

DIE VERWENDUNG VON DATENVERARBEITENDEN MASCHINEN BEI DER UNTERSUCHUNG VON LINEAR-B-INSCHRIFTEN¹

Der Einsatz von datenverarbeitenden Systemen gilt in den letzten Jahren als überaus modern; auch wird er als wirtschaftlich sehr vorteilhaft bezeichnet. Also sollten – so könnte man schlussfolgern – solche Systeme auch bei der Bearbeitung der Linear-B-Inschriften eingesetzt werden. Nun ist allerdings der Begriff der Modernität kein wissenschaftliches Kriterium, ebensowenig wie Rationalität der Arbeitsprozesse; beide sind nicht qualitativ, sondern allenfalls quantitative Kategorien einer geistigen Disziplin. Wir wollen daher hier begründen, warum wir datenverarbeitende Maschinen für die Untersuchung der Linear-B-Tafeln verwenden. Diese Erwägungen sind für uns dringlich, da wir mit diesen Verfahren schon 1958 begannen, also noch bevor solche Methoden modern wurden, und weil ausserdem gerade in dem ersten Stadium, also bis etwa 1962, kein nennenswerter ökonomischer Effekt dabei zu erzielen war. Vielmehr richteten sich unsere Absichten fast ausschliesslich auf qualitativen Gewinn, wenngleich dieser bei einzelnen Arbeitsschritten nur durch die Subsumierung grösserer Materialmengen zu erzielen ist. Auch haben wir bei Beginn der maschinellen Arbeiten weder Konsequenzen noch Zielsetzung so klar gesehen, wie sie uns heute, nach fast 10 Jahren theoretischer Überlegungen und etwa 6 Jahren praktischer Erfahrungen, erscheinen. Wir wollen zuerst eine begriffliche Klärung des heute so weit verbreiteten Wortes "Datenverarbeitung" versuchen. Zunächst ist vor auszuschicken, dass wir dieses

* Vortrag, gehalten auf dem Mykenologischen Kolloquium der Eirene-Konferenz, Berlin, 16. 10. 1967.

Gebiet in unserem internen Gebrauch nicht als Daten-, sondern als Informations- Verarbeitung bezeichnen. Denn tatsächlich werden ja nicht Daten – also Zahlenwerte – berechnet, sondern es werden Sachverhalte sprachlicher Natur – als Informationen verschlüsselt – von uns systematisiert. Soweit dabei Zählwerke in den Maschinen mitlaufen, dienen diese, wie immer bei natürlichen Sprachen, nicht der Berechnung im mathematischen Sinne, sondern der einfachen Zählung von Proportionen, die man lediglich als sprachstatistische Angaben bezeichnen sollte. Da sich jedoch der Begriff “Datenverarbeitung” international durchgesetzt hat, und zwar wegen seiner Bedeutung bei mathematischen und betriebswirtschaftlichen Aufgaben, wollen wir ihn auch hier gelten lassen; korrekter wäre es jedoch, mit H. Spitzbardt die Bezeichnung “Sprachdatenverarbeitung” zu verwenden.

Für die Datenverarbeitung sind – wenn wir uns auf die hier interessierenden Anwendungsbereiche beschränken – 4 Kriterien kennzeichnend :

1. Die Maschinen lesen eine nahezu unbegrenzt grosse Menge von Daten und Informationen selbständig ab, um sie weiter zu bearbeiten, sofern diese zuvor auf ein dafür geeignetes Speichermedium (Lochstreifen, Lochkarte, Magnetband usw.) übertragen wurden.
2. Art und Umfang der innermaschinellen Bearbeitung des Materials richtet sich nach dem Programm, das zuvor vom Menschen ausgearbeitet wird. Die Maschine arbeitet jeweils ein Programm vollständig ab und stoppt erst, wenn alle eingegebenen Befehle ausgeführt sind. Obwohl bei kleineren, besonders elektromechanischen Anlagen die Zahl der Programme, die gleichzeitig ablaufen können, beschränkt ist – also evtl. mehrere Teil- und Unterprogramme zur Lösung eines Aufgabenkomplexes erforderlich werden können – obwohl also Beschränkungen aus technologischen Erwägungen eintreten können, hängt die Qualität der

gewonnenen Ergebnisse u. E. ausschliesslich davon ab, wie gut oder wie ungeschickt diese Programme aufgebaut wurden.

Alle Aufgaben müssen für die Maschine auf logische Operationen zurückgeführt werden, also vorwiegend auf Systematisierung, Selektion, Vergleichung, Berechnen. Bei der Sprachdaten- oder Informationsverarbeitung erwächst ein besonderes Problem daraus, dass natürliche Sprachen in wesentlichen Elementen nicht logisch strukturiert sind. Hier besteht eine zusätzliche Aufgabe der Programmierung darin, die nichtlogischen Elemente natürlicher Sprachen sinnvoll in die logischen Operationselemente der Maschine zu überführen.

