entscheidungen darüber, ob eine Bildung zum Lexikon einer Sprachgemeinschaft gehört, oder ob sie nur für bestimmte Anlässe in einem Text vorkommt und deshalb als Neubildung auffällt, sind wegen der individuellen Abweichungen bei der Beantwortung dieser Frage problematisch.“ (Motsch 2004, 19).

Und genau darin liegt ein Ansatzpunkt für das Vorhaben dieser Arbeit: es existieren bisher keine genauen Angaben zur Klassifizierung von Produktivitätsgraden. Die Angaben der linguistischen Literatur stützen sich nicht auf systematisch erhobene empirische Befunde. Die Arbeit zielt darauf ab, die Aussagen zur Produktivität der untersuchten Suffixe²¹ zu überprüfen.

Das qualitative Kriterium ist, zumindest numerisch gesehen, nicht zu fassen, denn die Restriktionen, die ein Affix in seiner Produktivität einschränken können, können mathematisch nicht gewichtet werden. Sie betreffen unterschiedliche Ebenen (Phonologie, Morphologie, Syntax etc) und werden im Laufe dieses Kapitels detailliert behandelt und mit Beispielen belegt. Das quantitative Kriterium hingegen kann sehr gut erfasst werden. Auch die Anzahl an Neubildungen ist ein entscheidender Faktor für eine hohe Produktivität und wird in dieser Arbeit mit einbezogen.

Doch neben der Triebkraft Produktivität trägt auch die Kreativität ihren Teil zur Erweiterung eines Wortschatzes bei. Sie ist – im Gegensatz zur Produktivität – nicht zuständig für regelbasierte Neubildungen, sondern Resultat intentionaler Akte eines Sprachbenutzers (vgl. Bründl 2001).

Exemplarisch für kreative Bildungen können zum Beispiel Akronyme (*ADAC* für Allgemeiner Deutscher Automobil-Club e.V.) und Kontaminationen (z.B. *gruscheln*²² als Zusammenfügung aus *grüßen* und *kuscheln*) genannt werden.²³

Im Zusammenhang mit den beiden Begrifflichkeiten Ausdrücken Produktivität und Kreativität steht auch die Analogie. Der Zusammenhang zwischen den drei Konzepten ist folgender: Analogie wird als kognitives Prinzip aufgefasst, das sich auf sprachlich-lexikalischer Ebene dadurch bemerkbar macht, dass es als treibende Kraft in produktiven und kreativen Prozessen wirksam ist und auf diese Weise zur Erweiterung des Wortschatzes beiträgt (ebd.).

Bründl (2001) unterscheidet zwei Arten von Analogien: die morphologische/formale und die semantische Analogie. Durch erstere entstehen neue Lexeme

²¹Die kategorischen Aussagen zur Produktivität der Wortbildungssuffixe können Tab. 5.1 in Kapitel 5.1.14 entnommen werden.

²²Das Verb *gruscheln* bezeichnet eine spezielle Funktion in einem digitalen sozialen Netzwerk. Dazu siehe <http://de.wikipedia.org/wiki/StudiVZ#Gruscheln>, abgerufen am 12.06.2011 um 13:33 Uhr.

²³Gerade in der Werbesprache spielen Innovation und Kreativität eine entscheidende Rolle. Exemplarisch können hier Namen von Friseursalons genannt werden, die das Morphem *-Haar-* (oder das anglizistische *-hair-*) in ihren Salonnamen integrieren, sodass kreative Bildungen wie zum Beispiel *Verhaaren* (http://www.branchenbuchsuche.de/verhaaren-frisoer-in-dortmund_1080304) oder *Haarmonie*²⁴ produziert werden.

bzw. Wortbildungen: ein oder mehrere Morpheme eines existierenden Modellwortes werden durch Analogiebildung auf ein Neuwort übertragen, während das übrige Morphem des Modellwortes durch ein Neues ersetzt wird. So entstehen aus dem Modellwort *einsam* Worte wie *zweisam* oder *dreisam*.²⁵ Durch semantische Analogie hingegen entstehen vor allem neue lexikalische Einheiten wie Metaphern.²⁶

Zusammenfassend veranschaulicht Abb. 2.1 grafisch die Beziehungen zwischen Analogie, Produktivität und Kreativität. Wie die Abbildung zeigt, sind die Triebkräfte in Interaktion zu betrachten.

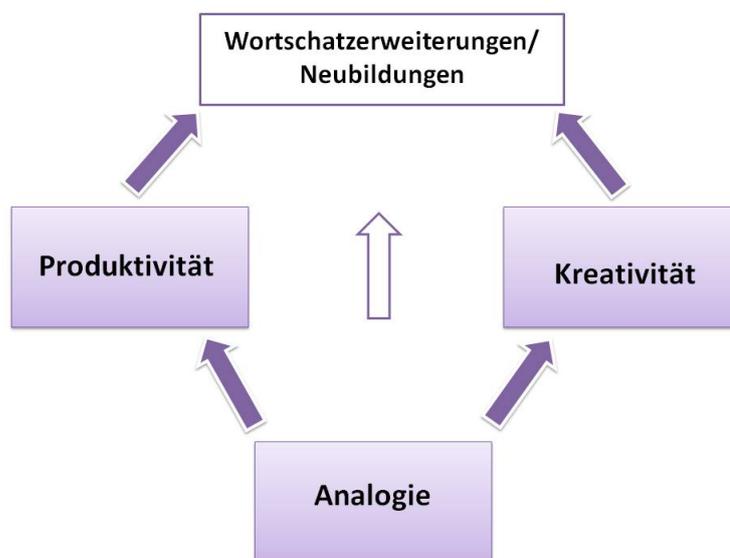


Abbildung 2.1: Wortschatzerweiterung nach Bründl (2001, 64)

In der Wortbildung wird Analogie mitunter als Gegenteil zur Produktivität angesehen; es wird davon ausgegangen, dass eine bestimmte Bildung aufgrund eines schon bestehenden Musters produziert wird, wohingegen eine produktive Regel unbegrenzt angewendet werden kann. Diese Ansicht ist kritisch zu betrachten, da Analogiebildung auch als Anstoß für eine produktive Regel fungieren kann. Als Beispiel sei hier erneut das Suffix *-sam* genannt, das als unproduktiv beschrieben wird (Lohde 2006). 1994

²⁵Ein weiteres Beispiel ist die analoge Bildung *Mutbürger* SPIEGEL 42/2010 vom 18.10.2010, <http://www.spiegel.de/spiegel/print/d-74549707.html>, die in Anlehnung an den *Wutbürger* (Wort des Jahres 2010, <http://www.gfds.de/presse/pressemitteilungen/171210-wort-des-jahres-2010/>) entstanden ist.

²⁶In der deutschen Sprache existieren eine Vielzahl an Wörtern, die Zustände oder Vorgänge gleichzeitig abstrakt und auch konkret beschreiben. Als Beispiele können *untermauern* oder *einbinden* genannt werden.

erschien ein Kinofilm mit dem Titel *Einsam, zweisam, dreisam* sodass *zweisam* und *dreisam* durch Analogiebildung produziert werden. Die Reihe auf *-sam* kann fortgesetzt werden, was die Bildung von *viersam*²⁷ oder *fünfsam*²⁸ erklärt.

Die traditionelle Unterscheidung zwischen Produktivität und Kreativität besteht darin, dass die Produktivität auf der Ebene der *langue* anzusiedeln ist, während sich die Kreativität auf der Ebene der *parole* wiederfindet (Marchand 1969, Bauer 1988, Lyons 1977).²⁹ Dennoch lässt sich daraus keineswegs ableiten, dass die Wahrscheinlichkeit zur Bildung von neuen Wörtern von produktiven Wortbildungsmustern nur durch Mittel der linguistischen Kompetenzbeschreibung zu erfassen ist; vielmehr wird angenommen, dass diese durch pragmatische und psycholinguistische Faktoren beeinflusst wird. So zum Beispiel besagt das „Prinzip des sinnvollen Wortes“ (Demske 2001, 77), dass nur diejenigen Wörter in das Lexikon einer Sprache aufgenommen werden, die als kommunikatives sinnvolles Konzept beschrieben werden können. Exemplifizierend werden die Beispiele **kopfig*, **armig* und **beinig* (ebd.) als unmögliche Bildungen benannt.³⁰ Blockierungen von Neubildungen wie zum Beispiel die Unmöglichkeit, im Deutschen das Wort **Kocher* ('Person, die kocht') zu bilden, weil das Wort *Koch* existiert, gehören zum Bereich der Performanz.³¹ Dieses Phänomen ist Gegenstand von Kapitel 2.4.5; darauf sei hier verwiesen.

²⁷Dazu siehe <http://fem.twoday.net/topics/viersam/>, abgerufen am 13.12.10 um 16:56 Uhr.

²⁸http://www44.sb6.os-community.de/Forum/Sonstiges/Thema/275001/Was_hoerst_Du_gerade?page=2191.

²⁹Kreativität ist als Erscheinung der *parole* als Auslöser für nicht-regelbasierte Bildungen zu betrachten. Produktive Bildungen hingegen sind regelbasiert. Des Weiteren muss ein Sprecher einer Sprachgemeinschaft sein Wissen über Aktivitäts- bzw. Produktivitätsunterschiede zwischen konkurrierenden Wortbildungsmitteln einsetzen, um neue Wortbildungen erzeugen zu können.

³⁰Eine interessante Diskussion zum Suffix *-ig* im Gegenwartsdeutschen findet man in Sick (2009, 138ff). Unter dem Titel „Manche mögen's apfelig“ widmet der Spiegel Online-Autor dem Phänomen den „originellen“ *-ig*-Bildungen ein ganzes Kapitel in seinem Buch. Er stellt Folgendes fest: „Immer wieder geraten Wörter in den Umlauf, die es eigentlich nicht geben dürfte: Adjektive wie möhrig oder rosinig. [...] Man kann darüber schmunzeln oder verwundert den Kopf schütteln. Auf jeden Fall sind Wortbildungen wie apfelig oder karottig originell. Das Wunderbare an unserer Sprache ist ja, dass sie so elastisch ist und viel mehr Möglichkeiten zulässt, als unser willkürliches und nicht immer besonders logisches Regelwerk erlauben will.“

³¹Angemerkt sei an dieser Stelle, dass sich die Blockierung eines Wortes nur auf dessen Lexikalisierung bezieht, die individuelle Verwendung einer sonst blockierten Form ist indes möglich. Als Beispiel sei hier der Gebrauch des Wortes *Stehler* genannt, das durch die existente Form *Dieb* blockiert wird. Der Begriff findet sich in der Überschrift eines Artikels (<http://www.dasinvestment.com/berater/news/datum/2010/07/28/der-hehler-ist-nicht-besser-als-der-stehler-leserreaktionen-zur-steuersuender-cd/>) auf einer Internetseite (05.02.2010, 10:06 Uhr) und wird an dieser Stelle eingesetzt, um ein Reimschema zu erzeugen.

2.4 Regeln und Einschränkungen der Wortschatzerweiterung

Sowohl beim Produzieren als auch beim Rezipieren von (neuen) Wörtern werden Wortbildungsmuster aus dem Lexikon aktiviert.³² Motsch (2004, 23) nennt einige Faktoren, die die Möglichkeit, Wortbildungsmuster aus lexikalisierten Formen zu abstrahieren, erleichtern:

1. **Typenhäufigkeit**³³, die er als „die Menge der zu einem Muster gehörenden lexikalisierten Wortbildungen“ beschreibt.

Seine These in diesem Zusammenhang lautet wie folgt: Je höher die Typenhäufigkeit ist, desto leichter ist das jeweilige Wortbildungsmuster aktivierbar.

2. **Aktivierungsgrad**

Er vertritt zudem die These, dass eine höhere Aktivierung eines Wortbildungsmusters zu einer höheren Bewusstseinspräsenz führt, die wiederum zu neuen regelbasierten Formen führen kann.

3. **Transparenz der Wortbildungsmuster**

Laut Motsch lässt sich sagen, dass ein durchschaubares Muster³⁴ leichter zur Produktion und zur Rezeption von Neubildungen führt.

4. **Singuläre Analogien**

Dieses sind für ihn Wortbildungen, die aufgrund schon bestehender Muster produziert werden, indem ein Teil des lexikalisierten Wortes erhalten bleibt: *Einsamkeit* vs. *Zweisamkeit*.

Der von Motsch (2004) genannte Zusammenhang zwischen frequenten Wortbildungsmustern und der damit einhergehenden hohen Wahrscheinlichkeit zur Bildung neuer Wörter wird in Kapitel 7 und 8 Teil der Auswertung und Diskussion dieser Arbeit sein. Die Frage, die sich in diesem Rahmen stellt, ist, inwieweit eine hohe Ty-

³²Siehe dazu Kap. 6, in welchem die Ergebnisse eines Produktionsexperimentes vorgestellt werden und in diesem Zusammenhang auch auf den Punkt der Lexikalisierung und Aktivierung aus dem Lexikon eingegangen wird.

³³In der strukturalen Linguistik dienen die Begriffe zur Unterscheidung zwischen konkreten sprachlichen Äußerungen (Token) und abstrakten Einheiten der Metaebene (Typen), die sie repräsentieren. So enthält der Satz *Eine Rose ist eine Rose, auch in gelb* zwei Token Rose, aber nur ein Type. Das Begriffspaar dient also auf Wortebene zur Unterscheidung zwischen *langue* und *parole*.

³⁴Exemplifizierend dazu nennt er Derivate mit *-haft*, die einen Vergleich herstellen. Als semantisch komplexer bezeichnet er Kollektiva und gibt dazu das Beispiel *Hegelianertum*, *Altherrenschaft* und *Gebeine*.

penhäufigkeit mit hoher Produktivität gleichgesetzt werden kann. Einerseits scheint es mehr als verständlich, davon auszugehen, dass ein Muster, das produktiv sein soll, viele Wortformen hervorbringen sollte, andererseits ist es gut möglich, dass ein Wortbildungsmuster zunächst sehr produktiv ist, aber durch ein anderes Muster ersetzt wird, sodass das zunächst produktive Muster an Produktivität verliert. Dennoch bleibt eine große Anzahl an lexikalisierten Wörtern im Wortschatz erhalten, die aus produktiven Zeiten des Affixes stammen.

Um neue Dinge auszudrücken und Benennungseinheiten zu schaffen, greifen Sprecher einer Sprache zu den für sie verfügbaren Mitteln der Wortbildung. Gleichwohl steht jedem Sprecher des Deutschen bei der Wahl seines sprachlichen Gebildes ein gewisser Spielraum zur Verfügung, da die Sprache auch alternative Strukturen anbietet; dies bedeutet, dass sich derselbe Inhalt durch unterschiedliche sprachliche Mittel ausdrücken lässt.

Der Satz *Anna treibt Sport*, der eine Prädikation beinhaltet, kann auch durch eine andere Prädikation mit einem desubstantivischen Adjektiv wiedergegeben werden, wie es in *Anna ist eine sportliche Frau* der Fall ist. Weitere Varianten wären durch *Anna ist eine Sportlerin* oder *Anna ist eine Sportskanone* gegeben.³⁵

Auch das deutsche Derivationssystem bietet Möglichkeiten an, bestimmte Affixe synonymisch zu verwenden, indem verschiedene Affixe an dieselbe Basis gehängt werden können ohne, dass es zu einem Sinnverlust kommt. So kann man Briefmarken im Deutschen sowohl *an-* wie auch *befeuchten* (Erben 2000, 39).

Doch Affixe³⁶ können bestimmten Restriktionen unterliegen, die unterschiedlich motiviert sein können. In den folgenden Abschnitten wird dies genauer skizziert und exemplifiziert.

³⁵Dazu siehe auch Motsch (2004, 18), der in diesem Zusammenhang von „syntaktischen Alternativen von Wortbildungsmustern“ spricht und derivierte Adjektive (*ärztliches Attest*) zu Genitiv- (*Attest eines Arztes*) und Präpositionalphrasen (*Attest von einem Arzt*) gegenüberstellt.

³⁶Doch nicht nur für das Derivationssystem kann man Restriktionen beobachten. Als Beispiel nennt Erben (2000) hier Kopulativkomposita, die seiner Meinung nach der Einschränkungregel unterliegen, dass nur „wortfähige Grundmorpheme gleicher Wortklasse“ paarig miteinander verbunden werden können, da eine Umstellung ohne Sinnverlust möglich ist. Dabei ist die Wortart an sich variabel, es eignen sich sowohl Substantive für eine derartige Zusammenfügung (*Dichter-Sänger*) wie auch Adjektive (*süß-bitter*).

2.4.1 Phonologische Beschränkungen

Phonologische Beschränkungen können sich prinzipiell auf alle Aspekte der Basis beziehen; zudem sind Suffixe in stärkerem Maße von ihnen betroffen als Präfixe (vgl. Rainer 2000, 881).

So zum Beispiel kann der segmentale Aufbau eines Wortes Beschränkungen begünstigen. Mit dem Morphem *-ität* lässt sich im Deutschen *Rapidität* und *Durabilität* bilden, nicht aber **Akutität* oder **Stereotypität*.

Des Weiteren wird das Suffix *-heit* in der Regel nur dann mit mehrsilbigen Adjektiven verbunden, wenn eine Endbetonung vorliegt. Exemplifizieren lässt sich dies durch das Wort *Gesamtheit*.³⁷

Als Beispiel zur Vermeidung des Gleichklangs kann die Restriktion vom Diminutivsuffix *-chen* genannt werden. Es ist im Deutschen abhängig von der Basis: nach *l* wird *-chen* verwendet während man nach *ch* im Deutschen *-lein* nimmt, da es sonst zu einer Verdopplung des Phonems kommt (ebd.).³⁸

2.4.2 Morphologische Beschränkungen

Bauer (2001, 130) nennt drei Möglichkeiten einer morphologischen Restriktion:

„First, the base may have to belong to a particular morphologically defined class before affixation can occur. Second, the base may have to show a particular morphological structure before affixation can occur. Third, the base may have to end in a particular affix or not end in a particular affix before affixation can occur.“

Den ersten morphologischen Constraint exemplifiziert Bauer (2001) mit einem Beispiel aus dem Hebräischen. Dort verwendet man ein Diminutivsuffix nur bei maskulinen Nomen.³⁹

Die Verwendung oder Nicht-Verwendung eines bestimmten Affixes kann auch strukturell motiviert sein: So wird das Suffix *-heit* zur Substantivierung von Adjek-

³⁷Wörter wie z.B. *Vollkommenheit* oder *Angetrunkenheit* zeigen, dass ein wortfinales unbetontes *-en* nicht zur Restriktion führt.

³⁸Exemplifizierend wird an dieser Stelle auf ein lateinisches Beispiel verwiesen: So z.B. bildet man im Lateinischen nach Basen auf *r* das Adjektiv nicht mit *-aris*, sondern mit *-alis*. Im Altspanischen nahm man nach Basen mit einem Lateral des Diminutiv *-ejo* und nicht *-iello*.

³⁹Dabei ist ausschließlich die Form des Wortes entscheidend, die dann zu CCaCCaC wird. Als Beispiel wird *zkankan* versus *zakan* genannt, was soviel wie '(kleiner) Bart' heißt.

tiven bei Komposita (*Rede-faul-heit*), bei präfigierten Adjektiven (*Un-wahr-heit*), bei adjektivischen Partizipien, die mit einem Zirkumfix gebildet worden sind (*Ge-schlossen-en-heit*) verwendet, aber es kann nicht mit suffigierten Adjektiven verbunden werden.

Die dritte Beschränkung darf als *repeated morph constraint* (Bauer 2001, 132) bezeichnet werden. Sie besagt, dass es nicht möglich ist, dasselbe Affix zweimal hintereinander zu hängen. Allerdings scheint im Deutschen eine Verletzung der Restriktion möglich, da Bildungen wie *Ur-ur-großvater* oder *Ur-ur-großmutter* zulässig sind.⁴⁰

2.4.3 Syntaktische Beschränkungen

Es gibt zwei Arten von syntaktischen Beschränkungen: eine wortartbezogene und eine, die auf syntaktische Merkmale der Basis zurückzuführen ist (vgl. Rainer 2000, 882).

Die erste Beschränkung wird in vielen Grammatikbüchern deutlich zum Ausdruck gebracht, indem Ausdrücke wie *deadjektivisch* oder *denominal* anzeigen, inwieweit man ein Affix verwenden oder auch nicht verwenden darf.

Der zweite Punkt bezieht sich auf Merkmale von Wortarten wie z.B. [transitiv], [ergativ] etc. Im Deutschen können viele Affixe nur mit bestimmten Wortklassen⁴¹ verbunden werden, z.B. *-bar* mit transitiven Verben wie in *lesbar*.⁴²

2.4.4 Semantische Beschränkungen

Auch existieren semantische Beschränkungen. So äußert sich Bauer (2001, 134) diesbezüglich wie folgt:

„In some cases, semantic restrictions in the base are clearly a matter of what it makes sense to have a word for.“

⁴⁰Siehe <http://www.duden.de/rechtschreibung/Ururgroszvater>, abgerufen am 20.06.2011 um 18:39 Uhr.

⁴¹Diese Wortklassenbeschränkung gehört nach Bauer (2001, 133) zu den häufigsten syntaktischen Beschränkungen. Er referiert in diesem Zusammenhang auf Koehn und Koehn (1986), die ein Beispiel aus der Sprache Apalai nennen, in welcher es zwei verschiedene Affixe gibt, um „agents of present action“ zu beschreiben; *-ne* markiert die transitiven Verben und *-kety* die intransitiven.

⁴²Auch für das Niederländische gilt, dass sich das *-baar* nur mit transitiven, nicht aber mit intransitiven Verben verbindet.

Das deutsche Suffix *-tum* kann in der Regel nur mit Nomen verbunden werden, die Personen bezeichnen, z.B. ist die Bildung von *Strebertum*⁴³, nicht aber die Bildung von **Armtum* möglich.⁴⁴

Ebenso kann an dieser Stelle das Präfix *re-* genannt werden, das den durch die Basis ausgedrückten Zustand rückgängig macht. Im Deutschen lässt sich etwas *reanimieren* oder *reimportieren*, aber nicht **retöten* oder **relachen*.⁴⁵

2.4.5 Lexikalische Beschränkungen

Bei nicht produktiven Affigierungen lässt sich von einer lexikalischen Beschränkung sprechen, da die Anzahl der Wörter, die durch ein nicht produktives Affix entsteht, sehr gering ist.

Bauer (2001) exemplifiziert dieses durch das englische Suffix *-th*, bei dem es möglich ist, die gesamte Liste der Wörter mit diesem Suffix aufzuschreiben. Auch bei dem Suffix *-ter* ist die Anzahl der Bildungen sehr gering, gebildet werden nur *laughter* und *slaughter*, sodass die Typenhäufigkeit exakt zwei beträgt.

Manchmal lässt sich in der linguistischen Literatur auch den Begriff *Blocking* als lexikalische Restriktion finden. Der Begriff *Blocking* wird gleich in zwei verschiedenen Zusammenhängen angewendet. Aronoff (1976, 43) definiert *Blocking* als „nonoccurrence of one form due to the simple existence of another“. Gemeint ist damit das Phänomen, dass es möglich ist, dass Synonyme existieren, die das Bilden einer – vielleicht möglichen und regelbasiert geformten – Form verhindern: So z.B. würde man das Wort *Stehler* im Deutschen eher selten bilden, da das Wort *Dieb* existiert.

Die andere Möglichkeit, dass ein Wort nicht gebildet wird, ist das Vorhandensein eines im Wortschatz frequent genutzten Homonyms. Die Betitelung eines menschlichen Organs durch den Begriff *Leber* blockiert die durchaus plausible und regelbasierte Form **Leber* als 'Mensch, der lebt'.

⁴³Dazu siehe <http://www.frag-caesar.de/synonyme/Strebertum.html>, abgerufen am 22.12.2010 um 16:25 Uhr.

⁴⁴Im Russischen z.B. existieren markierende Affix *-stv.o* wird nur an Lehrer- und Professorenbezeichnungen angehängt, um bestimmte Pflichten auszudrücken.

⁴⁵Rainer (2000, 882) nennt das englische Suffix *un-*, das „den von der Basis herbeigeführten Zustand rückgängig macht“. Wörter wie *unhold* oder *unscrew* können problemlos gebildet werden, während es nicht möglich ist, **unkill* zu bilden, da es faktisch nicht realisiert werden kann.

Damit eine „Wort-zu-Wort-Blockierung“ wirksam ist, müssen nach Rainer (2000, 878) gleich drei Bedingungen erfüllt sein: „das blockierende und das blockierte Wort müssen synonym sein, das blockierte Wort muß nach einer produktiven Regel bildbar sein, und das blockierende Wort muß eine ausreichende Gebrauchshäufigkeit besitzen.“

Motsch (1999, 331) benennt konkrete Beispiele des Deutschen und listet in diesem Kontext eine Reihe von lexikalisierte Verbalnomen auf, die eine regelbasierte Form mit dem hoch produktiven Suffix *-ung* verhindern:

- **Bummelung* vs. *Bummel*
- **Fahrung* vs. *Fahrt*
- **Fragung* vs. *Frage*
- **Führung* vs. *Gefühl*
- **Glaubung* vs. *Glaube* etc.

2.4.6 Pragmatische Produktivitätsbeschränkungen

Obwohl es in der Praxis schwierig ist, zwischen semantischen und pragmatischen Beschränkungen zu unterscheiden, erklärt Bauer (2001) diesen Unterschied wie folgt:

„[...] but in principle semantic effects deal with the linguistic nature of the base, while pragmatic effects deal with the way in which the words are used or the nature of the real-world referent of the word.“

In diesen Bereich gehören soziolinguistisch motivierte Beschränkungen: es ist möglich, dass Wortbildungsregeln eine registermotivierte Beschränkung besitzen. So z.B. ist es relativ wahrscheinlich, dass neuere deutsche Affixe/Affixoide wie *mega-*, *Bomben-*, *Drecks-*, *Heiden-*, *Höllen-*, *Monster-*, *Mords-*, *Sau-* und *Todes-* in der Jugendsprache Verwendung finden (Ascoop 2005, 23), während man sie in der Wissenschaftssprache vermutlich nicht sehr häufig finden wird.⁴⁶

⁴⁶Vgl. dazu auch Rainer (2000, 883), der beschreibt, dass diese Art von Beschränkungen eine hohe Komplexität aufweisen können. Diminutive des Spanischen werden beispielsweise in der Unterschicht häufiger als in der Oberschicht, ländlich häufiger als städtisch, von Frauen häufiger als von Männern, gegenüber Kindern häufiger als gegenüber Erwachsenen und in vertrauten Gesprächen häufiger als in formellen verwendet.

2.4.7 Fazit

Zusammenfassend lässt sich festhalten, dass das Blockieren einer regelbasiert gebildeten morphologischen Form synonymisch und homonymisch motiviert sein kann. Produktivitätsbeschränkungen limitieren die Anzahl der potentiellen Wörter. Diese Beschränkungen können allerdings nur auf qualitativer Ebene festgehalten werden und nicht auf quantitativer.

Außerdem sei Folgendes ergänzend hinzugefügt: Sind mehrere Affixe ähnlicher Funktion vorhanden, besteht zwar generell eine Wahlfreiheit in der Verwendung, dennoch ist diese Freiheit nicht uneingeschränkt. Erben (2000, 47) exemplifiziert dieses mit den sogenannten Vorgangskollektiva, die ein wiederholtes Geschehen beschreiben. Schaut man sich den Satz *Es pfeift wiederholt*⁴⁷ an, so hat der Sprecher auch die Möglichkeit, eine Konstruktion mit einem substantivierten Infinitiv zu bilden: *das wiederholte Pfeifen*. Dieser grammatischen Möglichkeit stehen aber auch andere Wortbildungsmuster gegenüber; erstens das Muster *Ge- BV -(e)* und zweitens das Muster *BV -(er)ei*. Bei Verwendung eines dieser Muster wird das Geschehen als negativ und lästig empfunden und erhält somit eine eher negative Konnotation: *das Gepfeife* oder *die Pfeiferei*.

Wie man an diesem Beispiel erkennt, sind die zur Wahl stehenden Wortbildungsaffixe nicht synonym. Bei präfigierten Verben beispielsweise ergibt sich nur eine der beiden Möglichkeiten zur Affigierung, sie sind nur mit *-erei* kombinierbar, sodass hier eine morphologische Beschränkung wirksam ist.

Dieses Phänomen darf als 'komplementäre Distribution' bezeichnet werden: dort, wo das eine Affix nicht anhängbar ist, darf das andere verwendet werden.

Anzumerken ist an dieser Stelle, dass diese Aussage für einen Großteil von Affixen und ihre Verwendungen zutreffend zu sein scheint. Das *Zeugnis* besitzt eine andere Bedeutung als die *Zeugung*, indem ersteres eine gewisse Abgeschlossenheit angibt, während die *Zeugung* durchaus prozesshaft zu beurteilen ist. Dasselbe gilt für *Bedrängnis* versus *Bedrängung* oder auch *Hindernis* und *Hinderung*.

⁴⁷Erben (2000, 47).

2.5 Konkurrenz zwischen Derivationsaffixen

Entwicklungsprozesse von unterschiedlichen Derivationsaffixen laufen parallel zueinander ab.⁴⁸ Dies soll im Folgenden kurz anhand eines historischen Beispiels dargestellt werden.

Gleichzeitig mit der Entwicklung von *-bar* entstehen neue andere Adjektivsuffixe wie *-sam*, *-haft* und *-lich*. Nübling et al. (2006, 79) zeigen, dass *bāri* im Althochdeutschen mit nur 13 Bildungen unproduktiv gewesen ist, während man bei *-lich* 900 Bildungen mit Substantiven und bei *-haft* immerhin 100 Bildungen findet. *-sam* verhält sich mit 47 Bildungen auch nur schwach produktiv. Interessant ist, dass den *bāri*-Bildungen Konkurrenten, nämlich *-lich*, *-sam* und *-haft*, gegenüberstehen: *danclich* versus *dancbāri* oder z.B. *liohtsam* versus *liohtbāri* für 'hell'.

Bāri und *-lich* als Affixoide werden zu Konkurrenten, von welchem sich im Mittelhochdeutschen letztlich *-lich* durchsetzen kann, da bei *-bære* einige morphologische Beschränkungen wirksam sind.⁴⁹ Im Nhd. dagegen entwickelt sich *-bar* dann zu einem produktiven Suffix, *-lich* bleibt produktiv. Die beiden ehemaligen Konkurrenzsuffixe finden sich aber heute jeweils in anderen Bereichen wieder: *-lich* erscheint nach präfigierten Verbalbasen⁵⁰, *-bar*⁵¹ nach einfachen Basen. Außerdem behält *-lich* die Fähigkeit, sich mit nominalen Basen zu verbinden.⁵²

Dass es aber auch in der Gegenwartssprache des Deutschen Parallelkonstruktionen gibt, die mit unterschiedlichen Suffixen gebildet werden, zeigen Fleischer und Barz (1995, 267) in ihrer Aufstellung zu den semantischen Relationen von adjektivischen Suffigierungsmodellen:

⁴⁸Der Systemgedanke ist für die vorliegende Arbeit wichtig, denn er wurde bisher nicht in die Produktivitätsforschung integriert. Das Wortbildungssystem ist eine „organisierte Gesamteinheit verschiedener miteinander verbundener Subsysteme“ (Uluchanov 1997, 25), das aus produktiven Modellen (Komposition, Derivation) und einem Inventar an Wortbildungsmorphemen besteht. Mit Blick auf den Systemgedanken kommt die Frage auf, ob das Derivationssystem als System von 'kommunizierenden Säulen' [Anmerkung: Der Begriff wurde in Analogie zu den 'kommunizierenden Säulen' in der Physik gebildet.] beschrieben werden kann: wird das eine Affix produktiver, so schwindet vielleicht die Produktivität eines anderen Suffixes und vice versa.

⁴⁹Es verbindet sich im späten Mhd. nur mit Substantiven und nur in Ausnahmefällen mit Adjektiven. Außerdem existieren viele austauschbare Bildungen wie z.B. *genisbære* und *genislich*, was heilbar bedeutet. Zur Erläuterung von morphologischen Beschränkungen siehe Kap. 2.4.2.

⁵⁰Z.B. in *unglaublich*.

⁵¹Z.B. in *machbar*.

⁵²*Brüderlich*, *schwesterlich* etc.

- Deverbativa mit passivischer bzw. nichtpassivischer Beziehung auf das zugehörige Substantiv: *begehrbar/begehrlich, verantwortbar/verantwortlich*
- Desubstantiva, das Vorhandensein bzw. Nichtvorhandensein des durch die Basis Ausgerückten bezeichnend: *skrupelhaft/skrupellos, schwunghaft/schwunglos, waldig/waldlos, fleckig/fleckenlos*
- Ableitungen von Stoffbezeichnungen, das Bestehen aus diesem Stoff bzw. einen Vergleich ausdrückend: *gläsern/glasig, seiden/seidig*
- Ableitungen von Personenbezeichnungen, eine Beziehung ohne bzw. mit pejorativer Expressivität ausdrückend: *kindlich/kindisch, weiblich/weibisch, bäuerlich/bäuerisch*
- Zeitangaben der Dauer bzw. regelmäßigen Wiederkehr: *viertelstündig/viertelstündlich, zweiwöchig/zweiwöchentlich*
- freies Adjektiv bzw. Relativadjektiv nur in Zusammensetzungen: *krimineller Assistent/Kriminalassistent*

Anhand der Beispiele aus Fleischer und Barz (1995) lässt sich erkennen, dass zwar einige Parallelkonstruktionen für das Deutsche existieren, sich diese allerdings in den meisten Fällen allerdings semantisch leicht unterscheiden oder antonym zueinander sind. Dennoch lassen sich auch Beispiele finden, die sich semantisch sehr nahe stehen: *Kindlein* versus *Kindchen* oder *schmerzhaf* versus *schmerzvoll*.

Im Rahmen dieses Abschnittes wurden einige wichtige Phänomene skizziert, die die Wortbildung betreffen. Es wurde deutlich gemacht, welche Funktion die Wortbildung im Deutschen besitzt, welche Einheiten zur Wortbildung beitragen, wie der Wortschatz durch die Triebkräfte erweitert wird und welche Konsequenzen sich daraus ergeben, wenn Affixe miteinander konkurrieren.

Dennoch ist es schwierig, wissenschaftlich begründete Voraussagen darüber zu treffen, welche Wortbildungen wahrscheinlich sind und welche von ihnen eine hohe Akzeptabilität erlangen. In diesem Zusammenhang macht Erben (2000, 50) darauf aufmerksam, dass eine „umfassende Materialbasis“ notwendig ist, um statistische Aussagen über Wortbildungsprodukte zu machen. Er spricht von einer „Wechselbeziehung der Prozesse und der Ergebnisse, die selbst wieder zu Bedingungen neuer Prozesse werden“.

Überdies fügt er hinzu, dass auch unproduktive Wortbildungsstrukturen nicht unbeachtet bleiben dürfen, „da die Wahrscheinlichkeit eines weiteren Ausbaus der Reihe

nicht auszuschließen ist und sie allein durch ihre Kopräsenz im Sprachschatz als potentiell Mittel wirken oder das eventuelle Üblichwerden anderer Morphemgefüge hemmen können“ (ebd.).

Der Forderung Erbens wird in dieser Arbeit gefolgt, indem die umfassende Materialbasis eines großen deutschen Korpus (DWDS) verwendet wird, um statistische Aussagen über die Produktivität deutscher Derivationssuffixe zu treffen. Zudem wird auch der Empfehlung nachgegangen, unproduktive Wortbildungsstrukturen zu beachten.

Grammatikalisierung und Wortbildungswandel

Im nachfolgenden Abschnitt werden weitere theoretische Überlegungen gemacht und diskutiert, die im Zusammenhang mit Produktivität stehen. Dazu wird zunächst die Grammatikalisierungstheorie erläutert, die zur Beschreibung von Veränderungen in der Grammatik herangezogen wird. Mithilfe der Grammatikalisierungstheorie kann der Weg eines sprachlichen Zeichens von der Lexik in die Grammatik anhand von verschiedenen Stufen beschrieben werden.

Für die vorliegende Arbeit ist entscheidend, inwieweit die Produktivitätsentwicklung bei Grammatikalisierungsprozessen eine Rolle spielt.

Im Anschluss daran wird der Begriff des Wortbildungswandels eingeführt, da auch dieser im Zusammenhang mit diachronen Veränderungen in der Grammatik und im Wortbildungssystem steht. Anhand eines Forschungsberichtes werden die bereits zu den Thema Wortbildungswandel im Deutschen existierenden Arbeiten vorgestellt. Hier gilt es darzustellen, welche Ideen und Konzepte im Hinblick auf die Produktivitätsforschung bereits entwickelt worden sind. Deutlich werden wird insbesondere, dass eine (empirische) Forschungslücke existiert, die mit dieser Arbeit verkleinert werden soll.

3.1 Grammatikalisierungstheorie

Einen Linguisten, der sich mit Affixen beschäftigt, wird auch ihre Entstehung interessieren. Aus diesem Grund wird er zweifelsohne nicht daran vorbeikommen, sich mit dem Begriff und dem dahinter stehenden Konzept zu beschäftigen: **Grammatikalisierung**. Welcher Zusammenhang zwischen der Entstehung von Affixen und ihrer Produktivität besteht, soll nunmehr in diesem Abschnitt deutlich gemacht werden.¹

Die Entstehung von Affixen ist ein Wandelprozess, bei dem ein Element seinen lexikalischen Status verliert und zu einer grammatischen Kategorie wird (Diewald 1997, Girnth 2000, Heine 2003, Hopper und Traugott 2003, Lehmann 2005). Die Grammatikalisierung darf als Prozess beschrieben werden, bei dem die Grammatik durch sogenannte Grammemen angereichert wird.² Unter einem Grammem versteht man in der Linguistik ein grammatisches Morphem. Dazu zählen sowohl grammatisch gebundene Morpheme als auch grammatisch freie wie z.B. Funktionswörter, zu denen beispielsweise Präpositionen und Artikel gehören. Dem Grammem steht dabei das Lexem gegenüber.

Durch Grammatikalisierungsprozesse kommt es zu Veränderungen in der Grammatik. Der Weg eines lexikalischen Elements zu einem grammatischen ist in der Regel

¹Obwohl es klare Gemeinsamkeiten zwischen der Entwicklung von Derivationsuffixen und dem Prozess der Grammatikalisierung gibt, setzen nicht alle Linguisten die Prozesse einfach gleich. Als Hauptgrund der Unterscheidung nennen Nübling et al. (2006, 72) „den Grad an semantischer Reduktion“. Die Entwicklung der Derivationsuffixe schreitet schneller voran als die Grammatikalisierung von Flexionsmorphemen. Bei einer Umwandlung eines Wortes zu einem Flektiv kommt es auch zu einer Umwandlung der lexikalischen Bedeutung: Zunächst wird das freie Lexem zu einem freien grammatischen Morphem, dann zum Klitikon und letztendlich zum Flektiv. Als Beispiel sei hier auf ein Beispiel aus Lehmann (1989, 12) verwiesen: die Umwandlung des Althochdeutschen *haidus* in das mittelhochdeutsche Suffix *-heit* bezeichnet er als Lexikalisierung. In der vorliegenden Arbeit wird die Terminologie der Grammatikalisierung auch im Bezug auf Wortbildungswandel angewandt, die u.a. auch Wischer (2008) vertritt.

²Der Terminus der Grammatikalisierung wird seit über 100 Jahren verwendet; er findet sich bereits 1912 in einem Aufsatz von A. Meillet, *L'évolution des formes grammaticales*. Für Meillet (1948, 131) ist die Grammatikalisierung 'l'attribution du caractère grammatical à un mot jadis autonome.' Er geht auch von einem diachronen Prozess aus und unterscheidet in seinem Aufsatz drei Wortklassen: *mots principales*, *mots accessoires* und *mots grammaticales*. Zwischen diesen Wortklassen bestehen für ihn graduelle Unterschiede, die er wie folgt beschreibt:

„L'affaiblissement du sens et l'affaiblissement de la forme des mots accessoires vont de pair; quand l'un et l'autre sont assez avancés, le mot accessoire peut finir par ne plus être qu'un élément privé de sens propre, joint à un mot principal pour en marquer le rôle grammatical. Le changement d'un mot en élément grammatical est accompli.“

ein sehr langer Weg mit graduellen Abstufungen, der mehrere hundert Jahre andauern kann, da unterschiedliche Phasen durchlaufen werden. Heine (2003, 579) unterscheidet vier solcher Phasen:

- Desemantisierung
- Extension
- Dekategorisierung
- Erosion

Die Desemantisierung bezeichnet die Phase, in der eine noch lexikalische morphologische Einheit die ersten Wandelerscheinungen aufweist, indem ein Verlust an semantischem Inhalt vorzufinden ist. Vereinfacht gesagt, wird der Inhalt demnach abstrakter. Hopper und Traugott (2003, 100) sprechen in diesem Zusammenhang auch von einer Bedeutungsverallgemeinerung oder einer semantischen Ausbleichung.

In der zweiten Phase, der Extension, kommt es zu einer Verwendung des sprachlichen Zeichens in bisher nicht verwendeten Kontexten. Dennoch sind Phase 1 und Phase 2 nicht unabhängig voneinander, da die Desemantisierung erst nach Abschluss der Extension deutlich wird.

Die dritte Phase der Grammatikalisierung, die Dekategorisierung, zeichnet sich durch eine Veränderung auf morphosyntaktischer Ebene aus, was sich dadurch erklären lässt, dass die zu verändernde Einheit zunehmend in mehreren Kontexten zu finden ist und damit die Häufigkeit, sich an andere Wörter zu binden, ansteigt. Dabei kann es dann zu einem Statusverlust als unabhängiges Wort kommen, weil eine Verschmelzung an andere Wörter erfolgt.

Das letzte Stadium auf dem Weg zur grammatisch abstrakten Einheit ist die Phase der Erosion. Mit der Erosion ist eine Abschwächung oder ein totaler Schwund der Lautsubstanz gemeint. Die Grammatikalisierung erfolgt demnach auch auf der phonologischen Ebene.

Die Phase der Erosion ist für einen Grammatikalisierungsprozess jedoch keineswegs obligatorisch (Nübling et al. 2006, 224). Ferner besteht die Möglichkeit, dass eine Desemantisierung stattfindet, ohne dass weitere Grammatikalisierungsschritte folgen. Als Beispiel kann hier das Morphem *Höllen-/höllen-* genannt werden, das sich in den letzten Jahren stark verändert hat. So lassen sich im Gegenwartsdeutschen Wörter wie

*Höllenhitze*³ oder *Höllspreis*⁴ finden: Während bei der *Höllenhitze* durchaus assoziiert werden kann, dass die negative Konnotation der Hölle weiterhin Bestand hat, so ist unklar, inwieweit es sich bei einem *Höllspreis* um einen besonders hohen (negativ) oder niedrigen (positiv)⁵ Preis handelt.

Lehmann (1995) (Abb. 3.1) entwirft ein Modell, das auf Givón (1971) basiert und untergliedert den Prozess der Grammatikalisierung in verschiedene Phasen, in denen das Sprachzeichen langsam an Selbstständigkeit verliert und in einen grammatischen Kontext eingebettet wird. Am Ende dieses Prozesses bleibt entweder ein Grammem oder ein abstraktes Konzept zurück wie z.B. 'Präteritum'.

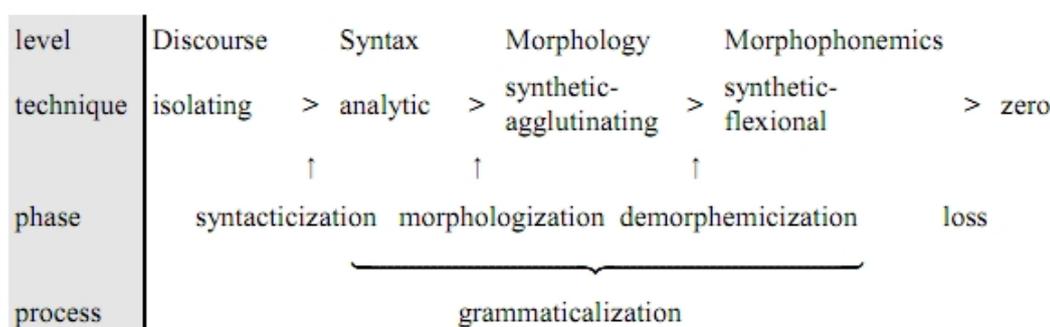


Abbildung 3.1: Phasen der Grammatikalisierung nach Lehmann (1995, 12)

Als erste Stufe zur Grammatikalisierung beschreibt Lehmann die Phase der Syntaktisierung. Exemplifizierend für dieses Phänomen kann die deutsche Verbstellung genannt werden, die durch das sogenannte „Topologische Feldermodell“ determiniert ist.⁶ Im Deutschen ist die Position des Verbs durch dieses Modell vorgegeben, sodass ein Sprecher stets entscheiden muss, ob er einen V1-, einen V2- oder einen VL-Satz bildet. Dementsprechend muss dann eine richtige Platzierung des Verbs erfolgen. Die

³Dazu siehe <http://presstext.de/news/050701021/hoellenhitze-mittelmeer-duestere-klimaprognose-fuer-urlaubsziele/>, abgerufen am 31.07.2010 um 12:18 Uhr.

⁴Zu finden auf <http://blog.zeitenwende.ch/hansruedi-ramsauer/steigende-kreativitaet-bei-den-bankbilanzen.>: „Mittlerweile dürfte diese Verfahren insbesondere bei den ABS und seinen Angehörigen vielfach kaum noch möglich sein, weil entweder Höllenpreise gestellt werden oder gar keine Preise mehr, zumindest aber keine Preise mehr, zu denen auch jemand zum Handel bereit ist.“

⁵Siehe dazu zum Beispiel <http://aktuell.meinestadt.de/geisenheim/2011/05/05/guenstige-vvk-tickets-fuer-hoellen-mitglied/>: „Wer Mitglied in der KulturHölle e.V. ist, der erhält auf alle Tickets der Vereinsveranstaltungen für Erwachsene einen Rabatt von 25%. Der derzeitige Endpreis liegt aktuell bei 18 Euro (aufwendigen Highlights 20 Euro). Höllenpreis also: 15 Euro.“

⁶Zur Geschichte des topologischen Feldermodells siehe Drach (1937) und für weitere Informationen siehe Altmann (1981).

Syntaktisierung darf als Positionsfestigung eines sprachlichen Zeichens im Satz beschrieben werden.

Die zweite Phase bezeichnet Lehmann als Morphologisierung, die er synonymisch zum Begriff der *Agglutinierung* verwendet. Die Morphologisierung „reduces the analytic construction to a synthetic one, so that grammatical formatives become agglutinating affixes.“⁷

In dieser Phase findet eine Veränderung auf der morphosyntaktischen Ebene statt, indem z.B. Wörter miteinander verschmelzen; dazu gehört auch das Phänomen der Univerbierung, bei welcher syntaktische Konstruktionen zu zwei- oder mehrgliedrigen Wörtern miteinander verbunden werden. Soll die Entstehung von Affixen beschrieben werden, so kann festgehalten werden, dass sie mit der Morphologisierung und der Fusion und der damit verbundenen Phase der Entwicklung einer grammatischen Funktion beginnt.

Nach Lehmann folgt daraufhin eine Phase der Veränderung von der Ebene der Morphologie zur Morphophonologie, die er als Demorphemisierung bezeichnet.⁸ Diese Phase ist gekennzeichnet durch eine phonologische Anpassung eines Affixes an einen Stamm. Anschaulich lässt sich dieses Phänomen durch Affixe, die in Flektive übergehen, erklären: im Althochdeutschen existierte die Pluralendung *-i* wie z.B. in **gast-i* (Gäste), die im Laufe der Zeit mit dem Stamm fusionierte, sodass eine Umlautung des Stammvokals vollzogen und aus *gasti gest-i* wurde. Im Gegenwartssdeutschen wird der Plural des Lexems *Gast* sogar an zwei Stellen gleichzeitig ausgedrückt: sowohl durch eine Veränderung des Stammvokals als auch durch eine Suffigierung (*Gäste*). Zudem ist eine Abschwächung des Endvokales erfolgt.

Die letzte Phase der Grammatikalisierung, die Lehmann als *loss* bezeichnet, ist geprägt von dem Schwund eines sprachlichen Zeichens. So z.B. kann es sein, dass ein Flektiv im Laufe der Zeit verschwindet.

Szczepaniak (2009, 18) exemplifiziert dies durch den Abbau der einstigen Pluralendung *-e*, die nach zweisilbigen Substantivstämmen geschwunden ist. Während es im Mittelhochdeutschen noch *engel* (Sg.) und *engele* (Pl.) heißt, existiert im Neuhochdeutschen nur noch eine einzige Form für Singular und Plural zugleich: *Engel*. Die deutsche Pluralform markiert seinen Plural durch ein Nullmorphem. Wird eine gram-

⁷Siehe dazu Lehmann (1995b, 12).

⁸Vgl. dazu aber auch Givon (1971), der diese Phase als Lexikalisierung bezeichnet.

matische Kategorie abgebaut, so kommt es zu einem Schwund aller Allomorphe dieses Morphems.

Doch auch in Bezug auf Derivationsuffixe lässt sich ein Beispiel zur Anschauung heranziehen. Betrachtet man beispielsweise das Suffix *-bar*, wird schnell deutlich, dass es im Laufe der deutschen Sprachgeschichte sehr produktiv geworden ist.⁹ Mit diesem Derivationsmorphem lässt sich eine Vielzahl an Adjektiven bilden (*ableitbar, wunderbar, machbar,...*). Jedoch lässt sich heutzutage nicht mehr erkennen, dass dieses Suffix auch mit *Geburt* oder *Bahre* verwandt ist. Das Suffix *-bar* geht – wie bereits in Kap. 2.5 beschrieben – auf das althochdeutsche Adjektiv *bāri* zurück.

Ähnlich ist auch die Entwicklung von *-lich*, *-tum* oder *-heit* anzusehen, die aus selbstständigen Wörtern entstanden sind. Der Übergang eines Wortes ist ein gradueller Prozess, bei dem in einem Zwischenstadium sogenannte Affixoide oder Halbsuffixe entstehen.¹⁰ Sie können als Brücke zwischen Wort und Affix angesehen werden¹¹ (Abb. 3.2).

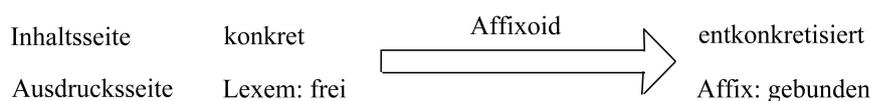


Abbildung 3.2: Das Affixoid als Brücke zwischen Lexem und Affix. Modifiziert nach Nübling et al. (2007, 71)

Spricht man von Grammatikalisierung, so wird angenommen, dass es sich dabei um einen **unidirektionalen Prozess** handelt. Der unidirektionale Weg eines Sprachzei-

⁹Zur Geschichte dieses Suffixes siehe Nübling et al. (2006).

¹⁰Stevens (2005, 19) schlägt fünf Kriterien für Affixoide vor. Sie lauten wie folgt: 1. *Koexistenz*: Affixoide koexistieren immer mit einem identischen freien Morphem, 2. *Abstraktheit*: Affixoide sind semantisch abstrakt und können auch sehr spezielle Verwendungen besitzen, 3. *Nicht-Determinierend*: Die Bedeutung des Gesamtwortes wird durch das andere Element bestimmt und nicht durch das Affixoid. Als Beispiel kann man hier *saukalt* nennen. Insgesamt trägt das Wort die Bedeutung 'besonders kalt', 4. *Konkurrenz und komplementäre Verteilung*: Affixoide stehen immer in Konkurrenz oder aber auch in komplementärer Verteilung. Exemplifizieren lässt sich dies an den von Ascoop (2005) genannten Beispielen: *textitgefahr-los* (Affixoid) versus *un-gefährlich* (Affix), *sehnsuchtsvoll* (Affixoid) versus *sehnsücht-ig* (Affix), 5. *Reihenbildung*: Es lassen sich – und hier besteht eine Parallelität zum Affix – viele neue Wörter bilden (*saukalt, sauwarm, sauteuer,...*).

¹¹Der morphologische Status des Affixoids ist unklar und strittig. Zur Diskussion dieses Problems siehe Ascoop (2005). Es wird angemerkt, dass es kein Wortbildungsverfahren gibt, das man als *Affixoidierung* bezeichnen würde und der Komposition und der Derivation zur Seite stellen könnte, dass Affixoide allerdings als Komposita analysierbar und als Ergänzung eines Derivationsmusters angesehen werden.

chens von der Lexik in die Grammatik wird in der Literatur auch mit der Metapher der Einbahnstraße verbildlicht (Norde 2002, 45).

Das Prinzip der Unidirektionalität schließt eine Degrammatikalisierung aus. Dennoch darf diese Annahme auch kritisch betrachtet werden, da im Deutschen auch Entwicklungen in die andere Richtung existieren: Beispielsweise gibt es im Gegenwartsdeutschen das Phänomen, dass ursprünglich stark gebundene Wortbildungsmorpheme wie *bio*-¹², *psycho*-¹³ und *öko*-¹⁴ frei verwendet werden, obwohl für Konfixe Gebundenheit als oberstes Kriterium gilt (Elsen 2005).

Und genau dieser Aspekt der Unidirektionalität sollte auch im Hinblick auf Produktivität und den Produktivitätsveränderungen nicht unbeachtet bleiben, denn der Begriff der Unidirektionalität lässt zunächst annehmen, dass ein Zeichen auf dem Weg in die Grammatik an Abstraktheit zunimmt. Die Zunahme an Abstraktheit impliziert eine Zunahme an Produktivität, indem das Zeichen auf seinem Grammatikalisierungsweg die Fähigkeit, sich an Stämme zu binden, ausweitet und somit Neubildungen schafft.¹⁵

Sollte die empirischen Untersuchungen dieser Arbeit zeigen, dass es Affixe gibt, deren Produktivität stark schwankt, so ist zu überlegen, wie sich dieses Phänomen mit der Unidirektionalität des Grammatikalisierungsprozesses vereinbaren lässt.

3.2 Morphologischer Wandel

Im Rahmen der vorliegenden Arbeit wird der derivationelle Wandel an Wortbildungssuffixen untersucht. Zu diesem Forschungsbereich liegen, wie der nachfolgende Forschungsbericht deutlich machen wird, bisher nur sehr wenige Arbeiten vor. Dennoch

¹²Vgl. „Wo fängt Bio an und wo hört Bio auf?, fragt Schmitz.“, in: Rhein-Zeitung, 29.07.2002.

¹³Z.B. in: „oh man, das ist voll psycho!“, abgerufen unter: <http://www.gutefrage.net/frage/lie-d-wie-jede-zelle-meines-koerpers-ist-gluecklich> am 07.09.2009 um 12:15 Uhr.

¹⁴Vgl. dazu „Und plötzlich sind alle öko...“, in: Nürnberger Nachrichten, 24.06.2008.

¹⁵Ein Beispiel des Gegenwartsdeutschen ist *-frei*. Adjektive wie *alkoholfrei* oder *fettfrei* sind eindeutig zu interpretieren: *Alkoholfreies* Bier enthält keinen Alkohol, *fettfreier* Joghurt kein Fett. Schwieriger wird es, wenn man *scheinfrei* oder *bügelfrei* interpretiert, denn *scheinfrei* ist ambivalent. Im ersten Fall ist es möglich, einen Kontext zu finden, in welchem man davon ausgeht, dass eine Person keine Scheine besitzt, im anderen Fall spricht man von *Scheinfreiheit*, wenn ein Student alle Scheine vorlegen kann, die er zur Anmeldung seines Studienabschlusses benötigt. Demnach ist ein Student, der *scheinfrei* ist, im engen Sinne nicht frei von Scheinen. In diesem Fall liegt demnach eine semantische Verschiebung und eine Zunahme der Abstraktheit vor.

existieren zahlreiche Arbeiten, die sich mit der Thematik des morphologischen Wandels¹⁶ und/oder dem Versuch, eine Theorie dafür zu entwickeln¹⁷, beschäftigen.

Ein Versuch, den Wortbildungswandel auf theoretischer Ebene zu beschreiben, kommt von Munske (2002, 23) (Abb. 3.3), der sich zum Ziel gesetzt hat, „eine erste Übersicht über die Haupttypen des Wortbildungswandels zu geben“.

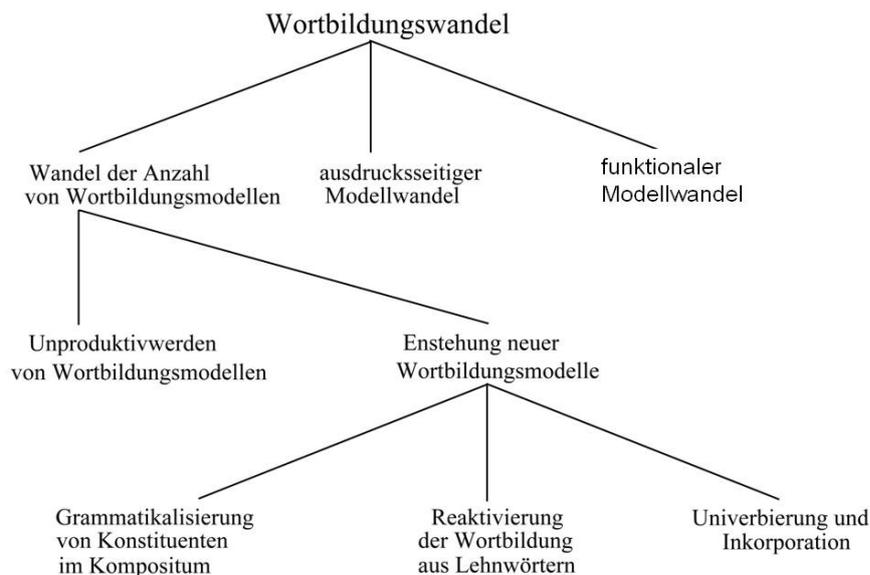


Abbildung 3.3: Munskes Wortbildungstypen (2002, 26)

Zunächst einmal unterscheidet Munske zwischen dem *funktionalen* und dem *ausdrucksseitigen Modellwandel* und einem *Wandel der Anzahl von Wortbildungsmodellen*.

Unter dem ausdrucksseitigen Modellwandel fasst er z.B. Phänomene zusammen, die den phonologisch motivierten Schwund beschreiben und auf der anderen Seite gleich-

¹⁶Betrachtet man die Veröffentlichungen zum *morphologischen Wandel* genauer, so wird schnell deutlich, dass gerade der flektivische Wandel beschrieben wird und selten der derivationale (Anderson 1980).

¹⁷Joseph (1998, 366) hält eine Theorie zur Vorhersage des morphologischen Wandels für problematisch und macht darauf aufmerksam, dass es dabei höchstens Tendenzen geben kann:

„Thus, it may well be that for morphological change, a general theory - that is, a predictive theory - is not even possible, and that all that can be done is to catalogue tendencies, which, however valid they may be, do not in any sense constitute inviolable predictions about what types of changes will necessarily occur in a given situation.“

zeitig aber eine Suffixerweiterung implizieren. Als Beispiel kann das althochdeutsche Suffix *-ari* genannt werden, das sich im Laufe der Zeit zu *-er* entwickelt hat und im Gegenwartsdeutschen als hoch produktiv gilt (Fleischer und Barz 1995, 151).

Weiterhin gibt es nach Munske auch einen funktionalen Wandel. Realisiert wird dieser beispielsweise durch das Entstehen von Fugenelementen aus einem Genitiv-S, denn diesbezüglich liegt eine Veränderung eines Flektionselementes zu einem Wortbildungsmorphem vor; die Funktion des sprachlichen Zeichens hat sich verändert. Ebenso ist der Bedeutungswandel als funktionaler Wandel zu klassifizieren. Er tritt genau dann auf, wenn z.B. das Suffix *-ung* im Gegenwartsdeutschen nicht nur Handlungen beschreibt, sondern auch Personen (z.B. *Schulleitung*) bezeichnen kann.

Doch der für die vorliegende Arbeit interessante Teil seines Modells ist der folgende: Munske macht darauf aufmerksam, dass ein Wechselspiel zwischen dem Unproduktivwerden von bestimmten Wortbildungsmodellen und dem Ansteigen der Anzahl jener Wortbildungsmodelle existiert, die das Umwandeln von freien Stämmen zu Affixen schaffen und begründet dies wie folgt:

„Das Aussterben von Wörtern hat bisher weit weniger Aufmerksamkeit gefunden als deren Vermehrung. Ähnlich ist auch der Untergang von Wortbildungsmodellen ein rechtes Stiefkind historischer und diachroner Wortbildungsforschung. Stets hat das Neue und haben die Regeln der Innovation mehr Interesse gefunden. Bevor man jedoch erkennt, was untergegangen ist, muß beschrieben werden, was es gab. Am ehesten erkennt man die untergegangenen Präfixe und Suffixe, obwohl eben diese Ausdrucksweise fragwürdig ist. Denn die Affixe wie z.B. *-nis*, *-sal* und *-icht* leben in Bildungen wie *Finsternis*, *Labsal* und *Dickicht* fort. Untergegangen sind nur die betreffenden Modelle als produktive Muster.“¹⁸

Munske erklärt, dass das Untergehen von bestimmten Wortbildungsmodellen mit der Konkurrenz anderer Modelle, die den abschwächenden irgendwann überlegen sind, zu begründen ist. Die alten Bildungen bleiben stark lexikalisiert erhalten.

Des Weiteren fügt er hinzu, dass das Untergehen eines Modells ein sehr langsamer Prozess ist, „der auch manchmal revidiert werden kann“ (Munske 2002, 27).

Auch diese Annahme ist für diese Arbeit höchst interessant, denn diese These lässt einen Interpretationsspielraum zu; die These impliziert nämlich, dass Produktivitäts-

¹⁸Munske (2002, 27).

schwankungen möglich sind, sodass zu fragen ist, inwieweit man in solchen Fällen überhaupt von Unidirektionalität, die für Grammatikalisierungsprozesse grundlegend sein soll, sprechen kann.

Nach Munske ist der Wortbildungswandel sehr komplex, denn ebenso können auf syntaktischer Ebene Veränderungen vorliegen, indem freie Lexeme zu gebundenen werden.

Neben der Grammatikalisierung kann es auch zu einer Reaktivierung der Wortbildung aus alten Lehnwörtern kommen. Geschieht dieses, indem beispielsweise bestimmte Konfixe produktiv werden, so liegt hier ein Wandel der Konfixproduktivität vor. Exemplifizieren lässt sich dieses am Konfix *Bio-*, das im Gegenwartsdeutschen sehr produktiv ist und, wie schon vorab erwähnt, ein Trend der Degrammatikalisierung besteht.

In seinem Modell macht Munske zudem deutlich, dass die Prozesse Univerbierung und Inkorporation als Möglichkeiten für eine Entstehung neuer Wortbildungsmuster in Frage kommen.¹⁹

Kritik an Munskes Modell kommt von Scherer (2006), die bemängelt, dass das Aussterben eines Wortbildungsmusters eine Senkung der Produktivität auf Null impliziert. Dies bedeutet, dass ein Anstieg von derivationaler Morphologie durch Entstehung neuer Wortbildungsaffixe zu einer Eliminierung andere Modelle führen kann.

Auch lässt sich anmerken, dass Munskes Modell viele unterschiedliche Teilgebiete der Wortbildung beinhaltet: das Wortbildungssystem, die Wortbildungsmuster, aber auch die Produkte der Wortbildung.

Scherer (2006) fordert eine Trennung der unterschiedlichen Prozesse. Um den Wortbildungswandel theoretisch zu beschreiben, schlägt sie vor, die Ebene der Wortbildung zu bestimmen, die von dem Wortbildungswandel betroffen ist; in Frage kommen die

- Wortbildungskonstituenten,
- die Wortbildungsmuster,
- die Wortbildungsprodukte und

¹⁹Als Univerbierung bezeichnet man den „Vorgang und das Ergebnis des Zusammenwachsens mehrgliedriger syntaktischer Konstruktionen zu einem Wort“ (Bußmann 1990, 819). Als Beispiel kann das Wort *Langeweile* genannt werden, das seinen Ursprung in der Phrase *lange Weile* hat. Eisenberg (1998, 316ff) spricht von Inkorporation, wenn ein Element in eine Wortbildungsstruktur eingliedert wird. Er nennt auch Beispiele für einen solchen Prozess, zu dem beispielsweise die Konstruktion Substantiv + Verb (eislaufen) mit Inkorporation eines Objekts gehört.

