

Noch einmal: Wie viele Wörter hat die deutsche Sprache?

DOI: 10.14232/fest.bassola.22

ABSTRACT

Nur ein paar Überlegungen und Beobachtungen zur Frage nach dem Umfang des deutschen Wortschatzes. Zahlenangaben aus Wörterbüchern und Korpusrecherchen werden referiert. Anhand von Beispielen veralteter, alternder, neuer Wörter und produktiver Wortbildungsmuster wird die prinzipielle Offenheit des Wortschatzes demonstriert.

1. Wörter

Sprachinteressierte Laien fragen immer wieder nach der Größe des deutschen Wortschatzes, d.h., nach der Gesamtzahl der deutschen Wörter. Ich musste auf solche Fragen schon oft eingehen, konnte aber nie mit einer konkreten Zahl antworten, sondern nur mit Erklärungen, warum sich keine bestimmte Zahl nennen lässt.¹ Aber auch eine Erklärung, warum sich keine Gesamtzahl der Wörter der deutschen Sprache angeben lässt, kann, wie ich erfahren habe, sprachliches Interesse bedienen. Deshalb stelle ich im Folgenden etwas ausführlicher zusammen, was ich bei verschiedenen Gelegenheiten auf die thematische Frage geantwortet habe² und was mir auch anhand von Wortsuchen in computergespeicherten Texten noch zu dieser Frage eingefallen ist.

Wörter interessieren mich seit langem, sind aber nicht mein linguistisches Spezialgebiet. Als ich vor langen Jahren in die Sprachwissenschaft eingeführt wurde, war neben Phonetik und Phonologie die Grammatik das Kerngebiet der

¹ In manchen linguistischen Arbeiten werden Zahlen genannt. Kommentierte Zahlenangaben finden sich u.a. in Haß-Zumkehr (2001: 381–385), Best (2006: 13ff.) und Duden (2017: 148ff.).

² Unter anderem bei Gesprächen mit Germanistikstudenten während verschiedener Aufenthalte an der Universität Szeged.

Linguistik mit der Königsdisziplin Syntax. Wörter waren dabei meist nur wichtig, um Beispielsätze für generative, konstitutionelle oder dependenzielle Strukturbeschreibungen zu füllen. Auf die zentrale Bedeutung der Wörter für unser sprachliches Tun bin ich erst nach und nach aufmerksam geworden, und das eher anhand alltäglicher Beobachtungen: Kinder können sich in ihren ersten Jahren entsprechend ihren Bedürfnissen mit Wörtern verständlich machen, die grammatisch gar nicht oder nur schwach verbunden sind. Bei Reisen in Ländern mit anderen Sprachen kommt man manchmal mit ein paar Wörtern und Wendungen recht weit, ohne die Grammatik der betreffenden Sprachen gelernt zu haben. Flüchtlinge und Arbeitsmigranten aus anderen Ländern können ihre grundlegenden Bedürfnisse zur Not mit isolierten oder ungrammatisch verbundenen Wörtern äußern. Natürlich geht es dann ohne Grammatik nicht viel weiter; deshalb braucht eine sprachliche Integration auch Grammatik.

Im Vergleich zu der überschaubaren Anzahl der Phoneme, den je nach Sprache 24 bis 30 Zeichen des lateinischen ABCs und der ebenfalls überschaubaren Menge der Satzstrukturen einer europäischen Sprache haben Wörter für die Wissenschaft einen großen Nachteil: Es gibt so viele davon. Sie sind in ihrer Menge und Vielfalt schwer zu überschauen, zu ordnen und prinzipiell nicht vollständig zu beschreiben. Schwierigkeiten bereitet schon die Bestimmung dessen, was als deutsche Wörter gezählt werden soll. Sind es die bedeutungstragenden deutschen Phonemsequenzen, zwischen denen bei langsamem Sprechen Pausen gesetzt werden können, oder – der Einfachheit halber auf die Schriftsprache beschränkt – die Buchstabenfolgen, die in Texten durch Leerzeichen (Spalten) und/oder Satzzeichen voneinander getrennt erscheinen? Gehören Zahlwörter, gesprochen, geschrieben oder in Ziffern notiert, dazu? Wie steht es mit chemischen oder mathematischen Formeln, die gesprochen werden können? Sind auch alle Eigennamen, die in deutschen Telefonbüchern stehen, und Ortsnamen, die in Registern deutschsprachiger Atlanten zu finden sind, mitzuzählen? Ich erspare den Lesern und mir, die reiche Literatur mit Definitionen dessen, was Wörter sind, zu referieren,³ und beschränke mich auf die alltagssprachliche Verwendung des Ausdrucks ‚Wörter‘ als Bezeichnung der sprachlichen Einheiten, die im Wörterbuch stehen

³ Eine knappe Übersicht findet sich in Bußmann (1990: 849f.); ausführlicher diskutiert Klein (2013: 22–28) verschiedene Wortbegriffe.

oder stehen könnten, also lexikalische Wörter (Grundformen). Diese unter-
scheide ich aber von den grammatischen Formen, in denen flektierbare Wör-
ter in Texten vorkommen.