3. Die Ausgabe der gewünschten Ergebnisse erfolgt ebenfalls maschinell, gleichgültig ob es sich um Zahlen, Summen, Wörter oder Sachverhalte handelt. In der Regel lässt man diese Ergebnisse auf fortlaufendes Papierband drucken; sie können jedoch gleichzeitig oder ausschliesslich auf einen neuen Maschinenspeicher übertragen werden.

Bei der Ausgabe können je nach Wunsch einzelne Partien umgestellt oder unterdrückt werden wie auch zusätzliche Informationen – z. B. eine fortlaufende Numerierung oder das Tagesdatum – mitgeschrieben werden.

4. Alle Operationen laufen in Geschwindigkeiten ab, die weit über dem menschlichen Vorstellungsvermögen liegen, das sich ja nur an der manuellen Arbeitsweise orientieren kann. Lochkarten-Sortiermaschinen sortieren ca. 45.000 Karten pro Stunde; moderne, grosse Elektronenrechner führen heute in einer Sekunde mindestens 1 Million Operationen aus. Die Druckwerke der Ausgabe-Einheit schreiben, je nach Maschinentyp, zwischen 9.000 und 80.000 Zeilen pro Stunde. Das

bedeutet, dass selbst das langsamste Druckwerk in der Stunde ca. 1 Million Anschläge ausführen kann, sofern man die Zeilenbreite voll ausnutzt. Mit Maschinen, für die diese 4 genannten Kriterien kennzeichnend sind, haben wir also seit mehreren Jahren an einem Ausschnitt der überlieferten Linear-B-Tafeln — genauer : an einer Tafelgruppe — experimentiert, um die verschiedenen Anwendungsmöglichkeiten zu studieren. Allerdings sind diese Erfahrungen auch wesentlich bereichert worden durch Arbeiten an mehreren anderen natürlichen Sprachen, da ja unsere Maschinenstation einen sehr weiten Aufgabenkreis auszufüllen hat.

Die erste Vorstellung, mit der ein Geisteswissenschaftler sich dazu entschliesst, seinen Materialfundus auf einen Maschinenspeicher zu übertragen, liegt normalerweise in der *Selektion*. Nachdem ihm die Operationszeiten, die wir unter 4. ganz grob umrissen, durch irgend einen Zufall bekannt werden und er sie mit dem riesigen Zeitaufwand mühsamen manuellen Suchens an seiner Handkartei in ein Verhältnis zu setzen sucht, muss der Wunsch wach werden, solche Maschinen für Recherche- und Selektionsarbeiten anwenden zu können. Für Fragen der Art : "Kommt ein bestimmtes Silbenzeichen überhaupt in dieser oder jener Tafelgruppe vor und wenn ja wie oft ?" können natürlich solche Maschinen auch verwendet werden, obwohl sie dabei — wie wir wiederholt nachzuweisen suchten — schlecht eingesetzt sind. Die Selektionsaufgabe wird schon sinnvoller, wenn sie eine kombinatorische Recherche umfasst, etwa der Art : "Suche alle Sachverhalte A, sofern ihnen ein B folgt, jedoch darf ein C in diesem Beleg nicht enthalten sein." Dies wäre eine Aufgabe, die manuell schon nicht mehr zu lösen ist, wenn die Materialsammlung umfangreicher ist, d. h. mehr als ca. 10.000 Belege zählt. In den Auseinandersetzungen um die Entzifferung von Linear B wäre es oft von Bedeutung zu wissen, ob

ein bestimmtes Zeichen in Verbindung mit einem anderen definierten Zeichen vorkommt, ohne dass z. B. eine Zahlenangabe, ein Ideogramm oder eine Abkürzung folgt.

Mit einer kombinierten Frage hat man jedoch schon den halben Schritt zur nächsten Kategorie getan, nämlich der *Systematik*. Wenn man den Vorzug maschineller Verfahren erst einmal erkannt hat, nämlich Systematiken in beliebiger Zahl anfertigen zu können, wird man die Recherche oder Selektion entweder ganz unterlassen oder auf wenige Sonderfälle beschränken. Die herkömmliche Handkartei ist nur nach jeweils einem Gesichtspunkt geordnet. Wird ein anderes Kriterium zusätzlich benötigt, muss entweder diese Kartei umsortiert oder eine zweite angelegt werden. Aus rein ökonomischen Erwägungen und keinesfalls aus sachlichen Gründen beschränkt sich daher die Handkartei des Wissenschaftlers – sofern sein Materialfundus umfangreich ist – auf eine oder allenfalls zwei Karteien. Damit bleiben ihm aber zugleich Einsichten verschlossen, die aus der möglichen einen oder anderen Anordnungsweise hervorgehen können.