➤ das Wortbildungssystem.

Ihrer Meinung nach sind drei unterschiedliche Arten des Wortbildungswandels voneinander zu differenzieren:

1. Veränderungen im Wortbildungssystem
 2. Der Wandel an mehreren Komponenten (Schnittstellenphänomen)
 3. Phonologischer und semantischer Wandel als Wandel an Wortbildungsprodukten
- Diese Teilung der Ebenen bedarf einer genauen Skizzierung.

Der Wandel am Wortbildungssystem und das damit verbundene Wechselspiel in verschiedenen Bereichen der Grammatik wird nicht nur an Munks Modell deutlich. Zu ähnlichen Annahmen kommt nämlich auch Müller (2002, 7), der zehn Sprachwandelerscheinungen auflistet:

1. Veränderungen im Inventar der heimischen und entlehnten Wortbildungselemente
2. Veränderungen bei Wortbildungsmustern, also etwa der Wandel von desubstantivisch gebildeten Adjektiven mit dem Suffix *-bar* zu deverbalen wie *machbar*
3. Veränderungen im Funktionspotential von Affixen einschließlich der Fokussierung bestimmter Wortbildungsbedeutungen
4. Die Reduzierung semantischer Konkurrenzbildungen wie der Lexeme *Abschneidung* und *Abschnitt* aufgrund des Veraltens oder der Bedeutungsdivergenz einer Bildung
5. Der Rückgang von Wortbildungen mit pleonastischem – also funktionslosem – Affix wie bei *Geschrift* (neben *Schrift*) oder *Feldung* (neben *Feld*)
6. Die Reduzierung allomorphischer Variation
7. Der Abbau sog. Halbkomposita des Typs *der tochter mann* als morphologische Mischgebilde, der aus der zunehmenden Nachstellung von Genitivattributen sowie einer konsequenteren Zusammenschreibung von Komposita resultiert.
8. Idiomatisierungsprozesse
9. Grammatikalisierungsprozesse und die Entwicklung von Wortbildungsmorphemen aus Lexemen
10. Die Entstehung von Pseudomorphemen wie *Him-* in *Himbeere* oder *Schorn-* in *Schornstein* als Zeugnisse einer Fossilisierung

Die Betrachtung vom Müllers vorgeschlagenen Punkten zeigt auf, dass die Punkte 2-10 direkt mit Punkt 1 in Verbindung stehen und eine Veränderung im System beschreiben. Demnach lässt sich daraus ableiten, dass Veränderungen von Wortbildungsmustern zu einer **Veränderung des gesamten Wortbildungssystem** führen.

Zur Phänomenen an einer Schnittstelle lässt sich Folgendes erklären: Sowohl quantitative als auch qualitative Veränderung im System der Wortbildung führen dazu, dass Einheiten umgeformt werden werden. Exemplifizierend können an dieser Stelle die Arbeiten von Habermann et al. (2002) und Wellmann (1997) genannt werden, die die Herausbildung von Affixen oder Fugenelementen aus Flektiven erklären und damit einen Austausch zwischen Lexikon und Wortbildungskomponente skizzieren (vgl. Scherer 2006). Dabei handelt es sich nach Scherer vordergründig um eine Art des lexikalischen Wandels:

„Da freie und gebundene Wortbildungskonstituenten sowie Wortbildungsprodukte Einheiten des Lexikons darstellen, Wortbildungsmuster hingegen Einheiten der Wortbildungskomponente, sind Prozesse, die die morphologische Struktur eines Wortes verändern, primär als lexikalischer Wandel einzuordnen.“²⁰

Doch auch die von Munske (2002) genannten Punkte der Univerbierung und Inkorporation finden an einer Schnittstelle statt, an der Schnittstelle zur Syntax. So z.B. entstanden deutsche Genitivkomposita wie *Frauenkirche* aus Nominalphrasen mit Genitivattribut. Bei der Inkorporation kann man davon ausgehen, dass syntaktische Konstruktionen wie z.B. ein Verb gefolgt von einem direkten Objekt in eine einzige Wortbildungskonstruktion verschmelzen: ([Rad NP fahren V]VP) wird zu *Radfahren*.

Wortbildungswandel im engen Sinne jedoch ist für Scherer der phonologische und semantische Wandel als Wandel an den Produkten der Wortbildung. Dieser Ansicht ist zuzustimmen: Wortbildungsprodukte besitzen die inhaltlichen Merkmale von freien Morphemen, da sie einer Wortart angehören, eine bestimmte Lautgestalt besitzen und in der Regel flektierbar sind.²¹ Sie sind Einheiten des Lexikons und können sich durch Bedeutungsverschiebungen²² und Lautwandel verändern. Veränderungen können nicht

²⁰Scherer (2006, 17).

²¹Ausnahmen bilden im Deutschen zum Beispiel Adverbien wie *ungern*.

²²Diese Wandelerscheinung wurde bereits für das Deutsche erforscht. Demske (2000) stellt fest, dass es bei der *-ung*-Derivation im Laufe der Zeit Veränderungen gegeben hat; sie untersucht diese Derivation ab dem Frühneuhochdeutschen und kommt zu dem Schluss, dass es im Gegenwartsdeutschen semantische Beschränkungen gibt, die zu einem Produktivitätsverlust führen.

nur den Input (die Basis des Derivats), sondern auch den Output betreffen. In diesem Fall verweist Scherer auf Müller:

„Müller (2002) nennt unter anderem den Aus- und Abbau von Wortbildungsbedeutungen, der sich beispielsweise in der Fokussierung bestimmter Wortbildungsbedeutungen, der Reduzierung semantischer Konkurrenzbildungen und allomorphischer Variation sowie dem Rückgang von pleonastischen Wortbildungen mit funktionslosem Affix wie bei *Feldung* (neben *Feld*) äußert.“ (Scherer 2006, 14).

Daraus leitet Scherer die These ab, dass es **formale und inhaltliche** Beschränkungen gibt, die sich auf die Konstituenten (Input) und auf die Produkte der Wortbildung (Output) beziehen können. Liegen diachrone Veränderungen der Restriktionen vor, so lässt sich von einem Wortbildungswandel sprechen.

3.3 Stand der Forschung in der (germanistischen) Linguistik

Einige Arbeiten aus der germanistischen Sprachwissenschaft, die sich mit dem Phänomen des Wortbildungswandels beschäftigen, wurden im vorangegangenen Abschnitt bereits genannt.

Für das Deutsche existieren allerdings auch Arbeiten, die sich speziell mit dem **Wandel von Affixen** befassen. Bei vielen dieser Arbeiten handelt es sich um qualitativ ausgerichtete Studien wie z.B. die Fallstudien von Kiese Wetter (1987; 1991) und Demske (2000).

Quantitativ ausgerichtete Studien liegen (bisher) nur in geringem Maße vor.²³ So z.B. untersucht Scherer (2006) den Wandel des *-er*-Suffixes im Laufe eines großen Zeitraums (1600 - 2000) und findet dabei heraus, dass sich dieses Suffix im Laufe der Zeit verändert und produktiver wird. Sie konzentriert sich in ihrer Studie auf die semantische Seite und beleuchtet die Veränderungen an der Wortart der Basis.

²³Für das Chinesische existiert eine erwähnenswerte Studie, die quantitativ ausgerichtet ist: Eiji Nishimoto (2003, 56 ff.) untersucht und vergleicht im Sinica Corpus die morphologische Produktivität von fünf chinesischen Suffixen, *-men*, *-hua*, *-r*, *-zi* und *-tou*. Er verwendet sowohl ein tokenbasiertes als auch ein typebasiertes Produktivitätsmaß. Das typenbasierte Maß nennt sich *deleted estimation method* und setzt die als ungesehen definierten Types eines Korpus zur Zahl aller Types des Korpus in Beziehung.

In der Rezension dieser Arbeit (Lüdeling, 2009) wird darauf hingewiesen, dass Scherer (2006) zwar einen umfassenden Überblick über die Thematik der Produktivität gibt, indem sie viele 'Facetten' wie z.B. die Regularität, Beschränkungen und Produktivitätsmaße skizziert, aber die quantitative Auswertung verbesserungswürdig ist, da Scherer in ihrer Arbeit nicht auf die Arbeiten von Baayen²⁴ aus der zweiten Hälfte der neunziger Jahre eingeht, was deshalb als problematisch gelten darf, da gerade in diesen Arbeiten auf die Probleme der Vergleichbarkeit von Produktivitätsmaßen eingegangen wird.

Lüdeling (2009, 336) erklärt, dass Scherer (2006) in ihrer Arbeit einen grundsätzlichen Fehler macht:

„Ein Problem, das Baayen in vielen späteren Arbeiten (zusammenfasst und mathematisch detailliert dargestellt in Baayen 2001) diskutiert, ist, dass sich P bei verschiedenen Korpusgrößen nicht vergleichen lässt, weil es sich mit der Korpusgröße (Tokenanzahl der durch das betrachtete Muster erzeugten komplexen Wörter) verändert. Diese Nichtvergleichbarkeit lässt sich nicht durch einfache Normierung (z.B. auf 100 Token) in den Griff bekommen, da eine sich verändernde Zahl sich auch in der Normierung verändert. [...] Man hat in einem solchen Fall nur zwei Möglichkeiten – entweder schneidet man auf die kleinste Korpusgröße ab (und verliert Information) oder man rechnet statistische Modelle, die auf der Basis der vorliegenden Zahlen eine künftige Entwicklung extrapolieren.“

Eine andere Studie zur morphologischen Produktivität kommt von Lüdeling und Evert (2004), die sich mit dem Suffix *-itis* beschäftigen und dieses Suffix auch auf seine Verwendung außerhalb des medizinischen Bereichs hin untersuchen. Sie wenden bisher nicht etablierte statistische Maße an, die unter dem Namen LNRE-Modelle zusammengefasst werden und finden dabei Folgendes heraus: Ausgehend von der These, dass das medizinische *-itis*, das regelbasiert zur deutschen Wortbildung beiträgt, produktiver sein sollte als das in Analogie dazu angewandte *-itis*, das in anderen, nicht-medizinischen Kontexten auftaucht (Telefonitis²⁵, Filmeritis²⁶), stellen sie fest,

²⁴Zu den von Baayen vorgeschlagenen Maßen siehe Kap. 4.

²⁵Dazu siehe <http://www.stupidedia.org/stupi/Telefonitis>, abgerufen am 29.11.10 um 17:47 Uhr.

²⁶Ein Beleg findet man zum Beispiel auf http://www.rattinger-wochenblatt.de/ratingen/content/news/newsDetail.html;jsessionid=8D12DD7AFCEA57B53D37EFE2E192999B?command=display&btUid=bt_NewsArticle&iDf_id=d583e45:-4fe48bd8:12085896d21:4722, abgerufen am 28.11.10 um 17:49 Uhr.

dass ihre These auf Basis ihrer Ergebnisse nicht zu halten ist, da gerade das nicht-medizinische *-itis* im Gegenwartsdeutschen produktiv zur Wortbildung beiträgt.

Eine diachron ausgerichtete Untersuchung für deutsche Suffixe in der schweizerdeutschen Varietät findet sich in Schneider-Wiejowski (2009). Untersucht werden vier Suffixe (*-heit*, *-ung*, *-nis* und *-sal*) der schweizerdeutschen Schriftsprache, deren Wortbildungen anhand des Schweizer Textkorpus²⁷ über einen Zeitraum von 100 Jahren (1900-2000) extrahiert werden konnten.

Die Studie zeigt, dass das Suffix *-heit* mit seinen Allomorphen zu allen Zeitpunkten die höchste Produktivität besitzt; das Suffix *-ung* ist trotz hoher Frequenzen von Token und Typen unproduktiver als die Allomorphe von *-heit*. Dennoch macht dieses Suffix auf der anderen Seite die stärkste Entwicklung durch: insbesondere ab 1950 steigt die Produktivität von *-ung* stark an. Die Produktivität des Suffixes *-heit* hingegen bleibt während des gesamten Untersuchungszeitraums konstant.

Zu *-nis* wird festgestellt, dass es nur als sehr schwach produktiv bezeichnet werden kann. Außerdem wird anhand der *P*-Werte deutlich, dass die Produktivität im Laufe der untersuchten Jahre stetig sinkt.

Schaut man sich die Werte und die Entwicklung von dem als unproduktiv beschriebenen Suffix *-sal* an, so wird man auf den ersten Blick annehmen, dass es insbesondere um 1900 als 'produktiv' bezeichnet werden müsste. Doch die Autorin erklärt in diesem Zusammenhang, warum das Betrachten der *P*-Werte als einziges Produktivitätsmaß problematisch ist: die Datenmenge, die für *-sal* extrahiert wurde, ist zu gering ist, um solide Aussagen treffen zu können.

3.4 Arbeitshypothesen

Wortbildungswandel steht im Zusammenhang mit Produktivität, und Produktivität von Affixen kann, wie im nachfolgenden Kapitel erläutert werden wird, quantitativ gemessen werden.

Die Grammatikalisierungstheorie, die bereits skizziert wurde, wird zur Beschreibung der Entwicklung eines freien Morphems zu einem grammatischen herangezogen.

²⁷ Abrufbar unter <http://www.schweizer-textkorpus.ch/>.

Da sich die empirischen Untersuchungen in der vorliegenden Arbeit nur auf Affixe und nicht auf Affixoide stützten, wird nur der in Abb. 3.4 rot dargestellte Bereich untersucht werden.

Plausibel ist, dass Affixe im Zuge ihrer Grammatikalisierung an Abstraktheit dazu gewinnen. Je abstrakter das Zeichen, das Affix, wird, in desto mehr Kontexten ist es einsetzbar; die zunehmende Abstraktheit führt demnach zu einer Zunahme an Produktivität.



Abbildung 3.4: Messbereich der Produktivität in Bezug auf Grammatikalisierungsprozessen bei Affixen

Die Grammatikalisierungstheorie geht davon aus, dass es sich dabei um einen unidirektionalen Prozess handelt. In Bezug auf die Produktivitätsentwicklung ist diese Ansicht problematisch, denn das Prinzip der Unidirektionalität impliziert, dass es unmöglich ist, dass ein Affix Produktivitätsschwankungen ausgesetzt ist, indem es zeitweise an Produktivität zu- oder abnimmt.

Es ist natürlich nicht auszuschließen, dass es 'prototypisch grammatikalisierte' Affixe gibt, die sich genau nach dem unidirektionalen Muster entwickeln.

Auch ist es möglich, dass es am Ende ihrer Entwicklung zu einer Erosion von sprachlichem Material kommt. Nübling et al. (2006, 77) bezeichnen ein segmentierbares morphologisches Gebilde, das völlig unproduktiv geworden ist, als *morphologischen Rest*:²⁸

²⁸Ein Beispiel für einen morphologischen Rest im Gegenwartsdeutschen ist das Suffix *-t*, das auch heute noch in Wörtern wie z.B. *Fahr-t* oder *Fluch-t* existiert. Den meisten Sprachverwendern wird jedoch nicht bewusst sein, dass es sich hierbei um ein Suffix handelt. Es trug vor allem im Althochdeutschen zur Bildung von Neologismen und der Erweiterung des Wortschatzes bei (Lohde 2006, 114).

„Dies ist eine weitere Stufe im Kreislauf der Morphologie: Am Beispiel des Suffixes *-bar* haben wir gesehen, dass selbstständige Wörter als Quelle für neue Derivationsuffixe fungieren. Wie gezeigt, entstehen sie in einem langen Prozess, in dem die Bedeutung entkonkretisiert, die syntaktische Freiheit aufgegeben und die Lautung reduziert wird. Sie entwickeln sich zu produktiven Derivationsuffixen. Mit der Aufgabe der Produktivität werden sie zum segmentierbaren morphologischen Rest: Lexeme > Affixide > Derivationsuffixe > morphologischer Rest“

Sowohl Entwicklungen hinsichtlich steigender wie auch Entwicklungen abfallender Produktivität sind statistisch messbar, sofern genügend Material zum Untersuchen vorhanden ist.²⁹

In dieser Studie werden Daten des 20. Jahrhunderts untersucht. Die zentrale Fragen, die zu beantworten sind, lauten wie folgt:

1. Wie hat sich der Wortschatz des Deutschen im 20. Jahrhundert aufgrund der in dieser Arbeit untersuchten Derivationen verändert?
2. Liegen Schwankungen vor, die statistisch messbar sind?
3. Können die Aussagen in der linguistischen Literatur zu der Produktivität der Wortbildungssuffixe bestätigt werden oder gibt es empirische Befunde, die darauf hindeuten, dass einige Aussagen revidiert werden sollten?

Im nächsten Schritt wird deshalb notwendig sein, die Produktivitätsforschung genau zu skizzieren.

²⁹Die Diskussion zur Größe der Datenmenge ist Gegenstand von Kapitel 7 und 8 der vorliegenden Arbeit.

Produktivität in der Morphologie

Das Phänomen der morphologischen Produktivität ist erst in den letzten 30 Jahren intensiv beschrieben worden. So sagt z.B. Mayerthaler zu Beginn der 80er Jahre noch, dass die Produktivität „zu den unklarsten Begriffen der Linguistik“¹ gehöre.

Ähnlich beschreibt Aronoff (1976, 36) die Produktivität, nämlich als „one of the central mysteries of derivational morphology“.

Das folgende Kapitel soll dazu dienen, Ideen und theoretische Ansätze von morphologischer Produktivität zusammenzufassen. Dabei wird chronologisch vorgegangen. Weiterhin sollen auch die bisher entwickelten quantitativen Methoden zur Produktivitätsmessung erläutert werden.

4.1 Geschichte der morphologischen Produktivität

Als Schwerpunkt der (quantitativen) Produktivitätsforschung muss die Forschung von Harald Baayen an erster Stelle genannt werden. Sie stammt vordergründig aus den 90er Jahren.²

¹Mayerthaler (1981, 124).

²Vgl. Baayen (1989, 1992, 1993, 2001) und Baayen und Lieber (1991).

Die Vorstellung der Produktivität taucht allerdings auch in älteren Forschungsarbeiten – wenn auch oft nur implizit – im Zusammenhang mit der Bindungsfähigkeit von Affixen auf.

Zur Beschreibung des Phänomens liegt keine homogene Terminologie vor. So z.B. spricht Kruisinga (1932, 22) von „living suffixes“, die sie als produktiv beschreibt, wohingegen es ihrer Meinung nach auch sogenannte „dead suffixes“ gibt, die unproduktiv sind. Diese Aussage impliziert eine Einteilung in produktiv und unproduktiv, die inhaltlich nicht ohne Weiteres abgestritten werden kann, aber die Gradualität der Produktivität von Suffixen, ein zentraler Aspekt dieser Arbeit, bleibt unberücksichtigt. Es ist unzureichend, Suffixe nach 'toten' und 'lebenden' zu klassifizieren, vielmehr ist eine detaillierte feingradige Unterscheidung notwendig.

Jespersen (1942, 466) beschreibt einen anderen wichtigen Aspekt von morphologischer Produktivität, indem er auf das Phänomen aufmerksam macht, dass einige englische Adjektive nicht mit dem Präfix *un-* verbunden werden können. Er leitet daraus ab, dass nicht jedes Affix willkürlich an jede Basis anhängbar ist. Er gibt keine Erklärung für seine Beobachtung, beschreibt aber demnach dennoch, dass Restriktionen wirksam sein können, die die Produktivität beeinflussen können. An dieser Stelle sei auf Kapitel 2.4 verwiesen.

Ein weiterer wichtiger Schritt der Produktivitätsforschung kommt aus den Niederlanden; Schultink (1961, 113) beschreibt das Phänomen wie folgt:

„Onder produktiviteit als morfologisch fenomeen verstaan we dan de voor taalgebruikers bestaande mogelijkheid door middel van het morfologisch procédé dat aan de vorm-betekenis-correspondentie van sommige bekende woorden ten grondslag ligt, onopzettelijke in principe niet telbaar aantal nieuwe formaties et vormen.“³

Bei Betrachtung von Schultinks Definition fällt auf, dass einige wichtige Aspekte beschrieben werden, die zu einem späteren Zeitpunkt unter anderem auch von Evert und Lüdeling (2001) aufgenommen werden, die darauf hinweisen, dass es sich dabei um Aspekte handelt, die unabhängig voneinander existieren:

³Unter Produktivität als morphologisches Phänomen verstehen wir die Fähigkeit der Sprachbenutzer, morphologische Prozesse aufgrund anderer bereits zugrunde liegende Wortformen anzuwenden und unbewusst eine im Prinzip unbegrenzte Anzahl an neuen Wörtern zu bilden. [Eigene Übersetzung]

- *Unbewusstheit*: Hier erfolgt eine Unterscheidung zwischen Produktivität und Kreativität.⁴ Durch Produktivität werden neue Wörter nach Regeln gebildet, die oftmals nicht als neue Wörter realisiert werden, sofern Hörer und Sprecher die Regelanwendung nachvollziehen können. Mit Kreativität ist die Eigenschaft gemeint, neue Wörter mithilfe anderer Prozesse zu bilden. Dazu gehören z.B. die Wortschöpfung und die Bildung von Wörtern durch Analogie (*einsam*, *zweisam*, *dreisam*...)
- *Infinitheit*: Im Grunde ist es möglich, nahezu unbegrenzt neue Wörter zu bilden, sodass ihrer Meinung nach keine konkrete Anzahl festgelegt werden kann.
- *Regelmäßigkeit*: Produktive Wortbildungen werden regelbasiert gebildet und erlauben in vielen Fällen sogar eine Reihenbildung.

Dennoch darf Schultinks Definition auch kritisch betrachtet werden, denn fraglich ist, ob produktive Wortbildungen zwangsweise immer unbewusst und nicht-intentional gebildet werden müssen (Bauer 2005).

Ein andere wichtige theoretische Überlegung, die zu einem späteren Zeitpunkt von weiteren Linguisten aufgenommen wurde, ist die von Zimmer (1964): Er beschreibt den Zusammenhang von Frequenz und Produktivität. Sehr frequente Wortformen, die mit vielleicht auch produktiven Affixen gebildet worden sind, werden als gesamtes Wort wahrgenommen und können weniger schnell in ihre Bestandteile zerlegt werden. Für ihn ist Produktivität ein „psychologically real phenomenon“ (Zimmer 1964, 20), das eher semantischer oder pragmatischer Natur ist und nicht nur formal beschrieben werden sollte. Außerdem macht er ebenfalls die Beobachtung, dass es einige semantische Restriktionen hinsichtlich der Produktivität gibt, die nicht quantitativ ermittelt werden können.⁵

4.2 Die jüngere Produktivitätsforschung

Zur quantitativen Messung von Produktivität können unterschiedliche Modelle ihre Anwendung finden, die nacheinander besprochen werden:

- Frequenzmodelle

⁴Zu den Triebkräften der Wortschatzerweiterung und zur Erläuterung des Kreativitäts-Begriffes siehe Kapitel 2.3.

⁵Vgl. dazu Zimmer (1964, 84). Exemplifizierend dazu erwähnt er das nicht existierende englische Wort *unwell*, das nicht gebildet wird, da es die schon vorhandenen Wörter *sick* und *ill* gibt.

➤ Wahrscheinlichkeitsmodelle

Aus frequentieller Sicht lässt sich zunächst Folgendes festhalten: Mit einem Suffix, das produktiv ist, lassen sich viele Wörter bilden, was sich sowohl in der Typen- als auch in der Tokenhäufigkeit niederschlagen kann.

Sollen morphologische Sprachwandel-Tendenzen untersucht werden, muss eine diachron ausgerichtete Studie durchgeführt werden. In Bezug auf die Typenhäufigkeit ist die Anzahl der Typen zu unterschiedlichen Zeitpunkten zu ermitteln⁶ und die erhobenen Werte sind miteinander zu vergleichen.

Diese Methode fand ihre Anwendung insbesondere in Zeiten jenseits der digitalen Korpora⁷ und war dort unter dem Begriff *Wörterbuchstudie* bekannt; ein Zuwachs an neuen in das Lexikon aufgenommen Wörtern wurde folglich mit einem Zuwachs an Produktivität gleichgesetzt.

Problematisch an diesem Ansatz jedoch ist, dass Wörterbücher auch Wörter beinhalten, die in der Gegenwartssprache nicht aktiv genutzt werden (Baayen 1992, 111).

Auch die Tokenanzahl wird manchmal zur Produktivitätsmessung hinzugezogen. Die Frage, die man sich dabei allerdings stellen muss, ist, ob eine hohe oder eine geringe Tokenfrequenz für eine hohe Produktivität steht.

Aronoff (1976) erklärt dazu, dass die Tokenhäufigkeit von Wörtern mit produktiven Affixen geringer sein muss als die Tokenhäufigkeit von Wörtern, die von weniger stark produktiven Affixen gebildet worden sind.

Dieses Phänomen lässt sich damit erklären, dass Wörter, die mit unproduktiven Affixen gebildet worden sind, lexikalisiert sein können. Ein Aspekt der Lexikalisierung ist die Idiomatisierung, die dazu führt, dass lexikalisierte Wörter mehrere verschiedene Bedeutungen haben können. Wörter, die mit produktiven Affixen regelbasiert gebildet worden sind, können auch nur ein einziges Mal vorkommen, da sie eventuell für ein spezielles Phänomen gebildet wurden, das in einem Text nur einmalig benannt wird.

Token- und Typenhäufigkeit können auch interagieren (vgl. Wimmer 2005). Eine Möglichkeit, die Produktivität nach dieser Methode statistisch festzuhalten, sieht so aus, dass die Anzahl Typen und die Anzahl der Token in Relation zueinander gesetzt werden:

⁶Man spricht dann in diesem Zusammenhang von realisierter Produktivität (Baayen 2009, 905).

⁷In Kapitel 5.3 findet sich ein Exkurs zur Korpuslinguistik.

$$TTR = \frac{Typen}{Token}$$

Der Wert, der im Folgenden als **TTR** (Typen-Token-Relation) bezeichnet wird,⁸ lässt sich wie folgt interpretieren: ein produktives Affix wird, wie auch schon im vorangegangenen Kapitel beschrieben, viele verschiedene Types hervorbringen: **Je höher der Wert ist und je näher der Wert an 1 herankommt, desto höher ist die Produktivität des Affixes.**

Unter synchroner Perspektive können Affixe untereinander verglichen werden, um Aussagen über die Produktivitätsgrade zu machen. Unter diachroner Perspektive lässt sich die Typen-Token-Relationen dazu verwenden, mehrere Zeitpunkte zu untersuchen, um die prozessuale Veränderung statistisch skizzieren zu können.

Exemplifizierend wird das folgende Beispiel aufgeführt: gesetzt den Fall, die beiden Suffixe *-nis* und *-ung* sollen in demselben Textkorpus untersucht werden und insgesamt werden 500 Token mit *-ung* und 400 Token mit *-nis* gefunden; das Suffix *-ung* ist in 250 unterschiedlichen Wörtern vertreten, *-nis* in 100. Dies führt für *-ung* zu dem Ergebnis 0.5 (250/500), für *-nis* zum Ergebnis 0.25 (100/400). Die Ergebnisse zeigen somit, dass *-ung* im Vergleich zu *-nis* das produktivere der beiden Wortbildungsmodelle ist.

Selbstverständlich ist allerdings nur dann Vergleichbarkeit gewährleistet, sofern dieselbe Textlänge bzw. Tokenmenge, im Idealfall dasselbe Korpus, zur Untersuchung herangezogen wird. Und genau hier liegt deshalb der Nachteil der Methode, der nicht verkannt werden darf.

Aronoff (1976, 36) kritisiert diese Form der Produktivitätsmessung auch aus dem Grunde, dass keine Aussagen über die *möglichen* Wortbildungen gemacht werden können, die ein Affix theoretisch hervorbringen kann.⁹ Er orientiert sich deshalb an den sogenannten *belegten* und *potentiellen* Wortbildungsprodukten:

„Just as the simplest goal of a syntax is the enumeration of the class of possible sentences of a language, so the simplest task of a morphology, the least demand of it, is the enumeration of the class of possible words of a language. The greatest difference between the syntax and morphology with respect to his enumeration is that in derivational morphology there is

⁸Oftmals findet man die Anwendung der TTR als Charakteristik des Vokabelreichtums um beispielsweise den Wortschatz von Kindern zu ermitteln (vgl. Peuser 1993).

⁹Siehe dazu beispielsweise Plag (1999, 22) oder Baayen (1992, 111).

a distinction to be made between the classes of possible words and actual words.“ (Aronoff1976, 17f.)

Er spricht dabei von einer Anwendung einer Wortbildungsregel (*word-formation rule WFR*):

„There is a simple way to take such restrictions into account: we count up the number of words which we feel could occur as the output of a given WFR (which we can do by counting the number of possible bases for the rule), count up the number of actually occurring words formed by that rule, take the ratio of the two and compare this with the same ratio for another WFR.“¹⁰

Er entwickelt den folgenden Produktivitätsindex:

$$I = \frac{V}{S}$$

V stellt die in dem Korpus tatsächlich gefundenen Formen dar; S ist die mögliche Anzahl der Typen eines Wortbildungsmodells. I bezeichnet demnach den „Ausschöpfungsgrad“ (*degree of exhaustion*).

Bei diesem Ansatz entsteht das Problem, dass – zumindest auf den ersten Blick – nicht deutlich wird, wie die Zahl der möglichen Formen ermittelt werden kann.

Durch komplexe statistische Berechnungen gelingt es Baayen (1992), die möglichen Formen in einem Textkorpus zu berechnen, dennoch sind diese Berechnungen sehr aufwendig. Aus diesem Grund schlägt er ein einfaches Produktivitätsmaß, das sich auf zählbare Wörter anwenden lässt.

4.3 Wahrscheinlichkeitsmodelle

Die Anwendung von Wahrscheinlichkeitsmodellen zur Ermittlung von Produktivität soll erlauben, Aussagen über die Entwicklung von Produktivität (und damit der Bildung von Neologismen) zu machen, indem statistische Werte interpretiert werden.

Baayen unterscheidet zwischen der **Produktivität im engeren Sinn** (P) und der **globalen Produktivität** (P^*). Beide Ansätze werden im Folgenden nacheinander diskutiert.

¹⁰Aronoff (1976, 36).

4.3.1 Baayens Produktivitätsmaß P

Baayens Produktivitätsindex P basiert auf der Anzahl der sogenannten Hapax Legomena¹¹ und der Anzahl der Token des jeweiligen Wortbildungsmodells. Hapax Legomena repräsentieren jene Formen mit dem zu untersuchenden Affix, die nur ein einziges Mal vorkommen. In diesen Fällen beträgt sowohl die Token- wie auch die Typenhäufigkeit genau 1.

Der im Nenner stehende Wert stellt die Gesamtanzahl der Token dar, die mit einem Wortbildungsmodell gefunden werden können; im Zähler befindet sich die Anzahl der einmaligen Bildungen, die anhand der Tokenreihe ermittelt werden konnte:

$$P = \frac{Hapaxe}{Token}$$

Es wird angenommen, dass eine hohe Anzahl an einmaligen Bildungen darauf hinweist, dass das jeweilige Affix produktiv ist, zumal die einmaligen Bildungen mit hoher Wahrscheinlichkeit Neubildungen repräsentieren.¹²

Ein Affix, das nicht mehr produktiv ist, wird auch keine Neubildungen mehr hervorbringen.¹³

Soll das Verhalten von Typen in Relation zu der entsprechenden Tokenfrequenz untersucht werden und lassen sich diese Beobachtungen anhand von Daten eines großen Korpus machen, können die jeweiligen Entwicklungen in Kurven grafisch darstellen.

¹¹Wörtlich bedeutet *hapax legomenon* „etwas, das einmal gesagt wird“, der Plural dazu lautet *hapax legomena*. Im Englischen hat sich in der einschlägigen Literatur die Kunstform *hapax* mit der eigentlichen falschen Pluralform *hapaxes* („die Einmale“) herausgebildet. In der Produktivitätsforschung wird „hapax“ als neues Wort angesehen. Im Folgenden der Arbeit wird der Begriff in der nach der deutschen Rechtschreibung angepassten Variante *Hapax* (pl. *Hapaxe*) verwendet.

¹²Aufgrund ihrer Wichtigkeit widmen Aronoff und Fudeman (2005) den Hapaxen gleich ein ganzes Kapitel in ihrem Buch *What is morphology?*. Begründet wird das hapaxorientierte Maß P dadurch, dass eine produktive Wortbildungsregel mit einer Maschine vergleichbar sei, die Wörter ausspuckt, welche einmalig in die Luft geschleudert werden, ohne eine Spur zu hinterlassen. Allerdings weisen die Autoren auch darauf hin, dass die Wahrscheinlichkeit, dass es sich bei den Hapaxen zugleich um Neubildungen der jeweiligen Sprache handelt, gerade dann besonders hoch ist, wenn diese in einem sehr großen Korpus gefunden werden. Diese Überlegung erscheint plausibel, denn je geringer die Textmenge ist, die untersucht wird, desto größer ist die Wahrscheinlichkeit, dass es sich bei einem einmal vorkommendem Wort nicht um eine Neubildung handelt, sondern lediglich um eine selten verwendete Form.

¹³So ist es sehr unwahrscheinlich, dass wir heutzutage neue Wörter mit *-nis* (oder *-sal*) bilden werden: *Erlebnis*, aber nicht **Riechnis/*Riechsal*.

Wachstumskurven und Entwicklungen von Typenentwicklungen erlauben eine Unterscheidung von produktiven und unproduktiven Prozessen:

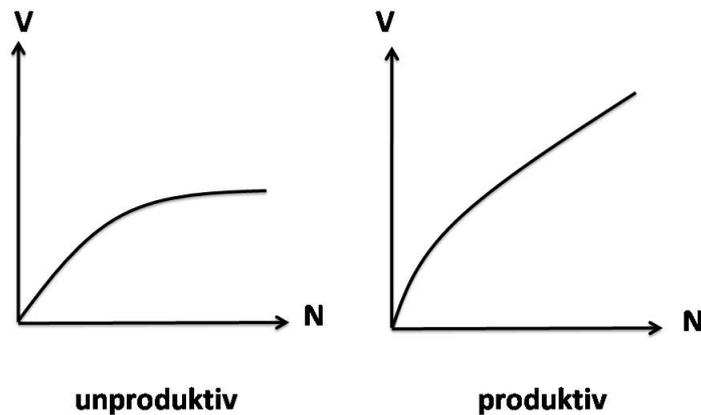


Abbildung 4.1: Wachstumskurven von Typen (V) für einen prototypisch unproduktiven und einen produktiven Wortbildungsprozess

Abb. 4.1 zeigt das Wachstum der Typen in zwei unterschiedlichen Wortbildungsprozessen. Bei einem nur schwach produktiven Modell steigt die Anzahl der Typen (V) erst an, bis ein gewisses Niveau erreicht ist. Dann allerdings erfolgt ein Abflachen der Kurve; die Kurve bleibt auf einem konstanten Plateau, was bedeutet, dass obwohl die Anzahl der Token (N) weiter steigt, sich die Anzahl der Typen kaum oder gar nicht weiter verändert.

Anders sieht es bei produktiven Prozessen aus. Dort steigt mit der Anzahl der Token auch die Anzahl der Typen konstant an.

In diesem Zusammenhang ist es äußerst wichtig, darauf hinzuweisen, dass P von N abhängig ist. Je größer das Korpus, desto kleiner ist demnach der Wert P . Um eine diachrone Untersuchung durchführen zu können, ist es deshalb wünschenswert, vergleichbare Korpora zu verwenden.

Auch in dieser Arbeit wird P als mathematisches Maß zur Beschreibung der deutschen Derivationssuffixe hinsichtlich ihrer Wortbildungspotenz angewendet.

4.3.2 Kritik an P

Der Wert P basiert auf der Anzahl der Hapax Legomena. Dies ist insofern als problematisch anzusehen, als dass es durchaus möglich sein kann, dass zwei unterschiedliche Sprecher einer Sprachgemeinschaft dasselbe Wort (neu) bilden. Im Korpus würde dieses Wort demnach zweimal erscheinen und nicht zur Anzahl der Hapax Legomena gezählt werden.

In der Literatur findet sich eine Auseinandersetzung mit diesem Problem. Scarborough (2001), die Ergebnisse von Baayen und Lieber (1991) (Tabelle 4.1) aufnimmt, erklärt, dass die von ihnen dargestellten Ergebnisse die Intuition der Linguisten in Bezug auf die Produktivität der untersuchten Affixe nicht bestätigen können.¹⁴ Scarborough bezieht sich insbesondere auf die Werte der Produktivität der englischen Suffixe *-ous* und *-able*:

Affix	Token	Hapaxe	Dislegomena	P_1	P_2
<i>-ous</i> (n)	21861	13	10	0.0006	0.0011
<i>-able</i> (v)	15004	10	8	0.0007	0.0012
<i>-er</i> (v)	57683	40	40	0.0007	0.0014
<i>-ness</i> (n)	17481	77	54	0.0044	0.0075
<i>-ish</i> (n)	1602	8	4	0.0050	0.0075
<i>-ish</i> (adj)	290	1	2	0.0034	0.0103

Tabelle 4.1: Ergebnisse zur Produktivität englischer Affixe nach Baayen 2001

Die Tabelle von Baayen und Lieber (1991) zeigt, dass das Suffix *-ous* nur gering weniger produktiv ist als die Suffixe *-able* und *-er*, was sich nicht mit der Intuition der Sprecher des Englischen deckt, da *-able* und *-er* im Englischen sehr produktive Suffixe sind. Scarboroughs Meinung nach kommen zwei unterschiedliche Fehlerquellen in Betracht: die falsch gezählte Anzahl der Token oder die der Hapax Legomena.

Eine Möglichkeit besteht darin, dass die Anzahl der Token überzählt wurde: Es könnten bestimmte Wortformen vorliegen, die nicht kompositionell gebildet werden, sondern im Lexikon gespeichert sind, sodass es nicht sinnvoll erscheint, diese Wörter

¹⁴Es sei an dieser Stelle angemerkt, dass es sich bei der hier zitierten Quelle um ein Unterrichts-Skript im Rahmen eines Morphologie-Kurses handelt. Dennoch erscheinen mir die in diesem Skript präsentierten Ergebnisse geeignet, um sie mit in die Argumentation aufnehmen zu dürfen.

als regelbasiert gebildete zu zählen. Auch ist es möglich, dass die Anzahl der Hapax Legomena als zu hoch und damit falsch deklariert wird, da, so Scarborough (2001, 5) drei Gründe aufgeführt werden können, dass ein Wort in einem Korpus nur einmalig vorkommt:

- 1. Das Wort IST ein Hapax Legomenon.
- 2. Die Gebrauchshäufigkeit eines Wortes ist nur sehr gering und das Wort ist nur im Lexikon einzelner Sprachbenutzer gespeichert. In diesem Fall handelt es sich um Archaismen oder Wörter aus dem Fachvokabular verschiedener sprachlicher Register.
- 3. Es kann ein statistischer Rechenfehler vorliegen.

Scarborough (2001) schlägt folgende Modifizierung für Baayens Produktivitätsindex vor: erstens sei es ihrer Meinung nach notwendig, Wörter mit einer sehr hohen Häufigkeit, die mehr als 100 Token beträgt, nicht mitzuzählen; zweitens müssen auch jene Wörter aus der Zählung herausgenommen werden, die weniger als zwanzigmal vorkommen, da dies ihrer Meinung nach nicht frequent genug sei; drittens schlägt sie vor, auch jene Wörter zu dem Wert des Zählers dazuzuzählen, die zweimal vorkommen. Diese werden als Dislegomena bezeichnet.

Auch wenn sich Scarboroughs Kritik nachvollziehen lässt und ihre Ergebnisse plausibel erscheinen, so sind ihre Ausführungen auch problematisch. Dies hat folgende Gründe:

1. Bei einem sehr großen Korpus lässt es sich vermutlich nicht vermeiden, dass es Wörter mit einer Tokenanzahl über 100 gibt.
2. Andersherum kann bei kleinen Korpora das Problem entstehen, dass Wörter weniger als zwanzigmal vorkommen. Es liegt kein nachvollziehbarer Grund vor, gerade die Zahl 20 als Schwellenwert zu definieren. Genauso gut könnte die Grenze auch bei 15, 19 oder 30 gesetzt werden.
3. Zum Vorschlag, nicht nur die Anzahl der Hapax Legomena, sondern auch die der Dislegomena zu zählen, lässt sich Folgendes sagen: auch die Zahl 2 ist kritisch zu betrachten, denn auch bei einem dreimalig auftretenden Wort kann es sich um eine Neubildung handeln, indem z.B. ein Redakteur einer Zeitung einen neuen (und aufmerksamkeitsregenden) Begriff erfindet und innerhalb seines Artikels dreimal oder sogar noch öfter auf diesen referiert.

Als weiterer Kritikpunkt muss an dieser Stelle darauf hingewiesen werden, dass die Typenfrequenz bei Berechnung der P -Werte unberücksichtigt bleibt. Um dieses Problem zu lösen, lässt sich neben der Produktivität P , auch die globale Produktivität P^* zur Analyse morphologischer Produktivität hinzuziehen.

4.3.3 Globale Produktivität

Die globale Produktivität ist eine von Harald Baayen entwickelte Modifikation der Formel P unter Berücksichtigung der Typenfrequenz.

Letztere setzt man auf der y-Achse mit der Produktivität P auf der x-Achse ins Verhältnis, sodass ein zweidimensionales Maß P^* entsteht. Damit wird sowohl die Wahrscheinlichkeit der Neubildungen als auch die Anzahl der vorhandenen Bildungen mitberücksichtigt.

Nachfolgend soll ein konkretes Beispiel zur Veranschaulichung besprochen werden: Abb 4.2.

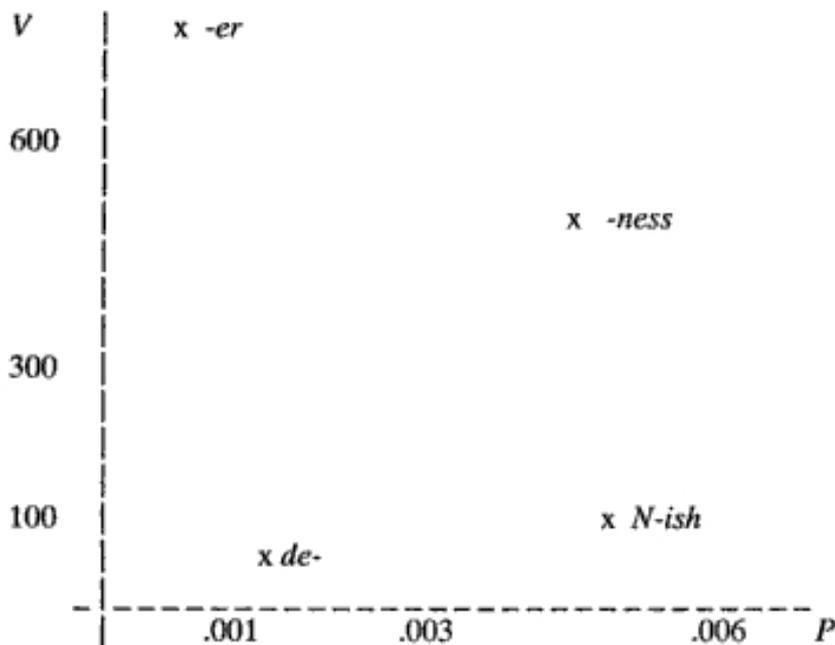


Abbildung 4.2: Globale Produktivität für einige englische Suffixe nach Baayen (1999, 124)

Die Interpretation der Ergebnisse in Abb. 4.2 gestaltet sich schwierig, denn die von Baayen dargestellten Ergebnisse zeigen Folgendes: *-ness* besitzt im Vergleich zu *-de*

die höhere Produktivität P^* , da bei *-ness* sowohl die Typenfrequenz als auch die Produktivität P höher ist als bei *-de*. Sollen allerdings die Suffixe *-ness* und *-er* miteinander verglichen werden, so fällt auf, dass dies problematisch ist: *-ness* zeigt eine höhere Produktivität im Bezug zu P , *-er* jedoch besitzt die höhere Typenfrequenz.

Einen Vorschlag, wie diesbezüglich nun eine klare Aussage zur Produktivität gemacht werden kann, findet sich in Baayen (1999) nicht. Deshalb soll dieses Problem im Folgenden ausführlich diskutiert werden.

Ein Idealfall für einen Anstieg oder einen Abfall der globalen Produktivität (in einer auf ein einziges Suffix ausgerichtete diachrone Studie oder im Vergleich mehrerer Suffixe untereinander) ist Abb. 4.3 zu entnehmen: mit Anstieg oder Abfall von P steigt oder fällt auch die Typenfrequenz.

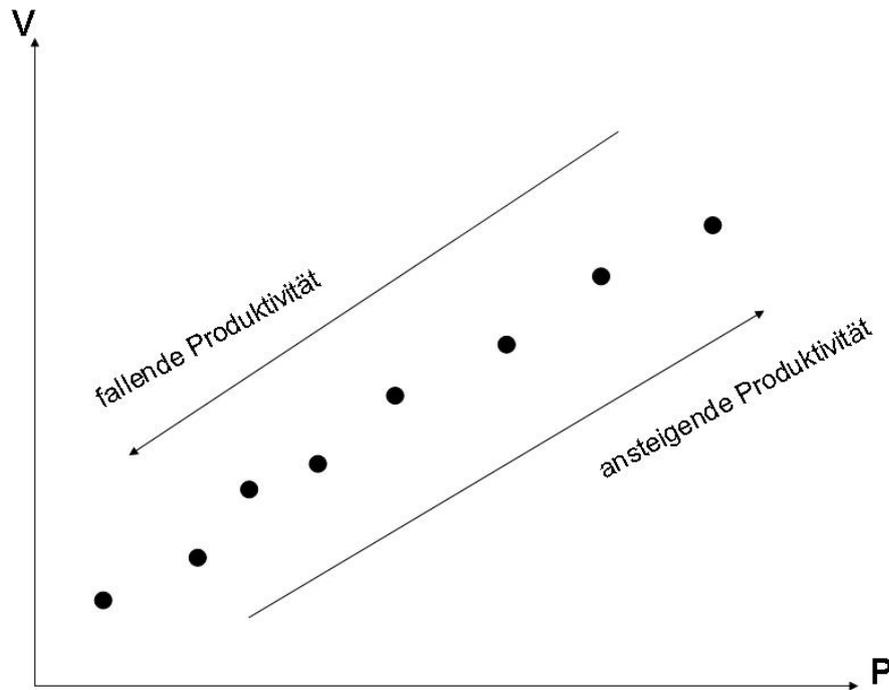


Abbildung 4.3: Ein Idealfall für eine Steigerung/einen Abfall der globalen Produktivität im Laufe eines Untersuchungszeitraums

4.3.4 Kritik und Diskussion von P^*

Wie bereits erklärt, kann P^* nicht ohne Weiteres rechnerisch ermittelt, sondern nur grafisch dargestellt werden. Im Idealfall erfolgt eine Abstufung, die bei größerem Wert P gleichzeitig auch eine größere Typenfrequenz hervorbringt.

Liegen hohe P -Werten bei gleichzeitig geringer Typenfrequenz vor, so ist die Interpretation der Ergebnisse problematisch, denn was bedeutet es, wenn ein Affix 'links oben' oder 'rechts unten' platziert ist?

Das Diagramm, das zur Darstellung der globalen Produktivität verwendet wird, kann man zunächst in vier Quadranten unterteilt werden.

Einfach und problemlos lassen sich zunächst unproduktive und deutlich produktive Affixe identifizieren (Abb. 4.4)

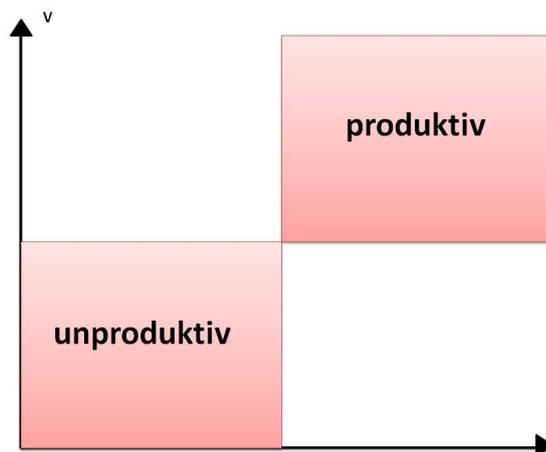


Abbildung 4.4: Produktive und unproduktive Affixe im Zusammenhang mit der globalen Produktivität

Affixe, die sich sowohl durch eine hohe Produktivität P wie auch durch eine hohe Typenanzahl auszeichnen, dürfen ohne Zweifel als produktive Affixe deklariert werden. Diese Affixe sind im oberen rechten Quadranten verortet.

Affixe, die weder eine hohe Produktivität, noch eine hohe Typenanzahl hervorbringen, werden sich in einem Diagramm der globalen Produktivität im linken unteren Quadranten befinden; sie sind unproduktiv.

Schwieriger wird es nun zu interpretieren, was es heißt, wenn die globale Produktivität eines Affixes im 'linken oberen' oder 'rechten unteren' Quadranten liegt (Abb. 4.5).

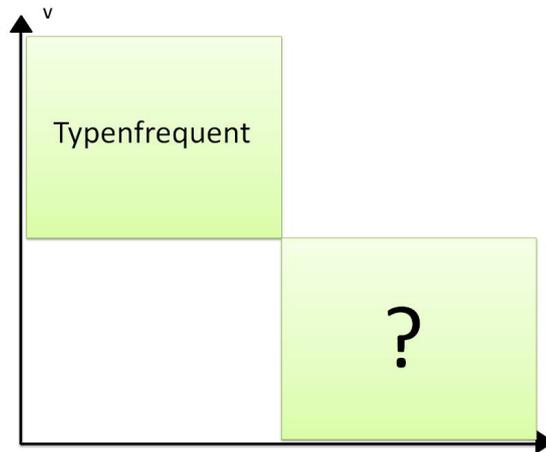


Abbildung 4.5: Probleme im Zusammenhang mit der globalen Produktivität

Festhalten lässt sich zunächst eine These: Affixe, die sich im linken oberen Quadranten verorten lassen, müssen eine hohe Anzahl an realisierten Bildungen aufweisen. Die entscheidende Frage, die zu beantworten ist, ist, ob die Etikettierung als typenfrequent mit einer hohen Produktivität gleichgesetzt werden darf oder nicht.

Dies ist nicht der Fall, denn denkbar wäre Folgendes: ein Affix war zunächst eine lange Zeit (z.B. über mehrere Jahrhunderte hinweg) produktiv; es entstanden viele Wortbildungen, die meisten davon wurden lexikalisiert. Auch heute noch sind diese lexikalisierten Bildungen im deutschen Wortschatz vorhanden und werden von den Sprechern einer Sprachgemeinschaft verwendet. Konnte sich allerdings ein Konkurrent dieses Suffixes mit der Zeit durchsetzen, so ist es gut möglich, dass dieses Wortbildungsmodell, was die Bildung von Neologismen angeht, nicht mehr aktiv ist. Demnach sollte der Bereich 'links oben' nur als typenfrequent etikettiert werden; dies steht zumindest für eine frühere Produktivität. Je weiter links sich das Affix in diesem Quadranten befindet, desto größer ist zudem die Wahrscheinlichkeit, dass es sich um ein unproduktives Affix handelt.

Wesentlich problematischer ist die Interpretation von Verortungen von Affixen im unteren rechten Quadranten. Ein Affix, dessen globale Produktivität hier eingeordnet wird, besitzt eine hohe Produktivität, ist aber nicht in vielen unterschiedlichen Typen

realisiert. Es ist möglich, dass es sich hierbei um ein produktives Affix und die Erklärung dafür lautet wie folgt: ein Affix, das (im Zuge der Grammatikalisierung) an Abstraktheit gewinnt, wird bindungsfreudiger und damit auch produktiver. Es ist denkbar, dass das Affix erst seit kurzer Zeit als Affix fungiert und parallel dazu ein freies Lexem existiert. In einem solchen Fall muss kein großes Typeninventar vorhanden sein, vielmehr wird sich dieses erst im Laufe der nächsten Jahre ergeben.

Auf der anderen Seite muss auch immer überlegt werden, dass *P* an sich nicht unkritisch zu betrachten ist, denn eine hohe Anzahl an Hapaxen tritt in der Regel auch genau dann auf, wenn nur eine geringe Datenmenge zur Untersuchung bereit steht. Des Weiteren muss es sich, wie vorab schon erklärt, nicht bei jedem Hapax Legomenon um eine Neubildung handeln.

4.4 Psycholinguistische Aspekte im Bezug auf morphologische Produktivität

Beschäftigt man sich mit dem Phänomen der morphologischen Produktivität, so sollten auch Überlegungen hinzugezogen werden, die aus dem Bereich der Psycholinguistik kommen.

In diesem Zusammenhang macht Plag (2004, 12) auf eine Reihe von Fragen aufmerksam:

„First, how do speakers know whether a process can be used to coin new words or not? In other words, what do productive processes have in common that unproductive processes do not have? And which properties of affixes give rise to different degrees of productivity?“

Er selbst gibt dazu die folgende Antwort:

„The central point of this research with regard to the nature of productivity is that the productivity of an affix is a reflection of the storage and processing of the complex words containing this affix.“ (ebd.)

In der Psycholinguistik existieren zwei unterschiedliche Modelle zur Verarbeitung morphologischer Strukturen (McQueen und Cutler 1998). Beim ersten Modell wird davon ausgegangen, dass ein morphologisch komplexes Wort im Ganzen abgerufen und auch wieder abgespeichert werden kann, sodass von einer 'whole word route'

gesprochen wird.¹⁵ Bei Annahme des zweiten Modells wird von einer Dekomposition ('decomposition route') der einzelnen Morpheme ausgegangen.¹⁶

Und obwohl es sich um zwei unterschiedliche Zugänge zu einem Wort handelt, nehmen die meisten Linguisten an,¹⁷ dass die Wahl der einen oder anderen Variante wortabhängig ist und die Häufigkeit von Basis oder Derivat darüber entscheidet, ob eine Dekomposition stattfindet oder nicht.

Hay (2001) beschäftigt sich mit dieser Frage und berechnet dabei die so genannte *relative frequency*. Sie schlägt vor, die Häufigkeit des Derivats und die der Basis des Derivats zu ermitteln, um beide Zahlen in Relation zueinander zu setzen:

$$F_{\text{relativ}} = \frac{F_{\text{Derivat}}}{F_{\text{Basis}}}$$

Im Normalfall besitzt die Basis ihrer Annahme nach eine höhere Frequenz als das Derivat mit dieser Basis¹⁸, was folglich einen Wert kleiner als 1 impliziert.

Exemplifizierend dazu erklärt Plag (2004), dass das englische Wort *detectable* in einer Korpusanalyse 281 Vorkommen hervorbringen konnte, während die Basis *detect* ganze 3083 Tokens besaß. Dennoch lassen sich auch derivierte Wörter mit *-able* finden, bei welchen der umgekehrte Fall zutrifft: so z.B. bei *deplorable* oder *comfortable*.

Schaut man sich den Normalfall an, bei welchem die Anzahl der Basis höher ist als die des Derivates, so ist nach Plag (2004, 14) auch die Ruheaktivierung im mentalen Lexikon größer, sodass die Basis stark repräsentiert ist. Den Zusammenhang zwischen den beiden Ansätzen des Parsings beschreibt Plag (2004) wie folgt:

„This leads to a decomposition bias, since due to its high resting activation, the base will be accessed each time the derivative enters the system. If, to the contrary, the derived word is much more frequent than the base, there is a whole word bias in parsing, because the resting activation of the base is lower than the resting activation of the derivative.“ (Plag 2004, 14)

Daraus lässt sich wiederum folgern, dass es innerhalb einer morphologischen Kategorie, bei der es eine Vielzahl an Wörtern gibt, die aufgrund der relativen Häufigkeit beim Parsen zerlegt werden müssen, dazu kommt, dass das entsprechende Affix ei-

¹⁵Vgl. dazu Butterworth (1993).

¹⁶Diesen Ansatz vertritt unter anderem Taft (1985).

¹⁷Dazu siehe beispielsweise Baayen (1992) und Frauenfelder und Schreuder (1992).

¹⁸Siehe dazu auch Harwood und Wright (1956), die bereits in den 50er Jahren herausfinden, dass eine derivierte Form generell weniger häufiger auftritt als die zur Derivation verwendeten Basis im Allgemeinen.

ne starke Repräsentation impliziert und damit für weitere Wortbildungsmöglichkeiten bereit steht.

Im Gegenzug dazu ist ein Affix aus psycholinguistischer Perspektive dann unproduktiv, wenn der Großteil der Wörter im Ganzen abgerufen und erschlossen werden kann, da in einem solchen Fall keine dekompositionelle Struktur vorhanden ist, die es ermöglicht, das Affix als eigenständige morphologische Einheit wahrzunehmen und dann analogisch weiterzuverwenden.¹⁹

Aus diesen Thesen lassen sich auch Zusammenhänge zu dem von Baayen (1992) vorgeschlagenen Maßen herstellen, da auch hier die Annahme vertreten wird, dass ein produktiver Prozess eine hohe Zahl an Wörtern hervorbringt, die insgesamt nur eine geringe Häufigkeit besitzen, bei welchen es sich in hohem Maße auch um Hapax Legomena handelt; andersherum ist zu erwarten, dass ein produktiver Wortbildungsprozess eine nicht unerhebliche Anzahl an Wörtern mit hoher Tokenfrequenz hervorbringt.

Aus psycholinguistischer Sichtweise taucht der Begriff der morphologischen Produktivität manchmal auch im Zusammenhang mit 'semantic and phonological transparency' auf.²⁰ Die Annahme, die hier zugrunde liegt, ist jene, dass ein produktiver Prozess sehr viele semantisch und phonologisch transparente Formen beinhaltet. Auch dafür gibt Plag eine Erklärung:

„Productive categories are semantically transparent, because their derivatives favor the decomposed route, and decomposed storage strengthens the individual semantic representations of the constituent morphemes. This leaves little room for semantic drift and opacity, which frequently arise under whole word access and storage.“ (Plag 2004, 16)

Den Zusammenhang zwischen Phonotaktik und Produktivität erklären Hay und Baayen (2003) dadurch, dass sie in ihrer Untersuchung feststellen können, dass eine Phonemkombination, die innerhalb eines Wortes nur selten auftritt, den Zerlegungsprozess einfacher gestaltet als eine Phonemkombination, die relativ häufig zu finden ist. Die Morphemgrenze sollte demnach aus einer nicht häufigen Phonemkombination bestehen, damit die Zerlegung einfacher stattfinden kann und die Produktivität demzufolge höher ist.

¹⁹Hay und Baayen (2002) untersuchen in diesem Kontext 80 englische Affixe und finden dabei heraus, dass die Produktivität des Affixes umso höher ist, je mehr zerlegbare Wortformen mit einem Affix vorhanden sind.

²⁰Dazu siehe Cutler (1981).

4.5 Fazit

In diesem Kapitel wurden verschiedene Möglichkeiten zur statistischen Messung von Produktivität vorgestellt. Zudem erfolgte gegen Ende dieses Abschnittes – in aller Kürze – eine Skizzierung von psycholinguistischen Annahmen in Bezug auf morphologische Produktivität, die vordergründig aufgeführt wurden, um das Phänomen Produktivität verständlicher darzustellen.

Als ersten Schritt zur Ermittlung der morphologischen Produktivität kann es sinnvoll sein, Häufigkeitsdaten zu ermitteln; zu diesen zählen zum einen Token und Typen, zum anderen die Anzahl der Hapax Legomena. Letzteres wird zudem benötigt, um das von Baayen vorgeschlagene Maß P anzuwenden. Weil die Typenfrequenz keinen Einfluss auf die jeweiligen P -Werte hat, sollte dann in einem weiteren Schritt auch die globale Produktivität P^* betrachtet werden.

Bei Beschreibung und Diskussion der einzelnen Methoden und Modelle wurde deutlich, dass diese allerdings auch mit bestimmten Nachteilen verbunden sind.

Doch trotz Nachteile ist es möglich, die Produktivitätsmaße anzuwenden und die daraus resultierenden Daten zu interpretieren. Im Verlauf dieser Arbeit soll deutlich gemacht werden, dass es sinnvoll ist, sich nicht nur auf ein Maß zu beschränken, sondern mehrere Maße im Zusammenspiel zu betrachten, um valide Aussagen zu treffen.

Die empirischen Studien

Das folgende Kapitel stellt die Rahmenbedingungen der empirischen Studie dar. Nach Definition des Untersuchungsgegenstandes, einer Beschreibung der Datenbasis und des Untersuchungszeitraums erfolgen auch Angaben zur methodischen Auswertung der Studie. Zudem findet sich in diesem Kapitel auch ein Exkurs zur Korpuslinguistik und den verfügbaren Korpora des Gegenwartsdeutschen.

5.1 Der Untersuchungsgegenstand

In der vorliegenden Arbeit werden Ergebnisse empirischer Studien zu nativen Suffixen vorgestellt. Es werden sowohl substantivbildende Suffixe wie auch adjektivbildende Suffixe untersucht, die nachfolgend skizziert werden. Insgesamt handelt es sich um 13 Suffixe, acht adjektivbildende und fünf substantivbildende Suffixe, die exemplarisch ausgewählt worden sind, um die Untersuchung in ihrer Ausweitung einzugrenzen.¹

Zunächst einmal sollen die Geschichte, die Semantik und die möglichen Basen der Suffixe im Vordergrund stehen. In der linguistischen Literatur (vgl. Fleischer und Barz 1995, Motsch 1999, Lohde 2006) werden dazu relativ homogene Aussagen gemacht, die zusammengefasst wiedergegeben werden sollen. Auch Hinweise zur Produktivität der Affixe lassen sich finden; sie werden in Tab. 5.1 gegenüber gestellt.

¹Die Untersuchung aller nativen Affixe des Deutschen hätte den Rahmen einer Dissertation gesprengt. Dennoch bleibt zu hoffen, dass die hier vorliegende Arbeit als Anstoß für weitere Forschungsprojekte dieser Art fungiert.

5.1.1 Das Suffix *-ung*

Geschichte

Das Suffix *-ung* hat sich aus dem mittelhochdeutschen *-unge* entwickelt (Wilmanns 1899, 369ff). Bereits im Frühneuhochdeutschen bilden Verben vorzugsweise die Basis-morpheme für *-ung*-Derivate. Das gilt auch für das Gegenwartsdeutsche, sodass diesbezüglich keine Veränderung festgestellt werden kann (Demske 2000, 371).² Dennoch merkt Demske (2000, 374) an, dass es im Frühneuhochdeutschen Bildungen gibt, die im Gegenwartsdeutschen nicht mehr gebräuchlich sind: *steiffung*, *zehrung*, *ansehung*, *umbkommung*, *auffhenckung* etc.

Basen

Wie bereits aufgeführt, dienen insbesondere Verben als Input für *-ung*-Derivationen. Basen können Simplexe (*Äsung*), Präfixverben (*Aufladung*), doppelte Präfigierungen (*Anerkennung*), suffigierte Verben (*Ängstigung*), kombinatorische Derivate (*Verstaatlichung*), verbale Komposita (*Entgegenstellung*) oder verbale Wortgruppen (*Zugrundelegung*) sein (vgl. Fleischer und Barz 1995).³

Doch auch eine substantivische Basis (*Waldung*, *Holzung*) oder eine adjektivische Basis kann suffigiert werden (*Dickung*, *Festung*).

Als vierte Möglichkeit werden isolierte Bildungen genannt, bei welchen sich sich nur schwer Motivationsbeziehungen zum Grundmorphem herstellen lassen, da die Basis oftmals nicht mehr im Gebrauch ist oder es sich um eine noch nicht alte Entlehnung handelt: *Innung* z.B. lässt sich auf das mittelhochdeutsche *innen* zurückführen, was soviel heißt wie „in einen Verband aufnehmen“ (Fleischer und Barz 1995, 172).

²Im 18. Jahrhundert erklärt Adelung (1782):

„**ung**, gleichfalls für Substantiva. 1. Concreta und Collectiva von Beschaffenheitswörtern und anderen Substantiven zu bilden: (...) 2. Abstracta, von Verbis mit Wegwerfung des *en* der Infinitive, die Handlung des Verbi als selbstständig darzustellen.“

³Vgl. dazu auch Motsch (2004, 332), der zur morphologischen Analyse des Suffixes angibt, dass es an „den Stamm beliebiger, auch morphologisch komplexer Verben“ tritt.