2. Zahlen

Auch eine modifizierte Frage wie „Wie viele lexikalische Wörter hat die deut-
sche Sprache?“ erlaubt als Antwort keine Zahlenangabe. Sicher ist nur, dass
ein vollständiges Verzeichnis deutscher Wörter nirgends zu finden ist. Selbst
die größten deutschen Wörterbücher sind nicht vollständig, wengleich sie
in den aufgeführten Wortmengen beachtlich und nützlich sind. Der neueste
Rechtschreib-Duden enthält etwa 145 000 Wörter (Duden 2017: Vorwort). Das
größte moderne, wenn auch schon leicht veraltete Wörterbuch, das 10-bän-
dige „Große Wörterbuch der deutschen Sprache“ von 1999, ebenfalls ein Du-
den-Produkt, umfasst etwa 200 000 Wörter (Duden 1999: Vorwort). Das his-
torische „Deutsche Wörterbuch“ der Brüder Grimm beschreibt „weit mehr als
300.000 Stichwörter“ (Grimm-Online: o.S.), von denen aber viele nicht mehr
gebräuchlich sind. Von ungebräuchlichen Wörtern später noch.

Wörterbuchredaktionen wie die des Duden nehmen für den modernen
Gesamtwortschatz des Deutschen etwa 300 000 bis 500 000 Wörter an, lassen
dabei aber viele nur Experten bekannte Spezialwörter aus Fächern wie der Bio-
logie und der Chemie mit Ausnahme gebräuchlicherer Fachwörter beiseite (s.
Duden 2007: 13; Duden 2013: 134; Duden 2017: 150). Duc Ho Lee verzeich-
net in seinem als E-Buch veröffentlichten „Großen Deutschen Wortschatz“ so-
gar einen „Wortbestand von rund 700 000 Stichwörtern“ (Lee 2014: Vorwort).
Diese Wörtermengen sind abstrakte Größen. Der jeweils gesammelte und
verzeichnete Wortschatz ist eine Art kollektiver Speicher von verschiedenen
Wörtern, die von vielen verschiedenen Menschen in ihren verschiedenen Tex-
ten gebraucht worden sind. Einzelne Menschen, die 500 000 oder gar 700 000
Wörter kennen, gibt es nicht.

Forscher an der Berlin-Brandenburgischen Akademie der Wissenschaften
haben anhand großer Mengen computergespeicherter Texte aus den Jahren
1994 bis 2004 eine Gesamtmenge von 5,3 Millionen verschiedener Wörter er-

mittelt, wie sie in einem virtuellen Wörterbuch stehen könnten.⁴ Für den Beginn des 20. Jahrhunderts, also die Jahre 1905–1914, wurden nach dem gleichen Verfahren 3,7 Millionen ermittelt. Das lässt vermuten, dass der deutsche Gesamtwortschatz im Verlauf des 20. Jahrhunderts um mehr als eine Million Wörter zugenommen hat. Das Zählverfahren der Berliner Wortforscher war freilich unscharf, wie Klein (2013: 32f.) darlegt. Das liegt unter anderem daran, dass viele Wörter in Texten nicht nur als lexikalische Einheiten vorkommen, in der sie im Wörterbuch stehen oder stehen könnten. Das Verb *liegen* z.B. erscheint im Textzusammenhang auch als *liegt*, *liegtst*, *lag*, *gelegen* usw., *groß* auch als *großer*, *großes*, *größer* und *größten*, *Haus* auch als *Häuser* oder *Häusern*.⁵ Das heißt, wenn man die Anzahl der verschiedenen (lexikalischen) Wörter ermitteln will, muss man ihre in Texten vorkommenden grammatischen Formen auf ihre Grundformen zurückführen, sie also lemmatisieren, wie die Lexikographen sagen. Bei einer großen Textmenge, wie sie in Berlin genutzt wurde, würde das menschliche Bearbeiter Jahre oder gar Jahrzehnte kosten. Deshalb haben es die Berliner mit einem speziellen Computerprogramm gemacht, das aber wegen der vielen Besonderheiten unserer Wörter sehr ungenau ist. Eine große Anzahl der gezählten Wörter waren vermutlich auch Zahlen oder Namen von Menschen oder Orten usw. Viele Wörter kamen auch nur ein- oder zweimal vor, würden also nie in ein Wörterbuch aufgenommen werden. Darunter waren vermutlich auch manche Fehlschreibungen, die als gesonderte Wörter gezählt wurden. Für eine Sichtung und quantitative Bereinigung wäre eine gedruckte oder digitale Liste der ermittelten Wörter nützlich, wie sie etwa Lee (2014) bietet.

Einen ähnlichen Zuwachs an Wörtern lassen die Zahlenangaben in den beiden neuesten Ausgaben des Rechtschreib-Dudens vermuten. Die 26. Auflage von 2013 nennt als Datenbasis ein Textkorpus aus insgesamt 2,1 Milliarden Wortformen für das Frühjahr 2012, die seit 1995 erfasst worden sind. Daraus

⁴ Dabei wurde diese Zahl anhand von Befunden zu kleineren Korpora »hochgerechnet«, und nicht etwa durch Abzählen von tatsächlich im Gesamtkorpus ermittelten Wörtern. Zum Verfahren s. Klein (2013: 31f.).

