

Philologische Erfahrungen mit datenverarbeitenden Maschinen

Von Hans Eggers

Im Jahre 1956 habe ich erstmals begonnen, mich ernsthaft mit der Frage zu beschäftigen, ob und wieweit datenverarbeitende Maschinen für sprachwissenschaftliche Arbeiten in Anspruch genommen werden können, und im Herbst 1958 habe ich auf dem Germanistenkongreß in Hamburg im Rahmen eines Einführungsvortrages zum erstenmal die Arbeit mit einer Sortiermaschine, einem mechanischen Gerät, demonstrieren können. Ich kann damit wohl für mich in Anspruch nehmen, als einer der ersten Vertreter unseres Faches in Deutschland den Einsatz von Maschinen nicht nur erwogen, sondern auch praktisch erprobt zu haben. Seit mehr als zehn Jahren bin ich nun auf diesem Gebiet tätig, zwar nicht ständig, und durch die vielseitigen Pflichten akademischer Forschung und Lehre auf manches andere Interessengebiet geführt, aber doch intensiv genug, daß ich mir heute wohl erlauben darf, in diesem Kreise über „Philologische Erfahrungen mit datenverarbeitenden Maschinen“ zu berichten. Mir kommt dabei zugute, daß ich seit etwa vier Jahren durch großzügige Hilfe der Deutschen Forschungsgemeinschaft einen wissenschaftlichen Mitarbeiter und fünf studentische Hilfskräfte beschäftigen kann, an deren fortschreitenden Erfahrungen und Leistungen ich ständig Anteil nehme.

Davon, daß ich anfangs viel Lehrgeld zu zahlen hatte, will ich nicht berichten, auch nicht von den mancherlei überflüssigen Umwegen, auf die ich zuweilen geriet. Ich will aber keineswegs verschweigen, daß meine Mitarbeiter und ich auch heute noch lange nicht am Ziele stehen. Man braucht sehr viel Zeit, bis ein umfangreiches Vorhaben „maschinenreif“ programmiert ist, und mein Ziel, die automatische Analyse des syntaktischen Baus beliebiger deutscher Sätze zu errei-

chen oder jedenfalls die Grenzen der Maschinenarbeit auf diesem Gebiet abzustecken, ist vielleicht besonders weit gespannt und anspruchsvoll. Zwar sind meine Mitarbeiter und ich in den letzten Jahren auf diesem Weg ein gutes Stück vorangekommen, aber er muß Schritt für Schritt erkämpft werden, und auch heute noch beobachten wir immer wieder, daß jeder neue Schritt sorgfältig erwogen und vorbereitet werden muß. Oftmals glaubt man, alle Möglichkeiten in Betracht gezogen zu haben, arbeitet ein entsprechendes Programm aus und stellt dann in der praktischen Erprobung mit dem gegebenen Sprachmaterial doch eine unvorhergesehene Schwierigkeit fest, meist eine nicht hinlänglich beachtete Einzelheit der Sprachstruktur, der gegenüber das ausgearbeitete Programm versagt. Manchmal kann man dann die Wege zur Lösung noch in das Programm einbauen. Es ist aber ebenso möglich, daß man alles umstoßen und den ganzen Schritt noch einmal tun muß. Es entwickelt sich also – und das ist eine grundlegende Erfahrung – sozusagen ein Dialog zwischen dem Sprachforscher und der Maschine. Der Forscher meint, alles sprachlich Mögliche berücksichtigt und im Maschinenprogramm erfaßt zu haben, und die Maschine gehorcht seinen Anweisungen bis zu dem Punkt, wo ihr Versagen anzeigt: „Die an dieser Stelle vorliegende Sprachstruktur hast du noch nicht gemeistert.“ Neue geistige Arbeit hat dann die nun erst erkannte Schwierigkeit zu überwinden.