Semantik

Fleischer und Barz (1995, 172) sprechen bei den *-ung*-Derivaten von einer „semantisch offenen und wenig festgelegten Bildungsweise“.⁴

Ihrer Meinung nach ist es nicht erforderlich, dass die Wortbildungsprodukte alle Sememe der Basis aufnehmen und sie exemplifizieren dies durch das Wort *decken*: das Verb vererbt die Sememe 'Nachfrage befriedigen' (*Deckung des Bedarfs*), 'Schutz gewähren' (*Deckung des Rückzugs*) und 'für etwas aufkommen, Sicherheit gewähren' (*Deckung des Wechsels*), nicht jedoch 'etwas auf etwas legen' (**Deckung des Tisches*).

Die Wortbildungsbedeutung der Derivate zeichnet sich auch durch die Ausbildung verschiedener Wortbildungsreihen aus: Zum einen können Handlungsbezeichnungen (Nomen actionis) gebildet werden, bei denen die Merkmale der Transitivität und Temporalität des Verbs übernommen werden (*Die Professorin besprach die Dissertation ausführlich...* versus *während der ausführlichen Besprechung der Dissertation*), zum anderen ist es möglich, dass Vorgangsbezeichnungen (Nomen actionis) produziert werden, bei denen intransitive Verben als Basis zur Verfügung stehen (ebd.).

Auch ist die Möglichkeit der Temporalpräposition gegeben; in einem solchen Fall wird das Agens als Genitivattribut angeschlossen (*Der Zug fährt ein* versus *während der Einfahrt des Zuges*).

Ferner kommt eine Verflechtung der Nomen actionis „mit der Resultatsbezeichnung (Nomen acti) in der Bedeutungsstruktur eines Derivates“ in Betracht (ebd.).⁵

Als vierte Möglichkeit können, so Fleischer und Barz (1995), Sach- und Gegenstandsbezeichnungen gebildet werden (*Sammlung, Kupplung, Erfindung*), die dann oftmals auch als Kollektivum auftreten: *Bedienung, Leitung* etc.

Produktivität

Neben dem Suffix *-er* ist das Suffix *-ung* das produktivste substantivbildende Suffix des Deutschen. 1970 existieren, so Mater (1970), etwa 9575 Derivate auf *-ung*.

Während das Modell, bei welchem eine verbale Basis zur Affigierung verwendet wird, stark produktiv ist, ist das Modell mit einer substantivischen Basis nur als

⁴Vgl. auch Shippan (1967, 63), der von einer „Synthese der Wortarten Verb und Substantiv mit wechselnder Dominanz der verbalen und der substantivischen Eigenschaften“ spricht.

⁵Als Beispiele werden hier unter anderem *Schwächung des Körpers* oder *Verfeinerung der Sitten* genannt.

schwach produktiv zu bezeichnen; das Wortbildungsmuster mit einer adjektivischen Basis wird sogar als unproduktiv deklariert (vgl. Fleischer und Barz 1995).

Auch wenn bei *-ung* von einer hohen Bindungsfähigkeit und damit hohen Produktivität ausgegangen wird, so gibt es nach Fleischer und Barz (1995) dennoch auch Vorzugsbildungen und Einschränkungen; ein Derivat aus einem verbalen Simplex ist im Deutschen nicht immer geläufig: *Anhörung* - **Hörung*.

Es liegen zudem Einschränkungen hinsichtlich der präfigierten Verben vor: dort, wo bereits geläufige Simplizia existieren (z.B. *Abgabe*), wird eine *-ung*-Bildung blockiert (**Abgebung*).

5.1.2 Das Suffix *-nis*

Geschichte

In früheren Zeitstufen existierte das Suffix *-nis* in unterschiedlichen Varianten: *-nassi*, *-nessi*, *-nissi*, *-nussi*, *-nisse* oder auch *-nusse* (Wilmanns 1899, Krahe und Meid 1960).

Fleischer und Barz (1995, 166) merken an, dass *-nis* zunächst die mittel- und *-nus* die oberdeutsche Form gewesen ist und sich *-nus* seit dem 15. Jahrhundert auch im Mitteldeutschen durchsetzen konnte.

Basen

Am häufigsten dient eine (präfigierte) verbale Basis zur Suffigierung (*Befugnis*, *Erlaubnis*, *Ereignis* oder *Bedürfnis*); in seltenen Fällen kann auch eine substantivische Basis als Input dienen (*Bildnis*, *Bündnis*). Ebenso kommt eine adjektivische Basis in wenigen Fällen in Frage (*Bitternis*, *Wildnis*) (Fleischer und Barz 1995, 166).

Semantik

Das Suffix *-nis* bildet feminine oder neutrale Abstrakta.

Aus semantischer Sicht lässt sich festhalten, dass durch eine Derivation mit *-nis* Vorgangsbeschreibungen (Nomen actionis) oder Beschreibungen von Zuständen (Nomen acti) entstehen. Neben einem *-nis*-Derivat existiert oftmals ein paralleles *-ung*-Derivat, das den „Prozeßcharakter als Abstraktum stärker zum Ausdruck bringt“ (ebd.).

Als Erklärung für die Parallelbildungen (*Erzeugung- Erzeugnis, Hemmung - Hemmnis, ...*) nennt Shippan (1967, 80) die Unfähigkeit seitens der *-nis*-Derivate, objektive Genitive an sich zu binden. In der deutschen Sprache gibt es sowohl Paare dieser Art, die sich semantisch stärker voneinander unterscheiden (*Gleichung* und *Gleichnis*), als auch Wörter, die synonym zueinander sind, wie es beispielsweise bei *Verlobung* und *Verlöbnis* der Fall ist.

Produktivität

Heutzutage bezeichnet man es nicht mehr als produktiv (Motsch 2004, 359).

5.1.3 Die Allomorphe *-heit* und *-keit*

Geschichte

Das Suffix *-heit* geht auf ein mittelhochdeutsches Substantiv *heit* zurück, das vor seiner Grammatikalisierung 'Art und Weise, Beschaffenheit, Eigenschaft, Person und Stand' bezeichnete (Fleischer und Barz 1995, 158).

Ein kurzer Überblick dazu findet sich in Erben (2000, 136), der anhand von *-heit* die Tendenz zur Grammatikalisierung bevorzugter Kompositionsglieder zu Affixen beschreibt. Erben beginnt in seiner Darstellung beim Gotischen und erläutert die Entwicklung des heutigen Suffixes sehr detailliert. Dabei stellt er fest, dass das Substantiv *heit* im Mittelhochdeutschen nur noch sehr vereinzelt vorkommt.

Zur Geschichte von *-keit* lässt sich sagen, dass es aus der Verschmelzung von zwei mittelhochdeutschen Morphemen, *-ec* und *-heit*, entstanden ist. Kluge (1925, 20) spricht von einem Vordringen von *-keit* seit dem 12. Jahrhundert.

Fleischer und Barz (1995, 159) sehen in der Form *-igkeit* ein eigenständiges Allomorph⁶ und machen zu diesem die folgenden Angaben:

„Die Form *igkeit* entstand als 'Verdeutlichung' von *ekeit* [...] mit Ausrichtung auf das jeweilige *-ig*-Adjektiv [...], so daß sich eine klare Exemplifizierung des Derivationsverhältnisses ergab. Die Verdeutlichungsform dehnte sich auch auf Derivate aus, deren Basis kein Adjektiv auf *-ig*, sondern ein Simplex ist: *genau - Genau-igkeit* [...].“

⁶In dieser Arbeit wird *-igkeit* nicht separat betrachtet, es wird in den empirischen Studien nur zwischen *-heit* und *-keit/-igkeit* unterschieden.

Basen

Alle Allomorphe können an eine adjektivische Basis affigiert werden (*Traurigkeit, Sauberkeit, Kühnheit*). Wird eine substantivische Basis als Ausgangspunkt der Derivation verwendet, so kommt vorzugsweise *-heit* in Frage.

Erwähnenswert ist, dass bei *-heit* und seinen Allomorphen Numerale die Basis für eine Derivation bilden können, wie es zum Beispiel bei *Einheit* der Fall ist. Dieses Bildungsmuster ist analogischer Natur, sodass gerade in diesem Bereich Neubildungen möglich sind.

Die drei Varianten *-heit/-keit/-igkeit* sind komplementär zueinander, obwohl Überlappungen vorliegen. Die Vorkommensverhältnisse jedoch sind klar, da Unterschiede in den Akzentmustern zu finden sind (vgl. Fleischer und Barz 1995).

Bereits Willmanns (1899, 386) stellt fest, dass die Variante *-keit* hinter einer unbetonten Silbe stehen muss.⁷ Aus diesem Grund bietet es sich an, suffigierete Basen auf *-bar, -ig, -lich, -isch* oder *-sam* zu verwenden. Es verbindet sich zudem mit simplizischen Basen, wie es z.B. bei *Heiser-* der Fall ist.

Die Variante *-keit* findet sich nach Basen auf *-ig* und auch nach Basen auf *-haft* und *-los*.⁸ Außerdem lassen sich im Deutschen auf *-e* auslautende Adjektive mit *-igkeit* suffigieren.

Bei Suffigierungen mit *-heit* gibt es sehr unterschiedliche Akzentmuster (Fleischer und Barz 1995, 160).

Zusammenfassend lässt sich die folgende Verteilungstabelle aufstellen⁹:

⁷Vgl. dazu auch Kolb (1985), der die verschiedenen Akzentmuster beschreibt.

⁸Zur Veränderung des Akzentmusters siehe dazu Kolb (1985).

⁹Entnommen aus canoo.net; abgerufen am 19.07.2010 um 15:54 Uhr unter <http://canoo.net/services/WordformationRules/Derivation/To-N/Suffixe/Var-heit-keit.html;jsessionid=31346967804EFF106B3588547BA78E7C>.

5 Die empirischen Studien

-heit		
immer bei adjektivi- schem Partizip Perfekt	<i>Bewegtheit, Ergebenheit, Verliebtheit</i>	
immer bei Adjektiven auf unbetontes -en oder -ern	<i>Trockenheit, Verlegen- heit, Nüchternheit</i>	
meist bei einsilbigen Ad- jektiven	<i>Derbheit, Schwachheit, Schönheit, Neuheit</i>	seltener auch -igkeit Sü- ßigkeit, Schnelligkeit
meist bei einfachen mehrsilbigen Adjekti- ven mit Endakzent	Gesundheit, Kaputtheit, Saloppheit	seltener auch -igkeit Ge- schwindigkeit, Genauig- keit
seltener bei Adjektiven auf unbetontes -er	<i>Sicherheit, Lockerheit</i>	meist -keit Sauberkeit, Tapferkeit
-keit		
nie unmittelbar nach be- tonter Silbe		
immer bei -bar	<i>Fruchtbarkeit</i>	
immer bei -ig	<i>Flüssigkeit</i>	
immer bei -isch	<i>Störrischkeit</i>	
immer bei -lich	<i>Ehrlichkeit</i>	
immer bei -sam	<i>Einsamkeit</i>	
bei Adjektiven auf un- betontes -el	<i>Eitelkeit, Übelkeit</i>	Ausnahme: Dunkelheit
meist bei Adjektiven auf unbetontes -er	<i>Heiserkeit, Sauberkeit, Tapferkeit</i>	seltener auch heit: Sicherheit Locker- heit
-igkeit		
immer bei -los	<i>Hilflosigkeit</i>	
immer bei -haft	<i>Ernsthaftigkeit</i>	
seltener bei einsilbigen Adjektiven	<i>Neuigkeit, Kleinigkeit, Süßigkeit</i>	meist heit Schwachheit, Schönheit, Neuheit
seltener bei einfachen mehrsilbigen Adjekti- ven mit Endakzent	<i>Geschwindigkeit, Genau- igkeit</i>	meist heit Gesundheit, Saloppheit

Semantik

Zwischen den drei Variationen existieren aus semantischer Sicht keine Unterschiede: durch die Allomorphe werden substantivische Eigenschaftsbezeichnungen und Verhaltensweisen und Handlungen geschaffen. Ebenso ist es möglich, Bezeichnungen für Gegenstände (*Flüssigkeit*) oder Personen (*Und es kam eine Schönheit hinein.*) zu bilden (ebd.).

Produktivität

Das Wortbildungsmodell mit einer adjektivischen Basis ist sehr produktiv, während das Modell mit einer substantivischen Basis als nur schwach produktiv bezeichnet wird (Fleischer und Barz 1995, 161).

5.1.4 Das Suffix *-sal*

Geschichte

Für das Suffix *-sal* nimmt man an, dass es sich aus Formen entwickelt hat, bei denen *s* als Wurzeldeterminativ fungiert hat (*ahsala* 'Achsel' zu *ahsa* 'Achse' aus *ag+s* zur Wurzel *ag*). Im Gotischen erfolgt der Anschluss an eine Wurzelsilbe (Wilmanns 1899), was auch im Althochdeutschen beobachtet wird. Im Althochdeutschen ist allerdings zudem auffällig, dass sich das Suffix insbesondere in der Form *-isal* wiederfindet.

Im Mittelhochdeutschen erfolgt eine Abschwächung des Suffixes zu *-sel* (*amsel*), gleichzeitig kann sich das Suffix auch anderweitig behaupten und findet sich in vor allem auch in nominalen Ableitung wieder (*flühtesal*, *fluchtsal* 'Flucht') (ebd.).

Im Neuhochdeutschen lassen sich noch einige Formen auf *-sel* wiederfinden: *Anhängsel*, *Rätsel*, *Streusel* oder *Überbleibsel* etc.

Basen

Das Suffix *-sal* erscheint im Gegenwartsdeutschen nur noch in wenigen Substantiven, und wenn, dann tritt es vorzugsweise in Kombination mit verbalen und simplizischen Basis wie z.B. in *Labsal* oder *Rinnsal* auf (Fleischer und Barz 1995).

Es lassen sich auch – wenn auch nur wenige – Belege mit einer substantivischen (*Drangsal*, *Mühsal*) und einer adjektivischen Basis (*Trübsal*) aufführen.

Semantik

Lohde (2006, 114) bezeichnet die noch existierenden Bildungen mit *-sal* als „Archaismen oder stilistisch gehobene Bildungen“. Der Begriff Archaismus wird in der Lexikologie verwendet, um Wörter zu beschreiben, dessen Gebrauchshäufigkeit abnimmt. Archaismen werden vor allem von älteren Menschen einer Sprachgemeinschaft verwendet, da sie von vielen jüngeren Sprechern als altmodisch empfunden werden. Des Weiteren ist es möglich, die veralteten Wörter aus rhetorischen oder pragmatischen Gründen bewusst einzusetzen.¹⁰

Produktivität

Das Suffix *-sal* ist heute unproduktiv (Fleischer und Barz 1995, 168). Doch trotz der eindeutigen Aussage zur Unproduktivität ist es höchst interessant, dass es dennoch möglich ist, kreative Neubildungen wie z.B. *Rührsal* zu schaffen, welche sich am 07.08.1990 in der TAZ findet.¹¹ Der Kontext dazu lautet „von Rührsal triefend“. Warum sich die Autoren gerade für das Wort *Rührsal* entschieden haben, ist unklar, denn auch das Wort *Rührung*¹² wäre vorhanden, denselben Inhalt zu beschreiben.

5.1.5 Das Suffix *-tum*

Geschichte

Das Suffix *-tum* hat seinen Ursprung im mittelhochdeutschen Grundmorphem *tuom*. Die Bedeutung dieses Morphems war in dieser Zeit sehr vielseitig, Meinecke (1994, 501ff) nennt z.B. *Besitz*, *Herrschaft*, *Urteil*, *Ruhm* und *Ehre* als Beispiele.¹³

¹⁰Aussterbende Wörter des Deutsche finden sich in Mrotzek (2005). So verwendet man im Gegenwartsdeutschen das Wort *Single* anstatt *Hagestolz* oder *Brunch* anstatt *Gabelfrühstück*. Bei beiden neuen Varianten handelt es sich um Anglizismen, die die veralteten deutschen Wörter ersetzen.

¹¹Siehe: *taz.de*. Dieser Beleg ist über die Korpora der geschriebenen Sprache des IdS Mannheim abrufbar.

¹²Eine Korpusabfrage mit dem DWDS (06.06.2010, 16:28 Uhr) zeigt, dass insgesamt 453 Tokens zum Treffer *Rührung* gefunden werden, für *Rührsal* hingegen wird kein Beleg gefunden.

¹³Dabei kann an dieser Stelle kurz angemerkt werden, dass die Entwicklung von *-tuom* im Englischen morphologisch und semantisch anders verlaufen ist als im Deutschen. Im Gegenwartsendglichen

Basen

Das heutige *-tum* kann substantivische (*Königtum*), verbale (*Wachstum*) und adjektivische Basen (*Reichtum*) ableiten, wobei die häufigste Ableitung mit einer substantivischen Basis erfolgt (Lohde 2006, 107f).

Semantik

Durch das Suffix werden gewöhnlich nur Neutra gebildet¹⁴ und es wird an Personenbenennungen affigiert, die „eine bestimmte Verhaltensweise (Verhaltensabstrakta)“ (ebd.) bezeichnen. Als Beispiel kann an dieser Stelle *Bürokratentum* oder *Strebertum* genannt werden.

Es existiert auch eine Reihe an substantivischen Basen, die sich nicht auf Personen bezieht. Dieses semantische Muster, das sich noch in lexikalischen Formen wie z.B. *Besitztum* oder *Brauchtum* findet, wird allerdings keine Neubildungen mehr hervorbringen (Lohde 2006, 108).

Produktivität

Das Suffix *-tum* wird als schwach produktiv bezeichnet (Fleischer und Barz 1995, Lohde 2006).

5.1.6 Das Suffix *-bar*

Geschichte

Das Suffix *-bar* geht auf das althochdeutsche *bāri* (von *beran* 'tragen') zurück und entwickelte sich im Mittelhochdeutschen zu *baere* weiter (Lohde 2006, 177).

Basen

Als Basis kommen vorzugsweise verbale Basen in Frage (*brauchbar*, *trinkbar*). Es steht ein breites Inventar an Möglichkeiten zur Verfügung: Infinitivstämme, starke und

existiert sowohl eine gebundene Variante *-dom* in z.B. *kingdom*, aber auch ein freies lexikalisches Morphem *doom*, was soviel heißt wie *verurteilen*, *verdammnen* oder *Verdammnis*.

¹⁴Ausnahmen bildet zum Beispiel *Besitz* (mask.) und *Brauch* (mask.).

schwache Verben, simplizische (*drehbar*) und auch komplexe Basen (*auffindbar*) können deriviert werden.

In seltenen Fällen dienen auch Substantive als Basis (*mittelbar*).

Semantik

Fleischer und Barz (1995, 252) sehen das Suffix als Vermittler zwischen Verben und Substantiven an:

„Es ermöglicht die Substantivierung von Verben mit der Wortbildungsbedeutung des Potentiellen: *machen* - *machbar* - *Machbarkeit*, *verwerten* - *verwertbar* - *Verwertbarkeit*.“

Die meisten Basisverben sind transitiv und passivfähig, sodass die durch das Verb bezeichnete Handlung in Bezug auf ein Objekt ausgeführt werden kann (*trinkbares Wasser*).

Interessant an *-bar* ist vor allem die Konkurrenz zu anderen Suffixen. Das Modell mit *-bar* löste die deverbalen Modelle mit *-lich* und *-sam* ab. Dennoch, so merken Fleischer und Barz (1995, 252) an, existieren im heutigen Wortschatz noch Parallelkonstruktionen wie z.B. bei *erklärlich/-bar*. In diesem Zusammenhang wird auch von „synonymischer Konkurrenz“ gesprochen. Das Nebeneinanderexistieren von *-bar* und *-sam* wird als „unsystematische semantische Differenzierung“ beschrieben (ebd.). Den aktivischen *-sam*-Bildungen stehen die passivischen *-bar*-Bildungen gegenüber: *heilbar* - *heilsam*.¹⁵

Produktivität

Das Suffix *-bar* wird als das produktivste Suffix deverbalen Adjektivbildung beschrieben. Dabei steht allerdings nicht eine Erweiterung des Wortschatzes im Vordergrund, sondern vielmehr die „Bildung syntaktischer Alternativkonstruktionen“ (Fleischer und Barz 1995, 252).

Deshalb werden viele Bildungen letztendlich auch nicht lexikalisiert. Die hohe Produktivität lässt sich vor allem auf die geringen Restriktionen hinsichtlich der Bindungsmöglichkeiten zurückführen.

¹⁵An dieser Stelle wird von Fleischer und Barz (1995) auch die synonymische Konkurrenz von *-fähig* und *-bar* erwähnt, die allerdings eher selten ist. Diese Konkurrenz tritt vor allem bei Personenbezeichnungen auf.

Das substantivische und das adjektivische Modell sind nicht produktiv.¹⁶ Erklären lässt sich diese Tatsache dadurch, dass es zwischen suffigierten Adjektiven wie z.B. *kundbar* und *kund* kaum einen semantischen Unterschied gibt, sodass eine Bildung mit *-bar* keinen Mehrwert darstellt und redundant erscheint (ebd.).

5.1.7 Das Suffix *-sam*

Geschichte

Dieses Suffix ist aus dem althochdeutschen *sama*, *samo* entstanden, was soviel hieß wie 'derselbe' oder 'ebenso' (Erben 2000, 139).

Basen

Es bindet sich an Verben (*bedeutsam*), an Nomen (*gewaltsam*) und an Adjektive (*gemeinsam*).

Semantik

Hinsichtlich der Wortbildungsbedeutung von *-sam* gibt es verschiedene Wortbildungsreihen: potentiell-passivisch (z.B. *biegsam*), zu einer Handlung geneigt, die durch das Verb beschrieben wird (z.B. *sparsam*) und ornativ (z.B. in *wirksam*) (ebd.).

Produktivität

Das Suffix *-sam* ist heute nicht mehr produktiv.

Das einzige Modell, das man als noch schwach produktiv bezeichnen kann, ist das Modell mit verbaler Basis, in welchem hauptsächlich simplizische Verbstämme als Basis fungieren (Fleischer und Barz 1995, 266).¹⁷

¹⁶Siehe dazu auch Flury (1964, 94), der feststellt, dass der Anteil an Deverbativa vom 15. Jahrhundert von 33% im 20. Jahrhundert auf 98% wächst. Er findet heraus, dass heutzutage nur noch 34 desubstantivische Bildungen gebräuchlich sind.

¹⁷Siehe dazu auch Willmanns (1899, 493).

5.1.8 Das Suffix *-haft*

Geschichte

Das Suffix *-haft* ist aus dem noch im alt- und mittelhochdeutschen existierenden freien Morphem *haft* entstanden, was in seiner ursprünglichen Bedeutung 'gefesselt, gebunden' hieß (Wilmanns 1899, 501).

Basen

Vorzugsweise tritt das Suffix an eine substantivische Basis (*mangelhaft*) (Fleischer und Barz 1995). Dabei ist es möglich, dass es zu einer Tilgung von sprachlichem Material kommt (*stimmhaft*).

Die meisten auf ein unbetontes *e* endenden Substantive werden über das Fugenelement *en* mit *haft* verbunden (*katzenhaft*). Die substantivische Basis kann zudem sowohl eine simplizische als auch eine komplexe Basis sein: *bildhaft*, *bruchstückhaft*.

Auch verbale (*naschhaft*) und adjektivische Basen (*boshaft*) können belegt werden.

Als vierte Möglichkeit nennen Fleischer und Barz (1995, 256) „Relikte einer besonders seit dem 14. Jahrhundert zunehmenden Erweiterung von *-haft* durch *-ig*“ (*tugendhaftig*), wobei angemerkt wird, dass sich die einfache *-haft*-Bildung wieder durchsetzen konnte.

Semantik

Auf der einen Seite ist es möglich, komparative Wortbildungen zu schaffen, indem vorwiegend Personen- und Tierbezeichnungen als Basis fungieren (*lehrerhaft*, *bärenhaft*), auf der anderen Seite können ornative Wortbildungen produziert werden (*fieberhaft*, *krampfhaft*).

Als Konkurrenz zu *-haft* können die Suffixe *-voll* und *-reich* angesehen werden: *grauenhaft/voll* oder *heldenhaft/heldenreich* (ebd.).

Produktivität

Das Suffix ist reihenbildend und damit produktiv (Motsch 2004, Lohde 2006).

5.1.9 Das Suffix *-lich*

Geschichte

Das Suffix *-lich* ist aus dem noch im mittelhochdeutschen existierenden freien Lexem *līch* entstanden, was soviel wie 'Leib, Körper' bedeutete (Fleischer und Barz 1995, 260).

Basen

Das adjektivbildende Suffix *-lich* bindet sich an Substantive simplizischer (*täglich*) oder kompositioneller Basis (*arbeitsgerichtlich*), an Verben (*bedrohlich*), an Adjektive (*kleinlich*) und an Adverbien (*sämtlich*).

Einen Sonderfall stellen die Bildungen mit Konfixen als Fremdbasis (*kontinuierlich*) dar (ebd.).

Semantik

Zur Semantik des Wortbildungsmodells geben Fleischer und Barz (1995, 260) ornative (*gegensätzliche Meinung*) und komparative Bildungen (*väterlicher*) an.

Des Weiteren existieren Parallelbildungen zu *-ig*, durch welche synonymische Doppelungen entstehen: *schaurig* versus *schauerlich*. Doch die semantische Nähe zwischen zwei Bildungen mit gleicher Basis kann unterschiedlich ausgeprägt sein. So zum Beispiel bezeichnet das Adjektiv *farbig* etwas, das mit Farbe versehen ist, während sein 'Konkurrent' *farblich* etwas in Bezug auf Farbe beschreibt (ebd.).

Auch lassen sich Konkurrenzen zwischen *-lich*-Adjektiven und dem Partizip I finden: *maßgeblich* versus *maßgebend* (ebd.).

Produktivität

Das Suffix wird als stark produktiv bezeichnet, da es „zu den wichtigsten Suffixen für die desubstantivische Adjektivderivation“ und „in starkem Maße auch mit Verben und Adjektiven verbindbar“ ist (Fleischer und Barz 1995, 260).

5.1.10 Das Suffix *-ig*

Geschichte

Das Suffix *-ig* gehört im gegenwärtigen Deutsch zu den zentralen Adjektivsuffixen und hat seinen Ursprung in den althochdeutschen Formen *-ag*, *-īg* und *-ig* (vgl. Fleischer und Barz 1995).

Basen

Die Bindungsfähigkeit dieses Suffixes ist, wie bei vielen anderen Suffixen auch, variabel, da es sich an substantivische, verbale, adjektivische und aber auch an adverbiale Basen anhängen lässt.

Wellmann (1995, 529) stellt fest, dass die substantivischen Basen die dominantesten sind, weil es sich bei den *-ig*-Derivaten zu 83% um desubstantivische Ableitungen handelt.

Liegt eine Ableitung vom Substantiv vor, so kann das Substantiv simplizischer (*hungrig*) oder kompositioneller Natur sein (*schwachsinnig*). Außerdem ist es möglich, dass eine 'substantivische Wortgruppe' aus Adjektiv und Substantiv (*breitschultrig*) oder Numeral und Substantiv (*zweisprachig*) als Ableitungsbasis fungiert (Lohde 2006, 183).

Wird eine verbale Basis deriviert, so geschieht dies vorzugsweise an einen Infinitivstamm (*Taumelig*, *rutschig*,...). Aber auch eine verbale Gruppe kann als Input dienen, wie es sich beispielsweise durch die deutschen Wörter *schwerhörig* oder *luftdurchlässig* demonstrieren lässt (ebd.).

Dient ein Adverb als Basis, so liegt oftmals ein Konsonantenausfall oder -wechsel vor.¹⁸ Auch Tilgungen, insbesondere des adverbialen Suffixes *-s*, sind geläufig und frequent: *damals* vs. *damalig*.

Als unproduktives Muster nennt Lohde (2006, 184) Derivation mit Adjektiven: *lebendig* und *faulig*.

¹⁸Siehe dazu Erben (2000, 116), der dieses durch *jetzt* und *jetzig* und *hier* und *hiesig* belegt.

Semantik

Als Basis der desubstantiven Bildungen können ornative (*wolkiger Himmel*, *kantiger Tisch*) oder komparative Muster (*affig*) genannt werden (Fleischer und Barz 1995, 265).

Auch *-ig* besitzt Konkurrenten. In diesem Zusammenhang nennt Lohde (2006, 183) kompositionelle Zweitglieder wie *-artig* (*wellenartig* vs. *wellig*) und auch das Partizip II, das dann aus komplexen präfigierten Verben, vorzugsweise mit *ver-*, gebildet worden ist (*dreckig* vs. *verdreckt*). Weiterhin kann auch das Suffix *-haft* synonymisch konkurrieren: *krüppelig* vs. *krüppelhaft*.

Produktivität

Weil es sich „um ein stark genutztes Modell“ (Fleischer und Barz 1995, 256) handelt, ist das Suffix als produktiv zu bezeichnen.

5.1.11 Das Suffix *-mäßig*

Geschichte

Der Ursprung des heutigen Suffixes *-mäßig*¹⁹ liegt in der mittelhochdeutschen Pluralform *diu māze*, das die Bedeutung 'Verhältnismäßigkeit' und 'das Ausmaß von etw. habend' besaß (Seibicke 1963, 33f).

Basen

Das Suffix bindet sich an vorzugsweise an simplizische (*streitmäßig*) und komplexe (*vaterlandmäßig*) substanvische Basen. Ähnlich wie auch bei dem Suffix *-los* liegen auch bei der Nutzung von Fremdwörtern kaum Restriktionen vor (Lohde 2006, 192).

Es existieren auch einige Derivate, die auf eine verbale Basis zurückführbar sind: *werbemäßig*.

¹⁹Zur umstrittenen Debatte des Suffixstatus siehe von Polenz (1994, 293).

Semantik

Als Basen dienen hauptsächlich Personen (*merkelmäßig*), Tiere (*katzenmäßig*), Sachen/Gegenstände (*bleistiftmäßig*) und Abstrakte (*hassmäßig*). In diesem Zusammenhang macht Lohde (2006, 192) auf die Distinktion dreier semantischer Muster aufmerksam, wobei das 'limitative' das für ihn am häufigsten verwendete Muster ist. Weiterhin sei es seiner Meinung nach möglich, Vergleichs- oder Ähnlichkeitsbildungen zu konstruieren wie z.B. *robotermäßig*. Als dritte semantische Gruppe werden Derivate genannt, die eine 'verbindliche Entsprechung oder Norm' kennzeichnen: *ordnungsmäßig*.

Konkurrierend zu *-mäßig* kann das Suffix *-haft* und das Kompositazweitglied *-artig* aufgeführt werden. Lohde (2006, 193) exemplifiziert dieses Konkurrenzphänomen an den Wörtern *schauspielermäßig*, *schauspielerartig* und *schauspielerhaft*.

Produktivität

Das Suffix kennt keine Strukturbeschränkungen (Fleischer und Barz 1995, 265), so dass eine hohe Produktivität angenommen werden kann.

5.1.12 Das Suffix *-isch*

Geschichte

Das Suffix *-isch* geht auf eine althochdeutsche Form *isc* zurück (Krahe und Meid 1969, 194ff).

Basen

Es bindet sich an substantivische, verbale und adjektivische Basen. Liegt eine substantivische Basis vor, welche vorzugsweise als Input dient,²⁰ so kann sie sowohl simplizischer (*hündisch*) wie auch komplexer Natur (*abendländisch*) sein.

Sehr produktiv kann das Suffix auch an Eigennamen, in besonderem Maße an Länder- und Landschaftsnamen, affigiert werden.²¹

²⁰Siehe dazu Wellmann (1995, 529f.), der feststellt, dass 95% aller Derivate auf *-isch* als desubstantivisch angesehen werden müssen.

²¹*Englisch, Hessisch, Hamburgisch* etc.

Eine weitere Verwendungsmöglichkeit bieten Fremdwörter als Basis (*demokratisch*, *kommunistisch*), darunter auch viele Konfixe (Lohde 2006, 186). Außerdem darf an-gemerkt werden, dass auch relativierende Adjektive wie *ironisch* oder *mechanisch* ge-bräuchlich sind.

Bei Betrachtung der adjektivischen Basen fällt auf, dass vor allem nicht-native Ad-jektive wie *deskriptivisch* oder *substantivisch* gebildet werden.

Bei den wenigen verbalen Basen des Deutschen handelt es sich um Infinitivstämme, die oftmals durch ein Interfix *-er* erweitert worden sind: *quälerisch*.

Semantik

Häufig findet man (neben den Länder- und Landschaftsnamen) auch Personen- oder Tierbezeichnungen, die das semantische Muster komparativ ausdrücken, das dann eine negative Konnotation besitzt (ebd.): *schweinisch*.

Als Konkurrenz zu *-isch* kann *-haft* aufgeführt werden: *heldisch* vs. *heldenhaft*.

Produktivität

Das Suffix war bereits im Althochdeutschen produktiv (Schlaefer 1977) und ist seit-dem stark produktiv und reihenbildend geblieben (Fleischer und Barz 1995, Lohde 2006).

5.1.13 Das Suffix *-los*

Geschichte

Das Morphem *los* ist zugleich Suffix und Verbalpräfix. Das Suffix *-los* geht auf ein schon im Althochdeutschen existierendes freies Lexem *los* zurück, das die Bedeutung 'abgelöst' besaß (Lohde 2006, 191).

Basen

Suffigiert werden hauptsächlich simplizische (*papierlos*) und komplexe Substantive (*grundsatzlos*) nativer und nicht-nativer Herkunft. Dabei kann das Basislexem eine

Veränderung durchleben, indem oftmals ein *-e* getilgt wird, wie es beispielsweise in *freudlos* (aus: Freude) der Fall ist (ebd.).

Unter bestimmten Bedingungen ist eine Interfigierung mit einem Fugenelement notwendig: *reibungslos, frauenlos, kinderlos*.

Des Weiteren existieren, wenn auch nur sehr vereinzelt, auch deverbale Formen wie z.B. *neidlos*.

Semantik

Basis für eine Derivation mit *-los* kann sowohl eine Personen- oder Sachbezeichnung als auch ein Abstraktum sein. Durch das Suffix wird das Fehlen des dadurch Bezeichneten angegeben. Fleischer (1982, 274) spricht in diesem Zusammenhang auch von einer wertenden Funktion, denn das Nichtvorhandensein von einer Sache oder einer Person kann sowohl positiv im Sinne eines Vorzugs, als auch negativ im Sinne eines Mangels aufgefasst werden.²²

Synonymisch konkurriert das Suffix mit den beiden kompositionellen Zweitgliedern *-frei* und *-leer*; aus antonymischer Sicht kann angemerkt werden, dass eine Beziehung zu dem Zweitglied *-voll*²³ besteht (Lohde 2006, 192).

Produktivität

-los wird als äußerst produktiv, bzw. reihenbildend, beschrieben; es besteht kaum eine Beschränkung durch die morphologische Struktur (Fleischer und Barz 1995, 264).

5.1.14 Zusammenfassung

In den vorangegangenen Kapiteln wurden die zu untersuchenden Suffixe hinsichtlich ihrer Verwendung beschrieben. Neben Angaben zur semantischen Bindungsmöglichkeit findet man in der linguistischen Literatur auch Aussagen über die Produktivität der einzelnen Suffixe.

²²Als positive Konnotation nennt Lohde (2006, 191) die Beispiele *schuldlos* und *furchtlos*, als negative werden *respektlos* und *arbeitslos* genannt.

²³Dass diese Morpheme auch als sogenannte Halbaffixe oder Affixoide bezeichnet werden können, wird in Fandrych (1993, 101) diskutiert. Diese Frage soll allerdings an dieser Stelle unbeantwortet bleiben.

Als Zusammenfassung der deskriptiven Aussagen der Literatur lässt sich Tab. 5.1 aufstellen. Die Aussagen zur Produktivität von Fleischer und Barz (1995) dürfen nicht wörtlich genommen werden: Zum Teil liegen klare Beschreibungen vor, dass ein Affix produktiv sei oder nicht, jedoch gilt dies nicht für alle Fälle; bei einigen Suffixen allerdings finden sich Aussagen wie z.B., dass es '(stark) reihenbildend' sei. Derartige Aussagen wurden interpretiert und dann mit der entsprechenden Etikettierung wie beispielsweise '(stark) produktiv' gleichgesetzt:

	Fleischer und Barz (1995)	Lohde (2006)	Motsch (2004)
-tum	schwach produktiv	gering produktiv	beschränkt aktiv
-heit/-keit	produktiv	produktiv	stark aktiv
-nis	nicht produktiv	gering produktiv	inaktiv
-ung	hoch produktiv	produktiv	stark aktiv
-sal	unproduktiv	keine Produktivität	keine Aussage
-lich	(stark) produktiv	hoch produktiv	stark aktiv
-ig	stark produktiv	hoch produktiv	stark aktiv
-mäßig	stark produktiv	hoch produktiv	stark aktiv
-bar	hoch produktiv	hoch produktiv	stark aktiv
-sam	nicht produktiv	gering produktiv	inaktiv
-isch	produktiv	hoch produktiv	stark aktiv
-los	stark produktiv	hoch produktiv	stark aktiv
-haft	keine klare Aussage	produktiv	stark aktiv

Tabelle 5.1: Zusammenfassung der Aussagen der linguistischen Literatur zur Produktivität der zu untersuchenden substantiv- und adjektivbildenden Suffixe

Bei Betrachtung der Aussagen in Tab. 5.1 wird deutlich, dass nahezu alle adjektivbildenden Suffixe – die Ausnahme bildet *-sam* – als produktiv und/oder aktiv bezeichnet werden; damit scheinen sie fast alle ein vergleichbares Wortbildungsverhalten aufzuweisen.

Nun stellt sich die Frage, ob eine quantitative Analyse der jeweiligen Frequenzen dies auch numerisch belegen kann. Liegt für alle produktiven Suffixe eine ähnliche Anzahl an Wortbildungen und Neubildungen vor oder gibt es bei den produktiven Suffixen diesbezüglich vielleicht Unterschiede, die klar mit Zahlen belegt werden können?

Es gilt zu beantworten, was es bedeutet, wenn in der Literatur von einem produktiven Suffix gesprochen und wie sich z.B. produktive von stark produktiven Suffixen unterscheiden.

Des Weiteren gilt es natürlich auch zu prüfen, ob und wie sich die Unproduktivität von *-sal* und die schwache Produktivität von *-nis* und *-tum* durch eine quantitative Studie belegen lässt.

5.2 Datengrundlage

Die Daten, die in dieser quantitativen Studie untersucht werden, wurden aus einem frei verfügbaren Korpus des Deutschen, dem Kernkorpus des DWDS (Digitale Wörterbuch der deutschen Sprache), erhoben.

Das von der *Deutschen Forschungsgemeinschaft* geförderte Projekt gilt als erstes zeitliches und nach Textsorten ausgewogenes Textkorpus der deutschen Sprache des 20. Jahrhunderts.

Die Korpora des DWDS und damit auch das Kernkorpus sind lemmatisiert, mit Wortinformationen²⁴ versehen und mit einer linguistischen Suchmaschine²⁵ abfragbar. Das Kernkorpus des DWDS soll den deutschen Wortschatz möglichst ausgewogen repräsentieren.²⁶ Aufgrund dessen wurde die folgende Aufteilung des Korpus gewählt: schöne Literatur (ca. 26%), journalistische Prosa (ca. 27%), Fachprosa (ca. 22%), Gebrauchstexte (ca. 20%), (Transkribierte) Texte gesprochener Sprache (ca. 5%).²⁷

5.3 Exkurs: Deutsche Korpora

Der folgende Exkurs in die deutsche Korpuslinguistik ist für diese Arbeit von großer Relevanz, denn unabhängig davon, dass es notwendig ist zu beschreiben, wie eine Korpusstudie durchgeführt wird, ist es ebenso wichtig, die derzeit verfügbaren Ressourcen der Korpuslinguistik der deutschen Gegenwartssprache vorzustellen.

²⁴Dazu siehe http://www.dwds.de/erschliessung/pos_tagger.

²⁵Zur Verwendungsmöglichkeit der Suchanfragen siehe <http://www.dwds.de/erschliessung/suchmaschine>.

²⁶Dazu siehe auch Klein (2004a), der die Entstehungsgeschichte und die Zusammensetzung des DWDS beschreibt.

²⁷Zur genauen Aufstellung der Texte siehe <http://www.dwds.de/textbasis/kerncorpus>.

Per definitionem ist ein Korpus „eine Sammlung schriftlicher oder gesprochener Äußerungen“ (Lemnitzer und Zinsmeister 2006). Idealerweise sind die Korpusdaten digitalisiert und über eine benutzerfreundliche Schnittstelle abrufbar.

Korpora können sehr unterschiedlich aufgebaut sein. Merkmale zur Klassifizierung benennt Scherer (2006), indem sie skizziert, dass man aus formaler Perspektive zwischen computerlesbaren und Papierkorpora, Gesamt- und Teilkorpora²⁸, Proben- und Volltextkorpora²⁹, statische Korpora und Monitorkorpora, unterscheidet.³⁰ Korpora lassen sich allerdings auch hinsichtlich ihrer Aufbereitung unterscheiden, denn sie können annotiert oder nicht-annotiert sein.³¹ Auf der inhaltlichen Ebene sind Korpora der mündlichen versus der schriftlichen Sprache zu unterscheiden³², ebenso, ob es sich um ein ein- oder mehrsprachiges Korpus handelt und, welchen Zeitraum das Korpus abdeckt (historisches Korpus versus Korpus der Gegenwartssprache)³³.

²⁸Ein Teilkorpus liegt beispielsweise dann vor, wenn man ein Korpus zu einer literarischen Epoche zusammenstellt und dabei mehrere Autoren berücksichtigt. Die einzelnen Teilkorpora der jeweiligen Autoren bilden zusammen das ganze Korpus.

²⁹Bei einem Volltextkorpus sind die Texte in vollständiger Länge enthalten; bei einem Probenkorpus hingegen sind die Texte auf eine bestimmte Größe reduziert worden, sodass sie möglicherweise nur ausschnittsweise vorliegen.

³⁰Ein weiteres Klassifikationskriterium für Korpora ist das der *Beständigkeit*. Es existieren Korpora, deren Zusammensetzung von Anfang an feststeht. Ist das Korpus erstellt worden, wird im Nachhinein nichts mehr an ihm verändert. Deshalb bezeichnet man diese Art von Korpora auch als statische Korpora. Anders sieht es bei sogenannten *Monitorkorpora* aus. Ihre Zusammensetzung verändert sich im Laufe der Zeit. Ein Beispiel für ein deutsches Monitorkorpus ist das Mannheimer-Morgenkorpus, abrufbar unter <http://www.ids-mannheim.de/kl/projekte/korpora/archiv/mm.html>, 1999 enthielt es 64 Millionen Wörter und am Ende des Jahres 2003 bereits 141,7 Millionen Textwörter.

³¹Es gibt mittlerweile eine Vielzahl an Korpora, die linguistisch annotiert ist. Am häufigsten findet man als Annotation morphologische oder syntaktische Informationen, dennoch gibt es für den deutschen Bereich auch Korpora, die prosodisch annotiert sind. So zum Beispiel ist es bei den *Korpora des Bayerischen Archivs für Sprachsignale* (BAS), abrufbar unter <http://www.phonetik.uni-muenchen.de/Bas/BasKorporadeu.html>.

³²Dass die geschriebenen Korpora dominieren, ist verständlich, da der Aufwand zur Korpusprache der gesprochenen Sprache enorm hoch ist. Außerdem ist die Verfügbarkeit von geschriebenen Daten wesentlich höher.

³³Dies hängt natürlich auch davon ab, wie der Begriff Gegenwartssprache definiert wird. Der Kreis der dazugehörigen Dokumente lässt sich sehr klein fassen und es lässt sich annehmen, dass Texte, die aus der Zeit der neunziger Jahre bis heute stammen, hineingehören oder aber auch jene Texte dazuzählen, die in den Fünfzigern und Sechzigern entstanden sind. Lassen sich Korpora mit Texten von Goethe oder Kant finden, würden sie zweifelsohne als historische Korpora bezeichnet werden. Wie allerdings mit Texten umzugehen ist, die zu Beginn des 20. Jahrhunderts entstanden sind, ist Definitionssache des anwendenden Linguisten.

5.3.1 Das Durchführen einer Korpusstudie

Eine korpuslinguistische Untersuchung besteht in der Regel aus mehreren Phasen:

1. Formulierung einer Fragestellung/Hypothese
2. Korpusuche und/oder Korpuserstellung mit anschließender Aufbereitung des Korpus
3. Datenerhebung
4. Datenaufbereitung
5. Auswertung und Interpretation der Daten

Bei der Formulierung einer oder mehrerer für diese Arbeit relevanten Fragestellungen ist vorab zu überlegen, welche Strukturen und/oder Einheiten relevant sind, und ob es sich um eine qualitative oder quantitative Studie handelt. Ebenso ist die zeitliche Dimension zu überdenken, indem entschieden werden muss, ob synchron oder diachron gearbeitet werden soll, da diese Entscheidung Einfluss auf den Inhalt und die Größe des Korpus haben kann.

Die nächste Entscheidung, die sich daran anschließt, ist jene, ob man mit einem bereits vorhandenen Korpus arbeiten will oder nicht, denn sind die vorhandenen Korpora für die angedachte Studie ungeeignet, so muss ein neues Korpus erstellt werden.

Die Aussagekraft einer Studie hängt allerdings nicht primär von der Größe des Korpus oder der elektronischen Form ab, sondern von der Interpretation der Korpusdaten. Wurde beschlossen, ein eigenes Korpus zu erstellen, so müssen Überlegungen zur Größe, zum Inhalt, zur Beständigkeit und zur Repräsentativität erfolgen.

Zur Aufbereitung eines Korpus lässt sich festhalten, dass grundsätzlich zwei Möglichkeiten vorhanden sind: entweder eine *Korpusannotation* oder eine *Belegklassifikation*. Bei einer Annotation wird das gesamte Korpus bearbeitet, was bedeutet, dass jedes einzelne Wort annotiert wird, während bei einer Belegklassifikation nur diejenigen Wörter klassifiziert werden, die für die Studie relevant sind.

Für die Datenerhebung ist vorab zu überlegen, ob eine automatische Datenerhebung möglich ist, um effektiv und ökonomisch arbeiten zu können. Dabei gilt in der Regel, dass sich der Linguist mit der (Abfrage-)Syntax des Korpus vertraut macht. Exemplifizierend soll an dieser Stelle exemplarisch die Suchabfrage mit dem DWDS aufgeführt werden³⁴: wird bspw. das Wort *Apfel* eingegeben, so werden alle flektierten Formen

³⁴Siehe dazu http://www.dwds.de/help/ddc/#part_2, abgerufen am 28.02.2011 um 15:53 Uhr.

des Wortes gefunden (*Apfels, Äpfel, Apfel*); die Suche nach einer bestimmten Flexionsform wird durch das Voransetzen des Klammeraffens ermöglicht, die Suche nach *Apfel** bringt alle Wörter mit der Buchstabenfolge *Apfel* in initialer Stellung hervor, sodass bspw. auch Komposita mit *Apfel* als Erstglied gefunden werden.

In vielen Fällen ist es erforderlich, die Daten nach ihrer Erhebung weiter aufzubereiten. Auch für die Berechnungen in dieser Arbeit musste eine manuelle Aufbereitung der Daten erfolgen, die an späterer Stelle noch skizziert wird.

Im letzten Schritt steht der Linguist vor der für ihn schwierigsten Aufgabe, indem er die gefundenen Ergebnisse im Hinblick auf seine vorab gestellten Fragen interpretieren muss; mögliche Fragen lauten:

- Konnten die aufgestellten Hypothesen bestätigt werden?
- Welche Konsequenzen haben die gefundenen Ergebnisse für die untersuchte Sprache oder gar das Sprachsystem?
- Welche Forschungsfragen lassen sich anschließen und an welchen Stellen besteht ein besonderer Forschungsbedarf für die zukünftige Linguistik?

5.3.2 Verfügbare freie Korpora für das Deutsche und ihre Probleme

Obwohl für die deutsche Sprache einige große verfügbare und freie Korpora verfügbar sind wie z.B. das *Digitale Wörterbuch der deutschen Sprache des 20. Jahrhunderts* (DWDS)³⁵ oder die Korpora des IDS³⁶, ist schnell feststellbar, dass diese Korpora hauptsächlich zum Untersuchen der Gegenwartssprache geeignet sind und nur eine begrenzte Zeitspanne beinhalten.

Um Aussagen zum derivationalen Sprachwandel zu treffen, wäre es natürlich von Vorteil, die Produktivität von Derivationssuffixen über einen längeren Zeitraum zu untersuchen. Ein derartiges Korpus müsste ausgewogen sein und für jede Zeitspanne dieselbe Anzahl an Wörtern besitzen. Ein Korpus dieser Art existiert derzeit für das Deutsche noch nicht.

Allerdings sei an dieser Stelle bemerkt, dass das Deutsche Textarchiv (DTA) eine historische Ausweitung des DWDS bis 1650 plant:

³⁵ Abrufbar unter <http://www.dwds.de/>.

³⁶ Abrufbar unter <http://www.ids-mannheim.de/kl/corpora.html>.

„Ziel des Deutschen Textarchivs (DTA) ist die Erstellung eines disziplinenübergreifenden Volltextkorpus deutschsprachiger Texte. Grundlage hierfür bilden digitale Faksimiles historischer Druckwerke (Entstehungszeit der Drucke zwischen ca. 1650 und 1900). Das Korpus umfasst Werke verschiedener Textsorten, literarischer Gattungen und wissenschaftlicher Disziplinen. Ziel des Projekts ist die Bereitstellung und aufgrund der großen Textmenge weitestgehend automatische linguistische Aufbereitung eines vielseitigen, umfangreichen Textbestandes auf dieser heterogenen Grundlage.“³⁷

Im Mai 2011 sind nach der ersten Bearbeitungs- und Korrekturphase 532 Bücher im DTA online. Die Auswahl der Bild- und Textdigitalisierung der Werke erfolgte unter lexikographischen Gesichtspunkten. Primär wurden dabei originär deutschsprachige Werke und Übersetzungen von Werken, die einen großen Einfluss auf den deutschsprachigen Kulturraum besaßen, mit einbezogen.³⁸

Für das Deutsche existieren derzeit nur kleine historische Korpora, so z.B. das *Bonner Frühneuhochdeutsch Korpus*. Insgesamt besteht dieses Korpus aus 40 Texten, die sich aus Handschriften und bestimmten Editionen zusammensetzen.³⁹ Die Texte stammen aus zehn deutschen Sprachräumen und lassen sich jeweils vier Zeitabschnitten zuordnen: 1350 - 1400, 1450 - 1500, 1550 - 1600 und 1650 - 1700. Des Weiteren sind in drei Formaten vorhanden: Als XML- oder HTML-kodierte Fassung oder als Original. Wortformen wurden manuell mit einem eigenen Tag-Set ausgezeichnet. Verben, Nomen und Adjektive sind mit Informationen zu Wortart, Flexionsklasse und grammatischen Merkmalen versehen (Finitheit, Person, Numerus, Tempus und Modus für Verben, Genus und Kasus für Nomen und Adjektive). Das Korpus ist frei verfügbar.⁴⁰ Leider sind keine Suchmöglichkeiten vorhanden, was in hohem Maße nachteilig wirkt und das Arbeiten mit diesem Korpus einschränkt.

Das *Bochumer Mittelhochdeutsch Korpus*⁴¹ besteht aus Texten aus dem Zeitraum 1070-1350. Sie stammen aus verschiedenen Sprachräumen: Oberdeutsch, Bairisch, Bairisch-Alemannisch, Alemannisch Westmitteledeutsch, Hessisch-Thüringisch,

³⁷<http://media.dwds.de/dta/doku/dta-leitlinien.pdf>, abgerufen am 08.05.2011 um 11:29 Uhr.

³⁸Zur Textauswahl siehe http://www.deutschestextarchiv.de/project/#part_21, abgerufen am 08.08.2011 um 11:35 Uhr.

³⁹Vgl. dazu auch Lender und Wegera (1982).

⁴⁰Siehe dazu <http://www.ikp.uni-bonn.de/dt/forsch/fnhd/>.

⁴¹Abrufbar unter www.ruhr-uni-bochum.de/wegera/archiv.

Schwäbisch Mittelfränkisch, Rheinfränkisch, Ostmittelfränkisch, Ostfränkisch. Die Texte sind in Editionen digitalisiert, aber Angaben zur Annotation und zu den Suchmöglichkeiten sind nicht vorhanden.

Eine andere, in diesem Zusammenhang erwähnbare Datenbank nennt sich *Mittelhochdeutsche Wörterbücher und Digitales Mittelhochdeutsches Textarchiv*⁴². Es handelt sich hierbei um ein DFG-Projekt, das zum Ziel hat, das Mittelhochdeutsche Wörterbuch von Benecke/Müller/Zarncke (BMZ), das Mittelhochdeutsche Handwörterbuch von Lexer und das Findebuch zum mittelhochdeutschen Wortschatz zu digitalisieren. Das Korpus soll 148 mittelhochdeutsche Texte beinhalten und die Einträge der Wörterbücher sollen mit Belegstellen in einem digitalen Belegarchiv verknüpft werden. Die Texte sind im TEI XML-Format codiert. Die Metainformationen bestehen u.a. aus Informationen zu Autor, Titel, Gliederung des Textes (Kapitel, Verse), Anmerkungen des Editors, Seite und Zeile. Es existiert eine Volltextsuche über den gesamten Inhalt. Die Vorkommen der Wortformen können einfach aufgelistet werden und als Konkordanz oder im Kontext angezeigt werden.

Bei dem Projekt *TITUS*⁴³ von Jost Gippert⁴⁴ handelt es sich um ein Textarchiv, das Texte aus dem gesamten indogermanischen Sprachraum aus verschiedenen Zeiträumen enthält.⁴⁵ Bei den Texten, die über TITUS verfügbar sind, handelt es sich um Digitalisierungen in Druckform vorliegender Texte. Zur Annotation lässt sich sagen, dass Informationen zu Sprache bzw. Sprachstufe, Text, Autor, Kapitel, Absatz, Vers, Seite, Zeile und Lemmainformationen zugänglich sind. Für den wissenschaftlichen Gebrauch sind die Materialien frei verfügbar; über ein Webinterface ist eine KWIC-Suche⁴⁶ nach Wortformen möglich.

Bei den *Codices Electronici Ecclesiae Coloniensis (CEEC)*⁴⁷ handelt es sich um Kodices der Erzbischöflichen Diözesan- und Dombibliothek Köln⁴⁸, die als Bilddaten eingescannt wurden, sodass die Texte nicht elektronisch verfügbar sind, was eine empirische Studie zum Sprachwandel erschwert.

⁴²Erreichbar unter <http://kompetenzzentrum.uni-trier.de/projekte/kooperationsprojekte/mhgta>.

⁴³Abrufbar unter <http://titus.uni-frankfurt.de/>.

⁴⁴Vgl. Gippert (1995) und Gippert (2002).

⁴⁵Die genaue Aufstellung der Texte ist unter <http://titus.uni-frankfurt.de/texte/texte.htm> zugänglich.

⁴⁶KWIC steht für Keyword In Context (Schlüsselwort im Kontext). Dadurch wird das alphabetische Sortieren der Wörter eines Textes erstellt. Für weitere Informationen siehe dazu Gaus (2001).

⁴⁷Dazu siehe <http://www.ceec.uni-koeln.de/>.

⁴⁸Siehe <http://www.ceec.uni-koeln.de>.

Die Mittelhochdeutsche Begriffsdatenbank (MHDBDB)⁴⁹ besteht aus 126 Texten; insgesamt wurden 5,7 Mill. Wörter digitalisiert. Die MHDBDB ist Grundlage einiger Konzeptwörterbücher und wird laufend erweitert. Suchen lässt sich nach einzelnen Wörtern, Lemmata und Wortarten; auch existiert eine Begriffsliste, die eine Konzeptsuche ermöglicht. Die Texte sind morphologisch annotiert und lemmatisiert. Die Recherche verläuft über ein Webinterface.

Das *GerManC*-Projekt wurde an der Universität Manchester entwickelt.⁵⁰ Ziel des Projektes ist es gewesen, ein repräsentatives historisches Korpus für das Deutsche für die Jahre 1650 bis 1800 aufzubauen. Dabei wurde das Korpus in drei Teilkorpora eingeteilt (1650-1700, 1701-1750, 1751-1800). Um regionale Varietäten auszuschließen, wurden die Texte auf fünf verschiedenen Regionen entnommen: 'North German', 'West Central German', 'East Central German', 'West Upper German', 'East Upper German'. Für jede Region gibt es drei Auszüge aus einer Zeitung mit jeweils etwa 2000 Wortformen; es handelt sich deshalb um ein recht kleines Korpus.

Die Bibliotheca Augustana⁵¹ ist ein Projekt, das einige digitale Dokumente vom 8. bis zum 20. Jahrhundert beinhaltet. Man findet demnach althochdeutsche, mittelhochdeutsche, frühneuhochdeutsche und neuhochdeutsche Literatur.

Nachteilig an diesem Korpus ist, dass diese Texte nur htmlisiert⁵² sind und man sie zur Weiterverarbeitung manuell kopieren muss.

Zusammenfassend lässt sich deshalb festhalten, dass es für die deutsche Sprache einige frei verfügbare Korpora gibt, die das Untersuchen einer bestimmten beschränkten Zeitspanne erlauben; ein großes diachron ausgerichtetes Korpus existiert bisher nicht. Die kleinen historischen Korpora sind aufgrund ihrer geringen Tokenzahl nicht für eine

⁴⁹ Abrufbar unter <http://mhdbdb.sbg.ac.at:8000/>.

⁵⁰ Für nähere Informationen siehe <http://www.llc.manchester.ac.uk/research/projects/germanc/>.

⁵¹ Anschaubar unter http://www.hs-augsburg.de/~harsch/germanica/Chronologie/d_chrono.html.

⁵² Bei der Hypertext Markup Language (HTML) handelt es sich um eine Auszeichnungssprache, mit welcher Texte und multimediale Elemente für das World Wide Web aufbereitet werden. Dazu müssen den Elementen und damit auch den Korpustexten vorab eine logische Struktur gegeben werden, indem sie mit Metainformationen (z.B. zur Markierung von Überschriften) versehen werden, die für den Nutzer des Webs später nicht sichtbar sind. Der Vorteil der Htmlisierung von Texten liegt darin, dass die Linearität aufgehoben werden kann: Hypertexte besitzen eine n:m-Struktur, die dem Leser Wahlfreiheit bezüglich der zu rezipierenden Texten ermöglicht. Insbesondere machen Online-Zeitungen seit Jahren davon Gebrauch, indem der Leser beim Rezipieren eines Zeitungsartikels in der Regel auf weitere ähnliche oder bereits zu diesem Thema erschienenen Artikel verwiesen wird. Dabei können die Beziehungsverhältnisse der Ausgangs- und Zielmodule unterschiedlicher Natur sein: 1:1, n:1, 1:n oder n:n.

diachrone Studie geeignet, da geringe Datenmengen nicht zur Produktivitätsmessung ausreichen.

Aus diesem Grund wurde entschieden, das DWDS als Korpus zu verwenden. Doch auch diesbezüglich müssen ein paar kritische Überlegungen aufgeführt werden. Zunächst einmal ist zu sagen, dass das Korpus mit 95% aus schriftlicher Kommunikation besteht und nur 5% mündlicher Kommunikation vorhanden sind. Gerade hier wäre es interessant, die Suffixe im Hinblick auf ihre Produktivität in der mündlichen und schriftlichen Kommunikation separat zu untersuchen. Es ist davon auszugehen, dass sich es diesbezüglich Produktivitätsunterschiede gibt.

Ein weiteres Problem ergibt sich aus der den unterschiedlichen Datenmengen der einzelnen Jahre. Im folgenden Abschnitt wird auf dieses Problem vertiefend eingegangen.

5.4 Untersuchungszeitraum

Wie bereits angemerkt, erfasst das DWDS sprachliche Daten des Deutschen in einem Zeitraum von 1900 bis 1999. Dieser Zeitraum wurde für diese Studie insofern komplett genutzt, indem fünf Untersuchungszeitpunkte, bzw. -räume, verwendet worden sind: 1900, 1925, 1950, 1975 und 1999.

Dennoch muss an dieser Stelle kritisch hinzugefügt werden, dass für die Zeiträume unterschiedlich große Dokumentenmengen zur Verfügung standen:

- 1900: 1.332.089 Token
- 1925: 1.195.773 Token
- 1950: 1.190.450 Token
- 1975: 588.810 Token
- 1999: 1.621.890 Token

Dieser Auflistung kann entnommen werden, dass die Textmenge (hinsichtlich der Tokenanzahl) für 1900, 1925 und 1950 vergleichbar ist. Für 1975 dagegen war nur etwa die Hälfte der Textmenge verfügbar. Für 1999 liegt eine etwas größere Tokenmenge, die untersucht werden kann.

Als Linguist hat man in diesem Fall zwei Möglichkeiten, mit diesen Daten zu arbeiten: die erste Möglichkeit besteht darin, die Tokenanzahl zu neutralisieren, um für alle

Messzeitpunkte dieselbe Anzahl an Wortformen zur Verfügung zu haben. Im Fall der Nutzung des DWDS würde dieses bedeuten, für alle Zeitpunkte nur 588.810 Token zu betrachten. Dabei bleibt allerdings nicht aus, dass Information verloren geht, sodass für diese Studie bewusst entschieden wurde, alle verfügbaren Token zu nutzen.

Begründet werden kann die Entscheidung damit, dass zumindest für 1900, 1925 und 1950 eine vergleichbare Datenmenge verfügbar ist. Die Ergebnisse, die auf Basis der Daten von 1975 erzielt werden, werden im Folgenden zwar hinzugezogen, aber mit Vorsicht betrachtet und dementsprechend kritisch geprüft.

5.5 Methode der Auswertung

Um quantitative Aussagen zur Produktivität und zum Produktivitätswandel treffen zu können, wird wie folgt vorgegangen: vor Beginn der korpusbasierten Arbeit ist ein kleines Vorexperiment durchgeführt worden, bei welchem Sprecher des Deutschen möglichst viele Wörter mit den dreizehn Suffixen in einem begrenzten Zeitraum von fünf Minuten produzieren sollten. Die vorliegenden Ergebnisse sollen als Intuition dienen. Geprüft werden soll insbesondere, ob die Ergebnisse des Experimentes die Aussagen der Literatur decken. Ein Suffix, das produktiv ist, wird viele Wortbildungen hervorbringen. Mit einem Suffix, das unproduktiv ist, lassen sich nur stark lexikalisierte Bildungen abrufen und keine neuen mehr produzieren.

Im einem nächsten Schritt wird korpusbasiert gearbeitet. Die Daten des DWDS werden quantitativ untersucht, indem die in Kap. 4 vorgestellten Maße zur Produktivitätsmaße angewendet wurden.

Die Auswertung der Daten erfolgt nach den in Kap. 4.2 und Kap. 4.3 vorgestellten Produktivitätsmaßen:

- Tokenhäufigkeit
- Typenhäufigkeit
- die Relation von Typen und Token
- P
- P^* und im letzten Schritt auch durch ein
- LNRE-Modell⁵³

⁵³Die Erklärung des Modells erfolgt im Zuge der Diskussion in Kapitel 8.

Um die Datenreihen zur Produktivitätsmessung anwenden zu können, war eine Bereinigung der Daten erforderlich. Im ersten Schritt wurden alle Tokenlisten der zu untersuchenden Affixe für alle Messzeiträume extrahiert. Im zweiten Schritt wurden dann alle Komposita (halbautomatisch) in ihre Einzelteile zerlegt, da sonst eine Verfälschung der Ergebnisse nicht auszuschließen ist. Die Erstglieder wurden nicht weiter beachtet, sodass nur die determinierenden Zweitglieder in die Berechnungen einfließen konnten.⁵⁴

Im nächsten Schritt wurden Tippfehler und orthografische Fehler beseitigt. Danach wurde das Problem der Homophonie betrachtet und entsprechende Homophone wurden gelöscht. Das Löschen der Homophone stellte dabei eine zeitaufwendige Aufgabe dar, da einzelne Belege sorgfältig überprüft werden mussten. So z.B. ist bei der Buchstabenfolge *Jugendhaft* nicht kontextfrei zu entscheiden, ob es sich um ein Adjektiv oder ein Substantiv handelt. Letzteres zeigt ein Kompositum und darf nicht als Wortbildung des Derivationsuffixes *-haft* mitgezählt werden.

