

Marek Konopka 

# Grundlegende Aspekte der Wortbildung

**Abstract** This chapter begins with a sketch of the specifics of our approach, an overview of the contents of the chapters on word formation and some methodological notes. It then discusses the general characteristics of word formations and of their overall inventory, comparing word formations to primary words. Furthermore, the chapter explores the relative frequencies of word formations in different vocabulary areas and traces the word formation profiles of individual parts of speech. Finally, it compiles the characteristic word formation rules for different parts of speech.

**Keywords** word formation, simple word, complex word, productivity, part of speech, word formation types

## 1 Fragestellung und Methodisches

Die deskriptiv-grammatische Perspektive auf die Wortbildung<sup>1</sup> verlangt es, dass wir uns auf die Darstellung von Regularitäten konzentrieren und nicht auf die Behandlung von Idiosynkrasien. Dazu gehört es, dass ins Zentrum der Beschreibung Klassen von Wörtern gerückt werden, die auf relativ hohen Abstraktionsebenen definiert sind, und nicht etwa Bildungen auf der Basis eines bestimmten Lexems oder einiger weniger Lexeme eines bestimmten Wortfeldes. In diesem Sinne nehmen wir eine regelorientierte und nicht eine lexemorientierte Perspektive ein. Dennoch bleiben bei Kategorisierungen – etwa in primäre Wörter und Wortbildungen – lexembasierte Entscheidungen nicht aus. Die beiden Perspektiven sind damit nicht strikt getrennt. Allerdings wird die lexemorientierte Sicht nur so weit verfolgt, wie es für die regelorientierte Beschreibung notwendig ist.

Bei unserer Beschreibung von Wortbildungen können verschiedene Abstraktionsebenen unterschieden werden. Auf der obersten Ebene werden

1 Zum Wesen der Wortbildung als Gegenstand der Grammatik vgl. z.B. Fleischer & Barz (2012: 1–3), Eisenberg (2013: 201–208), Dudengrammatik (2016: 645–649).

Wortbildungen<sup>2</sup> nicht weiter analysiert, sondern als Gesamtgruppe den übrigen (nicht „wortgebildeten“) Wörtern gegenübergestellt. Die weiteren wichtigen Beschreibungsebenen sind – mit absteigendem Abstraktionsgrad:

- die Ebene der **Wortbildungsarten**, auf der wortartenübergreifende Bauweisen unterschieden werden können wie Komposition, Derivation oder Konversion (vgl. z. B. Fleischer & Barz 2012: 83–90),
- die Ebene der **Wortbildungsregeln**, auf der die Wortbildungsprodukte und ihre Bausteine kategorial spezifiziert werden, oft in Bezug auf Wortart oder Affixkategorie wie in den schematischen Darstellungen SBST + SBST → SBST<sup>3</sup> (z. B. *Umsatz|steuer*) oder ADJ + SUFF → SBST (z. B. *Aufmerksam|keit*),
- die Ebene der **Wortbildungsmuster**, auf der die Wortbildungsprodukte teils kategorial, teils formal spezifiziert bzw. als individuelle Ausdrücke angegeben werden wie in der Formulierung „Adverbbildungen auf -weise“ (*behelfs|weise, haufen|weise* etc.) oder in der entsprechenden schematischen Darstellung SBST + *weise* → ADV.

Unsere empirische Exploration von Wortbildungsphänomenen orientiert sich vor allem an folgenden Fragen:

1. Welche grundlegenden Regularitäten sind in der Wortbildung des Gegenwartsdeutschen zu beobachten?
2. Wie verteilen sie sich auf verschiedene Klassen von Wörtern?
3. Welche Wortbildungsarten und Wortbildungsregeln sind in den behandelten Klassen von Wörtern besonders frequent?
4. Welche Wortbildungsmuster sind besonders produktiv und wie ist dies zu erklären?
5. Welche sprachlichen und außersprachlichen Variablen beeinflussen die Realisierung produktiver Wortbildungsmuster?
6. Welche Regeln stehen hinter den relevanten Wortbildungsmustern, welchen Wortbildungsarten entsprechen die Regeln?

Die Beschreibung betrifft dabei ausdrücklich nur das Gegenwartsdeutsche. Jedoch macht die Erklärung von Regularitäten im Aufbau von Ausdrücken, die als Wortbildungen kategorisiert werden, gelegentlich auch kurze Exkurse in die Sprachgeschichte notwendig.

2 *Wortbildung* wird hier also nicht nur für die verschiedenen Aspekte des Wortaufbaus und für die linguistische Disziplin gebraucht, sondern auch für ein Wort, das infolge von Wortbildungsprozessen entstanden ist. Letzteres wird zuweilen verdeutlichend als *Wortbildungsprodukt* bezeichnet.

3 SBST = Substantiv, ADJ = Adjektiv, ADV = Adverb, SUFF = Suffix.

Eine ausführliche Behandlung aller Bereiche der deutschen Wortbildung ist in dieser Grammatik nicht möglich.<sup>4</sup> Wir beschränken uns auf einen empirisch fundierten Gesamtüberblick in den Kapiteln „Grundlegende Aspekte der Wortbildung“ und „Wortbildungsarten“ und widmen uns im Anschluss vertiefenden und auf besondere Detailschärfe abzielenden Beschreibungen ausgewählter Spezialbereiche.

Im vorliegenden Kapitel werden die allgemeinen Charakteristika der einschlägigen Wortbildungen und ihres Gesamtinventars vorgestellt, die relativen Häufigkeiten von Wortbildungen in verschiedenen Wortschatzbereichen erörtert, die Wortbildungsprofile einzelner Wortarten nachgezeichnet und die für einzelne Wortarten charakteristischen Wortbildungsregeln genannt (Fragen 1 bis 3 oben). Das darauffolgende Kapitel „Wortbildungsarten“<sup>5</sup> liefert eine formal-kompositionale Beschreibung der Wortbildungsarten, ergänzt um die Ermittlung wortartenübergreifender Regularitäten und um die Erörterung von Prinzipien der semantisch-relationalen Interpretation komplexer Wörter (Fragen 3 bis 4). Die weiteren Kapitel behandeln einerseits besonders Variationsträchtiges, andererseits wenig Untersuchtes und schwer Klassifizierbares (Fragen 4 bis 6). So wird im Kapitel „Substantivkomposition und Fugenelemente“ ein Phänomenbereich aufbereitet, der besonders stark durch ein Nebeneinander verschiedener formaler Möglichkeiten gekennzeichnet ist. In den Kapiteln „Bau von und Umbau zu Adverbien“ und „Desubstantivische Zweitglieder in der Wortbildung des Adverbs“ werden wiederum Wortbildungsregeln und -muster behandelt, die sich mit den Wortbildungsarten wie Komposition, Derivation und Konversion in ihrem klassischen Verständnis nur schwer in Einklang bringen lassen. Im Kapitel „Partikel- und Präfixverben und ihre Komplemente“ schließlich werden Wortbildungsphänomene beschrieben, deren Auswirkungen weit in die Syntax hineinreichen.

In methodischer Hinsicht haben die Ausführungen ihren Ausgangspunkt in Betrachtungen quantitativer Verhältnisse zwischen Klassen unterschiedlich gebildeter Wörter. Besonders wenn sich diese Betrachtungen auf das gesamte Untersuchungskorpus (ca. 7 Mrd. Tokens, davon knapp 6 Mrd. Worttokens) beziehen, ist mit einer gewissen Unschärfe zu rechnen. Mit einer derart großen Datenmenge kann man nämlich nur dann effektiv arbeiten, wenn Ergebnisse automatischer Annotationen genutzt werden. Die Güte solcher Annotationen lässt sich in einem großen Korpus allerdings nur stichprobenartig überprüfen. Korrekturen können

4 Für umfassende Darstellungen des Gegenwartsdeutschen vgl. u. a. Deutsche Wortbildung (1973–1992, Bd. 1–5), Motsch (1999), Eichinger (2000), Donalies (2002), Fleischer & Barz (2012), nicht unerwähnt bleiben soll hier auch die online verfügbare, lexikographische Quelle *elexiko* (<https://www.owid.de/docs/elex/start.jsp>, abgerufen am 18.6.2020, vgl. Haß & Klosa 2015). Als wichtige ältere Werke sind vor allem Grimm (1826), Wilmanns (1899), Paul (1920) und Henzen (1965) zu nennen.

5 In dieser Ausgabe der „Bausteine einer Korpusgrammatik des Deutschen“ nicht mehr enthalten.

oft nur punktuell vorgenommen werden, sodass meist ein Recall- und evtl. auch ein Präzisionsverlust in Kauf genommen werden muss. Das Gesamtkorpus wird daher vor allem zur Exploration der Phänomenvielfalt und für globale Hypothesen zur Reichweite und Produktivität von Wortbildungsregeln genutzt. Bei spezielleren Fragestellungen werden öfter Teilkorpora herangezogen, und zwar

- das Korpus *40k*, in dem 40.000 zufällig ausgewählte Sätze aus dem Gesamtkorpus zusammengestellt wurden (750.643 Tokens) und
- die zufällig zusammengestellten Achtel des Gesamtkorpus (Korpora *8th-1* bis *8th-8*, je über 877 Mio. Tokens).

Manche komplexen Korpusrecherchen sind erst durch die Benutzung dieser Teilkorpora in einer vertretbaren Zeit durchführbar. Ihre Verwendung macht es möglich, die Qualität der Recherchen zu steigern, da die Ergebnisse automatischer Suchen auf Umfänge reduziert werden, bei denen sie genauer überprüft und im Bedarfsfall korrigiert bzw. durch Nachrecherchen ergänzt werden können. Im Folgenden machen wir im Allgemeinen Gebrauch von der TreeTagger<sup>6</sup>-basierten Lemmatisierung, einer Wortbildungsannotation, die auf den Wortbildungsanalysen von Canoo<sup>7</sup> aufbaut, und den Wortarten-Annotationen von TreeTagger und Canoo. Da sich unsere Darstellung der substantivischen Wortbildungen auf die den Kern dieses Bereichs bildenden Appellative beschränkt, wird außerdem bei den Recherchen grundsätzlich mithilfe von Stanford NER (Finkel et al. 2005) eine Einschränkung auf Nicht-Eigennamen vorgenommen.

Im Teilkorpus *40k*, das im vorliegenden Kapitel eine besondere Rolle spielt, finden sich der Canoo-Klassifizierung zufolge 623.130 Wörter. Nach Ausschluss von Eigennamen, von Wörtern, deren Wortart nicht erkannt wurde oder die falsch bzw. gar nicht lemmatisiert wurden, sowie nach manuellen Korrekturen<sup>8</sup> werden 558.374 Tokens berücksichtigt, die sich auf 48.901 TreeTagger-Lemmata verteilen. Bei diesen Daten wurde die Wortbildungsannotation an neuralgischen Stellen überprüft. Vor allem wurden alle Lemmata in Augenschein genommen, die Canoo nicht als Wortbildungen analysierte. Handelte es sich bei einem Lemma doch zweifelsfrei um eine synchron analysierbare Wortbildung, wurde es manuell als solche annotiert.<sup>9</sup>

6 Vgl. Schmid (1995) – unterstützend wurden ein Lexikon (<https://github.com/rsling/cow/blob/master/src/de/data/lexicon>, abgerufen am 7.6.2020) aus dem COW-Webkorpusprojekt (Schäfer & Bildhauer 2012, Schäfer 2015) und ein mithilfe von SMOR (Schmid et al. 2004) generiertes Ad-hoc-Lexikon verwendet.

7 Siehe <http://canoonet.eu/services/WordformationRules/ueberblick/index.html?lang=de> (abgerufen am 6.12.2019).

8 Manuell wurden weitere Eigennamen ausgeschlossen und Kardinalzahlwörter annotiert, die in der Wortartenklassifizierung von Canoo nicht erfasst wurden.

9 Solche Lemmata waren z.B. *Orkanbö*, *abrechenbar*, *gestreng*, *beprobten*, *derweil*, *sowas*.

## 2 Der Beschreibungsgegenstand in Theorie und Praxis

Bei der Bestimmung des Gegenstands der linguistischen Disziplin Wortbildung ist eine Abgrenzung der zu fokussierenden Bildungen von sonstigen Wörtern gefragt. Hilfreich ist hier eine Unterscheidung von primären und sekundären Wörtern (vgl. Dudengrammatik 2016: 645, Meibauer et al. 2007: 33–34). Erstere sind nicht auf andere sprachliche Einheiten zurückzuführen, letztere hingegen bauen auf anderen sprachlichen Einheiten auf und sind damit Wortbildungen. Diese Unterscheidung ist in der Praxis manchmal schwer anzuwenden, weil ältere Wortbildungen nicht immer als solche erkannt werden. Das Wort *heute* etwa wird von Fachleuten auf die althochdeutsche Phrase *hiu tagu* ‚an diesem Tag‘ zurückgeführt (vgl. z.B. DWB: Lemma „heute“). Für „normale“ Sprachbenutzer/innen ist die Wortbildungs­vergangenheit von *heute* aber nicht mehr zu erkennen. Verdunkelte Wortbildungen sind auch *Adler*, *Welt*, *erst*, *nein*<sup>10</sup> und viele andere Ausdrücke, die auf den ersten Blick wie primäre Wörter aussehen. Überhaupt schrumpft die Gruppe der Kandidaten für den Status eines primären Wortes immer weiter, je genauer man den Ursprung der Wörter erforscht.

Solche alten, kaum oder gar nicht mehr als analysierbar erkennbaren Bildungen stehen jedoch nicht im Fokus von Beschreibungen der gegenwarts­sprachlichen Wortbildung. In diesen wird vielmehr eine „synchrone Erklärung des gegenwärtigen Wortbildungssystems“ (Fleischer & Barz 2012: 4) angestrebt. Das kann man so verstehen, dass in den zu behandelnden Ausdrücken Bausteine erkennbar sein sollen, die heute (das meint *synchron* in diesem Fall) interpretierbar sind und gleichzeitig nach heute nachvollziehbaren Regeln verbaut werden. Mit einer derartigen Vorgabe neigt man dazu, diejenigen Wortbildungen behandeln zu wollen, die von unseren Zeitgenoss/inn/en als solche empfunden werden. Das kann aber zu keiner klaren Konturierung des Beschreibungs­gegenstands führen. Das Wortbildungswissen und die wortbildungsanalytischen Fähigkeiten Einzelner sind nämlich sehr unterschiedlich, und gleichzeitig ist es unklar, welche Sprachbenutzer/innen die Referenzgruppe bilden sollen: Manche erkennen schon in *heute* eine Wortbildung, einige mehr, jedoch bei Weitem nicht alle, werden in *Fall*, *Wurf*, *statt* und *während*<sup>11</sup> eine Wortbildung sehen, ein breiter Konsens, dass es sich um Wortbildungen handelt, wird sich jedoch womöglich nur in Bezug auf eine zu sehr eingeschränkte Gruppe besonders transparenter Bildungen einstellen, die aus mehreren Konstituenten bestehen, welche entweder Wörtern gleichen (z.B. *Haus|tür*, *Schnell|zug*, *honig|gelb*, *klein|kariert*,

10 Vgl. Nübling et al. (2017: 185) und entsprechende Einträge in Pfeifer (1993).

11 Entstanden wohl durch Umdeutung syntaktischer Verhältnisse: *während der Zeit* < (in) *währender Zeit* (vgl. Lemma „während“ in Pfeifer 1993 und DWB).

nach|dem, so|mit) oder häufige Wortbildungsmorpheme darstellen (z. B. *Zeit|ung*, *ver|arbeiten*, *Ge|red|e*).

