

«Bericht zur Lage der deutschen Sprache»

Akademien legen quantitative Studien zu Teilaspekten vor

Von Daniel Goldstein u. a.¹

Wenn ein Staatsoberhaupt zur Lage der Nation redet, erwartet man eine Standortbestimmung und ein Programm. Was aber darf man von einem «Bericht zur Lage der deutschen Sprache» erwarten, wie er jetzt erstmals vorliegt²? Dass hier jemand diese Sprache in ihrer Gesamtheit erfasst, ihr Wohl und Wehe beurteilt und auch noch sagt, was zu tun sei? Wohl kaum. Die Akademien, die sich die Herausgabe einer Reihe solcher Berichte vorgenommen haben, sind in ihrem Anspruch bescheidener. Sie stellen fest, es werde «über den Einfluss des Englischen, über den Verfall oder die Verarmung des Deutschen, die schwindende internationale Geltung unserer Sprache und den Verlust von Verwendungsdomänen wie der Wissenschaft gestritten»; es fehlten aber «zu einem guten Teil ... fundierte Diagnosen».

Zu Teilthemen «von besonderem Interesse» sollen nun «wissenschaftlich fundierte Informationen» zusammengetragen werden. Der erste Bericht gilt Klagen, die seit Langem für Aufregung sorgen: angebliche Verarmung in Wortschatz und Grammatik, Bedrohung durch Anglizismen und «Papierdeutsch» (mit Ausdrücken wie «zur Anzeige bringen»). Die vier Studien, die in der Folge vorgestellt werden, sind in Projektgruppen unter professoraler Leitung entstanden. Gemeinsam ist ihnen, dass sie auf der computergestützten Analyse grosser Textmengen beruhen. Auf die Schwierigkeiten dieses Verfahrens gehen die vier Projektverantwortlichen selber ein. Wichtiger als die Präzision der gewonnenen Zahlen sind ohnehin die linguistischen Überlegungen, welche die Autoren an Trends und Textbeispiele anknüpfen.

1 Vom «Sprachspiegel»-Redaktor stammen Einleitung sowie Abschnitte 1 und 2; der Autor und die Autorin der Abschnitte 3 und 4 sind dort angegeben.

2 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. De Gruyter, Berlin 2013. 233 Seiten, ca. Fr. 42.– (auch als E-Book erhältlich).

1. Von Reichtum und Armut des deutschen Wortschatzes³

«Weisst du, wie viel Wörtlein stehen in dem Buch der deutschen Sprach'?» Gäbe es diese Liedzeile, so müsste die Fortsetzung heute lauten: «Com-pu-ter hat sie gezählet . . .». Wolfgang Klein sagt es nicht so, aber seine Projektgruppe geht die Frage nach dem Umfang des Wortschatzes und seiner Veränderung im 20. Jahrhundert mithilfe von Korpora an, elektronisch durchsuchbaren Textsammlungen. Mit dem «Buch» der deutschen Sprache ist also nicht ein Wörterbuch gemeint, obwohl Klein eingangs eine Zählung anführt, die im umfangreichsten, dem Digitalen Wörterbuch der deutschen Sprache (DWDS), auf 336 925 Stichwörter kommt. Diese auf die Brüder Grimm zurückgehende Sammlung «spiegelt jedoch im Wesentlichen nur den deutschen Wortschatz bis etwa 1900 wider». Die heutige deutsche «Alltagssprache» wird von der Dudenredaktion auf eine halbe Million Wörter geschätzt, von denen 70 000 den «zentralen Teil» ausmachen. «Nimmt man den «Fachwortschatz» dazu, so kommt man hinwieder auf einige Millionen Wörter.»

Was da gezählt wird, sind Lexeme, wie sie uns in Wörterbüchern als Stichwörter (Lemmata) begegnen; bei flektierbaren Wörtern in der Grundform. Gleichlautende Wörter verschiedener Wortarten, etwa «sein», oder unterschiedlicher Bedeutung, etwa «Absatz», erhalten separate Wörterbuch-Einträge. Die vorliegende Korpus-Auswertung orientiert sich «an den geschriebenen Wortformen», ohne Unterschiede der Bedeutung oder grammatischen Funktion, weil es dazu keine verlässlichen Computerprogramme gebe. Immerhin wird versucht, Beugungen und getrennte Verbformen zu erkennen.