Ganz anders verhält es sich bei den maschinellen Methoden. Die Herstellung einer neuen Systematik erfordert im allgemeinen nicht mehr Zeit – und damit auch Geld – als die Recherche nach einem bestimmten Merkmal. Sobald man diese Einsicht gewonnen hat, wird man sich von der Maschine Systematiken nach allen möglichen interessierenden Gesichtspunkten anfertigen lassen. So haben wir für die genannte Tafelgruppe von Linear B in kurzer Zeit Systematiken nach dem "Alphabet" (d. h. dem aufsteigenden Zahlenschlüssel) vorwärtslaufend und rückläufig, nach dem Stellenwert des Zeichens im Wort, nach der Wortlänge, der Zeichenfrequenz, der Wortfrequenz der Zeichenkombination usw. usw. angelegt. Dazu gehörten auf Fragen der Art : Welches Zeichen kann nur am Wortanfang, welches nie am Anfang, welches nur am Wortende, welches nie am

Ende stehen. Welches Zeichen kann nur mit welchem anderen zusammen stehen, welches nie mit einem bestimmten anderen usw. Es dürfte einleuchten, dass derartige Systematiken, sinnvoll angelegt und richtig ausgewertet, einen sehr hohen Informationsgehalt besitzen. Andererseits darf die Leichtigkeit ihrer Produktion nicht dazu verführen, mehr Tabellen anzulegen, als ein Mensch vernünftigerweise noch gründlich auswerten kann.

Bevor wir in diesem Gedanken fortfahren, müssen noch zwei Besonderheiten maschineller Verfahren erwähnt werden. Wir hatten oben bei der Aufzählung der 4 Kriterien jeder Sprachdatenverarbeitung unter Punkt 2 sehr nachdrücklich darauf hingewiesen, in welcher hohen Masse die Qualität aller Ergebnisse von dem jeweiligen Programm abhängig ist. Jedes Programm ist jedoch teilweise wieder abhängig davon, mit welchen Daten oder Informationen man arbeiten kann, weil sie bereits im Speicher verzeichnet sind, und welche Informationen aus den bereits bekannten Fakten neu gewonnen werden müssen. Auch ohne dieses Problem, das übrigens sehr weitreichende Konsequenzen hat, nur annähernd zu überblicken, haben wir von Anfang an grosse Mühe und sehr, sehr viel Zeit auf die Frage verwandt, welche möglichen Fragestellungen im Laufe der Arbeit auftreten könnten und welche Informationen folglich als Minimum im Speicher enthalten sein müssen. Wir werden auf diese Frage weiter unten noch mal zurückkommen müssen. Hier soll jedoch auf ein Phänomen hingewiesen werden, das jeder kennenlernte, der sich eingehend mit Datenverarbeitung befasste. Es zeigt sich nämlich, dass man sein ureigenes Fachgebiet durchaus nicht so überblicken kann, wie man es glaubte und wie es für eine systematische Daten- bzw. Informationsaufbereitung unerlässlich ist. Fragen etwa der Art : Wie lang wird das längste Wort sein; wieviele Untergruppen werden die einzelnen grammatischen Kategorien etwa enthalten; welche Arten von textkritischen Angaben werden auftreten; wieviel Platz wird für die unterschiedlichen Formen der Referenz zu

jedem Wort nach den einzelnen Editionen benötigt; wieviele Ideogramme, Abkürzungen oder Zahlenangaben können einem Beleg folgen usw. usw., Fragen dieser Art also müssen erwogen werden, bevor man auch nur mit der Übertragung der Informationen auf Maschinenspeicher beginnt. Natürlich kann man all diese Fragen nicht vorher präzise beantworten — könnte man dies, so wüsste man ja bereits, was man durch die maschinellen Arbeiten erst erfahren will — aber man muss ungefähre Proportionen angeben können, muss Werte von "mindestens" bis "höchstens" nennen. Denn die überaus exakte, streng logisch orientierte Arbeitsweise der Maschinen zwingt uns, sowohl Art und Umfang unserer Materialien einerseits wie auch Arbeitsprozess und Ziel unseres Themas andererseits vorher zu durchdenken. Jede Nachlässigkeit in diesem vorbereitenden Arbeitsstadium zwingt uns später, wenn die Maschinen uns unsere Fehler unbittlich vorweisen, zu einem zusätzlichen Aufwand an manueller Tätigkeit, zu Mehrarbeit und Umdisponierungen, die eigentlich durchaus vermeidbar waren. Und so ergab sich bei allen Projekten, die wir bisher maschinell bearbeiteten — beginnend mit Linear B —, eine gründliche gedankliche Durchdringung des Aufgabenkomplexes allein schon aus den Vorbereitungsarbeiten, d. h. ein Grad der geistigen Auseinandersetzung mit dem Thema, der normalerweise bei geisteswissenschaftlichen Aufgaben erst mit Abschluss der Arbeit erforderlich ist und erreicht wird.