⁵ Homographen wie *rasten* (*Wir wollen rasten / Sie rasten davon*), trennbare Verben wie *fortfahren* (*Fahren Sie fort / Sie sollen fortfahren*) und Einsparungen (*Reisebeginn und -ende; Ein- oder Ausgang*) spielen vermutlich für die Zählung des Wortinventars großer Textmengen nur eine geringe Rolle.

ermittelte die Duden-Redaktion „rund 10 Millionen unterschiedliche Wörter (Grundformen)“ (Duden 2013: 133f.). In den wenigen Jahren bis zur 27. Auflage von 2017 ist das Dudenkorpus offensichtlich erheblich gewachsen. Genannt wird eine Gesamtmenge von Texten im Umfang von „mehr als vier Milliarden Wortformen (Stand Frühjahr 2017)“ (Duden 2017: 148). Hierzu heißt es: „Das Dudenkorpus hat nach heutigem Stand (Frühjahr 2017) einen Umfang von knapp 23 Millionen Wörtern (Grundformen)“ (Duden 2017: 148, 150). Demnach müsste eine Erweiterung des Dudenkorpus um etwa 1,7 Milliarden Wortformen in nur vier Jahren zu einer Vergrößerung der Menge der verschiedenen Wörter von 10 auf 23 Millionen geführt haben. Da auch diese Zahlen nicht aufgrund einer Lemmatisierung von Hand, sondern durch Computerprogramme gewonnen worden sind, sind hierbei noch größere Unschärfen anzunehmen als bei den Zählungen der Berliner Wortforscher. Wie Sibylle Stadler (2014: bes. 17f.) durch vergleichende Untersuchungen ermittelt hat, liefern verschiedene Lemmatisierungsprogramme aus ein und demselben Korpus Wortbestände (Lemmastrecken), die sich nach Art und Menge stark voneinander unterscheiden. Die vom Duden genannten Wortmengen sind wie die der Berliner Gruppe vermutlich auch (noch) nicht als überprüfbare Listen verfügbar.

Doch abgesehen von solchen Unschärfen ist eine gewaltige Menge von Wörtern anzunehmen, die derzeit im Gebrauch sind – natürlich nicht im Gebrauch einzelner Menschen, sondern all derer, die an den riesigen Textmengen mitgeschrieben haben, auf die sich die Zählungen beziehen. Anzunehmen ist auch, dass die Anzahl der verschiedenen Wörter mit der Vergrößerung der hierzu genutzten Textkorpora weiterhin zunehmen wird, weil auch Wörter, die nur ein- oder zweimal verwendet worden sind, genauso gezählt werden wie neu hinzukommende Wörter, die wiederholt gebraucht werden.

Der kaum abzuschätzende Gesamtwortschatz ist ständig in Bewegung. Es bleibt zwar ein Kernwortschatz von mehreren zehntausend Wörtern über viele Jahre in Gebrauch; Teilwortschätze aber, die zu bestimmten Lebens- und Sachbereichen gehören, in denen viel Neues geschieht, verändern sich mehr oder weniger rasch. Gerade das macht die Untersuchung und Beschreibung von Wörtern so schwierig.

An den Veränderungen des deutschen Gesamtwortschatzes sind wir alle beteiligt, wenn auch die meisten von uns nur einen geringen Anteil daran haben. Die meisten Menschen verstehen zwar einige zigtausend Wörter, sprechen

und schreiben selbst aber nur einige Tausend. Und auch das sind schwankende Größen. Der Wortschatz des einzelnen Menschen ändert sich ja mit dem Lebensalter. Das gilt besonders für den passiven Wortschatz, also die Wörter, die man versteht ohne alle auch aktiv zu gebrauchen. Auch unser aktiver Wortschatz, also die Wörter, die wir sprechen und schreiben, auch dieser Wortschatz ändert sich im Lauf unseres Lebens: während der Schulzeit und des Studiums und mit immer neuen sprachlichen Erfahrungen im Beruf. Er ändert sich manchmal sogar im Lauf eines Tages. Wörter, nach denen man morgens sucht, fallen einem erst am Abend wieder ein. Mit unserem individuellen Wortschatz und seinen Veränderungen sind wir mit den anderen deutschsprachigen Menschen am Gesamtwortschatz beteiligt: an dem, was an Wörtern bleibt, was vergessen wird und was hinzukommt. Weil diese Veränderungen sich von einem Menschen zum anderen unterscheiden, sind sie als Gesamtentwicklung kaum zusammenzufassen.

Den geschriebenen und gedruckten Wortschatz einzelner Schriftsteller hat man ausgezählt. Goethe hat in seinen Schriften etwa 93 000 deutsche Wörter gebraucht (Goethe-Wörterbuch), Theodor Storm rund 22 000, Hesse in seinen Dichtungen 15 000, Rilke aber nur 5000 Wörter.⁶ Damit sind die tatsächlichen Wortschätze dieser Autoren aber nicht erfasst. Was Goethe mit seiner Mutter oder der Freifrau von Stein gesprochen hat, oder Rilke mit seiner Freundin Lou Andreas-Salomé, konnte bei diesen Zählungen natürlich nicht berücksichtigt werden. Was sich an den aktiven und passiven Wortschätzen der Autoren im Laufe ihres Lebens geändert hat, lässt sich auch kaum ermitteln.