Indessen könnte postuliert werden, einschlägig seien bei der Zielsetzung „synchrone Erklärbarkeit“ diejenigen Wortbildungen, die sich sinnvoll und vollständig in solche sprachlichen Einheiten (Wörter, Morpheme, Fugenelemente u. ä.) analysieren lassen, die heute regulär verwendet werden. Eine korpusorientierte Ideallösung würde dann nicht auf das individuell variierende Sprachempfinden setzen, sondern versuchen, die Korpusdaten sprechen zu lassen. Ein Kriterium für eine sichere Klassifizierung einer Wortbildung als einschlägig wäre etwa der Befund, dass sich jeder ihrer Bausteine in der gleichen Bedeutung bzw. Funktion frei oder innerhalb von anderen Lexemen im Korpus wiederfindet. Einen entsprechenden intelligenten Abgleich von Korpuswörtern automatisch vorzubereiten und durchzuführen erscheint für unsere Zwecke aber als zu aufwendig, und „manuell“ ist er für größere Datenmengen nicht durchführbar.

Übrigens wären bei dieser Lösung Wortbildungen mit unikalen lexikalischen Morphemen ausgeschlossen worden. Betroffen hätte dies bekannte Fälle wie *Brom|beere*, *Lind|wurm* oder *Un|flat*, aber auch viele weniger bekannte Fälle wie *Mein|eid*, *gewieft*, *reüss|ieren*. Unikale Morpheme treten in genau einer Verbindung auf<sup>12</sup> und sind in diesem Sinne unproduktiv. Da ihre Bedeutung verblasst ist, können alle oben genannten Fälle als synchron nicht vollständig erklärbar betrachtet werden. Aber eigentlich sollte die Bestimmung der einschlägigen Wortbildungen im Sinne unserer regelorientierten Perspektive dazu dienen, dasjenige sprachliche Material zu sichern, in dem heute wirksame Regularitäten bzw. erkennbare Wortbildungsregeln zu finden sind, und die genannten Wörter mit unikalen Morphemen scheinen gerade nach solchen Regeln aufgebaut zu sein. Eine Lösung, die sie daher doch mit einbezieht, würde allerdings nicht mehr auf eine synchrone Erklärbarkeit der Bildungen abzielen, sondern nur auf ihre synchrone Analysierbarkeit. Bei dieser neuen Lösung wäre es für die Einschlägigkeit einer mehrgliedrigen Wortbildung unter Umständen ausreichend, wenn nur **eine** ihrer unmittelbaren Konstituenten heute regulär<sup>13</sup> vorkäme. Allerdings lässt sich auch diese Lösung korpuslinguistisch nicht direkt umsetzen.

Vor dem Hintergrund solcher theoretischen Optionen muss nach einem pragmatischen Ausweg gesucht werden. Dabei wollen wir die zuletzt genannte Lösung nicht gänzlich aus den Augen verlieren: Wir möchten die Regularitäten

12 Evtl. auch in den auf dieser Verbindung aufbauenden Wortbildungen (z. B. *Brom-beer|saft*).

13 Also nicht nur im vorliegenden Wort und in den darauf aufbauenden Wortbildungen. So erscheint das Morphem {beere} nicht nur in Verbindung mit dem unikalen Morphem {brom-} in *Brombeere* oder *Brombeersaft*, sondern auch frei in *Beere* sowie in *Blaubeere*, *Stachelbeere* etc. Entsprechend ist *Brombeere* eine einschlägige Wortbildung.

in den heute synchron analysierbaren Wortbildungen beschreiben. In der Praxis greifen wir dann, um die großen Datenmengen zu bewältigen, auf die Canoo-basierte Annotation zurück. Diese verwendet vorzugsweise gelistete Wortbildungsanalysen und leitet ggf. nicht gelistete Bildungen mithilfe von Regeln ab. Die automatische Annotation wird, wann immer möglich und notwendig, durch manuelle Korrekturen präzisiert bzw. – sofern Recall-Prüfungen dies nahelegen – durch zusätzliche, etwa formorientierte Recherchen ergänzt. Angemerkt werden muss dabei, dass die manuelle Analyse bzw. Annotation in einigen Fällen Interpretationsspielraum zulässt (etwa bezüglich der Zugehörigkeit verschiedener Formen zu ein und demselben Morphem). Dennoch glauben wir uns insgesamt der auf synchrone Analysierbarkeit der Bildungen ausgerichteten Ideal-lösung anzunähern. Gemäß dieser sind von den in diesem Abschnitt genannten Beispielen die Wörter *Fall*, *Wurf*, *statt*, *während*, *Haustür*, *Schnellzug*, *honiggelb*, *kleinkariert*, *nachdem*, *somit*, *Zeitung*, *verarbeiten*, *Gerede*, *Brombeere*, *Lindwurm*, *Unflat*, *drollig*, *Meineid*, *gewieft*, *reüssieren* als einschlägige Wortbildungen einzu-stufen, *heute*, *Adler*, *Welt*, *erst*, *nein* dagegen nicht.

### 3 Primäre Wörter und Wortbildungen – Simplizia und komplexe Wörter

Die einschlägigen Wortbildungen im Sinne von Abschnitt 2 werden im Weiteren der Kürze wegen schlicht als „Wortbildungen“ bezeichnet. Ähnlich wird auf die Komplementärgruppe mit „primäre Wörter“ Bezug genommen, auch wenn eigentlich „tendenziell primäre Wörter“ passender gewesen wäre, da sich in dieser Gruppe neben echten primären Wörtern auch verdunkelte Wortbildungen finden. Einige dem Korpus entnommenen Vertreter beider Gruppen enthält Tab. 1. Sie ist die Hauptquelle für die Beispiele, die weiter unten unsere Ausführungen veranschaulichen.

Tabelle 1: Primäre Wörter vs. Wortbildungen: Beispiele.

Primäre Wörter		Wortbildungen	
<i>Tor</i>	<i>Vater</i>	<i>Torfrau</i>	<i>Schwiegermutter</i>
<i>Frau</i>	<i>Kabarett</i>	<i>Landesgrenze</i>	<i>Stiefvater</i>
<i>Land</i>	<i>heute</i>	<i>gelblich</i>	<i>(das) Heute</i>
<i>Grenze</i>	<i>laufen</i>	<i>Volkstum</i>	<i>(der) Lauf</i>
<i>gelb</i>	<i>werfen</i>	<i>vernehmen</i>	<i>(der) Wurf</i>
<i>Volk</i>	<i>leben</i>	<i>Gerede</i>	<i>(das) Leben</i>
<i>nehmen</i>	<i>grün</i>	<i>Dialekt</i>	<i>Grüne(r)</i>
<i>reden</i>	<i>das</i>	<i>Telearbeit</i>	<i>demnach</i>
<i>Arbeit</i>	<i>nach</i>	<i>kabarettophil</i>	<i>Unflat</i>
<i>Mutter</i>	<i>[...]</i>	<i>identisch</i>	<i>[...]</i>

Konstitutive Bausteine von Wortbildungen sind kleinste bedeutungs- bzw. funktionstragende Formen, sog. Morpheme, die **wortfähige Morpheme**, **gebundene lexikalische Morpheme** und gebundene grammatische Morpheme der Kategorie **Wortbildungsaffixe** repräsentieren.<sup>14</sup>

## I. Wortfähige Morpheme

Wortfähige Morpheme im engeren Sinn können freistehend als (primäre) **Wörter** realisiert werden – sie stellen somit sog. freie Morpheme dar. Die meisten freien Morpheme weisen eine selbständige Bedeutung auf – als Beispiele können {tor}, {frau}, {land}, {grenze}, {gelb}, {volk}, {arbeit}, {mutter}, {vater}, {kabarett}, {heute} oder {grün} dienen. Sie werden als lexikalische Morpheme<sup>15</sup> angesehen (vgl. z. B. Brandt et al. 2006: 136–137) und bilden sog. Inhaltswörter<sup>16</sup> bzw. Autosemantika (vgl. Meibauer et al. 2007: 133), die mehrheitlich flektierbar sind. Wortfähig im engeren Sinn ist aber auch eine Reihe von Morphemen, die wie {das}<sup>17</sup> und öfter auch {nach} als grammatisch bezeichnet werden, da ihre Bedeutung unselbständiger erscheint und sie ziemlich spezifische grammatische Funktionen ausüben. Freistehend werden solche Morpheme als sog. Funktionswörter<sup>18</sup> bzw. Synsemantika eingestuft. Bereits am Übergang zu gebundenen lexikalischen Morphemen sind schließlich Basismorpheme von Verben zu verorten. Sie können nämlich standardsprachlich nur im Imperativ Singular schwacher Verben und starker Verben ohne *e/i*-Wechsel frei – d. h. hier auch – ohne eine explizite Flexionsmarkierung realisiert werden (Dudengrammatik 2016: 441–446), vgl. *leb!* bzw. *lauf!* Linear nicht segmentierbare Verbformen<sup>19</sup> wie *nahm*, *warf* oder *nimm!* *wirf!* hingegen können bereits als Kombinationen aus Basismorphem und Flexionsmarkierung mittels Vokalwechsel aufgefasst

14 Zur Spezifik der Kurzwörter, die sich nicht direkt in Morpheme zerlegen lassen, siehe weiter unten.

15 Lexikalische Morpheme sind nicht mit Lexemen zu verwechseln, zu denen hier alle Wörter im Sinne von Lexikoneinheiten gezählt werden, also auch Funktionswörter, vgl. Dudengrammatik (2016: 136–140).

16 Zu den Inhaltswörtern werden Substantive, Adjektive, die meisten Verben und Adverbien sowie manchmal auch Präpositionen gezählt.

17 Als Determinierer und Pronomen.

18 Zu den Funktionswörtern werden vor allem Artikel, Partikeln, subordinierende und koordinierende Konnektoren (zuzüglich Adverbkonnektoren, vgl. HDK 1), meist auch Präpositionen, Hilfs- und Modalverben, bestimmte Modalwörter und Pronomina gezählt, vgl. Willems (2013).

19 Der Vokalwechsel begründet kein nur als Flexionsmorphem interpretierbares Segment des Wortes.



werden.<sup>20</sup> Basismorpheme von Verben können dennoch als wortfähig in einem erweiterten Sinn betrachtet werden: Sie bilden (primäre) Wörter ohne Zuhilfenahme eines anderen Morphems der Gruppen I bis III.

## II. Gebundene lexikalische Morpheme

Lexikalische Morpheme, die nur gebunden vorkommen, sind fast immer als **Konfixe** einzustufen.<sup>21</sup> Sie sind typischerweise aus anderen europäischen Sprachen entlehnt und wurzeln besonders häufig im Griechischen, wie {dia}, {lekt}, {tele}, {phil}, oder Lateinischen, wie {ident}. Den Konfixen können außerdem heimische Morpheme wie {schwieger} oder {stief} zugeschlagen werden. Zu gebundenen lexikalischen Morphemen sind am Rande auch die in Abschnitt 2 bereits angesprochenen unikalen Morpheme wie {flat} in *Unflat* zu zählen.<sup>22</sup>

Die Morpheme der Gruppen I und II können miteinander sowie mit gebundenen grammatischen Morphemen der Kategorie Wortbildungsaffixe kombiniert werden.

## III. Wortbildungsaffixe

Wir sprechen hier einfach von **Affixen** (vgl. z.B. Fleischer & Barz 2012: 52–63) bzw. spezifischer von Präfixen wie in *ver|nehmen* oder *Un|flat*, Suffixen wie in *gelb|lich* oder *ident|isch* und Zirkumfixen wie in *Ge|red|e*. Diese Bezeichnungen verwenden wir in diesem Kapitel, um Verwechslungen zu vermeiden, nicht für Flexionsmorpheme (wie in *nimm|st*), die übrig gebliebene Kategorie der gebundenen grammatischen Morpheme.

Nur die wortfähigen Morpheme (Gruppe I) können ganz alleine in einer Wortbildung eingesetzt werden, vgl. (*das*) *Heute* oder (*der*) *Lauf* (zur sog. morphologischen Konversion vgl. z.B. Eisenberg 2013: 280). Baut eine Wortbildung demgegenüber auf zwei oder mehr Morphemen der Gruppen I bis III auf, betrachten wir sie als **komplexes Wort**. Jedes komplexe Wort muss mindestens ein Morphem enthalten, das entweder ein wortfähiges Morphem oder ein gebundenes lexikalisches Morphem repräsentiert.

20 Donalies (2002: 20) fasst dennoch die „Verbstämme des Typs *ess-*, *geb-* oder *les-*“ (folglich wohl auch *nehm-* oder *werf-*) als frei vorkommend auf.

21 Vgl. Donalies (2002: 21–23), Fleischer & Barz (2012: 63–64). Verschiedene Konfixtheorien werden in Donalies (2009) diskutiert.

