Wolfgang Klein (Nijmegen)

Das Wörterbuch der Zukunft ist kein Wörterbuch*

Abstract: Den Wortschatz einer Sprache auf hohem Niveau zu dokumentieren und in all seinen Eigenschaften zu beschreiben, ist gleichermaßen wichtig wie schwierig. Verschiedene Gründe haben dazu geführt, dass die Tradition der großen Wörterbücher derzeit zusammenbricht. An ihre Stelle werden in der Zukunft flexibel handhabbare digitale lexikalische Systeme treten.

1 Vorbemerkung

Als mich Ludwig Eichinger im letzten Herbst gebeten hat, auf der Jahrestagung zum fünfzigjährigen Bestehen des Instituts für Deutsche Sprache über Lexikografie zu sprechen, habe ich mich sehr geehrt, zugleich aber auch etwas unbehaglich gefühlt. Ich bin kein gelernter Lexikograf, und wenn ich mich zweimal im Leben ein wenig mit der Kunst des Wörterbuchmachens befasst habe, so hat sich das eher durch Zufall als aus eigenem Antrieb ergeben. Immerhin habe ich bei diesen beiden Gelegenheiten die eine oder andere Erfahrung auf diesem Gebiet gemacht und mir das eine oder andere gedacht; davon will ich im Folgenden reden. Es hat ja auch seine Vorteile, wenn man die Dinge aus der Warte des Außenseiters betrachtet: Man weiß weniger, hat aber eine bessere Chance als der Innenseiter, den Wald zu sehen und nicht oder nur die Bäume; man fährt nicht auf den eisernen Schienen der Tradition, sicher geführt aber mit wenig Freiheiten, nach anderem zu schauen, das vielleicht auch wichtig wäre; und man kann sich ungescheut zur eigenen Laienhaftigkeit bekennen, statt mit dem Anspruch der papaler Unfehlbarkeit zu reden. So ist, was ich hier zur schwierigen Gegenwart und zur vielleicht hoffnungsvolleren Zukunft der Lexikografie sagen möchte, ganz auf meine eher untypischen Erfahrungen gegründet – erklärtermaßen Ausdruck einer subjektiven Sehweise, die man teilen kann oder auch nicht. Aber

* Das Folgende ist eine nachträgliche Rekonstruktion des Vortrags, den ich auf der Jahrestagung des IDS gehalten habe. Leider habe ich keine Aufzeichnungen dazu, und so ist es möglich, dass es einige kleinere Abweichungen von dem, was ich dort gesagt habe, gibt. Die Grundgedanken sind aber mit Sicherheit dieselben. Ich danke den Teilnehmern herzlich für die sehr lehrreiche Diskussion.

das ist ja der Sinn des wissenschaftlichen Diskurses, anderen dar zu tun, was man für richtig hält, und sie dazu aufzufordern, es zu prüfen und sich dazu zu stellen.

2 1967 ff.

Im Jahre 1967, als das IDS noch in den Inkunabeln lag, wurde neben vielem anderen auch ein Vorhaben mit dem Namen ‚Grunddeutsch‘ auf den Weg gebracht. Angeregt wurde es von dem amerikanischen Germanisten Alan Pfeffer, und es hatte zum Ziel, für den Sprachunterricht ein Inventar der wichtigsten deutschen Wörter zusammenzustellen, das nicht auf barer Einschätzung, sondern auf harten Fakten über die tatsächliche Nutzung der Wörter beruht. Wie soll man den Deutschunterricht sinnvoll gestalten, wenn man nicht weiß, welche Wörter am wichtigsten zu wissen sind? Anders als bei Pfeffer sollte es dabei nicht um die gesprochene, sondern um die geschriebene Sprache gehen. Einer der Beteiligten war auch mein Lehrer Hans Eggers, freilich nicht in seiner Rolle als Mediävist, sondern weil er mit seiner Arbeitsgruppe in Saarbrücken als einer der ersten ein gut ausgesuchtes digitales Korpus der deutschen Gegenwartssprache aufgebaut hat, das statistische Analysen erlaubte. Gedacht war dieses Korpus ursprünglich für grammatische Untersuchungen; aber genauso gut, so die Überlegung, könnte man es auch für die Lexik verwenden. Und so hat Hans Eggers mich, empfohlen von Alois Brandstetter, gefragt, ob ich nicht als studentische Hilfskraft an diesem Vorhaben mitarbeiten wolle. Das kam ein wenig überraschend, denn ich verstand nicht das Geringste von der Sache und habe mich damals für ganz andere Dinge interessiert. Aber ich war jung und brauchte das Geld, und so habe ich „adsum“ gesagt. Das Projekt war wissenschaftlich ein mäßiger Erfolg, aber ich selbst habe viel daraus gelernt, unter anderen, weshalb es ein so mäßiger Erfolg geworden ist. Eine der Erfahrungen hat nichts mit der Lexikografie im Besonderen zu tun, aber da sie meine Denkweise überaus geprägt hat, will ich sie doch erzählen, vielleicht dass sie auch anderen von Nutzen ist. Damals hat „Der Spiegel“ einen ausführlichen Bericht über die Eggers’sche Arbeitsstelle gebracht; darin war nichts falsch, und alles war falsch. Jeder Satz traf zu, aber als jemand, der die Sache von innen kannte, hatte man den Eindruck, es sei von einem ganz anderen Vorhaben die Rede. Da habe ich gelernt, dass eines die Wahrheit der Sätze ist, ein anderes die Wahrheit eines Textes. Und ich habe gelernt, dass man den Medien nicht trauen darf, selbst wenn die Absichten die besten sind.

Aber in erster Linie habe ich einiges über Lexikografie gelernt, über die Schwierigkeiten, aus digitalen Korpora verlässliche Informationen zu ziehen, aber

auch über die grundsätzlichen Probleme, mit denen man zu kämpfen hat, wenn man den Wortschatz einer Sprache umfassend und wissenschaftlich sauber beschreiben will. Das hat den Keim zu einer Überzeugung gelegt, die sich im Laufe der Jahre immer weiter verfestigt hat und die mit der herrschenden Meinung der strukturellen Linguisten einerseits, der traditionellen Lexikografen andererseits nicht ganz im Einklang steht. Es ist dies die der Überzeugung, dass (a) die systematische und adäquate Beschreibung des Lexik nicht nur von höherer praktischer Bedeutung ist als jede andere Aufgabe der Sprachwissenschaft (vielleicht ist sie sogar die einzige, die einen wirklichen Nutzen für die Menschheit hat), (b) dass es die wichtigste wissenschaftliche Aufgabe der Sprachwissenschaft ist, und (c) dass es auch die schwierigste unter allen sprachwissenschaftlichen Tätigkeiten ist. Ersteres würden vielleicht die meisten Linguisten bejahen, wenn auch nicht gern, denn man möchte ja die eigene Arbeit allemal für besonders wichtig halten. Das Zweite würden nur wenige so sehen, das Verfertigen von Wörterbüchern gilt heutzutage vielen nicht als die Krone der linguistischen Forschung, sondern eher als Fleißarbeit, nach Schillers Motto: „Wo die Könige bauen, da haben die Kärner zu tun.“ Das gilt auch für das Dritte: Wieso soll es schwieriger sein, die Bedeutung oder die sonstigen Eigenschaften von Wörtern zu beschreiben als beispielsweise die Abfolge im Mittelfeld oder die Funktion des Perfekts? Das ist ein weites Feld. Ich will aber immerhin meine seit damals gewachsene und verfestigte Meinung an drei Punkten erläutern und begründen.