Auch die größten Wörterbücher, gedruckt oder digital, können also gar nicht vollständig sein, weil die Menge aller Wörter der Sprache nie abgeschlossen ist, sich besonders in den Lebensbereichen ändert, in denen viel Neues geschieht, wofür die schon vorhandenen Wörter nicht ausreichen, zwar auch Altes bleibt, für das man aber gerne neue Wörter verwenden möchte. Wo kommen all die neuen Wörter her? Ergiebige Quellen für neue deutsche Wörter sind bekanntlich andere Sprachen, aus denen Wörter ins Deutsche übernommen werden. Unsere germanischen Vorfahren haben sich schon bei lateinischen und grie-

⁶ Die Zahlen zu diesen und anderen Schriftsteller nennt der Psycholinguist und Wissenschaftsjournalist Dieter E. Zimmer (1988/2012). Siehe zudem Best (2006: 16f.), der auch Zahlen zum Wortschatz anderssprachiger Autoren anführt.

chischen Wörtern bedient, und das haben auch die Deutschen im Mittelalter getan. Später kamen dann vor allem französische Wörter hinzu, und im Verlauf des 20. Jahrhunderts wurden es zunehmend Wörter aus dem britischen, später auch aus dem amerikanischen Englisch.

Anglizismen und andere ‚Fremdwörter‘ möchte ich aber hier nicht erörtern. Ich will lieber mit Beispielen aus drei sehr kleinen Ausschnitten auf den steten Wandel des deutschen Wortschatzes eingehen: der Alterung mancher Wörter, dem Aufkommen neuer Wörter und einem laufend genutzten Verfahren zur Vermehrung unseres Wortschatzes durch Schaffung neuer Wörter aus schon vorhandenen Wörtern und deren Teilen.

3. Alte und alternde Wörter

Ein wichtiger Bereich des Sprachwandels ist das Altern, das heißt, der Wegfall bestimmter Wörter aus dem allgemeinen Sprachgebrauch. Wenn wir Literatur aus dem 19. Jahrhundert oder gar dem 18. lesen, stoßen wir auf Wörter, die wir allenfalls im Textzusammenhang verstehen, aber selbst nicht gebrauchen würden. Das gilt auch etwa für Briefe unserer Großeltern, die wir nicht so leicht lesen können. Das auch, weil sie vielleicht in deutscher Schrift geschrieben sind, die zuletzt Sütterlin-Schrift genannt wurde, aber auch wegen einzelner Wörter, die wir heute nicht mehr gebrauchen. Ganz verschwunden sind aber nicht alle diese Wörter. Manche tauchen auch in neuerer Literatur wieder auf oder werden im Spaß gebraucht, wenn jemand bewusst etwas altmodisch redet oder schreibt. Hier nun meine persönliche Liste von lexikalischen Archaismen, also schon veralteten Wörtern, und Wörtern, die nach meinem Eindruck heute kaum oder nur selten noch gebraucht werden.

<i>anheischig</i>	<i>Bückling</i>	<i>frohlocken</i>
<i>Backfisch (->Teenager)</i>	<i>Dirne</i>	<i>Fuchtel</i>
<i>Barbier</i>	<i>dünken</i>	<i>fürderhin</i>
<i>Base</i>	<i>Eidam</i>	<i>Gabelfrühstück</i>
<i>Behuf/ behufs</i>	<i>Firlefanze</i>	<i>Galan</i>
<i>bisweilen</i>	<i>Flausen</i>	<i>Galoschen</i>
<i>Blaustrumpf</i>	<i>freien</i>	<i>garstig</i>

<i>Gesinde</i>	<i>meschugge</i>	<i>Schlamassel</i>
<i>Griesgram</i>	<i>Muckefuck</i>	<i>Schnurre</i>
<i>Groll</i>	<i>Muhme</i>	<i>Schwank</i>
<i>Hader</i>	<i>Müßiggang</i>	<i>spornstreichs</i>
<i>Hagestolz</i>	<i>Ölgötze</i>	<i>Subhastation</i>
<i>Hahnrei</i>	<i>Pappenstiehl</i>	<i>Taugenichts</i>
<i>Hoffart</i>	<i>pardauz</i>	<i>Tohuwabohu</i>
<i>hoffärtig</i>	<i>piesacken</i>	<i>Ulk</i>
<i>hold</i>	<i>Potztausend</i>	<i>Vagant</i>
<i>Kerbholz</i>	<i>Rabauke</i>	<i>Veitstanz</i>
<i>Kopfputz</i>	<i>Remedur</i>	<i>Wams</i>
<i>Kratzfuß</i>	<i>Rentier (Rentner)</i>	<i>weiland</i>
<i>kujonieren</i>	<i>Sapperlot</i>	<i>Wollust</i>
<i>Larifari</i>	<i>saumselig</i>	<i>Zeter</i>
<i>löcken</i>	<i>Schelm</i>	<i>(sich) ziemen</i>
<i>Maulaffen (feilhalten)</i>	<i>Schindluder</i>	<i>Zwist</i>
<i>Maulschelle</i>	<i>Schlafittchen</i>	

Ich bin auf solche Wörter vor allem bei der Lektüre älterer Romane und anderer Texte gestoßen und habe mich dabei jeweils gefragt, ob wir den betreffenden Ausdruck heute noch sagen oder schreiben würden. Aus dem deutschen Wortschatz sind solche Wörter aber nicht ganz verschwunden, eben weil sie auch heute bei der Lektüre älterer Texte wenigstens passiv gebraucht werden. Sie müssten also auch bei Wortzählungen berücksichtigt werden. Im Übrigen könnte vermutlich jeder Leser oder jede Leserin meine Liste aufgrund eigener Beobachtungen ergänzen, vielleicht auch kürzen, weil ihnen einige der aufgeführten Wörter noch nicht veraltet vorkommen. Der persönliche Wortschatz ist jedenfalls auch in seiner historischen Ausdehnung für verschiedene Menschen unterschiedlich groß und durch Lesen und Lernen veränderbar.