Nach Bereinigung der Daten wurden die Tokenlisten in Typenlisten umgewandelt, was automatisch – mit einem für die vorliegende Arbeit geschriebenen Programm – erfolgte. Danach wurden die oben genannten Häufigkeitsmaße angewandt.

Die Ergebnisse werden aus zweierlei Hinsicht interpretiert:

1. **Synchron:** Um die Aussagen der Literatur (Tab. 5.1) zu überprüfen, werden die Ergebnisse des Jahres 1999 interpretiert.
2. **Diachron:** Um mögliche Sprachwandel-Tendenzen im Wortschatz des 20. Jahrhunderts sichtbar zu machen, werden die Ergebnisse der fünf Messzeitpunkte miteinander verglichen.

⁵⁴Es ist anzumerken, dass in den meisten bisherigen Arbeiten zu dieser Thematik keine Angaben dazu gemacht werden, ob Komposita zuvor in ihre Einzelteile zerlegt wurden, sodass nur die determinierenden Zweitglieder zur statistischen Auswertung hinzugezogen wurden. Eine Ausnahme ist die Dokumentation der Datenaufbereitung in Evert und Lüdeling (2001, Kap. 3).

Das Vorexperiment

In nachfolgenden Abschnitt sollen erste Hinweise auf die Produktivität der zu untersuchenden Suffixe anhand der Ergebnisse eines kleinen Vorexperimentes erfolgen.

6.1 Versuchsbeschreibung

Mit einem Suffix, das produktiv ist, lassen sich viele (neue) Wörter bilden. Dagegen wird ein unproduktives Suffixes nicht zu (regelbasierten) Neubildungen führen. Die Suffixe, die in der vorliegenden Korpusstudie untersucht werden, wurden im vorangegangenen aufgelistet und beschrieben. Interessant ist in diesem Zusammenhang nicht nur die Frage, welche Schlüsse sich durch die Korpusstudie ziehen lassen, sondern auch die Frage nach dem Zusammenhang von Produktivität und der Produktion von Wörtern mit einem bestimmten Suffix.

Um diesem Problem auf den Grund zu gehen, wurden sechs Versuchspersonen gebeten, innerhalb eines vorgegebenen Zeitraums von fünf Minuten möglichst viele Wörter mit den jeweiligen Suffixen zu bilden.

Den Versuchsteilnehmern wurde vorher die Information gegeben, dass es sowohl möglich sei, schon bestehende und häufig genutzte Wörter zu bilden als auch Wortbildungen zu schaffen, die 'man in keinem Lexikon' finden würde, die ihnen aber sinnvoll erscheinen, sodass andere Sprachbenutzer in der Lage sein würden, das Wort (kontextfrei) zu verstehen.

Das jeweilige Suffix wurde den Versuchspersonen auf einem Schild gezeigt; mit Hochhalten eines Schildes seitens des Moderators begannen die fünf Minuten der Produktionsphase.

6.1.1 Untersuchungsgegenstand

In diesem Versuch wurden dabei dieselben Suffixe untersucht, die auch Gegenstand der vorliegenden Korpusstudie (Kap. 7) sind, zum einen die substantivbildenden Suffixe *-nis*, *-heit/-keit*, *-tum*, *-sal* und *-ung*, und zum anderen die adjektivbildenden Suffixe *-ig*, *-isch*, *-lich*, *-bar*, *-sam*, *-mäßig*, *-los* und *-haft*.

6.1.2 Versuchspersonen

An dem Vorexperiment nahmen sechs Versuchspersonen teil. Dabei handelte es sich um Studenten und Studentinnen der Germanistik im Alter von 22 bis 27 Jahren. Alle von ihnen waren Muttersprachler des Deutschen und besaßen keine detaillierten Kenntnisse über die statistische Messung der Produktivität von Affixen. Der Versuch wurde rechnerbasiert durchgeführt, die zu erstellenden Listen wurden von den Versuchspersonen direkt in eine Excel-Datei geschrieben.

6.1.3 Methode der Auswertung

Die erhobenen Daten wurden unter mehreren unterschiedlichen Gesichtspunkten ausgewertet: Häufigkeitsdaten (Wortformen und Typen) und Lexikalisiertheit. Um eine Unterscheidung zwischen Lexikalisiertheit und Nicht-Lexikalisiertheit treffen zu können, wurde jede Bildung mit einem Online-Lexikon (canoo.net) überprüft.

6.2 Ergebnisse

6.2.1 Gesamtanzahl der Bildungen

Zu allererst soll die Anzahl der Bildungen betrachtet werden, die jede Versuchsperson mit einem bestimmten Suffix produziert hat. Bei den substantivbildenden Suffixen ergeben sich die in Tab. 6.1 ersichtlichen Frequenzen.

	VP1	VP2	VP3	VP4	VP5	VP6	Ø
<i>-tum</i>	11	7	9	11	7	15	10
<i>-heit/-keit</i>	38	28	19	21	20	28	25.7
<i>-nis</i>	14	12	11	5	10	14	11
<i>-ung</i>	76	52	26	30	39	32	42.5
<i>-sal</i>	4	3	2	2	2	2	2.5

Tabelle 6.1: Anzahl der Wortbildungen mit den substantivbildenden Suffixen *-tum*, *-heit/-keit*, *-nis*, *-ung* und *-sal* der einzelnen Versuchspersonen.

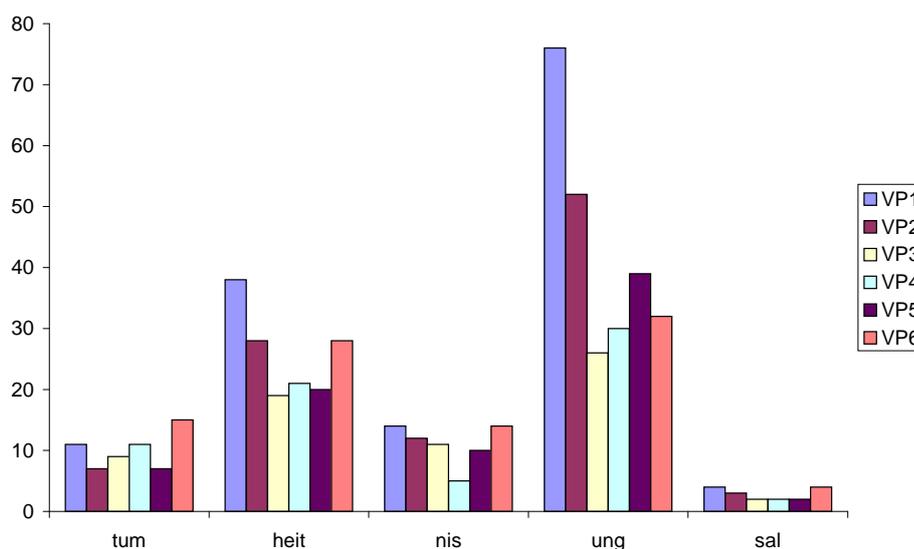


Abbildung 6.1: Anzahl der Wortbildungen mit den substantivbildenden Suffixen *-tum*, *-heit/-keit*, *-nis*, *-ung* und *-sal* der einzelnen Versuchspersonen.

Bei Betrachtung der Suffixe im Einzelnen fällt auf, dass *-ung* dasjenige Suffix ist, das bei allen Versuchspersonen zu den meisten Wortbildungen führt – im Durchschnitt 42.5. Doch auch mit *-heit/-keit* werden durchschnittlich vergleichsweise viele Wörter gebildet: ca. 26.

Mit *-tum* und *-nis* wurden nur mäßig viele Wortbildungen erzeugt und mit *-sal* in allen sechs Fällen sehr wenige. Lediglich eine Versuchsperson konnte vier Derivate mit *-sal* nennen; in der Regel wurden allerdings nur zwei Formen gebildet.

Für die adjektivbildenden Suffixe konnten die folgenden Häufigkeiten erhoben werden:

	VP1	VP2	VP3	VP4	VP5	VP6	Ø
<i>-lich</i>	33	26	29	20	25	18	25.2
<i>-ig</i>	34	25	13	8	14	17	18.5
<i>-mäßig</i>	34	25	17	11	23	12	20.4
<i>-bar</i>	45	37	30	25	28	23	31.4
<i>-sam</i>	9	7	6	9	8	8	7.8
<i>-isch</i>	48	30	13	21	12	22	24.4
<i>-los</i>	41	29	31	17	25	28	28.5
<i>-haft</i>	9	18	18	13	13	10	13.5

Tabelle 6.2: Anzahl der produzierten Wortbildungen mit den adjektivbildenden Suffixen *-lich*, *-ig*, *-mäßig*, *-bar*, *-sam*, *-isch*, *-haft* und *-los* der einzelnen Versuchspersonen.

Anhand von Tab. 6.2 und Abb. 6.2 lässt sich erkennen, dass *-sam* mit durchschnittlich 7.8 Bildungen ein unproduktives Verhalten aufweist. Die Suffixe *-haft* und *-ig* sind mit ihren durchschnittlich 10 bis 20 Bildungen mäßig produktiv.

Vier Suffixe, *-lich*, *-mäßig*, *-isch* und *-los*, bringen einen Durchschnittswert hervor, der zwischen 20 und 30 Token liegt und weisen in diesem Kontext ein produktives Produktionsverhalten auf. Das Suffix mit den meisten Bildungen – durchschnittlich 31 – ist *-bar*; damit kann es unter diesem Aspekt als besonders produktiv bezeichnet werden.

6.2.2 Typenhäufigkeit

Im nächsten Schritt werden die gebildeten Wörter der einzelnen Versuchspersonen unter dem Aspekt der Typenhäufigkeit zusammengefasst. Dabei liegt die folgende These zugrunde: Je produktiver ein Suffix ist, desto mehr unterschiedliche Wörter können mit diesem Suffix gebildet werden.

Tab. 6.3 zeigt die Typenhäufigkeiten der substantivbildenden Suffixe.

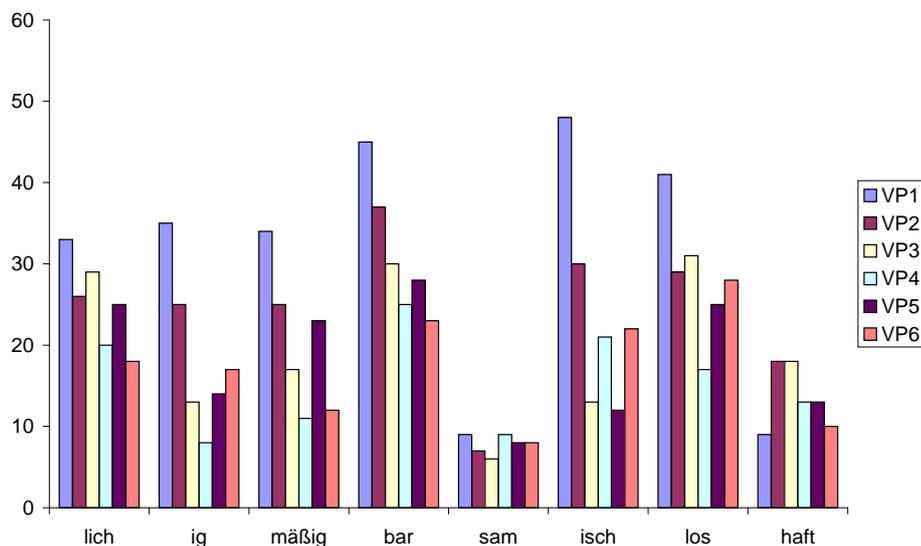


Abbildung 6.2: Anzahl der Wortbildungen mit den adjektivbildenden Suffixen *-lich*, *-ig*, *-mäßig*, *-bar*, *-sam*, *-isch* und *-los* der einzelnen Versuchspersonen.

Das Suffix *-sal* bringt dabei mit insgesamt nur sechs Typen die wenigsten unterschiedlichen Wortbildungen hervor. Die sechs Versuchspersonen produzierten zusammen nur sechs Wortbildungen.¹

Die beiden Suffixe *-tum* und *-nis* scheinen hinsichtlich ihrer möglichen Wortbildungen ein sehr ähnliches Verhalten aufzuweisen, indem sie eine vergleichsweise hohe Anzahl an Wörtern produzieren lassen. Die Allomorphe *-heit* und *-keit* hingegen können unter diesem Aspekt als weitaus produktiver angesehen werden, da etwa viermal mehr unterschiedliche Wörter mit ihnen gebildet werden als mit *-nis* und *-tum*.

Doch besonders herausragend ist das Suffix *-ung*, das mit 220 Typen Platz 1 in der Rangfolge belegt.

Für die adjektivbildenden Suffixe konnten die in Tab. 6.4 ersichtlichen Zahlen erhoben werden.

¹*Schicksal, Rinnsal, Trübsal, Mühsal, Labsal und Scheusal.*

Suffix	Typen
<i>-tum</i>	36
<i>-heit/-keit</i>	124
<i>-nis</i>	38
<i>-ung</i>	220
<i>-sal</i>	6

Tabelle 6.3: Typenhäufigkeit der substantivbildenden Suffixe *-tum*, *-heit/-keit*, *-nis*, *-ung* und *-sal*.

Suffix	Typen	Suffix	Typen
<i>-lich</i>	103	<i>-sam</i>	36
<i>-ig</i>	91	<i>-isch</i>	100
<i>-mäßig</i>	102	<i>-los</i>	122
<i>-bar</i>	146	<i>-haft</i>	62

Tabelle 6.4: Typenhäufigkeit der adjektivbildenden Suffixe *-lich*, *-ig*, *-mäßig*, *-bar*, *-sam*, *-isch*, *-haft* und *-los*.

Um die Typenhäufigkeit zu bewerten, wird die folgende Kategorisierung angenommen: bei 0 bis 30 Typen wird von einem unproduktiven Verhalten (-) gesprochen, bei über 30 bis 100 Typen von einer mäßigen Produktivität, bei 100 bis 200 Typen von einem produktiven Verhalten und bei über 200 Typen ist davon auszugehen, dass es sich um eine sehr hohe Produktivität handelt. Die Bewertung dieses Parameters erfolgt in der Zusammenfassung des sechsten Kapitels.

6.2.3 Andere Überlegungen

Typen können Aufschluss darüber geben, wie stark ein Suffix in vergangenen oder gegenwärtigen Zeitstufen zur Erweiterung des Wortschatzes beigetragen hat oder möglicherweise noch beiträgt. In Kapitel 4 wurde das Baayensische Maß *P* vorgestellt, das auf der Anzahl der Hapax Legomena basiert. In diesem Zusammenhang wurde kritisch darauf hingewiesen, dass es durchaus sinnvoll wäre, auch die zweimaligen Bildungen, die sogenannten Dislegomena, oder vielleicht auch die dreifachen Vorkommen, mit einzubeziehen.

Eine Methode, sich die das Produktionsverhalten von Affixen anzuschauen, ist die Erstellung eines Frequenzspektrums, aus welchem ersichtlich wird, wie hoch die Anzahl der einmaligen, zweimaligen oder auch sechsmaligen Bildungen etc. jeweils ist.²

Tabelle 6.5 repräsentiert die Frequenzen der substantivbildenden Suffixe.

Suffix	einmal	zweimal	dreimal	viermal	fünfmal	sechsmal
<i>-tum</i>	18	14	1	0	1	1
<i>-heit/-keit</i>	98	20	4	1	0	0
<i>-nis</i>	26	6	5	2	1	0
<i>-ung</i>	189	28	2	1	0	0
<i>-sal</i>	1	2	0	1	0	1

Tabelle 6.5: Frequenzen der Wortlisten zu den substantivbildenden Suffixen in absoluter Anzahl

Bei Betrachtung von Tab. 6.5 fällt auf, dass deutliche Unterschiede im Frequenzverhalten der substantivbildenden Suffixe vorliegen. Für *-ung* und *-heit/-keit* ist ersichtlich, dass sie viele niederfrequent auftretende Wörter erzeugen lassen; es lässt sich der Tabelle entnehmen, dass kein einziges Wort vorhanden ist, das fünfmal oder sechsmal gebildet wird.

Für das Suffix *-sal* lässt sich die Beobachtung machen, dass insgesamt nur sehr wenige Wörter gebildet werden, die auch gleichzeitig in hoher Anzahl vorkommen.³

Sowohl für die substantivbildenden Suffixe als auch für die adjektivbildenden Suffixe soll die Frequenzverteilung bewertet werden. Dazu ist es notwendig, ein geeignetes Maß zu finden, das hier angewandt werden kann.

Es bietet sich an, jene Werte zu betrachten, die sich ergeben, wenn die Anzahl der einmalig vorkommenden Bildungen durch die Typenanzahl geteilt wird.

Für die substantivbildenden Suffixe ergeben sich unter diesem Aspekt die folgenden Werte:

- *-tum*: 0.5
- *-heit/-keit*: 0.79

²An dieser Stelle sei angemerkt, dass die Anwendung von Frequenzspektren zur Bewertung der morphologischen Produktivität eine entscheidende Rolle spielen, indem LNRE-Modelle hinzugezogen werden. Auch in dieser Arbeit erfolgt eine Anwendung eines LNRE-Modells, sodass an der entsprechenden Stelle in aller Genauigkeit auf die Rolle von Frequenzspektren einzugehen ist.

³Bei dem von allen Versuchspersonen genanntem Wort handelt es sich um das Wort *Schicksal*.

- *-nis*: 0.68
- *-ung*: 0.86
- *-sal*: 0.17

Auch für die adjektivbildenden Suffixe soll die Verteilung der Frequenzen detailliert betrachtet werden (Tab. 6.6).

Suffix	einmal	zweimal	dreimal	viermal	fünfmal	sechsmal
<i>-lich</i>	70	21	6	4	1	0
<i>-ig</i>	77	8	5	1	0	0
<i>-mäßig</i>	90	7	3	1	1	0
<i>-bar</i>	121	16	9	2	0	0
<i>-sam</i>	27	5	2	1	0	0
<i>-isch</i>	75	13	6	4	1	1
<i>-los</i>	84	30	6	1	1	0
<i>-haft</i>	52	8	3	1	0	0

Tabelle 6.6: Frequenzen der Wortlisten zu den adjektivbildenden Suffixen in absoluter Anzahl

Aus den Ergebnissen in Tab. 6.6 ergeben sich die folgenden Werte in Bezug auf die Relation von Einmalbildungen und Typen:

- *-lich*: 0.68
- *-ig*: 0.85
- *-mäßig*: 0.88
- *-bar*: 0.83
- *-sam*: 0.75
- *-isch*: 0.75
- *-los*: 0.69
- *-haft*: 0.84

Zur Interpretation dieser Werte existiert kein etabliertes Schema, vielmehr müssen die Werte in Relation zueinander interpretiert werden.

Für die in dieser Studie zu untersuchenden Suffixe ist es sinnvoll, die Abstufung wie folgt vorzunehmen: bei einem Wert bis zu 0.59 darf von Unproduktivität ausgegangen werden (-); liegt der Wert zwischen 0.6 und 0.69, so wird das Suffix in diesem Zusammenhang als mäßig produktiv betrachtet (O), bei Ergebnissen zwischen 0.7 und 0.79

kann ein produktives Verhalten abgeleitet werden (+) und ab einem Wert ab 0.8 darf für diesen Parameter von einer hohen Produktivität (++) gesprochen werden.

6.2.4 Lexikalisiertheit

Schaut man sich die von den Versuchspersonen gebildeten Wörter genauer an, so stößt man möglicherweise auf Wörter, die bisher nicht zum eigenen aktiven Wortschatz gehörten. Beispielfhaft können die folgenden Derivate genannt werden:

- **-los:** *autolos, ticketlos, helmlos, uniformlos, deckellos*
- **-heit/-keit:** *Wörtlichkeit, Polierbarkeit, Bezahlbarkeit, Vernehmbarkeit, Begehbarkeit*
- **-mäßig:** *unimäßig, löffelmäßig, kaffeemäßig, rabattmäßig, flaschenmäßig, suppenmäßig, zuckermäßig, papiermäßig, geldmäßig, konzertmäßig*
- **-bar:** *leerbar, bedruckbar, umklappbar, markierbar*

All diese Wortformen sind jedoch in sich verständlich: ein Soldat, der abends ausgeht, geht möglicherweise *uniformlos* aus, die *Begehbarkeit* der Plantage wurde durch ein Raubtier gefährdet, die Veranstaltung wurde *konzertmäßig* gestaltet etc.

Und obwohl die regelbasierten Formen nachweislich existieren und von Sprechern der deutschen Sprache gebildet worden sind, sind sie zum Teil nicht in einem Wörterbuch zu finden. In diesem Zusammenhang kommt die Frage auf, ob die von den Versuchspersonen gebildeten Wörtern im Hinblick auf ihre Lexikalisiertheit untersucht werden können und wenn ja, wie.

Bei Lexikalisierungsprozessen besteht ein Kontinuum, sodass es zwischen klar lexikalisierten Wörtern und brandaktuellen Innovationen einen breiten Übergangsbereich gibt, den man nur schwer abgrenzen kann, da von unterschiedlichen Vorkommenshäufigkeiten in unterschiedlichen Registern, kommunikativen Gattungen, Sprechsituationen usw. ausgehen ist.

Sicher jedoch ist Eines: **ein Wort, das in einem Lexikon aufgelistet ist, kann als lexikalisiert gelten.**⁴

⁴Es wird an dieser Stelle auch darauf hingewiesen, dass alle Wörterbücher unabhängig von ihrer Größe nur eine unvollständige Abdeckung besitzen. Die Neubearbeitung des größten deutschen Wörterbuches (Grimm) (http://dwb.uni-trier.de/Projekte/WBB2009/DWB/wbgui_py?lemid=GA00001) nimmt in aller Regel nur diejenigen Wörter auf, die in mindestens zehnfacher Ausführung vorhanden sind, viele Wörter auch erst bei dreißigfacher Belegung. Während das Grimm'sche Wörterbuch

Fehlende Lexikalisiertheit ist ein klares Definitionsmerkmal von Wortneubildungen (Barz 1999, 22). Ein Suffix, das ein produktives Wortbildungsverhalten aufweist, wird viele Bildungen schaffen, die nicht lexikalisiert sind.

Das Problem, das sich für die Auswertung der Wortlisten der Versuchspersonen ergibt, ist die klare Operationalisierung und Messbarkeit der Lexikalisierung, denn gerade gedruckte Wörterbücher können aufgrund ihres Defizits an Aktualität und Vollständigkeit mit der tatsächlichen Entwicklung des Wortschatzes nicht Schritt halten. Des Weiteren ist anzumerken, dass (gedruckte) Wörterbücher nicht alle Bildungen umfassen, die möglich und auch belegt sind, denn Reihenbildungen, die an offene Paradigmen anschließen, werden manchmal nicht vollständig aufgelistet. Im Wörterbuch finden sich in diesen Fällen oftmals nur Repräsentanten der Wortbildungsmuster. Aber gerade die Möglichkeit der Reihenbildung lässt viele neue Wörter entstehen und darf als Merkmal für eine hohe Produktivität angesehen werden.

Um die Wörter der Versuchspersonen auf ihre Lexikalisiertheit hin zu testen, wurde *canoo.net*, ein Online-Portal, „Deutsche Wörterbücher und Grammatik“, verwendet.

Ein Ergebnis lautet wie folgt: keines der obigen Beispiele findet sich im Wörterbuch von *canoo.net*. Die Google-Suche der Wörter allerdings zeigt, dass sie, obwohl sie zumindest nach *canoo.net* nicht lexikalisiert sind, in niedrigen bis mäßigen Häufigkeiten belegt werden können. So zum Beispiel lässt sich über Google das Wort *helmlos* insgesamt 14.700 Mal auffinden⁵.

Auf ihr Vorkommen im Wörterbuch wurden alle Wörter der Versuchspersonen getestet. Daraufhin wurde ermittelt, wie hoch die jeweilige Anzahl der nicht-lexikalisierten Wörter mit dem jeweiligem Suffix ist.

Als erstes sollen die substantivbildenden Suffixe in Tab. 6.7 in Betracht gezogen werden.

Die jeweiligen Prozentwerte verdeutlichen, dass die Anzahl der nicht-lexikalisierten Wörter in allen Fällen sehr gering ist. Den höchsten Wert findet das Suffix *-heit* hervor, den geringsten *-sal*, weil alle wenigen mit *-sal* gebildeten Wörter auch lexikalisiert sind.

etwa 330.000 Wörter beinhaltet, finden sich im gesamten Textbestand des DWDS insgesamt rund 4,5 Millionen Types.

⁵Abgerufen am 29.11.10 um 10:45 Uhr. Auch die Duden-Suche zu *helmlos* bleibt im Übrigen erfolglos.

Suffix	Nicht-Lexikalisiertheit
<i>-tum</i>	6.7%
<i>-heit/-keit</i>	7.1%
<i>-nis</i>	1.5%
<i>-ung</i>	3.5%
<i>-sal</i>	0%

Tabelle 6.7: Anzahl der mit substantivbildenden Suffixen produzierten Wörter, die nicht lexikalisiert sind

Die 1.5% bei *-nis* ergeben sich durch das Wort *Fairnis*, das von einem einzigen Sprecher gebildet worden ist. Dieser Fall ist als kritisch zu betrachten, da diese Bildung in Analogie zu *fairness* entstanden ist. Dennoch wäre es denkbar, dass sich die deutsche analoge Variante dieses Wortes aufgrund der phonologischen Ähnlichkeit verbreiten könnte. Immerhin liefert Google am 11.09.09 um 15:09 Uhr 85.700 Bildungen des Wortes *Fairnis*, macht aber gleichzeitig auch den Vorschlag für das Wort *fairness*, das zum Abrufzeitpunkt 16.800.000 Bildungen im Web und 727.000 Bildungen auf deutschen Webseiten erscheinen lässt.

Auch für die adjektivbildenden Suffixe wurde eine Unterscheidung zwischen Lexikalisiertheit versus Nicht-Lexikalisiertheit vorgenommen (Tab. 6.8).

Suffix	Nicht-Lexikalisiertheit
<i>-lich</i>	0.66%
<i>-ig</i>	5.4%
<i>-mäßig</i>	59%
<i>-bar</i>	30%
<i>-sam</i>	6.4%
<i>-isch</i>	0.7%
<i>-los</i>	26.3%
<i>-haft</i>	28.4%

Tabelle 6.8: Anzahl der nicht-lexikalisierten Wörter der mit den adjektivbildenden Suffixen produzierten Wörter in Prozent

Hier zeigt sich ein deutlicher Unterschied zu den den substantivbildenden Suffixen. Obwohl auch unter den adjektivbildenden Suffixen einige fast nur lexikalisierte Bildungen hervorbringen, insbesondere *-lich* und *-isch*, sticht das Suffix *-mäßig* heraus, weil 59% aller Bildungen mit diesem Suffix nicht lexikalisiert sind.

6.3 Zusammenfassung

Das in den vergangenen Abschnitten präsentierte Vorexperiment soll einen ersten Überblick über die zu untersuchenden substantiv- und adjektivbildenden Suffixe geben.

Nachfolgend dienen Tab. 6.9 und Tab. 6.10 als schematische Zusammenfassung der untersuchten Merkmale: - steht dabei für Unproduktivität, O für mäßige Produktivität, + für Produktivität und ++ für eine besonders hohe Produktivität.

Die Ergebnisse können wie folgt analysiert werden: Werte zwischen 0 und 10% zeigen eine hohe Lexikalisiertheit und damit Unproduktivität (-), Werte zwischen 11 bis 30% der nicht-lexikalisierten Wörter werden als Indiz für mäßige Produktivität (O) gewertet, Werte zwischen 31 und 50% deuten auf ein produktives Verhalten (+) hin und bei mehr als über 50% nicht-lexikalisierten Wortbildungen ist davon auszugehen, dass das Suffix in der Lage ist, viele Neubildungen zu schaffen und es wird als hoch produktiv (++) beschrieben.

Suffix	Wortformen	Typen	Frequenzverteilung	Nicht-Lexikalisiertheit	Gesamt
<i>-tum</i>	-/O	O	-	-	-/O
<i>-heit</i>	+	+	+	-	+
<i>-nis</i>	O	O	O	-	(-)/O
<i>-ung</i>	++	++	++	-	++
<i>-sal</i>	-	-	-	-	-

Tabelle 6.9: Zusammenfassung der Ergebnisse der substantivbildenden Suffixe *-nis*, *-ung*, *-sal*, *-heit/-keit/-igkeit* und *-tum*

Betrachtet man nun die Zusammenfassung der Ergebnisse, so lässt sich schlussfolgern, dass *-sal* als völlig unproduktiv, *-nis* und *-tum* als unproduktiv bis mäßig produktiv, *-heit/-keit* als mäßig produktiv und *-ung* als produktiv bezeichnet werden darf.

Diese ersten Ergebnisse bestätigen nicht alle in der Literatur beschriebenen intuitiven Behauptungen zur Produktivität der fünf Suffixe, denn für *-ung* und die Allomorphe *-heit/-keit* wird angenommen, dass sie hoch produktiv zur Wortbildung beitragen. Zwar zeigen die Ergebnisse des Versuchs, dass die sechs Versuchspersonen mit *-ung* viele verschiedene Wortbildungen produzieren, doch es handelt sich bei diesen fast ausschließlich um lexikalisierte Bildungen, demnach Bildungen, die bereits fest im Wortschatz des Deutschen verankert sind. Dasselbe kann auch für *-heit/-keit* beob-

achtet werden. Für *-nis*, *-tum* und *-sal* hingegen scheinen die Versuchsergebnisse die Aussagen der Literatur zu bestätigen.

Tab. 6.10 zeigt die Zusammenfassung für die adjektivbildenden Suffixe.

Suffix	Wortformen	Typen	Frequenzverteilung	Nicht-Lexikalisiertheit	Gesamt
<i>-lich</i>	+	+	O	-	O
<i>-ig</i>	O	O	++	-	+
<i>-mäßig</i>	+	+	++	++	+ / ++
<i>-bar</i>	++	++	++	+	++
<i>-sam</i>	-	O	+	-	O
<i>-isch</i>	+	+	+	-	(O) / +
<i>-los</i>	+	+	O	+	+
<i>-haft</i>	O	O	++	+	+

Tabelle 6.10: Zusammenfassung der untersuchten Merkmale der adjektivbildenden Suffixe *-bar*, *-sam*, *-haft*, *-lich*, *-los*, *-ig*, *-mäßig* und *-isch*

Die Auswertung macht deutlich, dass *-ig* und *-sam* nur als beschränkt produktiv gelten. Für *-sam* ergibt sich damit eine Parallele zu den in der Literatur angegebenen Beschreibungen, es sei nach Fleischer und Barz (1995) „nicht produktiv“, nach Lohde (2006) „gering produktiv“ oder nach Motsch (2004) „inaktiv“. Bei *-ig* allerdings sprechen alle Autoren von einer starken Produktivität (Tab. 5.1), die an dieser Stelle so nicht erkannt werden kann. Ein ähnliches Bild ergibt sich bei dem Suffix *-isch*, das in der Literatur als produktiv bis stark produktiv beschrieben wird und nach Meinung von Fleischer und Barz (1995, 260) „zu den wichtigsten Suffixen für die Adjektivderivation“ gehört; auch diese Aussage lässt sich nicht bestätigen. Dasselbe gilt auch für *-lich*, denn nach Angaben der Literatur gehört dieses Wortbildungsmorphem „zu den wichtigsten Suffixen für die desubstantivische Adjektivderivation“ und ist „in starkem Maße auch mit Verben und Adjektiven verbindbar“ (ebd.). Für *-los* und *-haft* kann nach den untersuchten Kriterien geschlossen werden, dass sie produktiv zur Wortbildung beitragen. Die Aussagen der Literatur sprechen sogar für eine hohe Produktivität, die allerdings in dem Vorexperiment nicht gefunden werden konnte; erst recht nicht, wenn die zwei Suffixe mit den Suffixen *-bar* und *-mäßig* verglichen werden, die sich insofern von den anderen beiden unterscheiden, weil sie noch produktiver erscheinen. Während sich *-bar* insbesondere durch eine hohe Typenanzahl profiliert, heben sich die Ergebnisse von *-mäßig* durch einen anderen Faktor ab: die hohe Anzahl der nicht-lexikalisierten Wörter.

Vornehmlich die letzte Beobachtung ist auffällig. Aus diesem Grund sollen nun weitere Anmerkungen zu diesem Suffix gemacht werden. Auf der Online-Seite des Dudens⁶ findet man einen Hinweis zur Verwendung und damit auch Produktivität des Suffixes *-mäßig*:

„Das Suffix *-mäßig* wird neben den Bedeutungen »in der Art von, wie« (z. B. *gewohnheitsmäßig*, *serienmäßig*) und »entsprechend, gemäß« (z. B. *gesetzmäßig*, *planmäßig*) auch im Sinne von »in Bezug auf, hinsichtlich« verwendet, wobei dieser Gebrauch eine Modeerscheinung ist, die erst in letzter Zeit in der Umgangssprache aufgekommen und erstaunlich verbreitet ist. Aus Bequemlichkeit werden häufig ganz willkürlich Neuschöpfungen geprägt, die oft eher zweifelhaft sind. *Farbenmäßig* hat sie einen wirklich guten Geschmack. *Personalmäßig* ist der Betrieb unterversorgt. *Notenmäßig* hat er sich in diesem Schuljahr erheblich verbessert. *Intelligenzmäßig* ist sie ihm haushoch überlegen. Solche Formulierungen sind zwar recht häufig, sie sind aber sehr unschön und oft schlicht überflüssig. Wenn durch Bildungen mit *-mäßig* nur knappe, präpositionale Fügungen oder substantivische Wendungen ersetzt werden, die den Bezug erläutern sollen, sollte man sie vermeiden: Bei der Auswahl der Farben hat sie einen wirklich guten Geschmack. Der Betrieb hat zu wenig Personal. Seine Noten sind in diesem Schuljahr wesentlich besser geworden. Sie ist ihm an Intelligenz haushoch überlegen.“

Wenn *-mäßig* gerade in der Umgangssprache und damit insbesondere in der mündlichen Kommunikation als besonders produktiv gilt, ist es nicht verwunderlich, dass sich deshalb auch viele nicht-lexikalisierte Wörter in diesem Versuch zeigen lassen.

Die in diesem Abschnitt vorgestellten Ergebnisse skizzieren nun ein erstes Bild der Produktivität der einzelnen Suffixe, das im Folgenden durch die Korpusstudie präzisiert werden soll.

⁶Siehe dazu http://www.duden.de/deutsche_sprache/sprachberatung/newsletter/archiv.php?id=55; abgerufen am 13.09.09 um 14:53Uhr.

Ergebnisse der Korpusstudie mit dem DWDS

7.1 Die substantivbildenden Suffixe

7.1.1 Tokenhäufigkeit

Die durch das DWDS-Kernkorpus verfügbaren Datenmengen der einzelnen Suffixe unterscheiden sich in ihrer Größe; insbesondere für das Suffix *-ung* stehen große Datenreihen zur Verfügung, von denen die 5818 Token des Jahres 1975 die kleinste Datenreihe darstellen. Es wurde deshalb an dieser Stelle entschieden, für jeden Zeitraum 5000 *-ung*-Token für die weiteren Analysen zu berücksichtigen.¹ Die Auswahl der Token erfolgte zufällig.

Insgesamt konnten die in Tab. 7.1 präsentierten Datenmengen extrahiert werden.

Aufgrund der unterschiedlich großen Datenreihen bietet es sich an, die Daten zu normieren; gäbe es genau 1.000.000 Million Token für jeden Messzeitraum, sähen die Tokenzahlen folgendermaßen aus:

¹Dass die tatsächliche Tokenanzahl für jeden Zeitpunkt geringfügig kleiner ist als 5000, ergibt sich durch die Bereinigung der Daten. Dazu siehe Kap. 5.5.

	1900	1925	1950	1975	1999
<i>-ung</i>	4963 (14482)	4937 (15633)	4958 (9190)	4952 (5818)	4910 (16185)
<i>-tum</i>	281	250	87	121	334
<i>-nis</i>	2245	1459	735	768	1902
<i>-heit/-keit</i>	7181	4676	3289	2572	5735
<i>-sal</i>	89	151	97	31	122

Tabelle 7.1: Tokenhäufigkeit der substantivbildenden Suffixe *-ung*, *-tum*, *-nis*, *-heit/-keit* und *-sal* für die Zeiträume 1900, 1925, 1950, 1975 und 1999

	1900	1925	1950	1975	1999
<i>-ung</i>	10862	13312	13359	9891	10035
<i>-tum</i>	211	210	73	206	207
<i>-nis</i>	1684	1226	617	1306	1179
<i>-heit/-keit</i>	5386	3928	2763	4372	3556
<i>-sal</i>	67	127	81	53	76

Tabelle 7.2: Tokenhäufigkeit (normiert auf eine Million Token) der substantivbildenden Suffixe *-ung*, *-tum*, *-nis*, *-heit/-keit* und *-sal* für die Zeiträume 1900, 1925, 1950, 1975 und 1999

Für das Suffix *-ung* ergeben sich Schwankungen von bis zu knapp 400 Token; die Tokenmenge von *-tum* erscheint bis auf das Jahr 1950 in seiner Größe konstant zu sein; es werden jeweils etwas über 200 Wortformen gefunden.

Ebenso wird man auch für *-nis* beobachten, dass im Jahre 1950 die wenigsten Wortformen extrahiert werden können. Für die anderen Jahre kann festgestellt werden, dass – ähnlich wie bei *-ung* – Schwankungen vorhanden sind. Und obwohl die Tokenfrequenzen zu allen Zeitpunkten wesentlich geringer als bei *-ung* sind, liegt ein Schwankungsbereich einer Größe von 400 Token vor.

Die Tokenfrequenzen für *-sal* sind für jeden Zeitraum sehr gering und liegen nach Normierung zwischen 53 und 127 Wortformen.

Die Werte von *-heit/-keit* zeigen, dass auch die Tokenzahlen der Allomorphe schwanken und auch im Jahre 1950 die geringsten Häufigkeiten erhoben werden. In der Gesamtschau impliziert dies, dass im Jahre 1950 (mit Ausnahme von *-ung*) insge-

samt weniger Derivationen gefunden werden und das Jahr demnach eine Ausnahme darstellt. Gründe können hier nur spekulativ aufgeführt werden. In Kapitel 2.4 wurde beschrieben, dass das Deutsche zu derivationellen Strukturen auch Alternativen anzubieten hat. Möglicherweise wurden diese in diesem Jahr bevorzugt produziert. Ein anderer Grund kann darin liegen, dass sich die Zusammensetzung der Textsorten über das DWDS im Jahre 1950 von den anderen Erhebungsjahren unterscheidet, was zu unterschiedlichen Textstilen führt.²

Bei Betrachtung der Frequenzen von *-heit* und *-keit* in Tabelle 7.1 ist auffällig, dass es zwei Zahlen gibt, die die 5000er-Grenze überschreiten: die Werte von 1900 und von 1999. Die hohen Werte ergeben sich aus der Tatsache, dass *-heit* und *-keit* separat betrachtet wurden, um die Möglichkeit nutzen zu können, die Produktivität von Allomorphen anhand einer möglichst großen Datenmenge zu beschreiben. Schaut man sich nun die geringen Werte für die Jahre 1950 und 1975 an, steht man vor der Entscheidung, ob die Tokenzahlen zu neutralisieren sind. Dies würde zu einem erheblichen Informationsverlust führen. Die Entscheidung für oder gegen eine Neutralisierung darf demnach gleichgesetzt werden mit der Entscheidung für oder gegen eine separate Betrachtung der Allomorphe. Es wurde entschieden, mit allen verfügbaren Daten und ohne Informationsverlust zu arbeiten.

Tab. 7.3 listet die Frequenzen der Allomorphe *-heit* und *-keit* separat auf.

	1900	1925	1950	1975	1999
<i>-heit</i>	3068	1945	1625	1205	2436
<i>-keit</i>	4113	2731	1664	1367	3299

Tabelle 7.3: Tokenhäufigkeit der Allomorphe *-heit* und *-keit* für die Zeiträume 1900, 1925, 1950, 1975 und 1999

Anhand der Daten lässt sich zunächst feststellen, dass das Allomorph *-keit* häufiger im Korpus vertreten ist als *-heit*; das gilt für alle Zeiträume. Während sich die Werte von 1950 und 1975 der beiden Suffixe nur schwach voneinander unterscheiden, gibt es insbesondere für 1900 und 1999 einen größeren Unterschied bezüglich der Tokenmengen. Auch für die Allomorphe erfolgte eine Normierung der Werte; sie sind Tab. 7.4 zu entnehmen. Durch die Normierung der Daten wird deutlich, dass die geringste Tokenanzahl wiederholt für das Jahr 1950 zu verzeichnen ist, sodass ein deutlicher

²Zum Zusammenhang von Produktivität und Textsorte siehe Baayen (1994).

	1900	1925	1950	1975	1999
<i>-heit</i>	2301	1634	1365	2049	1510
<i>-keit</i>	3085	2294	1398	2324	2045

Tabelle 7.4: Tokenhäufigkeit der Allomorphe *-heit* und *-keit* für die Zeiträume 1900, 1925, 1950, 1975 und 1999

Unterschied zum Jahre 1975 ersichtlich wird. Dennoch sind auch im Vergleich zu den anderen Messzeitpunkten frequentielle Schwankungen vorhanden.

An dieser Stelle muss darauf hingewiesen werden, dass eine hohe Tokenzahl nicht gleichzusetzen ist mit einer hohen Produktivität eines Affixes, vielmehr kann es sein, dass eine (oder mehrere) Bildungen, die stark lexikalisiert sind, in einem Text in hoher Anzahl erscheinen können, unter anderem deshalb, weil sie polysem verwendet werden.

Aus diesem Grund ist es im nächsten Schritt der Analyse notwendig, die Typenhäufigkeiten hinzuzuziehen, die selbstverständlich anhand der nicht-normierten Daten erhoben wurden.

7.1.2 Typenhäufigkeit

Für die substantivbildenden Suffixe konnten die in Tab. 7.5 präsentierten Häufigkeiten ermittelt werden.

	1900	1925	1950	1975	1999
<i>-ung</i>	592	660	578	558	689
<i>-tum</i>	39	55	22	18	60
<i>-nis</i>	45	47	39	28	46
<i>-heit/-keit</i>	847	721	454	284	793
<i>-sal</i>	7	7	2	1	5

Tabelle 7.5: Typenhäufigkeit der substantivbildenden Suffixe *-ung*, *-tum*, *-nis*, *-heit/-keit* und *-sal* für die Zeiträume 1900, 1925, 1950, 1975 und 1999

Unstrittig ist, dass das Suffix *-sal* als unproduktiv bezeichnet werden darf. Es ist zu allen Zeitpunkten nur in sehr wenigen Wortbildungen vorhanden, 1975 sogar nur in

einer einzige Wortbildung: *Schicksal*. Während 1900 und 1925 noch sieben verschiedene Wortbildungen vertreten sind, finden sich 1999 nur fünf unterschiedliche Wörter im Korpus.

Schwieriger gestaltet sich die Interpretation der Typenhäufigkeiten von *-nis* und von *-tum*. Beide Suffixe produzieren jeweils nur eine geringe Anzahl an unterschiedlichen Bildungen, wobei ein Unterschied zu verzeichnen ist: der Schwankungsbereich mit *-tum* ist deutlicher als jener mit *-nis*. *-nis* ist in 28 bis 47 Wortbildungen vertreten, Bildungen mit *-tum* hingegen weisen insbesondere im Vergleich von 1975 und 1999 einen relativ großen Unterschied auf: 1975 findet man 18 Typen, 1999 jedoch 60 Wortbildungen. Im Vergleich zu *-nis* im Jahre 1999 bringt *-tum* somit 14 Wortbildungen mehr hervor. Das Suffix *-ung* scheint – zumindest auf den ersten Blick – für alle Messzeiträume produktiv zu sein, da zwischen 558 und 689 Typen erzeugt werden. Ein Suffix, das über 500 unterschiedliche Typen generieren kann, lässt schon zu diesem Zeitpunkt die Vermutung aufkommen, dass es als produktiv deklariert werden kann.

Auch die Allomorphe *-heit* und *-keit* scheinen einen wichtigen Beitrag zur deutschen Wortbildung zu leisten, auch wenn enorme Schwankungen in der Realisierung von möglichen Bildungen erkennbar sind: sie sind 1975 lediglich in 284 Bildungen vertreten, 1900 jedoch werden insgesamt 847 Wörter gebildet. Auch für 1999 deutet eine sehr hohe Typenanzahl darauf hin, dass es sich bei den Allomorphen um sehr produktive Suffixe handelt. Die Typenanzahl scheint nicht korpusunabhängig zu sein und es gibt eine Korrelation zwischen Token- und Typenanzahl, was sich beim Vergleichen der Jahre 1975 und 1900 und 1999 zeigt. Interessant ist nun, wie sich die einzelnen Allomorphe im Vergleich zueinander verhalten. Die nachfolgende Tabelle soll zur Betrachtung hinzugezogen werden:

	1900	1925	1950	1975	1999
<i>-heit</i>	265	245	158	93	199
<i>-keit</i>	581	476	296	191	594

Tabelle 7.6: Typenhäufigkeit der Allomorphe *-heit* und *-keit* für die Zeiträume 1900, 1925, 1950, 1975 und 1999

Während die Tokenhäufigkeit der beiden Allomorphe für einige Zeiträume vergleichbar gewesen ist, werden nun Unterschiede deutlich: *-keit* ist in Bezug auf seine

Typenhäufigkeiten weitaus produktiver als *-heit*. Die Differenz ist insbesondere in den Jahren 1900 und 1999 auffällig.

7.1.3 Typen-Token-Relationen

Um weitere Interpretationen vornehmen zu können, ist es nun notwendig, Typen und Token in Beziehung zueinander zu setzen und die Typen-Token-Relationen (TTR) zu betrachten. Die entsprechenden Werte sind in Tabelle 7.7 zusammengefasst.

	1900	1925	1950	1975	1999
<i>-ung</i>	0.12	0.13	0.12	0.11	0.14
<i>-tum</i>	0.14	0.22	0.25	0.15	0.18
<i>-nis</i>	0.01	0.03	0.05	0.04	0.02
<i>-heit/-keit</i>	0.12	0.15	0.14	0.11	0.14
<i>-sal</i>	/	/	/	/	/

Tabelle 7.7: Typen-Token-Relationen (TTR) der substantivbildenden Suffixe *-ung*, *-tum*, *-nis*, *-heit/-keit* und *-sal* für die Zeiträume 1900, 1925, 1950, 1975 und 1999

Für das Suffix *-sal* wurden keinerlei Berechnungen vorgenommen, da die Betrachtung der einzelnen Frequenzen von Token und Typen bereits gezeigt hat, dass es sich hierbei um ein unproduktives Suffix handelt. Weitere Interpretationen hinsichtlich der Produktivität sind deshalb nicht erforderlich. Insbesondere bei sehr niedrigen Häufigkeitsdaten ist sowohl die Betrachtung der TTRs als auch das nach Baayen anwendbare Maß *P* problematisch.

Für das Suffix *-ung* fällt auf, dass die TTR-Werte relativ konstant sind, insbesondere dann, wenn man 1975 aufgrund der geringeren Tokenanzahl ausschließt: 1900 und 1950 erhält man einen Wert von 0.12, 1952 einen leicht höheren Wert von 0.13; der höchste Wert wird im Jahre 1999 erreicht und liegt bei 0.14.

Schwieriger ist die Interpretation der Werte für die anderen substantivbildenden Suffixe, da deutlichere Schwankungen zu verzeichnen sind. So z.B. liegen bei *-tum* Schwankungen zwischen 0.14 und 0.26 vor. Ungeachtet der Tatsache, dass eine geringe TTR für den Zeitraum 1975 festgestellt wird, so ist die TTR auch im Jahr 1900 mit 0.14 vergleichbar niedrig. Im Jahre 1999 wird ebenfalls ein im Vergleich zu 1925 und 1950 niedriger Wert erkannt: 0.18. Den höchsten Wert mit 0.25 erhält man für die Zeiträume 1950 und 1925.

Ebenso ist es sehr schwierig, eindeutige Interpretationen für *-nis* vorzunehmen, da die Werte zwischen 0.01 und 0.05 schwanken und der höchste oder niedrigste Wert von 0.05 nicht für 1999 berechnet wird, sondern für 1950. Diese Daten zeigen daher keinen klaren Aufwärts- oder Abwärtstrend hinsichtlich der Produktivität.

Auch *-heit* und sein Allomorph sind hinsichtlich ihrer TTRs nicht konstant. Selbst nach Ausklammerung des 1975er-Wertes bleibt eine Spanne von 0.12 und 0.15 stehen. Auch hier sollen die einzelnen Werte von *-heit* und *-keit* zur Interpretation hinzugezogen werden (Tab. 7.8).

	1900	1925	1950	1975	1999
<i>-heit</i>	0,09	0,13	0,1	0,08	0,08
<i>-keit</i>	0,14	0,17	0,18	0,14	0,18

Tabelle 7.8: TTRs der Allomorphe *-heit* und *-keit* für die Zeiträume 1900, 1925, 1950, 1975 und 1999

Bei Betrachtung der Ergebnisse nach Separation der Morphe fällt auf, dass *-keit* im Gegensatz zu *-heit* höhere Werte hervorbringt, was als Indiz für eine höhere Produktivität gelten kann. *-keit* ist geringen Schwankungen ausgesetzt, die zwischen 0.14 und 0.18 liegen. Für *-heit* lässt sich festhalten, dass die Produktivität auf Basis der TTRs im 20. Jahrhundert einen Abwärtstrend erfährt, da für 1999 ein geringer Wert erreicht wird als für 1900, 1925 und 1950.

Mit den bisherigen Ergebnissen jedoch kann sich ein Linguist noch nicht zufrieden geben, sodass weitere statistische Analysen erfolgen müssen.

7.1.4 Hapax Legomena

Die Anzahl von Hapax Legomena gibt Hinweise darüber, inwieweit es möglich ist, Neubildungen mit einem Affix zu kreieren. Ein Wortbildungsmuster, das produktiv ist, wird viele niederfrequente Bildungen hervorbringen und damit möglicherweise auch eine hohe Anzahl an Hapaxen, während mit einem unproduktiven Wortbildungsmodell vorzugsweise stark lexikalisierte Bildungen abgerufen werden und Neubildungen rar oder gar unwahrscheinlich sind.

Anhand der untersuchten Korpusdaten können die in Tab. 7.9 aufgelisteten Häufigkeiten der Hapax Legomena ermittelt werden.

	1900	1925	1950	1975	1999
<i>-ung</i>	219	279	224	226	286
<i>-tum</i>	18	30	9	10	34
<i>-nis</i>	7	3	5	6	5
<i>-heit/-keit</i>	382	333	209	133	352
<i>-sal</i>	5	(1)	1	0	1

Tabelle 7.9: Häufigkeiten der Hapax Legomena der substantivbildenden Suffixe *-ung*, *-tum*, *-nis*, *-heit/-keit* und *-sal* für die Zeiträume 1900, 1925, 1950, 1975 und 1999

Erneut wird auf den ersten Blick deutlich, dass *-sal* die niedrigsten Frequenzen hervorbringt: Bei den fünf Hapaxen des Jahres 1900 handelt es sich um die Wörter *Drangsal*, *Labsal*, *Rinnsal*, *Triübsal* und *Wirrsal*, Archaismen des Deutschen; 1925 findet man lediglich den Namen der Stadt *Bruchsal* als Derivat³, 1950 das Wort *Labsal* und das Hapax Legomenon des Jahres 1999 lautet *Drangsal*.

Auch für *-nis* ist der Befund klar: es finden sich nur wenige Hapaxe und dies gilt für alle Zeiträume. Zwischen drei und sieben Einmalbildungen lassen sich im 20. Jahrhundert finden, was auf ein unproduktives Wortbildungsmuster hindeutet. Insbesondere im Vergleich zu *-tum* weist *-nis* somit einen entscheidenden Unterschied auf, denn bei Betrachtung der Typen in Tab. 7.5 schien es zunächst Parallelen zwischen *-nis* und *-tum* zu geben. Das Suffix *-tum*, deren Hapaxfrequenzen zwischen neun und 35 liegen, weist damit relativ große Schwankungen diesbezüglich auf, allerdings zeigt sich, dass sowohl im Jahre 1925 als auch im Jahre 1999 über 30 Wörter nur einmalig im Korpus auftauchen. Das Wort *Rowdytum*, das eine Bildung des Jahres 1999 darstellt, zeigt, dass *-tum* nicht völlig unproduktiv zu sein scheint, da es sich bei der Basis des Derivates um einen Anglizismus handelt.

Das Suffix *-ung* ist im Hinblick auf die Frequenzen seiner einmaligen Bildungen durchgängig konstant, auch die Korpusgröße scheint keinen Einfluss auf die Anzahl Hapaxe zu haben, da zu jedem Zeitpunkt zwischen 219 und 286 Einmalbildungen vorhanden sind.

³Zur Geschichte des Namens der Stadt siehe <http://www.belle-alliance.com/bruchsal/stadtgeschichte/stadtgeschichte.html> (Abruf am 28.06.2011 um 13:35 Uhr). Er unterlag im Laufe von ca. 900 Jahren verschiedenen Veränderungen: *Bruhosella* über *Bruohsele*, *Bruohsela*, *Brochsale*, *Broxole*, *Brucsel*, *Brusela*, *Bruoselle*, *Bruhsel*, *Bruchsel* (16. Jahrhundert), *Prussel* zu *Bruchsal*.

Anders ist es bei *-heit* und *-keit*: sie zeichnen sich insbesondere zwischen 1900 und 1925 durch sehr hohe Frequenzen aus, die im Vergleich zu 1975 sogar dreimal höher sind. Allerdings liegt für den Zeitraum 1900 auch eine etwa dreimal so hohe Tokenanzahl vor (Tab. 7.1), die analysiert werden konnte.

	1900	1925	1950	1975	1999
<i>-heit</i>	115	119	74	40	83
<i>-keit</i>	267	214	135	93	269

Tabelle 7.10: Häufigkeiten der Hapax Legomena der der Allomorphe *-heit* und *-keit* für die Zeiträume 1900, 1925, 1950, 1975 und 1999

Betrachtet man nun die beiden Allomorphe im Einzelnen (Tab. 7.10), so fällt erneut auf, dass *-keit* das produktivere der beiden Allomorphe ist: in Relation zu *-heit* tritt es in etwa doppelt so vielen Hapaxen auf, im Jahre 1999 sogar in einer dreimal höheren Anzahl.

Auch für den weiteren Verlauf dieser Arbeit stellt sich die Frage, inwieweit es möglich ist, dass sich Allomorphe hinsichtlich ihrer Produktivität in unterschiedliche Richtungen entwickeln können. Die bisherigen Analysen deuten zum jetzigen Zeitpunkt darauf hin, dass dies der Fall ist.

7.1.5 Baayens *P*

Vorzüge und Nachteile des von Baayen (1992) entwickelten Maßes *P* zur Messung von Produktivität wurden in Kap. 4 diskutiert; darauf sei hier verwiesen.

Tabelle 7.11 listet die Werte von *P* auf, die auf Basis der Frequenzen von Token und Hapax Legomena berechnet worden sind.

Wie schon für die Analyse der Typen-Token-Relationen wurde *-sal* nicht weiter betrachtet, da seine Unproduktivität schon anhand der Token- und Typenfrequenzen festgestellt wurde.

Für das besonders produktive Suffix *-ung* gibt es deutliche Schwankungen hinsichtlich *P*: 1900, 1950 und 1975 liegt der Wert jeweils bei etwa 0.045, 1925 und 1999 mit 0.057 und 0.058 deutlich höher.

Auch die Werte zu *-nis* unterliegen gewissen Schwankungen; allerdings sind die Werte mit 0.002 bis 0.008 durchweg sehr niedrig.

	1900	1925	1950	1975	1999
<i>-ung</i>	0,044	0,057	0,045	0,046	0,058
<i>-tum</i>	0,064	0,120	0,103	0,083	0,102
<i>-nis</i>	0,004	0,002	0,008	0,008	0,003
<i>-heit/-keit</i>	0,053	0,071	0,064	0,052	0,061
<i>-sal</i>	/	/	/	/	/

Tabelle 7.11: *P* der substantivbildenden Suffixe *-ung*, *-tum*, *-nis*, *-heit/-keit* und *-sal* für die Zeiträume 1900, 1925, 1950, 1975 und 1999

Schwierig ist nun die Beurteilung der Produktivität von *-tum*, da *-tum* starke Schwankungen hinsichtlich der *P*-Werte aufweist. Während der Wert um 1900 noch bei etwa 0.6 liegt (und damit höher ist als bei *-ung*), liegt er nach Ausschluss von 1975 zu allen Zeitpunkten über dem Wert 0.1, einem Wert, den sonst kein anderes Wortbildungsmodell zur Substantivbildung erreicht. Dies zeigt, dass es im Vergleich zur gesamten Tokenanzahl viele einmalige Bildungen gibt, sodass ein mögliches Potential zur Wortbildung vorhanden sein könnte.

Die Allomorphe *-heit* und *-keit* hingegen zeigen Schwankungen, die selbst nach Ausschluss des problematischen Jahres 1975 noch vorhanden sind: 1900 wird ein Wert von 0.053 erreicht, welcher in Relation zu den anderen (vergleichbaren) Jahren niedrig ist. Der höchste Wert kann für 1925 berechnet werden.

Nach Allomorphen aufgetrennt ergeben sich für *P* die in Tabelle 7.12 präsentierten Werte.

	1900	1925	1950	1975	1999
<i>-heit</i>	0,037	0,061	0,046	0,033	0,034
<i>-keit</i>	0,065	0,078	0,081	0,068	0,082

Tabelle 7.12: *P* der der Allomorphe *-heit* und *-keit* für die Zeiträume 1900, 1925, 1950, 1975 und 1999

Nicht überraschend kann Tab. 7.12 abgelesen werden, dass erneut Unterschiede zwischen *-heit* und *-keit* vorliegen und *-keit* das produktivere der beiden Allomorphe ist.

7.1.6 P^*

Die globale Produktivität als zweidimensionales Maß versteht sich als Relation von Hapaxen und Token (P) und der Anzahl der Typen eines Wortbildungsmodells.

Die globale Produktivität soll nun aus zwei Perspektiven betrachtet werden: synchron und diachron. Zunächst einmal sollen die einzelnen Suffixe diachron betrachtet werden. Das Jahr 1975 wird nicht zur Betrachtung hinzugezogen und ausgeklammert, da die Tokenanzahl des Teil-Korpus für diesen Zeitraum um die Hälfte geringer ist als für die anderen Messzeiträume.

Das Suffix *-sal* wird, wie bereits in einigen vorangegangenen Analysen, auch an dieser Stelle nicht weiter betrachtet, da seine Unproduktivität bereits festgestellt werden konnte.

Um Vergleichbarkeit der Ergebnisse zu gewährleisten, wird für alle Suffixe dieselbe Skalierung der x- und y-Achse genommen. Diese Maßnahme ist deshalb von Notwendigkeit; die Wertebereiche der Achsen ergeben sich aus den Werten für P und der Anzahl der Typen; der höchste Wert P stellt das Maximum der x-Achse, die höchste Typenanzahl das Maximum der y-Achse dar.

Es sei an dieser Stelle auf die in Kap. 4.3.4 geführte Diskussion verwiesen.

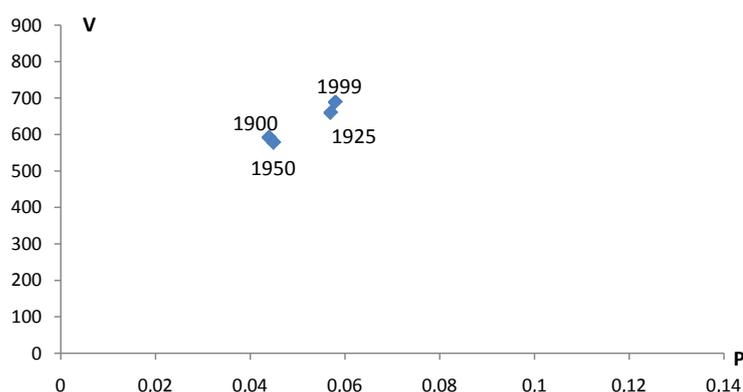


Abbildung 7.1: Globale Produktivität von *-ung* im 20. Jahrhundert

P** für das Suffix *-ung

Abb. 7.1 zeigt die globale Produktivität von *-ung*. Deutlich wird, dass es sich bei *-ung* um ein konstant produktives Suffix handelt, das kleinen Schwankungen ausgesetzt ist, die man vermutlich als Zufallsschwankung deuten darf.⁴

Bei Ausklammerung des Jahres 1950 könnte ein Aufwärtstrend erfasst werden, der skizziert, dass die Produktivität von *-ung* im 20. Jahrhundert leicht zunimmt.

P** für das Suffix *-nis

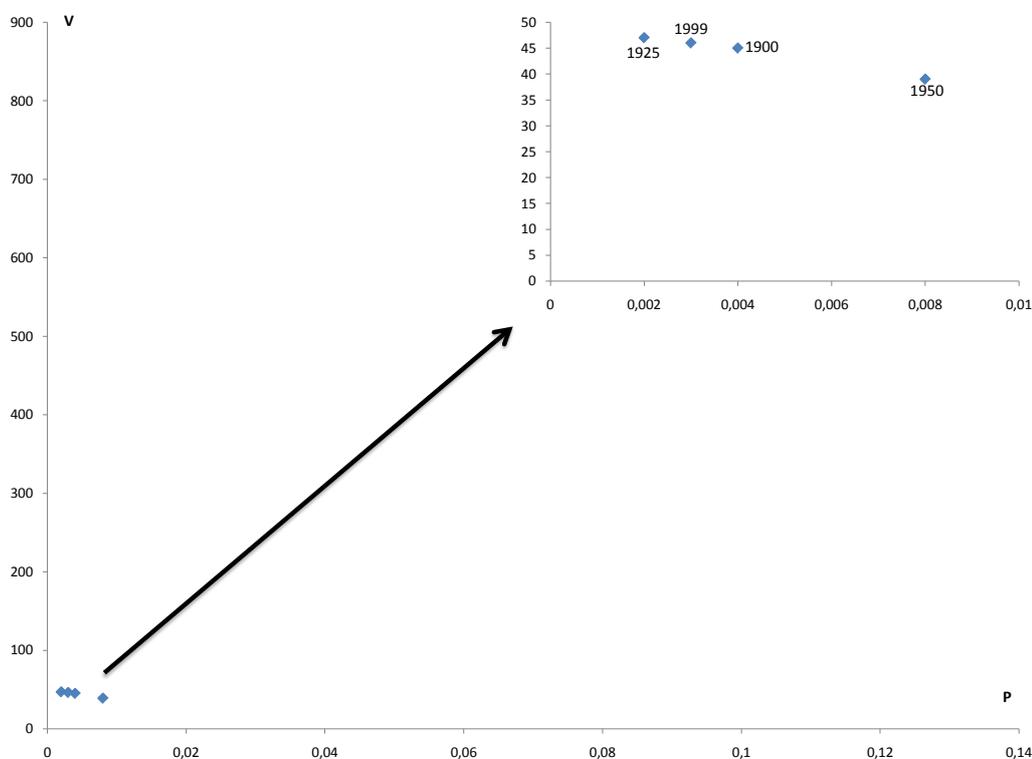


Abbildung 7.2: Globale Produktivität von *-nis* im 20. Jahrhundert

Für das Suffix *-nis* ist es einfach, eine Aussage zur globalen Produktivität zu machen (Abb. 7.2), denn das Suffix ist unproduktiv und kann in Bezug auf die globale Produktivität für alle Zeiträume im 'linken unteren' Viertel der Grafik verortet werden.

⁴Dazu siehe auch Abb. 8.14.

***P** für das Suffix -tum**

Für -tum zeigt sich nach Abb. 7.3 ein anderes Bild.