A. Wie kann man die Eigenschaften eines Wortes korrekt beschreiben? Man denkt darüber nach, man schaut sich einige oder auch viele Belege an, in denen das Wort vorkommt, und dann schreibt man hin, was es bedeutet. Das aber ist unheimlich schwierig, wenn man es näher besieht, vor allem dann, wenn man sich nicht mit ein paar ersten Angaben bescheidet, sondern wenn die Beschreibung den üblichen Maßstäben wissenschaftlicher Absicherung und Präzision genügen soll. Das gilt auch dann, oder vielleicht gerade dann, wenn man sich auf viele und gute Belege für die Verwendung in unterschiedlichen Textsorten und in unterschiedlichen kommunikativen Zusammenhängen stützen kann. Was wir außer unserer erlernten Sprachkenntnis haben, ist (bei geschriebenen Belegen) eine Folge von Zeichen auf dem Papier, die in einer längeren Folge von solchen Zeichen eingebettet ist: *ist zwivel herzen nachgebür, daz muoz der sele werden sur*. Was wir beschreiben möchten, sind nicht die Eigenschaften der Zeichenfolge *zwivel*, sondern die ihr zugrunde liegende lexikalische Einheit. Die kann man nicht sehen oder hören. Es ist eine höchst komplexe, in der Regel nur in den Köpfen der Sprecher gespeicherte abstrakte Entität, die aus mehreren Bündeln von Eigenschaften besteht: (a) phonetischen; (b) graphematischen, wenn es sich – wie im betrach-

teten Fall – um geschriebene Sprache handelt; (c) morphologischen; (d) syntaktischen; (e) semantischen. Hinzu treten Eigenschaften, die nicht definierend, aber dennoch für eine umfassende Charakterisierung der lexikalischen Einheit wichtig sind, etwa jene, die ihre Herkunft, Verbreitung oder Stilebene betreffen. Woher weiß man denn eigentlich, dass eine solche Einheit „veraltend“, „gehoben“ oder „obszön“ ist, und wie kann man beispielsweise ihre Bedeutung – neben der Orthografie das Herzstück der klassischen Lexikografie – so angeben, dass sie sich von der Bedeutung aller anderen Wörter unterscheidet? Denn das würde man ja von einer wirklich sauberen semantischen Analyse verlangen. Wirklich synonyme Wörter gibt es nur, wenn man sich mit einer gewissen Tiefe der Analyse begnügt und konnotative und andere Besonderheiten der Verwendung außer Acht lässt. Aber auch die gehören zur lexikalischen Bedeutung, auch sie bestimmen das Ausdruckspotenzial einer Sprache. Eine wirklich umfassende Analyse müsste daher auch den Unterschied von *schon* und *bereits*, von *beginnen* und *anfangen*, von *Zahnarzt* und *Dentist* angeben, auch wenn dies mehr ist, als sich der gewöhnliche Nutzer von einem Wörterbuch erwartet, der es vielleicht gar nicht so genau wissen will (oder eben doch). Aber die Schwierigkeiten fangen schon bei viel bescheideneren Ansprüchen an, bzw. dort beginnen sie bereits.

Vor vielen Jahren hat mich einer meiner Söhne, damals vier Jahre alt und seit einiger Zeit mit einer anderen Sprache, dem Niederländischen, konfrontiert, gefragt: „Was ist denn eigentlich ein Körper?“ Was er offenbar wissen wollte, war nicht, was das Wesen eines Körpers ausmacht, sondern was das deutsche Wort *Körper* im Gegensatz zum niederländischen *lichaam* bedeutet. Da hat er mich kalt erwischt, als Vater wie als Sprachwissenschaftler. Was eigentlich ist die Bedeutung des deutschen Wortes *Körper*? Ich weiß es natürlich, ich kann ja Deutsch, und es ist ein ganz alltägliches Wort, das ein jeder kennt, der Deutsch kann. Es ist wie mit des Augustinus' bekanntem Wort über die Zeit: *Quid est tempus? Scio, si nemo ex me quaeret; sed si quaerenti explicare velim, nescio*. Ich weiß auch, dass das Wort *Körper* viele Bedeutungen hat. Manche davon stellen sich rasch ein, wenn man nach seiner Bedeutung gefragt wird; man denkt vielleicht zuerst an *Körper* im Gegensatz zu *Geist*. Aber woran man zunächst denkt, kann ja nicht ausreichen. Ein Körper kann lebendig und tot sein (ursprünglich galt, wie heute noch für *corpse* im Gegensatz zu *body*, nur letzteres), organisch und unorganisch (Planeten sind Körper), massiv oder nicht (der Wein hat Körper), materiell und immateriell (die platonischen Körper sind nicht materiell). Ich glaube sogar zu verstehen, was Goethe sagen wollte, als er über Wilhelm Grimms Ausgabe dänischer Heldenlieder lobend schrieb: „Es ist Ihnen gelungen, durch eine geschickte Behandlungsweise aus dem vielen Einzelnen einen ganzen Körper zu bilden.“ Und wenn ich all das weiß und dann noch viele Belege zu Rate ziehe, wie kann ich dieses Wissen – ein winziger Ausschnitt des lexikalischen Wissens all jener,

die Deutsch können –, so beschreiben, dass es klar und verständlich und überprüfbar ist? Hier mache ich eine Pause, damit Sie ein wenig nachdenken können. ... Ich habe meinem Sohn gesagt: „Das ist das, was man bei dir so anfassen kann.“, und ihm über die Schulter gestrichen. Das hat ihm eingeleuchtet, und er war zufrieden. Es hat ihm deshalb eingeleuchtet, weil man als Kind die Bedeutung eines Wortes immer so lernt, dass man in einem bestimmten Zusammenhang versteht, was gemeint ist, und dieses Wissen Schritt für Schritt auf andere Vokommen ausweitet, in denen das Wort vielleicht gleich, vielleicht auch ein wenig anders verwendet wird. So ist vielleicht auch die Aufgabe eines geschickten Lexikografen, der den praktischen Nutzen und nicht die Ansprüche des Wissenschaftlers vor Augen hat, den gedachten Nutzer auf die rechte Spur zu setzen, also einen Anfangsschritt zu definieren, der es diesem Nutzer erlaubt, sich allmählich die verschiedenen Verwendungen zu erschließen und so schrittweise in seinem Kopf einen „ganzen Körper“, nämlich die lexikalische Einheit zu bilden. Das tun wir alle im Kindesalter wie im Erwachsenenalter, denn neue Wörter lernt man bis zum Lebensende, und ebenso justiert man sein bisheriges Wissen nach (*zeitnah* statt *rasch*). Die semantischen Eigenschaften auch nur einer einzigen lexikalischen Einheit systematisch und erschöpfend zu analysieren und dann darzustellen, ist hingegen ein unendlich schwieriges, ein schier hoffnungsloses Unterfangen. Ich will das an einem anderen Beispiel erläutern, gleichfalls einem gängigen Wort, das ein jeder kennt und ständig benutzt.