Die Grösse der verwendeten Korpora bemisst sich nicht nach der Anzahl darin vorkommender Lexeme – wie (un)präzis auch definiert und erkannt –, sondern nach der Anzahl Wörter, wie sie auch ein Schreibprogramm zählt: Nach jedem Leerschlag kommt ein neues. Da die gesprochene Sprache erst schlecht in elektronischen Sammlungen erfasst ist, beschränkte man sich auf die geschriebene. Den Grundstock

3 Das Buchkapitel ist abrufbar: <http://pubman.mpdl.mpg.de> (Suche: Klein Armut).

bildet das Kernkorpus der DWDS-Redaktion, das etwa zu gleichen Teilen aus Belletristik, Zeitungen, Gebrauchstexten und wissenschaftlichen Texten besteht. Ergänzt durch neuere Archive von vier deutschen Zeitungen, umfasst es eine Milliarde Wörter.

Da die rückwirkend erfassten Texte aus vordigitaler Zeit weniger umfangreich sind, wurden für die Analyse der Wortschatz-Entwicklung drei «Zeitscheiben» à 10 Millionen Wörter gebildet, für Anfang, Mitte und Ende des 20. Jahrhunderts. Nun ist es aber so, dass mit der Grösse des Korpus auch die Zahl der darin vorkommenden Lexeme stark steigt, weil die Chance wächst, dass auch sehr seltene Wörter einmal erfasst werden. Es gibt dafür eine Umrechnungsformel, und so wurden die drei Zeitscheiben auf ein Korpus von einer Milliarde Wörter hochgerechnet, wie es eben fürs Ende des Jahrhunderts tatsächlich vorliegt und den errechneten Wert ungefähr bestätigt.

Rechnet man Zahlen, Abkürzungen, Namen, Fehlschreibungen, Zitate aus Fremdsprachen aufgrund von Stichproben ab, so kommt man ums Jahr 2000 auf 5,3 Millionen Lexeme, die als «deutsche Wörter» gelten können. Im Jahrzehnt vor dem Ersten Weltkrieg waren es 3,7 und Mitte des Jahrhunderts 5,0 Millionen. Weniger wichtig als die – mit den genannten Unsicherheiten behafteten – absoluten Zahlen ist für den Autor der Trend (einmal angenommen, die Mängel wirkten sich für alle Zeitscheiben etwa gleich aus): «Dieses Anwachsen ... wird man nicht als eine Verarmung betrachten wollen».

Auch wenn eine Detailanalyse per Computer unmöglich ist, lässt sich sagen, dass der «weitaus grösste Teil des Zuwachses» aus Ableitungen und Zusammensetzungen bestehender Wörter besteht. Überall dort, wo sich der Sinn solcher Wortbildungen erst aus dem Zusammenhang erschliesst, nennt Klein sie «eine echte Erweiterung der Lexik». Nach Textsorten betrachtet, war der Zuwachs in der Belletristik am kleinsten, bei den Zeitungen am grössten, was der Autor mit der Themenvielfalt erklärt. «Rätselhaft» ist für ihn, dass der ermittelte Wortschatzumfang seit der Jahrhundertmitte bei der Belletristik stagniert; zur Verlangsamung des Zuwachses insgesamt äussert er sich nicht.

Vielmehr betont Klein, «dass der deutsche Wortschatz sehr viel umfangreicher geworden ist. Nur sehr wenige Sprachen weisen einen solchen Reichtum in ihrer Lexik auf.» Zwar nähmen auch in andern Sprachen Komposita zu, aber selten so stark wie im Deutschen (kein Wunder, denn es zeichnet sich ja gerade durch diese Möglichkeit aus). Wenn trotzdem zuweilen der Eindruck der Verarmung entstehe, so liege das nicht an der Sprache, «sondern an denen, die von ihr Gebrauch machen. Es reicht nicht, einen Bösendorfer in der Stube stehen zu haben; man muss ihn auch spielen können.»

2. Anglizismen im Deutschen

Auch Peter Eisenberg unterscheidet in seinem Kapitel über Anglizismen zwischen Sprachkritik und Sprachgebrauchskritik, denn er befürchtet, die Diskussion über die Einflüsse des Englischen könnte das Vertrauen in die eigene Sprache beschädigen. Die Frage, was die deutsche Sprache «an sich» sei, ist zwar theoretisch schwierig, aber für die Praxis dieser Studie einfach zu beantworten: Sie ist das, was im Textkorpus erfasst ist. Verwendet werden diesmal die 1. und die 3. Zeitscheibe (siehe oben).