22 Zum Unterschied zwischen unikalen Morphemen und Konfixen vgl. Meibauer et al. (2007: 31).

In ihrer Minimalform basieren komplexe Wörter auf zwei Morphemen. Die Zusammensetzungs- und Reihenfolgeoptionen des Morphempaars sind folgender Auflistung zu entnehmen:<sup>23</sup>

1. zwei wortfähige Morpheme, z. B. *Tor|frau*
2. ein gebundenes lexikalisches Morphem (A) und ein wortfähiges Morphem (B):

- a. A + B, z. B. *Tele|arbeit*
- b. B + A, z. B. *kabarett(o)|phil*<sup>24</sup>

3. zwei gebundene lexikalische Morpheme, z. B. *Dia|lekt*
4. ein wortfähiges Morphem (A) und ein Affix (B):<sup>25</sup>

- a. A + B, z. B. *gelb|lich* (Suffigierung)
- b. B + A, z. B. *ver|nehm-* (Präfigierung)

5. ein gebundenes lexikalisches Morphem (A) und ein Affix (B):

- a. A + B, z. B. *ident|isch* (Suffigierung)
- b. B + A, z. B. *Un|flat* (Präfigierung)

Es werden alle Reihenfolgemöglichkeiten genutzt, auch wenn die Optionen 2b und 5b weit seltener erscheinen als die anderen.<sup>26</sup> Die Verbindungstypen 1 bis 3 werden der Komposition zugerechnet, 4 und 5 der Derivation. Komposition und Derivation werden zusammen mit Konversion in weiteren Abschnitten differenzierter betrachtet und im Kapitel „Wortbildungsarten“ genauer analysiert.

Wortbildungsprozesse sind im Allgemeinen rekursiv. Das heißt hier, dass sich Verbindungen von Morphemen der Gruppen I bis III zum einem mit weiteren Morphemen der Gruppen I bis III, zum anderen mit weiteren Verbindungen von

23 Für eine Übersicht, die sich nicht nur auf komplexe Wörter in ihrer Minimalform beschränkt und auch Fugenelemente zur Unterscheidung von Kombinationsmöglichkeiten verwendet, vgl. Donalies (2002: 49).

24 Eingeklammert ist das Fugenelement, das keinen Morphemstatus besitzt – dazu weiter unten.

25 Weggelassen werden hier der Einfachheit halber Verbindungen mit Zirkumfixen, die als Kombinationen aus Präfigierung und Suffigierung betrachtet werden können.

26 Wird ein wortfähiges Morphem vor ein gebundenes lexikalisches Morphem gesetzt (2b), muss das vorangehende Wort häufig erst durch ein Fugenelement (vgl. -o- in *kabarett(o)|phil*) anschlussfähig für das darauf folgenden Konfix gemacht werden (wobei die Zusammensetzung oft immer noch ungewöhnlich klingt); die Verbindung eines Präfixes mit einem gebundenen lexikalischen Morphem (5b) scheint sich wiederum auf unikale Morpheme wie {flat} zu beschränken.

Morphemen der Gruppen I bis III zu noch komplexeren Bildungen vereinen können. Etwas pauschaler ausgedrückt: Komplexe Wörter können sich zusammen mit anderen Wörtern sowie mit Konfixen und Affixen zu noch komplexeren Wörtern verbinden.<sup>27</sup> Dabei können die Verbindungen im Deutschen bekanntlich eine sehr hohe Komplexität erreichen wie in den Korpusbeispielen *Gemeinnützigkeit(s) entbürokratisierung(s)gesetz* oder *Bund(es)verfassung(s)gericht(s)entscheidung(s) sammlung*. Wie solche Bildungen zu analysieren und hinsichtlich ihres Komplexitätsgrades zu bewerten sind, wird später im Kapitel „Wortbildungsarten“ diskutiert.

Neben den Morphemen der Gruppen I bis III können komplexe Wörter auch Fugenelemente und Flexionsmarkierungen enthalten. **Fugenelemente**<sup>28</sup> wurden in den gerade genannten Beispielen eingeklammert (vgl. dementsprechend auch *Land(es)|grenze*, *Volk(s)|tum*, *kabarett(o)|phil*), da ihnen der Morphemstatus meist abgesprochen wird (dazu Kopf 2018: 8–9). Sie sorgen nicht für eine Komplexitätserhöhung, sondern sind lediglich deren mögliche Begleiterscheinung.<sup>29</sup> Die heimischen Fugenelemente wie *-s-* oder *-es-* haben ihren Ursprung in Flexionsendungen, werden aber aus synchroner Sicht in der Regel nicht als deren Realisierungen eingestuft (vgl. z. B. Nübling & Szczepaniak 2009: 197).<sup>30</sup>

Flektierbare Wörter gehen in die Wortbildungen meist als unflektierte Stämme ein. **Flexionsmarkierungen** sind für Wortbildungen auch nicht konstitutiv. Das heißt vor allem, dass Wortformen, die als Verbindungen aus einem lexikalischen Morphem und einem Flexionsmorphem erscheinen (z. B. *Tor|es*, *Länd|er*, *nimm|st*, *leb|te*) oder durch Vokalwechsel markiert sind (z. B. *nahm*, *warf* oder *nimm! wirf!*), nicht als komplexe Wörter betrachtet werden. Entsprechend werden alle Konversionen aus primären Wörtern nicht als komplexe Wörter gewertet. Dies gilt für die schon erwähnten morphologischen Konversionen (vgl. Eisenberg 2013: 280–284) wie (*der*) *Lauf*, bei denen eine unflektierte Wortform (ein Stamm) direkt zur unflektierten Grundform einer anderen Wortart „konvertiert“. Das gilt aber auch für sog. implizite Ableitungen wie *Wurf* (vgl. Eisenberg 2013: 280)<sup>31</sup> oder offenkundig syntaktisch entstandene Infinitivkonversionen

27 Vgl. Donalies (2002: 15) zu den Möglichkeiten der Wortschatzerweiterung.

28 Sie werden genauer in einem getrennten Kapitel „Substantivkomposition und Fugenelemente“ behandelt.

29 Dies ist bei Minimalpaaren wie *Vermögen|steuer* vs. *Vermögen(s)|steuer* unmittelbar sichtbar. Zur Funktion von Fugenelementen s. die Übersicht in Nübling & Szczepaniak (2009: 201–208).

30 Speziell zu Fugenelementen als Nicht-Kasus-Suffixen s. Gallmann (1999). Für eine differenzierte Sicht auf die Semantik der Fugenelemente vgl. Donalies (2002: 46–48) und Kopf (2018: 54–72).

31 Donalies (2002: 135) wertet solche Fälle als Konvertate aus spezifischen, u. U. veralteten Verbstämmen (z. B. dem mhd. Präteritumstamm *wurf*, vgl. auch Fleischer & Barz 2012: 89). Wir werden sie im Kapitel „Wortbildungsarten“ diskutieren.

wie (*das*) *Leben*, die einen (ehemals) flexionsrelevanten Vokalwechsel respektive die Infinitivendung konservieren. Substantivierte Adjektive wie (*die/der*) *Grüne* (etwa im Sinne einer Person, die durch ihre politische Orientierung charakterisiert wird) stellen hier nur insofern einen Sonderfall dar, als sie nicht lediglich ein bestimmtes Flexionsmerkmal des Ausgangswortes behalten, sondern große Teile des adjektivischen Flexionsparadigmas in den nominalen Gebrauch integrieren (zur syntaktischen Konversion vgl. z. B. Eisenberg 2013: 280). Produkte aller genannten Konversionsarten gehören in unserer Auffassung also nicht zu den komplexen Wörtern, sondern sind zusammen mit primären Wörtern zu den einfachen Wörtern (den Simplicia) zu zählen.<sup>32</sup> Sie bleiben nichtsdestoweniger Wortbildungen, denn sie gehen auf andere Lexeme bzw. auf Wörter einer anderen Wortart zurück.

In die weiter oben beschriebene, aus formaler Sicht rein additive „Aufbau-mechanik“ komplexer Wörter fügen sich auch einige etwas weniger typische Bildungen ein. In diesen können an die Stelle eines Wortes, welches das Erstglied eines Kompositums bildet, zum einen nicht wortfähige Gruppen von Morphemem (*Kurzzeit|gedächtnis*), Wortgruppen (*Saure-Gurken-Zeit*) und sogar satzartige Konstrukte (*Trimm-Dich-Pfad*), zum anderen etwa einzelne Buchstaben bzw. ihre Bezeichnungen (*S-Kurve*) treten.<sup>33</sup> Einige dieser Ausdrucksformen können auch als Basis für Konversionen dienen, vgl. (*das*) *Stelldichein*, das auf einen Imperativsatz zurückgeführt werden kann (Fleischer & Barz 2012: 65, Henzen 1965: 83–84).

Es gibt jedoch auch Bildungen, die den additiven Regularitäten nicht direkt folgen. Zu denken ist hier vor allem an Kurzwörter und Kontaminationen (vgl. z. B. Fleischer & Barz 2012: 91, 93–94). Bei den eher randständigen Kontaminationen sind die Ausgangskomponenten meist immer noch leicht rekonstruierbar (vgl. *jein*, *Denglisch* oder – aus dem Englischen – *Brexit*). Insofern liegt die Klassifizierung der Kontaminationen als komplex nahe. Die viel häufigeren Kurzwörter (vgl. Fleischer & Barz 2012: 91, Balnat 2011: 287–289) sind ein komplizierterer

32 Konversionen bzw. sog. implizite Ableitungen aus primären Wörtern als einfache Wörter zu betrachten, hat eine lange Tradition. Schon nach Grimm hebt die „[i]nnere Wortbildung [...] die Einfachheit des Wortes nicht auf“ (Grimm 1826: 1).

33 Die Reihe kann mit weiteren, eher randständigen Erstgliedtypen fortgesetzt werden, z. B. Kombinationen aus Zahlwörtern und Präpositionen (geschrieben als Ziffern mit Trennzeichen) wie in *1:8-Debakel*. Beachtenswert ist der häufige Gebrauch des Bindestriches zwischen Erst- und Zweitglied des Kompositums bei Wortgruppen, Sätzen, Buchstaben etc. als Erstglied. Der Bindestrich wird bei Wortgruppen und Sätzen auch innerhalb des Erstglieds eingesetzt. Zu den Gründen für die Bindestrichsetzung in ähnlichen Fällen vgl. Kopf (2018: 119–120).

Fall.<sup>34</sup> Sie stellen zwar Reduktionen von fast ausschließlich komplexen Wörtern oder von Wortgruppen dar, die Ausgangskomponenten sind aber selten aus dem Kurzwort erschließbar und den Sprachbenutzer/inne/n in der Kommunikationssituation öfter nicht präsent (vgl. Dudengrammatik 2016: 680). Außerdem kann ein Kurzwort die Langform weitgehend aus dem Sprachgebrauch verdrängen und sich damit noch stärker einem primären Wort annähern (tendenziell bei *Labor* gegenüber *Laboratorium*). Dies zeigt einerseits, dass Kurzwörter tatsächlich eigenständige Wörter sind (auch wenn sie die Bedeutung der Langform und ihre grammatischen Merkmale behalten).<sup>35</sup> Es macht andererseits die Entscheidung schwer, sie pauschal den Simplizia oder den komplexen Wortbildungen zuzuschlagen (vgl. Dudengrammatik 2016: 748).

Wir fassen Kurzwörter als mit ihren Langformen teilweise synonyme Lexeme auf und verwenden zur Bewertung ihrer Komplexität wie bei anderen Bildungen das Kriterium der formalen Repräsentation der Ausgangskomponenten. So zählen wir unisegmentale Kurzwörter,<sup>36</sup> die auf ein Morphem oder einen Morphemteil der Langform gekürzt sind – wie *Labor[atorium]*, *Ober[kellner]*, *Zoo[logischer Garten]* bzw. *Alu[minium]*, *Krimi[nalroman/-film]* –, zusammen mit primären Wörtern zu den einfachen Wörtern bzw. den Simplizia. Alle anderen Kurzwörter betrachten wir als komplex. Hierzu gehören alle multisegmentalen Kurzwörter – unabhängig davon, ob sie als Silben- oder Buchstabenwörter (z. B. *Schiri*, *Kita* bzw. *GmbH*, *Kfz*, *GAU*, *TÜV*) erscheinen und ob sie gebunden (*Schiri*, *Kita*, *GAU*, *TÜV*) oder mit den Buchstabennamen (*GmbH*, *Kfz*) ausgesprochen werden. Da alle Kurzwörter aus unserer Sicht Wortbildungen sind, geht die Unterscheidung von primären Wörtern und Wortbildungen auch an dieser Stelle nicht mit der Unterscheidung von Simplizia und komplexen Wörtern einher.

34 Fleischer & Barz (2012: 277) sehen in der Kurzwortbildung eine Wortbildungsart. Eichinger (2000: 34) schreibt differenzierter: „In gewisser Weise haben [...] die Kurzwörter mehr vom Phraseologismus als von der eigentlichen Wortbildung – das Kurzwort ist eine Art Metapher der Ausdrucksform“.

35 Bei Wortgruppen richten sich Kurzwörter grammatisch meist nach dem Kopf der Langform. Zwischen Kurzwort und Langform kann es aber Bedeutungsverschiebungen und konnotative Unterschiede geben (vgl. Balnat 2011: 110–117). Allerdings ist deren Entstehung dem eigentlichen Wortbildungsprozess nachgeordnet. So kann *BAföG* anders als *Bundesausbildungsförderungsgesetz* auch die aufgrund des *Bundesausbildungsförderungsgesetzes* gezahlte Unterstützung bedeuten (<https://www.duden.de/rechtschreibung/BAfoeG>, abgerufen am 8.6.2020). Wie nah sich ein Kurzwort und seine Langform semantisch stehen, könnte korpuslinguistisch automatisiert geprüft werden, indem in einem Word-Embedding-Modell die Distanz der beiden Formen zueinander gemessen würde (vgl. Bubenhofer et al. 2019: 222–226).

36 Vgl. Fleischer & Barz (2012: 277–279) und – etwas differenzierter – Donalies (2002: 144–146).

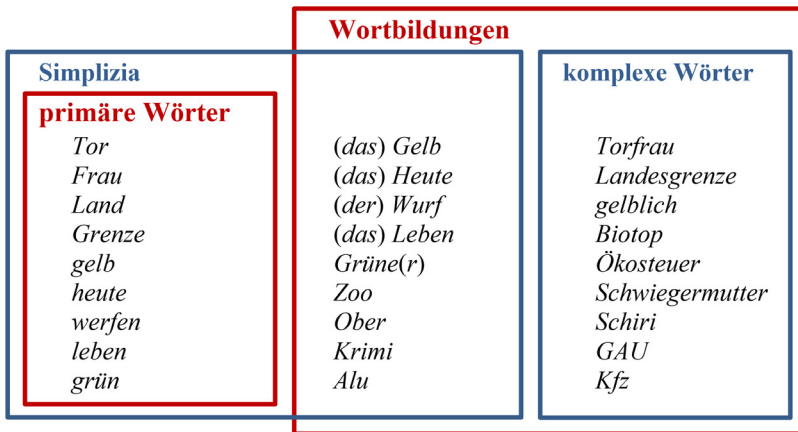


Abbildung 1: Primäre Wörter und Wortbildungen vs. Simplizia und komplexe Wörter.