Ich sprach von „gegebenem Sprachmaterial“ und möchte mein Verfahren in aller Kürze erläutern. Wir arbeiten in Saarbrücken mit zwei Katalogen, zwei Corpora mit je etwa 5500 Einzelsätzen aus populärwissenschaftlicher und journalistischer Prosa, die nach bestimmten äußerlich-formalen Gesichtspunkten zusammengestellt sind. Und zwar haben wir aus einem Gesamtmaterial von zweimal 50000 Sätzen und von 100 Autoren – je 1000 Sätze von jedem Autor – alle Sätze von bestimmter Länge, gemessen nach der Anzahl der Wörter, zu diesen beiden Katalogen zusammengestellt. Ausnahmslosigkeit wurde dabei angestrebt; kein Satz wurde beiseite geschoben, weil er unbequem oder grammatisch undurchsichtig war. Wir glauben daher, in den 11000 Sätzen unserer Kataloge über ein einigermaßen repräsentatives Material aus einer mittleren schriftsprachlichen Schicht zu verfügen, in dem alle irgend wichtigen syntaktischen Erscheinungen der deutschen Gegenwartssprache vorkommen müssen. Dabei erlaubt es unsere Methode auch, die Häufigkeit des Vorkommens jeder Einzelerscheinung statistisch festzustellen, was für die maschinelle Bear-

beitung und die Reihenfolge der dafür notwendigen Untersuchungsgänge wichtig ist.

Dieses Verfahren muß notwendigerweise strukturalistisch sein; denn es sind in erster Linie formale Kriterien der Wort- und Satzgliedfolge einschließlich der Flexionsregeln, durch deren Anwendung die automatische syntaktische Analyse ermöglicht wird. Dabei ist unsere Methode ausgesprochen induktiv. Sie geht von der Analyse der Einzelercheinungen aus und versucht, auf dieser Grundlage zur Synthese immer größerer Einheiten zu gelangen. Dieses Verfahren steht in diametralem Gegensatz zu jener anderen Methode der sogenannten „generativen Grammatik“, mit der sich heute die führenden Strukturalisten, vor allem auch die mathematischen Linguisten, vornehmlich befassen. Dieser Richtung geht es darum, sprachrichtige Sätze nach erkannten und festgelegten Regeln zu „erzeugen“. Demgegenüber sucht unser Verfahren schon vorhandene Sätze zu analysieren. Ich halte es für völlig unnötig, über die eine oder die andere Methode zu streiten, was oft genug geschieht. Denkt man beispielsweise an das Problem der automatischen Übersetzung, so müssen freilich in der Zielsprache Sätze erzeugt werden, und hier ist die generative Grammatik ganz und gar am Platze. Bei der Ursprungssprache dagegen handelt es sich um ein bereits fertig vorliegendes Sprachmaterial, das analysiert werden muß, bevor man es der Übersetzung zuführen kann. Und hier möchte ich auf Grund meines jahrelangen Umgangs mit aktuellem Sprachmaterial, der mich vor täglich neue überraschende Einsichten stellt, behaupten: Die deutsche Sprache und jede andere Sprache verfügt über so viele syntaktische Möglichkeiten, daß man ihrer in einem theoretischen Deduktionsverfahren niemals zur Gänze habhaft wird. Hier kann nur die Fülle der Einzelbeobachtungen die sichere Erkenntnisgrundlage liefern.

Aber wenden wir uns konkreten Erörterungen zu, und dem Erfahrungsbericht, den ich Ihnen mit meiner Themenwahl versprochen habe. Es sollte zunächst erwähnt werden, daß es außer dem Elektronenrechner auch andere datenverarbeitende Maschinen gibt. Im Vorhof der Elektronik steht ein elektromechanisches Gerät, die Sortiermaschine, die Lochkarten verarbeitet – etwa 600–700 Stück in der Minute – und nach allen gewünschten Gesichtspunkten ordnet. Das zugehörige Tabelliergerät ermöglicht es, auf Grund ziemlich einfacher Programmierung die durch den Sortiervorgang gewonnenen Ergebnisse in Listen auszudrucken und unmittelbar benutzbar zu

machen. Die Lochkartensammlung (Lochkartei) kann daher den bei uns Philologen so beliebten Zettelkasten mit großem Vorteil ersetzen. Man denke etwa an die ungeheure Beschleunigung aller rein mechanischen Ordnungsarbeiten in einem Wörterbuch-Archiv, das mit Hunderttausenden und Millionen von Belegen zu arbeiten hat. Noch rascher und noch differenzierter läßt sich allerdings auch ein solcher Sortiervorgang mit Hilfe des Elektronenrechners bewerkstelligen. Denn hier lassen sich die einzelnen Phasen des Sortierens, die auf dem mechanischen Sortiergerät nacheinander ablaufen müssen, kombinieren. Die Zwischenergebnisse können gespeichert werden, und das Ausdrucken kann genau in der differenzierten Form erfolgen, die der Bearbeiter für notwendig hält. Damit wären wir beim Elektronenrechner angelangt und ich möchte Ihnen nun an drei ausgewählten Beispielen von Erfahrungen berichten, die ich als Philologe mit den modernen Automaten gemacht habe.