Vor einiger Zeit habe ich mich einmal mit den syntaktischen und semantischen Eigenschaften von Partikeln wie *auch*, *nur*, *schon*, *noch*, *wieder* und ähnlichen befasst. Dazu gibt es eine ausgedehnte linguistische Literatur und natürlich mehr oder minder detaillierte Angaben in allen Wörterbüchern unserer Sprache. Viele dieser Analysen sind hilfreich, keine ist wirklich befriedigend. Eine gemeinsame syntaktische Eigenschaft dieser Partikel ist, dass sie – hier ist das Deutsche weitaus flexibler als viele andere Sprachen – in verschiedenen Positionen vorkommen und damit einhergehend die Gesamtbedeutung des Satzes in unterschiedlicher Weise beeinflussen:

- (1) (a) Goethe trug noch einen Zopf.
- (b) Noch Goethe trug einen Zopf.
- (c) Noch trug Goethe einen Zopf.

Das ist, was die syntaktischen Eigenschaften der Partikel angeht, ein Ärgernis, weil es, wie (1b) und (1c) zeigen, gegen die eisenharte Vorfeldregel des Deutschen verstößt: Ist *noch* ein Satzglied oder nicht, oder mal dies, mal das? Immerhin, man kann diese syntaktische Eigenschaft leicht beschreiben, was immer man daraus für die deutsche Grammatik macht. Was die semantischen Eigenschaften

betrifft, so ist der Unterschied bei den drei Positionen für jeden, der Deutsch kann, fasslich, aber sie ist sehr schwer präzise anzugeben. (Man versuche es.) Und wieso ist (2b), anders als (1b), so seltsam?

- (2) (a) Goethe war noch krank.
- (b) Noch Goethe war krank.
- (c) Noch war Goethe krank.

Der einzige Unterschied liegt zwischen *einen Zopf tragen* und *krank sein* und der Art und Weise, wie beide Prädikate mit *noch* interagieren. Steht *noch* in einer anderen Position, ergibt sich kein solcher semantischer Effekt. Welche lexikalische Bedeutung müssen wir dem Wort *noch* zuweisen, um ihn zu erklären? Selbst in den besten Wörterbüchern werden diese und ähnliche Fragen nicht einmal andeutungsweise behandelt. Das heißt nicht, dass diese Wörterbücher schlecht sind. Es heißt aber, dass sie allesamt die lexikalische Bedeutung von *noch* – und das ist für die anderen Fokuspartikel ganz ähnlich – nicht wirklich erfassen, und das nicht deshalb, weil die Lexikografen unfähig sind, sondern weil die Aufgabe, die lexikalische Bedeutung eines Wortes wirklich in den Griff zu bekommen, so unendlich schwierig ist. Es ist, und das ist die eigentliche Lehre für mich gewesen, es ist die schwierigste Aufgabe der Linguistik überhaupt. Das mag nicht für alle Wörter gelten, aber für sehr viele, sehr vertraute, die jedes Kind im Schulalter kennt und richtig verwendet.

B. Soviel zu dem Problem, eine einzelne lexikalische Einheit befriedigend zu beschreiben. Nun hat eine Sprache wie das Deutsche sehr viele solcher Einheiten. Das zehnbändige „Große Wörterbuch der deutschen Sprache“ des Dudenverlags hat, Querverweise und Namen nicht gerechnet, etwa 170.000 Stichwörter. Die Schöpfer des von den Brüdern Grimm begonnenen (oder angefangenen) „Deutschen Wörterbuchs“ haben im Verlauf von über hundert Jahren 337.500 Stichwörter in Bedeutung und Entwicklung beschrieben; hinzu kommen etwa 15.000 weitere, die nicht als Stichwörter ausgezeichnet, aber doch mitbehandelt sind, insgesamt also 350.000. Das nicht nur nach meiner Meinung beste Wörterbuch einer Sprache überhaupt, das „Oxford English Dictionary“ (OED), beschreibt etwa 600.000 lexikalische Einheiten des Englischen. Hat das Deutsche auch so viele Wörter? Nein, es hat wesentlich mehr, wenn man sich anschaut, wieviele solcher Einheiten nach Ausweis größerer Korpora tatsächlich verwendet werden. Eine neuere Untersuchung (Klein 2013) kommt auf etwa fünf Millionen lexikalischer Einheiten, Dialektwörter, ausgesprochene Fachtermini und Kollokationen (im Sinne von Einheiten, die syntaktisch zusammengesetzt sind, aber Wortbe-

deutung haben wie *sich die Kante geben*) nicht gerechnet. Fast alle davon sind natürlich Ableitungen oder Komposita. Die Zahl der „einfachen Wörter“ liegt eher im vierstelligen Bereich. Es wäre aber eine Illusion, zu glauben, die Art und Weise, wie Ableitungen und Komposita verwendet werden, ließe sich normalerweise aus der Bedeutung ihrer Bestandteile ableiten. Ein Übersetzer, der die Bedeutung von *ein* und *Waage*, von *Bürger* und *Amt*, von *Plan* und *Quadrat*, von *fremd* und *schämen* kennt, kann mit diesem Wissen allein noch lange nicht die Wörter *Einwaage*, *Bürgeramt*, *Planquadrat* oder *fremdschämen* übersetzen. Er kann vermuten, was sie bedeuten könnten, aber nicht sagen, wie sie tatsächlich verwendet und verstanden werden. Sie müssen daher als eigenständige lexikalische Einheiten gelten und, will man in der Tat das lexikalische Repertoire einer Sprache vollständig erfassen, in ihrer besonderen Bedeutung beschrieben werden. Das berührt nicht den Umstand, dass fast alle Komposita auch eine kompositionelle Bedeutung haben und in dieser verwendet werden können. Wie soll man das Bein einer Elfe anders nennen als Elfenbein?

Wie lange würde es wohl dauern, auch nur, sagen wir, ein Fünftel der nachweislich im gegenwärtigen Deutschen verwendeten Wörter, also eine Million, korrekt zu beschreiben und mit einigen Belegen zu veranschaulichen? Nehmen wir an, es stünde bereits ein umfassendes, gut ausgewähltes Korpus zu Verfügung, dem alle erforderlichen Informationen entnommen werden können, und nehmen wir an, dass ein erfahrener Lexikograf im Schnitt eine Stunde für ein Wort benötigt. Das sind acht am Tag. Ein Jahr hat 220 Arbeitstage, macht 1760 Wörter im Jahr. Er schafft also in hundert Jahren 176.000, soviel wie sich im zehnbändigen „Großen Wörterbuch der deutschen Sprache“ finden, aber weitaus weniger als die erwähnte Million. Immerhin: 5,68 Lexikografen könnten es bei dieser Geschwindigkeit in einem Jahrhundert schaffen, etwa ein Fünftel der derzeit tatsächlich verwendeten lexikalischen Einheiten des Deutschen zu beschreiben. Aber sie müssten zügig arbeiten, tagaus tagein pro Wort eine Stunde, und es darf niemand krank werden. Jetzt kann man diese einfache Rechnung in Beziehung zu dem unter A. genannten Problem setzen, auch nur die semantischen Eigenschaften so alltäglicher Wörter wie *Körper* oder *noch* anzugeben, und man verzagt vollends. In hundert Jahren gibt es ganz andere Wörter, wenn bis dahin das Deutsche nicht außer Gebrauch gekommen ist. Man beginnt, eine gewisse Sinnlosigkeit in dem Unterfangen zu sehen, den deutschen Wortschatz relativ umfassend (damit sind hier 20% gemeint) lexikografisch darzustellen – jedenfalls wenn man an die Methoden der tatsächlich vorkommenden klassischen Lexikografie denkt.