Auch die Frage, was ein Anglizismus sei, wirft viele weitere Fragen auf. Hier wurde nicht auf die Herkunft abgestellt, sondern «im Sinne der Projektziele ist ein Anglizismus ein Wort, das erkennbar Eigenschaften des Englischen hat, die in der Kerngrammatik des Deutschen nicht erfasst werden können». Das schliesst «Pseudoanglizismen» wie «Handy» ein, ebenso Zusammensetzungen wie «hochladen». Als «relevante Wortarten» sind Substantive, Adjektive, Verben und Adverbien erfasst; es sind (in dieser Reihenfolge) jene mit der grössten «Entlehnbarkeit».

Unter den gut 10 Millionen Wörtern des Korpus 1905–1914 erkennt der manuell unterstützte Computer rund ein halbes Promille derartige Anglizismen. Der einzelne Anglizismus wird relativ selten verwendet (in zwei Dritteln der Fälle gar nur ein einziges Mal). Für den Wortschatz zählt aber die Rarität ebenso als Bestandteil wie etwa

das massenhaft verwendete «und». Unter den rund 370 000 Lemmata beträgt der Anglizismen-Anteil 0,35 Prozent, unter jenen der relevanten Wortarten 0,5 Prozent. Im Durchschnitt taucht eines dieser Wörter auf jeder fünften Seite (à 500 Wörter) auf.

Am Ende des Jahrhunderts haben sich die Anteile etwa verzehnfacht: gut zwei Anglizismen pro Seite, 3,5 Prozent des Wortschatzes (5,2 beim relevanten). Dass auch der Duden von 1986 rund 3,5 Prozent Anglizismen aufweist, wertet der Autor «erst einmal als Zufall», der näher zu untersuchen wäre. Innerhalb der Anglizismen ist die Hierarchie insofern gewahrt, als immer noch zwei Drittel der überhaupt vorkommenden im ganzen Korpus nur einmal festgestellt werden. Am andern Ende der Skala hat sich aber eine Anzahl «hochfrequenter» Wörter etabliert, die bestimmte englische Eigenschaften bewahren.

In jenem Viertel des Korpus, das aus Presstexten besteht, ist der Anglizismenanteil deutlich höher: 1,2 Prozent der insgesamt gezählten Wörter (hier passend Tokens genannt), gegenüber 0,5 Prozent im Durchschnitt; in Belletristik und Gebrauchsprosa dürfte der Wert deutlich darunter liegen, in der Wissenschaft nicht unbedingt. Dazu schreibt Eisenberg: «Die Presse ist seit jeher ein Einfallstor für Fremdwörter aus der jeweils dominierenden Gebersprache. (...) Es besteht keine Berechtigung, aus einer repräsentativen Mischung von Presstexten auf «das geschriebene Standarddeutsche» zu schliessen.»

Hier zeigt sich wieder die Sorge des Autors, nicht die Zunahme der Anglizismen, wohl aber die Kritik daran könnte die Sprache schwächen: «Abgesänge auf das Deutsche sind nicht nur fehl am Platz, sondern sie schädigen die Loyalität der Sprecher zu ihrer Sprache.» Da ist es nicht mehr weit zum Vorwurf, die Kritiker seien Nestbeschmutzer. Dabei teilt Eisenberg die Kritik am Sprachgebrauch, «etwa wo einem präventösen Globalismus gehuldigt wird oder wo Texte gezielt unverständlich gemacht werden». Aber er hält die deutsche Sprache für widerstandsfähig: «Anglizismen stehen unter erheblichem Integrationsdruck der Kerngrammatik»; ihr «struktureller Einfluss (ist) marginal».

3. Die Entwicklung der Flexion⁴

Nicht umsonst ist die Wendung vom «Dativ, der dem Genitiv sein Tod» sei, in der neueren populären Sprachkritik zum geflügelten Wort geworden. Die Verhältnisse im Kern der Grammatik – und darum geht es ja in diesem griffigen Spruch – scheinen so grundlegend und fest zu sein, dass Veränderungen in diesem Bereich als Symptome dafür gelesen werden, wie tief das tradierte System der Sprache von einem Wandel betroffen sei, in dem früher vorhandene Ausdrucksmöglichkeiten verloren gingen.

So werde zum Beispiel der Genitiv von Konstruktionen mit der Präposition *von* verdrängt (*das Haus von meinem Vater*), das Bewusstsein dafür schwinde, ob nach Präpositionen wie *wegen*, *trotz* oder auch *entlang* der Genitiv oder der Dativ zu stehen habe, statt eigentlich richtigen *dieses Jahres* oder *jedes Jahres* heiße es *diesen Jahres* und *jeden Jahres*, bei den Verben nutze niemand mehr die starken alten Formen wie *buk* (→ *backte*) oder *schwor* (→ *schwörte*).