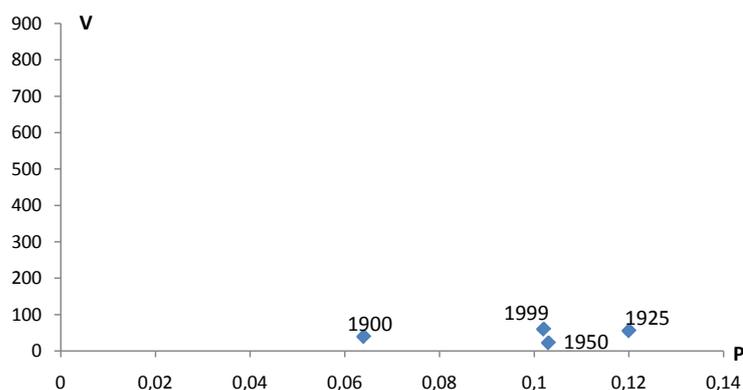


Abbildung 7.3: Globale Produktivität von -tum im 20. Jahrhundert

Die globale Produktivität schwankt stark, was eine klare Aussage einer steigenden oder fallenden Produktivität verbietet; auch eine konstante Produktivität kann an dieser Stelle nicht deklariert werden: Es liegen hohe *P*-Werte vor, die sich durch eine große Anzahl an Hapaxen ergeben, die Typenanzahl jedoch ist gering. Die daraus resultierende Verortung im rechten unteren Quadranten macht eine Interpretation, wie bereits in Kap. 4.3.4 erörtert, deshalb schwierig.

Die zentrale Frage lautet in diesem Zusammenhang wie folgt: Darf ein Suffix, dessen Typenfrequenz niedrig ist, als produktiv betrachtet werden, wenn sich im Gegensatz dazu ein hoher Wert von *P* beobachten lässt?

Auf diese Frage ist tendenziell mit ja zu antworten, da in der vorliegenden Arbeit angenommen wird, dass sich Produktivität eines Wortbildungsmusters in der Eigenschaft zeigt, neue Bildungen zu schaffen, die nicht lexikalisiert sein müssen. Demnach sollte als Zwischenergebnis festgehalten werden, dass -tum scheinbar durchaus das Potential besitzt, produktiv zur deutschen Wortbildung beizutragen.

Um eine klare Aussage treffen zu können, wäre es allerdings wünschenswert, eine größere Datenmenge zu betrachten, um zu schauen, wie sich die Produktivität ent-

wickeln würde, wenn weitere Daten zur Verfügung stünden. Diese Datenmenge steht für den Untersuchungszeitraum und den Daten des DWDS nicht zur Verfügung. Wie jedoch eine andere Lösung zur Klärung dieses Problems aussehen kann, ist Gegenstand des achten Kapitels.

P** für die Allomorphe *-heit* und *-keit

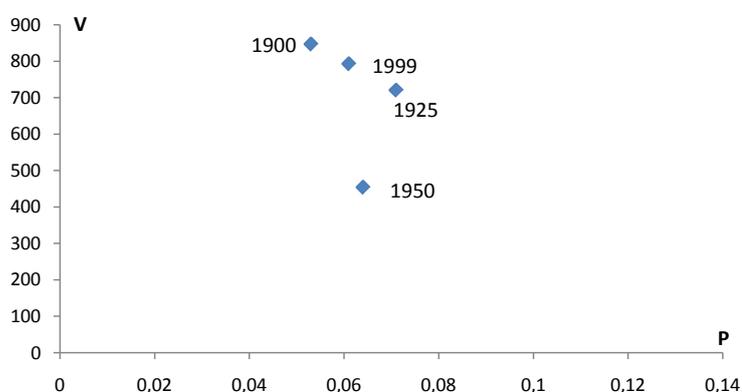


Abbildung 7.4: Globale Produktivität von *-heit* und *-keit* im 20. Jahrhundert

Zunächst einmal stehen die Allomorphe – in Bezug auf die globale Produktivität P^* – gemeinsam zur Betrachtung (Abb. 7.4). Der Grafik kann entnommen werden, dass es sich im Jahre 1950 um ein 'Ausreißer'-Wertepaar zu handeln scheint, welches nicht in das sonst erkennbare Muster der globalen Produktivität hineinpasst. Diese Entdeckung konnte auch schon für das Suffix *-nis* gemacht werden.

Schaut man sich die Wertepaare von 1900, 1925 und 1999 an, so wird deutlich, dass geringe Schwankungen vorhanden sind.

Die Produktivität von Allomorphen ist bisher nicht in den linguistischen Fokus geraten, soll aber in dieser Arbeit an dem Beispiel von *-heit* und *-keit* diskutiert werden.

Betrachten wir zunächst *-heit* (Abb. 7.5). Wie schon in vielen anderen Fällen beobachtet, lässt sich auch hier keine Tendenz verzeichnen, das Wertepaar von 1999 jedoch

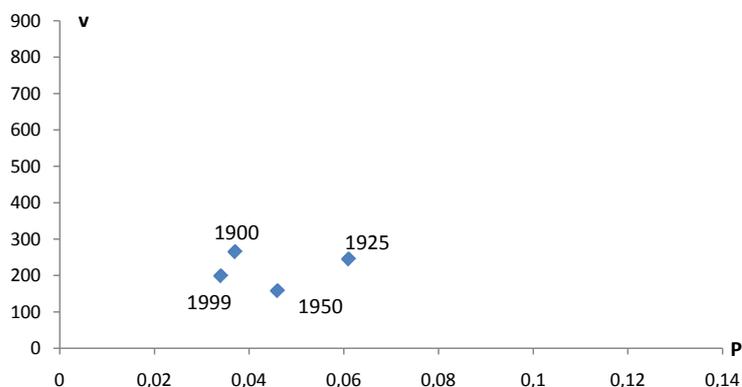


Abbildung 7.5: Globale Produktivität von *-heit* im 20. Jahrhundert

weist eine geringere Produktivität auf als die beiden Jahre 1900 und 1925. Auch für *-heit* zeigt sich, dass 1950 einen Sonderfall darstellt.

Für das Allomorph *-keit* kann Abb. 7.6 Folgendes entnommen werden: Bekannterweise sind auch in dieser Abbildungen Schwankungen zu beobachten. Während es für die Jahre 1900, 1925 und 1999 eine Anstiegstendenz hinsichtlich P gibt, so trifft dies in Bezug auf die Typenhäufigkeiten nicht zu, was die Interpretation erneut problematisch macht. Nur der Vergleich von 1900 zu 1999 zeigt einen Anstieg der globalen Produktivität P^* .

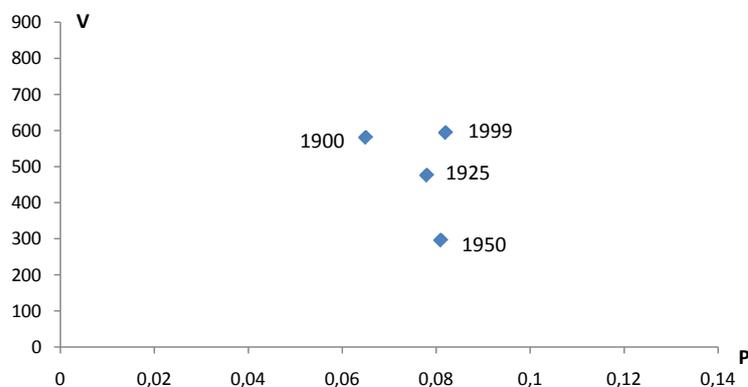


Abbildung 7.6: Globale Produktivität von *-keit* im 20. Jahrhundert

Zusammenfassung

Zusammenfassend lässt sich festhalten, dass die substantivbildenden Suffixe auch im Bezug auf ihre globale Produktivität Schwankungen aufweisen und keine eindeutigen Produktivitätsanstiege oder -abfälle ermittelt werden konnten.

Das Einzige, was sich festhalten lässt, ist die Tatsache, dass sich *-nis* und *-ung* jeweils für alle Messzeitpunkte jeweils etwa im gleichen Bereich des Diagramms verorten lassen.

Für *-tum* gilt, dass es sich durch eine globale Produktivität auszeichnet, die nah an der x-Achse des rechten unteren Quadranten angesiedelt ist, die sich aus einer geringen Typenfrequenz, aber hohen Produktivität *P* ergibt. Dieses Ergebnis steht im Widerspruch zu den in Tab. 5.1 zusammengefassten Aussagen der linguistischen Literatur, denn die Linguisten etikettieren *-tum* nur als schwach produktiv.

Ein Problem der hohen Produktivitätswerte *P* könnte darin liegen, dass die Tokenmengen für *-tum* gering waren, was vermuten lässt, dass andere *P*-Werte zu erwarten wären, sofern größere Datenmengen zur Verfügung stünden.

Für *-heit* und *-keit* lassen sich den jeweiligen Diagrammen mäßige Schwankungen entnehmen, deren Interpretation schwer fällt und vermutlich auch mit den ungleichen Tokenmengen zusammenhängen kann.

Nachfolgend sollen auch synchrone Analysen durchgeführt werden. Dabei können nun alle fünf Zeiträume zur Interpretation hinzugezogen werden, da in diesem Fall jeweils die gleiche Tokenanzahl untersucht werden konnte.

***P** für den Zeitraum 1900**

Die erste synchrone Betrachtung wird für das Jahr 1900 durchgeführt (Abb. 7.7). Zunächst einmal fallen die Allomorphe auf, die sich mit ihrer sowohl hohen Produk-

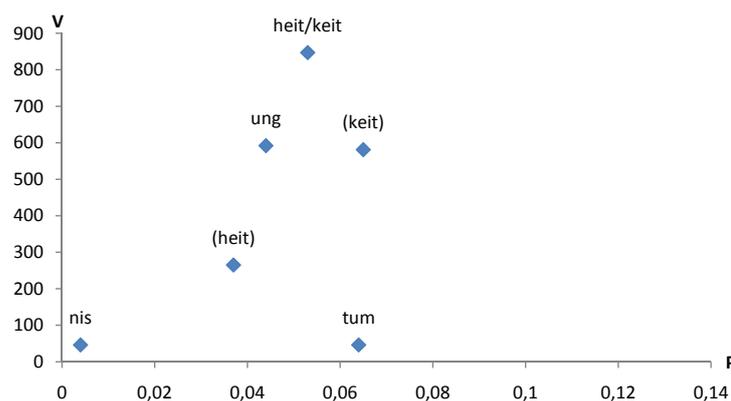


Abbildung 7.7: Globale Produktivität der substantivbildenden Suffixe im Jahre 1900

tivität *P* als auch der hohen Anzahl an Typen im oberen Bereich des Diagramms befinden. Nicht verkannt werden darf, dass die zur Betrachtung herangezogenen Token in diesem Fall besonders hoch waren.

Im Gegenzug dazu sollte in dieser Arbeit aber auch insbesondere die Allomorphie betrachtet werden. Dazu wurde entschieden, mit wenig Informationsverlust zu arbeiten und eine möglichst große Datenmenge zur Analyse zu verwenden.

Nach Trennung der Allomorphe wird ersichtlich, dass sich *-keit* deutlich von *-heit* abhebt und das produktivere der beiden Allomorphe ist, da sowohl *P* als auch die Typen jeweils höhere Werte aufweisen.

Auch wenn *-ung* in Bezug auf die Typenfrequenz ähnlich produktiv erscheint wie *-keit*, so existiert ein deutlicher Unterschied hinsichtlich *P*.

Besonders interessant ist der Vergleich von *-nis* und *-tum*, zwei Suffixen, die bezüglich ihrer realisierten Produktivität vergleichbar waren. Abb. 7.7 macht den Unterschied der Produktivität nun grafisch sichtbar: *-tum* scheint das produktivere Wortbildungsmodell zu sein.

***P** für den Zeitraum 1925**

Auch für den Zeitraum 1925 wurde die globale Produktivität grafisch dargestellt (Abb. 7.8).

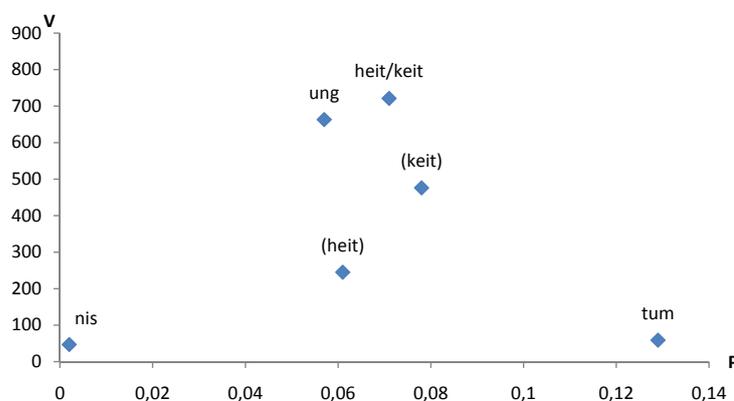


Abbildung 7.8: Globale Produktivität der substantivbildenden Suffixe im Jahre 1925

Die Struktur der globalen Produktivität, die für das Jahr 1900 beschrieben worden ist, findet sich auch für 1925 wieder: die Allomorphe *-heit* und *-keit* zeigen zusammen betrachtet, dass es sich um ein produktives Wortbildungsmodell handelt, bei welchem erneut festzustellen ist, dass *-keit* das produktivere der beiden Morphe ist.

Das Suffix *-ung*, das in vielen unterschiedlichen Wortbildungen realisiert wird, weist eine hohe Produktivität auf, welche jedoch kleiner ist als die der Allomorphe. Für *-nis* und *-tum* darf dasselbe festgehalten werden wie für das Jahr 1900: in realisierten Wortbildungen unterscheiden sie sich frequenziell nur sehr schwach, beim Heranziehen von *P* jedoch zeichnet sich ein deutlicher Unterschied ab, sodass *-tum* als produktiv gilt.

***P** für den Zeitraum 1950**

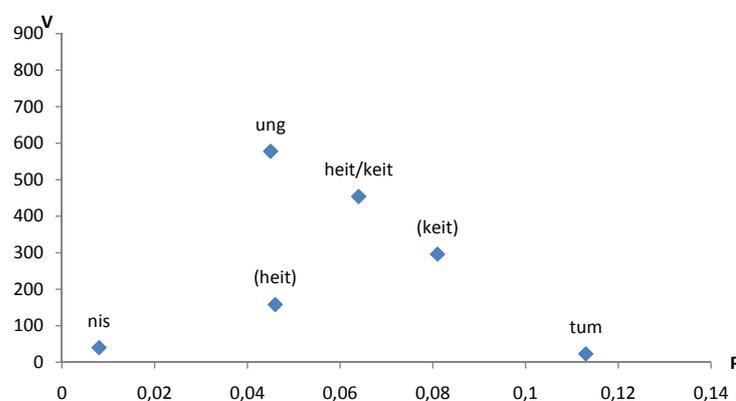


Abbildung 7.9: Globale Produktivität der substantivbildenden Suffixe im Jahre 1950

Auch für das Jahr 1950 (Abb. 7.9) zeigt sich die bekannte Struktur von *-nis* und *-tum*, wobei jedoch angemerkt werden muss, dass die Typenhäufigkeit von *-tum* in diesem Jahr geringer ist als die von *-nis*.

Ebenso ist die bekannte Struktur von *-heit* und *-keit* – beide als separate Allomorphe betrachtet – sichtbar: *-keit* ist produktiver als *-heit*.

Im Vergleich zu *-ung* weisen die Allomorphe eine höhere Produktivität *P* auf, *-ung* jedoch wird frequenziell stärker in unterschiedlichen Wortbildungen realisiert, indem man im Korpus im Jahre 1950 578 Bildungen findet.

P* für den Zeitraum 1975

Dass die Daten des Jahres 1975 Probleme bereiten, wurde bereits an mehreren Stellen erwähnt, zur synchronen Analyse jedoch sind sie geeignet. Deshalb soll Abb. 7.10 betrachtet und diskutiert werden.

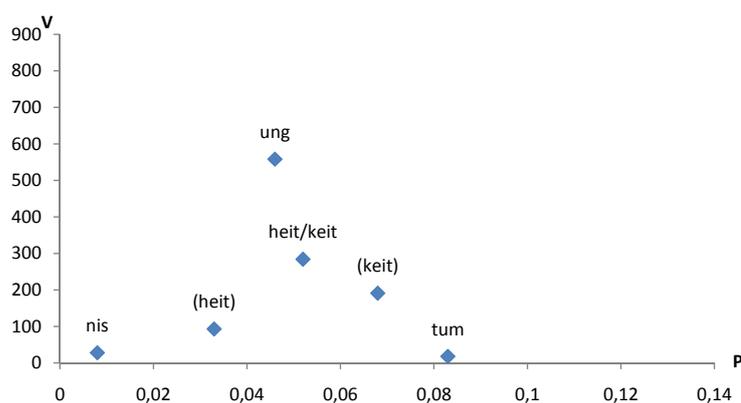


Abbildung 7.10: Globale Produktivität der substantivbildenden Suffixe im Jahre 1975

Der Vergleich von *-tum* und *-nis* zeigt das schon bekannte Verhalten; die Typenhäufigkeit ist beiderseits gering und dennoch wird mit *-tum* eine größere Anzahl an Hapax Legomena gebildet.

Auch die Allomorphe *-heit* und *-keit* zeigen den schon bekannte Produktivitätsunterschied, der auch bei relativ geringerer Tokenanzahl nicht unerkannt bleibt.

Der Unterschied von *-ung* zu den Allomorphen *-heit* und *-keit* jedoch tritt nun stärker zum Vorschein, indem *-ung* in 558 Typen realisiert wird, die Allomorphe insgesamt aber nur 284 Typen hervorbringen. Dieses Ergebnis jedoch resultiert aus der betrachteten Tokenanzahl, die bei *-ung* auf knapp unter 5000 Token limitiert wurde, für die Allomorphe allerdings keine Limitierung erfolgen konnte, da insgesamt nur knapp 3000 Token zur Verfügung standen.

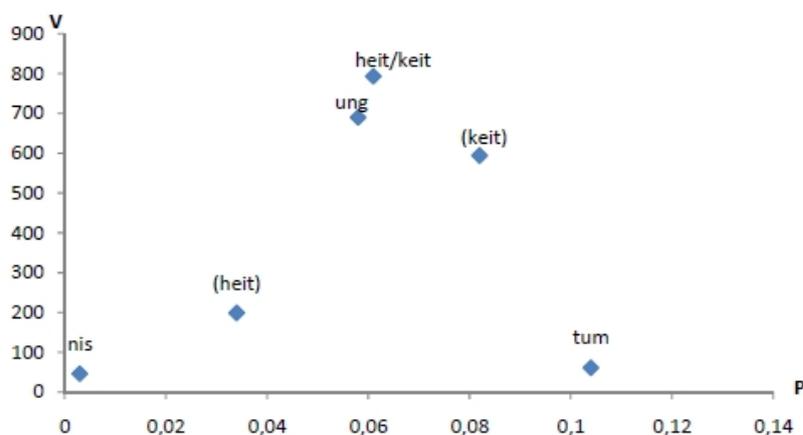


Abbildung 7.11: Globale Produktivität der substantivbildenden Suffixe im Jahre 1999

P^* für den Zeitraum 1999

Als letztes soll nun P^* für den Zeitraum 1999 visualisiert werden (Abb. 7.11).

Für das Jahr 1999 zeigt sich dasselbe Bild, das für die Jahre 1900, 1925 und 1950 beschrieben werden konnte.

Das Gefälle zwischen den Allomorphen erscheint für diesen Untersuchungszeitraum auffällig groß, die Diskrepanz zwischen den beiden Suffixen *-nis* und *-tum* bleibt unverändert. *-ung* und *-heit/-keit* scheinen sich sehr ähnlich zu verhalten; die etwas größere globale Produktivität der Allomorphe kann aus der leicht höheren Tokenanzahl resultieren.

Das Jahr 1999 soll zu einem späteren Zeitpunkt dieser Arbeit erneut betrachtet werden, wenn es um die Validität der linguistischen Aussagen geht.

Zusammenfassung der Ergebnisse von P^*

Die einzelnen synchron durchgeführten Analysen lassen sich zu einem Schema abstrahieren (Abb. 7.12).

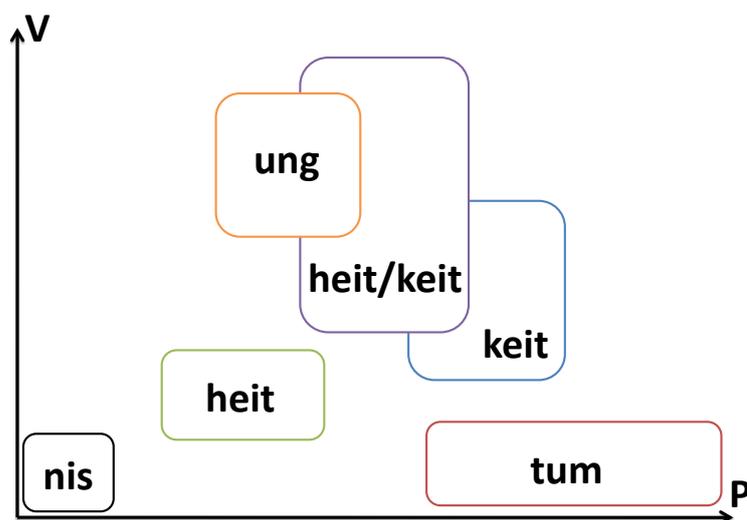


Abbildung 7.12: Schematische Zusammenfassung der globalen Produktivität der substantivbildenden Suffixe textit-nis, -ung, -sal, -heit/-keit/-igkeit und -tum im Zeitraum von 1900 bis 1999

Es lassen sich an dieser Stelle drei Zwischenergebnisse festhalten:

1. Im Deutschen gibt es substantivbildende Suffixe, die sowohl durch eine hohe Typenanzahl als auch durch einen hohen Wert P als produktiv gelten, was sich insbesondere durch Betrachtung von P^* bestätigen lässt.
2. Es ist möglich, dass zwei Allomorphe unterschiedlich stark produktiv zur Wortbildung beitragen.
3. Zwei Suffixe, die eine geringe Typenfrequenz im Korpus aufweisen, können dennoch unterschiedliche Verhaltensweisen aufzeigen: eine niedrige Typenfrequenz muss nicht, wie anhand von -tum für mehrere Zeitpunkte beobachtet werden konnte, mit einer geringen Produktivität P und damit auch P^* einhergehen.

7.2 Die adjektivbildenden Suffixe

Im folgenden Abschnitt werden nun die Ergebnisse der Berechnungen zur Produktivität der acht untersuchten adjektivbildenden Suffixe betrachtet.

7.2.1 Tokenhäufigkeit

Tab. 7.13 präsentiert die aus dem DWDS extrahierten Tokenlisten für die acht Suffixe.

	1900	1925	1950	1975	1999
<i>-ig</i>	3339 (11422)	3365 (10374)	3474 (7353)	3587 (3988)	3489 (13066)
<i>-isch</i>	3876 (8436)	3668 (6145)	3596 (7109)	3698 (5503)	3072 (13034)
<i>-lich</i>	4993 (14362)	4987 (10618)	4988 (9468)	4997 (5036)	4990 (16179)
<i>-mäßig</i>	447	292	154	92	319
<i>-los</i>	827	803	815	205	1170
<i>-haft</i>	559	552	295	156	612
<i>-sam</i>	1074	818	890	551	1030
<i>-bar</i>	1828	1059	710	426	1840

Tabelle 7.13: Tokenhäufigkeit der adjektivbildenden Suffixe *-ig*, *-isch*, *-lich*, *-mäßig*, *-los*, *-haft*, *-sam* und *-bar* für die Zeiträume 1900, 1925, 1950, 1975 und 1999

Für *-haft*, *-mäßig* und *-los* gelangt man über die Schnittstelle des DWDS an unterschiedlich große Datenmengen. Eine Limitierung auf die niedrigste Tokenfrequenz, die im Jahre 1975 zu verzeichnen ist, erscheint problematisch, sodass, wie auch schon für die substantivbildenden Suffixen erklärt wurde, mit den tatsächlich verfügbaren Zahlen gerechnet wird.

Wenn das Problemjahr 1975 ausgeblendet wird, findet man sowohl für *-bar* als auch für *-sam* ähnlich hohe Tokenfrequenzen, die in einem Rahmen zwischen 700 und 1800 Token liegen.

Für die drei Suffixe *-isch*, *-lich* und *-ig* stehen für alle Messzeiträume relativ große Datenmengen zur Verfügung, sodass an diesen Stellen zur Weiterberechnung normiert werden kann: bei *-isch* und *-lich* auf 5000 und bei *-ig* immerhin auf 4000 Token.

	1900	1925	1950	1975	1999
<i>-ig</i>	8567	8714	6177	6780	8101
<i>-isch</i>	6327	5162	5972	9355	8081
<i>-lich</i>	10772	8919	7953	8561	10031
<i>-mäßig</i>	335	245	129	156	198
<i>-los</i>	620	675	685	349	725
<i>-haft</i>	419	464	248	421	379
<i>-sam</i>	806	687	748	937	639
<i>-bar</i>	1371	890	596	724	1141

Tabelle 7.14: Tokenhäufigkeit (normiert auf 1 Million Token) der adjektivbildenden Suffixe *-ig*, *-isch*, *-lich*, *-mäßig*, *-los*, *-haft*, *-sam* und *-bar* für die Zeiträume 1900, 1925, 1950, 1975 und 1999

Analog zu den substantivbildenden Suffixen erfolgt auch bei den adjektivbildenden Suffixen im nächsten Schritt eine weitere Normierung der Werte. In der Annahme, es stünden 1 Million Token zur Verfügung, würde man die in Tab. 7.14 aufgelisteten Werte erhalten.

Aus diesen Frequenzen ergibt sich das Bild, dass nicht unerhebliche Schwankungen zu beobachten sind: bei *-ig* bis zu 2500 Token, bei *-isch* bis zu etwa 4000 Token, bei *-lich* bis zu 3000 Token, bei *-mäßig* bis zu etwa 200 Token und für *-los* sind Schwankungen von über 50% vorhanden (bis zu knapp 400 Token).

Auch für *-haft* kann Ähnliches beobachtet werden, obwohl die Tokenfrequenzen etwas niedriger sind als bei *-los*; für das Suffix *-sam* sind Schwankungen bis zu 300 Token sichtbar und für *-bar* auch wieder bis zu 50% (knapp 600 Token).

Doch zum jetzigen Zeitpunkt können keine Schlüsse in Bezug auf die morphologische Produktivität der adjektivbildenden Suffixe gezogen werden, da unklar ist, wie sich die Tokenfrequenzen zusammensetzen. Es ist möglich, dass es viele stark lexikalisierte Wörter gibt, die in einer hohen Anzahl vorkommen, was, sofern insgesamt wenige unterschiedliche Bildungen und keine oder nur sehr wenige Neubildungen gefunden werden, nicht für eine hohe Produktivität sprechen kann.

7.2.2 Typenhäufigkeit

Auf Basis der verfügbaren Token wurden auch die Frequenzen der Typen ermittelt (Tab. 7.15).

	1900	1925	1950	1975	1999
<i>-ig</i>	267	320	306	237	354
<i>-isch</i>	368	441	317	314	377
<i>-lich</i>	354	433	384	338	390
<i>-mäßig</i>	55	57	40	22	71
<i>-los</i>	190	191	154	91	229
<i>-haft</i>	94	125	77	40	167
<i>-sam</i>	54	55	43	21	46
<i>-bar</i>	274	178	141	123	324

Tabelle 7.15: Typenhäufigkeiten für die adjektivbildenden Suffixe *-ig*, *-isch*, *-lich*, *-mäßig*, *-los*, *-haft*, *-sam* und *-bar* für die Zeiträume 1900, 1925, 1950, 1975 und 1999

Bei Betrachtung von *-ig*, *-isch* und *-lich* wird deutlich, dass sich diese drei Suffixe ähnlich verhalten und jeweils zwischen 237 und 390 Typen hervorbringen.

Dabei ist *-ig* dasjenige Suffix, das im Vergleich zu seinen beiden Konkurrenten die kleinsten Werte aufweist, was mit hoher Wahrscheinlichkeit mit der niedrigeren Gesamttokenzahl von 4000 zusammenhängt.

Die beiden Suffixe *-bar* und *-sam*, für welche eine ähnlich hohe Tokenanzahl zur Verfügung stand, unterscheiden sich stark im Hinblick auf ihre realisierte Produktivität: während mit *-sam* nur zwischen 21 und 55 unterschiedliche Wörter gefunden wurden, schwankt die Typenfrequenz von *-bar* enorm und liegt in einem Rahmen zwischen 123 und 324 Typen. Die Werte von 1900 und 1999 können mit den jeweiligen Werten der drei erstgenannten Suffixe verglichen werden, für welche allerdings dreibis viermal so große Tokenmengen betrachtet wurden. Dies erscheint auffällig und soll im späteren Verlauf der Arbeit weiter diskutiert werden.

Schwierig gestaltet sich die Interpretation für *-mäßig*, das nur wenige Typen hervorbringt: es finden sich jeweils zwischen 40 und 71 unterschiedliche Bildungen. Weitaus produktiver, zumindest was die Typenfrequenzen anbelangt, sind *-haft* und *-los*; allerdings ist *-los* nahezu doppelt so produktiv wie *-haft*.

7.2.3 Typen-Token-Relationen

Wie schon für die substantivbildenden Suffixe wurden auch die Typen-Token-Relationen der adjektivbildenden Affixe berechnet (Tab. 7.16).

	1900	1925	1950	1975	1999
<i>-ig</i>	0,08	0,1	0,09	0,07	0,1
<i>-isch</i>	0,1	0,12	0,09	0,09	0,12
<i>-lich</i>	0,07	0,09	0,08	0,07	0,08
<i>-mäßig</i>	0,123	0,2	0,26	0,24	0,23
<i>-los</i>	0,24	0,24	0,21	0,2	0,2
<i>-haft</i>	0,17	0,23	0,26	0,26	0,28
<i>-sam</i>	0,05	0,07	0,05	0,04	0,05
<i>-bar</i>	0,15	0,17	0,2	0,3	0,18

Tabelle 7.16: Typen-Token-Relationen der adjektivbildenden Suffixe *-ig*, *-isch*, *-lich*, *-mäßig*, *-los*, *-haft*, *-sam* und *-bar* für die Zeiträume 1900, 1925, 1950, 1975 und 1999

Für die drei Suffixe mit 'großen' verfügbaren Datenreihen ergeben sich relativ konstante Werte, die in einem Bereich zwischen 0,07 (*-lich* im Jahre 1975) und 0,12 (*-isch* im Jahre 1925) liegen. Demnach scheinen diese drei Suffixe ein ähnliches Verhalten aufzuweisen, was die Relation der jeweiligen Typen zu ihren Token betrifft. Erst die nachfolgenden Berechnungsmethoden werden zeigen, ob das ähnliche Wortbildungsverhalten erhalten bleibt.

Für *-sam* wird deutlich, dass die errechneten Werte ebenfalls relativ konstant sind und zudem als niedrig eingestuft werden dürfen.

Die Betrachtung der anderen vier Suffixe gestaltet sich wesentlich schwieriger. Sowohl *-bar*, *-mäßig*, *-los* als auch *-haft* zeichnen sich an vielen Stellen durch eine hohe TTR aus. Dabei scheint es bei *-haft* eine Anstiegstendenz der morphologischen Produktivität zu geben.

Für *-los* dagegen fallen die Werte konstant von 0,24 auf letztlich 0,2.

Das Suffix *-mäßig* zeichnet sich durch schwankende TTR-Werte aus und scheint im Jahre 1950 produktiver zu sein als in den anderen Jahren. Zudem ist die TTR im Jahr 1900 relativ gering.

Auch die Typen-Token-Relationen von *-bar* sind wenig konstant: selbst nach Ausklammerung des problematischen Jahres 1975 bleiben Schwankungen übrig, die zwischen 0.15 und 0.2 liegen.

7.2.4 Hapax Legomena

Als nächstes Kriterium sollen nun die Hapaxfrequenzen der adjektivbildenden Suffixe in den linguistischen Fokus geraten (Tab. 7.17).

	1900	1925	1950	1975	1999
<i>-ig</i>	96	122	108	93	127
<i>-isch</i>	143	173	114	115	140
<i>-lich</i>	114	130	109	94	117
<i>-mäßig</i>	23	30	21	11	41
<i>-los</i>	88	87	54	50	75
<i>-haft</i>	39	55	33	19	97
<i>-sam</i>	16	14	13	6	15
<i>-bar</i>	123	76	57	62	164

Tabelle 7.17: Häufigkeiten der Hapax Legomena der adjektivbildenden Suffixe *-ig*, *-isch*, *-lich*, *-mäßig*, *-los*, *-haft*, *-sam* und *-bar* für die Zeiträume 1900, 1925, 1950, 1975 und 1999

Erneut fallen bestimmte adjektivbildende Suffixe durch konstante Werte auf: zu diesen zählen in erster Linie *-lich* und *-ig*, da die Schwankungen hinsichtlich ihrer Frequenzen der Hapax Legomena gering sind, es sich aber dennoch gleichzeitig um recht hohe Werte handelt. Es sind jeweils zwischen 93 und 130 Bildungen vorhanden.

Für *-isch* hingegen werden im Jahr 1925 sogar 173 Hapaxe gefunden, was besonders auffällig wirkt.

Das Suffix *-sam* ist dasjenige Suffix, mit welchem zu allen Zeitpunkten nur sehr wenige Hapaxe produziert werden: die jeweilige Anzahl liegt zwischen 13 und 16, 1975 sogar nur bei 6.

Für die anderen vier zu betrachtenden adjektivbildenden Suffixe steht man als Linguist erneut vor dem Problem, eindeutige Interpretationen durchzuführen, da in allen Fällen Besonderheiten, bzw. Auffälligkeiten, vorhanden sind, die erklärt werden müssen. Begonnen wird dabei mit *-mäßig*. Das Suffix *-mäßig* zeichnete sich durch eine

niedrige Typenfrequenz aus. Die genaue Betrachtung der Hapaxe führt zu einem ähnlichen Bild. Die 11 Hapaxe als niedrigste Zahl der fünf Messzeiträume sind der geringen Tokenmenge des Jahres 1975 zuzuschreiben; die Jahre 1900, 1925 und 1950 zeigen Schwankungen, die in einem Wertebereich zwischen 21 und 30 schwanken. Herausstechend ist die Anzahl der 41 Hapaxe des Jahres 1999.

Für *-los* hingegen finden sich zwei Messzeiträume, 1950 und 1975, in welchen die Hapaxfrequenzen mit etwa 50 jeweils geringer sind als zu den anderen Zeitpunkten.

Weitaus auffälliger erscheinen die Schwankungen für *-bar* und *-haft*. Mit dem Suffix *-bar* konnte nicht im problematischen Jahr 1975 die geringste Frequenz gefunden werden, sondern im Jahr 1950. Und diese ist sogar dreimal so klein wie jene Frequenz des Jahres 1999. Im Jahr 1999 werden 164 Hapaxe gefunden, eine Zahl, die trotz geringer Tokenfrequenz mit den Häufigkeiten der Hapax Legomena von *-ig*, *-lich* und *-isch* konkurrieren kann. Ebenso ist die Zahl der Hapaxe für 1900 mit 123 bemerkenswert hoch.

Für das Suffix *-haft* finden sich im Jahre 1975 der geringste Wert; er liegt bei 19. Die Jahre 1900 und 1950 unterscheiden sich nicht entscheidend, weil 1900 39 Einmalbildungen und 1950 33 Hapax Legomena gefunden wurden; der hohe Wert von 97 Hapax Legomena im Jahre 1999 sticht jedoch heraus.

7.2.5 Baayens *P*

Auch für die adjektivbildenden Suffixe wurde die Produktivität im engeren Sinne (*P*) nach Baayens Index berechnet (Tab. 7.18).

	1900	1925	1950	1975	1999
<i>-ig</i>	0,03	0,04	0,03	0,03	0,04
<i>-isch</i>	0,04	0,05	0,03	0,03	0,05
<i>-lich</i>	0,02	0,03	0,02	0,02	0,02
<i>-mäßig</i>	0,05	0,1	0,14	0,12	0,13
<i>-los</i>	0,11	0,11	0,08	0,07	0,07
<i>-haft</i>	0,07	0,1	0,11	0,12	0,16
<i>-sam</i>	0,02	0,02	0,02	0,011	0,02
<i>-bar</i>	0,07	0,07	0,08	0,15	0,09

Tabelle 7.18: *P* der adjektivbildenden Suffixe *-ig*, *-isch*, *-lich*, *-mäßig*, *-los*, *-haft*, *-sam* und *-bar* für die Zeiträume 1900, 1925, 1950, 1975 und 1999

Die Auflistung der Werte zeigt für *-ig*, dass P zwischen 0.03 und 0.04 liegt; damit kann hier eine konstante Produktivität festgehalten werden.

Ein ähnliches Bild ergibt sich auch für das Suffix *-isch*, auch wenn eine etwas größere Schwankung zwischen 0.03 und 0.05 vorliegt.

Konstanz ist ebenfalls für *-lich* feststellbar; mit einer Ausnahme um 1925 durch den Wert 0.03 zeigen die Datenreihen zu allen anderen Messzeiträumen einen stabilen Wert von 0.02.

Das Suffix *-mäßig* hingegen zeigt Schwankungen hinsichtlich der Produktivität P auf, die zwischen 1925 und 1999 gering sind und zwischen 0.1 und 0.14 liegen, allerdings findet man um 1900 einen zu den anderen Ergebnissen relativ kleinen Wert von 0.05.

Die Produktivität von *-los* zu bewerten, erscheint unproblematisch, da es einen stetigen Abfall der Produktivität von 0.11 auf 0.07 gibt und dieser Abfall selbst nach Einbeziehung des Problemjahres zu beobachten ist.

Für *-haft* zeigt sich ein Aufwärtstrend von 0,07 auf 0,16.

Ebenso eindeutig erscheint die Interpretation von *-sam*, dessen Werte P für alle Jahre bei 0.02 liegen, wenn man das Problemjahr nicht zur Betrachtung mit hinzuzieht.

Dieselbe Methode lässt sich auch für *-bar* anwenden, um zu einer eindeutigen Interpretation zu kommen; nach Ausschluss des Jahres 1975 ist ein eindeutiger Trend erkennbar: der Anstieg der Produktivität von 0.07 auf 0.09.

Zusammenfassend kann demnach festgehalten werden, dass einige adjektivbildende Suffixe vorhanden sind, die eindeutige Sprachwandel-Tendenzen in Bezug auf die morphologische Produktivität erkennen lassen: Die Suffixe *-bar* und *-haft* gewinnen im 20. Jahrhundert an Produktivität, während *-los* an Produktivität verliert.

-ig, *-isch*, *-lich* und *-sam* zeichnen sich vorzugsweise durch eine konstante Produktivität aus; es sind nur geringe Schwankungen zu beobachten.

Das Suffix *-los* zeigt einen Abwärtstrend im Hinblick auf P ; P von *-mäßig* in hohem Maße schwankt, sodass keine eindeutige Interpretation möglich ist.

7.2.6 P^*

Im Folgenden soll nun auch die globale Produktivität grafisch dargestellt und betrachtet werden, um weitere Interpretationen vorzunehmen. Zunächst erfolgt, wie auch schon für die substantivbildenden Suffixe, eine diachrone Betrachtung, auf welche – dann im Anschluss – eine synchrone folgt.

P^* für das Suffix *-ig*

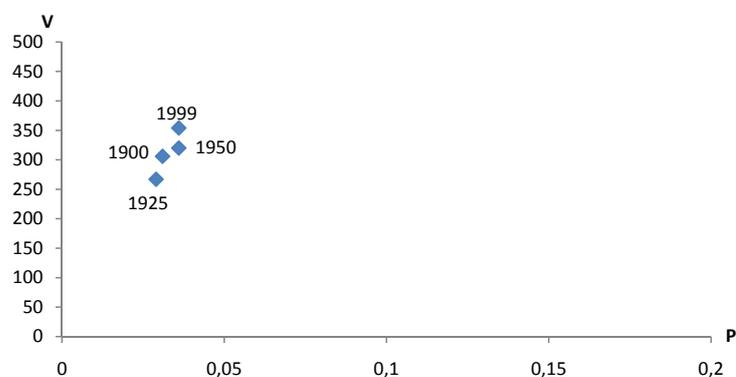


Abbildung 7.13: Globale Produktivität von *-ig* im 20. Jahrhundert

Ein eindeutiger Trend lässt sich anhand Abb. 7.13 nicht erkennen: die globale Produktivität ist zwar im Jahre 1999 höher als um 1950 und 1900, jedoch lässt das Wertepaar des Jahres 1925 auf eine niedrigere Produktivität schließen als im Jahre 1900.

Dennoch sei an dieser Stelle angemerkt, dass alle Wertepaare dicht beieinander liegen, sodass die globale Produktivität von *-ig* im 20. Jahrhundert relativ konstant wirkt.

P^* für das Suffix *-isch*

Auch bei Betrachtung des Suffixes *-isch* (Abb. 7.14) wird deutlich, dass keine eindeutigen Sprachwandel-Tendenzen erkannt werden können, vielmehr schwankt die Pro-

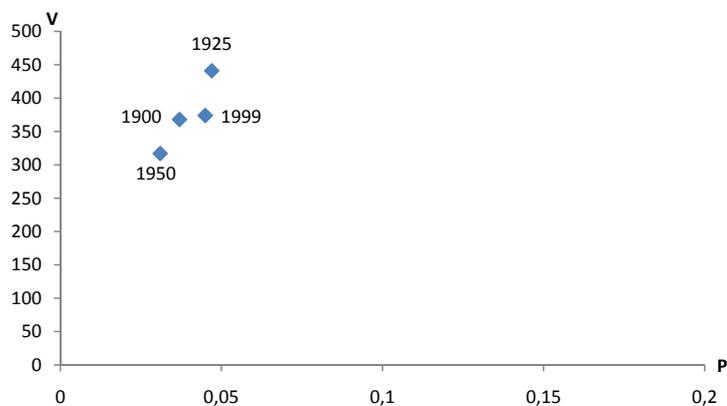


Abbildung 7.14: Globale Produktivität von *-isch* im 20. Jahrhundert

duktivität in einem bestimmten Rahmen; die Wertepaare lassen sich alle im 'linken oberen' Quadranten des Diagramms verorten.

P** für das Suffix *-lich

Wie schon für die beiden zuvor betrachteten Suffixe ist auch für *-lich* ersichtlich (Abb. 7.15), dass kleine Schwankungen vorhanden sind. Auffällig ist, dass das Wertepaar des Jahres 1925 auf eine besonders hohe globale Produktivität hinweist. Die anderen drei Wertepaare zeigen, dass die globale Produktivität sehr dicht beieinander liegt, sodass auf eine konstante Produktivität geschlossen werden darf; das Suffix darf durch die Verortung im oberen linken Quadranten als typenstark etikettiert werden.

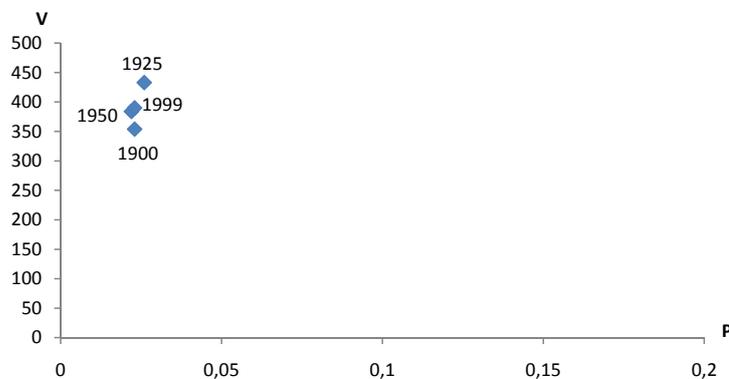


Abbildung 7.15: Globale Produktivität von *-lich* im 20. Jahrhundert

P** für das Suffix *-mäßig

Für das bisher in Bezug auf seine Werte schwankende Suffix *-mäßig* wird nun das Schaubild zur globalen Produktivität zur Betrachtung herangezogen (Abb. 7.16).

Die Betrachtung der Jahre 1900, 1925 und 1999 verdeutlicht, dass es einen Trend in Richtung ansteigender Produktivität gibt; lediglich das Jahr 1950 bereitet bei der Interpretation Schwierigkeiten, da zwar ein hoher Wert *P* berechnet wurde, die Typenfrequenz jedoch sehr niedrig ist. Dies lässt sich mit hoher Wahrscheinlichkeit auf die geringe Tokenanzahl zurückführen.

Werden im Jahre 1950 nur 40 Typen gebildet, so finden sich im Jahre 1999 bereits 71 realisierte Wortbildungen im Korpus. Besonders interessant bei der Betrachtung der globalen Produktivität von *-mäßig* ist die Tatsache, dass die 'Spanne der Produktivität' auffällig groß wirkt; insbesondere der Unterschied zwischen 1900 und 1999 sticht heraus.

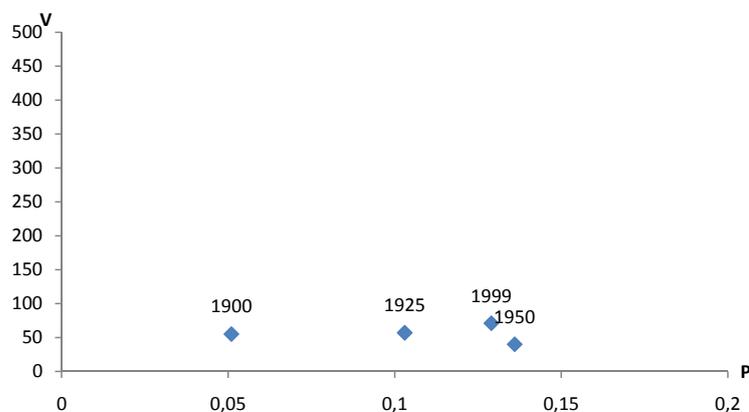


Abbildung 7.16: Globale Produktivität von *-mäßig* im 20. Jahrhundert

P** für das Suffix *-los

Für das Suffix *-los* zeigen die berechneten *P*-Werte eine abfallende Produktivität. Die globale Produktivität von *-los* im 20. Jahrhundert kann anhand Abb. 7.17 sichtbar gemacht werden.

Die Interpretation der Ergebnisse gestaltet sich ebenfalls schwierig, denn die Produktivitätsschwankungen im 20. Jahrhundert sind groß; zwar ähneln sich auf der einen Seite die Wertepaare von 1900 und 1925, und die *P*-Werte von 1950 und 1999, die geringer sind als zu den erstgenannten Zeiträumen, doch auf der anderen Seite liegen im Hinblick auf die Typenanzahl für die Zeiträume 1950 und 1999 deutliche Unterschiede vor.

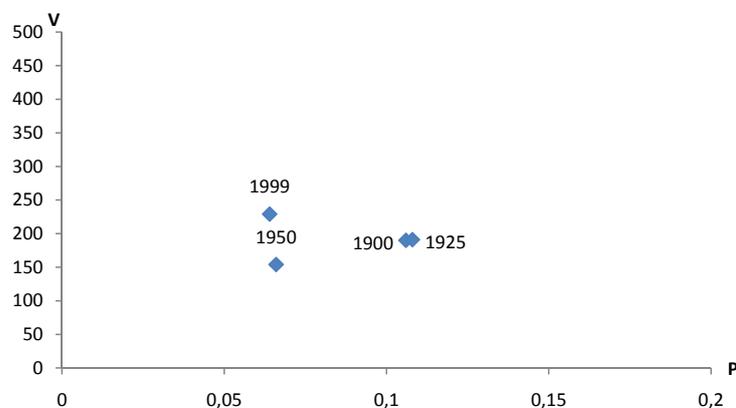


Abbildung 7.17: Globale Produktivität von *-los* im 20. Jahrhundert

P** für das Suffix *-haft

Wie schon für das Suffix *-mäßig* beobachtet werden konnte, zeigt sich auch für *-haft* eine ansteigende Produktivität (Abb. 7.18), allerdings nur unter Ausschluss des Jahres 1950, das erneut heraussticht, da weitaus weniger Typen gefunden werden konnten als für die anderen Untersuchungszeiträume.

Dennoch zeigt der Vergleich der Zeiträume 1900, 1925 und 1999 einen prototypischen Anstieg der globalen Produktivität, weil sowohl die Typenhäufigkeiten wie auch die Produktivitätswerte von *P* stetig zunehmen: Die Wertepaare verschieben sich vom linken unteren in den rechten unteren Bereich des Diagramms. Dabei kommt es allerdings nicht nur zu einer klaren Verschiebung von links nach rechts, sondern auch zu einem leichten Anstieg nach oben.⁵

⁵In diesem Zusammenhang wird auf Abb. 4.3 verwiesen.

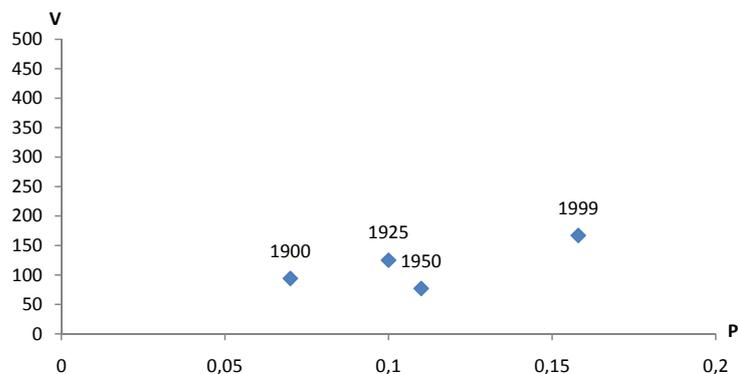


Abbildung 7.18: Globale Produktivität von *-haft* im 20. Jahrhundert

P** für das Suffix *-sam

Für das bisher als unproduktiv beschriebene Suffix *-sam* zeigt Abb. 7.19, dass sich dieses Bild auch im Hinblick auf die globale Produktivität bestätigen lässt.

Auch wenn für die fünf Zeiträume unterschiedliche Typenfrequenzen gefunden werden, so scheint hier eine Normierung durch *P* erfolgt zu sein. Geringe Schwankungen sollten nicht als Sprachwandeltendenzen interpretiert werden, vielmehr gilt für *-sam*, dass es sich um ein unproduktives Suffix handelt, das in seiner Unproduktivität im gesamten Jahrhundert konstant bleibt; alle Wertepaare können eindeutig im linken unteren Quadranten eingeordnet werden.

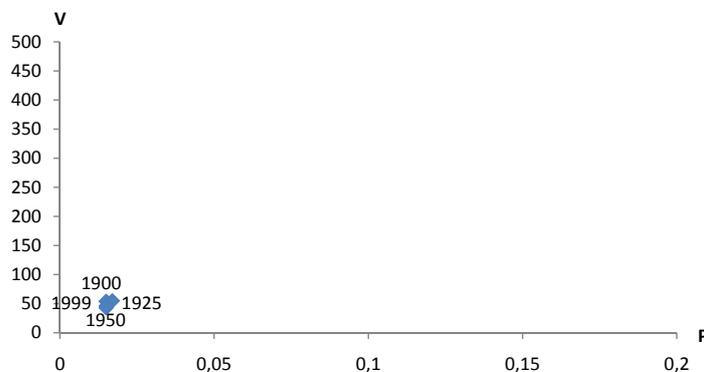


Abbildung 7.19: Globale Produktivität von *-sam* im 20. Jahrhundert

P** für das Suffix *-bar

Für die adjektivbildenden Suffixe soll nun als letztes das Suffix *-bar* betrachtet werden (Abb. 7.20).

Während die Interpretation von *-bar* hinsichtlich der Produktivität im engeren Sinne *P* unproblematisch gewesen ist, weil eine ansteigende Produktivität beobachtet wurde, so lässt sich dieses Bild durch Visualisierung der globalen Produktivität nicht aufrecht erhalten: Die Typenanzahl ist inkonstant und im Jahre 1950 besonders niedrig. Doch auch im Jahre 1925 findet man eine relativ geringe Anzahl an unterschiedlichen Bildungen, wenn man die Werte in Relation zu den Zeiträumen 1900 und 1999 betrachtet. Eine besonders hohe Anzahl an Typen findet man für den Zeitraum 1999.

Demnach kann für das Suffix *-bar* zusammenfassend festgehalten werden, dass die globale Produktivität im 20. Jahrhundert nicht unerheblich schwankt; sie ist für zwei Zeiträume im unteren und für zwei Zeiträume im oberen linken Quadranten (und in allen Fällen tendenziell rechts innerhalb der Quadranten) zu verorten.

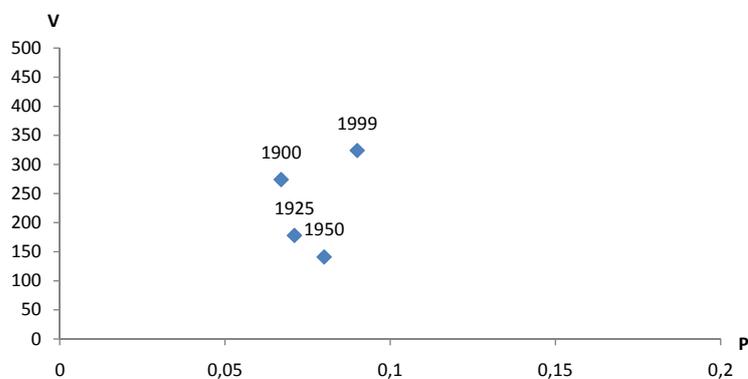


Abbildung 7.20: Globale Produktivität von *-bar* im 20. Jahrhundert

***P** für den Zeitraum 1900**

Die erste synchrone Darstellung der adjektivbildenden Suffix hinsichtlich *P** erfolgt für das Jahr 1900 (Abb. 7.21). Die drei Suffixe *-ig*, *-lich* und *-isch*, die zuvor parallele Verhaltensweise aufzeigten, lassen eine globale Produktivität erkennen, die dicht beieinander liegt: sie befinden sich im linken oberen Viertel der Grafik. Dennoch werden kleine Unterschiede sichtbar: *-isch* ist das produktivste der drei Suffixe; für *-ig* und *-lich* ist es schwierig, eine Rangfolge aufzustellen, da *-ig* die größere Produktivität *P*, *-lich* aber die höhere Anzahl der Typen hervorbringt.

Im Vergleich zu seinen adjektivbildenden Konkurrenten ist sichtbar, dass *-sam* durch seine Verortung im linken unteren Viertel unproduktiv ist.

P von *-haft* und *-mäßig* ist zwar höher als die der vier bereits beschriebenen Suffixe, aber die geringe Typenanzahl weist daraufhin, dass zum einen im Jahre 1900 noch nicht viele lexikalisierte (und natürlich auch nicht-lexikalisierte) Wortbildungen im Korpus zu finden waren.

Weiterhin zeigen *-los* und *-bar*, dass sie produktiv sind, auch wenn an dieser Stelle ebenfalls keine eindeutigen Aussagen gemacht werden können, welches der beiden

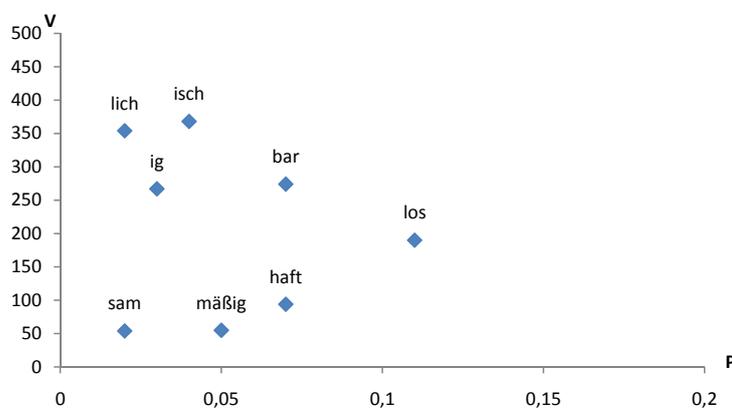


Abbildung 7.21: Globale Produktivität der adjektivbildenden Suffixe im Jahre 1900

Suffixe das produktivere ist: *-bar* findet sich in vielen Typen wieder, allerdings ist *P* von *-los* sehr hoch.

***P** für den Zeitraum 1925**

Die bereits bekannte Struktur der vier Suffixe *-lich*, *-ig*, *-isch* und *-sam* lässt sich auch anhand Abb. 7.22 für den Zeitraum 1925 wiederfinden: *-sam* darf als unproduktiv, die anderen drei Suffixe als frequenziell produktiv, aber hinsichtlich *P* als unproduktiv beschrieben werden.

Interessant ist die Struktur von *-lich*, *-ig* und *-isch* im Vergleich untereinander, da diese identisch zum Jahre 1900 ist, indem *-isch* als das produktivste Suffix der drei hoch frequenten bezeichnet werden darf und sich *-ig* und *-lich* durch unterschiedliche Verhaltensweisen voneinander differenzieren lassen: *-ig* durch einen höheren *P*-Wert und *-lich* durch die höhere Anzahl an Typen.

Die globale Produktivität des Suffixes *-bar* wird in der Mitte der Grafik verortet; sie ergibt sich durch eine mäßig hohe Typenanzahl und einen relativ hohen *P*-Wert.

Die drei Suffixe *-mäßig*, *-los* und *-haft* hingegen befinden sich alle drei im unteren rechten Viertel der Grafik. Dieses legt die folgende Interpretation nahe: berechnet werden konnte eine hohe Produktivität *P*. Dennoch existieren zwischen den drei Suffixen

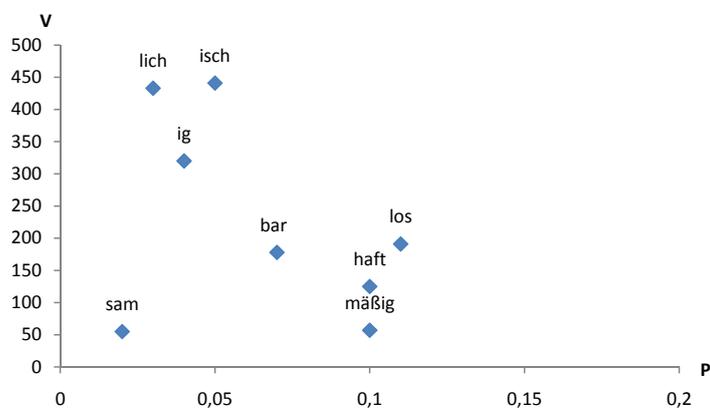


Abbildung 7.22: Globale Produktivität der adjektivbildenden Suffixe im Jahre 1925

Unterschiede hinsichtlich ihrer realisierten Typen *-mäßig* ist in nur in 57 Wortbildungen zu finden, während *-haft* immerhin in 125 Typen und *-los* sogar in 191 Typen zu finden ist.

***P** für den Zeitraum 1950**

Das Wortbildungsverhalten von *-sam* erweist sich auch im Jahre 1950 als unproduktiv (Abb. 7.23).

Die Betrachtung der drei hochfrequenten Suffixe *-lich*, *-ig* und *-isch* hingegen zeigen nun ein abweichendes Verhalten zu den beiden bereits betrachteten Jahrgängen 1900 und 1925, da *-isch* in diesem Jahr weniger typenstark auftritt und zudem auch eine vergleichbare Produktivität *P* zu *-ig* aufzeigen lässt.

Des Weiteren ist eine Veränderung der globalen Produktivität des Suffixes *-los* zu verzeichnen, das nun eine fast identische globale Produktivität wie *-bar* erkennen lässt und in der unteren Mitte der Grafik platziert ist.

Die Suffixe *-haft* und *-mäßig* befinden sich beide im unteren rechten Viertel des Schaubildes, wobei *-haft* im Jahre 1950 eine niedrige Produktivität *P* besitzt, aber typenfrequenter auftritt als *-mäßig*.

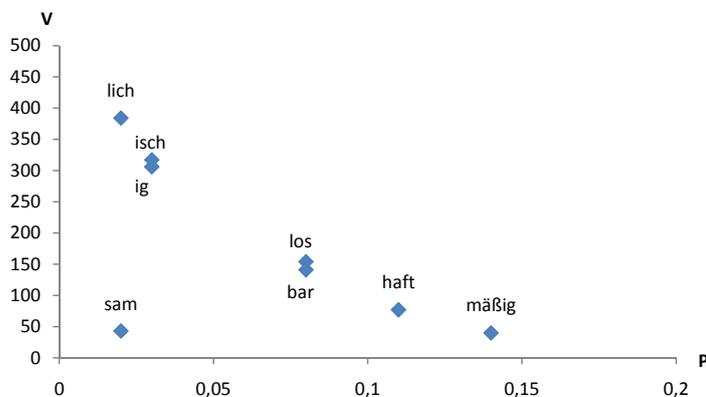


Abbildung 7.23: Globale Produktivität der adjektivbildenden Suffixe im Jahre 1950

P* für den Zeitraum 1975

Bei Betrachtung von Abb. 7.24 fällt auf, dass die Struktur der Anordnung Veränderungen zu den bereits betrachteten Jahrgängen aufzeigt: So kann zwar die bekannte Struktur von *-lich*, *-ig* und *-isch* wiedergefunden werden, aber für die anderen Suffixe können Veränderungen in der Verortung ihrer globalen Produktivität erkannt werden; lediglich das Suffix *-sam* scheint den gewohnten Platz im unteren linken Viertel gegen einen Platz im unteren rechten Viertel eingenommen zu haben.

Eine Veränderung der Platzierungen für *-haft*, *-mäßig* und *-los* ergibt sich durch den besonders hohen Wert *P* von *-bar* mit 0.15, der die drei genannten Suffixe nun links von *-bar* platzieren lässt.

In Abb. 7.24 rücken die drei Suffixe dicht zusammen und erwecken den Anschein, sie verhalten sich im Jahre 1975 ähnlich produktiv. Es ist in diesem Fall nicht auszuschließen, dass die vergleichbar geringe Tokenanzahl für das strukturell auffällige Ergebnis verantwortlich ist, sodass nun – im letzten Schritt der globalen Produktivitätsbetrachtung – die Ergebnisse des Jahres 1999 besprochen werden sollen.

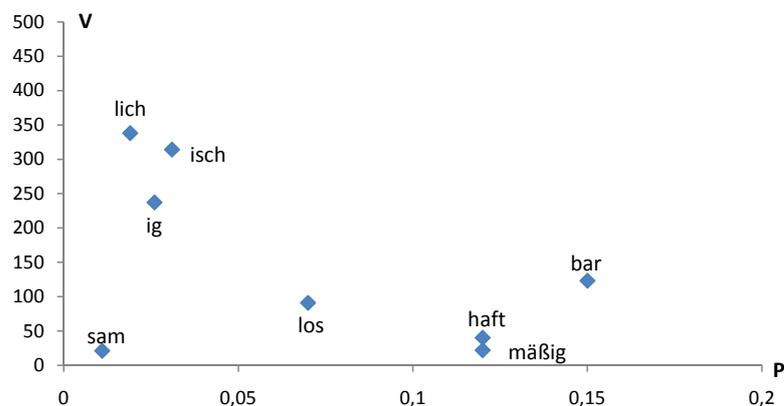


Abbildung 7.24: Globale Produktivität der adjektivbildenden Suffixe im Jahre 1975

P^* für den Zeitraum 1999

Abb. 7.25 zeigt die bekannte Struktur von *-lich*, *-ig* und *-isch* durch ihre Verortung im oberen linken Viertel des Diagramms. Auch das bereits bekannte Verhalten untereinander lässt sich wiederfinden: die Produktivität P von *-isch* ist höher als von *-ig* und *-lich*, hinsichtlich der realisierten Produktivität, die sich durch die Anzahl der Typen ergibt, verhalten sich *-isch* und *-lich* ähnlich. Für *-ig* konnten gering weniger unterschiedliche Bildungen gefunden werden.

Ebenfalls lässt sich *-sam* auf seinem gewohnten Platz im unteren linken Viertel wiederfinden – auf dem Platz für ein unproduktives Suffix.

-bar und *-los*, beide mittig verortet, zeigen im Jahre 1999, dass *-los* weniger stark zur Wortbildung beiträgt als *-bar*. *-mäßig* und *-haft* hingegen, die im unteren rechten Viertel platziert sind, fielen vorab durch hohe P -Werte auf, auch wenn die Typenhäufigkeit, insbesondere von *-mäßig*, als gering einzustufen war. Auch bei Betrachtung von Abb. 7.25 lässt sich dieses Bild aufrecht erhalten.

