KYBERNETIK UND ALTAISTIK

VON

HANS-PETER VIETZE

Berlin

0. Vorbemerkungen zur Kybernetik

0.1. *Gegenstand der Kybernetik*

Die Kybernetik entstand in der Zeit des 2. Weltkrieges, als eine Forschungsgruppe um den amerikanischen Kybernetiker Norbert Wiener die Analogie von Steuerungs- und Regelungsprozessen sowie von Prozessen der Nachrichtenübertragung im Lebewesen und in der Maschine erkannte. Norbert Wiener war es auch, der dieser neuen Wissenschaft den Namen gab und ihren Gegenstand im wesentlichen abgrenzte. Die Kybernetik erforscht

- A. Transformationen von Informationen und
- B. Steuerungs- und Regelungsprozesse in organischen und nichtorganischen Systemen.

Beispiele für Transformationen von Informationen:

- a. alle bedingten und unbedingten Reflexe in biologischen Systemen,
- b. jede Kommunikation,
- c. jede geistige Verarbeitung eines gebotenen Stoffes, z.B. Entzifferung eines verstümmelten Textes, Erlernen einer Fremdsprache,
- d. Human- und Maschinenübersetzung,
- e. die maschinelle Herstellung eines rückläufigen Wörterbuches.

Beispiele für Steuerungs- und Regelungsprozesse:

- a. jede gezielte menschliche Tätigkeit (Steuerung und Regelung erfolgt durch das Gehirn),

- b. Erziehung von Schülern und Studenten,
- c. Steuerungs- und Regelungsprozesse in einer Übersetzungsmaschine (z.B. Abtasten des Lexikons der Zielsprache u.a.).

Der Gegenstand der Kybernetik zeugt von ihrem hohem Abstraktionsgrad. Sie abstrahiert von allem Gegenständlichen und ist – ähnlich wie die Mathematik – eine Strukturwissenschaft.

0.2. *Die theoretische Bedeutung der Kybernetik*

Die Kybernetik liefert die theoretische Grundlage für die maschinelle Informationsverarbeitung und zwar sowohl für die ingenieurtechnische Seite des Computerbaus wie auch für die Problemanalysen und Programmierarbeiten. Daneben führt die Anwendung kybernetischer Theorien und Methoden in einigen bestimmten Fällen auch ohne den unmittelbaren Zielbezug einer maschinellen Abarbeitung zu neuen Erkenntnissen und Problemstellungen.

0.3. *Die praktische Bedeutung der Kybernetik*

Die Zahl der aktiven elektronischen Bauelemente kann als Gradmesser für den Stand der Industrialisierung eines Landes angesehen werden. Mit der zur Zeit vorstatten gehenden Einführung von Computern der III. Generation, die in der Lage sind, integrierte Leitungs- und Informationssysteme aufzubauen, wird die Zahl der aktiven elektronischen Bauelemente gleichzeitig zu einem Maßstab für die Effektivität der Innen- und Außenpolitik sowie vieler Wissenschaftsbereiche eines Landes. Nach der mit großen Aufwand erarbeiteten naturwissenschaftlichen Prognose der RAND-Corporation (Santa Monica, USA) ist für 1972 mit der Maschinenübersetzung und für 1978 mit automatischen Dolmetschern zu rechnen (mit einer Toleranz von + 3 und – 6 Jahren).¹

¹ Die Prognose wurde stichpunktartig von Rotta und Schmid in der »Naturwissenschaftlichen Rundschau« 20 (1967), 7, Stuttgart, veröffentlicht.

1. Möglichkeiten der Anwendung der Kybernetik in der Altaistik

Auf Grund des hohen Abstraktionsgrades der Kybernetik sind fast alle kybernetischen Theorien und Methoden in der Altaistik anwendbar, wobei der Grad der Effektivität sich mit dem unmittelbaren Zielbezug der maschinellen Abarbeitung erhöht. Wir wollen uns auf eine kurze Auslese beschränken.