Diese Überlegung galt den im Gegenwartsdeutschen verwendeten Wörtern. Sie gilt nicht anders für die historische Lexikografie. Die erste Ausgabe des „Deutschen Wörterbuchs“ zu erarbeiten, hat mehr als hundert Jahre in Anspruch genommen: Die erste Lieferung ist 1852 erschienen, die letzte im Jahre 1961. Es ist

trotz vieler Schwächen im Einzelnen in seiner Gesamtheit ein ungemein bewundernswertes Werk, in meinen Augen die bedeutendste Leistung der gesamten wissenschaftlichen Germanistik. Aber zum Zeitpunkt seines Abschlusses war es bereits völlig überholt; vom „Nachsommer“ zur „Blechtrommel“ ist der deutsche Wortschatz nicht gleich geblieben. Der letzte noch von Jacob Grimm selbst bearbeitete Buchstabe F ist 1863 im Druck erschienen; die gesamte Entwicklung des Deutschen seither ist nicht berücksichtigt, und auch in den späteren Lieferungen stammen nur die wenigsten Belege aus dem 20. Jahrhundert. So haben denn Ende der Fünfzigerjahre die Akademien in Berlin und in Göttingen eine gemeinsame Neubearbeitung zumindest der Buchstaben A–F beschlossen und in Angriff genommen. Die erste Lieferung mit dem Buchstaben A ist 1965 erschienen, die letzte mit dem Buchstaben A im Jahre 2007; insgesamt sind in diesen 42 Jahren etwa 8000 mit A beginnende Stichwörter beschrieben worden. Man kann sich ausrechnen, wie lange eine vollständige Neubearbeitung mit der Abdeckungsdichte der Erstbearbeitung (wie gesagt, etwa 350.000 Wörter) dauern würde. Jacob Grimm hat für ebendiesen Buchstaben A etwa zwei Jahre gebraucht, auch hier nach Erscheinungszeit der Lieferungen gerechnet. Das liegt nicht an den heutigen Lexikografen; es liegt daran, dass sie mit so unendlich viel mehr Material zu tun haben als Jacob Grimm. Er war so schnell fertig, weil er im Verhältnis relativ wenige Daten zur Verfügung hatte. Es macht einen Unterschied, ob man für ein Wort 10 Belege oder 10.000 Belege bearbeiten soll. Je mehr es zu berücksichtigen gilt, je mehr man weiß, desto schwieriger wird die Aufgabe, und wenn man, wie wir heute, riesige Korpora zur Verfügung hat und damit eine unvergleichlich bessere Datenbasis, dann wird sie vollends utopisch. Die Sprache entwickelt sich schneller, als man sie beschreiben kann. Die Neologismen, die heute beschrieben werden, sind die von gestern und vorgestern.

C. Gibt es für dieses Massenproblem, sei es in der gegenwartsbezogenen, sei es in der historischen Lexikografie, überhaupt eine Lösung? Wenn ja, dann nur mit einer anderen Vorgehensweise, beispielsweise mit Methoden der Computerlinguistik. Das war schon die Vision derer, die seinerzeit das IDS-Projekt ‚Grunddeutsch‘ begonnen haben: Wir bauen ein möglichst repräsentatives Korpus auf und zählen mit den Instrumentarien der „elektronischen Sprachverarbeitung“ aus, welche davon besonders häufig vorkommen. Das ist ein weitaus bescheideneres Ziel als eine umfassende lexikografische Analyse dieser „wichtigsten Wörter“; eine wirklich umfassende syntaktische oder semantische Beschreibung war auch gar nicht avisiert. Aber Morphologie, Syntax und Semantik kommen unweigerlich ins Spiel, wenn man nicht nur die vorkommenden Wortformen, reine Zeichenfolge also, zählen will, sondern die ihnen zugrunde liegenden lexikalischen Ein-

heiten. Dazu muss man zumindest (a) die Wortformen auf ihre Grundform zurückführen – *wäscht, gewaschen, wüscht* auf *waschen*; (b) sie desambiguieren – *sein* auf das Hilfsverb *sein* oder auf das Possessivpronomen *sein*, und (c) in gewissen Grenzen auch syntaktisch analysieren – *vor* als Präposition oder als abgetrennten Teil eines Verbs (*er schlägt xxx vor*). Der Lexikograf, wenn er Belege durchsieht, macht all dies mit seinem Kopf, und das zumeist ohne große Mühe. Dem Computer fiel es damals und fällt es heute schwer. Die automatische Lemmatisierung, die seinerzeit in der Eggers'schen Arbeitsstelle entwickelt wurde, war für die Zeit erstaunlich gut, wenn die Flucht der Jahre hier nicht meinen Blick verklärt; aber die Fehlerquote war dennoch sehr hoch. Gut zwanzig Jahre haben wir später am Max-Planck-Institut für Psycholinguistik die Datenbank CELEX aufgebaut. Ziel war es, für experimentelle Zwecke verlässliche Angaben über die Frequenz wichtiger Wörter (im Deutschen wie in anderen Sprachen) machen zu können, eine Aufgabe, die also manches mit dem Projekt ‚Grunddeutsch‘ gemein hatte, bei der man sich aber auf weitaus größere, nicht zuletzt vom IDS stammende Datenbestände stützen konnte. Anders als bei ‚Grunddeutsch‘ kam das Vorhaben zu einem guten Abschluss, und von den Ergebnissen wurde in der Psycholinguistik rege Gebrauch gemacht. Aber manches war schon bizarr, beispielsweise, dass die der bäuerlichen Welt zugehörigen Verben *heuen* und *rechen* offenbar zu den häufigsten Verben deutscher Zunge zählen. Das hätte man sich so nicht gedacht. Es lag aber daran, dass der Computer das Adverb *heute* als Flexionsform von *heuen* („die Bäurin heute den ganzen Tag“) ansah und die flektierten Adjektive *rechte* und *rechten* als Flexionsformen von *rechen*. Seither sind wiederum dreißig Jahre verstrichen, und man mag hoffen, dass die nunmehr in der Computerlinguistik entwickelten Verfahren zur automatischen morphologischen und syntaktischen Analyse stabil mit einer Fehlerquote von, sagen wir, fünf Prozent funktionieren. Dann hätte man bei einem Korpus von 100 Millionen Wortformen (das Deutsche Referenzkorpus des IDS hat derzeit 24 Milliarden) also nur noch 5 Millionen Fehler. Leider sind wir nach meinem Wissensstand von diesem Ziel noch weit entfernt; die Fehlerquote schwankt je nach Verfahren, Textsorte und Zählweise, liegt aber allemal deutlich höher. Und ebensowenig gibt es meines Wissens – ich würde mich aber freuen, unrecht zu haben – eine wirklich befriedigende automatische syntaktische Analyse des Deutschen, die auf größere Textmengen anwendbar wäre.

Das heißt nicht, dass solche automatischen Analysen für die Zwecke des Lexikografen sinnlos sind, ganz im Gegenteil: Ohne eine solche Zuarbeit kann man nicht hoffen, das unter B. angedeutete Massenproblem in den Griff zu bekommen. Automatische Analysen können aber entweder nur ein erster Schritt sein, auf den die manuelle Arbeit aufsetzen muss, oder aber man muss versuchen, die Aufgaben der Lexikografie überhaupt mit anderen Mitteln anzugehen.