1. Wo es um reine Materialsammlung und -ordnung geht, kann uns der Rechenautomat die Arbeit weitgehend abnehmen. Anstatt z. B. Zettel handschriftlich oder mit der Schreibmaschine anzulegen, kann man auch Lochkarten oder Lochstreifen benützen, wozu man sich einer Schreibmaschinentastatur mit einigen wenigen Sondereinrichtungen bedient. Wer maschineschreiben kann, erlernt dieses Verfahren in kürzester Zeit. Die so entstehenden Informationsträger können dann nach entsprechender Programmierung in jeder gewünschten Weise verarbeitet werden.

In günstig gelagerten Fällen kann man sich sogar die vorbereitende Stufe der maschinellen Herstellung von Informationsträgern ersparen. Als ich seinerzeit daranging, mir mein Untersuchungsmaterial zu schaffen, habe ich in mühsamer Arbeit mit freiwilligen Helfern von 100 Autoren je 1000 Sätze abgezählt und für jeden einzelnen Satz die Anzahl der darin vereinigten Wörter ermittelt. Die Ergebnisse wurden dann in Strichlisten eingetragen. Wir stellten fest, daß der Satz mit 16 Wörtern häufiger als jede andere Satzlänge vorkam, und vereinigten dann die sämtlichen Sätze dieser Länge und zum Vergleich sämtliche Sätze verschiedener anderer Längen in Schreibmaschinenarbeit zu den beiden vorerwähnten Katalogen. Diese mechanischen Arbeiten nahmen qualifizierte Kräfte, die dadurch wertvollerer Arbeit entzogen wurden, viele Monate lang in Anspruch. Dabei waren, wie wiederholte Kontrollen bewiesen haben, die Ergebnisse keineswegs

fehlerfrei, was bei so ausgedehnter mechanischer Arbeit gewiß nicht verwunderlich ist.

Inzwischen haben wir ein Verfahren gefunden – was uns damals noch nicht möglich war –, die 6-Kanal-Lochstreifen der automatischen Setzmaschinen der Druckereien in 5-Kanal-Lochstreifen umzukodieren, so daß sie durch Elektronenrechner bearbeitet werden können. Heute könnten wir daher mit leichter Mühe ein elektronisches Programm entwerfen, das die Texte vollständiger Bücher entsprechend aufbereiten würde. Es könnte in einem Arbeitsgang die Gesamtzahl der Sätze bestimmen oder auch 1000 Sätze abzählen, könnte die Anzahl der in jedem Satz enthaltenen Wörter ermitteln und die Ergebnisse in geordneten Tabellen ausgeben. Es könnte außerdem noch zugleich alle Sätze von bestimmter Länge ausdrucken, wobei diese automatisch sowohl eine laufende Katalognummer wie auch die Angabe der Belegstelle erhalten könnten. Ein Unternehmen, das seinerzeit eine Anzahl von Menschen monatelang zu rein mechanischer Tätigkeit zwang, ließe sich also heute in wenigen Stunden automatisch erledigen. Dabei würden „Stunden“ nur für die Ausgabe des Katalogs, für das Ausdrucken, erforderlich sein. Die eigentliche elektronische Verarbeitung würde viel schneller ablaufen. Da rechnet man heute nicht mehr mit Tausendstelsekunden, sondern mit Mikrosekunden, d. h. mit unvorstellbaren Geschwindigkeiten, bei denen sich die einzelnen elektronischen Vorgänge in Millionstelsekunden abspielen. Rein mechanische Tätigkeiten, die dem qualifizierten Bearbeiter nur zu vorbereitender Materialsammlung und -ordnung dienen und die ihn seiner eigentlichen Aufgabe entziehen, vertraut man daher besser der Maschine an. Auf diesem Gebiet arbeitet sie um ein Vielfaches rascher – und fehlerfreier – als der bald ermüdete Mensch.