Diese und eine Reihe verwandter sprachlicher Erscheinungen werden in ihrer Menge gerne als Symptome eines Verlusts an systematischen Möglichkeiten der Sprache wahrgenommen. Wenn man die davon berührten grammatischen Zusammenhänge genauer betrachtet, sieht man zum Ersten, dass all diese Erscheinungen und das Aufkommen von Varianten Folge und Ausläufer einer ganz langfristigen grammatischen Reorganisation und **Resystematisierung** im Rahmen der Schriftsprachentwicklung sind. Ein Teil der Erscheinungen kommt zum Zweiten erst jetzt so recht ins Bewusstsein. Erst in neuerer Zeit benutzen ganz viele Sprecher eine dem Standarddeutschen angenäherte Form und bringen dabei Erscheinungen ein, die Regularitäten der **Sprechsprache** darstellen, in der Schriftsprache aber allenfalls am Rand eine Rolle spielen.

4 Die Herausgeber des Buchs haben einige Monate vor der Publikation Kurzfassungen der Projektleiter mit «vorläufigen Ergebnissen» als Pressematerial abgegeben. Diese werden im Folgenden abgedruckt, mit minimalen Anpassungen aufgrund der entsprechenden Buchkapitel. Abschnitt 3 stammt von Prof. Ludwig M. Eichinger, Direktor des Instituts für Deutsche Sprache.

So läuft es – was das System angeht – z. B. bei der Bildung der Kasus des Substantivs darauf zu, dass die entsprechende Endung eigentlich nur einmal (und das nicht direkt am Substantiv, sondern am Artikel oder am Adjektiv) auftritt. So ist es systemgerecht, wenn wir nicht mehr wie Schiller *dem Manne* (*kann geholfen werden*) schreiben, sondern *dem Mann* ohne das Endungs *-{e}*. Eigentlich ist dann aber eine Form wie *des Jahres* auffällig; sie bleibt aber bestehen (und wird nicht zu: *des Jahr*). Das ist schon ein Indiz dafür, dass die Sprachgemeinschaft für diesen Kasus einen eigenen, anderen Platz gesucht hat.

Was den Gebrauch angeht, so kommen hier oft an der Oberfläche Dinge zusammen, die dann doch nicht ganz dasselbe sind. Zum Beispiel ist in vielen alltäglichen gesprochenen Formen des Deutschen der Genitiv nicht so geläufig, so dass es dort nicht *der Besuch der Tante*, sondern *der Besuch von der Tante* heisst (was schriftsprachlich zumindest deutlich stilistisch markiert ist). Andererseits gibt es auch im schriftlichen System Kontexte, in denen man (fast) nicht anders kann, als die Konstruktion mit *von* zu verwenden (*die Lieferung von Kohle* oder *die Lieferung Obst[s?]/von Obst*; vgl. *ein Glas Weins/ein Glas Wein*). Was gleich aussieht, ist also nicht immer dasselbe.

In den Texten der **Korpora** geschriebener Sprache von Anfang, Mitte und Ende des zwanzigsten Jahrhunderts, die wir daraufhin untersucht haben, sind die «verschwindenden» Fälle schon zu Beginn des Jahrhunderts selten, und sie wurden dann noch seltener. Häufig haben sie aber inzwischen eine spezielle Funktion erhalten, bei der ihre «Auffälligkeit» eher nützlich ist. So bekommt zum Beispiel der Genitiv einen klaren Platz als Präpositionalkasus bei lange Zeit strittigen Fällen (etwa: *entlang*, *laut*⁵). Dass die Zahl der starken Verben nun seit längerem eigentlich stabil ist, zeigt andererseits, dass «System» nicht alles ist: Die jetzigen starken Verben sind so zentral und häufig, dass sie gerade durch Formenvielfalt (und Seltenheit) zeigen, wie wichtig sie sind.

5 Anmerkung der Redaktion:

Der Duden führt für *entlang* den Genitiv als Hauptvariante an, den Dativ als seltene bzw. (in Nachstellung) schweizerische; für *laut* nennt er den Dativ an erster Stelle, dazu «auch mit Genitiv».