Die drei Probleme, die ich hier skizziert habe, sind grundsätzlicher Art. Sie haben nichts oder doch wenig mit der menschlichen Unvollkommenheit auch der besten Lexikografen zu tun. Es ist einfach unendlich schwierig, auch nur eine einzelne lexikalische Einheit in ihren verschiedenen Eigenschaften verlässlich und umfassend zu beschreiben. Es ist unrealistisch, dies für den ganzen zu einer bestimmten Zeit verwendeten Wortschatz einer Sprache wie Deutsch tun zu wollen; dazu ist er zu umfangreich, und der Belege sind zu viele. Die Methoden der Computerlinguistik sind ein Hilfsmittel für die Arbeit des Lexikografen; ihr Nutzen ist jedoch begrenzt, denn sie sind nach wie vor sehr fehleranfällig; Besserung ist nicht in Sicht. All dies sind wenig ermutigende Einsichten, aber es sind Einsichten, denen man sich stellen muss.

3 Zwischenspiel

Das Motiv hinter dem Projekt ‚Grunddeutsch‘ war ebensosehr ein praktisches wie ein wissenschaftliches, so wie es denn immer schon zu den Zielen des IDS gehört hat, nicht nur Wissenschaft für Wissenschaftler zu betreiben, sondern auch all jenen, die sich ansonsten für die deutsche Sprache interessieren, von Nutzen zu sein. Es sollte ein Grundbestand von lexikalischen Einheiten ermittelt werden, den man dann dem Fremdsprachenunterricht zugrunde legen konnte. Wenn man eine Sprache lernt, sei es die erste im Kindesalter von seinen Eltern, Geschwistern und der sonstigen sozialen Umgebung, sei es eine zweite im Unterricht oder gleichfalls durch den alltäglichen Kontakt, dann muss man eine ganze Reihe von Dingen lernen – Aussprache, Intonation, Morphologie, Syntax, Wortschatz, schließlich die verschiedenen Diskursfähigkeiten. All dies ist wichtig, wenn man sich in dieser Sprache andere verstehen und sich selbst verständlich machen will. Am wichtigsten aber ist der Wortschatz. Das kann man sich sehr einfach überlegen. Wenn jemand von einer Sprache, sagen wir Russisch, hunderttausend Wörter kennt, aber keine einzige für das Russische spezifische grammatische Regel (also nicht jene, die universal sind, denn die kann man ja ohnehin), dann könnte man in Omsk, Tomsk und Novosibirsk einkaufen oder nach dem Weg fragen, man könnte erzählen, was einen dorthin verschlagen hat, dass es einem gefällt oder auch, dass man eine unglückliche Kindheit gehabt hat, man könnte die Zeitung lesen oder sogar Tolstoi – all dies mit Einschränkungen, Missverständnissen und ohne Glanz, aber man könnte sprachlich überleben, und mehr als das. Wenn man hingegen alle morphologischen und syntaktischen Regeln des Russischen fein säuberlich gelernt hätte, aber kein einziges Wort oder vielleicht gerade so viele, als nötig sind, um diese Regeln zu lernen, dann käme man nicht sehr weit. Um es

etwas zugespitzt zu sagen: Von den beiden fundamentalen Komponenten einer Sprache, der Grammatik und der Lexik, ist keine entbehrlich; aber es ist die Lexik, die für das Ausdrucksvermögen einer Sprache maßgeblich ist und die die menschliche Kommunikation trägt. Dieses offenkundige Faktum steht in einem eigentümlichen Missverhältnis zu den Präferenzen der meisten strukturellen Linguisten wie auch der Spracherwerbsforschung. Die wenigsten Untersuchungen gelten den Eigenschaften der Lexik oder ihrem Erwerb im Kindesalter oder, bei einer Zweitsprache, im Erwachsenenalter. Auch in der modernen Semantik geht es seit vielen Jahren fast ausschließlich um die Bedeutung von einigen morphologisch oder syntaktisch zusammengesetzten Konstruktionen. Die lexikalische Semantik, die doch allem zugrunde liegt, spielt hingegen, sieht man von einigen Funktionswörtern ab, nur eine sehr geringe Rolle. Ich weiß nicht wirklich die Gründe, die zu dieser Unterbelichtung der Lexik in der Forschung geführt haben. Wahrscheinlich ist es zum einen das Verharren in einer Denkschule, zum andern aber der Umstand, dass die wissenschaftliche Erforschung der lexikalischen Semantik zu schwierig ist – also ebendas, was im vorigen Abschnitt gesagt wurde. Und so kneifen wir.

4 1997 ff.

Im Frühjahr 1997 hat mich Dieter Simon, damals Präsident der Berlin-Brandenburgischen Akademie der Wissenschaften, also jene auf Leibniz zurückgehende Akademie, die auch Jacob und Wilhelm Grimm zu ihren Mitgliedern zählte und das von ihnen begonnene Wörterbuch vor gut hundert Jahren unter ihre Fittiche nahm, gefragt, ob ich bereit sei, an einem neuen Wörterbuch der deutschen Gegenwartssprache mitzuarbeiten. Die Anregung zu einen solchen Vorhaben stammte, wie ich später erfahren habe, von Hartmut Schmidt, zu jener Zeit Leiter der Abteilung Lexik beim IDS. Die Frage kam ein wenig überraschend, denn ich hatte zwar, anders als 1967, eine gewisse Meinung vom Wörterbuchmachen, habe mich jedoch für ganz andere Dinge interessiert. Aber die Interessen ändern sich, wenn man älter wird, und so habe ich nach einigem Überlegen „adsum“ gesagt und das auch nicht zurückgenommen, als Dieter Simon mich umgehend ungefragt zum Leiter des Projekts erklärt hat. Freilich habe ich daraufhin zwei Bedingungen genannt, nämlich zum einen, dass die Leitung eine ganz informelle sein soll, und zum anderen, dass dieses Wörterbuch von Anfang an kein klassisches Wörterbuch zum Ziel hat, sondern auf dem Computer entstehen soll, als ein über Internet zugängliches digitales System, von dem man vielleicht ein gedrucktes „Wörterbuch“ im klassischen Sinne ableiten kann, aber nicht muss. Dies war der Ursprung des

„Digitalen Wörterbuchs der deutschen Sprache“ (DWDS), das seither in Berlin entstanden ist. Nun will ich hier nicht im einzelnen über das DWDS reden, so nahe es meinem Herzen steht. Es ist, wie ich glaube, eine Erfolgsgeschichte, und es wird auch reich genutzt. Aber es gibt auch anderswo ähnlich gute Vorhaben, nicht zuletzt am IDS selbst, und um fair zu sein, müsste man dann einen Überblick geben. Das aber ist nicht das Ziel dieses Beitrags (Näheres zum DWDS findet sich in Klein/Geyken 2010 sowie unter www.dwds.de). Vielmehr will ich auf einige grundsätzliche Punkte eingehen, die bei allen Bemühungen in diese Richtung eine Rolle spielen. Ich möchte jedoch zumindest jene erwähnen, die über die Jahre hinweg das Projekt getragen haben, weil sonst der Eindruck entstehen mag, als sei es mein Kind. Das ist natürlich nicht so. Von Anfang an beteiligt waren Manfred Bierwisch, Alexander Geyken, Wolf-Hagen Krauth, Hartmut Schmidt, Angelika Storrer und ich; etwas später sind Thomas Gloning und Lothar Lemnitzer hinzugekommen. Wie man sieht, haben auch hier frühere Mitarbeiter des IDS einen nicht unwesentlichen Anteil.