Noch einige Beispiele aus einem weniger persönlichen Lesebereich. Deutlich wird die Alterung von Teilen des modernen Wortschatzes auch an Wörtern, die in der DDR vor der Wende üblich waren, nun aber auch in den ostdeutschen Bundesländern gar nicht mehr oder kaum noch gebraucht werden. Beispiele sind etwa: *abkindern*, *Broiler*, *Bückware*, *Datsche*, *Grilletta*, *Kombine*, *Intershop*, *Plaste*, *Subbotnik*, *Traktorist* und viele andere (s. u.a. Hellmann 1992), die man leicht

auch in langen Listen im Internet finden kann (s. Wikipedia: Sprachgebrauch in der DDR). Bezeichnenderweise werden manche davon vom Korrekturprogramm meines Computers als Schreibfehler markiert. Ganz verschwunden sind diese Wörter aber nicht, weil sie in Erzählungen älterer Menschen aus Ostdeutschland noch vorkommen und in Romanen aus der DDR-Zeit, wenn diese heute noch gelesen werden. Auch dies heißt, dass der deutsche Gesamtwortschatz ›nach hinten‹, also in die Vergangenheit hinein nur schwer zu begrenzen ist.

4. Neue Wörter

Bei den erwähnten Zählungen der Berliner Wortforscher und der Duden-Redaktion ging es nicht um das Altern des Wortschatzes, sondern um seine Vergrößerung in neuerer Zeit. Deshalb jetzt zu neuen Wörtern. Wir werden auf Wörter nicht nur aufmerksam, wenn wir auf veraltete Ausdrücke stoßen, sondern eher noch, wenn wir Ausdrücken begegnen, die wir vorher noch nicht gehört oder gelesen haben oder die eine neue Bedeutung haben. Mir ging das in jungen Jahren so mit dem Wort *Infrastruktur*, das in den 60er Jahren des letzten Jahrhunderts aufkam, besonders in der Politik virulent wurde und dessen Bedeutung ich erst später verstand. Dann waren es neue Wörter und Wendungen, die ich zwar bald verstand, die mir aber wegen ihres zunehmenden Gebrauchs besonders auffielen wie *vorprogrammieren*, *unverzichtbar*, *nachvollziehen* (statt *verstehen*), *davon ausgehen*, *optimieren*, *Akzeptanz*, *Thematik* (statt *Thema*), *Technologie* (statt *Technik*), *Handlungsbedarf*, *Attraktivität*, *vor Ort*, *im Vorfeld*, *ein Stück weit* und manche andere, die ich aus der Literatur und auch aus dem Sprachgebrauch meiner Eltern nicht kannte. Ich habe dann vor etlichen Jahren auch einmal einen Satz aus einer fiktiven Politikerrede konstruiert, und der lautet:

Ich gehe davon aus, dass die Akzeptanz der neuen Technologie unverzichtbar ist. Es besteht aber kein Handlungsbedarf, die Attraktivität der Thematik vor Ort oder schon im Vorfeld zu optimieren, weil sonst eine schwer nachvollziehbare Problematik ein Stück weit vorprogrammiert wäre.

Das ist nicht besonders sinnvoll, fällt aber beim nur flüchtigen Hinhören kaum auf, wie ich einige Male ausprobiert habe. Heute könnte man vielleicht noch Wörter und Wendungen hinzufügen wie *zeitnah*, *alternativlos*, *gut aufgestellt*

und *gut unterwegs*. Ich will damit aber keine Diskussion sprachlicher Stereotype eröffnen. Die Beispiele sollen nur auf den Wortschatzwandel in einem bestimmten Lebensbereich hinweisen: im öffentlichen Sprachgebrauch. Besonders bemerkenswerte Wörter, die 2001 bis 2010 in Gebrauch gekommen sind, werden von Steffens / al-Wadi (2013) beschrieben.⁷

5. Kultur-Wörter

Nun zu einem besonders produktiven Verfahren, das wir neben der Entlehnung aus anderen Sprachen zur Vermehrung unseres Wortschatzes nutzen, der Bildung neuer Wörter aus alten Wörtern oder deren Teile, dabei vor allem die Produktion zusammengesetzter Wörter. Hierzu beschränke ich mich auf eine zunächst sehr klein scheinende Gruppe von Beispielen, die ich mir etwas näher angesehen habe. Und das, weil sie mir in letzter Zeit besonders aufgefallen sind.