2. Es stellte sich im Laufe unserer syntaktischen Arbeiten bald heraus, daß wir mit Abstraktbegriffen wie Satzglied und Syntagma, Wortart und Formklasse allein nicht arbeiten konnten. Ein Verzeichnis der vorkommenden Wörter und ihrer Häufigkeit erwies sich als dringend notwendig. Die beiden Kataloge enthalten jeder rund 100 000 Wörter, und in meiner Ungeduld, rasch zu Ergebnissen zu gelangen, habe ich anfangs, ehe noch an automatische Bewältigung des Problems zu denken war, den einen der beiden Kataloge von Teilnehmern einer Seminarübung verzetteln lassen. Dutzende von Karteikästen füllten sich mit Zetteln, die dann mühsam von Hand sortiert werden

mußten. Das erwünschte Wörterbuch ist dabei niemals über die ersten vier, fünf Buchstaben des Alphabets hinausgediehen, weil die Arbeitskräfte fehlten; und was fertig wurde, erweist sich jetzt als reichlich fehlerhaft.

Inzwischen nämlich ist es uns nach langen Vorbereitungen gelungen, das Wörterbuch maschinell zu erstellen. Es ist freilich kein Wörterbuch im herkömmlichen Sinne geworden, sondern nur ein Index aller vorkommenden Wortformen, aber das reicht für unsere Zwecke auch vollkommen aus. Jetzt besteht dieses Opus aus je zwei Teilen für beide Kataloge: je einem alphabetischen Register mit Häufigkeitszählung und vollständigen Stellenangaben und je einem nach der Häufigkeit geordneten Index für beide Kataloge. In die beiden alphabetischen Register konnten wir infolge entsprechender Vorarbeiten auch gleich die Unterscheidung von Homographen einarbeiten, so daß etwa der Artikel *arbeiten* automatisch untergliedert wurde in *arbeiten* Subst. plur., *arbeiten* Infinitiv und *arbeiten* Verbum finitum, wobei den einzelnen Teilen die Häufigkeits- und die Stellenangaben hinzugefügt wurden. Zum Schluß wird dann die Gesamthäufigkeit des Homographen registriert.

Nachdem alle Vorbereitungen getroffen waren, dauerte die automatische Verarbeitung und Ausgabe der 200000 Wörter zu zwei Doppelregistern nur etwa je sechs Stunden, wobei wiederum die eigentliche Rechenzeit wesentlich kürzer war; die längste Zeit nahm das Ausdrucken der Stichwörter (rund 40000) mit ihren 200000 Belegstellen in Anspruch.

Man sollte nicht übersehen, was alles hierbei zur „automatischen Verarbeitung“ gehört. Wir hatten zuvor die Sätze unserer beiden Kataloge, jeden mit seiner Katalognummer und jedes Wort mit den erforderlichen grammatischen Angaben versehen, von Lochkarten auf ein Magnetband übertragen. Nach einem mit Unterstützung des Deutschen Rechenzentrums in Darmstadt ausgearbeiteten Programm mußte nun der Elektronenrechner die 11000 Sätze Wort für Wort vom Magnetband „ablesen“, jedes Wort mit Angabe der Satznummer in alphabetisch richtiger Folge „speichern“, mußte dabei noch die einzelnen Stichwörter, wie eben am Beispiel *arbeiten* gezeigt, untergliedern und schließlich die Summe der Belege in den Untergliederungen und die Gesamtbelegzahl jedes Stichwortes feststellen. Ein äußerst umfangreiches Programm also, das trotzdem in wenigen Stunden fehlerfrei bewältigt wurde. Es ist kaum zu berechnen, wie-

viel angespannte geistige und manuelle Arbeit und wieviel Zeit ein für philologische Arbeit qualifizierter Mensch hätte aufwenden müssen, um die gleiche Arbeit zu leisten, und wieviel Fehler ihm dabei unterlaufen wären.

Jetzt steht das Programm und ist mit geringfügigen Adaptionen für die Herstellung beliebiger Wortregister anwendbar. Es bedurfte aber langwieriger Vorbereitungen. Ich will gar nicht die vorausgehende Herstellung von Lochkarten in Anschlag bringen, die uns auch für alle anderen Arbeitsvorhaben notwendig sind. Auch daß meine Mitarbeiter zunächst die recht schwierige und größte Sorgfalt erfordernde Technik des elektronischen Programmierens zu erlernen hatten, sei nicht in Rechnung gestellt; denn das sind notwendige Voraussetzungen für jegliche Arbeit mit Rechenautomaten. Aber auch nach Erfüllung aller Vorbedingungen waren meine beiden tüchtigsten und selbständigsten Mitarbeiter fast ein halbes Jahr lang damit beschäftigt, die Programme für das alphabetische und das Häufigkeitsregister zu entwerfen. Sie durften sich dabei der tatkräftigen Unterstützung unseres örtlichen Saarbrücker Rechenzentrums wie auch des Deutschen Rechenzentrums in Darmstadt erfreuen. Das sei auch an dieser Stelle dankbar hervorgehoben. Besonders sei dabei Herrn Stickel in Darmstadt Dank und Anerkennung ausgesprochen. Seine Erfahrungen befähigten ihn sofort, unsere Absichten hinsichtlich des alphabetischen Registers zu verstehen und unser noch unvollkommenes Programm zu korrigieren. Für den Häufigkeitsindex stellte er uns sogar ein von ihm bereits ausgearbeitetes Programm zur Verfügung, so daß wir auf einen Entwurf verzichten konnten. Aber das alphabetische Programm hat uns jedenfalls monatelang beschäftigt.