Abb. 1 veranschaulicht zusammenfassend die Unterschiede zwischen den beiden Begriffspaaren anhand ausgewählter Beispiele.

#### 4 Allgemeine quantitative Verhältnisse und Wortschatzerweiterung

Die automatisch als Wortbildungen analysierten Wörter machen im gesamten Untersuchungskorpus 28,5% der über 5,8 Milliarden Tokens aus, die in der Canoo-basierten Annotation als Wörter behandelt werden. Der tatsächliche Anteil der Wortbildungen im Untersuchungskorpus muss höher liegen, da automatisch – insbesondere bei seltenen Wörtern – einige Wortbildungen nicht als solche analysiert werden. Um die quantitativen Verhältnisse zwischen Wortbildungen und primären Wörtern genauer zu untersuchen, wird daher das Teilkorpus *4ok* herangezogen, das manuell bearbeitet werden konnte und über eine halbe Million relevante Worttokens umfasst, die sich auf knapp 50.000 TreeTagger-Lemmata<sup>37</sup> verteilen (vgl. Abschnitt 1 oben). Der Anteil der Wortbildungen beträgt in diesem Teilkorpus bei den Tokens 31,8% (177.202). Bei den Lemmata liegt er mit 91,5% (44.746) viel höher. Der sehr deutliche Unterschied zwischen Tokens und Lemmata erklärt sich daraus, dass Wortbildungslemmata

37 Wir verwenden hier den Terminus *Lemma* anstatt des evtl. zu erwartenden *Typ* (engl. *type*), um klarzustellen, dass es sich um eine im Tagging zugewiesene Kategorie handelt, die einem Lexem entsprechen soll. *Typ* lässt ungünstigerweise die Ambiguität zwischen einer Wortform (der die Tokens im Korpus entsprechen) und einem Lexem (das etwa bei flektierenden Wortarten mehrere Wortformen umfasst) zu.

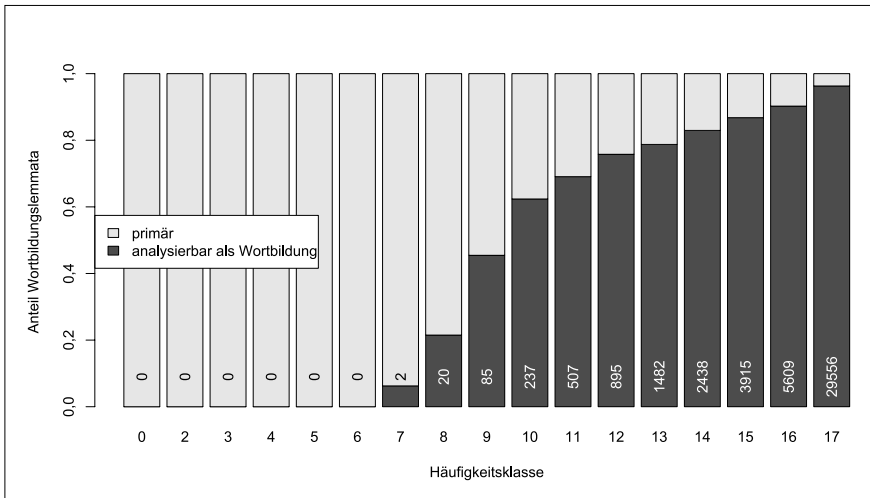


Abbildung 2: Anteil der Wortbildungslemmata nach Häufigkeitsklassen im Teilkorpus 40k: Je höher die Häufigkeitsklasse, desto tokenärmer die Lemmata, wobei:  $hk \approx \lfloor \log_2 (f(\text{der, die, das}) / f(\text{Lemma})) + 0,5 \rfloor$ . Die Zahlen in den Säulen geben die Anzahl der Wortbildungslemmata in der Häufigkeitsklasse an.

tendenziell tokenarm sind (66,0% davon sind Hapaxlegomena<sup>38</sup>) und es unter den besonders tokenreichen Lemmata gar keine Wortbildungen gibt. Teilt man alle Lemmata in Häufigkeitsklassen ein, so lässt sich in den Klassen, die Wortbildungen enthalten, Folgendes beobachten: Je weniger Tokens die Lemmata einer Häufigkeitsklasse aufweisen, desto größer der Anteil der Wortbildungen (vgl. Abb. 2). Es ist davon auszugehen, dass sich das auch in größeren Korpora so wiederfindet.<sup>39</sup>

In der Rangliste der 100 häufigsten Lemmata des Teilkorpus (vgl. Tab. 2 unten) findet sich das erste als Wortbildung analysierte Wort erst auf Rang 77. Das Fremdwort *Prozent* ist Canoo zufolge eine Zusammensetzung aus zwei „neoklassischen Formativen“.<sup>40</sup> Der erste Bestandteil der Bildung kann (allerdings anders als ein Konfix) auch frei auftreten wie in *pro Jahr* oder *pro Liter Benzin* (neben weiteren gebundenen Verwendungen wie in *prowestlich* oder *pro(-) amerikanisch*), der zweite Bestandteil lässt sich in *Zentimeter*, *Zentiliter* u. Ä.

38 Bei den primären Wörtern beträgt der Anteil der Hapaxlegomena 27,4%. Auch in Baayens (vgl. 2001: 135–138) Untersuchungen zu den Häufigkeiten niederländischer Wörter ist der Anteil der Hapaxlegomena bei den Ableitungen auf *-heid* viel höher als bei substantivischen Simplizia.

39 Vgl. auch dazu die prinzipiell in die gleiche Richtung zeigenden Ergebnisse von Baayen (2001: 135–138).

40 Vgl. <http://www.canoonet.eu/services/WordformationRules/Neoklassisch/Intro.html?MenuId=WordFormation112> (abgerufen am 3.11.2019).

Tabelle 2: Die Wortbildungen (hervorgehoben) unter den 100 häufigsten Lemmata im Teilkorpus 40k (HK=Häufigkeitsklasse).

<b>Rang</b>	<b>Lemma</b>	<b>Frequenz</b>	<b>HK</b>	<b>Rang</b>	<b>Lemma</b>	<b>Frequenz</b>	<b>HK</b>
1	<i>die</i>	68.695	0	34	<i>noch</i>	1.966	6
2	<i>in</i>	17.759	2	35	<i>vor</i>	1.916	6
3	<i>und</i>	14.508	3	36	<i>so</i>	1.863	6
4	<i>eine</i>	13.814	3	37	<i>alle</i>	1.765	6
5	<i>sein</i>	11.714	3	38	<i>über</i>	1.652	6
6	<i>zu</i>	8.628	3	39	<i>bis</i>	1.592	6
7	<i>werden</i>	7.406	4	40	<i>nur</i>	1.564	6
8	<i>von</i>	6.776	4	41	<i>müssen</i>	1.555	6
9	<i>haben</i>	6.658	4	42	<i>sollen</i>	1.516	6
10	<i>mit</i>	5.924	4	43	<i>sagen</i>	1.484	6
11	<i>an</i>	5.894	4	44	<i>geben</i>	1.476	6
12	<i>für</i>	4.825	4	45	<i>Uhr</i>	1.330	6
13	<i>auf</i>	4.779	4	46	<i>Sie</i>	1.318	6
14	<i>sich</i>	4.722	4	47	<i>mehr</i>	1.245	6
15	<i>nicht</i>	4.274	5	48	<i>man</i>	1.236	6
16	<i>es</i>	4.199	5	49	<i>neu</i>	1.228	6
17	<i>auch</i>	3.828	5	50	<i>wollen</i>	1.216	6
18	<i>er</i>	3.453	5	51	<i>oder</i>	1.203	6
19	<i>bei</i>	3.361	5	52	<i>gut</i>	1.200	6
20	<i>als</i>	3.109	5	53	<i>keine</i>	1.194	6
21	<i>dass</i>	2.790	5	54	<i>aber</i>	1.137	6
22	<i>aus</i>	2.705	5	55	<i>kommen</i>	1.093	6
23	<i>sein</i>	2.693	5	56	<i>unter</i>	1.041	7
24	<i>können</i>	2.665	5	57	<i>groß</i>	1.021	7
25	<i>dies</i>	2.660	5	58	<i>wenn</i>	1.016	7
26	<i>nach</i>	2.578	5	59	<i>gegen</i>	998	7
27	<i>ihr</i>	2.522	5	60	<i>gehen</i>	993	7
28	<i>wir</i>	2.477	5	61	<i>durch</i>	975	7
29	<i>ich</i>	2.344	5	62	<i>machen</i>	968	7
30	<i>sie</i>	2.328	5	63	<i>schon</i>	950	7
31	<i>um</i>	2.159	5	64	<i>andere</i>	938	7
32	<i>wie</i>	2.130	6	65	<i>viele</i>	889	7
33	<i>Jahr</i>	2.114	6	66	<i>dann</i>	860	7



Tabelle 2. Fortsetzung

Rang	Lemma	Frequenz	HK	Rang	Lemma	Frequenz	HK
67	<i>erst</i>	836	7	84	<i>jede</i>	561	7
68	<i>zwei</i>	821	7	85	<i>nun</i>	549	7
69	<i>wieder</i>	815	7	86	<b>bereits</b>	545	7
70	<i>was</i>	787	7	87	<i>heute</i>	538	7
71	<i>stehen</i>	756	7	88	<i>unser</i>	529	8
72	<i>aber</i>	730	7	89	<b>damit</b>	525	8
73	<i>lassen</i>	700	7	90	<b>dabei</b>	518	8
74	<i>immer</i>	699	7	91	<i>beide</i>	515	8
75	<i>ab</i>	685	7	92	<i>Frau</i>	515	8
76	<i>seit</i>	670	7	93	<i>Kind</i>	512	8
77	<b>Prozent</b>	650	7	94	<i>sehr</i>	509	8
78	<i>sehen</i>	641	7	95	<i>hier</i>	507	8
79	<i>weit</i>	632	7	96	<i>stellen</i>	498	8
80	<i>finden</i>	622	7	97	<i>Tag</i>	489	8
81	<i>jetzt</i>	620	7	98	<i>Mensch</i>	487	8
82	<i>drei</i>	611	7	99	<i>Zeit</i>	483	8
83	<i>ganz</i>	583	7	100	<i>weil</i>	477	8

erkennen. Die restlichen automatisch als Wortbildungen analysierten Lemmata der Rangliste aus Tab. 2 stellen Adverbien dar, deren synchrone Analysierbarkeit offensichtlich ist (*bereit|s, da|mit, da|bei*). Nicht als Wortbildungen analysiert wurden übrigens Fälle wie *dass* (Rang 21) oder *weil* (Rang 100), die man in einer historischen Betrachtung mit dem Pronomen *das* respektive dem Substantiv *Weile* hätte in Verbindung bringen müssen. Wir fassen sie aber als nicht einschlägig auf, da sie sich von den Basiswörtern bedeutungsmäßig bzw. funktional und teilweise auch formal zu sehr isoliert haben, als dass sie ohne Weiteres als synchron analysierbar betrachtet werden können.

Die obersten Positionen der Rangliste werden dominiert durch Funktionswörter, vor allem Determinierer, Präpositionen und Pronomina.<sup>41</sup> Auch die ersten auftauchenden Verblemmata (*sein, werden, haben, können, müssen, sollen*) werden zum großen Teil als Kopula-, Hilfs- oder Modalverben realisiert, die ebenfalls zu den Funktionswörtern gezählt werden können. Die häufigsten

41 Das Lemma *eine* (Rang 4) steht für Determinierer, Pronomen und Zahlwort, *sein* (Rang 5) für Verb, *sein* (Rang 23) für Determinierer und Pronomen und *dies* (Rang 25) ebenfalls für Determinierer und Pronomen.

Lemmata verweisen auf einsilbige Tokens – Ausnahmen finden sich unter den ersten 30 Lemmata bei einzelnen Wortformen der oben genannten Verben (z.B. *sein, bin, bist* etc. – einsilbig vs. *waren, gewesen* etc. – mehrsilbig) und des Determinierers/Pronomens/Zahlworts *ein(e)*. Die ersten sonstigen Lemmata, die für Inhaltswörter stehen, erscheinen auf der Liste erst später und gehören wie die Substantive *Jahr* (Rang 33) und *Uhr*<sup>42</sup> (Rang 45) oder die Verben *sagen* (Rang 43) und *geben* (Rang 44) den primären Wörtern an.

Festzuhalten ist, dass besonders häufige Wörter funktionale Ausdrücke darstellen, die kurz und primär sind. Auch die häufigsten Inhaltswörter sind kurz und primär. Bei besonders seltenen Wörtern hingegen überwiegen bei Weitem die Wortbildungen. Dabei sind unter den besonders seltenen Wörtern, insbesondere unter den Hapaxlegomena, die meisten Neuprägungen<sup>43</sup> zu erwarten (vgl. Baayen 2009: 905–906), und so ist die hohe Anzahl niederfrequenter Wortbildungen ein Indiz dafür, dass der Gesamtwortschatz durch Wortbildung immer weiter ausgebaut werden kann. Demgegenüber erscheint der Bestand an primären Wörtern nur beschränkt erweiterbar. Die Schöpfung neuer Wörter ohne Anschluss an vorhandenes Sprachmaterial (*Urschöpfung* oder auch einfach *Wort-schöpfung* genannt) spielt in der Gegenwartssprache kaum eine Rolle.<sup>44</sup> Sie ist weitgehend beschränkt auf Onomatopoetika<sup>45</sup> und Produktnamen (vgl. Donalies 2002: 17–18, Fleischer & Barz 2012: 18–20). An der Wortschatzerweiterung sind neben der Wortbildung vor allem die Entlehnung (*Computer, cool, Interface, Slot, Smartphone*) und die Bedeutungsveränderung bzw. -erweiterung (*geil* für ‚großartig‘, *Netz* für ‚Internet‘, *Rechner* für ‚Computer‘) beteiligt. Wohlgermerkt,

42 Das Lemma *Uhr* ist nur deshalb so häufig, weil es in Uhrzeitangaben wie *um 14.30 Uhr* auftritt. Gerade dort verhält es sich aber nicht wie ein typisches Substantiv (es ist nicht flektierbar, nicht erweiterbar etc.).

43 Neu- bzw. Spontanbildungen sind etwa in folgenden Hapaxlegomena aus dem Korpus zu vermuten: *ananasbeglückt, Bibliothekszwilling, kohlhäsisch, Kohl-Lafontaine-Weizsäcker-Vollmer-Mischung, valutadeutsch, wohlfahrtsvernünftig, wurmisch, (das) Zerlesenswerden*.