Eine ähnliche Vorbereitungsarbeit fordern einmal die Verteilung der Informationen auf die Speicherplätze sowie ferner die Ausarbeitung des Algorithmus für die einzelnen Arbeitsstufen bei der Auswertung der Materialien. Alle Maschinenspeicher erfordern im Prinzip eine Positionsschrift, d. h. die einzelnen Informationen müssen auf bestimmte gebundene Speicherplätze oder Adressen verteilt werden, wenn man sich später einen unsinnigen Aufwand technologischer Art ersparen will. Auch hierfür sind

Überlegungen von Seiten des Sprachwissenschaftlers unerlässlich, bei denen er sich Sachverhalte bewusst machen *und formulieren* muss, die er normalerweise nur unbewusst als allgemeine Grundkenntnisse seines Fachgebietes anwendet, ohne sie in ihren Verflechtungen und Konsequenzen zu überblicken. Die gedankliche Mehrarbeit, die hier von ihm gefordert wird, kommt ihm als Gewinn wieder unmittelbar zugute als systematische Kenntnis der Gegebenheiten seines Materials, jedoch auch mittelbar bei der maschinellen Auswertung.

Die Ausarbeitung eines Algorithmus, also eines schematisierten Ablaufes der einzelnen Arbeitsphasen, geht eigentlich schon über diejenigen Anforderungen hinaus, die man an einen Philologen billigerweise noch stellen kann. Hierfür ist der technische Sachverständige oder Programmierer zuständig.

Dieser jedoch ist nicht in der Lage, die einzelnen Fachgebiete – in unserem Falle die Inschriften von Linear B – so zu überblicken, dass die gespeicherten Materialien mit äusserstem Nutzen ausgewertet werden. Wie meist in Querschnittsdisziplinen ist es auch in der Sprachdatenverarbeitung unerlässlich, die Aufgaben in Gemeinschaftsarbeit zu lösen. Neben den intimen Kenner des Fachgebietes – in unserem Falle einen Spezialisten für Minoica – muss eine Art Dolmetscher treten, der so viel von allgemeiner Linguistik versteht, dass er die Vorstellungen des Fachphilologen in eine Maschinensprache und Schaltung umsetzen oder sie seinerseits wieder einem versierten Programmierer verständlich machen kann. Auch bei diesen gemeinsamen Vorbereitungsarbeiten werden vom Philologen immer wieder Angaben und Kenntnisse gefordert, die ihn zwingen, sich sein Fachwissen bewusst zu machen und präzise zu formulieren. Welcher Gewinn für ihn selbst aus solcher Form exakter Linguistik erwächst, ist oben schon erwähnt worden. Von der glücklichen Harmonie dieses Teams,

die sich kaum organisieren oder reglementieren lässt, hängt für die Qualität der schliesslichen Ergebnisse sehr viel ab : nämlich ob sie durchschnittlich brav ausfallen, oder ob sie über Erwarten gut gelingen. Tatsächlich fällt dem technischen Sachverständigen bei der Sprachdatenverarbeitung eine ungleich wichtigere Aufgabe zu als bei normaler Datenverarbeitung. Hier, wo zwei extrem auseinanderliegende Disziplinen, nämlich Sprachwissenschaft und Elektrotechnik, zum sinnvollen Ineinanderwirken gebracht werden sollen, reicht es nicht mehr aus, die Wünsche des Fachphilologen in ein Flussdiagramm oder eine Programmiersprache zu übertragen. Denn was intern in den Automaten vorgeht, entzieht sich ja durchaus dem Vorstellungsvermögen des Geisteswissenschaftlers. Daher müssen ihm weitreichende Vorschläge unterbreitet werden, wie seine Arbeitsvorhaben modifiziert, welche zusätzlichen Funktionen von der Maschine geleistet werden könnten und welche im linguistischen Sinne interessanten Effekte sich ausserdem erzielen liessen. Dem Fachwissenschaftler bleibt es dann überlassen, diese Vorschläge nach Prüfung gutzuheissen oder zu verwerfen.