Es ist freilich sicher, daß die Kräfte sich an wachsenden Aufgaben entfalten und üben, und ein neues Projekt gleichen Umfangs würde gewiß nicht mehr ebensoviel Zeit der Vorbereitung erfordern. Aber Monate würden doch vergehen müssen, bis ein größeres Programm entwickelt und erprobt ist und für ausführungsfähig erklärt werden kann. Man hat sich also als Philologe recht genau zu überlegen, ob der Umfang der gestellten Aufgabe den Aufwand der elektronischen Bearbeitung lohnt.

3. Von ganz anderer Art sind die Probleme, die bei dem dritten hier zu erörternden Arbeitsvorhaben auftreten, bei den vorbereitenden Arbeiten zur automatischen syntaktischen Analyse deutscher Sätze. Hier kommt es einerseits darauf an, z. B. Fügungen wie *die treuen*,

in vielen schwierigen Lagen erprobten Freunde nicht nur als syntaktische Einheit, sondern auch noch nach ihrer Gliedfunktion als Subjekt, Prädikativ, Objekt usw. zu bestimmen. Die dafür erforderlichen Programmierungsarbeiten sind umfangreich und zeitraubend. Hier kommt aber hinzu, daß intensive philologische Vorarbeit zu leisten ist. Es gibt ja noch keine Bestandsaufnahme all der vielen syntaktischen Möglichkeiten unserer Sprache. Ihre Untersuchung muß vorausgehen, bevor man überhaupt nur an Programmierung denken kann. Ich greife einige Beispiele heraus: Wir halten es für unerlässlich, vor Beginn einer auf Gliederung zielenden Analyse festzustellen, ob der zu untersuchende Satz ein einheitlicher einfacher Satz ist, ob mehrere Einfachsätze aneinandergereiht sind oder ob es sich um ein Gefüge aus Haupt- und Nebensätzen handelt. Die als bekannt geltenden, aber keineswegs wirklich erforschten Wortstellungsregeln reichen für eine solche Unterscheidung nicht aus. Man muß vielmehr auch den diakritischen Wert der Zeichensetzung dafür in Betracht ziehen. So wird in der Regel ein Komma vor *der, die, das* als Anhaltspunkt dafür gelten können, daß es sich um ein Relativpronomen und mithin um den Beginn eines Relativnebensatzes handelt, dessen Ende bei dem nächsten auf die finite Verbform folgenden Komma zu erwarten ist. Nun lassen sich die sämtlichen Kommata mit ihrer Umgebung sehr leicht und rasch automatisch auffinden und in Listen zusammenstellen. Bei der Bearbeitung der Listen zeigt sich dann aber, daß lange nicht jedes *der* hinter Komma ein Relativpronomen ist. Es kann sich z. B. auch um den Artikel am Anfang einer Apposition handeln (*Adenauer, der Altbundeskanzler, hat gestern erklärt...*). Hier hat man also zu differenzieren und hat nach formalen Kriterien zu suchen, die eine automatische sichere Unterscheidung ermöglichen.

Oder ein anderes Beispiel: Das vorhin schon erwähnte Wort *arbeiten* kann Plural eines Substantivs, Infinitiv oder Verbum finitum mit mehreren syntaktischen Funktionen sein. Wir werden uns dessen meist gar nicht bewußt, daß es Hunderte von semantischen und viele Tausende von syntaktischen Homographen gibt. Die Semantik macht (was zunächst verwunderlich erscheint) nicht allzu viele Schwierigkeiten, solange es um syntaktische Analyse geht; sie wird erst bei der Übersetzung relevant. Desto lästiger sind die syntaktischen Homographen; denn hier ist es natürlich von entscheidender syntaktischer Bedeutung, ob man es mit der 1. oder 3. Person Plural oder mit dem Infinitiv zu tun hat. Diese sehr vielfältigen Möglichkeiten syntak-