44 Anders als eine neue Wortbildung, die Assoziationen zu bereits bestehenden Wörtern wecken kann, kann ein gerade eingeführtes primäres Wort nur in einem absolut eindeutigen Kontext verstanden werden – etwa gestisch begleitet oder versehen mit einer Bedeutungserklärung wie „**Sitt** [...] ist ein Kunstwort, das als Adjektiv das Gegenteil von *durstig* bedeuten [...] soll“ (<https://de.wikipedia.org/wiki/Sitt>, abgerufen am 8.6.2020).

45 Allerdings sieht Donalies (vgl. 2002: 18, 91) in Verben aus Onomatopoetika (z.B. *miauen* oder *platschen*) bereits Wortbildungen. Hingegen betrachtet sie die Substantive *Kuckuck* und *Wauwau* als onomatopoetische Urschöpfungen (vgl. auch Fleischer & Barz 2012: 19). Henzen (vgl. 1965: 6–7) geht bei onomatopoetischen Verben dann von Wortbildungen aus, wenn sie verbale Suffixe *-r-* bzw. *-l-* wie in *plappern* oder *bimmeln* aufweisen.

die entlehnten Wörter können auf primäre Wörter (*cool, slot*) oder Wortbildungen (*computer, interface, smart phone*) der Gebersprache zurückgehen. Letztere werden nicht selten auch im Deutschen als Wortbildungen empfunden (*Interface, Smartphone*), da sie mit anderen geläufigen Wortbildungen (*Intercity, international* bzw. *Smartboard, Smartcard, Smart-TV*) leicht assoziierbar sind. Sie sind somit synchron analysierbar und auch erklärbar. Der Wortbildungsursprung eines Fremdworts kann im Deutschen manchmal aber auch kaum zu erkennen sein wie bei *Aids*<sup>46</sup> oder *Laser*, die im Englischen als multisegmentale Kurzwörter entstanden sind. Solche Entlehnungen können wie primäre Wörter wahrgenommen werden (vgl. dazu Dudengrammatik 2016: 744) und sind als synchron nicht analysierbar zu betrachten.

In einem umfangreichen Korpus können nicht alle Fremdwörter aufgespürt und daraufhin überprüft werden, ob sie synchron analysierbar sind, und selbst wenn das möglich gewesen wäre, hätte es bei einzelnen Entscheidungen öfter Interpretationsspielraum gegeben.<sup>47</sup> Wir stützen uns daher beim Gesamtkorpus auch bei Fremdwörtern auf die Ergebnisse der automatischen Wortbildungsanalysen von Canoo, die punktuell überprüft und korrigiert werden. In diesen werden von den oben genannten Fremdwörtern nur die ältere Bildung *international* sowie die neueren *Smartcard* und *Smartboard* als Wortbildungen dargestellt. Vereinbar mit unseren Überlegungen werden *Aids, Radar* und *Computer* nicht als Wortbildungen analysiert, ebenso wenig aber – diesmal nicht mit unseren Überlegungen konform – *Intercity, Interface, Smartphone* und *Smart-TV*. Die Annotation von Belegen wie den letzteren wird, sofern die Fehler gefunden werden, korrigiert. Was sich hier zeigt, gilt für die Nutzung der Wortbildungsannotation zur Suche nach Wortbildungen im Fremdwortbereich insgesamt. Die Annotation ist relativ präzise, zeigt jedoch Recall-Schwächen.

## 5 Quantitative Verhältnisse in verschiedenen Wortklassen

Die relative Häufigkeit von Wortbildungen variiert von Wortart zu Wortart und auch von einer Teilkategorie innerhalb einer Wortart zur anderen. Näher betrachtet werden hier in diesem Zusammenhang die Wortarten Substantiv, Adjektiv, Verb und Adverb, deren Lexembestand in größerem Maße erweiterbar ist. Sie

46 Durch die ebenfalls gebräuchliche Großschreibung *AIDS* wird das Wort als Kurzwort (Buchstabenwort) gekennzeichnet, was es formal wieder zur Wortbildung macht.

47 So könnte *Computer* von manchen als synchron nicht analysierbar und nicht erklärbar aufgefasst werden, von anderen aber als analysierbar in ein Wortbildungsmorphem {-er} und ein Konfix {comput-}, das in unserem Korpus noch in dem Ausdruck (*ubiquitäres*) *Computing* vorkommt.

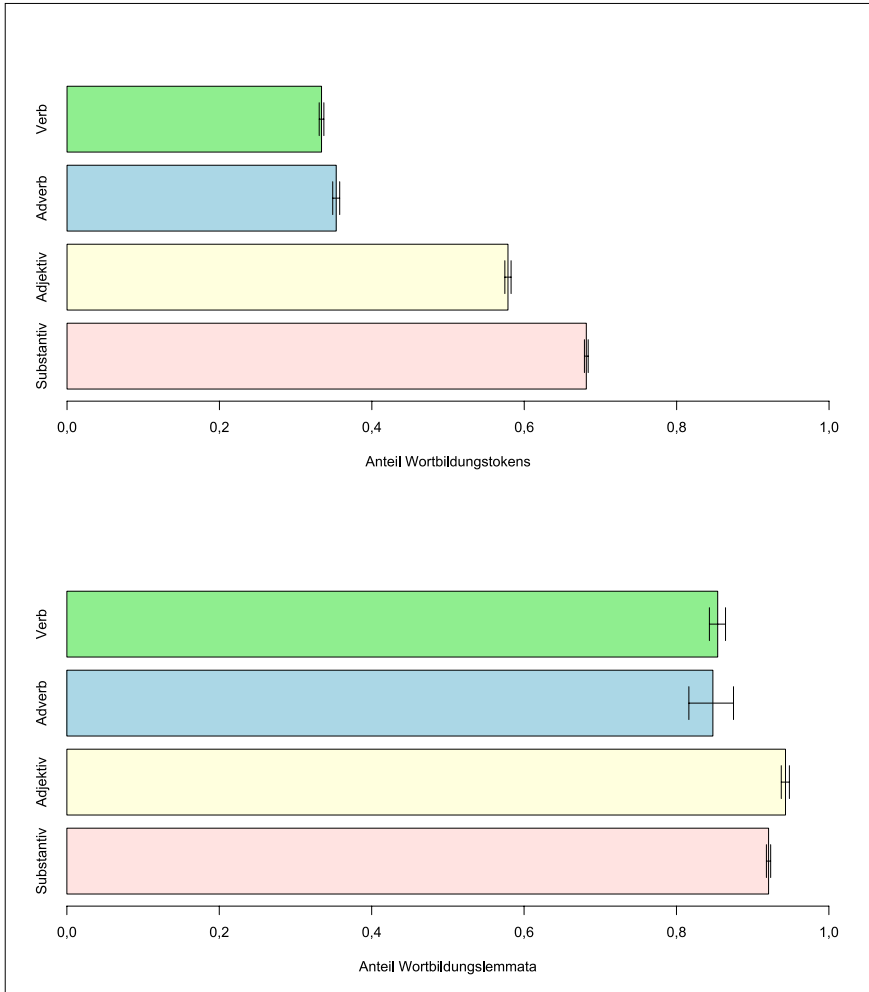


Abbildung 3: Anteil der Wortbildungen bei den erweiterbaren Wortarten im Untersuchungskorpus anhand des Teilkorpus *40k* (jeweils mit einem 95 %-Konfidenzintervall).

setzen sich in erster Linie aus Inhaltswörtern zusammen und werden allgemein als offene Klassen angesehen.

Substantive und Adjektive sind deutlich wortbildungsfreundlicher als Adverbien und Verben. Abb. 3 zeigt den für das Teilkorpus *40k* ermittelten Anteil der Wortbildungen an allen Tokens bzw. Lemmata der vier Wortarten.<sup>48</sup>

48 Zu der Zählweise ist zu bemerken, dass Partikelverben nur dann als *ein* Wort eingestuft werden, wenn sie im Text zusammengeschrieben erscheinen. Bei der Distanzstellung der Partikel (bzw. Getrennschreibung des Partikelverbs) wird die abgetrennte

Der Unterschied zwischen dem verhältnismäßig niedrigen Anteil der Wortbildungen an den Tokens und ihrem durchgehend deutlich höheren Anteil an den Lemmata ist besonders ausgeprägt bei Verben und Adverbien. Dies kann schon daran liegen, dass es einige hochfrequente Verb- und Adverb-lemmata gibt, die primäre Wörter sind (z. B. *sein*, *werden*, *haben* bzw. *auch*, *noch*, *so*, vgl. Tab. 2 weiter oben). Bei allen vier Wortarten ist der Anteil der Wortbildungen an allen Lemmata allerdings abhängig von der Größe des Teilkorpus. Wird diese erhöht, steigen die Anteile der Wortbildungen im Bereich der Lemmata bei allen vier Wortarten an, während ihre Anteile im Bereich der Tokens relativ unveränderlich bleiben. Hier wirkt sich die Produktivität der Wortbildungen aus: Mit wachsender Anzahl berücksichtigter Tokens kommen bei Wortbildungen schnell neue Lemmata dazu, wogegen bei primären Wörtern die Anzahl der Lemmata viel verhaltener ansteigt. Abb. 4 zeigt die Wortschatzwachstumskurven<sup>49</sup>, die mithilfe des Softwarepakets zipfR unter Anwendung des finiten Zipf-Mandelbrot-LNRE-Modells (vgl. Evert & Baroni 2007) allen Wortbildungen und allen tendenziell primären Wörtern im Teilkorpus *40K* zugeordnet werden können.

Während die Wachstumskurve für die Wortbildungslemmata auch am rechten Rand der Abb. 4 immer noch relativ steil ansteigt, flacht die Wachstumskurve für die Lemmata primärer Wörter sehr schnell ab, sodass ihr Anstieg bald kaum mehr bemerkbar ist. Ähnliches zeigt sich beim Vergleich der Kurven für die Hapaxlegomena (unter diesen sind die meisten Neuprägungen zu erwarten): Die Kurve für die Wortbildungen steigt am rechten Rand der Abbildung immer noch deutlich an, während die Kurve für die primären Wörter da sogar bereits abfällt. Aus diesen Vergleichen kann man erneut auf die grundsätzliche Produktivität der Wortbildungsprozesse und die vergleichsweise geringe Bedeutung der Wortschatzerweiterung durch primäre Wörter schließen.<sup>50</sup> Zu prinzipiell ähnlichen Ergebnissen kommen Roth (vgl. 2014: 177–182) beim Vergleich verschiedener

Partikel separat gezählt (z. B. in *Verspätungen setzen sich wie in einer Kettenreaktion weiter fort*) und das übrig gebliebene Verb wird, sofern es nicht weiter analysierbar ist, als primär klassifiziert. Die Überprüfung einer Stichprobe ergab, dass sich unter den als primär eingestuften Verbtokens  $5,8\% \pm 2,3\%$  ( $n=400$ , 95%-Konfidenzintervall) Verben befinden, zu denen eine Verbpartikel in Distanzstellung gehört. Wenn man die entsprechenden Partikelverben zu den Wortbildungen zählen würde, würde der Anteil der Wortbildungen an allen Verbtokens also leicht (um ca. 4 Prozentpunkte) ansteigen.

49 Auch *Vocabulary growth curve* (so bei Evert & Lüdeling 2001: 168), *Vokabularwachstumskurve* (bei Roth 2014: 169).

50 Die potenzielle Produktivität  $P$  (Quotient aus der Anzahl der Hapaxlegomena einer Kategorie und der Anzahl aller Tokens dieser Kategorie, vgl. Baayen 2009) beträgt bei 382.108 Tokens (beobachtete Tokenanzahl der primären Wörter) 0,0049 für primäre Wörter und 0,1213 für Wortbildungen.

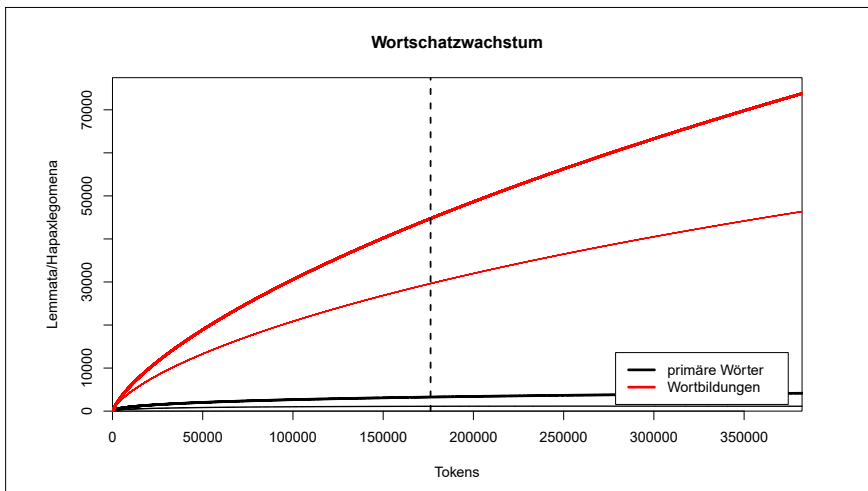


Abbildung 4: Anzahl der Lemmata (dickere Kurvenlinien) und der Hapaxlegomena (dünnere Kurvenlinien) bei steigenden Tokenzahlen berücksichtigter Wortbildungen bzw. primärer Wörter im Teilkorpus *40k* (gemäß dem finiten Zipf-Mandelbrot-Modell in *zipfR*). Bei den Wortbildungen wurden die Wachstumskurven über die beobachtete Anzahl der Tokens (176.266, gestrichelte Linie) hinaus so weit extrapoliert, bis die beobachtete Anzahl der Tokens der primären Wörter (382.108, Ende der Skala) erreicht wurde.

Kompositionsmuster mit Simplizia im Deutschen sowie bereits Baayen (vgl. 2001: 135–138) beim Vergleich von Derivationen auf *-heid* mit simplizischen Substantiven im Niederländischen.<sup>51</sup>

Wie zu erwarten, sind Wortbildungen verschiedener Wortarten unterschiedlich produktiv (zur Kompositionsproduktivität vgl. Roth 2014: 177–178). Die Wortschatzwachstumskurven in Abb. 5 unten lassen auf deutliche Unterschiede schließen. Im Teilkorpus *40k* sind die Wortbildungen besonders produktiv bei den Substantiven. Die anderen Wortarten folgen in der Reihenfolge Adjektive<sup>52</sup>, Verben und Adverbien. Die Wachstumskurve der Lemmata für Adverbien zeigt in Abb. 5 kaum mehr Steigung und die Kurven der Hapaxlegomena für Adverbien und Verben sinken am rechten Rand der Abbildung sogar schon.<sup>53</sup>

51 In Baayens Modellen steigen die Wachstumskurven der Worttypes und der Hapaxlegomena bei Derivationen auf *-heid* immer weiter an, während bei Simplizia die Anzahl der Worttypes nicht mehr wächst und die Anzahl der Hapaxlegomena abnimmt, sobald höhere Tokenzahlen erreicht werden.