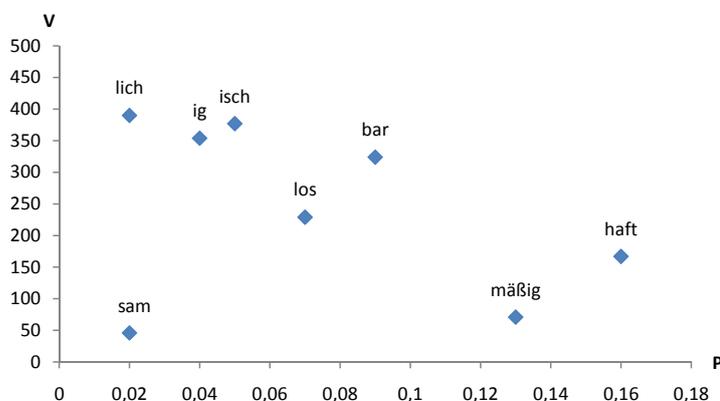


Abbildung 7.25: Globale Produktivität der adjektivbildenden Suffixe im Jahre 1999

Zusammenfassung der Ergebnisse von P^*

Auch für die adjektivbildenden Suffixe lässt sich ein Schema erstellen, das die Ergebnisse der globalen Produktivität zusammenfasst (Abb. 7.26).

Die drei Suffixe *-lich*, *-ig* und *-isch* nehmen das obere linke Viertel der Grafik ein; die jeweilige globale Produktivität ergibt sich durch eine geringe P -Werte, aber gleichzeitig hohe Typenfrequenzen.

In der Mitte und damit 'mäßig produktiv' nach P und zugleich 'mäßig typenfrequent' findet man die Suffixe *-bar* und *-los*.

Eindeutig kann das Suffix *-sam* durch die Verortung im linken unteren Viertel als unproduktiv deklariert werden.

-haft und auch *-mäßig* zeichnen sich durch eine hohe Produktivität P aus, sind aber gleichzeitig nur typenarm vertreten; klare Interpretationen hinsichtlich der globalen Produktivität sind aus den bereits dargelegten Gründen problematisch. Es sei an dieser Stelle auf die bereits geführte Diskussion in Kapitel 4.3.4 verwiesen.

Auffällig ist, dass sich die einzelnen abstrahierten Felder der Suffixe in ihrer Größe unterscheiden. Die Größe der Felder ergibt sich aus dem Schwankungsbereich der Suffixe; so gibt es einerseits Suffixe, deren Schwankungen über das 20. Jahrhundert

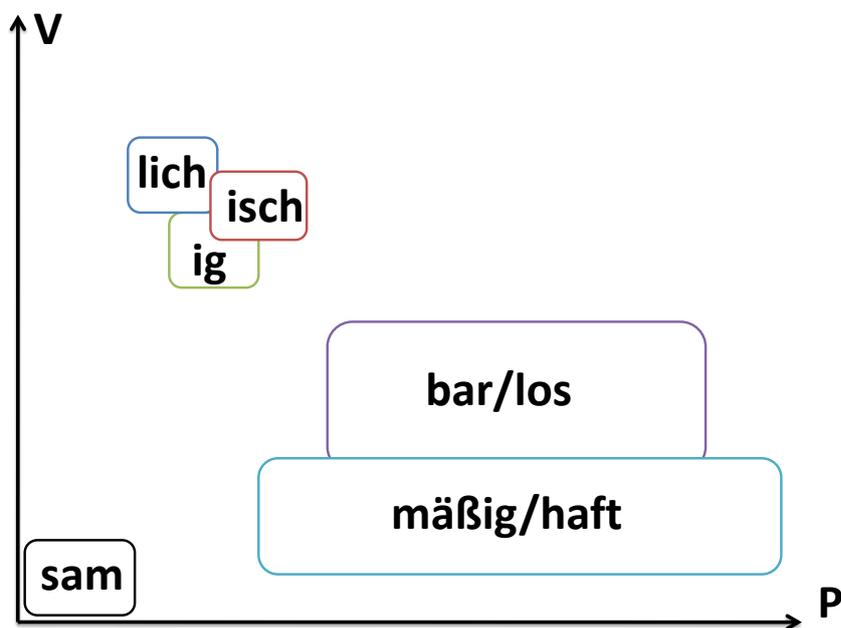


Abbildung 7.26: Schematische Zusammenfassung der globalen Produktivität der adjektivbildenden Suffixe *-bar*, *-sam*, *-haft*, *-lich*, *-los*, *-ig*, *-mäßig* und *-isch* im Zeitraum von 1900 bis 1999

besonders deutlich sind (z.B. *-bar* und *-los*) und andere Suffixe wie z.B. *-sam*, deren Schwankungen in Bezug auf die globale Produktivität relativ klein sind.

Eine derartige Darstellungs- und Herangehensweise ist in den bisherigen Arbeiten zur morphologischen Produktivität nicht zu finden und stellt deshalb ein Novität dar.

7.3 Fazit I

In den vorangegangenen Abschnitten wurden die für die Studie relevanten Suffixe zunächst deskriptiv betrachtet, indem die Aussagen, die in der linguistischen Literatur zur Produktivität der Wortbildungsmorpheme gemacht werden, zusammengefasst wurden.

Danach erfolgte die Darstellung und Auswertung und eines kleinen Produktionsexperimentes. Im dritten Schritt wurde eine Korpusstudie durchgeführt: Es wurden Datenreihen aus dem DWDS entnommen; die Auswertung der Ergebnisse erfolgte nach unterschiedlichen Methoden. Im Folgenden geht es nun darum, die Ergebnisse so zusammenzufassen, dass letztendlich die Frage überprüft werden soll, ob die in der

Literatur angegebenen deskriptiven Aussagen zur Produktivität durch die empirischen Studien bestätigt werden können.

7.3.1 Die substantivbildenden Suffixe aus synchroner Sicht

Tabelle 7.19 führt nun die Aussagen der Literatur zur Produktivität, die Ergebnisse des Vorexperimentes mit den Versuchspersonen und die Ergebnisse der Korpusstudie zusammen. Die Etikettierung mit ++ steht dabei für eine besonders hohe, + für eine hohe, O für eine mäßige und - für eine niedrige Produktivität.

Suffix	Aussagen der Literatur	Produktionsexperiment	Korpusstudie DWDS
<i>-tum</i>	O	-/O	?+?
<i>-heit/-keit</i>	+/(++)	+	+/>++
<i>-nis</i>	-/O	(-)/O	-
<i>-ung</i>	(+)/++	++	++
<i>-sal</i>	-	-	-

Tabelle 7.19: Zusammenfassung der Ergebnisse für die substantivbildenden Suffixe

Für *-sal* konnten die Aussagen der Literatur zur absoluten Unproduktivität des Suffixes gleich in zweierlei Weise bestätigt werden: sowohl bei der Produktion der Wörter durch die Probanden als auch in der Studie mit dem DWDS zeigte sich, dass die Aussagen der Literatur zutreffend sind.

Das Suffix *-nis* erwies sich ebenfalls als unproduktiv, auch wenn zwischen *-sal* und *-nis* ein Unterschied besteht, der sich schon bei Betrachtung der Typenhäufigkeit erkennen lässt. Immerhin gab es drei Zeiträume, in denen über vierzig unterschiedliche *-nis*-Bildungen gefunden wurden; dazu gehört z.B. das Jahr 1999. *-sal* hingegen bringt als höchste Typenfrequenz sieben Bildungen hervor, im Jahre 1999 allerdings nur fünf. Die Typen-Token-Relation ist für *-nis* im Jahre 1999 deutlich niedriger als die TTR der anderen Suffixe. Auch dies darf als Indiz für eine deutlich schwächere Produktivität gewertet werden. Insgesamt wurden nur fünf Hapaxe gefunden. In Relation zu der Anzahl der Token ergibt sich daraus auch ein niedriger Wert *P*. Besonders anschaulich zeigt sich der Unterschied der morphologischen Produktivität dann in den Abbildungen zur globalen Produktivität durch die Verortung im linken unteren Viertel

des Diagramms. Im Vergleich zu seinen substantivbildenden Konkurrenten darf *-nis* eindeutig als unproduktiv bezeichnet werden.

Das Suffix *-ung* hingegen erweist sich im Jahre 1999 als sehr produktiv. Insgesamt wurden 689 Typen gefunden, im Vergleich zu den Allomorphen *-heit* und *-keit* eine etwas niedrigere Anzahl, wobei sich dieser Wert nur bedingt vergleichbar ist, da für die Allomorphe zur Ermittlung der Typenfrequenzen höhere Tokenfrequenzen zur Verfügung standen.

Die Typen-Token-Relationen allerdings sind für die Suffixe *-ung* und *-heit/-keit* ähnlich und zugleich hoch. Das Suffix *-ung* findet sich in 286 Hapaxen wieder, was eine Produktivität P von etwa 0.06 ergibt. Dieser Wert ist mit dem Wert von *-heit/-keit* vergleichbar und darf als produktiv eingestuft werden. Die globale Produktivität von *-ung* und auch den Allomorphen ist ebenfalls hoch. Die sich daraus ergebende Konsequenz ist, dass sowohl bei *-heit/-keit* als auch bei *-ung* interpretiert werden kann, dass die Aussagen der linguistischen Literatur zutreffend sind, wenn in allen Fällen von hoch produktiven Suffixen gesprochen wird.

Der Problemfall der substantivbildenden Suffixe ist *-tum*, das in der Literatur als **schwach/gering produktiv** oder **beschränkt aktiv** deklariert wird. Diese Aussagen können nicht ohne Weiteres bestätigt werden, insbesondere nicht, was die Ergebnisse der Korpusstudie anbelangt.

Zunächst einmal zeigt die Tokenfrequenz, dass *-tum* als Suffix in sehr wenigen Wortformen belegt ist. Während das als unproduktiv beschriebene Suffix *-nis* im Jahre 1999 in immerhin in 1902 Token vertreten ist, kommt *-tum* nur auf eine Frequenz von 335, die demnach sechsmal geringer ist als die von *-nis*. Die Typenfrequenz jedoch zeigt, dass die Anzahl der tatsächlichen Bildungen von *-tum* mit jener von *-nis* vergleichbar ist, insbesondere dann, wenn man *-nis* und *-tum* in Relation zu *-ung* oder *-heit/-keit* betrachtet. Schaut man sich die Typen-Token-Relation von *-tum* an, so ist diese besonders hoch. Der hohe Wert jedoch ergibt sich einzig und allein aus der Tatsache, dass die Tokenhäufigkeit sehr gering ist.

Des Weiteren lässt sich Folgendes festhalten: es finden sich immerhin 35 Hapaxe, 30 Hapaxe mehr als im Vergleich zu *-nis*. Daraus ergibt sich ein hoher Wert P , der nahezu doppelt so hoch ist wie jene von *-heit/-keit* und *-ung*. Durch die Visualisierung der globalen Produktivität von *-tum* wird das oben beschriebene Bild auch grafisch sichtbar gemacht: das Suffix ist niederfrequent im Korpus vertreten, aber dem Wert P nach sehr produktiv.

Diese Tatsache wirft die Frage auf, wie dieses Verhalten zu bewerten ist: Gibt es Affixe des Deutschen, die wenige unterschiedliche Bildungen hervorbringen, aber dennoch produktiv sind?

Die Frage ist sehr schwierig zu beantworten, da es prinzipiell möglich wäre, dass zum Beispiel ein Affixoid/Halbaffix untersucht wird, dessen Frequenzen im Korpus zunächst belegen, dass nicht viele unterschiedliche Bildungen vorhanden sind. Gleichzeitig aber müsste sich ein Affixoid auf dem Weg seiner Grammatikalisierung durch eine gewisse 'Reihenbildung' auszeichnen und sich im Zuge der unidirektionalen Entwicklung nach und nach in immer mehr Wortbildungen finden. In solch einem Fall könnte man durchaus davon ausgehen, dass dieses sprachliche Zeichen ein produktives ist. Im Falle von *-tum* jedoch trifft dies nicht zu, denn bei *-tum* handelt es sich nicht um ein Affixoid, sondern um ein Affix des Deutschen, zu welchem in der Gegenwartssprache kein freies Lexem in paralleler Verwendung existiert.

Das bisher in der Literatur diskutierte Problem der ungleichen Datenmengen (Lieber 1992) scheint auch hier zu Interpretationsschwierigkeiten von *-tum* zu führen. Klare Aussagen hinsichtlich starker, mäßiger oder schwacher Produktivität im Jahre 1999 ließen sich vorzugsweise dann treffen, wenn eine größere Datenmenge vorhanden wäre.

Andererseits muss man an dieser Stelle festhalten, dass die Ergebnisse der Studie so sind, wie sie sind und deshalb auch etwas aussagen: wenn in einem Korpus eine bestimmte Menge an Token für alle Suffixe untersucht wird, so haben alle Morpheme denselben 'Lebensraum' zur Verfügung und können theoretisch in gleichem Maße zur Wortbildung beitragen. Warum jedoch wird *-ung* in sehr vielen Wortformen und auch unterschiedlichen Wortbildungen realisiert, während *-tum* nur in sehr wenigen Wortbildungen vorkommt?

Die Antwort ist sicherlich semantischer Natur sein: schaut man sich den Anwendungsbereich von *-tum* an, so wird deutlich, dass *-tum* hauptsächlich Personenbezeichnungen zu Abstrakta ableitet.⁶ Dieses Wortbildungsmuster scheint im Jahre 1999 durchaus noch produktiv zu sein, wenn man *P* betrachtet. Das Suffix *-ung*, das in jedem Falle als produktiv bezeichnet werden darf, kann deshalb – zumindest aus qualitativer Sicht – nicht zum direkten Vergleich mit dem Suffix *-tum* herangezogen werden, da *-ung* einen völlig **anderen Anwendungsbereich** besitzt, indem *-ung* verbale Basen zu

⁶Z.B. *Professorentum, Kaisertum, Lehrertum*.

Nomen ableitet. Damit ergeben sich folglich auch andere Regeln und Beschränkungen für die beiden Suffixe.⁷

Die Produktivität im Zusammenhang mit Allomorphie ist bisher nicht betrachtet worden. Die Ergebnisse der vorliegenden Arbeit von *-heit* und *-keit* lassen einige Thesen zu: Bereits die jeweiligen (unterschiedlichen) Frequenzen der Typen belegten, dass *-keit* häufiger in Wortformen vorhanden war als *-heit*. Nach Untersuchung der Token und Erstellung der Typenliste konnte dieses Bild bestätigt werden, indem im Jahre 1999 etwas dreimal so viele Typen mit *-keit* produziert worden sind wie mit *-heit*.

Die Berechnung der Typen-Token-Relationen hat gezeigt, dass auch diesbezüglich *-keit* das produktivere der beiden Allomorphe ist; Selbiges kann nach Analyse der Häufigkeiten der Hapax Legomena festgehalten werden. Insgesamt darf geschlossen werden, dass die Aussage, es handele sich bei *-heit* und seinem Allomorph um ein produktive Wortbildungsmorpheme, bestätigt werden kann.

7.3.2 Die adjektivbildenden Suffixe aus synchroner Sicht

Auch für die adjektivbildenden Suffixe werden die Ergebnisse aus synchroner Sicht – anhand von Tab. 7.20 – zusammengefasst. Es gelten hier dieselben Etikettierungen, wie sie bereits für die substantivbildenden Suffixe verwendet wurden.

Suffix	Aussagen der Literatur	Produktionsexperiment	Korpusstudie DWDS
<i>-ig</i>	++	+	O
<i>-isch</i>	++	+	O
<i>-lich</i>	++	O	-
<i>-mäßig</i>	++	++	++
<i>-bar</i>	++	++	+
<i>-sam</i>	-/O	O	-
<i>-los</i>	++	+	+
<i>-haft</i>	+ / ++	O	++

Tabelle 7.20: Zusammenfassung der Ergebnisse für die adjektivbildenden Suffixe

In der linguistischen Literatur wurden alle adjektivbildenden Suffixe bis auf *-sam* als produktiv etikettiert. Das Vorexperiment mit den Sprechern des Deutschen zeigte, dass *-bar* und *-mäßig* durchaus sehr produktiv zur Wortbildung beitragen, insbeson-

⁷Siehe dazu Kapitel 2.4.

dere mit *-mäßig* wurden sehr viele nicht-lexikalisierten Bildungen produziert. Andererseits konnte die hohe Produktivität von *-lich*, *-isch* und *-haft* anhand der Daten der Sprecher nicht belegt werden.

Nun ist zu überlegen, welche Schlüsse man aus der Korpusstudie ziehen kann. Wie bereits in den vorangegangenen Abschnitten der Arbeit skizziert wurde, gibt es drei Suffixe, die sich hinsichtlich ihrer Produktivität sehr ähnlich verhalten: *-ig*, *-isch* und *-lich*. Diese drei Suffixe sind im Jahre 1999 hoch frequent im Korpus vertreten; ihre Produktivität nach Baayen ähnelt sich. Daraus ergibt sich nach Visualisierung der globalen Produktivität eine Verortung der Suffixe im linken oberen Viertel des Diagramms.

Mit *-bar* werden mehr als 300 Typen gefunden; mit *-los* werden immerhin noch 229 unterschiedliche Wörter durch das Korpus belegt, mit *-haft* 167. Auch hier lässt sich festhalten, dass eine Etikettierung als 'typenfrequent' durchaus gerechtfertigt ist.

Mit einer Anzahl der Typen unter 100 erwecken sowohl *-sam* als auch *-mäßig* den Anschein, sie seien unproduktiv. Im Falle von *-sam* kann dies als Bestätigung der Aussagen der Literatur angesehen werden; für *-mäßig* allerdings ist das Ergebnis überraschend. Das Suffix *-mäßig* wurde von den Versuchspersonen des Vorexperimentes zur Bildung vieler nicht-lexikalisierten Wörter verwendet. Damit ergab sich unter diesem Aspekt eine hohe Produktivität.

Des Weiteren konnte festgestellt werden, dass *-los*, *-bar*, *-mäßig* und *-haft* jeweils einen sehr hohen TTR-Wert besaßen. Für die Produktivität P konnte dieses Bild nicht exakt aufrecht erhalten bleiben. Zwar kann man *-haft* mit einem 0.16-Wert als Rang-Erster vor *-mäßig* mit 0.13 als 'Gewinner' dieses Wettbewerbs bezeichnen, aber der P -Wert von *-los* ist im Vergleich zu den anderen beiden eher niedrig (0.07) und unterliegt damit auch *-bar*, das nach Berechnung einen Wert von 0.09 besitzt.

Der geringste Wert für P ergibt sich aus den Frequenzen von *-sam*, doch dieser Wert (0.015) ist nicht bedeutend niedriger als jener von *-lich* (0.023). Mit 0.0359 scheint *-ig* ein wenig produktiver zu sein als *-lich*, er wird allerdings durch den Wert 0.046 von *-isch* übertroffen. Die Produktivität P lässt sich gerade im Falle von *-lich*, *-ig* und *-isch* gut mit einander vergleichen, da genügend Datenmaterial zur Verfügung stand. Und an diesen Beispielen kann demonstriert werden, dass es dennoch kleine Unterschiede hinsichtlich ihres Wortbildungsverhaltens gibt: sie verhalten sich nicht gleich produktiv, auch wenn sie in der linguistischen Literatur dieselbe Etikettierung bekommen.

Zusammenfassend lässt sich Folgendes festhalten: die Werte der Korpusstudie stellen die Kategorisierung der Literatur in Frage. Wenn gesagt wird, dass *-mäßig*, *-los*, *-haft*, *-bar*, *-ig*, *-lich* und *-isch* alle (stark) produktiv sind, so ist unklar, warum die empirischen Ergebnisse jeweils unterschiedlich ausfallen. Während die einen hoch frequent vertreten sind, finden sich bei anderen hohe *P*-Werte.

Bei den typenstark vertretenden Suffixen ist davon auszugehen, dass eine Vielzahl an Bildungen existiert, die stark lexikalisiert ist. Die hohen *P*-Werte bei den anderen Suffixen, die sich nicht nur eine hohe Typenanzahl auszeichneten, lassen unterschiedliche Annahmen zu: einerseits könnte es sich tatsächlich um produktive Suffixe handeln, deren Bildungen zum Teil nicht lexikalisiert werden, andererseits bleiben Zweifel stehen, da das Datenmaterial bezüglich der Menge nicht für alle Suffixe identisch und zum Teil sehr gering gewesen ist.

7.4 Fazit II

Die schwierigere Aufgabe im Hinblick der Produktivitätsbetrachtung ist die Interpretation aus diachroner Sicht. Zunächst einmal ist festzuhalten, dass es für vier der fünf gewählten Jahre eine etwa vergleichbare und auch große Zahl an (Gesamt-)Token gab, was eine diachrone Herangehensweise gerechtfertigt. Lediglich im Jahre 1975 war die Gesamttokenanzahl bedeutend kleiner. Aus diesem Grund wurden die Werte für dieses Jahr mit Vorsicht betrachtet.

Es stellte sich allerdings erstaunlicherweise heraus, dass es sich beim Jahr 1950 um ein 'Ausreißer-Jahr' handelt oder zumindest um ein Jahr, in welchem die Produktivitätswerte abweichen.

Die aus der Reihe fallenden Werte sollen nicht einfach ignoriert werden, vielmehr stellt sich die Frage nach dem Warum der auffälligen Werte, obwohl die Tokenanzahl in den Jahren 1900, 1925, 1950 und 1999 vergleichbar ist. Ein genauerer Blick auf die Zusammensetzung der Jahrgänge im Korpus liefert eine mögliche Antwort:

- 1900 : 202 Dokumente : ca. 1,3 Mill. Token
- 1925 : 363 Dokumente : ca. 1,2 Mill. Token
- 1950: 1368 Dokumente : ca. 1,2 Mill. Token
- 1975 : 576 Dokumente : ca. 0,6 Mill. Token

➤ 1999 : 217 Dokumente : ca. 1,6 Mill. Token

Während es für die Jahre 1900, 1925 und 1999 eine vergleichbare Anzahl an Dokumenten gibt, so ist dieses für 1950 nicht zu verzeichnen: Mit 1368 Dokumenten bietet das DWDS in diesem Jahr eine Dokument-Vielfalt als Recherche-Möglichkeit an; diese Zusammensetzung jedoch scheint die Produktivität der untersuchten Suffixe global betrachtet zu beeinflussen.

Bisher existiert keine quantitativ ausgerichtete Studie, die klären kann, ob eine identische Menge an Token Unterschiede in der derivationalen morphologischen Produktivität impliziert, wenn es sich bei allen Teilkorpora um Korpora handelt, die sich hinsichtlich ihrer Anzahl an zusammengeführten Dokumenten unterscheiden. Allerdings konnte Baayen (1994) zeigen, dass sich die Produktivität von Affixen in verschiedenen Textsorten unterscheiden kann.⁸

In der Literatur lassen sich keine Aussagen über die Entwicklung der untersuchten Suffixe im 20. Jahrhundert finden, vielmehr werden – wie bereits an mehreren Stellen angerissen – kategoriale Aussagen über die Produktivität im Allgemeinen gemacht. Nun gilt es zu interpretieren und zu diskutieren, welche Informationen man aus den Ergebnissen der Korpusstudie mit dem DWDS festhalten kann.

7.4.1 Die substantivbildenden Suffixe aus diachroner Sicht

Begonnen wird wie gewohnt mit den substantivbildenden Suffixen (Tab. 7.21). Sowohl für Tab. 7.21 wie auch Tab. 7.22 gelten dieselben Etikettierungen: - für eine sinkende, o für eine konstant bleibende und + für eine steigende Produktivität.

Für die substantivbildenden Suffixe ist eine Interpretation aus diachroner Sicht schwierig; dies begründet sich wie folgt: Während es auf der einen Seite ein gänzlich unproduktives *-sal* und ein ziemlich unproduktives *-nis* gibt und beide Suffixe Konstanz in ihrer Unproduktivität aufzeigen, existiert allerdings auch ein produktives *-ung* und ein produktives *-tum*, dessen Ergebnisse schwierig zu deuten sind, eine kla-

⁸Dazu siehe auch Evert und Lüdeling (2001, 169):

„Since we want to compute the probability P of a new complex word formed by a given word formation pattern in a given text type (for example newspaper data or technical text) and because the productivity of word formation patterns is highly dependent on text type (neoclassical words, for example are much more likely in scientific texts than in everyday language, see also Baayen 1994) it is necessary to sample a corpus of that text type.“

Suffix	Entwicklung im 20. Jahrhundert
<i>-tum</i>	?
<i>-heit/-keit</i>	?
<i>-nis</i>	-
<i>-ung</i>	+
<i>-sal</i>	-

Tabelle 7.21: Zusammenfassung der Ergebnisse aus diachroner Sicht für die substantivbildenden Suffixe

re Tendenz hinsichtlich steigender oder fallender Produktivität wird nicht beobachtet. Zudem ist *-ung* kleineren Schwankungen ausgesetzt, *-tum* zeigt einen leichten Aufwärtstrend, was die Produktivität anbelangt, aber ist, wie bereits hinreichend skizziert, nur typenarm in den Teilkorpora vertreten. Auch für *-heit/-keit* ist es nicht möglich, einen Wandel anhand der berechneten Werte festzustellen.

7.4.2 Die adjektivbildenden Suffixe aus diachroner Sicht

Auch bei den adjektivbildenden Suffixen (Tab. 7.22) ist eine Interpretation aus diachroner Sicht problematisch.

Suffix	Entwicklung im 20. Jahrhundert
<i>-ig</i>	o
<i>-isch</i>	o
<i>-lich</i>	o
<i>-mäßig</i>	+
<i>-los</i>	-
<i>-haft</i>	+
<i>-sam</i>	o
<i>-bar</i>	+

Tabelle 7.22: Zusammenfassung der Ergebnisse aus diachroner Sicht für die adjektivbildenden Suffixe

Für die Suffixe *-lich*, *-ig*, *-isch* und *-sam* sind keine gravierenden Schwankungen zu vernehmen, sodass den Werten nach davon auszugehen ist, dass sich die Produktivität

dieser Suffixe nicht verändert hat. Dieses Ergebnis könnte durchaus zutreffend sein, denn Veränderungen von Wortbildungsmorphemen vollziehen sich in der Regel über einen langen Zeitraum. In dieser Arbeit konnten nur 100 Jahre untersucht werden. Es ist möglich, dass die morphologische Produktivität über ein gesamtes Jahrhundert konstant bleibt.

Bei drei Suffixen ist ein Anstieg der *P*-Werte ersichtlich. Dazu gehören *-mäßig*, *-bar* und *-haft*: *P* von *-mäßig* steigt im Laufe des 20. Jahrhunderts von 0.05 auf 0.13 an, *P* von *-haft* von 0.07 auf 0.16 und *P* von *-bar* von 0.07 auf 0.09. Nicht verkannt werden darf allerdings, dass sowohl für *-mäßig* als auch für *-haft* jeweils nur wenige Token zur Verfügung standen. Für *-bar* schwankte die Menge der Token, lag aber in einem mäßig hohen Bereich.

Das einzige Suffix, das scheinbar an Produktivität verliert, ist *-los*: die Produktivität fällt von 0.11 auf 0.07. Doch muss auch hier erwähnt werden, dass nicht besonders viele Derivationen für die Untersuchung bereit standen.

Als Fazit bleibt festzuhalten, dass sich im Falle der adjektivbildenden Suffix sowohl Tendenzen finden lassen, die auf eine absteigende, aber auch eine ansteigende Produktivität hinweisen. Die Frage, die sich in diesem Zusammenhang allerdings stellt, ist, inwieweit die vorgenommenen Interpretationen stichhaltig sind, denn hier können dieselben Zweifel wie im vorangegangenen Abschnitt aufgeführt werden, nämlich dass es wünschenswert wäre, größere Korpora und größere Zeiträume zu untersuchen.

Methodologische Diskussion

Anhand von Häufigkeitsdaten und etablierten Produktivitätsmaßen (TTR, P und P^*) wurden deutsche Derivationssuffixe hinsichtlich ihrer Produktivität betrachtet.

Die Daten, die zur Analyse verwendet wurden, stammen aus dem DWDS-Kernkorpus, das als ausgewogenes deutsches Korpus gilt. Dass die statistische Produktivitätsmessung bisher nicht kritiklos zu betrachten ist, wurde an vielen Stellen bereits angerissen, nun aber soll diese Diskussion vertieft werden.

Anhand der bisher betrachteten Ergebnisse haben sich insbesondere drei problematische Punkte gezeigt, die zu kritischen Ergebnissen führen können:

1. Lexikalisiertheit
2. Stichprobengröße
3. Zufallsschwankungen

Diese drei Aspekte werden im Folgenden nacheinander besprochen.

Die synchrone Analyse der Daten konnte zeigen, dass sich viele Suffixe frequentiell in ihrer Token- und Typenanzahl (und auch Anzahl der Hapaxe) voneinander unterscheiden, obwohl sie in den linguistischen Werken mit derselben Etikettierung 'produktiv' oder 'aktiv' versehen werden: Während die einen Suffixe typenstark vertreten sind, liegen wiederum andere vor, bei welchen eine Vielzahl an einmaligen oder niederfrequenten Bildungen beobachtet werden konnte. Man kann Folgendes vermuten:

Bei den typenstark vertretenden Suffixen ist mit großer Wahrscheinlichkeit davon auszugehen, dass sich darunter viele lexikalisierte Wörter befinden.

Dieses Problem, die lexikalisierten Formen als Parameter einer mathematischen Formel zu berücksichtigen, lässt sich derzeit noch nicht lösen. Wie schon zu Beginn der Arbeit diskutiert, ist die Ermittlung von lexikalischen Einheiten äußerst schwierig (Barz 1999), da es bei Lexikalisierungsprozessen ein Kontinuum zwischen stark lexikalisierten und neu gebildeten Wörtern gibt.

Doch gerade lexikalisierte Formen können die Ergebnisse beeinflussen, indem sie die Tokenanzahl vergrößern.

Es ist zu überlegen, ob es sinnvoll sein könnte, stark lexikalisierte Formen vor dem Berechnen der Typen-Token-Relationen und der *P*-Werte zu entfernen. In der linguistischen Literatur wird man allerdings nicht fündig, was *stark lexikalisiert* in Zahlen bedeutet; zudem sind nicht nur frequente Formen lexikalisiert.

Unklar ist aus diesem Grund, an welcher Stelle der Datenreihen ein Schnitt gemacht werden müsste (Abb. 8.1), indem bestimmte hoch frequente Wörter zur Messung der Produktivität entfernt werden:

1900	1925	1950	1975	1999
Deutung 298	Stellung 195	Regierung 533	Beziehung 274	Werbung 343
Stellung 240	Zeitung 166	Stellung 128	Regierung 173	Stellung 137
Beziehung 184	Achtung 104	Haltung 109	Entwicklung 121	Entwicklung 108
Wirkung 162	Haltung 88	Handlung 102	Stellung 107	Bewegung 104
Erscheinung 161	Teilung 87	Stimmung 102	Ordnung 92	Lösung 88
Entwicklung 135	Bewegung 86	Ordnung 101	Haltung 80	Regierung 85
Bewegung 130	Wohnung 81	Sammlung 80	Handlung 77	Leistung 73
Richtung 101	Ordnung 80	Erklärung 72	Stimmung 75	Führung 72
Wendung 100	Deutung 78	Fassung 62	Bewegung 69	Bedingung 69
Leistung 98	Stimmung 71	Zeitung 59	Deutung 64	Bevölkerung 67
Stimmung 80	Wirkung 64	Bewegung 59	Erklärung 63	Ordnung 66
Bedingung 75	Handlung 61	Hoffnung 55	Sammlung 62	Richtung 64
Ordnung 71	Beziehung 60	Wohnung 55	Bedingung 61	Forderung 64
Forderung 70	Meinung 59	Ladung 52	Wohnung 59	Beziehung 63
Lösung 63	Richtung 58	Leistung 52	Spannung 56	Wendung 63
Wertung 53	Führung 57	Entwicklung 50	Lösung 53	Wirkung 62
Befriedigung 52	Erfahrung 54	Führung 50	Teilung 52	Stimmung 54
Bildung 49	Bedingung 53	Achtung 47	Forderung 51	Zeitung 50
Teilung 48	Kleidung 51	Entscheidung 47	Änderung 50	Haltung 48
Empfindung 44	Erscheinung 50	Verteidigung 47	Bildung 49	Verbindung 47
Erfahrung 44	Wendung 50	Beziehung 46	Führung 49	Modernisierung 45
Voraussetzung 42	Leitung 48	Richtung 46	Richtung 48	Handlung 43

Abbildung 8.1: Ausschnitt des Token-Frequenz-Spektrums für das Suffix *-ung*, die mithilfe des DWDS' ermittelt werden konnten

Für die zukünftige Forschung jedoch wäre es wünschenswert, eine mathematisch begründete Lösung zu finden, um den Parameter Lexikalisiertheit in eine Formel zur Berechnung von Produktivität mit einzubeziehen.

In der vorliegenden Arbeit wird ein Produktivitätsbegriff angenommen, der besagt, dass ein Suffix genau dann als produktiv bezeichnet werden kann, sofern die Möglichkeit existiert, viele neue Wörter zu bilden. Zu den klassischen Maßen gehört neben der Typen-Token-Relation auch das von Baayen vorgeschlagene Produktivitätsmaß P .

Zunächst sollen nun einige Überlegungen zu den Kategorien Token und Typen skizziert werden. Je größer ein Korpus ist, desto größer ist auch die Anzahl der Token. Und je länger die Liste der Token ist, desto größer ist in der Regel auch die Anzahl der Typen: Die Typen (V) lassen sich folglich auch als Funktion von N auffassen (Baayen 1992, 113): $V(N)$. Da es sich dabei um eine nicht-lineare Funktion handelt, ist es nicht möglich vorauszusagen, wie sie sich weiterentwickelt. Für einen Linguisten können in diesem Zusammenhang insbesondere zwei Dinge von Interesse sein:

- die Steigung der Kurve(n) und
- die Entwicklung der Kurve(n) auf Basis großer Datenmengen.

Baayen fand heraus, dass P die Steigung der Wachstumskurven der Typen anzeigt: bei einem produktiven Wortbildungsprozess steigt die Kurve stetig an, bei einem unproduktiven Prozess kommt es schnell zu einem Abflachen der Kurve (Abb. 8.2).¹

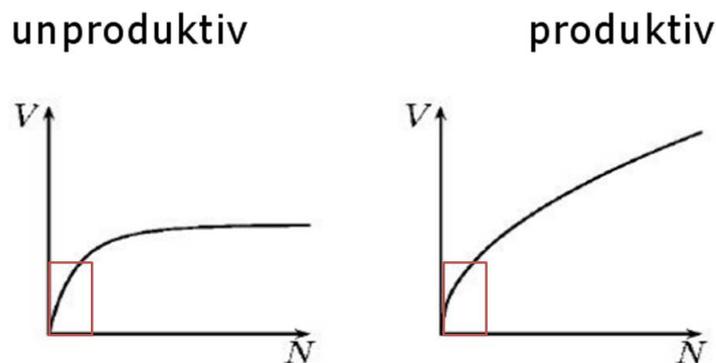


Abbildung 8.2: Wachstumskurven der Hapax Legomena für einen prototypisch unproduktiven und einen produktiven Wortbildungsprozess

¹Siehe Baayen (2008, 223): „The growth rate is a probability, the probability that, after having read N tokens, the next token sampled represents an unseen type, a word that did not occur among the preceding N tokens.“

Auffällig ist, dass es sowohl bei produktiven wie auch bei unproduktiven Wortbildungsmustern (roter Bereich von Abb. 8.2) zunächst einen steilen Anstieg der jeweiligen Kurve gibt. Dies impliziert folgende These:

Je kleiner die Tokenanzahl des Korpus und damit auch der Tokenliste des zu untersuchenden Suffixes ist, desto größer ist die Wahrscheinlichkeit, dass viele Wörter nur niederfrequent vorkommen, bei denen es sich allerdings nicht um Neubildungen handelt.

In diesen Fällen würde man sie aber dennoch zur Summe der Hapax Legomena hinzuzählen.² Demnach kann die Stichprobengröße Einfluss auf die Ergebnisse nehmen. Deshalb ist *P* nur dann als sinnvolle Methode zu bezeichnen, wenn der Input, die Tokenanzahl, für alle Suffixe identisch ist.

Als drittes Problem muss diskutiert werden, **welchen Einfluss Zufallsschwankungen auf die Ergebnisse haben können**. Mit Zufallsschwankungen sind in diesem Zusammenhang zwei Dinge gemeint:

- Dass es sich bei der Produktivität von Derivationsaffixen um eine Eigenschaft handelt, die sich in der Regel zwar nach einem bestimmten Schema entwickelt und verändert, aber dennoch Schwankungen möglich sind, die nicht auf eine gravierende 'Phasen-Veränderung' zurückzuführen sind und,
- dass es unwahrscheinlich ist, dass man beim Extrahieren von zwei Stichproben (trotz identischer Rahmenbedingungen) exakt dieselben Verteilungen vorfinden wird.

Die Phasen der Grammatikalisierung und der damit verbundenen Unidirektionalität wurden in Kapitel 3 erläutert: Im Zuge dieses Prozesses kommt es zunächst zu einem Anstieg an Produktivität, indem das Affix an Abstraktheit gewinnt. Danach sollte es sich in der Regel in einer 'Hoch-Phase' befinden, indem es rege zur Wortbildung beiträgt. Am Ende jedoch ist es möglich, dass die Produktivität absinkt und nur noch ein 'morphologischer Rest' übrig bleibt.

Dazu soll nachfolgend Abb. 8.3 betrachtet werden. Die x-Achse, die in diesem Fall den Zeitfaktor darstellt, soll als schematische Darstellung verstanden werden, denn konkrete Zeiträume in Zahlen können nicht angegeben werden, weil die Entwicklun-

²Dazu siehe auch Baayen (2008, 222): „If we read through a text or corpus, and at regular intervals keep note of how many different types we have encountered, we find that, unsurprisingly, the number of types increases first rapidly, and then more and more slowly.“

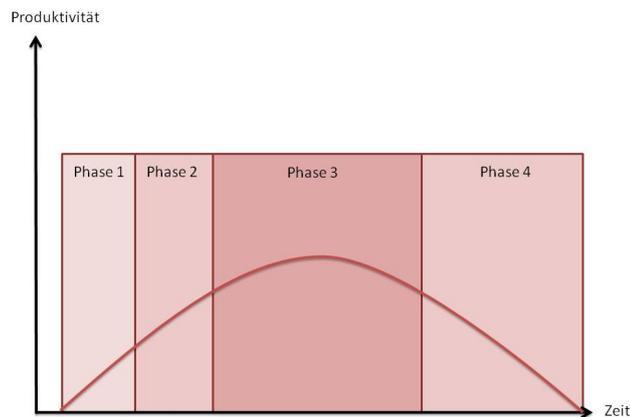


Abbildung 8.3: Beispiel für eine schematische Entwicklung der morphologischen Produktivität eines Affixes (begonnen mit seiner Entstehung als freies Lexem)

gen von Affixen – beginnend mit seiner Grammatikalisierung – unterschiedlich schnell ablaufen. In den meisten Fällen vollzieht sie sich über Jahrhunderte hinweg.³

In Phase 1 gewinnt ein freies Lexem an Bindungsfreudigkeit, indem es sich gerne und mit verschiedenen anderen freien Lexemen bindet und Komposita bildet. In Phase 2 kommt es zu einem Verlust des semantischen Inhalts, das Morphem findet sich in vielen neuen Gebrauchskontexten wieder und wird zunehmend abstrakter; es gewinnt damit an Produktivität. Es entstehen neue Wörter, jedoch mit hoher Wahrscheinlichkeit auch Ad-Hoc-Bildungen, bei welchen nicht klar ist, ob sie den Weg in das Lexikon der Sprecher finden werden. Beispielfhaft kann an dieser Stelle auf das deutsche Affixoid *Höll-en-/höll-en-* verwiesen werden. Gerade in Registern wie z.B. der Jugendsprache handelt es sich um einen beliebten Baustein zur Wortbildung. Begriffe wie *Höllenvirus*⁴ oder *Höllenanruf*⁵ sind akzeptabel und für andere Sprachbenutzer verständlich, obwohl sie derzeit nicht in einem deutschen Lexikon aufgelistet werden.

³In diesem Zusammenhang wird auf das in der linguistischen Literatur verwendete Zwiebelmodell verwiesen (Nübling et al. 2008, 2; Debus 1980, 188), das bildlich darstellt, dass die Morphologie neben der Phonologie und der Syntax zu den Kernbereichen gehört, was eine geringe Anfälligkeit für außersprachliche Einflüsse impliziert.

⁴Gefunden in einem Internet-Forum: *sehr enttäuschend con ihr hört man ja nix mehr...wahrscheinlich hat auch sie dein höllenvirus oder die pffife gegen sie in der liveshow haben sie gezähmt*, abgerufen am 04.03.2011 um 11:46 Uhr, <http://bbfun.de/modules.php?name=Forums&file=viewtopic&p=1143669&sid=5cab21c8aaeb75d77cdeed2957f4c3ed>.

⁵Gefunden auf <http://www.hirntumor.de/forum/index.php?topic=5560.60> (04.03.2011, 11:49 Uhr): *Eine milchige Flüssigkeit, oh nein.. 4 Tage später der Höllenanruf aus der Klinik: Bitte sofort zur Aufnahme kommen, wir müssen Sie morgen operieren.*

Ist der Affix-Status erreicht (Phase 3), so ist davon auszugehen, dass das Morphem einen gewissen Grad an Produktivität erlangt hat⁶ und zur Bildung neuer Lexeme beiträgt. In dieser Phase müsste es nach Baayens Maßen zu einer im Vergleich zu der Tokenanzahl hohen Anzahl an Hapaxen kommen, die dann wiederum zu einem hohen Wert *P* führt.

In der letzten Phase seiner Entwicklung verliert das Affix an Produktivität. Bemerkbar macht sich dies an der Überlegenheit konkurrierender Modelle und an der Tatsache, dass zunächst nur noch wenige, zu einem späteren Zeitpunkt keine Neologismen produziert werden. In Bezug auf die Anzahl der lexikalisierten Typen jedoch ist Phase 4 diejenige Phase, in der davon auszugehen ist, dass sie eine hohe Anzahl an Typen hervorbringt.

Da es sich beim obigen Schema um eine prototypische und **schematische Kurve** handelt, die nur eine von vielen möglichen Entwicklungsgeschichten eines Affixes nachzeichnet, sollte nun überlegt werden, was dies für die Messung von Produktivität bedeutet.

In der vorliegenden Arbeit standen die Daten des DWDS zur Verfügung, mit welchen ein Zeitraum von 100 Jahren abgedeckt wird. Gesetzt den Fall, es gäbe ein Suffix, das sich von 1900 bis 1999 in seiner vierten Entwicklungsphase befindet, so würde man mithilfe der Daten des DWDS empirisch belegen können, dass ein Abfall an Produktivität stattfindet. Ebenso gilt dies für die anderen Phasen, insbesondere die Phase der Entwicklung vom Affixoiden zum Affix, in welcher eine Zunahme an Produktivität erfolgen müsste.

Vergessen werden darf nicht, dass durchaus Produktivitätsschwankungen vorhanden sein können, ohne dass ein Phasenumbruch vollzogen wird; dazu soll Abb. 8.4 betrachtet werden. Die Punkte in der Vergrößerung repräsentieren die Stichproben, die Korpusdaten, die entnommen werden können.

Abb. 8.4 zeigt einen stark vergrößerten Ausschnitt der zuvor schematischen dargestellten Produktivitätskurve. Aus dieser Abbildung geht hervor, dass zu erwarten ist, dass zwei zeitlich nah beieinander liegende Stichproben zu minimal unterschiedlichen Ergebnissen der morphologischen Produktivität führen.

⁶Dabei sei an dieser Stelle angemerkt, dass nicht davon auszugehen ist, dass jedes Affix hoch produktiv ist, sondern dass es ebenso möglich ist, dass ein Suffix nur mäßig produktiv zur Wortbildung beiträgt. Dennoch, und das ist der entscheidende Punkt, ist es solange als produktiv zu bezeichnen, solange neue Typen produziert werden können.

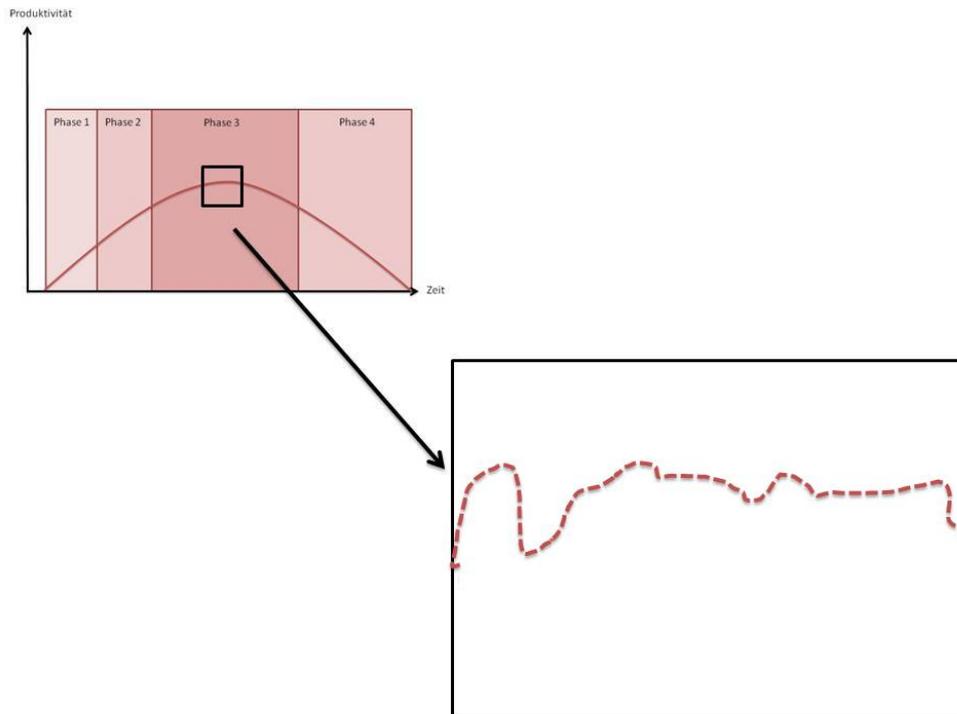


Abbildung 8.4: Zufallsschwankungen in der Produktivität eines Affixes

Des Weiteren ist es allerdings auch möglich, sogar wahrscheinlich, dass zwei Stichproben aus demselben Korpus und gleichzeitig demselben Zeitabschnitt extrahiert werden und sich die Anzahl der Typen und Hapaxe (und damit auch die Werte von P) trotz identischer Tokenanzahl voneinander unterscheiden.

Die korpuslinguistische Arbeit erlaubt es, feingradige statistische Unterschiede festzuhalten und zu interpretieren. Auch kleine statistische Verschiebungen können letztlich auf einen beginnenden Sprachwandel-Prozess im System hinweisen.

Selbstverständlich wäre es theoretisch ebenso möglich, (vollständige) Grammatikalisierungsprozesse durch Korpusstudien empirisch zu belegen und damit in Zahlen auszudrücken, doch fehlen derzeit noch die dafür notwendigen diachronen Korpora des Deutschen.

Zwei der drei skizzierten Problematiken, nämlich die Stichprobengröße und die Zufallsschwankungen, sollen im weiteren Verlauf dadurch gelöst werden, dass ein LNRE-Modell (*large number of rare events*) seine Anwendung findet. Dies stellt eine Novität in Bezug auf deutsche Wortbildungsmodelle dar.

8.1 LNRE-Modelle

In vielen vorangegangenen Abschnitten dieser Arbeit wurde die Rolle der Häufigkeiten von Token und Typen ausgiebig diskutiert. Auch bei den LNRE-Modellen spielen sie eine entscheidende Rolle, denn mit LNRE-Modellen (Baayen 2001, Kap. 3) werden in der quantitativen Linguistik Häufigkeitsverteilungen von Wörtern untersucht.

Mit diesen Modellen lassen sich, wie im Folgenden der Arbeit noch deutlich werden wird, Aussagen von beobachteten Daten auf nicht beobachtete Daten treffen. Dies ist insbesondere dann von großer Bedeutung, wenn wie im Fall dieser Korpusstudie, ungleiche Datenmengen zur Verfügung stehen. Den LNRE-Modellen liegt die Annahme der Zipf'schen Verteilung zugrunde.

Doch bevor im Detail auf das Zipf'sche Gesetz und eine Modifikation desselben (Zipf-Mandelbrot) eingegangen wird, sollen noch einmal Überlegungen zu den Häufigkeiten von Token und Typen im Vordergrund stehen.

Um die Produktivität nach Baayens Index (P) zu berechnen, wurden Tokenlisten aus dem DWDS-Korpus extrahiert, die in Typenlisten umgewandelt wurden. Um P zu berechnen, musste auch die jeweilige Anzahl der Hapax Legomena ermittelt werden. Doch gleichwohl stehen dem Linguisten auch weitere Informationen zur Verfügung: zum Beispiel die Anzahl der Wörter, die zweimal, dreimal, viermal oder z.B. exakt zehnmal im Korpus vorkommen.

Im Folgenden soll anhand von Tab. 8.1 ein einfaches Beispiel zur Frequenzverteilung herangezogen werden, in welcher f für Frequenz steht. Tabelle 8.1 zeigt eine

<i>Type</i>	<i>f</i>	<i>Type</i>	<i>f</i>
Ergebnis	2	Düsternis	1
Gläubnis	3	Besorgnis	1
Erkenntnis	1	Befugnis	7
Bildnis	2	Ersparnis	3
Zeugnis	1	Besäufnis	1
Kenntnis	2	Ärgernis	1
Finsternis	1		

Tabelle 8.1: Ein fiktives Frequenzspektrum für *-nis*-Derivate

mögliche Häufigkeitsverteilung von *-nis*-Derivaten auf. Die Daten der Tabelle lassen sich nach Rängen ordnen:

<i>Rang</i>	<i>f</i>	<i>Rang</i>	<i>f</i>
1	7	8	1
2	3	9	1
3	3	10	1
4	2	11	1
5	2	12	1
6	2	13	1
7	1		

Tabelle 8.2: Anordnung der *-nis*-Derivate nach Rank/Frequenz

Eine Frequenzliste zeigt an, wie viele Typen mit derselben Häufigkeit in einem Korpus vorkommen.

<i>f</i>	<i>V(f)</i>
1	7
2	3
3	2
7	1

Tabelle 8.3: Frequenzspektrum der *-nis*-Derivate

Zipf (1949) gilt der erster Wissenschaftler, der Häufigkeitsverteilungen von Wörtern in Texten beobachtete. Er fand heraus, dass sich die Schiefe der Verteilung durch eine nicht-lineare Funktion beschreiben lässt, die in Bezug auf die hohen Ränge steiler abfällt als bei den tiefen Rängen.

In einer mathematischen Formel ausgedrückt sieht das Zipf'sche Gesetz folgendermaßen aus:

$$f(w) = \frac{C}{r(w)^a} \quad (8.1)$$

$f(w)$ steht für Frequenz, $r(w)$ für Rang; C und a stellen Konstanten dar, die anhand des Korpus ermittelt werden können.

Zur Illustration wird das in Baroni (2009) aufgeführte Beispiel herangezogen: wird für a 1 angenommen, dann wird dem am häufigsten im Korpus auftretende Wort Rang

1 und die Frequenz C zugewiesen. Angenommen das am häufigste auftretende Wort ist 60.000 Mal vertreten, so gilt $C = 60.000$.

In Baroni's Beispiel wird dem zweithäufigsten Wort die Frequenz $C/2 = 30.000$, dem dritthäufigsten Wort $C/3 = 20.000$ zugewiesen. Das 100. am häufigsten auftretende Wort erhält $C/100 = 600$, das 101. am häufigsten auftretende Wort erhält $C/101 = 594.06$. Dies entspricht 99% von $C/100$. Für $C/102$ würde man 588.23 berechnen, was 98% der Häufigkeit von $C/100$ entspricht.

Dazu erklärt Baroni (2009, 813) Folgendes:

„Thus, the model predicts a very rapid decrease in frequency among the most frequent words, which becomes slower as the rank grows, leaving very long tails of words with similar low frequencies. Zipf's law does not predict frequency ties, since there are no ties among ranks, but it approximates the empirically attested plateaus by predicting a very large number of words with very similar non-integer frequencies. For example, the model above with a set to 1 and C set to 60,000 predicts that about 80,000 words will have frequencies between 1.5 and 0.5.“

Im Vergleich zu anderen mathematischen Modellen wie zum Beispiel der exponentiellen Verteilung fällt eine Potenzfunktion mit negativem Exponenten, zu denen die Zipf'sche Formel gehört, sehr langsam ab und hinterlässt einen langen „Schweif“ ('long tail') an niedrig frequenten Ereignissen.

Das Zipf'sche Gesetz hat die Eigenschaft, dass bei Logarithmierung beider Seiten eine lineare Funktion entsteht:

$$\log f(w) = \log C - a \log r(w) \quad (8.2)$$

Baroni (2009, 815), der die Verteilung von reellen Daten großer Korpora im Hinblick auf die von Zipf angenommene Verteilung testet, findet heraus, dass sich die Verteilungen der Häufigkeiten zum Beispiel im BNC-Korpus nicht exakt mit der Zipf'schen Verteilung beschreiben lässt:

„More worryingly, for the two largest corpora (BNC and la republica) we observe a curvature suggesting that frequency is dropping more rapidly than what is predicted by Zipf's law. [...] At the other end of the plot (low ranks, high frequencies) we observe, again, a downward curvature of the

empirical profile, i.e., the attested high frequencies tend to be lower than what would be predicted by their rank according to Zipf's law.“

Diese Beobachtung machte Mandelbrot (1953) bereits Mitte des 20. Jahrhunderts, er leitete eine Formel ab, indem er einen weiteren Parameter, b , hinzufügte:

$$f(w) = \frac{C}{(r(w) + b)^a} \quad (8.3)$$

In Bezug auf das von Baroni aufgeführte Beispiel würde man im Falle von $b = 1$ bei Anwendung des Zipf-Mandelbrot-Gesetzes für das am häufigsten vertretende Wort nicht 60.000, sondern $60.000/(1+1) = 30.000$ erhalten. Bei Betrachtung des am 1000. häufigste Wort fällt auf, dass die Differenz der Werte nach Zipf und Zipf-Mandelbrot nur minimal sind: 60 nach Zipf und 59.94 nach Zipf-Mandelbrot. Baroni (2009) stellt mithilfe von Untersuchungen fest, dass gerade die niedrigen Ränge eine wesentlich bessere Anpassung an das durch Mandelbrot modifizierte Gesetz erlangen.

Bezüglich Frequenz und Rang lassen sich für diese Arbeit die folgenden Thesen aufstellen:

1. Ein Affix, das produktiv ist, wird viele Wörter produzieren lassen, die statistisch gesehen in kleiner Häufigkeit vorkommen, und wenige, die in großer Anzahl vorhanden sind (Abb. 8.5 a).
2. Ein unproduktives Affix, das nicht mehr zur Wortbildung verwendet wird, wird sich, wenn überhaupt, nur in stark lexikalisierten und damit hoch frequenten Bildungen wiederfinden (Abb. 8.5 b)⁷.

Die Schiefe eines Frequenzspektrums kann zur Interpretation der morphologischen Produktivität hinzugezogen werden, Lüdeling und Evert (2005, 362) treffen dazu die folgende Aussage: „The stronger the skew, the more productive the process is.“

Daraus lässt sich ableiten, dass ein Frequenzspektrum, welches relativ flach verläuft, auf eine niedrige Produktivität hinweist.

Mithilfe des `zipfR`-Paketes in R ist es möglich, unterschiedliche LNRE-Modelle anzuwenden: ZM (*Zipf Mandelbrot*), fZM (*finite Zipf Mandelbrot*) und GIGP (*Generalized Inverse Gauss-Poisson*).

⁷Es sei an dieser Stelle darauf hingewiesen, dass es sich sowohl bei a) wie auch bei b) um schematische Darstellungen handelt und auch andere Verteilungen auf produktive und unproduktive Prozesse hindeuten können. Diese Verteilungen könnten durchaus noch deutlich schiefer angeordnet sein.

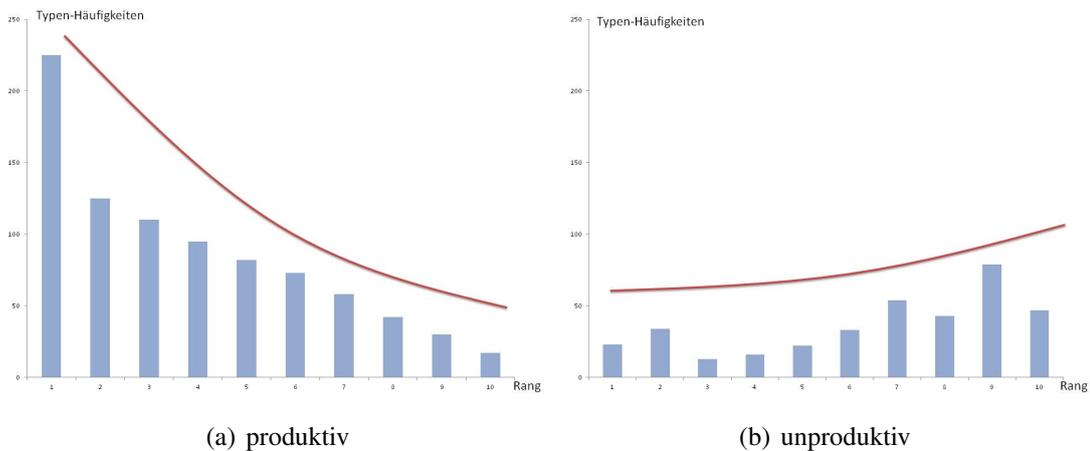


Abbildung 8.5: Prototypische Frequenzspektren und Skizzierungen der jeweiligen Schiefen der Typenverteilungen (rote Linien)

Zur genauen mathematischen Erläuterung des Zipf-Mandelbrot-Modell⁸ siehe Evert (2005, 4.2.2) und für das finite Zipf-Mandelbrot-Modell (fZM) siehe Evert (2005, 4.2.3). Zur genauen Betrachtung des GIGP-Modells siehe Baayen (1992, 89ff).

In der vorliegenden Arbeit wurde das finite Zipf-Mandelbrot-Modell ausgewählt, um die Derivationsuffixe exemplarisch zu untersuchen, denn während bei dem Zipf-Mandelbrot-Modell von einem unendlichen Typenreichtum auszugehen ist, sehen das fZM-Modell (und auch das GIGP-Modell) die mögliche Typenanzahl, die durch ein Wortbildungsmodell generiert werden kann, als endlich an, was eine genaue Berechnung des Typenreichtums S ermöglicht. Und dieser Parameter könnte bei der Interpretation der morphologischen Produktivität durchaus herangezogen werden, denn nur mit produktiven Suffixen lassen eine Vielzahl an Typen bilden, sodass numerische Unterschiede sichtbar gemacht werden können.

Arbeitet man mit dem zipfR-Paket, das in R verfügbar ist, so muss zunächst eine Datei, in unserem Fall, eine Datenreihe eines Suffixes, eingelesen werden. Exemplarisch wird an dieser Stelle die Datenreihe des Suffixes *-ung* für das Jahr 1999 genommen. Nach Einlesen der Datei wird eine Auflistung, aus welcher sich sowohl die Ränge

⁸Um die Ranghäufigkeitsverteilung für Wörter des Deutschen zu erstellen, wurde das Zipf-Mandelbrot-Gesetz bereits an einigen Stellen angewandt. So z.B. kommt Knüppel (2001) anhand ihrer Datengrundlage von 20 Briefen des Zeitraumes 1811 bis 1983 zu dem Schluss, dass sich die Frequenzen von Wörtern bestimmter Länge (einsilbig, zweisilbig, etc.) bestens mit der Zipf-Mandelbrot-Verteilung beschreiben lassen. Ebenso wird dies anhand von Altmanns Untersuchung (1988) auf Basis von Goethes *Erkönig* und Bests Studie (2001) zu Lichtenbergs *Sudelbücher* bestätigt.

(hier: k) wie auch die Frequenzen (f) der Wortbildungen ablesen lassen. Zudem wird – am Schluss der Liste – die Gesamttokenzahl (N) und die Anzahl der Typen (V) präsentiert:

	k	f	type
1	1	343	Werbung
2	2	137	Stellung
3	3	108	Entwicklung
4	4	104	Bewegung
5	5	88	Lösung
6	6	85	Regierung
7	7	73	Leistung
8	8	72	Führung
9	9	69	Bedingung
10	10	67	Bevölkerung
11	11	66	Ordnung
12	12	64	Richtung
13	13	64	Forderung
14	14	63	Beziehung
15	15	63	Wendung
16	16	62	Wirkung
17	17	54	Stimmung
18	18	50	Zeitung
19	19	48	Haltung
20	20	47	Verbindung

...

N	V
4911	691

Anschließend lässt sich automatisch ein Frequenzspektrum generieren (Abb. 8.6).

Im nächsten Schritt lassen sich nun weitere Berechnungen durchführen: In Frage kommen zum Beispiel **binomiale Interpolationen** (vgl. Baayen 2001, Kap.2), bei denen Erwartungswerte (z.B. $E(V)$ für die Anzahl der Typen, $E(V1)$ für die Anzahl der Hapax Legomena) anhand der tatsächlich zur Verfügung stehenden Daten mithilfe

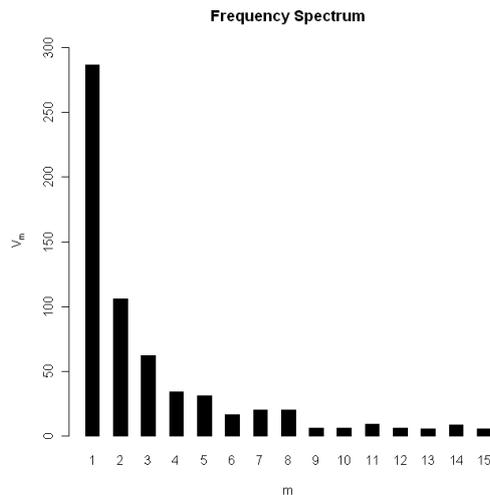


Abbildung 8.6: Frequenzspektrum auf Basis der Datenreihe des Suffixes *-ung* im DWDS im Jahre 1999

einer Funktion, die man auch Interpolante nennt, berechnet werden.⁹ Die Anpassungsgüte (engl. *goodness of fit*) gibt an, wie gut ein statistisches Modell zur Erklärungen von Beobachtungen herangezogen werden kann.

Des Weiteren ist es möglich, **Extrapolationen** durchzuführen. Damit ist gemeint, dass die Anzahl der Typen und der Hapax Legomena anhand des Modells vorausgesagt werden. Es lassen sich so letztendlich Datenmengen unterschiedlicher Größe miteinander vergleichen: Z.B. lässt sich exakt berechnen, wie viele Typen oder auch Hapax Legomena ein Wortbildungsmodell generieren würde, sofern 10.000 oder 100.000 oder 1 Million Token zur Verfügung stünden; vereinfacht gesagt handelt es sich demnach um eine Hochrechnung.

Die Eingabe

```
ung1999.zm <- lnre("zm", ungl1999.spc)
ung1999.zm
```

führt zu folgendem Ergebnis:

```
Zipf-Mandelbrot LNRE model.
Parameters:
```

⁹Siehe Baroni und Evert (2005, 2): „A (parametric) LNRE model is then simply a formula that describes the distribution of occurrence probabilities in the population.“

```
Shape:          alpha = 0.4410353
Upper cutoff:   B = 0.02046168
[ Normalization: C = 4.91483 ]
Population size: S = Inf
Sampling method: Poisson, with exact calculations.

Parameters estimated from sample of size N = 4911:
          V      V1      V2      V3      V4      V5
Observed: 691 287.00 106.00 62.00 34.00 31.00 ...
Expected: 691 332.07  92.81 48.23 30.85 21.96 ...

Goodness-of-fit (multivariate chi-squared test):
          X2 df          p
49.93609 11 6.427192e-07
>
```

Jedes LNRE-Modell besitzt Parameter, von welchen bei dem ZM- und dem fZM-Modell Alpha für die Aussagen zur Produktivität wichtig sein könnte, denn Alpha gibt an, wie schief das jeweilige Frequenzspektrum verteilt, sprich, wie ausgeprägt die Zipf-Eigenschaft der Verteilung ist¹⁰:

„The overall shape of the frequency spectrum predicted by the fZM model is mainly determined by the model parameter alpha. Its values range from alpha = 0 (indicating balanced spectrum, where the number of hapax legomena is not much larger than the number of types in higher frequency ranks) to alpha = 1 (indicating a highly skewed spectrum that is entirely dominated by the hapax legomena). We can thus tentatively use alpha as a quantitative measure for the degree of productivity.“¹¹

Je größer der Betrag von Alpha ist, desto mehr 'Wahrscheinlichkeitsmasse' entfällt auf seltene Typen – anschaulich gesprochen: desto häufiger taucht ein seltenes, mutmaßlich produktiv gebildetes Wort auf.