1.1. Systemtheorie

Die Systemtheorie ist von allen Theorien, die in den Rahmen der Kybernetik gehören, die abstrakteste und wohl auch diejenige, die am ehesten ohne mathematische Methoden auskommt. Sie untersucht die Beziehungen, die zwischen Systemen und innerhalb von Systemen bestehen, wobei unter einem System jeweils eine Menge von Elementen und von zwischen diesen Elementen bestehenden Relationen zu verstehen ist. So ist z.B. jeder Satz und jedes Wort ein System. Beim Wort können die Buchstaben als Elemente und ihre Anordnungen als Relationen verstanden werden. Die Systemtheorie ist bestrebt, die Menge aller Systeme unter dem Aspekt ihrer Struktur und Funktion in Klassen zu zerlegen. Da Sprache, Geschichte, Ökonomie und alle anderen für die Altaistik wichtigen allgemeinen Disziplinen eine Vielfalt von Systemen darbieten, können sie natürlich auch systemtheoretisch untersucht werden.

So kann z.B. die Anwendung der Systemtheorie zu der Erkenntnis führen, daß die Anfügung des mongolischen Suffixes *-нд* an einen Verbalstamm ein geschlossenes, die Anfügung des Verbsuffixes *-наа*, *-наә*, *-ноо*, *-ноо* aber ein offenes kybernetisches System darstellt, was jedoch zunächst noch nicht wesentlich weiterführt und keinen unmittelbaren ökonomischen Nutzen bringt, wenngleich es zu einer abstrakteren Betrachtungsweise und zur Entdeckung überraschender Analogien in Erscheinungen (Systemen) führt, die zunächst als völlig verschiedenartig angesehen wurden. Bei der maschinellen Abarbeitung, z.B. bei der automatischen Synthese mongolischen Sprachmaterials, wird dieser Umstand jedoch relevant, da die geschlossenen Systeme nur eines minimalen Programmieraufwandes bedürfen.

Die Systemtheorie kann weiterhin dazu beitragen, gesellschaftliche Erscheinungen und Prozesse in den Ländern, die den Gegenstand altaistischer Forschungen bilden, der automatischen Verarbeitung zugänglich zu machen.

1.2. *Informationstheorie*

Die Informationstheorie führt in vielen Fällen zu signifikanten Ergebnissen bei Redundanz- und Relevanzberechnungen für die Maschinenübersetzung, die automatische Exzerption und Dokumentation, die automatische Zeichenerkennung und bei jeder Kodierungs- und Dekodierungsarbeit, wobei alle diese Problemkreise (außer dem letzten) für die altaischen Sprachen allerdings wenig aktuell sind. In vielen anderen Fällen ist die Informationstheorie zwar anwendbar, ihre Anwendung aber ökonomisch nicht immer zu vertreten. So ist es z.B. ganz interessant, zu wissen, daß das Auffinden der nächsten Rune in dem verstümmelten alttürkischen Satz *Ynanču apa yarpan tarqan atı . . .* (Kül-tigin-Inschrift, Tafel XVI)² beim Philologen eine Ungewissheit beseitigt, die einem Informationsgehalt von rund 5 bit entspricht, der effektive Nutzen dieses Wissens ist jedoch gering.

1.3. *Spieltheorie*

Die Spieltheorie kann für die Altaistik nutzbringend sein, wenn man zeitgeschichtliche und prognostische Forschungen als Bestandteil der Altaistik versteht. Sie dient der Ermittlung jeweils optimaler Strategien für kybernetische Systeme in Konfliktsituationen und gibt Auskunft über den Wahrscheinlichkeitsgrad der möglichen Entscheidungsvarianten ökonomischer, politischer, militärischer u.a. Auseinandersetzungen. Derartige »Spiele« wären etwa die Auseinandersetzung der Interessengruppe, die für den Verbleib der Türkei in der NATO ist, mit derjenigen, die dagegen ist, die Distribution außenpolitischer Ambitionen der Türkei, Israels und der arabischen Staaten, die Haltung der Führung der Kommunistischen Partei Chinas zur Politik der Mongolischen Volksrepublik u.a.m. Derartige Untersuchungen verlangen den

² s. v. Gabain, *Altürkische Grammatik*, Harrassowitz Leipzig 1950, S. 257.

Einsatz größerer Kollektive und bleiben in der Regel speziellen Research-Zentren vorbehalten, von denen sie naturgemäß nicht publiziert werden.