tischer Homographien möglichst vollständig zu erkennen, zu gruppieren und gruppenweise zu registrieren, verlangt intensive philologische Vorarbeit. Dann erst kann man nach formalen Kriterien für die Unterscheidung suchen, und nur wenn diese sich finden, kann man überhaupt an programmierende Bewältigung der Aufgabe denken. Dabei ist es besonders verzwickelt, daß sich in sehr vielen Sätzen mehrere Homographen finden. In dem kleinen Sätzchen *das kann sein* kommen z. B. mit den drei Wörtern auch drei Homographen zusammen. *Das* ist entweder Artikel oder Demonstrativ- oder Relativpronomen. *Kann* bezeichnet sowohl die 1. wie die 3. Person, und *sein* kann als Infinitiv, Substantiv oder Possessivpronomen aufgelöst werden. Erst aus dem Zusammenspiel der drei Wörter wird die Geltung und syntaktische Funktion jedes einzelnen eindeutig bestimmt. Die Erfahrung hat uns gelehrt, daß manche Homographen leichter, andere schwieriger mittels formaler Kriterien zu bestimmen sind. Wir sind daher auch auf diesem Gebiet dabei, eine Rangliste zu entwerfen und festzulegen, welche Homographen mit Vorrang aufzulösen sind.

Es erübrigt sich, weitere Beispiele zu häufen. Ich habe eingangs schon angedeutet, daß wir von dem Endziel, der vollständig automatisch durchgeführten Analyse, noch weit entfernt sind. Aber wir machen ständig Fortschritte, und der Zwang zu scharfer Beobachtung ist nicht nur eine reizvolle wissenschaftliche Tätigkeit. Er führt auch zu immer neuen Einsichten, die, ganz abgesehen von dem eigentlichen Arbeitsziel, auch durchaus ihren philologischen Eigenwert haben. Ein Beispiel dafür ist etwa die Arbeit meines Mitarbeiters Rainer Rath „Trennbare Verben und Ausklammerung“ (Wirkendes Wort 15, 1965, S. 217 ff.), deren Ergebnisse in die Neuauflage der Duden-Grammatik Aufnahme fanden (S. 636–638, § 7040–7060).

Die drei Beispiele, die ich Ihnen hier vorgetragen habe, mögen für drei Typen von Arbeitsmöglichkeiten mit elektrotechnischen Maschinen stehen. In dem ersten Fall, bei der automatischen Aufnahme und Ordnung von Daten (Satz- und Wortzählung), kommt es im wesentlichen nur auf die statistischen Werte an. Hier ist der Rechenautomat ein idealer Helfer, der in kürzester Zeit ungeheure Mengen von Material verarbeiten kann, das dann unmittelbar für die Auswertung zur Verfügung steht. Dabei handelt es sich allerdings nicht um philologische Arbeiten im engeren Sinne, sondern höchstens um die Vorbereitungen dazu.

Ausgesprochen philologischen Wert haben dagegen die lexikalischen und die syntaktischen Arbeiten. Um mit den letzteren zu beginnen, so liegt hier das Schwergewicht auf der Vorbereitung, und zwar mehr noch auf der philologischen Bearbeitung als auf der Programmierung. Wenn es freilich gelingt, ein vollständiges Programm für die automatische Analyse zu entwickeln, so steht es ein für allemal und ist immer wieder verwendbar. Damit wäre dann die entscheidende Grundlage für die automatische Übersetzung aus dem Deutschen in andere Sprachen geschaffen. Es lohnt sich also schon, auf dieses Ziel hin von langer Hand zu arbeiten. Dabei mag es hinsichtlich der automatischen Technik als ein Nebenprodukt erscheinen, wenn dabei neue philologische Erkenntnisse herauspringen. Die Sprachwissenschaft hingegen wird gerade hier den Hauptvorteil sehen.

Zur Bewältigung dieses Programms sind allerdings noch sehr vielseitige Einzelfragen zu klären. Um einschlägige Aufgaben nur anzudeuten, so laufen an meinem Institut zur Zeit, jeweils von einzelnen bearbeitet, u. a. folgende Untersuchungen nebeneinander her:

Zeichensetzung

Aufbau nominaler Satzglieder

Syntaktische Funktionen der Adjektiva

Attributiv gebrauchte Partizipia

Zusammengesetzte Verbformen und Modalerweiterungen

Ausklammerung.