Es ist die Wortvermehrung mit Hilfe von *Kultur*, also die Bildung neuer Wortausdrücke, besonders Substantivkomposita, mit dem schönen Wort *Kultur*. Damit lassen sich nach meinem Eindruck besonders leicht neue Wörter bilden. Seit langem in Gebrauch sind Komposita wie *Baumkultur*, *Kulturvolk*, *Unkultur* und wohl auch *Nacktkultur*. Inzwischen ist erheblich mehr *Kultur* im Deutschen entstanden. Nicht als Kultiviertheit des Sprachgebrauchs, sondern als bloßer Ausdruck, als Teil zusammengesetzter Wörter. Die *Kulturwissenschaft*, zu der sich auch Teile der Germanistik entwickelt haben, kann sich nicht mehr beschränken auf *Sprachkultur*, auf *Schreib- und Lesekultur*, auf *Brief-, Buch- und Theaterkultur*. Es gibt längst auch die *Alltagskultur*, die *Medienkultur* und die *Streitkultur*.

Mir scheint die Ausbreitung von *Kultur* im Deutschen endemisch geworden zu sein etwa seit der Zeit, in der sich auch *interkulturelle* Kommunikation und *Interkulturalität* als Themen in den Geistes- und Sozialwissenschaften ausgebreitet haben. Also *Kultur* auch außerhalb der Botanik, der *Baum- und Pilzkultur*, und auch nicht beschränkt auf *Hochkulturen* wie die der Ägypter, Inder und Griechen, von denen schon lange geredet und geschrieben wird. *Kultur* scheint sich vielmehr auf nahezu alle Bereiche etwas länger andauernder menschlicher

⁷ Die betreffende Forschungsgruppe arbeitet weiter; neue Funde sind online zugänglich (s. Neologismenwörterbuch).

Aktivitäten auszubreiten, an denen mehr als eine Person beteiligt ist. Dies aber wohl nur in der deutschsprachigen Welt; denn – soweit ich von anderen Sprachen weiß – gibt es nirgendwo so viel *Kultur* wie im Deutschen. Rückläufige Wortlisten verzeichnen eine Vielzahl von Komposita mit *-kultur*, darunter altbekannte wie *Wohnkultur*, *Mono-*, *Trivial-* und *Agrikultur*, *Ess-* und *Reinkultur*, neuere Bildungen wie *Schuld-* und *Schamkultur*, *Konsum-*, *Nischen-*, *Familien-*, *Klang-*, *Schlaf-*, *Lauf-* und *Fahrkultur* bis hin zu relativ neuen Bildungen wie *Gefühls-*, *Bestattungs-*, *Verkaufs-*, *Biergarten-*, *Sexual-*, *Pop-*, *Spaß-* und *Eventkultur*, seit wenigen Jahren auch *Willkommenskultur*. Zur Erweiterung meiner eigenen kleinen Liste habe ich *-kultur* im „Rückläufigen Wörterbuch“ von Duk Ho Lee (2005: 969f.) nachgeschlagen. Dort finden sich 105 entsprechende Komposita.

Zu geradezu absurd vielen Belegen führt eine Suche in DeReKo, dem Deutschen Referenzkorpus des Instituts für Deutsche Sprache (IDS). Aus den im Archiv des IDS verfügbaren digitalen Korpora zur geschriebenen Sprache habe ich ein leichter erschließbares großes Teilarchiv im Umfang von 10,3 Milliarden Wortformen gewählt und mit COSMAS II, dem institutseigenen System zur Korpusverwaltung und -analyse, nach Wörtern durchsucht, die auf *-kultur* enden.⁸ Die automatische Suche in dieser riesigen Textmenge ergab die verblüffend große Anzahl von über 23 000 verschiedenen Komposita mit *-kultur* als Endglied. Darunter sind hochfrequente Wörter wie *Popkultur*, *Leitkultur*, *Subkultur* und *Unternehmenskultur*, die jeweils mehrere tausendmal in den Korpora vorkommen, aber auch Wörter wie *Affennierenzellkultur*, *Afterworkkultur*, *Aquisitionskultur* und *Alles-auf-Pump-Kaufkultur*, die je nur einmal belegt sind. 55,1%, das heißt, mehr als die Hälfte der insgesamt gefundenen *-kultur*-Wörter kommen in der ganzen Textmenge nur einmal vor.⁹ Sie sind einmalige Bildungen (Hapaxlegomena), die nicht weiter verwendet worden sind. Die Verteilung der verschiedenen *-kultur*-Komposita nach der Häufigkeit ihres Vorkommens in DeReko entspricht in etwa dem sog. Zipschen Gesetz, wie es auch für andere korpusbezogene Worthäufigkeiten ermittelt worden ist (s. Engelberg 2015). Hier sollen mit dem Hinweis auf den hohen

⁸ Der praktische Zugang zu den Korpora des IDS mit COSMAS II wird ausführlich beschrieben in Bodmer Mory (2014). Perkuhn / Belica (2016) beschreiben und diskutieren Recherchen in DeReKo und deren Auswertung am Beispiel des Lemmas *Konflikt*.

⁹ Ich danke Rainer Perkuhn (IDS) für die Berechnung des Anteils der Hapaxlegomena mit *-kultur*.

Anteil der Hapaxlegomena nur die schier unbegrenzten Erweiterungsmöglichkeiten betont werden, welche die Wortbildung für den deutschen Wortschatz bietet.