Wir hatten oben die Vorzüge einer Systematik gegenüber der reinen Recherche dargelegt. Diese Forderung muss – ohne dass wir hier in Einzelheiten gehen wollen – noch etwas weiter ausgeführt werden. Denn die Möglichkeiten einer wirklich sinnvollen Sprachdatenverarbeitung werden u. E. noch viel zu wenig ausgenutzt bzw. sind noch kaum bekannt, weil die Gesellschaftswissenschaftler zu Unrecht versuchen, mit den Maschinen lediglich den menschlichen Arbeitsprozess zu kopieren anstatt die weitreichenden Möglichkeiten datenverarbeitender Geräte auch nur annähernd auszuwerten.

Die Aussagekraft einer Systematik erhöht sich, wie jedermann weiss, ganz wesentlich, wenn man anstelle isolierter Gruppen eine Hierarchie aufbaut,

in der Über-, Haupt- und Unterbegriffe die Verflechtungen des Materials einerseits wie auch ihre unterschiedliche Wertigkeit andererseits deutlich hervortreten lassen. Nun sind aber solche Hierarchien von Hand nur unter unendlichen Mühen aufzubauen, während Maschinen derartige Ordnungssysteme geradezu begünstigen. Auch hier wachsen dem Fachwissenschaftler aus der Vorbereitung und Auswertung der gewonnenen Gruppen Erkenntnisse zu, die ihm beim manuellen System im allgemeinen verschlossen bleiben. Hinzu kommt ein weiteres Phänomen, das theoretisch schwer zu beschreiben ist und eigentlich nur in der praktischen Arbeit erkennbar wird. Die langen Papierbänder nämlich, die der Ergebnisdrucker auswirft, verführen mit ihrer Regelmässigkeit in der Anordnung der Wörter und Daten zum *vertikalen* anstelle des herkömmlichen *horizontalen* Lesens. Der Wissenschaftler verfolgt nicht mehr eine Zeile von links nach rechts, sondern statt dessen eine Kolumne von oben nach unten. Dabei treten Besonderheiten, also Abweichungen von der herkömmlichen oder erwarteten Norm, so deutlich hervor, dass sie geradezu ins Auge springen. Es bleibt einem weiteren Arbeitsgang überlassen zu untersuchen, ob derartige Abweichungen oder Häufungen nur zufällig sind oder ob sich hinter ihnen eine bisher noch nicht entdeckte Gesetzmässigkeit verbirgt. So dient gerade dieses vertikale Lesen, richtig angewendet, sowohl einer schrittweisen Optimierung der vorgesehenen Programme wie auch einem immer zielstrebigeren Hinwenden auf die entscheidenden Probleme. Man kann gar nicht oft genug betonen, welcher eminent heuristische Gewinn aus solchen Ergebnislisten erwächst, wenn sie vernünftig angelegt und klug ausgewertet werden. Dieses Prinzip, von vielen unterschätzt oder noch gar nicht erkannt, ist u. a. von H. Spitzbardt* hervorgehoben worden.

* Zur Algorithmierung des Nominalkomplexes im Indonesischen. Beiträge zur Sprachkunde und Informationsverarbeitung, 5, 1965, pp. 54-79, 6, 1965, pp. 33-64.

Es gibt wohl in jeder wissenschaftlichen Disziplin Vorurteile, die alle Gegebenweise – seien sie noch so einleuchtend geführt oder sachlich fundiert vorgebracht – hartnäckig überdauern. Die Beständigkeit solcher Vorurteile erklärt sich wahrscheinlich aus der Tatsache, dass ihnen Halbwahrheiten zugrunde liegen, d. h. Sachverhalte, die nur unter ganz bestimmten Bedingungen wirksam werden; verallgemeinert man nun diese Sachverhalte oder löst sie auch nur aus ihrer Kausalkette heraus, so blähen sie sich leicht zu drohender Allgemeingültigkeit auf. Im Gebiet der Datenverarbeitung heisst eines dieser Vorurteile, dass man aus dem Material mit Hilfe von Maschinen nur solche Fakten abfragen oder heraussuchen könne, die vorher eingespeichert worden seien. Diese Behauptung liesse sich unter extremen Bedingungen sogar rechtfertigen : Wenn man nämlich keine einzige Information einspeichert, kann man – selbst mit dem raffiniertesten Programm – nicht mehr als ein Ergebnis “Null” oder “Nichts” herausholen. Umgekehrt wird jeder Suchprozess umso einfacher ablaufen, je mehr Informationen bereits eingespeichert sind. Andererseits lässt sich die Unsinnigkeit des genannten Vorurteils an einem ganz simplen Exempel demonstrieren. Wenn man z. B. der Maschine eine beliebig grosse Menge von Zahlen zum Aufaddieren eingibt – also etwa für eine Materialabrechnung, Lohnberechnung oder eine vergleichbare Aufgabe, für die ja eigentlich diese Geräte konstruiert sind – so erhält man zum Schluss die Summe aus den Einzelposten, also aus den gespeicherten Summanden. Diese Summe war in keinem einzigen Einzelbeleg enthalten, sondern wurde neu in den selbständig arbeitenden Rechenwerken der Maschine gebildet, sofern mehr als ein Summand eingegeben wurde. Nicht anders verhält es sich mit komplizierten Rechenoperationen, die übrigens bei datenverarbeitenden Maschinen fast immer auf die Addition von Summanden oder deren Komplement reduziert werden. Es können – oder richtiger : es müssen – also von den Maschinen auch Ergebnisse ausgegeben werden, die ihr nicht über den Speicher eingegeben wurden; wäre dies nicht möglich, so gäbe es heute