¹⁰a = 1/α.

¹¹Lüdeling und Evert (2005, 362). Das Symbol des griechischen Buchstabens Alpha wurde in diesem Zitat durch das Wort 'alpha' ersetzt.

Damit beschreibt Alpha nahe bei 0 die am wenigsten produktive Verteilung, die sich mit diesem Modell darstellen lässt; ein Wert nahe 1 eine sehr produktive Verteilung, die dem ursprünglich von Zipf formulierten Gesetz entspricht.

Bisher wurde in der germanistischen Linguistik kein Versuch gewagt, den Parameter Alpha zur Produktivitätsanalyse hinzuzuziehen.

Im Anschluss an die Überprüfung der Alpha-Werte bietet es sich an, Wachstumskurve für Typen und Hapax Legomena sichtbar zu machen, um sie dann zur qualitativen Betrachtung heranzuziehen.

In der Annahme, es stünde eine Datenmenge von 10.000 Token zur Verfügung, sähen die Wachstumskurven der Hapax Legomena und der Typen wie in Abb. 8.7 skizziert aus. Die gestrichelte Linie in der Mitte der Grafik zeigt an, wie viele Token tatsächlich als Input eingelesen worden sind.

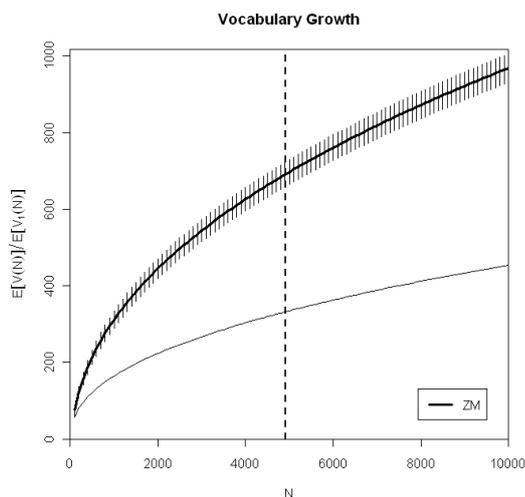


Abbildung 8.7: Typen- und Hapaxwachstum (Extrapolation) für das Suffix *-ung*

Dieselben Daten lassen sich auch mit dem finiten Zipf-Mandelbrot-Modell analysieren; auch in diesem Fall wird man mit dem Einlesen der Daten begonnen und im Anschluss daran dann das Frequenzspektrum erzeugt. Ebenso besteht auch bei diesem Modell die Möglichkeit, die Parameter des Modells und die Anpassungsgüte betrachten zu können:

finite Zipf-Mandelbrot LNRE model.

Parameters:

```
Shape:          alpha = 0.5458378
Lower cutoff:   A = 3.358647e-05
Upper cutoff:   B = 0.02243561
[ Normalization: C = 2.687845 ]
```

Population size: $S = 1323.352$

Sampling method: Poisson, with exact calculations.

Parameters estimated from sample of size $N = 4911$:

```
          V      V1      V2      V3      V4      V5
Observed: 691 287.00 106.00 62.00 34.00 31.0 ...
Expected: 691 285.45 116.83 59.49 36.62 25.3 ...
```

Goodness-of-fit (multivariate chi-squared test):

```
          X2 df          p
14.98953 11 0.1829769
```

>

Neben dem Parameter *Alpha* steht bei diesem Modell zudem die Populationsgröße ('population size') S zur Interpretation der morphologischen Produktivität zur Verfügung. Die Populationsgröße gibt an, wie viele Typen mit dem jeweiligen Suffix gebildet werden könnten.

Für das oben genannte Beispiel wird berechnet, dass 1323 Typen gebildet werden können. Ohne an dieser Stelle genauer auf Details einzugehen, kann angemerkt werden, dass ein Suffix, mit welchem sich über 1300 unterschiedliche Wörter bilden lassen, entweder einmal produktiv war oder noch immer ist.

Die Extrapolationen der Wachstumskurven nach dem fZM-Modell sehen folgendermaßen aus:

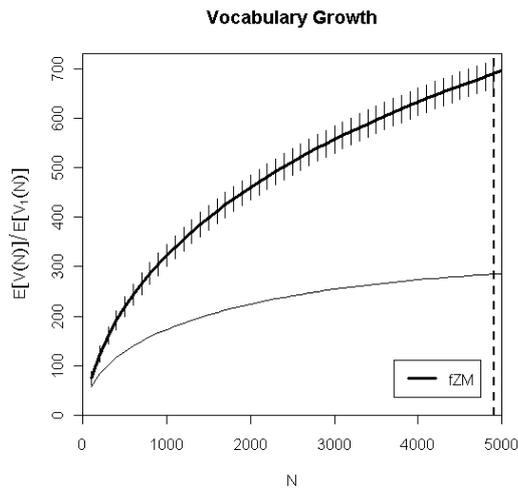


Abbildung 8.8: Typen- und Hapaxwachstum (Extrapolation) für das Suffix *-ung* nach dem finiten Zipf-Mandelbrot-Modell

Die obere Kurve skizziert das Wachstum der Typen und die untere die der Hapax Legomena – jeweils in der Annahme, dass 5000 Token zur Verfügung stünden.

Das GIGP-Modell darf als differenzierte Implementierung des Zipf-Mandelbrot-Gesetzes bezeichnet werden, es handelt sich um eine mathematisch komplexere Variante des fZM-Modells (Baayen 2001, 89ff). Die Ergebnisse nach dem GIGP-Modell für die bereits mit den anderen Modellen betrachtete Datenreihe lauten wie folgt:

```
> ungl1999.gigp <- lnre("gigp", ungl1999.spc, m.max=10)
> ungl1999.gigp
```

Generalized Inverse Gauss-Poisson (GIGP) LNRE model.

Parameters:

```
Shape:          gamma = -0.4264098
Lower decay:    B = 0.05129203
Upper decay:    C = 0.01561816
[ Zipf size:    Z = 64.02803 ]
```

Population size: S = 1862.559

Sampling method: Poisson, with exact calculations.

Parameters estimated from sample of size N = 4911:

```
V      V1      V2      V3      V4      V5
```

```
Observed: 691.00 287.00 106.00 62.00 34.00 31.00 ...
Expected: 694.34 286.34 110.15 59.41 38.19 27.09 ...
```

Goodness-of-fit (multivariate chi-squared test):

```
      X2 df      p
10.96166 11 0.4464809
```

>

Im GIGP-Modell wird der Parameter *Gamma* verwendet. Auch dieser beschreibt die Schiefe der Verteilung, sein Betrag dürfte zur Interpretation der Daten verwendet werden.

8.1.1 Vergleich des Zipf-Mandelbrot-, des finiten Zipf-Mandelbrot- und des GIGP-Modells

Abschließend sollen alle drei Modelle vergleichend in einem Diagramm gegenüber gestellt werden. Dazu kann Abb. 8.9 betrachtet werden.

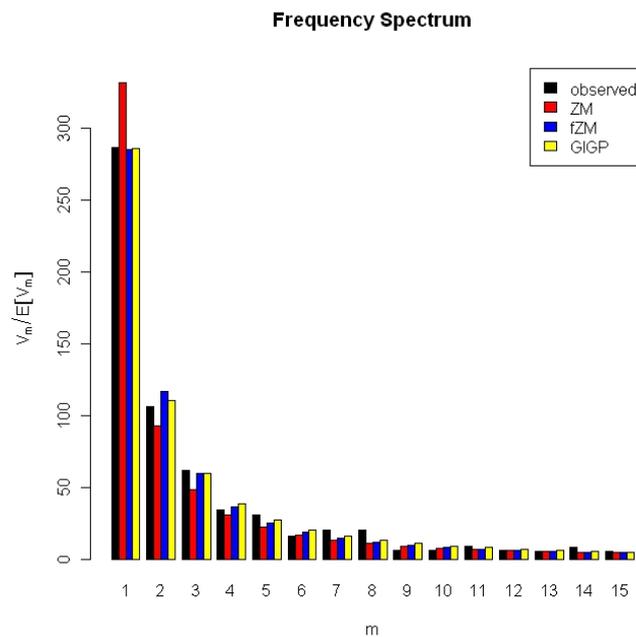


Abbildung 8.9: Vergleich der drei LNRE-Modelle ZM, fZM und GIGP

Die Entscheidung, im Folgenden das fZM-Modell anzuwenden, kann darin begründet werden, dass das fZM-Modell einfacher und numerisch robuster ist als das GIGP-Modell (vgl. Baroni und Evert 2007). Auch Abb. 8.9 macht deutlich, dass das ZM-Modell das tatsächliche Frequenzspektrum nicht gut wiedergeben kann und eine schlechte Anpassung vorliegt; die Ergebnisse des finiten Zipf-Mandelbrot-Modells und des GIGP-Modells hingegen ähneln sich. Es wurde entschieden, das robuste finite Zipf-Mandelbrot-Modell anzuwenden.

Die Entscheidung gegen das Zipf-Mandelbrot-Modell und für das fZM-Modell kann darin begründet werden, dass bei dem ZM-Modell eine unendliche Populationsgröße angenommen wird, aber gerade die Vorhersage, wie viele Typen mit einem Wortbildungsmodell gebildet werden könnten, für diese Arbeit besonders interessant sein könnte.

Damit eröffnen die LNRE-Modelle neue Zugänge zur Produktivitätsforschung. In der vorliegenden Arbeit wurde entschieden, verschiedene Parameter und Methodiken zu testen, die mithilfe des fZM-Modells möglich gemacht werden:

- P , Typen (V) und Hapaxe (V_1) anhand einer **normierten Stichprobengröße** zu betrachten,
- **Konfidenzintervalle** zur Hilfe zu nehmen, um Zufallsschwankungen zu überprüfen,
- die **vorausgesagte Anzahl der Typen** (S) zu betrachten,
- den Parameter Alpha, der die **Schiefe des Frequenzspektrums** angibt, zu testen und
- **Extrapolationen** von Wachstumskurven für Typen und Hapax Legomena (qualitativ) zu betrachten.

Die Berechnungen werden vorzugsweise und exemplarisch anhand der Datenreihen des Jahres 1999 gemacht, sodass hauptsächlich synchron vorgegangen wird. Zwar ist eine diachrone Betrachtungsweise mit den Modellen theoretisch möglich, doch soll in den nächsten Abschnitten nicht der Produktivitätswandel im Vordergrund stehen, sondern a) das Testen des fZM-Modells als (ergänzende) Methode zur Produktivitätsbetrachtung und b) die Überprüfung der Aussagen zur Produktivität der Suffixe (Tab. 5.1).

8.2 Normierung der Stichprobengröße

Um vergleichbare Aussagen über die Produktivität der Derivationssuffixe machen zu können, erfolgt eine Extrapolation auf 5000 Token.

Anhand der normierten Tokenmengen ergeben sich auch neue Typenfrequenzen, veränderte Frequenzen der jeweiligen Hapax Legomena und selbstverständlich können auch die P-Werte modifiziert werden. Von welcher tatsächlichen Stichprobengröße aus jeweils extrapoliert worden ist, kann Tab. 7.1 entnommen werden.

Mithilfe von des finiten Zipf-Modells ist zudem eine Erstellung von Konfidenzintervallen möglich; diese erlauben es, Aussagen anhand von **Wertebereichen** (V_{1min} bis V_{1max} , P_{min} bis P_{max} , V_{min} bis V_{max}) durchzuführen.

Konfidenzintervalle werden eingesetzt, um von verfügbaren Stichproben auf eine Grundgesamtheit schließen zu können, denn mithilfe von Konfidenzintervallen ist es möglich, Wertebereiche anzugeben, innerhalb derer sich die Werte in der Grundgesamtheit wahrscheinlich bewegen. Intuitiv beschreiben die Konfidenzintervalle, wie stark der beobachtete Wert von z.B. P zwischen verschiedenen Stichproben schwankt, sofern sie aus exakt derselben Grundgesamtheit kommen (gleiches Zeitfenster, gleiche Korpuszusammenstellung, usw.).

Somit ist das Konfidenzintervall ein Maß für die Unsicherheit bei der Schätzung der tatsächlichen Werte von P (oder einer anderen Größe) aus der Stichprobe.

Im Folgenden wird das Suffix *-sal* nicht weiter untersucht wird, da bereits die vorangegangenen Studien gezeigt haben, dass es sich um ein gänzlich unproduktives Suffix handelt.

Für die anderen substantivbildenden Suffixe ergeben sich nun – nach Extrapolation – die in Tab. 8.4 sichtbaren Werte.¹²

Tab. 8.4 präsentiert, welche Werte zu erwarten wären, wenn genau 5000 Token für jedes der Suffixe zur Verfügung stünden: Nach Analyse mithilfe des fZM-Modells erhält man für *-ung* ca. 700 Typen, was für eine hohe Produktivität spricht oder zumindest andeutet, dass das Suffix in früheren Zeiten produktiv gewesen ist. Derselbe Schluss lässt sich bei Betrachtung der Werte von *-keit* ziehen.

¹²Die Suffixe *-heit* und *-keit* werden im Übrigen nur separat betrachtet. In diesem Kapitel soll lediglich die Methodik des fZM-Modells im Vordergrund stehen. Wie an späterer Stelle der vorliegenden Arbeit deutlich werden wird, funktioniert das Modell noch nicht fehlerfrei, sodass lediglich exemplarische Ergebnisse und vorzugsweise methodische Ansätze diskutiert werden können.

Suffix	V	V1	P
<i>-ung</i>	696	286	0.057
<i>-tum</i>	276	151	0.030
<i>-nis</i>	48	1	0
<i>-heit</i>	265	98	0.020
<i>-keit</i>	703	255	0.051
Suffix	95%-K.-intervalle für V	95%-K.-intervalle für V1	95%-K.-intervalle für P
<i>-ung</i>	668...724	258...315	0.052...0.063
<i>-tum</i>	254...298	129...173	0.026...0.035
<i>-nis</i>	47...49	0...3	0.000...0.001
<i>-heit</i>	248...281	81...114	0.016...0.023
<i>-keit</i>	679...727	229...280	0.046...0.056

Tabelle 8.4: Auffistung der Extrapolations-Ergebnisse (V, V1, P) der substantivbildenden Suffixe nach dem finiten Zipf-Mandelbrot-Modell anhand der Daten des Jahres 1999 im DWDS

Die Suffixe *-tum* und *-heit* verhalten sich ähnlich, da beiderseits vorausgesagt wird, dass jeweils etwa zwischen 250 und 300 Typen gebildet werden müssten; dies impliziert eine mäßige Produktivität.

Das auch den bisherigen Ergebnissen nach unproduktive *-nis* ist auch nach der Analyse mit dem fZM-Modell unproduktiv: es werden lediglich 48 Typen vorhergesagt.

Auch die Betrachtung der Hapax Legomena bestätigt das eben skizzierte Bild und das Heranziehen der Häufigkeiten der jeweiligen Anzahl der Hapax Legomena kann als Indikator für eine hohe Produktivität geltend gemacht werden, auch wenn an vielen Stellen dieser Arbeit bereits diskutiert wurde, dass es sich nicht bei jedem Hapax Legomenon um ein neu gebildetes Wort handeln muss.

Die Interpretation der Anzahl der Hapax Legomena in Tab. 8.4 lässt den Schluss zu, dass es sich bei *-ung* und *-keit* um sehr produktive Suffixe, bei *-tum* und *-heit* um mäßig produktive und bei *-nis* um ein unproduktives Suffix handelt. Erwähnenswert ist in diesem Zusammenhang, dass für *-tum* zwischen 129 und 173 Hapaxe vorhergesagt werden, während die Werte für *-heit* zwischen 84 und 114 liegen, sodass trotz vergleichbarer Anzahl an Typen im 20. Jahrhundert davon auszugehen ist, dass *-tum* das produktivere Wortbildungsmodell ist. Für *-nis* wird angenommen, dass sich bei einer Tokenmenge von 5000 Textwörtern entweder keine,¹³ maximal aber drei Wörter

¹³Der an dieser Stelle eigentlich berechnete Wert ist -0.836, es handelt sich dabei nicht um einen Rechenfehler, sondern um asymptotische (d.h. angenäherte) Konfidenzintervalle. Vor allem bei kleinen

auffinden lassen, die einmalig auftreten, was selbstverständlich Unproduktivität impliziert.

Besonders wichtig jedoch ist die Betrachtung der jeweiligen *P*-Werte, denn nunmehr lassen sich diese Werte ohne Zweifel miteinander vergleichen. Das Bild, das sich nun abzeichnet, ist folgendes: Das produktivste substantivbildende Suffix ist *-ung*. Etwa vergleichbar produktiv ist *-keit*. Nur mäßig produktiv erscheinen *-tum* und *-keit*, *-nis* darf als gänzlich unproduktiv beschrieben werden.

Hier wird deutlich, dass die LNRE einen neuen Zugang ermöglichen, um Produktivität zu beschreiben: Während die Zusammenfassung der Ergebnisse in Kapitel 7.3.1 deutlich machte, dass *-tum* als Problemfall gilt, da die Grammatiken das Suffix als schwach/gering produktiv etikettieren, konnte diese Annahme nicht ohne Weiteres durch die Korpusstudie bestätigt werden, vielmehr schien das Suffix produktiv zu sein, auch wenn erkannt wurde, dass die in Kapitel 7 betrachteten Ergebnisse nicht kritiklos zu betrachten sind, da nur eine geringe Menge an Token zur Verfügung stand – für *-tum* im Jahre 1999 genau 334.

Die geringe Datenmenge führte zu einem hohen *P*-Wert (Tab. 7.11), bei welchem nun – nach Normierung der Stichprobengröße – deutlich wird, dass der Wert irreführend ist.

Das an dieser Stelle genannte Beispiel zeigt demnach explizit, was anhand von Abb. 8.2 erläutert worden ist, nämlich, dass die Wachstumskurve sowohl bei produktiven wie auch unproduktiven Prozessen zunächst sehr steil verläuft und eine große Datenmenge notwendig ist, valide Aussagen zu treffen.

Die Aussage, die hier getroffen werden kann, ist, dass die Etikettierung von *-tum* als beschränkt aktives Suffix durchaus plausibel erscheint, von schwacher oder geringer Produktivität jedoch lässt sich nicht sprechen.

Auch interessant ist das nun deutlich sichtbare Gefälle zwischen den Allomorphen *-heit* und *-keit*, die sich im Hinblick auf ihre Produktivität stark unterscheiden: Während *-keit* als hoch produktiv bezeichnet werden kann, scheint *-heit* nur schwach produktiv zu sein.

Erwartungswerten sind diese nicht sehr präzise. Die negativen Werte entstehen, wenn $V1$ gleich 0 oder $V1$ gleich 1 ist. Da diese Prozesse allerdings klar unproduktiv sind, braucht man an dieser Stelle aber keine genauen Konfidenzberechnungen.

Selbstverständlich kann anhand der Daten in Tab. 8.4 auch ein Diagramm zur Anschauung der globalen Produktivität herangezogen werden (Abb. 8.10).¹⁴

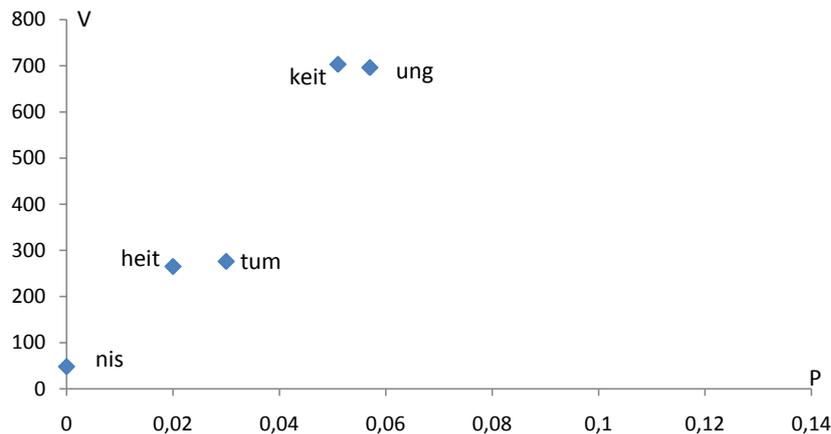


Abbildung 8.10: Globale Produktivität der substantivbildenden Suffixe im Jahre 1999 anhand der extrapolierten Daten durch das finite Zipf-Mandelbrot-Modell

Im Vergleich zu Abb. 7.9 zeigt sich nun teilweise etwas anderes Bild: lediglich das Suffix *-nis* erscheint – wie gewohnt – in gänzlicher Hinsicht unproduktiv, für *-heit* und *-keit* kann festgehalten werden, dass ihre Typenhäufigkeit höher, aber die Produktivität *P* niedriger ist als in Abb. 7.9, was zu einer 'Wanderung' nach 'links' und zugleich nach 'oben' führt. Dasselbe gilt auch für *-tum*. Für das Suffix *-ung* ergeben sich in Bezug auf Abb. 7.9 keine Veränderungen, weil mit *-ung* von Beginn an genügend Daten zur Verfügung standen.

Die Veränderungen, die sich anhand der Normierung ergeben, lassen sich Abb. 8.11 als Modifikation (von Abb. 7.9) entnommen werden. Die Bewegungen der Suffixe sind als Pfeile dargestellt: Während für *-nis* nur eine sehr kleine Bewegung nach links verzeichnet werden kann, sind die Modifikationen bei *-tum* besonders auffällig.

¹⁴Um das Schaubild, und insbesondere das daran anschließende Schaubild (Abb. 8.11), nicht zu überladen, wurde entschieden, die exakt ausgerechneten Werte als Koordinatenpaare in das Diagramm einzuzeichnen. Doch selbstverständlich könnte auch eine Darstellung mit Konfidenzstäben erfolgen in wie Abb. 8.14.

Die Länge der Pfeile zeigt das Ausmaß der Modifikation, was impliziert, dass für *-heit* und *-keit* – bei einer größeren Anzahl an Token – eine Veränderung der *P*-Werte und der globalen Produktivität *P** zu erwarten ist.

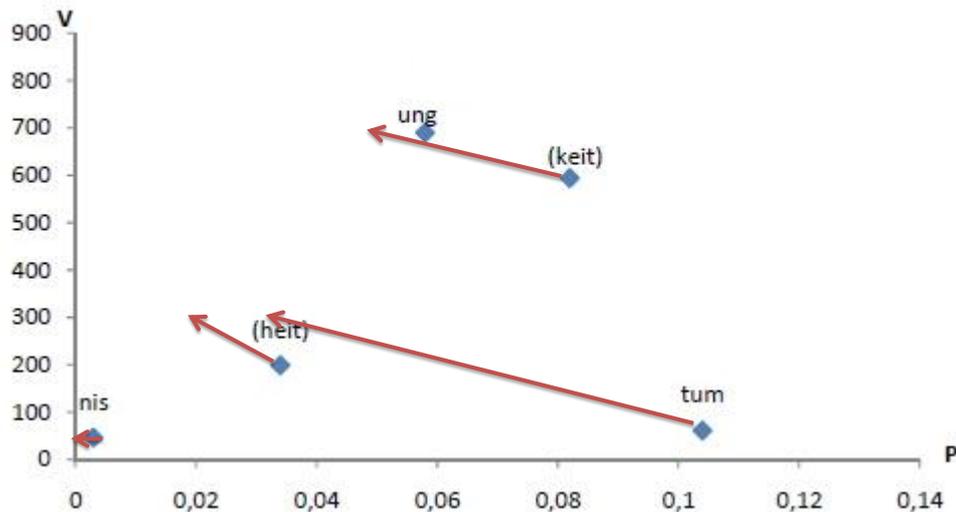


Abbildung 8.11: Veränderungen der globalen Produktivität der substantivbildenden Suffixe im Jahre 1999 anhand der extrapolierten Daten durch das finite Zipf-Mandelbrot-Modell

Auch die adjektivbildenden Suffixe wurden anhand einer normierten Stichprobengröße von 5000 Token in Bezug auf die daraus resultierenden Typenfrequenzen und die der Hapax Legomena untersucht; die Ergebnisse lassen sich Tab. 8.5 entnehmen.

Zunächst soll der Parameter Typenanzahl diskutiert werden. Zweifelsohne scheint an dieser Stelle *-haft* dasjenige Suffix zu sein, das am typenstärksten vertreten ist. Danach folgt *-bar* mit einer ebenfalls hohen Anzahl an Typen, die über die 500er-Grenze hinausgeht.

Für die drei bisher als typenstark deklarierten Suffixe *-ig*, *-lich* und *-isch* kann festgestellt werden, dass sich diese erneut ähneln, aber dennoch kleine Unterschiede zu verzeichnen sind, weil die Anzahl der Typen von *-isch* frequentiell höher ist als die der anderen beiden Suffixe.

Suffix	V	V1	P
<i>-isch</i>	436	119	0.024
<i>-lich</i>	390	113	0.023
<i>-ig</i>	398	126	0.025
<i>-bar</i>	513	194	0.039
<i>-sam</i>	76	23	0.005
<i>-mäßig</i>	305	145	0.029
<i>-los</i>	314	40	0.008
<i>-haft</i>	554	310	0.062
Suffix	95%-K.-intervalle für V	95%-K.-intervalle für V1	95%-K.-intervalle für P
<i>-isch</i>	420...452	101...136	0.020...0.027
<i>-lich</i>	374...407	95...130	0.019...0.026
<i>-ig</i>	380...416	108...145	0.022...0.029
<i>-bar</i>	491...535	171...216	0.034...0.043
<i>-sam</i>	68...84	15...31	0.003...0.006
<i>-mäßig</i>	284...326	125...166	0.025...0.033
<i>-los</i>	306...323	30...50	0.006...0.010
<i>-haft</i>	522...585	279...342	0.056...0.068

Tabelle 8.5: Auflistung der Extrapolations-Ergebnisse (V, V1, P) der substantivbildenden Suffixe nach dem finiten Zipf-Mandelbrot-Modell auf Basis der Daten des Jahres 1999 im DWDS

Für die Suffixe *-mäßig* und *-los* wird vorausgesagt, dass etwa 300 Typen auf 5000 Token auffindbar sind, was durchaus dafür spricht, dass es sich um produktive Suffixe handelt. Für *-sam* skizzieren die Werte ein unproduktives Bild, da nur zwischen 68 und 84 Typen gebildet werden.

Ein Suffix, das bei einer Datenmenge von 5000 Token zwischen 400 und 500 unterschiedliche Typen hervorbringt, darf aber nicht automatisch als hoch produktiv etikettiert werden, vielmehr müssen auch die anderen Frequenzen mit einbezogen werden. Und an dieser Stelle wird es interessant: Besonders produktiv ist *-haft*. Es werden über 300 Hapaxe prognostiziert; mit diesem Wert hebt sich das Suffix deutlich von *-bar* ab.

Für *-isch*, *-ig* und *-lich* zeigt sich, dass obwohl *-isch* typenstärker vertreten ist als seine beiden Mitstreiter, die jeweilige Anzahl der Hapaxe aller drei Suffixe vergleichbar ist.

Als weitere interessante Beobachtung muss festgehalten werden, dass *-mäßig*, dessen Typenanzahl geringer ist als die der drei eben beschriebenen Suffixe, dennoch dazu tendiert, mehr einmalige Bildungen zu produzieren als die Suffixe *-isch*, *-ig* und *-lich*.

Bei Betrachtung der Tabelle sticht eine weitere interessante Zahl heraus: die Anzahl der Hapax Legomena von *-los*, denn *-los* ist ein typenstarkes Suffix, dessen Anzahl der Hapaxe dennoch erstaunlich gering ist. Bei einem Input von 5000 Token sind nach Berechnungen des Modells nur etwa 40 einmalige Bildungen vorhanden, ein Wert, der mit der Anzahl der Hapax Legomena von *-sam* vergleichbar ist.

Selbstverständlich spiegeln sich die eben beschriebenen Beobachtungen auch in *P* wieder: Vergleichbar in ihrer Produktivität sind *-isch*, *-ig* und *-lich*; *-mäßig* ist leicht produktiver.

Mit 0.39 ist das Suffix *-bar* hoch produktiv; weitaus produktiver ist nur *-haft* mit einem Wert von 0.062. Im Vergleich zu den oben diskutierten Suffixen sind sowohl *-sam* wie auch *-los* unproduktiv.

Zusammenfassend lässt sich festhalten, dass **die Aussagen in den Grammatiken nicht ohne Weiteres für alle Suffixe bestätigt werden können**. In Tab. 5.1 wurden die Etikettierungen zum Grad der Produktivität der Suffixe gegenüber gestellt. Dabei wurde das Suffix *-los* von allen Linguisten als 'hoch/stark produktiv' deklariert. Die Berechnungen mit dem finiten Zipf-Mandelbrot-Modell zeigen, dass durchaus ein großes Inventar an Wortbildungen bereit steht, das Sprecher des Deutschen verwenden können, jedoch zeigt die Analyse auch, dass relativ wenige Neubildungen vorhanden sind, sodass davon auszugehen ist, dass *-los* in der Gegenwartssprache nicht (mehr) produktiv zur Wortschatzerweiterung beiträgt. Dieses schließt allerdings nicht aus, dass es in anderen Zeitstufen andersherum der Fall gewesen ist.

Interessant gestaltet sich in diesem Zusammenhang der Vergleich zu *-sam*, das nicht wesentlich unproduktiver erscheint als *-los*, aber in der Literatur deutlich als unproduktives/inaktives Suffix beschrieben wird.

Während Fleischer und Barz (1995) keine eindeutige Aussage zur Produktivität von *-haft* treffen (Tab. 5.1), merkt Lohde (2006) an, dass es sich um ein produktives Suffix handle; Motsch (2004) spricht von 'starker Aktivität'. Nach Analyse mit dem fZM-Modell ist *-haft* das von allen untersuchten Suffixen am produktivste.

Sowohl für *-mäßig* wie auch für *-haft* hat die Korpusstudie ergeben, dass es sich um sehr produktive Suffixe handelt. Allerdings wurden die Ergebnisse kritisch betrachtet, da in beiden Fällen nur sehr kleine Datenmengen verfügbar waren. Die Normierung der Stichprobengröße kann nun deutlich machen, dass *-mäßig* nun nicht mehr durch

eine besonders hohe Produktivität auffällt, sondern – gerade im Vergleich mit den anderen Suffixen – nur als produktiv erscheint.

Für *-bar* darf die Etikettierung der hohen Produktivität stehen bleiben, auch wenn nicht verkannt werden darf, dass sich die Werte im Vergleich zum auffälligen *-haft* unterscheiden.

Im Hinblick auf die Visualisierung der globalen Produktivität ergibt sich das in Abb. 8.12 dargestellte Bild.

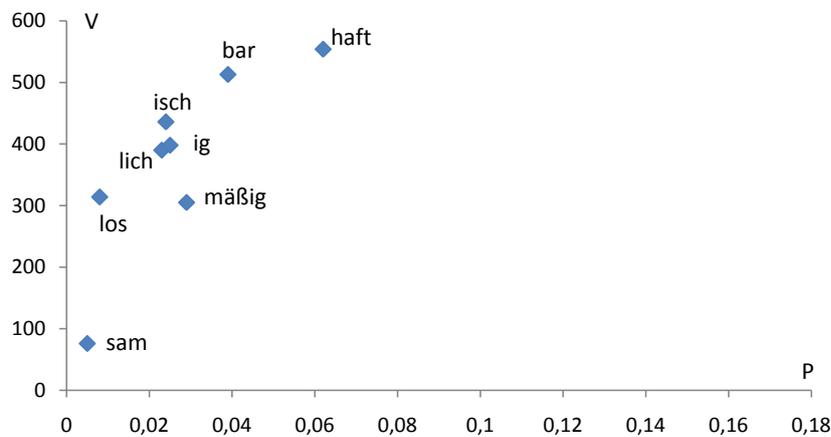


Abbildung 8.12: Globale Produktivität der adjektivbildenden Suffixe im Jahre 1999 anhand der extrapolierten Daten mithilfe des finiten Zipf-Mandelbrot-Modells

Es sei darauf hingewiesen, dass die Skalierung der x-Achse und der damit zusammengehörige Wertebereich aus Abb. 7.25 übernommen wurde, denn auch für die adjektivbildenden Suffixe sollen die 'Bewegungen', bzw. Berichtigungen, die sich durch die Normierung mithilfe des fZM-Modells ergeben, in einem Schaubild dargestellt werden (Abb. 8.13).

Abbildung 8.13 verdeutlicht nun, dass die Analyse nach dem fZM-Modell im Vergleich zu Abb. 7.25 große Unterschiede in der globalen Produktivität aufzeigt.

Insbesondere bei denjenigen Suffixen, von denen nur geringe Tokenmengen aus dem Korpus extrahiert werden konnten – und zu diesen Suffixen gehören *-haft*, *-mäßig* und

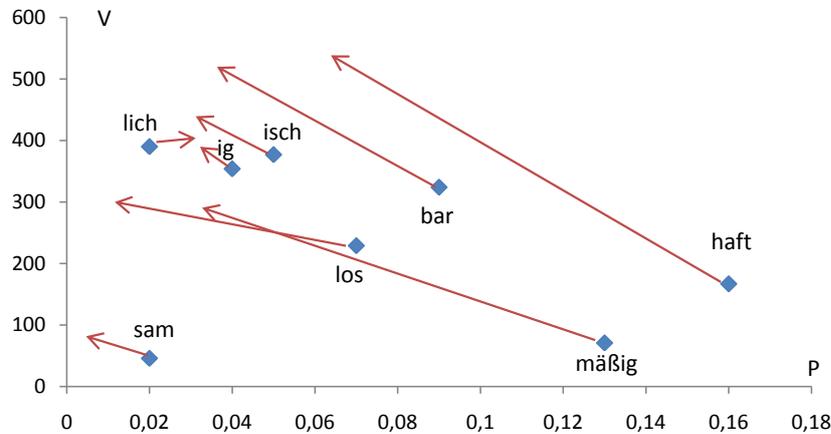


Abbildung 8.13: Veränderungen der globalen Produktivität der adjektivbildenden Suffixe im Jahre 1999 anhand der extrapolierten Daten nach dem finiten Zipf-Mandelbrot-Modell

-los – zeigen sich nun große Verschiebungen mit Bewegungen nach links und nach oben: nach links, weil die Produktivität deutlich gesunken ist, nach oben, weil sich durch die höhere Tokenanzahl von 5000 Token eine höhere Typenanzahl ergibt.

Am deutlichsten, und dies lässt sich grafisch an der Länge der Pfeile erkennen, erscheinen die Verschiebung von *-mäßig* und *-haft*, doch auch für *-bar* und *-los* ist ein Trend nach links und nach oben erkennbar.

Für das Suffix *-sam* wird anhand von Abb. 8.13 deutlich, dass es ebenfalls noch weniger produktiv ist als nach Abb. 7.25.

Das produktivste adjektivbildende Suffix ist *-haft*, das, sofern man Abb. 8.12 so darstellen würde, dass die x-Achse bei 0,08 endet, im oberen rechten Quadranten anzusiedeln wäre. Ebenfalls als sehr produktiv gilt nach Abb. 8.12 auch das Suffix *-bar*.

Die Modifizierungen für *-isch*, *-ig* und *-lich* sind nicht auffällig: für *-isch* und *-ig* darf jedoch im Gegensatz zu *-lich* festgehalten werden, dass diese beiden Suffixe nach Normierung ein wenig unproduktiver gelten und *-lich* ein klein bisschen produktiver

erscheint. Interessant ist, dass sich *-ig* und *-lich* aneinander annähern und man ableiten kann, dass sie etwa gleich produktiv sind.

Zusammenfassend lässt sich das folgende Zwischenergebnis festhalten: Insbesondere diejenigen Suffixe, die anhand der reellen Daten nur token- und typenarm vertreten waren, zeigten nach Normierung große Veränderungen hinsichtlich ihrer Frequenzen und damit verbunden Maße und P^* .

Abschließend soll an einem Beispiel getestet werden, ob das finite Zipf-Mandelbrot-Modell auch zur diachronen Analyse angewendet werden kann. Dazu wurden die Datenreihe von *-ung* für die einzelnen Zeitabschnitte untersucht, was zu den in Tab. 8.6 präsentierten Werten führt.

Jahr	V	V1	P
1999	696	286	0.057
1975	590	224	0.045
1950	580	226	0.045
1925	666	282	0.056
1900	593	218	0.044
Jahr	95%-K.-intervalle für V	95%-K.-intervalle für V1	95%-K.-intervalle für P
1999	668...724	258...315	0.052...0.063
1975	565...615	199...249	0.040...0.050
1950	555...605	201...251	0.040...0.050
1925	637...694	254...310	0.051...0.062
1900	570...616	194...241	0.039...0.048

Tabelle 8.6: Auflistung der Extrapolations-Ergebnisse für das Suffix *-ung* nach dem fZM-Modell (V, V1, P) anhand der Daten des DWDS'

Für *-ung* kann beobachtet werden, dass die Werte durchweg konstant sind und zudem auf eine hohe Produktivität hindeuten. Sowohl die Anzahl der Typen wie auch die Anzahl der zu erwarteten Hapaxe ist hoch.

Dennoch wird man feststellen, dass für die Zeitabschnitte 1925 und 1999 leicht höhere Werte als für die anderen Abschnitte berechnet werden, was impliziert, dass die Produktivität P in diesen Jahren ebenfalls höher ist als in den Jahren 1900, 1950 und 1975.

Schaut man sich dazu die Darstellung der globalen Produktivität an, so wird man unter Einbeziehung der Konfidenzintervalle zum Schaubild 8.14 gelangen.

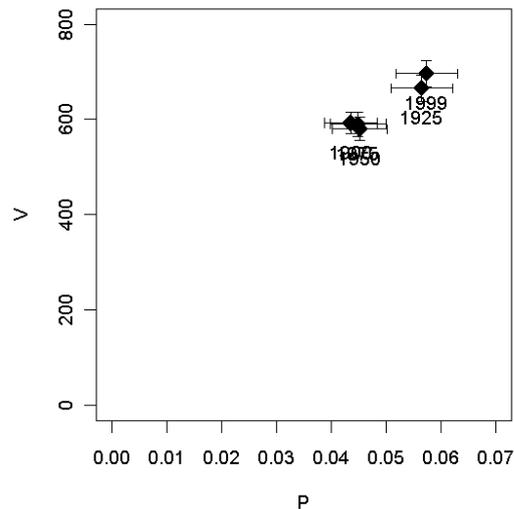


Abbildung 8.14: Darstellung der globalen Produktivität von *-ung* für das 20. Jahrhundert mithilfe von Konfidenzintervallen

Abb. 8.14 macht ersichtlich, dass es zu einer Überlappung der Konfidenzstäbe kommt, was impliziert, dass es sich um Zufallsschwankungen handelt.

Es ist davon auszugehen, dass der tatsächliche Wert von P über den beobachteten Zeitraum konstant ist und 0.05 beträgt, denn dieser Wert scheint für alle fünf Stichproben plausibel, indem er innerhalb oder sehr nahe an den Konfidenzintervallen liegt.¹⁵

8.3 Probleme mit dem Modell

Wie bereits im Vorfeld skizziert wurde, ist es mit dem fZM-Modell zum einen möglich, die Populationsgröße eines Suffixes zu berechnen, zum anderen ließe sich auch den Parameter Alpha, der die Schiefe des Frequenzspektrums beschreibt, zur Interpretation der Produktivität heranziehen.

Beides wurde überprüft, sowohl für die adjektiv- wie auch substantivbildenden Suffixe, doch die Ergebnisse waren nicht zufriedenstellend.

¹⁵Dazu siehe auch Baroni und Evert (2007, 906): „We complement rMSEs with reports on the average relative error (indicating whether there is a systematic under- or overestimation bias) and its asymptotic 95% confidence intervals, based on the empirical standard deviation of the e_i across the 20 trials (the confidence intervals are usually somewhat larger than the actual range of values found in the experiments, so they should be seen as 'pessimistic estimates' of the actual variance).“

Exemplarisch soll dazu eine tabellarische Aufstellung (Tab. 8.7) der berechneten Populationsgrößen für die substantivbildenden Suffixe präsentiert werden.

	1900	1925	1950	1975	1999
<i>-ung</i>	900	1371	1145	1250	1323
<i>-tum</i>	76	48	40	72237946028	4440567310
<i>-nis</i>	54	48	46	164	48
<i>-heit/-keit</i>	1795	1356	973	794	1349

Tabelle 8.7: Berechnete *S*-Werte auf Basis des finiten Zipf-Modells der substantivbildenden Suffixe *-ung*, *-tum*, *-nis* und *-heit/-keit* für die Zeiträume 1900, 1925, 1950, 1975 und 1999

Die Werte von Tab. 8.7 lassen keine soliden Aussagen zum Wandel der Produktivität zu, vielmehr scheint es so zu sein, dass die Parameter-Schätzung des Modells fehlerhaft ist, was sich insbesondere an den Werten von *-tum* für 1975 und 1999 zeigt. Festzuhalten bleibt deshalb, dass an dieser Stelle noch mathematische Entwicklungsarbeit notwendig ist, um das fZM-Modell robuster zu machen.

Für die Produktivitätsmessung wäre die Stabilität des Modells, auch im Hinblick auf die berechnete Populationsgröße, ein enormer Gewinn, denn Populationsgrößen könnten im Idealfall Aufschluss darüber geben, ob sich der Bestand an Wörtern im Laufe der Zeit vergrößert oder verkleinert, was als Indikator für steigende oder fallende Produktivität verwendet werden könnte.

8.4 Exemplarische Wachstumskurven der substantivbildenden Suffixe

Die Betrachtung von morphologischer Produktivität sollte nach Meinung von Lüdeling und Evert (2005, 351) sowohl qualitativ wie auch quantitativ erfolgen.¹⁶

¹⁶„Our claim is that the problem of productivity can only be understood when different kinds of evidence — quantitative and qualitative — are combined.“

Da in den vorangegangenen Abschnitten die quantitative Betrachtung im Vordergrund stand, soll nun skizziert werden, wie auch qualitative Analysen durch Anwenden des finiten Zipf-Mandelbrot-Modells möglich sind.

Exemplarisch werden dazu im Nachfolgenden Wachstumskurven (V und V1) interpretiert, die anhand der für das Jahr 1999 verfügbaren Daten erzeugt werden konnten. Die jeweilige obere Kurve skizziert den Verlauf der Entwicklung der Typen, die untere die der Hapax Legomena.

Aufgrund der derzeit noch bestehenden Probleme mit diesem Modell, die im vorangegangenen Abschnitt skizziert wurden, werden die Extrapolationskurven nicht für alle Suffixe aufgeführt und interpretiert. Vielmehr geht es darum aufzuzeigen, wie eine qualitative Analyse von Wachstumskurven aussehen würde, sofern die Modelle robust funktionierten.

Als erstes Suffix sollen die Wachstumskurven für das Suffix *-ung* analysiert werden. Zunächst wird angenommen, es stünden exakt 10.000 Token zur Verfügung (Abb. 8.15). Das Schaubild zeigt, dass die Wachstumskurve der Typen stetig ansteigt und

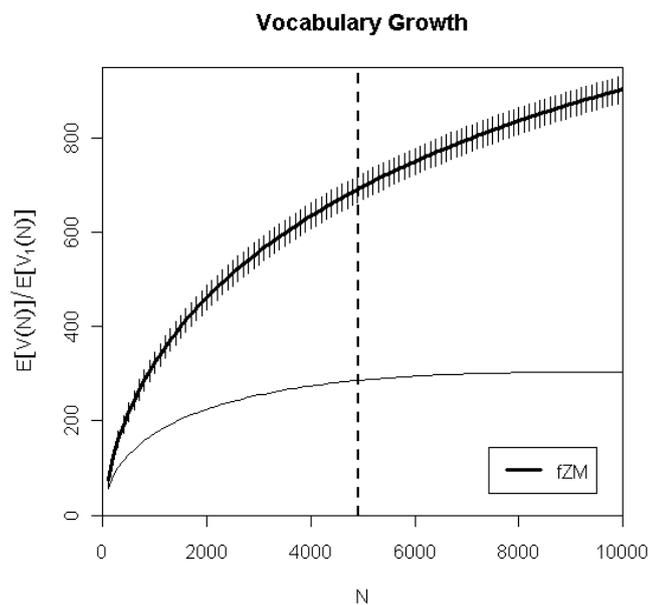


Abbildung 8.15: Extrapolation des Typenwachstums für das Suffix *-ung* als Prognose bis zu 10.000 Token

davon auszugehen ist, dass ein Sprecher des Deutschen selbst nach einem Input von 10.000 Token noch nicht auf alle möglichen Wörter mit *-ung* gestoßen ist. Die deutet

auf den ersten Blick darauf hin, dass es sich um einen produktiven Wortbildungsprozess handelt.

Bei Betrachtung der Wachstumskurve der Hapax Legomena fällt auf, dass die Kurve zunächst ansteigt, aber im Verlauf der Extrapolation relativ stark abflacht. Dennoch lassen sich auch nach einem Input von 10.000 Token noch immer über 200 Wörter finden, die einmalig vorhanden sind.

Erst bei einer Extrapolation auf 100.000 Token (Abb. 8.16) wird deutlich, wie sich dieses Wortbildungsmodell entwickelt: Ab einer Tokenmenge von etwa 20.000 Token würde die Kurve der Hapaxe deutlich abfallen, die Wachstumskurve der Typen würde nicht weiter ansteigen und sich bei einem Wert von etwas über 1200 Typen konstant halten.

Die Kurve der Hapaxe zeigt zudem, dass selbst nach 100.000 Token noch immer Neubildungen entstehen können, sodass *-ung* als Wortbildungsmodell weiterhin zur Erweiterung des Lexikons beiträgt, wenn auch nur im geringen Umfang.

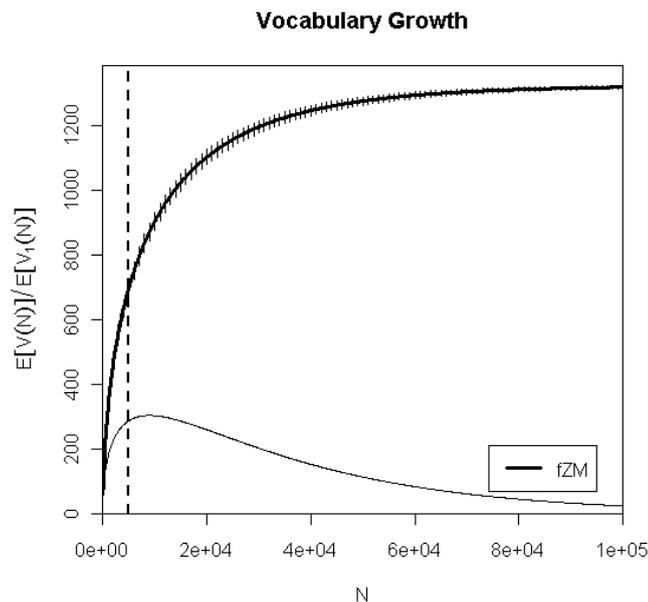


Abbildung 8.16: Extrapolation des Typenwachstums für das Suffix *-ung* auf Basis eines Inputs von 100.000 Token

Betrachten wir nun die Wachstumskurven von *-nis*. Es wird bis zu einem Input von 10.000 Token extrapoliert (Abb. 8.17).

Für das als nicht mehr aktive und produktive Suffixe *-nis* kann diese Unproduktivität auch durch die Modellierung mit dem fZM-Modell bestätigt werden. Vorausgesagt werden 48 Typen, die bereits nach einem Input von etwa 4000 Token erreicht wären. Eine Voraussage bis zu einem Input von 100.000 Token ist deshalb nicht notwendig.

Zu der Entwicklung der Hapax Legomena lässt sich festhalten, dass nach einem Input von etwas mehr als 6000 Token keine Bildungen auftauchen dürften, die einmalig vorkommen: Auch dies ist ein deutliches Zeichen für einen unproduktiven Wortbildungsprozess. Demnach lassen sich die Aussagen in der Literatur, *-nis* als unproduktiv zu bezeichnen, ohne Zweifel bestätigen.

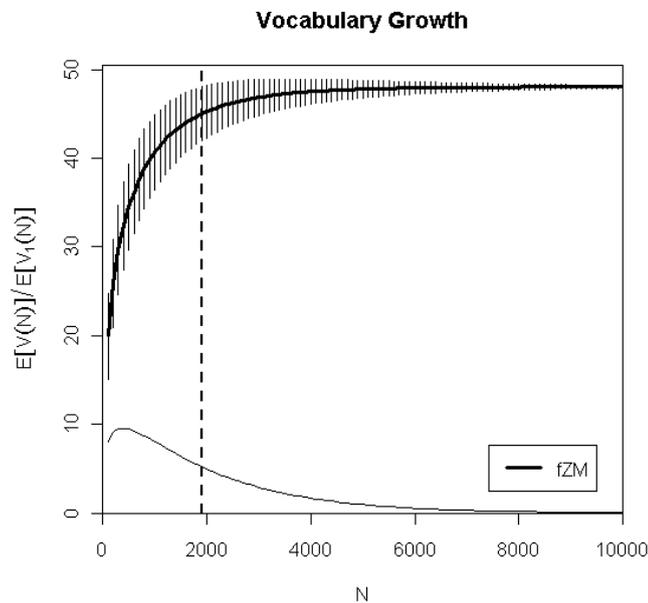


Abbildung 8.17: Extrapolation des Typenwachstums für das Suffix *-nis* auf Basis eines Inputs von 10.000 Token

Wie an einigen Stellen der vorliegenden Arbeit deutlich wurde, lässt sich das finite Zipf-Modell noch nicht einwandfrei anwenden. Dies zeigt sich auch in den Extrapolationen mit dem Suffix *-tum* (Abb. 8.18 und Abb. 8.19).

Sowohl bei einer Extrapolation auf 10.000 wie auch auf 100.000 Token scheint das Suffix in einem überdimensionalen Maße produktiv zu sein. Dies erscheint nicht realistisch. Demnach ist davon auszugehen, dass entweder die Stichprobengröße von 334 Token zu gering ist, um valide Ergebnisse herbeizuführen, oder aber es liegen hier mathematische Probleme des Modells vor.

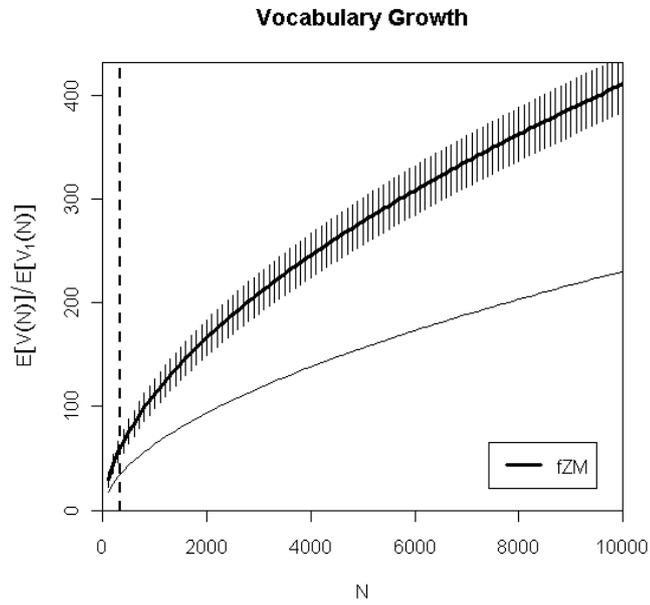


Abbildung 8.18: Extrapolation des Typenwachstums für das Suffix *-tum* auf Basis eines Inputs von 10.000 Token

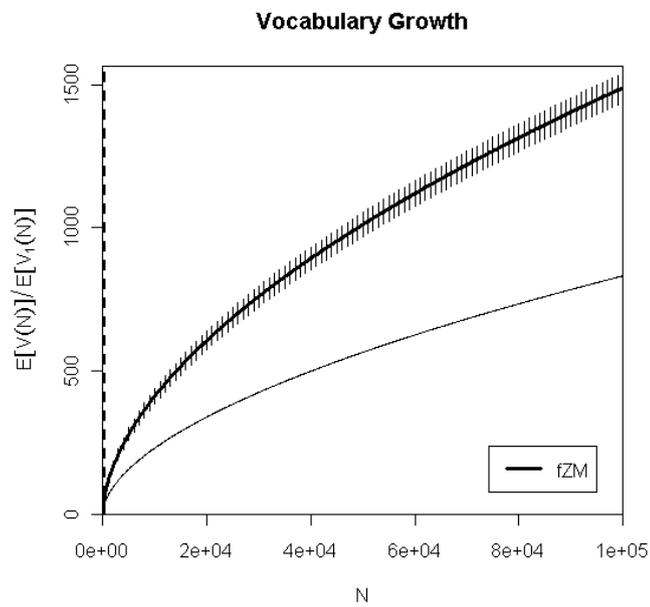


Abbildung 8.19: Extrapolation des Typenwachstums für das Suffix *-tum* auf Basis eines Inputs von 100.000 Token

8.5 Exemplarische Wachstumskurven der adjektivbildenden Suffixe

Auch für die adjektivbildenden Suffixe sollen exemplarisch Kurvenverläufe besprochen werden.

Für das typenstark auftretende Suffix *-ig* zeigen die Extrapolationen nach dem fZM-Modell nun (Abb. 8.20), dass es auch nach einem Input von 10.000 Token einen Anstieg des Typenwachstums gibt. Die Kurve der Hapaxe jedoch flacht nach Input einer derartigen Tokenmenge langsam ab. Das fZM-Modell berechnet, dass mit dem Suffix höchstens 579 Typen gebildet werden; dieser Wert wird dabei ungefähr ab einem Input von 30.000 Wörtern erreicht.

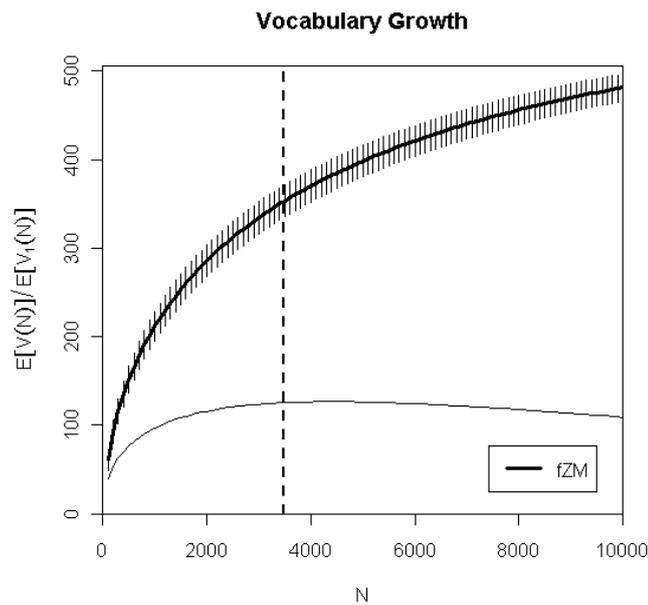


Abbildung 8.20: Extrapolation für das Suffix *-ig* als Prognose auf Basis eines Inputs von 10.000 Token

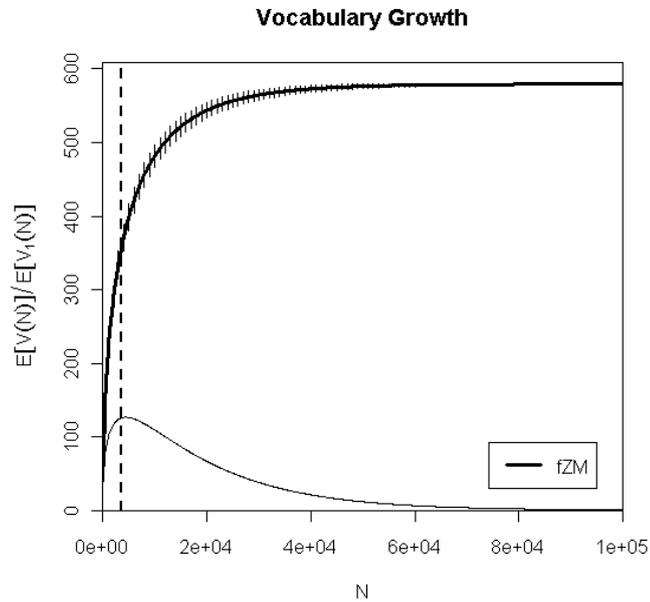


Abbildung 8.21: Extrapolation für das Suffix *-ig* als Prognose auf Basis eines Inputs von 100.000 Token

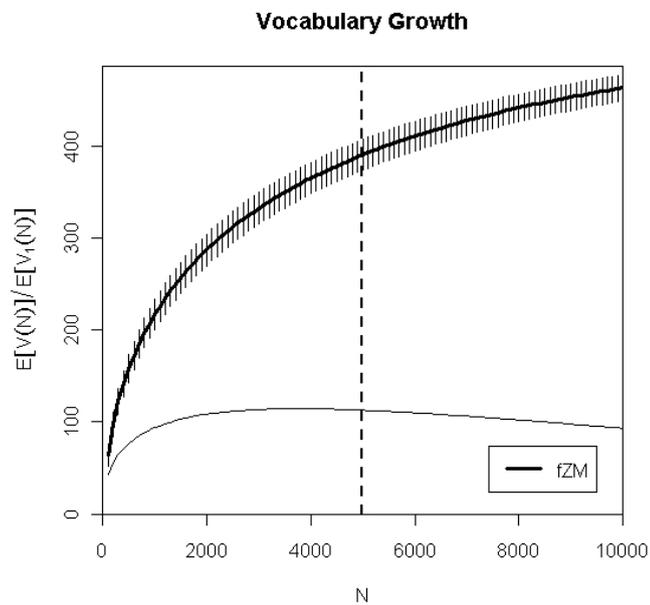


Abbildung 8.22: Extrapolation für das Suffix *-lich* als Prognose auf Basis eines Inputs von 10.000 Token

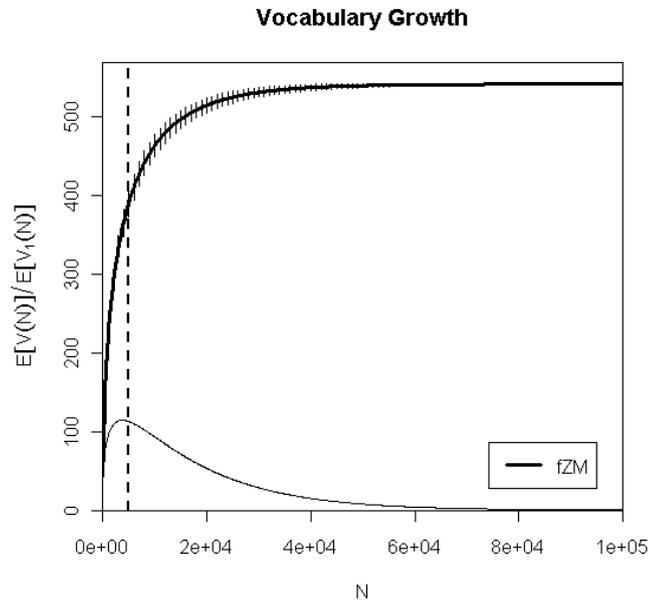


Abbildung 8.23: Extrapolation für das Suffix *-lich* als Prognose auf Basis eines Inputs von 100.000 Token

Auch für *-lich* wurden Kurven extrapoliert; die Ergebnisse können Abb. 8.22 und Abb. 8.23 entnommen werden.

Das Produktivitätsverhalten zeigt Parallelen zu dem zuvor beschriebenen Suffix *-ig* auf. Bei einer Datenmenge von 10.000 Token ist zu erwarten, dass die mögliche Typenanzahl noch nicht erreicht ist. Die Steigung der Kurve liegt zu diesem Zeitpunkt noch bei 0.009.

Die Kurve der Hapaxe ist nach 10.000 Token bereits abfallend, wenn auch nur sehr flach, was bedeutet, dass selbst nach einem Input von 10.000 *-lich*-Token noch knapp unter 100 Einmalbildungen gefunden werden. Erst bei einem Input von etwa 70.000 *-lich*-Token ist davon auszugehen, dass keine oder nur sehr wenige Neubildungen gebildet werden. Berechnet werden zudem 541 Typen, die das Modell generiert; es handelt sich dabei um eine nicht geringe Anzahl an Wortbildungen, die zudem mit der Anzahl von *-ig* und *-isch* vergleichbar ist.

Nach Analyse der Daten von *-mäßig* – ohne dabei das finite Zipf-Modell anzuwenden – stand ein Linguist bisher vor erheblichen Interpretationsschwierigkeiten, denn auf der einen Seite zeigte das in Kap. 5 präsentierte Vorexperiment, dass *-mäßig* produktiv ist, weil viele nicht-lexikalisierte Bildungen produziert werden – wesentlich

mehr als mit den anderen untersuchten Suffixen –, auf der anderen Seite jedoch standen mithilfe des DWDS jeweils geringe Datenmengen zur Verfügung, sodass die bisherigen Ergebnisse in Zweifel gezogen werden mussten.

Nunmehr soll auch für dieses Suffix eine qualitative Analyse anhand von Abb. 8.24 und Abb. 8.25 erfolgen. Das Suffix ist in der Lage, 463 Typen zu bilden. Bereits nach einem Input von 10.000 Textwörtern mit *-mäßig* wird man 3500 Typen analysieren; nach etwa 20.000 Token müsste bereits eine Typenanzahl von 400 Typen vorliegen.

Die Steigung der Typen-Kurve läge bei einer Tokenanzahl von 10.000 Wörtern bei 0.01, was verdeutlicht, dass die Kurve zu diesem Zeitpunkt nicht besonders steil ansteigt.

Zur Entwicklung der Hapax Legomena lässt sich festhalten, dass die Kurve bis zu einem Wert von 100 ansteigt, dann langsam abflacht, aber bis zu einem Input von 80.000 Token noch Neubildungen wahrscheinlich sind.

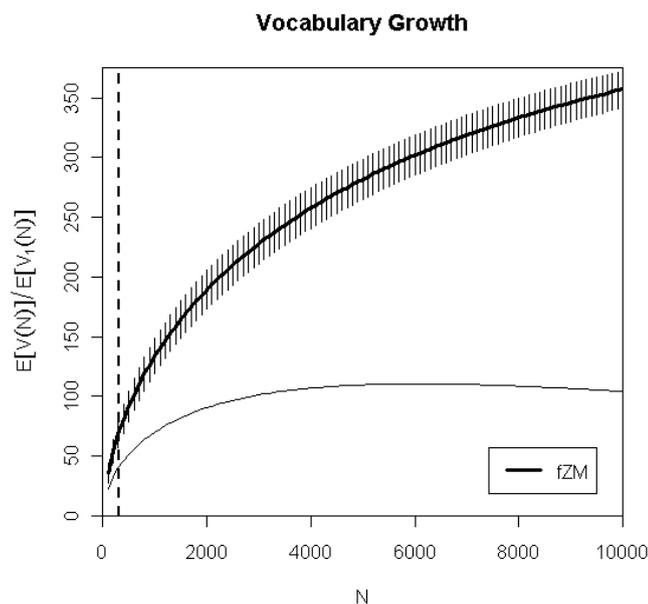


Abbildung 8.24: Extrapolation für das Suffix *-mäßig* als Prognose auf Basis eines Inputs von 10.000 Token)

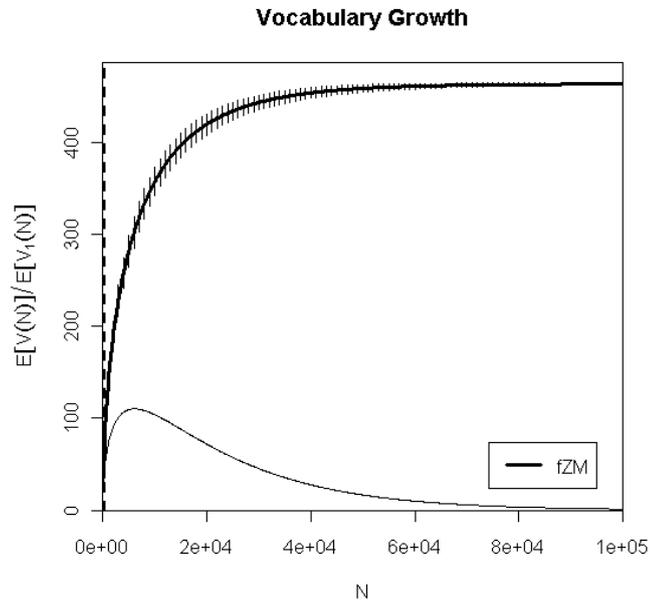


Abbildung 8.25: Extrapolation für das Suffix *-mäßig* als Prognose auf Basis eines Inputs von 100.000 Token)

8.6 Exemplarische Wachstumskurven zur diachronen Betrachtung morphologischer Produktivität

Die Betrachtung von Wachstumskurven kann auch unter diachroner Fragestellung angewandt werden, indem mehrere Kurven, die aus unterschiedlichen Datenreihen verschiedener Messzeiträume resultieren, miteinander verglichen werden. Möglich ist dann auch eine Darstellung mehrerer Kurven in einem einzigen Schaubild, was die Vergleichbarkeit erheblich erleichtert.