Die systematische Erforschung der Lexik ist nach meiner festen Überzeugung die wichtigste Teildisziplin der Sprachwissenschaft. Wie schon bemerkt, ist dies eine eher untypische Ansicht; aber es ist nun einmal die meine, die man teilen kann oder nicht (und im Übrigen kann man ja auch weniger Wichtiges besonders interessant finden; es ist ja auch für die Physik nicht sonderlich wichtig, ob es das Higgs-Boson gibt oder nicht, und wenn man es sich überlegt, sind die meisten Fragen, mit denen sich die Wissenschaftler schwerpunktmäßig beschäftigen, nicht wirklich wichtig). Aber auch wenn man die Lexikografie zwar nicht für die wichtigste, sondern nur für eine unter vielen wichtigen Teildisziplinen hält, und das tun sicher viele, dann kann man nur mit einer gewissen Bestürzung feststellen, dass die Lexikografie zwar auf der einen Seite bedeutende Leistungen vollbracht hat – das Grimmsche Wörterbuch, das OED, der *Grand Robert*, aber auch die vielen monumentalen zweisprachigen Wörterbücher bezeugen es. Aber auf der anderen Seite bleiben auch diese Werke weit hinter der Aufgabe zurück; das habe ich in Abschnitt 2 deutlich zu machen versucht. Und drittens ist die wissenschaftliche Lexikografie, und zwar die kommerzielle wie auch die an Universitäten und Akademien betriebene, zu Beginn des dritten Jahrtausends in eine fundamentale Krise geraten. Das ist leicht zu sehen, ich will es zunächst für die kommerzielle, dann für die akademische Lexikografie an ein paar Punkten illustrieren:

A. Das Bibliographische Institut („Duden“) hat im Verlaufe der letzten zwei Jahre seine angesehene Wörterbuchabteilung weitestgehend aufgelöst und die bisher rund 180 Mitarbeiter bis auf einige wenige entlassen. Der Verlag Wissensmedia, Träger des „Wahrig“ und damit der zweiten bedeutenden Wörterbuchreihe der

Gegenwart in Deutschland, hat seine lexikografische Arbeit eingestellt. Die großen zweisprachigen Wörterbücher bei Langenscheidt und Klett werden nicht oder allenfalls über werbefinanzierte Webportale in geringem Umfang weitergeführt; die Verlagsarbeit konzentriert sich auf kleinere Wörterbücher für den Sprachunterricht oder für Reisen. Diese Entwicklung, die nicht auf Deutschland beschränkt ist, kann man nicht den Verlagen oder gar den Mitarbeitern anlasten. Sie hat eine einfache Ursache: Kaum noch jemand kauft solche umfassenden Wörterbücher auf hohem wissenschaftlichem Niveau (und nur von diesen soll hier die Rede sein). Wir sind schuld, nicht die Verlage. Dafür mögen zum einen soziale Entwicklungen verantwortlich sein, etwa das Schwinden des klassischen Bildungsbürgerturns. Entscheidend ist jedoch das Internet, das viele Wörterbücher umsonst zur Verfügung stellt. Umsonst sind sie, weil sie auf älteren rechtfreien Wörterbüchern beruhen oder von Freiwilligen ohne Entgelt bearbeitet werden. Es ist dies nicht anders als beim Niedergang der großen Enzyklopädien wie „Brockhaus“, „Mayer“ oder „Encyclopedia Britannica“, die durch Wikipedia hinweggefegt wurden. Über die Qualität der so erhältlichen Informationen kann man geteilter Meinung sein, sie schwankt auch ganz erheblich. Kein Zweifel kann aber darüber bestehen, dass diese Entwicklung das Ende der von Verlagen getragenen und somit auf Erträge angewiesenen klassischen Lexikografie bedeutet. Eine der wenigen Ausnahmen ist – noch – das OED mit seinen etwa 120 Mitarbeitern (und vielen freiwilligen Helfern), das zwar gleichfalls nicht mehr gedruckt wird, aber in der Nutzung über Internet kostenpflichtig ist und auch genutzt wird. Nicht zuletzt lebt es aber davon, dass es das Flaggschiff eines bedeutenden, in anderen Bereichen gut verdienenden Verlags ist. Man kann nur hoffen, dass dies noch lange so bleibt.

B. Die akademische Lexikografie – damit meine ich die an Universitäten oder Akademien betriebene – wird aus öffentlichen Mitteln oder durch Stiftungen finanziert, ist also nicht auf Einnahmen von den Nutzern angewiesen. (Sie ist eigentlich überhaupt nicht auf Nutzer angewiesen.) Aber diese Quelle versiegt. Es werden so gut wie keine langfristigen Wörterbuchprojekte mehr bewilligt, und auch der Fortbestand der jetzigen ist gefährdet. Die Gründe für diese Krise, wenn nicht für das bevorstehende Ende der langen Tradition akademischer Wörterbuchvorhaben sind nicht so offenkundig wie bei den kommerziellen Wörterbüchern. Einer ist sicher, dass es immer schwieriger wird, gut ausgebildete Lexikografen für diese Arbeit zu finden; es gibt so gut wie keine einschlägigen Studiengänge, ja, es gibt sogar kaum noch Lehrveranstaltungen dazu. Der wichtigste Grund ist aber sicher, dass sie so endlos lange dauern und niemand mehr bereit ist, sich auf ein Projekt einzulassen, das zu seinem Abschluss wahrscheinlich hundert Jahre braucht und sich bis dahin überholt hat. Es ist mit dem Abschluss dieser monu-

mentalenen Vorhaben wie mit dem Ende des Regenbogens: Wenn man ihm näher kommt, entfernt er sich. Man kann sich dies an den großen historischen Wörterbüchern des Deutschen vor Augen führen. Das „Althochdeutsche Wörterbuch“ wurde Mitte der Dreißigerjahre an der Sächsischen Akademie der Wissenschaften begonnen und hat bis jetzt etwa die Hälfte der Strecke erreicht; derzeit sind M–N in Bearbeitung. Das „Mittelhochdeutsche Wörterbuch“, das gemeinsam von der Mainzer Akademie für Sprache und Dichtung (Arbeitsstelle Trier) und der Göttinger Akademie der Wissenschaften (Arbeitsstelle Göttingen) getragen wird, wurde Anfang der Neunzigerjahre des letzten Jahrhunderts begonnen und ist auf vier Bände und etwa 30 Jahre angelegt. Es ist nun beim Buchstaben G angelangt; das ist schnell, aber das Ende ist im Augenblick nicht abzusehen. Das „Frühneuhochdeutsche Wörterbuch“, dessen erste Lieferung 1989 erschien und das seit 2013, nachdem es teilweise auch vom IDS finanziert wurde, gleichfalls in die Obhut der Göttinger Akademie der Wissenschaften gelangt ist, ist inzwischen, eine bemerkenswerte Leistung, bis zu *st-* gediehen; es soll 2027 abgeschlossen sein. Wir drücken die Daumen.