4. Variation im Wortschatz am Beispiel *Streckverbgefüge*⁶

Wendungen wie *eine Entscheidung treffen* oder *zum Ausdruck bringen* gelten in der Sprachkritik als überflüssige «Streckformen» der zugrunde liegenden Verben (*entscheiden, ausdrücken*), als «Sprachbeulen» oder «Verfettung am Leibe unserer Sprache». Man wertet sie als Symptom für das sogenannte «Papierdeutsch», gegen das bis heute auch die im Internet publizierten Stil- und Schreibratgeber zu Felde ziehen. Die linguistische Forschung bemüht sich seit langem, die pauschale Abwertung dieser Gefüge durch eine empirisch fundierte Beschreibung zu ersetzen. Das Projekt setzt auf den bisherigen Ergebnissen auf und untersucht die folgenden Aspekte auf einer sehr breiten Datengrundlage:

- (1) **Zeitliche Entwicklung:** Wie verändern sich der Bestand und die Gebräuchlichkeit der Gefüge im Laufe des 20. Jahrhunderts?
- (2) **Textsortenspezifik:** Wie verteilen sich die Belege über verschiedene Textsortenbereiche?
- (3) **Funktionen und Verwendungsmotive:** Kann man Streckverbgefüge in allen Fällen durch einfache Verben ersetzen? Falls ja: Mit welchem Aufwand und Ergebnis?

Neben den drei «Zeitscheiben» des Berichtskorpus nutzt das Projekt das DWDS-Kernkorpus mit 100 Millionen Wortformen, das zehn Dekaden und vier Textsortenbereiche (Belletristik, Wissenschaft, Gebrauchsliteratur, Zeitung) in ausgewogenem Verhältnis abdeckt. Ausserdem konnten wir Daten aus einem 80 Millionen Textwörter grossen juristischen Korpus (Ulrich Heid, Universität Stuttgart) und aus einer linguistisch aufbereiteten Kopie der deutschen Wikipedia (Iryna Gurevych, Torsten Zesch, TU Darmstadt) mit ca. 550 Millionen Textwörtern nutzen.

Unsere Studien zur Frequenzentwicklung liefern keine Anhaltspunkte dafür, dass Streckverben seltener oder häufiger verwendet

6 Verfasserin dieses Abschnitts ist Prof. Angelika Storrer (Universität Mannheim).

werden als zu Beginn des 20. Jahrhunderts. Bemerkenswert ist, dass die Anteile im Wikipedia-Korpus ähnlich hoch liegen wie im DWDS-Kernkorpus: Offensichtlich werden die «papierdeutschen» Streckverben in digitalen Medien weiterhin verwendet, selbst beim dialogischen Schreiben auf Wikipedia-Diskussionsseiten. Für die Entwicklung der Gefüge gilt dasselbe wie für andere Wortschatzeinheiten auch: Sie kommen in Gebrauch und können auch wieder in Abgang kommen – das Gefüge *in Abgang kommen* selbst ist im DWDS-Kernkorpus letztmalig im Jahr 1957 belegt.

In der Belletristik findet man weniger Streckverben als in den anderen Textsortenbereichen; sie sind aber durchaus auch in literarischen Werken belegt. Verwunderlich ist der hohe Anteil in der journalistischen Prosa; man hätte einen höheren Anteil in wissenschaftlichen Fachtexten und im Bereich «Gebrauchsliteratur» erwartet. Einige Gefüge entwickeln eine Bedeutung, die gar nicht mehr mit dem zugrunde liegenden Verb paraphrasierbar ist. Ein Beispiel hierfür ist das Gefüge *Absage erteilen*, das in fast allen Belegen in einem Sinne verwendet, der sich mit *absagen* gar nicht mehr ausdrücken liesse: Wenn ein Politiker einem Krieg eine Absage erteilt, heisst das schliesslich nicht, dass er einen Krieg absagt.

Unsere qualitativen Studien zeigen generell, dass man die Gefüge im Kontext betrachten muss, wenn man ihren Wert oder Unwert beurteilen will. Oft sind sie im Textzusammenhang nur mit grossem Aufwand durch ein einfaches Verb ersetzbar, z. B. weil das Nomen um Adjektivattribute oder Relativsätze erweitert ist, wie in folgendem Beispiel: *Die Vereinigten Staaten begannen beizeiten, sich eingehend über die Deutschlandfrage zu informieren. Im Jahre 1946 traf^{en} wir dann eine schwere **Entscheidung**, die sich auf Material stützte, das damals noch unvollständig war.*

Die Feinanalysen zeigen, dass Streckverbgefüge erweiterte Möglichkeiten bieten, um Informationen zu komprimieren und Informationsteile zu gewichten; in diesen Funktionen bereichern sie den deutschen Wortschatz um zusätzliche Wahlmöglichkeiten.