52 Die Hapaxlegomena-Kurve für Adjektive ist in Abb. 5 durch die Lemmata-Kurve für Verben verdeckt.

53 Die potenzielle Produktivität  $P$  für das Ende der Skala (59.561 Tokens) erreicht für Substantive 0,2980, für Adjektive 0,0758, für Verben 0,0212 und für Adverbien 0,0009. Die Extrapolation der Kurven für Adverbien auf das Vierfache der beobachteten

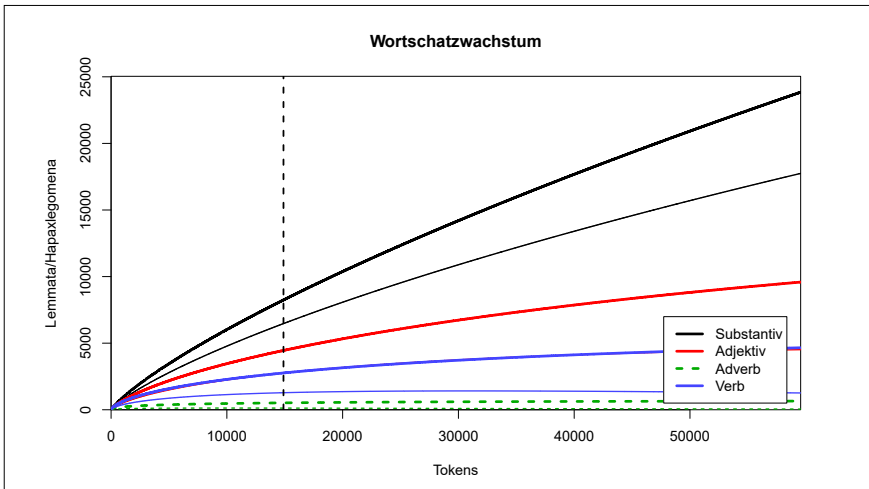


Abbildung 5: Anzahl der Lemmata (dickere Kurven) und der Hapaxlegomena (dünnere Kurvenlinien) für verschiedene Wortarten bei steigenden Tokenzahlen anhand des Teilkorpus *40k* (gemäß dem finiten Zipf-Mandelbrot-Modell in *zipfR*). Die Wachstumskurven wurden auf das Vierfache der beobachteten Anzahl der Tokens der seltensten Wortart Adverb extrapoliert. Die gestrichelte Linie zeigt die beobachtete Anzahl der Tokens für Adverbien (14.890).

Unsere Betrachtungen bedürfen einiger teilweise einschränkender Kommentare. Die traditionellen Wortarten sind auf hohen Abstraktionsniveaus definiert und stellen in mancher Hinsicht sehr grobe Oberkategorien dar. Die Unterkategorien können Teilklassen konstituieren, die bezüglich der Wortbildung differenziert zu betrachten sind. So ist die Wortart Substantiv wie bereits erwähnt durch die Einteilung in Appellative und Eigennamen geprägt. Beide Untergruppen unterscheiden sich deutlich in ihrer Semantik, Pragmatik und Schreibung. Eigennamen weisen auch Besonderheiten hinsichtlich der gegenwartssprachlich-synchronen Analysierbarkeit und Erklärbarkeit der Wortbildungsmuster auf (vgl. Fleischer & Barz 2012: 179). Hier ist noch einmal zu betonen, dass wir in den quantitativen Übersichten nur die „regulärere“ Wortbildung der Appellative fokussiert und die Eigennamen weitestmöglich ausgeschlossen haben. Ferner können in den Wortbildungen im Allgemeinen Bausteine erscheinen, die stark reihenbildend sind. Dabei handelt es sich nicht nur um Affixe. In appellativen Substantiven etwa sind es auch manche kompositionalen Zweitglieder wie das polyseme *Werk* in *Kraft|werk*, *Kunst|werk* oder *Studierenden|werk*

Tokenanzahl entspricht der Empfehlung von Evert & Baroni (2005) für noch plausible Ergebnisse.

(man vergleiche dazu auch das meist schon als Suffix betrachtete *-werk*<sup>54</sup> in *Mauer|werk*, *Netz|werk*, *Regel|werk*), *Stelle* in *Bau|stelle*, *Geschäfts|stelle* oder *Halte|stelle* und *Mann* in *Ehe|mann*, *Kauf|mann* oder *Vorder|mann*. Als Adjektive wiederum sind viele Kardinal- und Ordinalzahlverwendungen einzustufen, z. B. in *fünfzehn Jahre* oder *das fünfzehnte Jahrhundert*. Sie bilden potenziell unendliche Reihen aus Formen, die nach einigen wenigen, teilweise idiosynkratischen Wortbildungsregeln gebildet werden. Die Zahlwörter können im Übrigen außer mit Buchstaben auch mit Ziffern, u. U. unter Zuhilfenahme der Punktsetzung wie in *der 5. Mai* ausgedrückt werden, allerdings wurden solche mit Ziffern verschriftlichten Zahlwörter von der Untersuchung ausgenommen. Ausgeschlossen wurden dabei auch die von einem Punkt gefolgt Ziffern, die sich unter Umständen als aus Ordinalzahlen abgeleitete Adverbien interpretieren lassen (z. B. *1.* im Sinne von *erstens*) – solche Adverbkandidaten waren durch den Tagger nämlich ziemlich unzuverlässig klassifiziert worden.<sup>55</sup> Die Adverbien insgesamt zerfallen in verschiedene Teilklassen, die hinsichtlich der Wortbildung ganz unterschiedlich zu charakterisieren sind. So beträgt etwa der Anteil der Wortbildungen an Präpositionaladverbien – sie setzen sich aus einem Pro-Element und einer Präposition zusammen – per definitionem 100 % (Tokens und Lemmata). Allerdings bilden Präpositionaladverbien unter allen Adverbien eher eine kleine Gruppe (knapp 7 % der Lemmata), sodass sie die Zählungen zu Adverbien insgesamt nur geringfügig beeinflussen. Diese erscheinen also dennoch verhältnismäßig wenig wortbildungsfreudig (vgl. Abb. 3 weiter oben).

Anders als in den oben behandelten Wortarten ist der Lexembestand in denjenigen Wortarten, die sich aus Funktionswörtern zusammensetzen, sehr beschränkt ausbaufähig. Determinierer, Pronomina und Konjunktionen werden oft als geschlossene Klassen betrachtet (dazu kritisch Waßner 2002). Charakteristisch für diese Wortarten ist eine im Vergleich zu den berücksichtigten vier Wortarten überschaubare Anzahl an Lexemen, die teilweise mit sehr hohen Tokenzahlen (vgl. Tab. 2 weiter oben) realisiert werden. So entsprechen etwa den nur zwölf Formen des bestimmten und des unbestimmten Artikels bzw. der gleichlautenden Pronomen (*die*, *der*, *das*, *des*, *dem*, *den* einerseits und *eine*, *einer*, *ein*, *eines*, *einem*, *einen* andererseits) im Gesamtkorpus (knapp 6 Mrd. Wort-tokens) fast 700 Mio. Tokens.

54 So z. B. in Fleischer & Barz (2012: 56).

55 Sie werden im Kapitel „Bau von und Umbau zu Adverbien“ (Brandt in dieser Ausgabe) dennoch stichprobenartig untersucht.



## 6 Wortbildungsprofile verschiedener Wortarten im Vergleich

Die einzelnen Wortarten sind u. a. dadurch charakterisiert, dass sie verschiedene Wortbildungsarten unterschiedlich stark gewichten<sup>56</sup> und diese mit jeweils spezifischen Wortbildungsregeln realisieren. Die Verteilung der anhand der Wortbildungsannotation ermittelten Wortbildungsarten auf die Wortarten Substantiv, Adjektiv, Verb und Adverb im Teilkorpus *40k* wird für Tokens und Lemmata in Abb. 6 illustriert. Im Anschluss diskutieren wir die wichtigsten Gemeinsamkeiten und Unterschiede, die sich beim Vergleich der Wortarten zeigen. Dabei werden auch die jeweils auffälligsten Wortbildungsregeln genannt und durch Beispiele veranschaulicht. Zum Schluss wird in den Tab. 3 bis 6 eine Übersicht über die typischen Wortbildungsregeln der einzelnen Wortarten gegeben.

In einer ersten formal-kompositionalen Annäherung unterscheiden wir an dieser Stelle als Wortbildungsarten lexikalische Verbindung, Suffigierung, Präfigierung und Konversion. Die übrigen Bildungsweisen fassen wir unter „Rest“ zusammen. Es sind einige Spezifika dieser Unterteilung zu erläutern.

Unter *lexikalischer Verbindung* verstehen wir – provisorisch zusammengefasst – alle Verbindungen aus wortfähigen Einheiten<sup>57</sup> und/oder Konfixen als unmittelbaren Konstituenten (Fugenelemente werden hier vernachlässigt). Dazu gehören vor allem „klassische“ Komposita (vgl. z. B. Schlücker 2012: 8), in denen das Zweitglied die gleiche Wortart hat wie die gesamte Wortbildung (*Haus|tier*, *zahl|reich*, *darauf|hin*). Hier werden aber auch einige abweichende Bildungen berücksichtigt, etwa solche, in denen die Verbindung nicht die Wortart des Zweitglieds hat (*da|mit*, *dem|nach*) und die auch als Univerbierung<sup>58</sup> betrachtet werden können (*trotz|dem*, *in|zwischen*, vgl. Donalies 2018: 225–243). Unter Präfigierung findet sich außer den üblicherweise dazugerechneten Fällen (*Un|glück*, *Ge|flügel*, *er|klären*, *Ge|bet*) – bedingt durch die Spezifik der automatischen Annotation – auch die Zirkumfigierung (*Ge|länd|e*, *Ge|web|e*). Die Suffigierung wiederum umfasst neben den üblichen Fällen (*Musik|er*, *beruf|lich*) die meisten Zusammenbildungen mit Suffixen (*Vier|bein|er*, *groß|züg|ig*) und manche

56 Eichinger (2000: 74–114) beobachtet aus der umgekehrten Perspektive Präferenzen bei der Verteilung der Wortbildungsarten auf die Wortarten Substantiv, Adjektiv, Verb und erörtert eingehend die Gründe für solche Präferenzen.

57 Hierzu gehören wohlgermerkt auch Komponenten, denen lexikalische Bedeutung abgesprochen wird (zumindest *dem* in *trotz|dem*). Wir verwenden dennoch die Bezeichnung *lexikalische Verbindung*, weil wir hier Komponenten mit lexikalischer Bedeutung für prototypisch halten.

58 Mit *Univerbierung* kann der Prozess der Zusammenrückung von Wörtern einer Wortgruppe zu einem Einzelwort bezeichnet werden, aber auch ein komplexes Wort, das formal einer Wortgruppe gleicht – mit dem Unterschied, dass die Wortgruppe auseinander geschrieben wird.

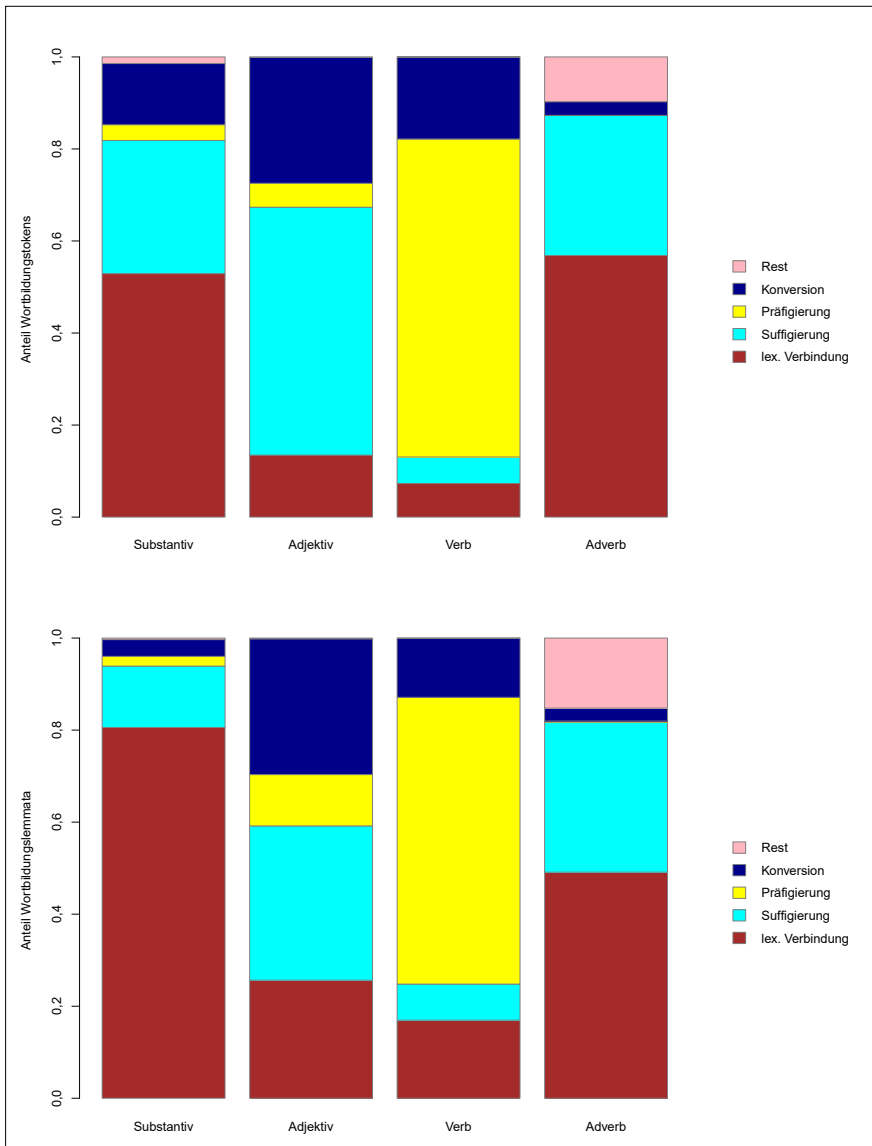


Abbildung 6: Wortbildungsprofile verschiedener Wortarten: Anteile der Wortbildungsarten.