Am deutlichsten kann man sich das Problem an der Neubearbeitung des „Grimm“ vor Augen führen, von der vorhin schon die Rede war. Sie wurde 1962 in Berlin, etwas später in Göttingen begonnen; die Strecke A–C (Berliner Anteil) sollte 43 Lieferungen umfassen, der Göttinger Anteil (D–F) 37. Es gab allerlei Verzögerungen, und nach der Wende haben die beiden Akademien gemeinsam noch einmal eine Verlängerung bis Ende 2006 beschlossen (anders als heute konnten zu dieser Zeit die Akademien noch selbst über die Finanzierung beschließen). Da war Göttingen in der Tat fertig, Berlin – aus durchaus nachvollziehbaren Gründen – noch nicht. So wurde vereinbart, dass nun die verbliebene Strecke neu aufgeteilt wird: Berlin sollte bis Mitte B arbeiten, Göttingen den Rest von B und C übernehmen. Mit einiger Mühe war es möglich, eine weitere Finanzierung für die insgesamt etwa 15 hauptamtlichen Mitarbeiter bis Ende 2012 zu erhalten. In Berlin ist dieser Abschluss beinahe gelungen, allerdings um den Preis einer in der Stichwortauswahl deutlich beschränkten letzten Lieferung, für die ein Jahr lang eine anderweitige Finanzierung gefunden werden konnte: A – *Betreuung* ist seit Anfang 2014 fertig. Der Berliner Teil der Neubearbeitung ist somit nach 52 Jahren beendet, das eine Ende des Regenbogens ist erreicht, der Topf mit den Goldstücken, der nach alter Sage dort wartet, steht jedermann zur Verfügung. In Göttingen konnte dies – wiederum aus guten Gründen – nicht erreicht werden, der Abschluss ist nun auf Ende 2016 festgelegt, und die Aussichten stehen gut, dass damit auch das andere Ende des Regenbogens erreicht sein wird.

Das Langzeitproblem der lexikografischen Langzeitvorhaben ist weder für die deutsche Sprache noch für Deutschland spezifisch. Manche dieser Vorhaben laufen noch, oft unter der Bedingung massiver Straffungen und Kürzungen, man-

che stehen vor einem ungesicherten Fatum, neu begonnen wird keines. Es wäre falsch zu sagen, dass hierzulande keine Langzeitvorhaben mehr gefördert würden. Sie sind aber seit einigen Jahren in der maximalen Laufzeit beschränkt, im gemeinsamen Programm der Wissenschaftlichen Akademien beispielsweise auf 25 Jahre (in Ausnahmefällen – deren ich keinen kenne – auf 28 Jahre). Bei der Deutschen Forschungsgemeinschaft oder der Volkswagen-Stiftung ist die Laufzeit deutlich geringer.

Man kann an dem Gesagten vielleicht das eine oder andere korrigieren, aber das ändert nichts an dem Ergebnis: In Zukunft werden keine großen wissenschaftlichen Wörterbücher klassischen Stils mehr geschaffen werden. Selbst wenn man das Geld fände – und ich wüsste nicht, wie –, es dauert einfach zu lang. Die Fülle des Materials, das man inzwischen bearbeiten kann, stellt die Erfassung der gesamten Lexik zwar auf eine bessere Basis, verlängert die Laufzeit aber noch mehr, und durch die fortwährende Entwicklung der Sprache ist ein solches Wörterbuch nur von geringem Nutzen. Der erste Band des alten „Deutschen Wörterbuchs“ mit seinen insgesamt 32 Foliobänden umfasst die Wortstrecke *A – Biermolke*. Die Wortstrecke *A – Betreuung* füllt in der Neubearbeitung vier Foliobände. Welchen Sinn soll es haben, ein Wörterbuch von 128 Foliobänden in Angriff zu nehmen? All das kann sich ein jeder leicht ausrechnen, der es will. Als ich dies einmal einem sehr kenntnisreichen älteren Lexikografen vorgerechnet habe, hat er mir geschrieben: „Ich weigere mich, solche Rechnungen anzustellen.“ Das kann man verstehen, aber das heißt nur, die Augen davor zu schließen, dass es in Zukunft keine umfassenden Wörterbücher des Deutschen mehr geben wird.

5 2014 ff.

Man kann den Titel dieses Beitrags „Das Wörterbuch der Zukunft ist kein Wörterbuch“ in zwei Weisen auffassen. Die eine ist, dass aus den genannten Gründen die systematische Darstellung der deutschen Lexik an ein Ende gekommen ist: Es wird keine Wörterbücher mehr geben (womit hier immer umfassende wissenschaftliche Wörterbücher gemeint sind, nicht kleinere für Schüler, Reisende und dergleichen). Die andere ist: Man muss nach ganz anderen Wegen suchen, die Aufgabe, umfassende und gründliche Informationen über die deutsche Lexik zu liefern, zu lösen. Dazu genügt es nicht, auf den Computer und die unglaublichen Möglichkeiten digitaler Korpora und Arbeitsmittel zu verweisen. Die sind nach meiner Ansicht in der Tat unabdinglich, aber sie reichen nicht. Einen der beiden zentralen Gründe habe ich weiter oben kurz erläutert; die derzeitigen Methoden sind zwar hilfreich, aber nach wie vor sehr fehlerbehaftet,

und Besserung zeichnet sich zumindest im Augenblick nicht ab. Der andere, noch wichtigere Grund ist, dass bislang niemand einen Weg gefunden hat, aus den Wortformen auf die semantischen Eigenschaften der zugrunde liegenden lexikalischen Einheiten zu schließen. Das Korpus mag noch so gut sein, die Lemmatisierung und die vorgängige syntaktische Analyse weitaus fehlerfreier als bisher – kein Computer kann aus noch so vielen Vorkommen von *Körper*, *noch*, *grinsen* ableiten, was diese Wörter bedeuten. Man kann vielleicht angeben, mit welchen anderen Wörtern sie bevorzugt vorkommen, und das zu wissen ist sicher auch nützlich. Aber es ist etwas anderes als die Bedeutung der Wörter anzugeben.

In meinen Augen sind drei miteinander eng zusammenhängende Schritte notwendig, um die darbende Lexikografie auf eine neue Bahn zu setzen, in der sie, denke ich, ihre Aufgaben in weitaus besserer Weise erfüllen kann als bisher.

A. Die Eigenschaften der Lexik müssen in einem „digitalen lexikalischen System“ (DLS) erfasst und dargestellt werden. Damit ist nicht das digitale Version eines klassischen Wörterbuchs gemeint, sondern ein anderes Instrument, um die Lexik einer Sprache abzubilden. Anders ist es vor allem in dreierlei Hinsicht (siehe dazu im einzelnen Klein/Geyken 2010):

1. Es steht eben nicht auf Buchseiten, sondern im Computer und ist über Internet zugänglich. Allerdings lassen sich jederzeit „Papierwörterbücher“ daraus ableiten, falls man das für wünschenswert hält.
2. Datenbasis und lexikalische Analyse sind konstant miteinander verknüpft, und zwar so, dass jeder Nutzer leicht von den Daten in die Analyse und umgekehrt gehen kann. Die Datenbasis ist keine Belegsammlung, sondern eine Menge gut aufbereiteter Textkorpora, die mit geeigneten Suchwerkzeugen zu erschließen ist.
3. Charakteristisch für ein DLS sind weiterhin vier Eigenschaften, letztlich allesamt Abwandlungen ein und derselben Eigenschaft, nämlich einer hohen Flexibilität: Es sind dies *Modularität*, *inkrementelle Funktionalität*, *kumulative Entwicklung*, *Methodenpluralität*.