keine Datenverarbeitung, sondern allenfalls primitive Büromaschinen. Wie verhält es sich aber nun bei Informationen, also bei Werten, die nicht in Rechenoperationen einbezogen werden können, wie z. B. Wörter, grammatische Kategorien, usw. ? Wie schon oben gesagt, werden Daten – also Zahlenwerte – *berechnet*, während Informationen – in unserem Falle vorwiegend sprachliche Zeichen – *systematisiert* werden. Nun wird niemand bestreiten, dass eine Systematik, die dem Wesen des Materials entspricht, in ihrer Kategorie einen vergleichbar ebensogrossen Aussagewert enthalten kann wie Ergebnisse aus der Kategorie der Rechenoperationen. Allerdings lassen sich mathematische Prozeduren leichter formalisieren und bewerten, aber dies ist ein gradueller und kein prinzipieller Unterschied, der hier nicht weiter verfolgt zu werden braucht. Beim heutigen Stand unserer Erfahrungen würden wir behaupten, dass maschinelle Sprachverarbeitung umso gewinnbringender ausfällt, je grösser die Spanne von vorgegebenen Informationen einerseits und dem aus der maschinellen Prozedur gewonnenen Ergebnis andererseits ist. Mehrere Laboratorien für Sprachdatenverarbeitung haben diesen Gesichtspunkt selbständig erkannt. Seine praktische Verwirklichung, die allerdings schwieriger ist als die theoretische Rechtfertigung, wurde u. a. in Liège bei der Identifizierung lateinischer Grundwörter und in Berlin bei der Bearbeitung der Deutschen Verben demonstriert. Diese Erkenntnis, deren Verwirklichung in fast jeder Fachrichtung anderen Gesetzen unterworfen ist, führt zum letzten, noch schwerer zu beschreibenden Schritt unserer Konzeption hinüber.

Wir gehen von dem Gedanken aus, dass man die Gesetze einer natürlichen Sprache sowohl im Bereich der Oberflächen- wie in dem der Tiefen-Struktur erst dann vollständig erkannt hat, wenn man eben diese Gesetze bis ins Detail formalisieren kann. Ein Mensch kann jedoch, allein auf sein Sprachgefühl gestützt, nicht feststellen, ob diese Gesetze vollständig und widerspruchsfrei sind. Er ist aber heute in der Lage, die formalisierten Regeln

einer Maschine, als Programm einzugeben und damit Sätze künstlich erzeugen zu lassen. Übrigens müssen das nicht Sätze sein, es können ebensogut Flexionsformen, Kompositions- oder Derivationsprozesse usw. vorgenommen werden. Am Ergebnis solcher generativen Prozeduren wird der Fachwissenschaftler Lücken oder Widersprüche seines Regelsystems leicht erkennen, diese berichtigen und sich mit solcher schrittweisen Optimierung einer kompletten Grammatik oder Semantik annähern. Der Linguist müsste sich also eine bestimmte Theorie zu einem Problem bilden, eine Methode zu deren Beweisführung ausarbeiten und am praktischen Ergebnis schliesslich Theorie und Methode auf ihre Richtigkeit überprüfen.

Mit diesen Grundsätzen zur Sprachdatenverarbeitung, wie wir sie heute sehen, ist noch nicht viel zum speziellen Problem von Linear B ausgesagt. Wir wollen versuchen, nachdem die Voraussetzungen grob umrissen sind, deren praktische Anwendung für Fragen der Entzifferung, oder richtiger : der Untersuchungen an Linear B, knapp zu umreissen.