Die Liste ist nicht vollständig. Viele detaillierte Untersuchungen (Funktion des *und*, Nebensatzeinleitende Konjunktionen u. ä.) werden außerdem stetig gefördert. Das sind zeitraubende Arbeiten, die große Sorgfalt verlangen, und damit stellt sich das Personalproblem. Je mehr geschulte Mitarbeiter zu solchen vorbereitenden Arbeiten angesetzt werden können, desto rascher wird man dem Ziele näher kommen. Da ich selbst zur Zeit nur einen hauptamtlichen Mitarbeiter und wenige studentische Hilfskräfte beschäftigen kann, machen meine Arbeiten bei weitem nicht die raschen Fortschritte, die im Interesse der Sache wünschenswert wären.

In ganz anderer Weise macht sich die Personalfrage bei unserem zweiten Beispiel geltend. Wie gesagt, nachdem das Programm einmal steht, könnten wir in kürzester Zeit eine beliebige Fülle von Material bereitstellen. Aber bleiben wir der Einfachheit halber bei unseren beiden vorhandenen Katalogen. Daß die Journalisten andere Substantiva verwenden als die Gelehrten, ergibt sich aus der Natur der

von ihnen behandelten Themata. Warum aber verwenden die Gelehrten mehr Demonstrativ- und Possessivpronomina, warum die Journalisten beim Verbum mehr das Präsens, die Gelehrten eher das Präteritum? Warum stehen die Journalisten im Gebrauch des Konjunktivs den Gelehrten weit voran, und warum kommen bei ihnen viel mehr Pluralformen vor? Das alles sind merkwürdige Erscheinungen, die die Häufigkeitsstatistik enthüllt. Ihnen allen, auf die man sonst gar nicht aufmerksam würde, kommt man durch die Statistik auf die Spur. Aber nun verlangen sie philologische Auswertung und versprechen dann viele neue Einsichten in unseren Sprachgebrauch. Wichtige Aufschlüsse über den Sprachstil unserer Zeit, über fachgebundene Sprachschichten, auch über den Personalstil könnten sich daraus ergeben. Aber wo sind die Bearbeiter, die alles untersuchen? Jeder neue Häufigkeitskatalog – und das ist ja nur ein Beispiel – würde neue Probleme aufwerfen. Und dies ist nun ein Faktum: Die datenverarbeitenden Maschinen werfen uns in Sekundenschnelle ein Material auf den Tisch, dessen wissenschaftliche Aufbereitung einen Bearbeiter jahrelang beschäftigen kann. Ich sage ohne Übertreibung: Die eine Maschine kann hundert Menschen beschäftigen, sehr viele zur Vorbereitung, vor allem der philologischen Vorbereitung weittragender Programme, noch viel mehr aber zur Aufbereitung des rasch und überreichlich gelieferten Materials. Solange wir in Deutschland, hier in diesem Mannheimer Institut, an meiner eigenen Forschungsstätte in Saarbrücken und an einer Reihe von anderen Arbeitsstellen immer nur mit einer Handvoll Mitarbeiter wirken können, darf von intensiver Ausnutzung der durch die Maschinen gebotenen Hilfsmittel noch nicht die Rede sein. Allerdings kann an diesen kleinen Forschungsstätten das erforderliche Personal besonders intensiv ausgebildet werden, und das wird nützlich sein. Wir müssen uns aber auf einen künftigen, sehr großen Personalbedarf einstellen, und wir sollten uns rechtzeitig darauf vorbereiten. Auch dies gehört zu den Erfahrungen eines Philologen, die vorgebracht und sehr ernst genommen werden müssen.

In lockerem Zusammenhang mit diesen allgemeinen Ausführungen möchte ich noch auf eine spezielle, die am deutschen Mittelalter Interessierten vor allem angehende Einzelfrage zu sprechen kommen. Seit mehreren Jahren arbeitet Dr. Roy Wisbey in Cambridge energisch und erfolgreich an Fragen der automatischen Herstellung mittelhochdeutscher Wörterbücher, oder sagen wir auch hier lieber vorsichtig,