Mit einer Erweiterung der digitalen Textmenge würden sicherlich neue (zuletzt) nur einmal belegte Wörter hinzukommen. Doch auch mit einer noch längeren Wortstrecke, die aus neuen Texten und Gesprächen zu ermitteln wäre, hätte die *Kultur* noch kein Ende, weil wir je nach Situation und Thema leicht eine weitere *-kultur* erfinden könnten. In einem Vortragsraum könnte man vielleicht von *Belüftungskultur* sprechen. Dieses Wort habe ich in den verfügbaren Wortlisten nicht gefunden. Nicht als ob es ein besonders nützliches Wort wäre; manch andere *Kultur*wörter lassen ohne ihre Kontexte auch keinen Nutzen erkennen.

Noch produktiver ist die Zusammensetzung von Wörtern, die mit *Kultur*-beginnen. Im „Großen deutschen Wortschatz“ von Duk Ho Lee (2014) sind 826 *Kultur*-Komposita aufgeführt, darunter sehr gebräuchliche wie *Kulturdenkmal*, *Kulturgut* und *Kulturgeschichte*. Mir waren bisher nicht geläufig unter anderen: *Kulturdolmetscher*, *Kulturfreak*, *Kulturfrau* und *Kulturroschen*. Wörter wie *Kulturnetzwerkträger* oder gar *Kulturwissenschaftskommunikationsportal* (Lee 2014) werden wahrscheinlich nicht mehr oft verwendet. Eine Suche mit COSMAS II in dem schon vorher genannten Teilarchiv von DeReko ergab über 40 000 verschiedene Komposita mit *Kultur*- als Erstglied, auf die einzugehen aber hier vom Thema wegführen würde.¹⁰ Mit den Komposita, die auf *-kultur* enden oder beginnen, sind nicht alle Wortbildungsmöglichkeiten erschöpft. Zu erwähnen ist ja noch *Kultur* innerhalb zusammengesetzter Wörter wie *Jugendkulturtag* und *Multikulturalismus*. Ich belasse es aber bei diesen beiden Beispielen.

Neben den offensichtlich gern genutzten Wortzusammensetzungen mit *-kultur*- gibt es viele andere produktive Muster zur Bildung neuer Wörter, etwa mit *-problem* oder *-krise*. Leser mögen vielleicht selbst auch die Zunahme von Adjektiven beobachten, die auf *-freundlich* enden, von *anlegerfreundlich*, über *autofreundlich* und *hautfreundlich* bis *magenfreundlich*, *nutzerfreundlich*, *reformfreundlich* und *umweltfreundlich*.¹¹

¹⁰ Sprach- und kulturgeschichtlich interessant wäre eine Untersuchung der *Kultur*wörter nach Vorkommen und Häufigkeit in verschiedenen Textsorten und das auch in zeitlicher Dimension.

¹¹ In den Texten des schon vorher genutzten DeReKo (Stand Frühjahr 2018) sind insgesamt 4487 verschiedene Adjektivkomposita zu finden, die auf *-freundlich* enden und ihrer Auswertung harren. Inzwischen sind vermutlich weitere ‚freundliche‘ Wörter hinzugekommen.

6. Der offene Wortschatz

Mit diesen Beispielen will ich eine Besonderheit der deutschen Sprache hervorheben, die der laufenden Erweiterung des deutschen Wortschatzes dient und für dessen kaum zu ziehende Grenzen verantwortlich ist, nämlich die schöne Möglichkeit, jederzeit und bei jeder Gelegenheit aus schon vorhandenen Wörtern oder deren Teilen neue Wörter zu bilden. Manche davon überleben den einmaligen Gebrauch nicht; andere geraten für kurze oder längere Zeit in den allgemeinen Sprachgebrauch und vielleicht auch in die Wörterbücher. All dies gehört zu den Gründen, warum sich die Frage „Wie viele Wörter hat die deutsche Sprache?“ nicht kurz und bündig beantworten lässt, mit einer belastbaren Zahl schon gar nicht. Der deutsche Wortschatz hat anders als eine Schatztruhe keinen Deckel. Er ist offen und wird je nach Bedarf der Sprecher und Schreiber in Teilbeständen umgeschichtet und erweitert. Dies fällt unter anderem wegen der hierfür günstigen Eigenschaften der deutschen Wortbildung besonders leicht.

Computerlinguisten und Lexikographen bleibt die Aufgabe, anhand der laufend vergrößerten Korpora die weiter wachsenden riesigen Wortmengen mit ihren sich ständig ändernden Rändern zu ermitteln und daraus für Forschungsprojekte wie für gedruckte und digitale Wörterbücher jeweils eine praktikable Auswahl von Wörtern zu treffen und deren Eigenschaften im Interesse der bekannten oder vermuteten Nutzer zu beschreiben.