Auf den ersten Blick stellen die auf den Tafeln überlieferten Texte einen geradezu idealen Versuchsgegenstand für maschinelle Sprachverarbeitung dar. Alle Verfahrensarten, die wir oben knapp darlegten, können von völlig neutralem, also vorurteilsfreiem Standpunkt aus experimentell angewendet werden, ohne dass Vorkenntnisse, die jeder Mensch bei Bearbeitung einer Sprache mitbringt und unbewusst wie auch ungewollt in die Untersuchung hineinträgt, den streng sachlichen Arbeitsablauf beeinträchtigen könnten. Die Sache hat nur einen Haken : Die Ergebnisse können nicht im herkömmlichen Sinne auf ihre Richtigkeit geprüft werden. Daher bestand von Anfang an die Gefahr, ungesicherte oder gar fragwürdige Hypothesen aus einem vorangehenden Arbeitsabschnitt als zuverlässiges Ergebnis in eine nächst höhere Etappe zu übernehmen. Wir müssten daher eine ungebräuchliche, weitgehend formalisierte Kategorie von "richtigen

Ergebnissen" zugrundelegen, die in späteren Arbeitsgängen durch andersartige Fragestellung immer wieder auf ihre Zuverlässigkeit zu prüfen war. Leider müssen wir daher auf semantische Kriterien so gut wie völlig verzichten, solange nämlich das Verhältnis von Ideogrammen zum vorangehenden Wort noch ungeklärt oder zumindest umstritten ist.

Man sollte annehmen, dass die Verwendung von datenverarbeitenden Maschinen im Bereich der Sprachenzifferung vorwiegend in statistischen Untersuchungen zu suchen sei. Tatsächlich sind ja auch die Maya-Texte im sibirischen Rechenzentrum der UdSSR entschlüsselt worden. Es darf aber nicht übersehen werden, dass bei diesen Arbeiten eine Bilingua zugrunde lag und folglich die Hauptaufgabe in der Zuweisung von Korrelaten nach Gesetzen der Häufigkeit bestand. Unsere Aufgabenstellung für Linear B bietet diese Möglichkeit bekanntlich nicht. Überhaupt scheint uns, dass die Anwendung mathematisch-statistischer Methoden bei natürlichen Sprachen in jüngster Zeit etwas überschätzt oder – um es präziser zu formulieren – nicht immer sachgerecht vorgenommen wird. Die statistisch häufigsten Buchstaben der deutschen Sprache sind z. B. nach Meyer-Eppler* e, n, s, i, r, a, d, t, u, h, l, c.

Die durchschnittliche Wortlänge liegt im Deutschen nach unseren Erfahrungen bei etwa 12 Buchstaben. Mathematisch gesehen müsste man also mit diesem Inventar einen sehr grossen Teil des deutschen Wortschatzes bilden können. Man könnte die Frequenz dieser Buchstaben auch noch nach ihrem Stellenwert im Wort untersuchen und dann etwa die 200 häufigsten Wörter – statistisch gesehen "häufigsten" Wörter – maschinell erzeugen. Vermutlich wäre kein einziges echtes deutsches Wort darunter,

* Grundlagen und Anwendungen der Informationstheorie : Berlin u. a., 1959, p. 64.

denn statistische Ergebnisse repräsentieren ja nur einen Mittelwert, der im Einzelfalle – sogar in jedem Einzelfalle ! – falsch sein kann. Natürlich unterschätzen wir die statistischen Ergebnisse, die wir bei den Untersuchungen an Linear B gewinnen, durchaus nicht. Aber wir bemühen uns – vielleicht aus einem ungerechtfertigten Vorurteil – ihnen nur die Funktion eines Mittelwertes zuzuerkennen.

Viel wichtiger scheinen uns Häufungen, Verdichtungen oder andererseits Unika im Material zu sein, aus denen man erste Anzeichen über Sprachtypus und Sprachstruktur entnehmen könnte. So haben wir u. a. von der Maschine Konstruktionen bilden lassen, die uns Aufschlüsse darüber geben sollten, ob es sich bei Linear B um eine suffigierende Sprache handeln könne. Es wurden dabei aus einem rückläufigen Index häufig vorkommende Silbenzeichen am Wortende entnommen und künstlich an andere "*Wörter*" angehängt. Die so entstandenen fiktiven Wortgebilde wurden dann maschinell mit dem echten Wortbestand verglichen, um zu prüfen, ob sie auch tatsächlich belegt sind. Abtrennungsoperationen und Distributionsanalysen dienten der gleichen Zielsetzung.