Dies soll auch im Folgenden so durchgeführt werden. Die bereits diskutierte Problematik des noch nicht tadellos funktionierenden finiten Zipf-Mandelbrot-Modells erlaubt es an dieser Stelle leider nicht, die Methodik für alle Suffixe anzuwenden und die anschließenden sichtbaren Ergebnisse kritiklos zu interpretieren. Deshalb werden auch hier – wie auch schon im vorangegangenen Kapitel – nur exemplarische Analysen vorgenommen, die das Ziel verfolgen zu skizzieren, wie es *theoretisch* möglich

ist, das fZM-Modell auch zur diachronen Betrachtung morphologischer Produktivität anzuwenden.

Als erstes wird das unproduktive Suffix *-nis* betrachtet (Abb. 8.26).

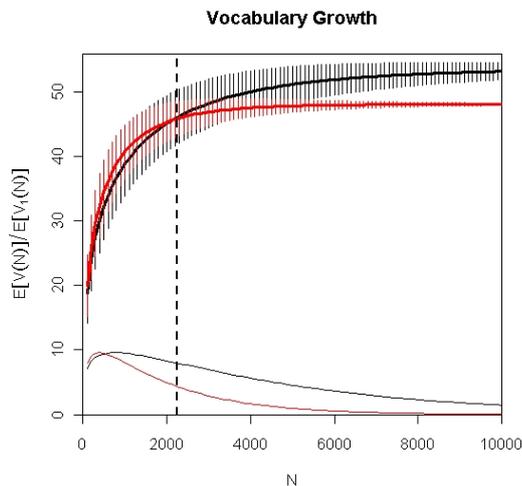


Abbildung 8.26: Extrapolationskurven von *-nis* anhand der Daten von 1900 (schwarz) und 1999 (rot)

Ersichtlich ist, dass die Kurven des Jahres 1900 auf einen leicht produktiveren Prozess hindeutet als die Kurve(n) des Jahres 1999. Die Unterschiede sind jedoch sehr gering.

Besonders interessant ist in diesem Zusammenhang aber auch, dass die Konfidenzstäbe darauf hinweisen, dass es sich um Zufallsschwankungen handelt.

Zudem skizziert die Wachstumskurve der Hapax Legomena auf Basis der Daten des Jahres 1999, dass ein Sprecher auch selbst nach Rezipieren von mehr als 10.000 Wörtern auf eine Derivation mit *-nis* stoßen könnte, die er bisher nicht aufgenommen hat, während die Extrapolationen anhand der Datenbasis aus dem Jahr 1900 aufzeigen, dass ab einer Datenmenge von 6000 Token keine einmaligen Bildungen auftauchen dürften.

Wenn es sich hier deshalb auch nur um sehr feine Unterschiede handelt, wäre es höchst interessant zu erfahren, ob der Trend zur Unproduktivität in den nächsten Jahren, Jahrzehnten und Jahrhunderten weiter fortgesetzt wird.

Sichtbar wird an diesem Beispiel auch, dass nur ein minimaler Unterschied in der Produktivität im Vergleich der beiden Zeitabschnitte vorliegt. Den Kurven nach lässt

sich ableiten, dass das Suffix im Jahre 1900 ein klein wenig produktiver gewesen ist als im Jahre 1999. Eindeutige Aussagen, ob es sich hierbei um einen Trend in Richtung abnehmende Produktivität handelt, ließe sich erst nach Ansicht weiterer Daten vornehmen. Dabei wären Daten aus den vorangegangenen Jahrhunderten von Vorteil, weil die Entwicklungen von Affixen im Zuge ihrer Grammatikalisierungsprozesse sehr langsam verlaufen.

Als zweites Exempel wird das Suffix *-ig* betrachtet (Abb. 8.27). Die Extrapolationen auf 100.000 Token machen deutlich, dass sich das Suffix für beide Messzeiträume sehr unterschiedlich verhält: Während es im Jahre 1999 recht schnell zu einem Abflachen der Kurven kommt, was impliziert, dass ein Sprecher des Deutschen nach Input von etwa 60.000 Wörtern auf kein neues *-ig*-Derivat stoßen wird, so zeigen die Kurven von 1900 einen anderen Verlauf auf. Denn selbst nach Aufnahme von 100.000 Token mit *-ig*, wird man weitere Typen, also bis zu diesem Zeitpunkt nicht rezipierte Wörter, finden.

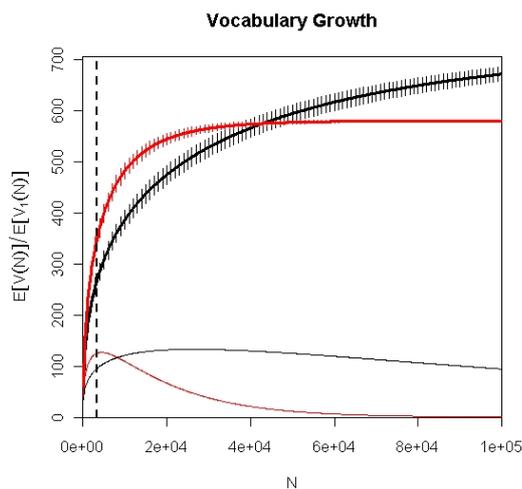


Abbildung 8.27: Extrapolationskurven von *-ig* anhand der Daten von 1900 (schwarz) und 1999 (rot)

Zusammenfassend lässt sich festhalten, dass ein Zusammenspiel von quantitativen und qualitativen Methoden sinnvoll erscheint, um Produktivitätsentwicklungen festzustellen und zu beschreiben. Mithilfe des finiten Zipf-Mandelbrot-Modells werden neue Zugänge zur Produktivitätsforschung eröffnet.

Die Gegenüberstellung von Wachstumskurven erlaubt es dem Linguisten, kleine oder auch größere Unterschiede zu visualisieren. Zur abschließenden Interpretation

der Kurven können dann verschiedene Fragestellungen ihre Berücksichtigung finden, die integrativ behandelt werden sollten:

1. Wie ist der Verlauf der Wachstumskurve(n) und wie groß muss der Input sein, damit es zu einem Abflachen der Kurven kommt? Welche Schlussfolgerungen lassen sich daraus (sowohl für die diachrone wie auch synchrone Analyse) ziehen?

2. Was sagt S über die Produktivität aus?

Hierbei handelt es sich um eine nicht einfach zu beantwortende Frage, denn letztendlich kann von einem hohen Wert S nicht zwangsweise auf eine hohe Produktivität geschlossen werden. Gründe, weshalb in diesem Fall weitere Betrachtungen notwendig sind, wurden bereits hinreichend aufgeführt: Möglich ist unter anderem, dass es sich um ein unproduktives Suffix handelt, das in vergangener Zeit produktiv gewesen ist, sodass bereits viele lexikalisierte Wörter im Wortschatz verankert sind.

Desiderata

Ziel dieser Arbeit ist es gewesen, dem Phänomen der derivationellen Produktivität vor allem aus quantitativer Sicht auf den Grund zu gehen. Empirisch untersucht wurde der Wortschatz des 20. Jahrhunderts, genauer gesagt, handelte es sich um ausgewählte substantiv- und adjektivbildenden Suffixe des Deutschen.

In der vorliegenden Arbeit wurde von einem Produktivitätsbegriff ausgegangen, bei welchem angenommen wird, ein Wortbildungsmodell ist genau dann als produktiv zu bezeichnen, wenn es stetig dazu beiträgt, neue Wörter zu bilden und damit das Lexikon erweitern kann.

Das Phänomen Produktivität in der Morphologie wurde in den letzten zwanzig Jahren ausgiebig thematisiert, aus quantitativer Sicht insbesondere durch die statistischen Modelle von *Harald Baayen*. Die Studie von Scherer (2006) zeigte in der Vergangenheit, dass die Anwendung der klassischen Methode *P*, die Hapaxe in Relation zu den Token zu setzen, problematisch ist, was in der Rezension von Lüdeling (2009) deutlich gemacht wird.

Aus diesem Grund wurde die von Lüdeling (2009) formulierte Kritik anhand ausgewählter Beispiele berücksichtigt, indem das *finite Zipf-Mandelbrot-Modell*, welches eines von derzeit unterschiedlichen verfügbaren LNRE-Modellen darstellt, im Zuge der methodologischen Diskussion exemplarisch angewandt wurde.

Die Intuition der deutschen Sprecher im Hinblick auf die Produktivität der zu untersuchenden Suffixe wurde in der vorliegenden Studie durch ein kleines Vorexperi-

ment berücksichtigt, indem es erste Erkenntnisse über das Verhalten der untersuchten Suffixe brachte. Danach erfolgte eine Datenerhebung auf Basis des DWDS, das als ausgewogenes deutsches Korpus gilt.

Ein Jahrhundert zu untersuchen, wie es das DWDS-Kernkorpus erlaubt, kann durchaus ausreichend sein. Doch traten Probleme auf, die sich vor allem aus den ungleichen Datenmengen der einzelnen Suffixe ergaben. Es ist deshalb offensichtlich, dass die jeweiligen Wachstumskurven (*V* und *VI*) auch bei unproduktiven Prozessen zunächst ansteigen und erst nach Betrachtung einer 'angemessenen' Datenmenge klare Interpretationen möglich sind. Deshalb erfolgte eine Normierung der Daten und u.a. eine Neuberechnung der jeweiligen *P*-Werte.

Die im letzten Schritt getätigten Extrapolationskurven der Suffixe machte visuell deutlich, dass in vielen Fällen eine Datenmenge von 10.000 Textwörtern nicht ausreicht, um festzustellen, ab wann ein Wortbildungsprozess unproduktiv wird. In einigen Fällen handelt es sich um dermaßen produktive Prozesse, dass ein Input von 100.000 notwendig ist, um zeigen zu können, wie viele Wörter ein Sprecher rezipieren muss, um auf alle Wörter eines Wortbildungsmodells zu stoßen.

Dennoch muss auch festgehalten werden, dass das finite Zipf-Modell derzeit noch nicht fehlerfrei funktioniert, sodass weitere mathematische Entwicklungsarbeit notwendig ist. Generell jedoch darf unstrittig stehen bleiben, dass das Modell einen neuen Zugang zur Produktivitätsforschung ermöglicht, der von großer Bedeutung ist, denn erst durch Intra- und Extrapolationen können ungleiche Datenmengen miteinander verglichen werden.

Deshalb müssen an dieser Stelle auch die noch existenten Probleme festgehalten werden:

- Es ist dringend erforderlich, dass große diachron ausgerichtete Korpora zur Verfügung gestellt werden, die es den Linguisten ermöglichen, Sprachwandel-Erscheinungen, zu denen auch Grammatikalisierungs-Phänomene gehören, über Jahrhunderte hinweg statistisch zu untersuchen.
- Die mathematisch ausgerichteten (Computer-)Linguisten und Mathematiker der Zukunft sollten überlegen, wie es möglich werden könnte, die Lexikalischerheit als Faktor in ein statistisches Produktivitätsmaß mit einzubauen. Dies wurde in der Vergangenheit gänzlich vernachlässigt, aber in der vorliegenden Arbeit

konnte deutlich gemacht werden, dass lexikalisierte und damit häufig verbunden hoch frequente Formen die Produktivitätsmaße beeinflussen.

- Das finite Zipf-Modell ist derzeit bereits anwendbar und für die Interpretation der morphologischen Produktivität sicherlich von großer Wichtigkeit, doch auch auf diesem Gebiet muss weitere mathematische Entwicklungsarbeit geleistet werden. Wünschenswert wäre es, wenn der Parameter Alpha zukünftig als robustes Maß zur Verfügung stünde und die Parameterschätzung der Modelle zuverlässig funktioniert.

Abschließend lassen sich die folgenden Schritte der quantitativen Untersuchung morphologischer Produktivität in Kürze zusammenfassen, um so eine Handlungsempfehlung für weitere Arbeiten zu formulieren.

1. Auswahl des Untersuchungsgegenstandes:

Zunächst einmal ist zu überlegen, welche Affixe untersucht werden sollen. Die Motivation, die hinter der jeweiligen Studie steckt, kann unterschiedlich fokussiert sein. Es kann sowohl interessant sein, ein einziges Suffix aus diachroner Perspektive zu betrachten und – sofern die Korpora bereit stünden (siehe dazu auch Punkt 2) – und die gesamte Entwicklung dieses einen Affixes (auch im Hinblick auf eine mögliche Grammatikalisierung und die Überprüfung der Unidirektionalität) empirisch zu betrachten, ebenso könnten aber auch Fragestellungen in Bezug auf das Wortbildungssystem vordergründig sein, was eine Untersuchung mehrerer Affixe impliziert. Des Weiteren könnte auch Interesse daran bestehen, ein oder mehrere Affixe in verschiedenen Registern zu untersuchen. Auch im Hinblick auf den Spracherwerb wären Untersuchungen zum Feststellen der Wortschatz-Entwicklung möglich.

2. Auswahl/Erstellung des Korpus:

Im zweiten Schritt sollte recherchiert werden, welche Korpora bereits verfügbar sind. Gerade was die diachrone Ausrichtung einer Studie betrifft, wird man derzeit – was die deutsche Sprache anbelangt – vor Problemen stehen, wenn eine Ausrichtung über mehrere Jahrhunderte erfolgen soll.

Selbstverständlich ist das Erstellen eines eigenen Korpus möglich: Beachtet werden muss in einem solchen Falle, dass Ausgewogenheit der Daten an primärer Stelle stehen muss.

3. **Datenerhebung:** Einige online-verfügbare Korpora bieten Suchmaschinen an, mit denen es mithilfe von regulären Ausdrücken möglich ist, schon von Beginn an eine Datenmenge zu erheben, die nur einen geringen Aufwand an Nachbereitung erfordert.

So z.B. wirft das Untersuchen des Suffixes *-ung* Probleme auf, weil die Abfrage nach der reinen Buchstabenfolge *u-n-g* auch zu Treffern wie *jung* etc. führt. Ein Ausschluss an Homophonen wäre vorteilhaft. Dieses kann allerdings nur bei morphologisch-annotierten Korpora durchgeführt werden.

4. **Datenbereinigung:**

Wie unter Punkt 3 beschrieben, ist es in vielen Fällen notwendig, die erhobenen Daten zu bereinigen.

In einem derartigen Schritt sollten nicht nur Homophone, sondern auch die Determinantien von Komposita entfernt werden, sodass nur die Determinata in die Analyse mit einfließen. Dies ist ein wichtiger Schritt, der in jedem Fall erfolgen muss, denn die Kompositabildung ist im Deutschen ein hoch produktiver Prozess, der zu verfälschten Ergebnissen im Hinblick auf derivationelle Produktivität führen kann, wenn alle gebildeten Komposita als Token in die Berechnung einfließen.

5. **Betrachtung der 'einfachen Frequenzen':**

Im nächsten Schritt sollte der sprachwissenschaftliche Forscher eine Betrachtung der Anzahl der Token, Typen und Hapaxe vornehmen. Förderlich wäre dabei eine automatische Ermittlung der Typen und der Hapax Legomena.

In der vorliegenden Arbeit konnte anhand der beschriebenen Frequenzen recht schnell festgestellt werden, dass *-sal* ein gänzlich unproduktives Suffix ist. Demnach waren in diesem Fall keine weiteren Berechnungen mehr notwendig. Dies jedoch ist sicherlich nicht der Regelfall, sodass weitere Überlegungen hinzugezogen werden müssen.

6. **Anwendung von P und P^* :**

Baayens Maße P und P^* können – gerade bei ausreichender Datenmenge – zu soliden Ergebnissen führen, sodass es sinnvoll ist, sowohl die Produktivität im engen Sinne (P) als auch die globale Produktivität P^* zu betrachten.

Beide Methoden sind für synchrone und auch für diachrone Analysen geeignet. Probleme entstehen allerdings dann, wenn die zu Verfügung stehende Daten-

menge zu klein ist. In solchen Fällen ist die Hinzuziehung von LNRE-Modellen unabdingbar.

7. Anwendung von LNRE-Modellen:

Wie in der vorliegenden Arbeit skizziert, eröffnen LNRE-Modelle neue Zugänge zur Produktivitätsforschung: Durch Extrapolationen werden unterschiedlich große Datenmengen miteinander vergleichbar, und dies sowohl im Hinblick auf Baayens P , die Steigung der Wachstumskurve wie auch Typenfrequenz und die Anzahl der Hapax Legomena.

Des Weiteren können auch qualitative Beschreibungen der Wachstumskurven zur Interpretation hinzugezogen werden. Es bleibt wünschenswert, auch die 'population size' und die Schiefe des Frequenzspektrums zu betrachten, denn gerade durch Hinzuziehen von Alpha als Parameter lässt sich der Konflikt umgehen, nur die einmaligen Bildungen, die Hapax Legomena, als Neubildungen in die Berechnung mit einfließen lassen zu können.

Schließlich ist es durchaus denkbar, dass neu gebildete Formen auch höher frequent als einmalig in einem Text auftauchen, indem ein Sprecher sein neu gebildete Wort wiederholt. Weiterhin ermöglicht die Anwendung von Konfidenzintervallen eine Überprüfung möglicher Zufallsschwankungen.

8. Zusammenfassung der Ergebnisse und Beobachtungen:

Resümierend lässt sich festhalten, dass die Interpretation der morphologischen Produktivität immer anhand mehrerer Maße und integrativ erfolgen sollte. Ergänzend dazu sollten auch qualitative Aspekte wie die Betrachtung der Wachstumskurven in die Schlussfolgerungen mit einfließen.

Literaturverzeichnis

- Adelung, J. C. (1782). Von Neuen Wörtern durch die Ableitung. In J. C. Adelung (Hrsg.), *Magazin für die Deutsche Sprache*, S. 36–78.
- Altmann, G. (1985). Sprachtheorie und mathematische Modelle. In *SAIS Arbeitshefte*, Band 8, S. 1–13. Kiel.
- Altmann, G. (1988). *Wiederholungen in Texten*. Bochum: Brockmeyer.
- Altmann, G. und G. Wimmer (1995). A model of morphological productivity. In *Journal of quantitative linguistics*, Band 1, S. 212–216.
- Altmann, H. (1981). *Formen der „Herausstellung“ im Deutschen. Rechtsversetzung, Linksversetzung, Freies Thema und verwandte Konstruktionen*. Tübingen: Niemeyer (Linguistische Arbeiten 106).
- Andersen, H. (1980). Morphological change: Towards a typology. In J. Fisiak (Hrsg.), *Historical Morphology*, S. 1–50. Mouton.
- Anderson, S. R. (1988). Morphological change. In F. J. Niemeyer (Hrsg.), *Linguistics. The Cambridge Survey*, S. 324–362. Cambridge Univ. Press.
- Anderson, S. R. (1992). *A-morphous morphology*. Cambridge: Cambridge University Press.
- Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge, Mass.: MIT Press.
- Aronoff, M. und K. Fudeman (2005). *What is morphology?* Blackwell Publishing.
- Ascoop, K. (2005). Affixoidhungrig? Skitbra! Status und Gebrauch von Affixoiden im Deutschen und Schwedischen. *Germanistische Mitteilungen* 62, S. 17–28.
- Baayen, H. (1989). *A corpus-based approach to morphological productivity: statistical analysis and psycholinguistic interpretation*. Amsterdam: Ph.D. thesis, Free

- University.
- Baayen, H. (1992). Quantitative aspects of morphological productivity. In G. E. Booij und J. van Marle (Hrsg.), *Yearbook of morphology 1992*, S. 109–149. Dordrecht: Kluwer Academic Publishers.
- Baayen, H. (1993). On frequency, transparency, and productivity. In G. E. Booij und J. van Marle (Hrsg.), *Yearbook of morphology 1992*, S. 181–208. Dordrecht: Kluwer Academic Publishers.
- Baayen, H. (2001). *Word Frequency Distributions*. Dordrecht: Kluwer.
- Baayen, H. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Baayen, H. (2009). Corpus linguistics in morphology: Morphological Productivity. In A. Lüdeling und M. Kytö (Hrsg.), *Corpus linguistics. An international handbook*, Band 2, S. 899–919. Berlin: Mouton de Gruyter.
- Baayen, H. und R. Lieber (1991). Productivity and English derivation: a corpus-based study. *Linguistics* 29, S. 801–843.
- Baayen, H. und A. Renouf (1996). Chronicling the times: productive lexical innovations in an english newspaper. *Language* 72, S. 69–96.
- Baayen, R. H. (1994). Derivational productivity and text typology. *Journal of quantitative linguistics* 1, S. 16–34.
- Baayen, R. H., R. Piepenbrock und H. van Rijn (1993). The CELEX lexical database. (Release 1) [CD-ROM].
- Baldi, P. und W. Schmalstieg (1990). Morphological change. *Research guide on language change*, S. 347–364.
- Baroni, M. (2009). Distributions in text. In A. Lüdeling und M. Kytö (Hrsg.), *Corpus Linguistics: An international handbook*, Band 2, S. 803–821. Berlin: Mouton de Gruyter.
- Baroni, M. und S. Evert (2005). Testing the extrapolation quality of word frequency models. In P. Danielsson und M. Wagenmakers (Hrsg.), *Proceedings of Corpus Linguistics 2005, volume 1 of The Corpus Linguistics Conference Series*.
- Baroni, M. und S. Evert (2007). Words and echoes: Assessing and mitigating the non-randomness problem in word frequency distribution modeling. In *Proceedings of the*

- 45th Annual Meeting of the Association for Computational Linguistics*, S. 904–911. Prag.
- Barz, I. (1999). Neologie und Wortbildung. Zum Neuheitseffekt von Wortneubildungen. In W. Teubert (Hrsg.), *Neologie und Korpus*, S. 11 – 29. Tübingen: Narr.
- Bauer, L. (1983). *English word formation*. Cambridge [u.a.]: Cambridge Univ. Pr.
- Bauer, L. (1988). *Introducing linguistic morphology*. Edinburgh: Edinburgh University Press.
- Bauer, L. (2001). *Morphological productivity*. Cambridge: Cambridge University Press.
- Bauer, L. (2004). Adjectives, compounds and words. *Nordic Journal of English Studies* 3(1), S. 7–22.
- Bauer, L. (2005). Productivity: Theories. In P. Stekauer und R. Lieber (Hrsg.), *Handbook of word formation*, S. 315–333.
- Bauer, L. (2007). Derivational morphology. *Language and linguistics compass*, S. 196–210.
- Beard, R. (1998). Derivation. In A. Spencer und A. Zwicky (Hrsg.), *The handbook of morphology*. Blackwell Publishers.
- Best, K. (2001). *Quantitative Linguistik: Eine Annäherung. 2., überarbeitete und erweiterte Auflage*. Göttingen: Peust Gutschmid.
- Booij, G., C. Lehmann und J. Mugdan (2002). *Morphologie. Ein internationales Handbuch zur Flexion und Wortbildung*. Berlin: 1. Halbbd., de Gruyter.
- Bründl, M. E. (2001). Lexikalische Dynamik: Kognitvlinguistische Untersuchungen am englischen Computerwortschatz. *Linguistische Arbeiten* 443, S. 59–67.
- Butterworth, B. (1993). Lexical representation. In B. Butterworth (Hrsg.), *Language production*, S. 254–294. London: Academic Press.
- Cowie, C. (1999). *Diachronic word-formation: a corpus-based study of derived nominalizations in the history of English*. Cambridge, Univ., Ph.D. thesis.
- Cowie, C. und C. Dalton-Puffer (2002). Diachronic word-formation over time: Theoretical and methodological considerations. In J. D. Vera (Hrsg.), *A changing world of words. Studies in English historical lexicography, lexicology and semantics*, S. S. 410–437. Amsterdam: Ropodi.

- Cutler, A. (1981). Degrees of transparency in word formation. *Canadian Journal of Linguistics* 20, 73–77.
- De Jong, N. H., R. Schreuder und R. H. Baayen (2000). The morphological family size effect and morphology. *Language and Cognitive Processes* 15, S. 329–365.
- de Saussure, F. (1967). *Grundlagen der allgemeinen Sprachwissenschaft*. Berlin/New York.
- Debus, F. (1980). Onomastik. In H. Althaus, H. Henne und H. Wiegand (Hrsg.), *Lexikon der Germanistischen Linguistik*, S. 187–198. Tübingen: Niemeyer.
- Demske, U. (2000). Zur Geschichte der *ung*-Nominalisierung im Deutschen. Ein Wandel morphologischer Produktivität. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 122, S. 365–411.
- Demske, U. (2001). Über Wolfgang Motsch: Deutsche Wortbildung in Grundzügen. *Zeitschrift für Germanistische Linguistik* 29/1, S. 75–84.
- Donalies, E. (2005). *Die Wortbildung des Deutschen*. Studien zur deutschen Sprache; 27. Tübingen: Narr.
- Dörfert, R. (1994). *Die Substantivableitung mit -heit, -keit, -ida, -î im Frühneuhochdeutschen*. Berlin: de Gruyter.
- Drach, E. (1937). *Grundgedanken der deutschen Satzlehre*. 1. Auflage. Darmstadt: de Gruyter.
- Dressler, W. U. (1995). Wege des morphologischen Wandels und der Entstehung morphologischer Regeln. *Natürlichkeitstheorie und Sprachwandel*, S. 93–109.
- Eichinger, L. M. (2000). *Deutsche Wortbildung. Eine Einführung*. Tübingen: Narr.
- Eisenberg, P. (1998). *Grundriß der deutschen Grammatik. Band 1: Das Wort*. Stuttgart/Weimar: Metzler.
- Elsen, H. (2005). Deutsche Konfixe. *Deutsche Sprache* 33, S. 133–140.
- Erben, J. (1993). *Einführung in die deutsche Wortbildungslehre*. 3. neubearbeitete Auflage. Berlin: Schmidt.
- Erben, J. (2000). *Einführung in die deutsche Wortbildungslehre*. 4. aktualisierte und ergänzte Auflage. Berlin: Schmidt.
- Evert, S. (2004). A Simple LNRE model for random character sequences. In *Proceedings of JADT 2004*, S. 411–422.

- Evert, S. (2005). *The statistics of word cooccurrences: Word pairs and collocations*. Stuttgart: Institut für maschinelle Sprachverarbeitung: Dissertation.
- Evert, S. und M. Baroni (2007). zipfR: Word frequency modeling in R. In *ACL. The Association for Computer Linguistics*.
- Evert, S. und A. Lüdeling (2001). Measuring morphological productivity: Is automatic preprocessing sufficient? In P. Rayson, A. Wilson, T. McEnery, A. Hardie und S. Khoja (Hrsg.), *Proceedings of the Corpus Linguistics 2001 conference*, Lancaster, S. 167–175.
- Fandrych, C. (1993). Wortart, Wortbildungsart und kommunikative Funktion. Am Beispiel der adjektivischen Privativ- und Possessivbildung im heutigen Deutsch. *Reihe Germanistische Linguistik*.
- Fernández-Domínguez, J., A. Díaz-Negrillo und P. Štekauer (2007). How is low morphological productivity measured? *Atlantis, revista de la Asociación Española de Estudios Anglo-Norteamericanos*.
- Fleischer, W. (1982). *Wortbildung der deutschen Gegenwartssprache*. Tübingen: Max Niemeyer Verlag.
- Fleischer, W. und I. Barz (1995). *Wortbildung der deutschen Gegenwartssprache*. Tübingen: Max Niemeyer Verlag.
- Flury, R. (1964). *Struktur- und Bedeutungsgeschichte des Adjektiv-Suffixes -bar*. Winterthur.
- Frauenfelder, U. H. und R. Schreuder (1992). Constraining psycholinguistic models of morphological processing and representation: the role of productivity. In G. Booij und J. van Marle (Hrsg.), *Yearbook of Morphology 1991*. Dordrecht: Kluwer.
- Fuhrhop, N. (1998). *Grenzfälle morphologischer Einheiten*, Band Studien zur deutschen Grammatik. Bd.57. Tübingen: Stauffenburg.
- Gaus, W. (2001). *Dokumentations- und Ordnungslehre. Theorie und Praxis des Information Retrieval*. Berlin: Springer.
- Geyken, A. (2005). Das Wortinformationssystem des Digitalen Wörterbuchs der deutschen Sprache des 20. Jahrhunderts (DWDS). *BBAW Circular* 32, 40.
- Geyken, A. (2007a). Automatische linguistische Erschließung großer Textdatenbanken am Beispiel des DWDS. In *Über den Tag hinaus. Berliner Zeitungskonferenz 2005. 400 Jahre Zeitung*. Staatsbibliothek Berlin.

- Geyken, A. (2007b). The DWDS corpus: A reference corpus for the German language of the 20th century. In *Idioms and Collocations: Corpus-based Linguistic, Lexicographic Studies*. Continuum Press.
- Geyken, A. und W. Klein (2001). Projekt "Digitales Wörterbuch der deutschen Sprache des 20. Jh.". In *Jahrbuch der BBAW 2000*, S. 263–270. Akademie Verlag.
- Gippert, J. (1995). TITUS. Das Projekt eines indogermanischen Thesaurus. *LDV-Forum* 12(1), S. 35–47.
- Gippert, J. (2002). Der TITUS-Server: Grundlagen eines multilingualen Online-Retrieval-Systems. *Festschrift für Winfried Lenders Sprachwissenschaft, Computerlinguistik und Neue Medien* 4, S. 81–86.
- Givón, T. (1971). Historical syntax and synchronic morphology. An archaeologist's fieldtrip. *Chicago Linguistic Society* 7, S. 394–415.
- Gotti, M. (1994). „Ware“: from compounding to suffixation. *Historical English word-formation*.
- Habermann, M. (1994). *Verbale Wortbildung um 1500. Eine historisch-synchrone Untersuchung anhand von Texten Albrecht Dürers, Heinrich Deichlers und Veit Dietrichs*. Berlin: de Gruyter.
- Habermann, M. (2002). Sprachwandel im Licht diachroner und synchroner Wortbildung. In M. Habermann, P. O. Müller und H. H. Munske (Hrsg.), *Historische Wortbildung des Deutschen*, S. 41–57.
- Habermann, M., P. O. Müller und H. H. Munske (2002). *Historische Wortbildung des Deutschen*. Tübingen: Niemeyer.
- Harwood, F. und A. M. Wright (1956). Statistical study of English word formation. In G. Booij und J. van Marle (Hrsg.), *Language*, S. 260–273.
- Hay, J. (2001). Lexical frequency in morphology: Is everything relative? *Linguistics* 39.4, S. 1041–1070.
- Hay, J. (2003). *Causes and consequences of word structure*. New York/London: Routledge.
- Hay, J. und I. Plag (2004). What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language and Linguistic Theory*, S. 565–596.

- Hay, J. B. und R. H. Baayen (2002). Parsing and productivity. In G. E. Booij und J. van Marle (Hrsg.), *Yearbook of Morphology 2001*. Dordrecht: Kluwer Academic Publishers.
- Hay, J. B. und R. H. Baayen (2003). Phonotactics, parsing and productivity. *Rivista di Linguistica* 15.1, S. 99–130.
- Heine, B. (2003). Grammaticalization. In B. D. J. und Richard D. Janda (Hrsg.), *The handbook of historical linguistics*, S. 575–601. Oxford: Blackwell.
- Himmelmann, N. P. (2004). Lexicalization and grammaticalization. Opposite or orthogonal? In H. N. W. B. Bisang, W. (Hrsg.), *What makes grammaticalization? A Look from its fringers and components*, S. 21–42. Berlin/New York.
- Hopper, P. J. (1987). Emergent grammar. *Berkeley Linguistics Society*, S. 139–157.
- Hopper, P. J. und E. C. Traugott (2003). *Grammaticalization*. Cambridge University Press.
- Jespersen, O. (1942). *A modern English grammar on historical principles*. Copenhagen: Munksgaard.
- Joseph, B. (1998). Diachronic morphology. In A. Spencer und A. Zwicky (Hrsg.), *The handbook of morphology*, S. 351–373. Blackwell Publisher.
- Kiesewetter, J. (1987). *Entwicklungsprozesse in der deutschen Wortbildung, in Auschnitten verfolgt und dargestellt an den verbalen Bildungen mit den Elementen bei-, dar-, und -ob*. Leipzig Universität – Dissertation.
- Kiesewetter, J. (1991). Sprachwandel in der Wortbildung – dargestellt an den verbalen Bildungen mit dem Element bei-. *Beiträge zur Erforschung der deutschen Sprache* 10, S. 133–175.
- Klein, W. (2004a). Das Digitale Wörterbuch der deutschen Sprache des 20. Jahrhunderts (DWDS). In J. Scharnhorst (Hrsg.), *Sprachkultur und Lexikographie*, S. 281–311. Berlin: Peter Lang.
- Klein, W. (2004b). Vom Wörterbuch zum digitalen lexikalischen System. *Zeitschrift für Literaturwissenschaft und Linguistik* 136, S. 10–55.
- Kluge, F. (1925). *Abriss der deutschen Wortbildungslehre*. Halle (Saale): Niemeyer.
- Knüppel, A. (2001). Untersuchungen zum Zipf-Mandelbrot-Gesetz an deutschen Texten. S. 248–280. Göttingen: Peust und Gutschmidt.

- Koehn, E. und S. Koehn (1986). Apalai. In D. C. Derbyshire und G. K. Pullum (Hrsg.), *Handbook of amazonian languages*, S. 33–127. Berlin/New York/Amsterdam: Mouton de Gruyter.
- Kolb, H. (1985). Über das Suffix *-igkeit*. *Studien zur deutschen Grammatik* 25, S. 159–167.
- Köpke, K.-M. (2002). Die sogenannte i-Derivation in der deutschen Gegenwartssprache ein Fall für outputorientierte Wortbildung. *Zeitschrift für germanistische Linguistik*, S. 293–309.
- Krahe, H. und W. Meid (1969). *Germanistische Sprachwissenschaft III. Wortbildungslehre*. Berlin/New York: de Gruyter.
- Kruisinga, E. (1932). *A handbook of present-day English*, Band 5. Groningen: Noordhoff.
- Kunkel-Razum, K. und P. Eisenberg (2005). *Duden. Die Grammatik*. Mannheim/Leipzig/Wien/Zürich: Dudenverlag.
- Lehmann, C. (1991). Grammaticalization and related changes in contemporary German. In E. Traugott und B. Heine (Hrsg.), *Approaches to grammaticalization. Volume 2. Focus in theoretical and methodological issues*, S. S. 493–535. Amsterdam/Philadelphia: J. Benjamins.
- Lehmann, C. (1995a). Theory and method in grammaticalization [Theorie und Methode in der Grammatikalisierung]. *Zeitschrift für germanistische Linguistik* 32, S. 493–535.
- Lehmann, C. (1995b). *Thoughts on grammaticalization*. LINCOM studies in theoretical linguistics ; 1. LINCOM Europa.
- Lehmann, C. (2002). New reflections on grammaticalization and lexicalization. In G. Diewald (Hrsg.), *New reflections on grammaticalization and lexicalization*, S. 1–18. Benjamins.
- Lemnitzer, L. und T. Ule (2005). *Die Wortwarte. Wörter von heute und morgen. Eine Sammlung von Neologismen täglich aktualisiert*. <http://www.sfs.uni-tuebingen.de/~lothar/nw/>, 17.03.2008.
- Lemnitzer, L. und H. Zinsmeister (2006). *Korpuslinguistik*. Tübingen: Gunter Narr.
- Lender, W. und K.-P. Wegera (1982). Maschinelle Auswertung sprachhistorischer Quellen. Ein Bericht zur computerunterstützten Analyse der Flexionsmorphologie

- des Frühneuhochdeutschen. *Sprache und Information Bd. 3*.
- Lenders, W. (1992). Wörter zwischen Welt und Wissen. In R. Kuhlen (Hrsg.), *Experimentelles und praktisches Information Retrieval: Festschrift für Gerhard Lustig*, S. 263–276. Konstanz: Universitätsverlag Konstanz.
- Lexikon der Sprachwissenschaft (1985). Stuttgart: Kröner.
- Li, W. (1992). Random texts exhibit zipf's-law-like word frequency distribution. In *IEEE Transactions on Information Theory*, S. 1842–1845.
- Lieber, R. (1981). *On the organization of the lexicon*. Bloomington: Indiana Univ. Linguistics Club.
- Lieber, R. (1992). *Deconstructing morphology*. Chicago [u.a.]: Univ. of Chicago Press.
- Lobin, H. und L. Lemnitzer (2004). *Texttechnologie*. Tübingen: Stauffenburg.
- Lohde, M. (2006). *Wortbildung des modernen Deutschen. Ein Lehr- und Übungsbuch*. Tübingen: narr studienbücher.
- Lüdeling, A. (2009). Wortbildungswandel und produktivität. eine empirische studie zur nominalen -er-derivation im deutschen. In *Beiträge zur Geschichte der deutschen Sprache und Literatur (PBB)*, S. 333–339. Tübingen: Niemeyer 2005.
- Lüdeling, A. und S. Evert (2003). Linguistic experience and productivity: corpus evidence for fine-grained distinctions. In D. Archer, P. Rayson, A. Wilson und T. McE- nery (Hrsg.), *Proceedings of the Corpus Linguistics 2003 conference*, Lancaster, S. 475–483. UCREL.
- Lüdeling, A. und S. Evert (2004). The emergence of productive non-medical -itis: corpus evidence and qualitative analysis. In *Proceedings of the First International Conference on Linguistic Evidence*, Tübingen.
- Lüdeling, A. und S. Evert (2005). The emergence of productive non-medical -itis. corpus evidence and qualitative analysis. In S. Kepser und M. Reis (Hrsg.), *Linguistic Evidence. Empirical, Theoretical, and Computational Perspectives*, S. 351 – 370. Berlin/New York: Mouton de Gruyter.
- Lüdeling, A., S. Evert und U. Heid (2000). On measuring morphological productivity. *KONVENS-2000*, S. 215–220.
- Lüdeling, A., T. Poschenrieder und L. C. Faulstich (2005). DeutschDiachronDigital - Ein diachrones Korpus des Deutschen. *Jahrbuch für Computerphilologie 2004*.
- Lyons, J. (1977). *Semantics/2*, Band 2. Cambridge: University Press.

- Mandelbrot, B. (1953). An informational theory of the statistical structure of languages. In W. J. (Betterworth) (Hrsg.), *Communication Theory*, S. 486–502.
- Mandelbrot, B. (1962). On the theory of word frequencies and on related Markovian models of discourse. In R. Jakobson (Hrsg.), *Structure of Language and its Mathematical Aspects*, S. 190–219. Providence. RI.
- Marchand, H. (1969). *The categories and types of present-day English word-formation (= Handbücher für das Studium der Anglistik)*. München: Beck.
- Mater, E. (1970). *Rückläufiges Wörterbuch der deutschen Gegenwartssprache*. Leipzig: Straelener Manuskripte Verlag.
- Mayerthaler, W. (1981). *Morphologische Natürlichkeit*. Wiesbaden: Aula-Verlag GmbH.
- McQueen, J. M. und A. Cutler (1998). Morphology in word recognition. In A. Spencer und A. Zwicky (Hrsg.), *The Handbook of Morphology*, S. 406–427. Oxford: Blackwell.
- Meibauer, J. (1995a). Komplexe Präpositionen - Grammatikalisierung, Metapher, Implikatur und *division of pragmatic labour*. *Implikaturen: Grammatische und pragmatische Analysen (Linguistische Arbeiten 343)*, S. 47–74.
- Meibauer, J. (1995b). Wortbildung und Kognition. Überlegungen zum deutschen -er-Suffix. *Deutsche Sprache 23*, S. 97–123.
- Meillet, A. ([1912]1948). L'évolution des formes grammaticales. *Linguistique historique et linguistique générale.*, S. 130–148.
- Meinecke, E. (1994). Abstraktbildung im Althochdeutschen. Wege zu ihrer Erschließung. In *Studien zum Althochdeutschen*, Band 24, S. 198–201. Göttingen: Vandenhoeck Ruprecht.
- Michel, S. (2006). Vom 'Terminator' zum 'TORminator'. Die Wortbildungseinheit '-minator': Strukturelle und sozio-pragmatische Analysen. *Muttersprache*, S. 289–307.
- Motsch, W. (1983). Überlegungen zu den Grundlagen der Erweiterung des Lexikons. In R. Ruzicka (Hrsg.), *Überlegungen zur Semantik*, S. 101–119.
- Motsch, W. (1996). Affixoide. Sammelbezeichnung für Wortbildungsphänomene oder linguistische Kategorie? *Deutsch als Fremdsprache 33*, S. 160–168.
- Motsch, W. (1999). *Deutsche Wortbildung in Grundzügen*. Berlin: de Gruyter.

- Motsch, W. (2004). *Deutsche Wortbildung in Grundzügen. 2. überarbeitete Auflage*. Berlin: de Gruyter.
- Mrotzek, B. (2005). *Lexikon der bedrohten Wörter*. Reinbek: Rowohlt.
- Müller, P. O. (1993). *Substantivderivation in den Schriften Albrecht Dürers. Ein Beitrag zur Methodik historisch-synchroner Wortbildungsanalysen*, Band 1. Berlin/New York: de Gruyter.
- Müller, P. O. (2002). Historische Wortbildung im Wandel. In M. Habermann, P. O. Müller und H. H. Munske (Hrsg.), *Historische Wortbildung des Deutschen*, S. 1–13.
- Munske, H. (1990). Über den Wandel des deutschen Wortschatzes. In W. Besch (Hrsg.), *Deutsche Sprachgeschichte. Grundlagen, Methoden, Perspektiven*, S. 387–402.
- Munske, H. (2002). Wortbildungswandel. In M. Habermann, P. O. Müller und H. H. Munske (Hrsg.), *Historische Wortbildung des Deutschen*, S. 23–40.
- Nishimoto, E. (2006). *Measuring and comparing the productivity of Mandarin Chinese suffixes*. Journal of Computational Linguistics and Chinese Language Processing.
- Norde, M. (2002). The final stages of grammaticalization. Affixhood and beyond. In G. Diewald und I. Wischer (Hrsg.), *New reflection on grammaticalization*, S. S. 46–65.
- Nübling, D., A. Dammel, J. Duke und R. Szczepaniak (2006). *Historische Sprachwissenschaft des Deutschen. Eine Einführung in die Prinzipien des Sprachwandels*. Tübingen: Gunter Narr Verlag.
- Olsen, S. (1990). Zur Suffigierung und Präfigierung im verbalen Bereich des Deutschen. *Papiere zur Linguistik* 42, S. 31–48.
- Paul, H. (1886). *Prinzipien der Sprachgeschichte. 2. Aufl.* Halle: de Gruyter.
- Peuser, G. (1993). *Einführung in die Patholinguistik. Teil 1: Sprachwissenschaftliche Grundlagen*. Köln: Reader für das Sommersemester.
- Plag, I. (1999). *Morphological productivity. Structural constraints in English derivation*. Berlin/New York: Mouton de Gruyter.
- Plag, I. (2004). Productivity. *Encyclopedia of Language and Linguistics*.
- Plag, I., C. Dalton-Puffer und H. Baayen (1999). Morphological productivity across speech and writing. *ENGLISH LANGUAGE AND LINGUISTICS* 3.2.

- Prell, H.-P. (1991). *Die Ableitung von Verben aus Substantiven in biblischen und nichtbiblischen Texten des Frühneuhochdeutschen*. Frankfurt am Main: Lang.
- Prell, H.-P. und M. Schebben-Schmidt (1996). *Die Verbableitung im Frühneuhochdeutschen*. Berlin: de Gruyter.
- Rainer, F. (2000). Produktivitätsbeschränkungen. In G. Booij, C. Lehmann und J. Mugdan (Hrsg.), *Ein Handbuch zur Flexion und Wortbildung*, Band 1. Halbband, S. S. 877–885. Berlin/New York.
- Ruzicka, R. (1983). *Untersuchungen zur Semantik*, Band 22, *Studia grammatica*. Akad.-Verlag.
- Scarborough, D. L., C. Cortese und H. S. Scarborough (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance* 3, S. 1–17.
- Scarborough, R. (2001). Modifying Baayen: Can a corpus count provide a reasonable measure of morphological productivity? *paper for Morphology course, taught by Donca Steriade*.
- Scherer, C. (2003). Diachronic word formation: -er nominals in German. *Proceedings of CIL* 17.
- Scherer, C. (2005). *Wortbildungswandel und Produktivität. Eine empirische Studie zur nominalen -er-Derivation im Deutschen*. Tübingen: Niemeyer.
- Scherer, C. (2006). Was ist Wortbildungswandel? *Linguistische Berichte* 205, S. 4–28.
- Schlaefter, M. (1977). *Die Adjektive auf -isch in der deutschen Gegenwartssprache*. Heidelberg: Carl Winter.
- Schneider-Wiejowski, K. (2009). Sprachwandel anhand von Produktivitätsverschiebungen in der schweizerdeutschen Derivationsmorphologie. *Linguistik online* 38.
- Schreuder, R. und R. H. Baayen (1997). How complex simplex words can be. *Journal of Memory and Language* 37, S. 118–139.
- Schultink, H. (1961). Produktivität als Morphologisch Fenomeen. *Forum der Letteren* 2, 110–125.
- Seibicke, W. (1963). Wörter auf '-mäßig'. Sprachkritik und Sprachbetrachtung. *Muttersprache* 73, S. 33–73.
- Shippan, T. (1967). *Die Verbalsubstantive der deutschen Sprache der Gegenwart*. Leipzig.

- Sick, B. (2009). *Der Dativ ist dem Genitiv sein Tod. Das Allerneueste aus dem Irrgarten der deutschen Sprache*. Hamburg: Kierperheuer und Witsch.
- Sinclair, J. (1998). Korpustypologie. ein Klassifikationsrahmen. In W. Teubert (Hrsg.), *Neologie und Korpus*, S. S. 111–128. Tübingen: Narr.
- Spencer, A. und A. Zwicky (1998). *The handbook of morphology*. Oxford: Blackwell.
- Stekauer, P. H. (2005). *Handbook of word formation*. Studies in natural language and linguistic theory; 64. Springer.
- Stevens, C. M. (2005). Revisiting the affixoid debate. On the grammaticalization of the Word. *Grammatikalisierung im Deutschen*. 4(9), S. 71–83.
- Stolz, T. (1991). Agglulationstheorie und Grammatikalisierungsforschung. Einige alte und neue Gedanken zur Entstehung von gebundener Morphologie. In *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, Band 44, S. 325–338.
- Stricker, S. (2000). *Substantivbildung durch Suffixableitung um 1800 : untersucht an Personenbezeichnungen in der Sprache Goethes*. Heidelberg: Universitätsverlag Winter GmbH.
- Szczepaniak, R. (2009). *Grammatikalisierung im Deutschen. Eine Einführung*. Tübingen: narr Studienbücher.
- Taft, M. (1985). The decoding of words in lexical access: a review of the morphographic approach. In D. Besner, T. G. Waller und G. E. Mackinnon (Hrsg.), *Reading Research: Advances in Theory and Practice*. London: Academic Press.
- Teubert, W. (1998). *Neologie und Korpus*. Tübingen: Narr.
- Štekauer, P. (1998). *An onomasiological theory of English word-formation*. Amsterdam/Philadelphia: John Benjamins.
- van Marle, J. (1991). The relationship between morphological productivity and frequency : a comment on Baayen's performance-oriented conception of morphological productivity. *Yearbook of morphology 1*, S. 151–163.
- von Humboldt, W. (1963). *Schriften zur Sprachphilosophie. Werke in fünf Bänden. Band III*. Darmstadt: Wissenschaftliche Buchgesellschaft WBG.
- von Polenz, P. (1994). *Deutsche Sprachgeschichte vom Spätmittelalter bis zur Gegenwart: 17. und 18. Jahrhundert*. Berlin/New York: Walter de Gruyter.

- Wellmann, H. (1995). Die Wortbildung. *Duden. Grammatik der deutschen Gegenwartssprache. 4*, S. 399ff.
- Wellmann, H. (1997). Wortbildung im Sprachwandel. In R. Wimmer und F. Berens (Hrsg.), *Wortbildung und Phraseologie*, S. 65–87. Tübingen: Narr.
- Willmanns, W. (1899). *Deutsche Grammatik. 2. Abteilung. Wortbildung*. Strassburg.
- Wimmer, G. (2005). The type-token-relation. In G. R. G. P. Reinhard Köhler, Gabriel Altmann (Hrsg.), *Quantitative Linguistik - Quantitative Linguistics. Ein internationales Handbuch.*, S. 325–348. Berlin/ New York: de Gruyter.
- Wischer, I. (2008). Zum kategorialen Status von Derivationsaffixen im Rahmen von Grammatikalisierungsprozessen. In T. Stolz (Hrsg.), *Grammatikalisierung und grammatische Kategorien*, S. 135–146. Bochum: Brockmeyer.
- Zimmer, K. E. (1964). *Affix negation in English and other languages*. New York.
- Zipf, G. K. (1949). *Human behavior and the principle of least effort*. H. Cambridge (MA): Addison-Wesley.

Tabellenverzeichnis

4.1	Ergebnisse zur Produktivität englischer Affixe nach Baayen 2001 . . .	51
5.1	Zusammenfassung der Aussagen der linguistischen Literatur zur Produktivität der zu untersuchenden substantiv- und adjektivbildenden Suffixe	80
6.1	Anzahl der Wortbildungen mit den substantivbildenden Suffixen -tum, -heit/-keit, -nis, -ung und -sal der einzelnen Versuchspersonen.	93
6.2	Anzahl der produzierten Wortbildungen mit den adjektivbildenden Suffixen -lich, -ig, -mäßig, -bar, -sam, -isch, -haft und -los der einzelnen Versuchspersonen.	94
6.3	Typenhäufigkeit der substantivbildenden Suffixe -tum, -heit/-keit, -nis, -ung und -sal.	96
6.4	Typenhäufigkeit der adjektivbildenden Suffixe -lich, -ig, -mäßig, -bar, -sam, -isch,-haft und -los.	96
6.5	Frequenzen der Wortlisten zu den substantivbildenden Suffixen in absoluter Anzahl	97
6.6	Frequenzen der Wortlisten zu den adjektivbildenden Suffixen in absoluter Anzahl	98
6.7	Anzahl der mit substantivbildenden Suffixen produzierten Wörter, die nicht lexikalisiert sind	101
6.8	Anzahl der nicht-lexikalisierten Wörter der mit den adjektivbildenden Suffixen produzierten Wörter in Prozent	101
6.9	Zusammenfassung der Ergebnisse der substantivbildenden Suffixe -nis, -ung, -sal, -heit/-keit/-igkeit und -tum	102

6.10 Zusammenfassung der untersuchten Merkmale der adjektivbildenden Suffixe -bar, -sam, -haft, -lich, -los, -ig, -mäßig und -isch	103
7.1 Tokenhäufigkeit der substantivbildenden Suffixe -ung, -tum, -nis, -heit/-keit und -sal für die Zeiträume 1900, 1925, 1950, 1975 und 1999	106
7.2 Tokenhäufigkeit (normiert auf eine Million Token) der substantivbildenden Suffixe -ung, -tum, -nis, -heit/-keit und -sal für die Zeiträume 1900, 1925, 1950, 1975 und 1999	106
7.3 Tokenhäufigkeit der Allomorphe -heit und -keit für die Zeiträume 1900, 1925, 1950, 1975 und 1999	107
7.4 Tokenhäufigkeit der Allomorphe -heit und -keit für die Zeiträume 1900, 1925, 1950, 1975 und 1999	108
7.5 Typenhäufigkeit der substantivbildenden Suffixe -ung, -tum, -nis, -heit/-keit und -sal für die Zeiträume 1900, 1925, 1950, 1975 und 1999	108
7.6 Typenhäufigkeit der Allomorphe -heit und -keit für die Zeiträume 1900, 1925, 1950, 1975 und 1999	109
7.7 Typen-Token-Relationen (TTR) der substantivbildenden Suffixe -ung, -tum, -nis, -heit/-keit und -sal für die Zeiträume 1900, 1925, 1950, 1975 und 1999	110
7.8 TTRs der Allomorphe -heit und -keit für die Zeiträume 1900, 1925, 1950, 1975 und 1999	111
7.9 Häufigkeiten der Hapax Legomena der substantivbildenden Suffixe -ung, -tum, -nis, -heit/-keit und -sal für die Zeiträume 1900, 1925, 1950, 1975 und 1999	112
7.10 Häufigkeiten der Hapax Legomena der der Allomorphe -heit und -keit für die Zeiträume 1900, 1925, 1950, 1975 und 1999	113
7.11 P der substantivbildenden Suffixe -ung, -tum, -nis, -heit/-keit und -sal für die Zeiträume 1900, 1925, 1950, 1975 und 1999	114
7.12 P der der Allomorphe -heit und -keit für die Zeiträume 1900, 1925, 1950, 1975 und 1999	114
7.13 Tokenhäufigkeit der adjektivbildenden Suffixe -ig, -isch, -lich, -mäßig, -los, -haft, -sam und -bar für die Zeiträume 1900, 1925, 1950, 1975 und 1999	127

7.14	Tokenhäufigkeit (normiert auf 1 Million Token) der adjektivbildenden Suffixe -ig, -isch, -lich, -mäßig, -los, -haft, -sam und -bar für die Zeiträume 1900, 1925, 1950, 1975 und 1999	128
7.15	Typenhäufigkeiten für die adjektivbildenden Suffixe -ig, -isch, -lich, -mäßig, -los, -haft, -sam und -bar für die Zeiträume 1900, 1925, 1950, 1975 und 1999	129
7.16	Typen-Token-Relationen der adjektivbildenden Suffixe -ig, -isch, -lich, -mäßig, -los, -haft, -sam und -bar für die Zeiträume 1900, 1925, 1950, 1975 und 1999	130
7.17	Häufigkeiten der Hapax Legomena der adjektivbildenden Suffixe -ig, -isch, -lich, -mäßig, -los, -haft, -sam und -bar für die Zeiträume 1900, 1925, 1950, 1975 und 1999	131
7.18	P der adjektivbildenden Suffixe -ig, -isch, -lich, -mäßig, -los, -haft, -sam und -bar für die Zeiträume 1900, 1925, 1950, 1975 und 1999	132
7.19	Zusammenfassung der Ergebnisse für die substantivbildenden Suffixe	148
7.20	Zusammenfassung der Ergebnisse für die adjektivbildenden Suffixe	151
7.21	Zusammenfassung der Ergebnisse aus diachroner Sicht für die substantivbildenden Suffixe	154
7.22	Zusammenfassung der Ergebnisse aus diachroner Sicht für die adjektivbildenden Suffixe	154
8.1	Ein fiktives Frequenzspektrum für -nis-Derivate	164
8.2	Anordnung der -nis-Derivate nach Rank/Frequenz	165
8.3	Frequenzspektrum der -nis-Derivate	165
8.4	Auflistung der Extrapolations-Ergebnisse (V, V1, P) der substantivbildenden Suffixe nach dem finiten Zipf-Mandelbrot-Modell anhand der Daten des Jahres 1999 im DWDS	178
8.5	Auflistung der Extrapolations-Ergebnisse (V, V1, P) der substantivbildenden Suffixe nach dem finiten Zipf-Mandelbrot-Modell auf Basis der Daten des Jahres 1999 im DWDS	182
8.6	Auflistung der Extrapolations-Ergebnisse für das Suffix -ung nach dem fZM-Modell (V, V1, P) anhand der Daten des DWDS'	186
8.7	Berechnete S-Werte auf Basis des finiten Zipf-Modells der substantivbildenden Suffixe -ung, -tum, -nis und -heit/-keit für die Zeiträume 1900, 1925, 1950, 1975 und 1999	188

Abbildungsverzeichnis

2.1	Wortschatzerweiterung nach Bründl (2001, 64)	14
3.1	Phasen der Grammatikalisierung nach Lehmann (1995, 12)	29
3.2	Das Affixoid als Brücke zwischen Lexem und Affix. Modifiziert nach Nübling et al. (2007, 71)	31
3.3	Munskes Wortbildungstypen (2002, 26)	33
3.4	Messbereich der Produktivität in Bezug auf Grammatikalisierungsprozessen bei Affixen	41
4.1	Wachstumskurven der Hapax Legomena (V1) für einen prototypisch unproduktiven und einen produktiven Wortbildungsprozess	50
4.2	Globale Produktivität für einige englische Suffixe nach Baayen (1999, 124)	53
4.3	Ein Idealfall für eine Steigerung/einen Abfall der globalen Produktivität im Laufe eines Untersuchungszeitraums	54
4.4	Produktive und unproduktive Affixe im Zusammenhang mit der globalen Produktivität	55
4.5	Probleme im Zusammenhang mit der globalen Produktivität	56
6.1	Anzahl der Wortbildungen mit den substantivbildenden Suffixen <i>-tum</i> , <i>-heit/-keit</i> , <i>-nis</i> , <i>-ung</i> und <i>-sal</i> der einzelnen Versuchspersonen.	93
6.2	Anzahl der Wortbildungen mit den adjektivbildenden Suffixen <i>-lich</i> , <i>-ig</i> , <i>-mäßig</i> , <i>-bar</i> , <i>-sam</i> , <i>-isch</i> und <i>-los</i> der einzelnen Versuchspersonen	95
7.1	Globale Produktivität von <i>-ung</i> im 20. Jahrhundert	115

7.2	Globale Produktivität von <i>-nis</i> im 20. Jahrhundert	116
7.3	Globale Produktivität von <i>-tum</i> im 20. Jahrhundert	117
7.4	Globale Produktivität von <i>-heit</i> und <i>-keit</i> im 20. Jahrhundert	118
7.5	Globale Produktivität von <i>-heit</i> im 20. Jahrhundert	119
7.6	Globale Produktivität von <i>-keit</i> im 20. Jahrhundert.	120
7.7	Globale Produktivität der substantivbildenden Suffixe im Jahre 1900 .	121
7.8	Globale Produktivität der substantivbildenden Suffixe im Jahre 1925 .	122
7.9	Globale Produktivität der substantivbildenden Suffixe im Jahre 1950 .	123
7.10	Globale Produktivität der substantivbildenden Suffixe im Jahre 1975 .	124
7.11	Globale Produktivität der substantivbildenden Suffixe im Jahre 1999	125
7.12	Schematische Zusammenfassung der globalen Produktivität der Substantivbildenden Suffixe im Jahr 1999	126
7.13	Globale Produktivität von <i>-ig</i> im 20. Jahrhundert	134
7.14	Globale Produktivität von <i>-isch</i> im 20. Jahrhundert	135
7.15	Globale Produktivität von <i>-lich</i> im 20. Jahrhundert	136
7.16	Globale Produktivität von <i>-mäßig</i> im 20. Jahrhundert	137
7.17	Globale Produktivität von <i>-los</i> im 20. Jahrhundert	138
7.18	Globale Produktivität von <i>-haft</i> im 20. Jahrhundert	139
7.19	Globale Produktivität von <i>-sam</i> im 20. Jahrhundert	140
7.20	Globale Produktivität von <i>-bar</i> im 20. Jahrhundert	141
7.21	Globale Produktivität der adjektivbildenden Suffixe im Jahre 1900 . .	142
7.22	Globale Produktivität der adjektivbildenden Suffixe im Jahre 1925 . .	143
7.23	Globale Produktivität der adjektivbildenden Suffixe im Jahre 1950 . .	144
7.24	Globale Produktivität der adjektivbildenden Suffixe im Jahre 1975 . .	145
7.25	Globale Produktivität der adjektivbildenden Suffixe im Jahre 1999 . .	146
7.26	Schematische Zusammenfassung der globalen Produktivität der adjektivbildenden Suffixe <i>-bar</i> , <i>-sam</i> , <i>-haft</i> , <i>-lich</i> , <i>-los</i> , <i>-ig</i> , <i>-mäßig</i> und <i>-isch</i> im Zeitraum von 1900 bis 1999	147

8.1	Ausschnitt der Token-Frequenzliste für das Suffix <i>-ung</i> , die mithilfe des DWDS ermittelt werden konnten	158
8.2	Wachstumskurven der Hapax Legomena für einen prototypisch unproduktiven und einen produktiven Wortbildungsprozess	159
8.3	Beispiel für eine schematische Entwicklung der morphologischen Produktivität eines Affixes (begonnen mit seiner Entstehung als freies Lexem)	161
8.4	Zufallsschwankungen in der Produktivität eines Affixes	163
8.5	Prototypische Frequenzspektren und Skizzierungen der jeweiligen Schiefen der Typenverteilungen (rote Linien)	168
8.6	Frequenzspektrum auf Basis der Datenreihe des Suffixes <i>-ung</i> im DWDS im Jahre 1999	170
8.7	Typen- und Hapaxwachstum (Extrapolation) für das Suffix <i>-ung</i> . . .	172
8.8	Typen- und Hapaxwachstum (Extrapolation) für das Suffix <i>-ung</i> nach dem finiten Zipf-Mandelbrot-Modell	174
8.9	Vergleich der drei LNRE-Modelle ZM, fZM und GIGP	175
8.10	Globale Produktivität der substantivbildenden Suffixe im Jahre 1999 anhand der extrapolierten Daten durch das finite Zipf-Mandelbrot-Modell	180
8.11	Veränderungen der globalen Produktivität der substantivbildenden Suffixe im Jahre 1999 anhand der extrapolierten Daten durch das finite Zipf-Mandelbrot-Modell	181
8.12	Globale Produktivität der adjektivbildenden Suffixe im Jahre 1999 anhand der extrapolierten Daten mithilfe des finiten Zipf-Mandelbrot-Modells	184
8.13	Veränderungen der globalen Produktivität der adjektivbildenden Suffixe im Jahre 1999 anhand der extrapolierten Daten nach dem finiten Zipf-Mandelbrot-Modell	185
8.14	Darstellung der globalen Produktivität von <i>-ung</i> für das 20. Jahrhundert mithilfe von Konfidenzintervallen	187

8.15	Extrapolation des Typenwachstums für das Suffix <i>-ung</i> als Prognose bis zu 10.000 Token	189
8.16	Extrapolation des Typenwachstums für das Suffix <i>-ung</i> auf Basis eines Inputs von 100.000 Token	190
8.17	Extrapolation des Typenwachstums für das Suffix <i>-nis</i> auf Basis eines Inputs von 10.000 Token	191
8.18	Extrapolation des Typenwachstums für das Suffix <i>-tum</i> auf Basis eines Inputs von 10.000 Token	192
8.19	Extrapolation des Typenwachstums für das Suffix <i>-tum</i> auf Basis eines Inputs von 100.000 Token	192
8.20	Extrapolation für das Suffix <i>-ig</i> als Prognose auf Basis eines Inputs von 10.000 Token	193
8.21	Extrapolation für das Suffix <i>-ig</i> als Prognose auf Basis eines Inputs von 100.000 Token	194
8.22	Extrapolation für das Suffix <i>-lich</i> als Prognose auf Basis eines Inputs von 10.000 Token	194
8.23	Extrapolation für das Suffix <i>-lich</i> als Prognose auf Basis eines Inputs von 100.000 Token	195
8.24	Extrapolation für das Suffix <i>-mäßig</i> als Prognose auf Basis eines Inputs von 10.000 Token)	196
8.25	Extrapolation für das Suffix <i>-mäßig</i> als Prognose auf Basis eines Inputs von 100.000 Token)	197
8.26	Extrapolationskurven von <i>-nis</i> anhand der Daten von 1900 (schwarz) und 1999 (rot)	198
8.27	Extrapolationskurven von <i>-ig</i> anhand der Daten von 1900 (schwarz) und 1999 (rot)	199

Versicherung

Hiermit versichere ich, dass ich meine Arbeit ohne unzulässige – insbesondere dem Betreuer unbekannte Hilfe – selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

Des Weiteren versichere ich auch, dass die Drucklegung der Arbeit vom Promotions-Ausschuss genehmigt ist und die elektronische Version mit der genehmigten und vervielfältigten Druckfassung vollständig übereinstimmt.

Bielefeld, den 8. März 2012

Karina Schneider-Wiejowski