7. Literatur

- Best, Karl-Heinz (2006): *Quantitative Linguistik – Eine Annäherung*. 3. Aufl. Göttingen: Peust u. Gutschmidt.
- Bodmer Mory, Frank (2014): *Mit COSMAS II „in den Weiten der IDS-Korpora unterwegs“*. In: Institut für Deutsche Sprache (Hrsg.): *Ansichten und Einsichten – 50 Jahre Institut für Deutsche Sprache*. Mannheim: Institut für Deutsche Sprache, 376–385.
- Bußmann, Hadumod (Hrsg.) (1990): *Lexikon der Sprachwissenschaft*. 2., völlig neu bearb. Aufl. Stuttgart: Kröner.
- DeReKo=Institut für Deutsche Sprache (2004ff.): *Deutsches Referenzkorpus – DeReKo*. Archiv der Korporageschriebener Gegenwartssprache. Mannheim: Insti-

- tut für Deutsche Sprache. <http://www1.ids-mannheim.de/kl/projekte/korpora> (gesichtet am 29.11.2018); Wortsuche mit COSMAS II: <https://cosmas2.ids-mannheim.de/cosmas2-web/> (Registrierung erforderlich).
- Duden (1999). Das große Wörterbuch der deutschen Sprache. 10 Bände. Mannheim / Leipzig / Wien / Zürich: Dudenverlag (Band 1).
- Duden (2007): Deutsches Universalwörterbuch. 6., überarb. und erw. Aufl. Mannheim / Leipzig / Wien / Zürich: Dudenverlag.
- Duden (2013): Die deutsche Rechtschreibung. 26., völlig neu bearb. u. erw. Aufl. Berlin / Mannheim / Zürich: Dudenverlag.
- Duden (2017): Die deutsche Rechtschreibung. 27., völlig neu bearb. u. erw. Aufl. Berlin: Dudenverlag.
- Engelberg, Stefan (2015): Quantitative Verteilungen im Wortschatz. Zu lexikologischen und lexikografischen Aspekten eines dynamischen Lexikons. In: Eichinger, Ludwig M. (Hrsg.): Sprachwissenschaft im Fokus. Positionsbestimmungen und Perspektiven. (Jahrbuch des IDS 2014). Berlin / München / Boston: De Gruyter, 205–230.
- Goethe-Wörterbuch = Berlin-Brandenburgische Akademie der Wissenschaften / Akademie der Wissenschaften in Göttingen / Heidelberger Akademie der Wissenschaften (Hrsg.) (1966ff.): Goethe-Wörterbuch. Stuttgart / Berlin / Köln / Mainz: Kohlhammer; Projektbeschreibung unter: <http://www.bbaw.de/forschung/gwb/> (gesichtet am 29.11.2018).
- Grimm-Online=Kompetenzzentrum Trier(2019):DasdeutscheWörterbuchvon Jakob und Wilhelm Grimm. <http://dwb.uni-trier.de/de/das-woerterbuch/das-dwb/> (gesichtet am 29.11.2018).
- Haß-Zumkehr, Ulrike (2001): Deutsche Wörterbücher. Berlin / New York: De Gruyter.
- Hellmann, Manfred W. (1992): Wörter und Wortgebrauch in Ost und West. Ein rechnergestütztes Korpus-Wörterbuch zu Zeitungstexten aus den beiden deutschen Staaten. Tübingen: Narr (Forschungsberichte des Instituts für deutsche Sprache 69).
- Klein, Wolfgang (2013): Von Reichtum und Armut des deutschen Wortschatzes. In: Deutsche Akademie für Sprache und Dichtung / Union der deutschen Akademien der Wissenschaften (Hrsg.): Reichtum und Armut der deutschen Sprache. Erster Bericht zur Lage der deutschen Sprache. Berlin / Boston: De Gruyter, 15–55.

- Lee, Duk Ho (2005): Rückläufiges Wörterbuch der deutschen Sprache. Berlin: De Gruyter.
- Lee, Duk Ho (2014): Großer deutscher Wortschatz. Einleitungsbroschüre und DVD. Mannheim: Institut für Deutsche Sprache.
- Neologismenwörterbuch (2005ff.). In: OWID-Online Wortschatz-Informationssystem Deutsch. Mannheim: Institut für Deutsche Sprache. <http://www.owid.de/wb/neo/start.html> (gesichtet am 29.11.2018).
- Perkuhn, Rainer / Belica, Cyril (2016): Konflikt, Sprache, korpuslinguistische Methodik. In: Vogel, Friedemann / Luth, Janine / Ptashnyk, Stefaniya (Hrsg.): Linguistische Zugänge zu Konflikten in europäischen Sprachräumen. Heidelberg: Winter, 339–364.
- Stadler, Heike (2014): Die Erstellung der Basislemmaliste der neuhochdeutschen Standardsprache aus mehrfach linguistisch annotierten Korpora. OPAL (Online publizierte Arbeiten zur Linguistik) 5/2014. Mannheim: Institut für Deutsche Sprache: <https://pub.ids-mannheim.de/laufend/opal/opal14-5.html> (gesichtet am 29.11.2018).
- Steffens, Doris / al-Wadi, Doris (2013): Neuer Wortschatz. Neologismen im Deutschen 2001–2010. 2 Bde. Mannheim: Institut für Deutsche Sprache.
- Wikipedia (o.J.): Sprachgebrauch in der DDR. https://de.wikipedia.org/wiki/Sprachgebrauch_in_der_DDR (gesichtet am 29.11.2018).
- Zimmer, Dieter E. (1988/2012): Weißt Du, wieviel Wörtlein – Zahlen, Fakten und Vermutungen über ein immer wichtigeres Thema: Die Menge der Wörter. In: Die ZEIT Nr. 45/1988, aktualisiert im ZEIT-Archiv unter: <https://www.zeit.de/1988/45/weisst-du-wieviel-woertlein/komplettansicht> (gesichtet am 29.11.2018).