mittelhochdeutscher Formenindizes. Er hat über die Fortschritte seiner Arbeiten an verschiedenen Stellen berichtet. Nunmehr kann er ein vollständiges, alphabetisches Formenregister zur Wiener Genesis vorlegen, und ich wünsche aufrichtig, daß die finanziellen Schwierigkeiten, die sich einer Veröffentlichung in Buchform entgegenstellen, überwunden werden können. Ich kenne sein Manuskript und habe im Interesse der Sache den lebhaften und ehrlichen Wunsch, daß es durch Veröffentlichung der Kritik der Fachgenossen zugänglich werden möge. Was also steht hier zu erwarten? Die einzelnen Wortformen werden in alphabetischer Folge verzeichnet. Während aber die bekannten amerikanischen Wortindizes, die im manuellen Verfahren hergestellt wurden, sich ohne Rücksicht auf Homographien mit Aneinanderreihung der Belegstellenangaben begnügen, gibt Wisbey zu jedem Beleg auch den Kontext im Umfang eines Verses an, was bei der oftmals stichischen Syntax der Wiener Genesis in den meisten Fällen wenigstens erste Einblicke in Bedeutung und Gebrauch der jeweiligen Form gewährt. Homographen trennt auch er nicht; aber der Benutzer erkennt sie jedenfalls aus dem mitangeführten Kontext. Bei reichlich 6000 Versen ergibt sich, durchschnittlich 5 Wörter auf den Vers gerechnet, ein Formenwörterbuch, das außer den Zeilen für die Lemmata 30000 Zeilen Belegstellen enthält. Der ganze Band enthält demnach reichlich 850 Seiten und wird trotz photographischer Herstellung des Satzes aus den vom Elektronenrechner ausgeworfenen Listen entsprechend teuer. Das Verfahren ist nun ausgearbeitet und läßt sich auf jeden beliebigen mittelhochdeutschen Text anwenden. Organisatorisch ließe sich durchaus ein Plan ausarbeiten, nach dem im Verlauf einiger Jahre die gesamte mittelhochdeutsche Literatur in entsprechender Weise indiziert werden könnte. Viele von uns, auch ich selber, würden das sogar für sehr wünschenswert halten. Aber wäre es gerechtfertigt, 4000 Seiten Hartmann von Aue, 8000–10000 Seiten Wolfram von Eschenbach usw. usw. zu drucken? Hier kann man vor den neuen Möglichkeiten einer Massenproduktion erschrecken. Archivierung (am besten nicht einmal ausgedruckt, sondern auf Magnetbändern) wäre hier wohl angebracht, vorausgesetzt, daß das Archivmaterial jedem Interessierten zugänglich gemacht wird, – und daß im Archiv eine hinlänglich große Gruppe von Fachgelehrten an der wissenschaftlichen Aufbereitung des ungeheuren Stoffes ständig arbeitet.

Denn auch hier erhebt sich die Frage nach dem rechten Verhältnis zwischen Material und bearbeitendem Personal. Roy Wisbey, wirklich ein Pionier auf diesem Gebiet, sagt in seinem Vorwort entschuldigend, er könne nicht mehr als einen voll belegten Formenindex vorlegen. Daraus ein wirkliches Wörterbuch herzustellen, dazu habe ihm und seinen Mitarbeitern die Zeit gefehlt. Das ist dieselbe Entschuldigung, die auch ich vorbringe, wenn ich aus meinen 200000 Wörtern nur ein Formenregister und kein Wörterbuch gemacht habe. Ich will die Frage offenlassen, ob sich überhaupt ein Wörterbuch auf vollautomatischem Wege herstellen läßt. Ein mehr oder minder großer Rest wird sich wohl immer nur in geistiger Arbeit bewältigen lassen, und legt man Wert auf eine irgendwie geartete intelligente Ordnung der Belege innerhalb der einzelnen Lemmata, dann muß der Geist und nicht die Maschine jeden einzelnen Artikel bearbeiten.

Noch einmal also dasselbe Problem: Die Maschine arbeitet rasch und fehlerfrei. Wo es nur auf Datenverarbeitung und mechanisches Erledigen ankommt, ist sie ein Helfer von unschätzbarem Wert. So auch für den Mathematiker und für den Techniker, für den sie die schwierigsten Berechnungen unbegreiflich schnell und fehlerfrei ausführt. Wir Philologen hingegen können oftmals von der Maschine nur die Bereitstellung des Materials verlangen, dessen geistige Durchdringung uns niemand abnimmt. Aber dieses Material wird in solcher Beschleunigung und Menge geliefert, daß der einzelne es nicht mehr bewältigen kann. Der Einsatz von Maschinen erfordert Gruppenarbeit, und ständiger Einsatz der Maschinen erfordert sehr große Gruppen von Bearbeitern.