1.4. *Automatentheorie*

Die Theorie der endlichen abstrakten Automaten dürfte für die Altaistik von großem Interesse sein. Sie liefert z.B. ein mathematisches Modell für einen Übersetzungsautomaten bzw. Modelle für seine Teile (Blöcke), wobei sie gleichzeitig zur Minimisierung (optimalen Ausstattung) eines solchen Automaten verhilft (z.B. durch Vermeidung äquivalenter Zustände). Daneben gibt sie u.a. auch die Möglichkeit zur exakten Definition linguistischer Grundeinheiten. Es sei uns erlaubt, hierfür ein Beispiel anzuführen:

In der strukturellen Linguistik wird ein Morphem meistens als »kleinste bedeutungstragende Einheit« bezeichnet. Das ist jedoch keine Definition, sondern eine Quelle mannigfaltiger Äquivokationen. Als eine »kleinste bedeutungstragende Einheit« kann z.B. der senkrechte Strich | bei der alttürkischen Rune D (y in einem hintervokalischen Wort der Orchon-Inschriften) verstanden werden, denn er ist zweifellos eine »kleinste Einheit« und trägt die Bedeutung 'Lautwert y' im Gegensatz zu der Rune) (n in hintervokalischen Wörtern der Orchon-Inschriften), bei der der senkrechte Strich fehlt. Auch die Bezeichnung »bedeutungstragend« ist prätheoretisch nicht abzugrenzen. Man gelangt zu Fragestellungen wie der folgenden: Ist HYB (mong. UNO) ein Morphem oder ein Wort, das aus 3 Morphemen besteht? Intersubjektiv kontrollierbare Entscheidungen kann hier ein abstrakter Automat mit dem äußeren Input »mongolische Wörter $\{W_1, W_2, \dots, W_n\}$ « und dem äußeren Output »morphologische Struktur der mongolischen Wörter $\{S(W_1), S(W_2), \dots, S(W_n)\}$ « bringen, wobei die Ausstattung eines solchen Automaten auf Grund der notwendigen Einbeziehung der Semantik zwar kompliziert, aber grundsätzlich möglich ist.³ Neben der Morphemdefinition wäre dieser Automat

³ An einem abstrakten Automaten zur monematischen (= morphologischen) Analyse mongolischer Wörter wird in der Forschungsgruppe »Anwendung der Kybernetik und EDV« der Sektion Asienwissenschaften der Humboldt-Universität zu Berlin gearbeitet.

theoretisch gleichzeitig ein Teilprogramm für die Maschinenübersetzung, falls ein kanonisches (herkömmliches) Wörterbuch gespeichert werden soll. Da sich jedoch immer mehr herausstellt, daß ein solches Verfahren ökonomisch ungünstig ist, wird der Automat vor allem dazu dienen, zu berechnen, ob es rationeller ist, in einem Übersetzungsautomaten ein nichtkanonisches oder ein gemischtes kanonisch-nichtkanonisches Wörterbuch zu speichern, wobei im letzten Fall noch der Anteil der kanonischen Eintragungen und der nichtkanonischen Eintragungskomplexe festgelegt werden kann. (Ein nichtkanonischer Eintragungskomplex ist z.B. mong. мал 'Vieh', малын 'des Viehs' малд 'dem Vieh' usw., insgesamt 24 Eintragungen). An eine mongolische Maschinenübersetzung ist zwar im gegenwärtigen Zeitpunkt noch nicht zu denken, die mit Hilfe des Automaten zu treffenden Aussagen können jedoch auf andere agglutinierende Sprachen, die der maschinellen Übersetzung näher sind (japanisch, ungarisch), extrapoliert werden.

2. Mögliche praktische Automatenarbeiten auf dem Gebiet der Altaistik

Die Möglichkeiten, Automaten zur Lösung wissenschaftlicher Probleme einzusetzen, sind praktisch unbegrenzt, wenn auch ihr Einsatz nicht immer effektiv ist. Aus diesem Grunde sollen hier nur einige Beispiele für rationelle Arbeiten angeführt werden.⁴

⁴ Wir wollen uns bei den folgenden Beispielen für ökonomisch vertretbare Computerarbeiten nur auf die Sprachen und hierbei vor allem auf die modernen Sprachen orientieren, deren Untersuchung wir als Teilgebiet der Altaistik verstehen. In den Möglichkeiten des Computereinsatzes für die Bearbeitung älteren altaistischen Sprachmaterials s. John R. Krueger, *The Application of Computer Processing to Altaic Studies, Feasibility, Procedure and Problems in: Central Asiatic Journal*, Vol. XI, No. 3, Sept. 1966, S. 161–186. John R. Krueger gibt in diesem Aufsatz neben vielen wichtigen Anregungen für weitere Arbeiten (z. B. Abarbeitung von Hauers Mandschu-Grammatik und Radloffs Wörterbuch der Türk-Dialekte, Koordinierung sämtlicher mongolischer Wörterbücher) Angaben über die Art und Weise der praktischen Durchführung der Arbeiten, z. B. über die Art der Eingabe der Daten auf Zwischenspeicher (Lochkarten) sowie Effektivitätsberechnungen.