Für jedermann einleuchtend werden die Vorteile sein, die Maschinen bei der Konjekturekritik bieten. Nehmen wir den einfachen Fall einer Lacuna zwischen zwei eindeutig identifizierbaren Zeichen. Man braucht sich jetzt nur von der Maschine alle Belege zusammenstellen zu lassen, in denen die beiden überlieferten Zeichen mit einem oder – je nach Breite der Lacuna – mehrere Zeichen vorkommen. Diese Wörter kann man noch nach ihrer Länge, dem Stellenwert der Zeichen usw. feiner sortieren lassen. Dann hat man alle Materialien beieinander, die sich ein Wissenschaftler bei der Textkonstitution nur wünschen kann, allerdings ohne jede semantische oder inhaltliche Kriterien. Diese Einschränkung ist jedoch primär nicht in der Verwendung von Maschinen, sondern in der Überlieferungslage von

Linear B begründet. Will man bei solchen maschinellen Materialhilfen mehr erzielen als einen simplen Selektionsvorgang, so muss man beginnen, Fragen zu stellen, warum dieses und kein anderes Zeichen am wahrscheinlichsten für die angenommene Lücke zu konjizieren sei. Diese Fragen müssen, wie wir oben schon zeigten, in ein Programm umgesetzt und regelrecht nach allen Varianten durchgespielt werden. Wenn man solchen schnell arbeitenden Maschinen als denkender Mensch gegenüber steht, wird einem erst bewusst, dass es eine Kunst ist, sinnvolle Fragen auszudenken und zu formulieren, dass jedoch andererseits in einer guten Frage schon ein Teil der Antwort enthalten sein kann. Wir wollen hier auf die Darstellung weiterer Anwendungsmöglichkeiten verzichten und statt dessen fragen, welche Aussichten bestehen, für Linear B mit Hilfe der Maschinen eine überzeugendere Entzifferung zu finden.

Die Antwort darauf kann kurz gefasst werden : Für die nächste Zeit bestehen kaum Aussichten. Der gespeicherte Fundus gibt interessante Einblicke, ist aber noch zu gering, um zuverlässige Aussagen über Linear B davon abzuleiten. Das ist aber nicht der wichtigste Grund für die Verzögerung. Vielmehr vertreten wir den Standpunkt, zuerst einmal die bisher vorliegende Einzelergebnisse oder geschlossenen Theorien zur Entzifferung auf ihre Konsequenz im Detail, d. h. auf ihre Widerspruchsfreiheit und Wahrscheinlichkeit hin zu untersuchen. Selbstverständlich spielt dabei die Entzifferung von Ventris und Chadwick eine hervorragende Rolle. So ist es z. B. für die Maschinen kein Problem, den internationalen Zifferncode für die einzelnen Zeichen in angenommene Lautwerte zu konvertieren. Die Untersuchung solcher Phonemdistributionen wiederum lässt Aussagen darüber zu, ob das System — selbst wenn man reichlich Sprossvokale konzidiert — vom Menschen noch artikuliert werden könnte, mithin wie hoch der Wahrscheinlichkeitsgrad für die Richtigkeit eines solchen Lautsystems sei. Erst wenn diese sowie viele weitere Fragen gründlich

untersucht sind, für deren zeitliche Verzögerung nicht die Maschinen, sondern die Unzulänglichkeit des Menschen die Ursache sind, könnte man an die Ausarbeitung einer eigenen Theorie zur Entzifferung gehen. Das liegt jedoch im Ermessen des Fachwissenschaftlers; unsere Aufgabe ist es lediglich, Wege und Verfahren zu finden, die dem Menschen ein Höchstmass an systematisiertem Material, an heuristischer Anregung und an praktischen Experimenten nach begründeten Theorien anbieten.

Werden also die Maschinen in der Zukunft eine nicht anfechtbare Entzifferung von Linear B ermöglichen ? Über den Zustand einer Brücke soll man besser erst urteilen, wenn man am anderen Ufer angelangt ist. Eines kann man jedoch mit Sicherheit vorhersagen : Richtig eingesetzt können solche Maschinen einen Fundus gesicherter Fakten aufbereiten, auf den sich schrittweise immer neue Erkenntnisse gründen lassen. In diesem Punkte übertreffen sie weit die Leistungsfähigkeit des einzelnen Wissenschaftlers wie auch die einer ganzen Gruppe. Eines jedoch scheint uns noch wichtiger zu sein : Bei jedem Misserfolg eines maschinellen Arbeitsprozesses kann man zugleich die Begründung ablesen, warum das Ergebnis ohne Gewinn blieb. Dies ist ein Faktor, der nicht nur die Maschinenarbeit von der menschlichen unterscheidet. In ihm liegt zugleich der Keim für einen weiteren Versuch : Aus dem Zustand wird eine Entwicklung, aus der These ein dialektischer Prozess. Es sollte Aufgabe der Fachwissenschaftler sein, aus diesen Gegebenheiten Gewinn zu ziehen.

Erich MATER
Deutsche Akademie der
Wissenschaften zu Berlin.