Rektionskomposita (*Feuer|spuck|er*, *Graffiti-|Sprüh|er*).<sup>59</sup> Unter Zusammenbildungen verstehen wir dabei – ähnlich wie u. a. Neef (2015: 583–585, zurückgehend auf Leser 1990: 4) – Dreierverbindungen, bei denen weder die ersteren beiden noch die letzteren beiden Komponenten mit der in die gesamte Bildung integrierbaren Bedeutung wortfähig sind (vgl. \**Vierbein*, \**Beiner* bzw. \**Großzug*, \**zügig*)<sup>60</sup> – in solchen Fällen geht die Basis der Suffigierung scheinbar auf eine Wortgruppe oder Fragmente einer größeren Wortgruppe zurück. Die Konversion umfasst an dieser Stelle neben den typischen Fällen wie *fertigen* oder (*das*) *Leben* auch departizipiale Adjektive und Substantive (*spannend*, (*der/ein*) *Betroffene(r)*) sowie auch implizite Ableitungen ((*das*) *Verbot* < *verbieten*, (*der*) *Auftritt* < *auftreten*). Zum „Rest“ schließlich gehören einige analysierbare, aber von Canoo aus verschiedenen Gründen nicht mit einer Regel verbundene Bildungen ((*das*) *Alter*, *ausfindig*, *schließlich*, *übrigens*). Eine genauere Differenzierung der Wortbildungsarten wird später im gleichnamigen Kapitel ausgearbeitet.

Der Vergleich des Anteils der lexikalischen Verbindung bei verschiedenen Wortarten legt auf den ersten Blick ihre paarweise Gruppierung nahe. **Substantive** und **Adverbien** ähneln sich darin, dass sie besonders durch lexikalische Verbindungen geprägt sind. Bei den Substantiven fallen insbesondere determinativ analysierbare Komposita aus zwei substantivischen Konstituenten wie *Wochen|ende*, *Stadt|rat*, *Verzichts|planung* ins Gewicht,<sup>61</sup> in größerem Abstand gefolgt von Komposita mit einer adjektivischen Erstkonstituente wie *Alt|papier*, *Neu|verschuldung*, *Halb|zeit* und solchen mit einer verbalen Erstkonstituente wie *Warte|zeit*, *Fahr|zeug*, *Halte|stelle*. Etwas seltener, aber immer noch sehr zahlreich sind Komposita mit einer präpositionalen Erstkonstituente wie *Mit|glied*, *Auf|takt*, *Gegen|rede* und Konfixverbindungen wie *Tele|fon*, *Dia|lekt*, *Gastro|nom*.

Bei den Adverbien sind Verbindungen aus Adverb und Präposition (Präpositionaladverbien u. ä.) wie *da|durch*, *her|aus*, *wo|nach* am auffälligsten – sie entsprechen nicht den typischen Komposita (zu diesen vgl. etwa Schlücker 2012: 8 oder Ortner & Ortner 1984: 36) – nicht zuletzt, weil das rechts stehende Element nicht die Wortart der gesamten Wortbildung aufweist. Häufig sind außerdem Komposita aus zwei Adverbien wie *wo|her*, *darauf|hin*, *immer|hin*. Zu weiteren relevanten lexikalischen Verbindungen, die übrigens ebenfalls

59 Die Zusammenbildungen aus (Zahl-)Adjektiv, Substantiv und Suffix werden oft auch zu den Komposita gezählt (so bei Neef 2015: 584–585, anders bei Donalies 2002: 95–96). Vor allem aber hätten Rektionskomposita auch der lexikalischen Verbindung zugeordnet werden können.

60 Somit zeigen sie eine Bildungsweise, in der mehrere Wortbildungsschritte zu einem zusammengefasst werden. Sie hat eine größere Reichweite als bisher angenommen und wird im Kapitel „Wortbildungsarten“ ausführlicher behandelt.

61 Hier sind auch Verbindungen mit Bindestrich(en) sehr zahlreich.

nicht den typischen Komposita entsprechen, gehören vor allem solche aus einer Präposition und einem Pronomen wie *trotz|dem, unter|dessen, ohne|dies* (unter ihnen sind auch Verbindungen mit einem komplexen Pronomen als Zweitglied zu finden wie *in|einander, gegen|einander, mit|einander*) und solche aus zwei Präpositionen wie *vor|bei, in|zwischen, mit|unter*. Die Ähnlichkeit der Substantive und der Adverbien hinsichtlich ihrer Prägung durch lexikalische Verbindungen entpuppt sich somit schnell als eine nur scheinbare. Die Positionierung der adverbialen Bildungen gegenüber den kanonischen, substantivischen Komposita wird ein wichtiger Punkt im Kapitel „Wortbildungsarten“ sein. Dem Aufbau von Adverbien aus Lexemen unterschiedlicher Provenienz, bei dem neben der lexikalischen Verbindung auch die Suffigierung eine sehr wichtige Rolle spielt, widmet sich das Kapitel „Bau von und Umbau zu Adverbien“ (Brandt in dieser Ausgabe).

Substantive und Adverbien haben darüber hinaus, was die Tokens angeht, einen ähnlichen Anteil an Suffigierungen, aber auch hier sind die Bildungen eigentlich grundsätzlich verschieden. Bei Substantiven dominieren die „klassischen“ Suffigierungen wie *Erklär|ung, Musik|er, Gewiss|heit*, bei denen das komplexe Substantiv mithilfe eines spezialisierten Suffixes vor allem von einem Verb, sehr häufig aber auch von einem anderen Substantiv oder einem Adjektiv abgeleitet wird. Bei Adverbien kommen die Suffixe *-s* und *-ens* hingegen in Verbindung mit sehr vielfältigen Basen zum Einsatz (*angesicht|s, bereit|s, zusehend|s, vergeben|s, erst|ens, höchst|ens*). Die Verwendung von *-s* überlagert dann auch einen Teil der Bildungen mit suffixartigen desubstantivischen Zweitgliedern wie *andern|fall|s, dies|seit|s, heim|wärt|s*. Zu solchen Bildungen gehören ansonsten auch etwa *beispiels|weise, glücklicher|weise* oder *gleicher|maßen*. Die sehr produktiven und theoretisch schwer einzuordnenden Adverbien mit desubstantivischen Zweitgliedern sind ebenfalls Gegenstand eines eigenen Kapitels.

Im restlichen Teil ihrer Wortbildungsprofile sind sich Substantive und Adverbien schon auf den ersten Blick unähnlich. Auffällig ist bei Substantiven ein merklicher Anteil an Konversionen im Bereich der Tokens (*(das) Leben, (der) Anfang, (der) Auftritt*), worin sie mit den Adjektiven und Verben vergleichbar erscheinen,<sup>62</sup> bei den Adverbien hingegen ein beträchtlicher Anteil an Bildungen, die nicht den etablierten Wortbildungsarten entsprechen wie *schließ|lich, übrig|ens, zu|nächst* (in Abb. 7 dem „Rest“ zugeordnet) – einige von ihnen können als Univerbierungen betrachtet werden (*seiner|zeit, zur|zeit, so|zu|sagen*, vgl. Donalies 2018: 176–262). Präfigierungen sind bei Adverbien absolut

62 Es ist aber nur eine Vergleichbarkeit hinsichtlich der Anteile und nicht in Bezug auf die Bildungsweisen. Letztere werden im Kapitel „Wortbildungsarten“ genauer dargestellt.

randständig,<sup>63</sup> bei Substantiven machen sie sich vor allem im Hinblick auf den Anteil der Tokens bemerkbar (*Miss|erfolg, Un|menge, Ur|aufführung*).

Bei **Adjektiven** und **Verben** sind die lexikalischen Verbindungen anteilmäßig schwächer ausgeprägt. Bei Adjektiven fallen vor allem die Verbindungen mit einem Substantiv als Erstglied ins Gewicht wie *zahl|reich, sinn|voll, welt|weit*, in einem deutlichen Abstand gefolgt von den Verbindungen mit einem Adjektiv als Erstglied wie *weit|gehend, offen|sichtlich, gleich|gültig*. Bildungen beider Typen sind oft (wie die meisten lexikalischen Verbindungen bei den Substantiven) als typische Komposita zu betrachten. Unter den Verbindungen mit einem Adjektiv als Erstglied finden sich aber auch viele, die auch als Suffigierungen bzw. Zusammenbildungen interpretiert werden könnten wie *recht|zeit|ig, voll|ständ|ig, frei|will|ig*. Bei Verben wiederum kommen spezielle morphosyntaktische Regularitäten zum Tragen. So machen sich hier die in Verbletzstellung befindlichen und daher zusammengeschriebenen Partikelverben besonders bemerkbar. Dabei fallen vor allem Verben mit depräpositionalen und deadverbialen Erstgliedern wie *mit|teilen, fort|setzen, hin|weisen* ins Gewicht, dann in einem größeren quantitativen Abstand (mit absteigender Häufigkeit) solche mit einem deadjektivischen (*bekannt|geben, sicher|stellen, offen|legen*), desubstantivischen (*statt|finden, gewähr|leisten, teil|haben*) und deverbalen (*kennen|lernen, stehen|lassen, liegen|bleiben*) Erstglied.

Was die Verben im Gesamtvergleich auszeichnet, ist der besonders hohe Anteil an Präfigierungen. Es dominieren die Ableitungen eines Verbs aus einem anderen Verb wie *er|klären, be|stehen, ent|scheiden*. Quantitativ sehr deutlich nachgeordnet erscheinen Ableitungen mit einer desubstantivischen (*be|fristen, ver|ursachen, ent|puppen*) oder deadjektivischen Basis (*be|treuen, er|ledigen, ver|öffentlichen*). Bei den Adjektiven ist wiederum der Anteil der Suffigierung im Vergleich mit den drei anderen Wortarten am höchsten. Dabei erfolgt die Ableitung vor allem aus Substantiven wie in *beruf|lich, beispiel|haft, witz|ig*, schon deutlich seltener aus Verben wie in *deut|lich, denk|bar, abhäng|ig* und noch seltener aus Adjektiven wie in *kürz|lich, stet|ig, gemein|sam*. Auf die Ableitung aus Adjektiven folgen, was die Häufigkeit angeht, Zusammenbildungen wie *gleich|zeit|ig, wahr|schein|lich, lang|frist|ig*. Sie können als Ableitungen aus Adjektiv-Substantiv-Verbindungen interpretiert werden, welche nicht als Wörter (Komposita) etabliert sind (*\*Gleichzeit, \*Wahrschein, \*Langfrist*).

Adjektive weisen im Weiteren auch den im Vergleich der Wortarten höchsten Anteil an Konversionen auf, was vor allem darauf zurückgeht, dass hierzu die große Anzahl von adjektivisch gebrauchten Partizipien bzw. partizipbasierten Bildungen wie *vergangen, entsprechend, bewusst* gehört. Bei Verben ist der Anteil

63 Zu Bildungen wie *zu|tiefst* als Präfigierungskandidaten vgl. Fleischer & Barz (2012: 370) und das Kapitel „Bau von und Umbau zu Adverbien“.

Tabelle 3: Wortbildungsregeln beim Substantiv.

<b>Substantiv</b>		
<b>Wortbildungsart</b>	<b>Typische Regeln</b>	<b>Beispiel</b>
Lexikalische Verbindung	1. SBST+SBST 2. ADJ+SBST, V+SBST 3. PRÄP+SBST 4. KONF+KONF, KONF+SBST 5. ADV+SBST, [ADJ]+SBST]+SBST	<i>Wehrmachts ausstellung</i> <i>National rat, Turn halle</i> <i>Ab glanz</i> <i>Disko thek, Atmo sphäre</i> <i>Auswärts spiel, Lang zeit projekt</i>
Suffigierung	1. V+SUFF 2. SBST+SUFF 3. ADJ+SUFF	<i>Besuch er</i> <i>Herr schaft</i> <i>Aufmerksam keit</i>
Konversion	1. V → SBST 2. ADJ → SBST	<i>Beginn</i> <i>Intellektuelle(r)</i>
Präfigierung	1. PRÄF+SBST 2. PRÄF+V(+SUFF)	<i>Un glück</i> <i>Ge sang, Ge bild e</i>

der Konversionen wie *landen*, *nutzen*, *sichern* ebenfalls beachtlich, wobei sich vor allem bei Stämmen, die sowohl substantivisch als auch verbal gebraucht werden, zumindest synchron gesehen öfter die Frage der Konversionsrichtung stellt, z. B. *Spiel* < *spielen* oder *spielen* < *Spiel* (vgl. Eisenberg 2013: 284). Die Präfigierung ist bei Adjektiven (*miss|mutig*, *un|schön*, *ur|alt*) anteilmäßig etwa so schwach ausgeprägt wie die Suffigierung bei Verben. Letztere wird durch das Suffix *-(is)ier* im Anschluss an fremde Basen dominiert. Gänzlich heimische Bildungen wie *fest|igen*, *rein|igen*, *genehm|igen* sind deutlich seltener.

In den Tab. 3 bis 6 werden die für die vier Wortarten typischen Wortbildungsregeln nach Wortbildungsarten geordnet gelistet und jeweils mit einem Beispielbeleg veranschaulicht. Die häufigeren Wortbildungsarten und Regeln werden zuerst aufgeführt.<sup>64</sup> Sind bestimmte Regeln innerhalb einer Wortbildungsart etwa gleich häufig, erscheinen sie in derselben Zeile.

64 Bei den Häufigkeiten für Tokens und für Lemmata ergeben sich in der Regel gleiche Reihenfolgen wie schon weiter oben in der Abb. 6 für Wortbildungsarten.

Tabelle 4: Wortbildungsregeln beim Adjektiv.

<b>Adjektiv</b>		
<b>Wortbildungsart</b>	<b>Typische Regeln</b>	<b>Beispiel</b>
Suffigierung	1. SBST+SUFF 2. V+SUFF 3. ADJ+SUFF 4. ADV+SUFF	<i>farb ig</i> <i>bedenk lich</i> <i>eigent lich</i> <i>heut ig</i>
Konversion	1. V → ADJ 2. SBST → ADJ	<i>bestehend</i> <i>ernst</i>
Lexikalische Verbindung	1. SBST+ADJ 2. ADJ+ADJ 3. PRON+ADJ, ADV+ADJ	<i>jahre lang</i> <i>halb voll</i> <i>all gegenwärtig, wohl bekannt</i>
Präfigierung	1. PRÄF+ADJ 2. PRÄF+[V+SUFF]	<i>in direkt</i> <i>un ermüd lich</i>


Tabelle 5: Wortbildungsregeln beim Verb.