Das will ich hier noch ein wenig ausführen. Mit Modularität ist gemeint, dass das gesamte System aus einer Reihe von einzelnen Komponenten besteht, die miteinander verknüpft sind, sich aber im Prinzip unabhängig voneinander bearbeiten und nutzen lassen. So kann man eigene Module für die phonologischen, die morphologischen, die syntaktischen und die semantischen Eigenschaften vorsehen. Eine weitere Komponente kann der Etymologie oder der historischen Entwicklung gewidmet sein. Ebenso können Angaben über die Verwendungshäufigkeit,

gefiltert nach Zeiten, Textarten, Themenbereichen und dergleichen, zusammengetragen oder durch den Nutzer ermittelt werden. Schließlich kann man auch ermitteln, mit welchen anderen Wörtern das gerade untersuchte besonders häufig gemeinsam vorkommt, auch dies nach Bedarf gefiltert nach Zeit, Textart und Themenbereich. Dies sind nur einige der Möglichkeiten, weitere kann man sich leicht überlegen. Mit ‚inkrementelle Funktionalität‘ ist gemeint, dass die Bearbeitungstiefe nicht von Anfang an feststeht, sondern je nach Bedarf, Nutzerwünschen, wissenschaftlichen Interessen und nicht zuletzt vorhandenen finanziellen Mitteln variabel ist. Manche Stichwörter werden sehr detailliert analysiert, zu anderen gibt es nur Basisinformationen, ergänzt vielleicht um ausgesuchte Belege. Das Gegenstück dazu auf Bearbeiterseite ist die Möglichkeit der kumulativen Entwicklung. Das klassische Wörterbuch wird in der Regel von A–Z – oder auch einem sonstigen, zu Beginn festgelegten Prinzip – bearbeitet. Ein DLS hingegen ist beliebig erweiterbar und in gewisser Weise daher auch niemals abgeschlossen. So können die einzelnen Module zu verschiedenen Zeiten von verschiedenen Experten bearbeitet werden. Die Methodenplurazität schließlich bezieht sich darauf, dass man für gleichwelchen Bereich verschiedene Beschreibungsverfahren verbinden kann; die phonetischen Eigenschaften kann man in konventioneller Weise mithilfe einer Lautschrift umschreiben; aber man kann die entsprechenden Wörter auch einsprechen lassen, und dies in verschiedenen regionalen Abwandlungen. Für jemanden, der die Sprache lernen will, ist das sicher viel hilfreicher. Ebenso kann man für die semantischen Eigenschaften die traditionellen Paraphrasen nutzen, aber auch andere Möglichkeiten, wie sie beispielsweise in WordNet oder in FrameNet entwickelt wurden. Schließlich kann man auch Bilder oder Videos benutzen, mit denen sich – auch hier für den Lerner – die Bedeutung von *grinsen* vielleicht schlagender darstellen lässt als im „Großen Wörterbuch der deutschen Sprache“.

B. Das bereits vorhandene lexikalische Wissen muss soweit als irgend möglich ausgenutzt und integriert werden. Es gibt zahllose Wörterbücher von oft sehr hoher Qualität; es ist ja nicht ungewöhnlich, dass die neueste Variante ein und desselben Wörterbuchs zwar aktueller, aber in der Beschreibung der lexikalischen Eigenschaften weniger klar und richtig ist. Auf dieses akkumulierte Wissen früherer Gelehrter muss man sich systematisch stützen. Eine einfache Möglichkeit sind Wörterbuchportale, die es bereits für viele Sprachen gibt. Sie sind auch nützlich, aber im vorliegenden Zusammenhang ist daran gedacht, sie sinnvoll in das Gesamtsystem zu integrieren. Das ist keine triviale Aufgabe, aber auch hier macht sich die Flexibilität eines DLS nieder: Man kann sie Schritt für Schritt angehen und so das System stetig verbessern.

C. All dies reicht nicht, wenn man eine Million Wörter – ein Fünftel der derzeit in einem großen Korpus vorkommenden – erfassen will. Zu weitaus den meisten dazu gibt es bislang keinerlei Analyse, und sie in Angriff nehmen zu wollen, führt unweigerlich in das einleitend geschilderte Massenproblem. Ein hehres Ziel, aber so sinnlos wie die Absicht von Ron Obvious (in dem berühmten Sketch von Monty Python), von Dover über den Kanal nach Calais zu springen. Für die große Mehrzahl der Wörter gibt es daher nur einen Weg: Man muss den Benutzer zu seinem eigenen Lexikografen machen. Das heißt nicht, dass er von nun an Artikel schreiben sollen. Vielmehr soll man ihn in die Lage versetzen, sich die Eigenschaften, die ihn interessieren, aufgrund einer gut ausgewählten Anzahl von Belegen aus dem Korpus selber zu erschließen. Für einen Übersetzer, der zum ersten Mal das Wort *Eierstich* liest, ist es vielleicht ausreichend, wenn er sich ein paar geschickt ausgewählte Vorkommen aus der gleichen Zeit und der gleichen Textsorte anschauen kann. Dazu muss man ihm einige leicht bedienbare Werkzeuge an die Hand geben, mit denen er das flexibel machen kann. Man kann aber auch noch einen Schritt weitergehen und zum Beispiel manchen Wörtern immerhin ein ungefähres Synonym begeben. Dazu bedarf es keiner umfänglichen Analyse, der Lexikograf kann es in einer Minute, statt in einer Stunde, erledigen, und dem Nutzer ist zumindest ein wenig weiter geholfen. Das ist keine wissenschaftliche Beschreibung, es ist, wenn man so will, ein „Hilfsmodus“; aber es ist mehr als nichts, und in vielen Fällen alles, was ein Nutzer braucht. Und auch hier gilt, dass man flexibel und nach Bedarf vorgehen kann: Die Aufarbeitung kann jederzeit vertieft werden.

Diese drei Schritte hängen eng miteinander zusammen, jeder ist wichtig, keiner allein würde ausreichen. Sie führen auf einen Weg, der es uns, anders als das Wörterbuch, das wir kennen, in der Tat erstmals erlaubt, auf lange Sicht den Wortschatz einer Sprache in großem Umfang zu erfassen und sukzessive in seinen verschiedenen Eigenschaften zu analysieren. Man kann sich für den Kernbereich eines solchen digitalen lexikalischen Systems beispielsweise vorstellen, dass ein bestimmter Ausschnitt des Wortschatzes, vielleicht 10.000–20.000 Wörter, in größter Ausführlichkeit und historischer Tiefe behandelt wird, vielleicht 100.000–200.000 in der Art, die man vom „Großen Wörterbuch“ des Duden oder dem „Wahrig“ gewohnt ist, und eine Million oder zwei Millionen in dem eben genannten „Hilfsmodus“; in allen drei Fällen sind Daten und lexikalische Analyse – die freilich im letztgenannten Fall minimal ist – miteinander verknüpft. Dieser Kern wird von einer Reihe weiterer Module, wie sie weiter oben angedeutet wurden, umgeben. Ein solches System ist, da völlig flexibel, erscheint praktisch machbar, lässt sich jederzeit weiter ausbauen, und es liefert mehr lexikalische Informationen über eine Sprache, als ein jedes klassische Wörterbuch vermöchte. Das Wörterbuch der Zukunft ist kein Wörterbuch, es ist mehr.

Literatur

Dem Charakter dieses Beitrags entsprechend habe ich hier keine Literaturliste angefügt, sondern verweise nur auf zwei Veröffentlichungen, in denen sich Näheres zu dem hier Gesagten findet.

Klein, Wolfgang (2013): Von Reichtum und Armut des deutschen Wortschatzes. In: Deutsche Akademie für Sprache und Dichtung/Union der deutschen Akademien der Wissenschaften (Hg.): Reichtum und Armut der deutschen Sprache. Erster Bericht zur Lage der deutschen Sprache. Berlin/Boston, S. 15–55.

Klein, Wolfgang/Geyken, Alexander (2010): Das Digitale Wörterbuch der Deutschen Sprache (DWDS). In: Lexicographica 26, S. 79–96.