<b>Verb</b>		
<b>Wortbildungsart</b>	<b>Typische Regeln</b>	<b>Beispiel</b>
Präfigierung	1. PRÄF+V 2. PRÄF+SBST 3. PRÄF+ADJ	<i>ver treten</i> <i>be flügeln</i> <i>ent schuldigen</i>
Konversion	1. SBST → V 2. ADJ → V	<i>planen</i> <i>sichern</i>
Lexikalische Verbindung	1. PRÄP/ADV+V 2. ADJ+V 3. SBST+V	<i>mit teilen, fort setzen</i> <i>fest stellen</i> <i>teil nehmen</i>
Suffigierung	1. SBST+SUFF 2. ADJ+SUFF	<i>bilanz ieren</i> <i>mobil isieren</i>

Tabelle 6: Wortbildungsregeln beim Adverb.

<b>Adverb</b>		
<b>Wortbildungsart</b>	<b>Typische Regeln</b>	<b>Beispiel</b>
Lexikalische Verbindung	1. ADV+PRÄP 2. ADV+ADV 3. PRÄP+PRON 4. PRÄP+PRÄP	<i>da bei</i> <i>wo anders</i> <i>zu dem</i> <i>zu vor</i>
Suffigierung	1. ADJ+SUFF 2. [ADJ+SBST]+SUFF 3. SBST+SUFF 4. ADV+SUFF, ADJ <sub>super</sub> +SUFF	<i>bereit s</i> <i>neuer ding s</i> <i>teil s</i> <i>damal s, mindest ens</i>
Konversion	1. ADJ <sub>super</sub> → ADV 2. SBST → ADV	<i>möglichst, längst</i> <i>morgen</i>
Präfigierung	1. PRÄF+ADJ <sub>super</sub>	<i>zu vorderst</i>

ORCID®

Marek Konopka  <https://orcid.org/0000-0001-9627-0918>

## Literatur

- Baayen, Harald R. 2001. *Word Frequency Distribution* (Text, Speech and Language Technology 18). Dordrecht: Springer.
- Baayen, Harald R. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (Hgg.), *Corpus Linguistics: An International Handbook* (Handbücher zur Sprach- und Kommunikationswissenschaft 29), Bd. 2, 899–919. Berlin: De Gruyter.
- Balnat, Vincent. 2011. *Kurzwortbildung im Gegenwartsdeutschen* (Germanistische Linguistik: Monographien 26). Hildesheim, New York & Zürich: Olms.
- Brandt, Patrick, Rolf-Albert Dietrich & Georg Schön. 2006. *Sprachwissenschaft: Ein roter Faden für das Studium der deutschen Sprache*, 2. Aufl. Köln: Böhlau.
- Bubenhofer, Noah, Selena Calleri & Philipp Dreesen. 2019. Politisierung in rechtspopulistischen Medien: Wortschatzanalyse und Word Embeddings. *OBST. Osnabrücker Beiträge zur Sprachtheorie* 95, 211–242.
- Deutsche Wortbildung 1 = Kühnhold, Ingeburg & Hans Wellmann. 1973. *Deutsche Wortbildung: Erster Hauptteil: Das Verb* (Sprache der Gegenwart 29). Düsseldorf: Schwann.
- Deutsche Wortbildung 2 = Wellmann, Hans. 1975. *Deutsche Wortbildung: Zweiter Hauptteil: Das Substantiv* (Sprache der Gegenwart 32). Düsseldorf: Schwann.
- Deutsche Wortbildung 3 = Kühnhold, Ingeburg, Oskar Putzer & Hans Wellmann. 1978. *Deutsche Wortbildung: Dritter Hauptteil: Das Adjektiv* (Sprache der Gegenwart 43). Düsseldorf: Schwann.
- Deutsche Wortbildung 4 = Ortner, Lorelies, Elgin Müller-Bollhagen, Hanspeter Ortner, Hans Wellmann, Maria Pümpel-Mader & Hildegard Gärtner. 1991. *Deutsche Wortbildung: Vierter Hauptteil: Substantivkomposita* (Sprache der Gegenwart 79). Berlin & New York: De Gruyter.
- Deutsche Wortbildung 5 = Pümpel-Mader, Maria, Elsbeth Gassner-Koch & Hans Wellmann. 1992. *Deutsche Wortbildung: Fünfter Hauptteil: Adjektivkomposita und Partizipialbildungen* (Sprache der Gegenwart 80). Berlin & New York: De Gruyter.
- Donalies, Elke. 2002. *Die Wortbildung des Deutschen* (Studien zur deutschen Sprache 27). Tübingen: Narr.
- Donalies, Elke. 2009. Stiefliches Geofaszintainment: Über Konfixtheorien. In Peter O. Müller (Hg.), *Studien zur Fremdwortbildung* (Germanistische Linguistik 197–198), 41–64. Hildesheim, New York, Zürich: Olms.



- Donalies, Elke. 2018. *Wetterbeobachter, Zeitlang, wahrsagen, zartfühlend, kurzerhand, dergestalt: Handbuch zur Univerbierung*. Heidelberg: Winter.
- Dudengrammatik. 2016 = Wöllstein, Angelika & Dudenredaktion (Hgg.). 2016. *Dudenband 4 – Die Grammatik*, 9. Aufl. Berlin: Dudenverlag.
- DWB = Grimm, Jacob & Wilhelm Grimm. 1854–1961. *Deutsches Wörterbuch von Jacob und Wilhelm Grimm*. 16 Bände in 32 Teilbänden. Quellenverzeichnis 1971. Leipzig: Hirzel. [http://woerterbuchnetz.de/cgi-bin/WBNetz/wbgui\\_py?sigle=DWB](http://woerterbuchnetz.de/cgi-bin/WBNetz/wbgui_py?sigle=DWB) (abgerufen am 6.6.2020).
- Eichinger, Ludwig M. 2000. *Deutsche Wortbildung: Eine Einführung*. Tübingen: Narr.
- Eisenberg, Peter. 2013. *Grundriss der deutschen Grammatik. Bd. 1: Das Wort*, 4. Aufl. Stuttgart & Weimar: Metzler.
- Evert, Stefan & Marco Baroni. 2005. Testing the extrapolation quality of word frequency models. In Pernilla Danielsson & Martijn Wagenmakers (Hgg.), *Proceedings from the Corpus Linguistics Conference Series*, Bd. 1. Birmingham: University of Birmingham. <https://www.birmingham.ac.uk/Documents/college-artslaw/corpus/conference-archives/2005-journal/Lexiconodf/EvertBaroniz2005.pdf> (abgerufen am 7.7.2020).
- Evert, Stefan & Marco Baroni. 2007. zipfR: Wordfrequency distributions in R. In Sophia Ananiadou (Hg.), *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*. 29–32. Prag: ACL. <https://www.aclweb.org/anthology/P07-2008.pdf> (abgerufen am 7.7.2020).
- Evert, Stefan & Anke Lüdeling. 2001. Measuring morphological productivity: Is automatic preprocessing sufficient? In Paul Rayson, Andrew Wilson, Tony McEnery, Andrew Hardie & Shereen Khoja (Hgg.), *Proceedings of the Corpus Linguistics 2001 conference*, 167–175. Lancaster: UCREL.
- Finkel, Jenny Rose, Trond Grenager & Christopher Manning. 2005. Incorporating non-local information into information extraction systems by Gibbs sampling. In Kevin Knight, Hwee Tou Ng & Kemal Oflazer (Hgg.), *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, 363–370. Ann Arbor: Association for Computational Linguistics. <http://nlp.stanford.edu/~manning/papers/gibbscrf3.pdf> (abgerufen am 18.7.2020).
- Fleischer, Wolfgang & Irmhild Barz. 2012. *Wortbildung der deutschen Gegenwartssprache*, 4. Aufl. Berlin & Boston: De Gruyter.
- Gallmann, Peter. 1999. Fugenmorpheme als Nicht-Kasus-Suffixe. In Matthias Butt & Nanna Fuhrhop (Hgg.), *Variation und Stabilität in der Wortstruktur (Germanistische Linguistik 141–142)*, 177–190. Hildesheim, New York & Zürich: Olms.
- Grimm, Jakob. 1826. *Deutsche Grammatik*. Bd. 2. Göttingen: Dietrich.

- Haß, Ulrike & Klosa, Annette (2015): *elexiko – Geschichte und Zukunft. Sprachreport* 3. 10–20.
- HDK 1 = Pasch, Renate, Ursula Brauße, Eva Breindl & Ulrich H. Waßner. 2003. *Handbuch der deutschen Konnektoren: Linguistische Grundlagen der Beschreibung und syntaktische Merkmale der deutschen Satzverknüpfers (Konjunktion, Satzadverbien und Partikeln)* (Schriften des Instituts für deutsche Sprache 9). Berlin & New York: De Gruyter.
- Henzen, Walter. 1965. *Deutsche Wortbildung*, 3. Aufl. Tübingen: Niemeyer.
- Kopf, Kristin. 2018. *Fugenelemente diachron: Eine Korpusuntersuchung zu Entstehung und Ausbreitung der verfügbaren N+N-Komposita* (Studia linguistica Germanica 133). Berlin & Boston: De Gruyter.
- Leser, Martin. 1990. *Das Problem der ‚Zusammenbildungen‘: eine lexikalistische Studie* (Fokus 3). Trier: WVT.
- Meibauer, Jörg, Ulrike Demske, Jochen Geilfuß-Wolfgang, Jürgen Pafel, Karl Heinz Ramers, Monika Rothweiler & Markus Steinbach. 2007. *Einführung in die germanistische Linguistik*, 2. Aufl. Stuttgart & Weimar: Metzler.
- Motsch, Wolfgang. 1999. *Deutsche Wortbildung in Grundzügen* (Schriften des Instituts für deutsche Sprache 8). Berlin & New York: De Gruyter.
- Neef, Martin. 2015. Synthetic compounds in German. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (Hgg.), *Word-Formation: An International Handbook of the Languages of Europe* (Handbücher zur Sprach- und Kommunikationswissenschaft 40), Bd. 1, 581–592. Berlin: De Gruyter.
- Nübling, Damaris, Antje Dammel, Janet Duke & Renata Szczepaniak. 2017. *Historische Sprachwissenschaft des Deutschen: Eine Einführung in die Prinzipien des Sprachwandels*, 5. Aufl. Tübingen: Narr.
- Nübling, Damaris & Renata Szczepaniak. 2009. *Religion+s+freiheit, Stabilität+s+pakt und Subjekt(+s)+pronomen*: Fugenelemente als Marker phonologischer Wortgrenzen. In Peter O. Müller (Hg.), *Studien zur Fremdwortbildung* (Germanistische Linguistik 197–198), 197–222. Hildesheim, New York & Zürich: Olms.
- Ortner, Hanspeter & Lorelies Ortner. 1984. *Zur Theorie und Praxis der Komposita-forschung* (Forschungsberichte des Instituts für deutsche Sprache 55). Tübingen: Narr.
- Paul, Hermann. 1920. *Deutsche Grammatik. Bd. 5: Wortbildungslehre*. Halle a. S.: Niemeyer.
- Pfeifer, Wolfgang (Hg.). 1993–heute. *Etymologisches Wörterbuch des Deutschen.: Digitalisierte und v. Wolfgang Pfeifer überarbeitete Version im Digitalen Wörterbuch der deutschen Sprache*. Bearbeitet von Wilhelm Braun, Gunhild Ginschel, Gustav Hagen, Anna Huber, Klaus Müller, Heinrich Petermann, Gerlinde Pfeifer, Dorothee Schröter & Ulrich Schröter. <https://www.dwds.de/d/wb-etymwb> (abgerufen am 8.5.2020).

- Roth, Tobias. 2014. *Wortverbindungen und Verbindungen von Wörtern: Lexikografische und distributionelle Aspekte kombinatorischer Begriffsbildung zwischen Syntax und Morphologie* (Basler Studien zur deutschen Sprache und Literatur). Tübingen: Francke.
- Schäfer, Roland. 2015. Processing and Querying Large Web Corpora with the COW<sub>14</sub> Architecture. In Piotr Bański, Hanno Biber, Evelyn Breiteneder, Marc Kupietz, Harald Lungen, & Andreas Witt (Hgg.), *Proceedings of Challenges in the Management of Large Corpora (CMLC-3)*, 28–34. Mannheim: Institut für Deutsche Sprache.
- Schäfer, Roland & Felix Bildhauer. 2012. Building Large Corpora from the Web Using a New Efficient Tool Chain. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (Hgg.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, 486–493. Istanbul: European Language Resources Association (ELRA).
- Schlücker, Barbara. 2012. Die deutsche Kompositionsfreudigkeit: Übersicht und Einführung. In Livio Gaeta & Barbara Schlücker (Hgg.), *Das Deutsche als kompositionsfreudige Sprache: Strukturelle Eigenschaften und systembezogene Aspekte* (Linguistik – Impulse & Tendenzen 46), 1–25. Berlin & Boston: De Gruyter.
- Schmid, Helmut. 1995. Improvements in Part-of-Speech Tagging with an Application to German. *Proceedings of the ACL SIGDAT-Workshop*. Dublin (Irland). <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/treetagger2.pdf> (abgerufen am 9.2.2020).
- Schmid, Helmut, Arne Fitschen & Ulrich Heid. 2004. SMOR: A German Computational Morphology Covering Derivation, Composition, and Inflection. In Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa & Raquel Silva (Hgg.), *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, 1263–1266. Lissabon: European Language Resources Association (ELRA). <https://www.cis.uni-muenchen.de/~schmid/papers/smor.pdf> (abgerufen am 18.6.2020).
- Waßner, Ulrich H. 2002. Geschlossene Klassen? In Reinhard Rapp (Hg.), *Sprachwissenschaft auf dem Weg in das dritte Jahrtausend: Akten des 34. Linguistischen Kolloquiums in Germersheim 1999. Teil II: Sprache, Computer, Gesellschaft*, 635–643. Frankfurt a. M., Berlin, Bern u. a.: Lang.
- Willems, Klaas. 2013. Funktionswort. In Stefan J. Schierholz & Pál Uzonyi (Hgg.), *Grammatik: Formlehre: Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK) Online*. Berlin & Boston: De Gruyter. [https://db.degruyter.com/view/WSK/wsk\\_id\\_wsk\\_artikel\\_artikel\\_21552](https://db.degruyter.com/view/WSK/wsk_id_wsk_artikel_artikel_21552) (abgerufen am 18.6.2020).
- Wilmanns, Wilhelm. 1899. *Deutsche Grammatik: Gotisch, Alt-, Mittel- und Neuhochdeutsch*. Straßburg: Trübner.