2.1. *Maschinenübersetzung*

Nachdem 1954 das erste Experiment auf einer IBM-Anlage durchgeführt worden war, nahm die Maschinenübersetzung vor allem in den USA und der Sowjetunion einen gewaltigen Aufschwung, wobei auch orientalische Sprachen einbezogen wurden. Allein auf diesem Gebiet wurden schon 1958 in der Sowjetunion folgende Probeübersetzungen durchgeführt:

Chinesisch-russisch: V. A. Voronin, Moskau.

Japanisch-russisch: M. B. Jefimov, Moskau; A. A. Babinzev, Leningrad.

Indonesisch-russisch: N. D. Adreev, Leningrad.

Vietnamesisch-russisch: N. D. Andreev, D. A. Batova, V. S. Panfilov, Leningrad.

Hindi-russisch: T. A. Katenina, Leningrad.

Birmanisch-russisch: O. A. Timofeeva, Leningrad.

Arabisch-russisch: O. B. Frolova, Leningrad.⁵

In zwei Unionsrepubliken (der Armenischen und der Grusinischen SSR) wurden Zentren für maschinelle Übersetzung geschaffen. Es ist bedauerlich, daß die Turksprachen von diesen Bemühungen unberührt blieben und noch keine Erfahrungswerte aus Probeübersetzungen altaischer Sprachen vorliegen.

Nach diesen Experimenten erkannte man bald, daß man zu weit gegangen war und eine vollwertige Maschinenübersetzung eine intensive Grundlagenforschung, vor allem auch auf dem Gebiet der Semantik, voraussetzt. In der gegenwärtigen Etappe ist man bemüht, die Computer für eben diese Grundlagenforschung einzusetzen, wie zuletzt auf dem 3. Symposium über Maschinelles Übersetzen in Jerewan 1967 deutlich wurde.⁶

Da keine der altaischen Sprachen zu den besonders übersetzungswürdigen gehört, wird man am besten daran tun, Erfahrungen bei der Maschinenübersetzung ähnlich strukturierter

⁵ Vgl. den Bericht über die »Allunionskonferenz über maschinelles Übersetzen«, Probleme der Kybernetik, Band 2, Akademie-Verlag Berlin 1963, S. 350 ff.

⁶ Vgl. Kunze, J. Bericht über das 3. Symposium über Maschinelles Übersetzen 1967 in Jerewan in: Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung, Band 20, Heft 5/6 (1967), S. 571.

Sprachen (japanisch, ungarisch)⁷ abzuwarten, ehe man an die direkte automatische Übersetzung der altaischen Sprachen herangeht, wobei am Anfang wahrscheinlich die automatische Synthese dieser Sprachen im Vordergrund stehen wird. Allerdings tragen fast alle praktischen und theoretischen linguistischen (und damit natürlich auch altaistisch-linguistischen) Automatenarbeiten in dieser oder jener Form zur Maschinenübersetzung bei und liefern extrapolierbare Ergebnisse, wie in Abschnitt 1.4. bereits hervorgehoben wurde.

2.2. Rückläufige Wörterbücher

In rückläufigen Wörterbüchern erfolgt die Anordnung der Lemmata nicht nach den 1., 2., . . . , n-ten, sondern umgekehrt nach den n-ten, n-1., n-2., . . . , 1. Buchstaben. Derartige Wörterbücher sind maschinell sehr schnell und einfach herzustellen, da die automatische Abarbeitung abgesehen von der Ein- und Ausgabe nur im Sortieren und in der Längenverschiebung besteht (die Längenverschiebung hat den Zweck, daß alle Eintragungen mit den letzten Buchstaben in der gleichen Spalte abschließen). Die maschinelle Herstellung bringt im Vergleich zur manuellen Herstellung für ein rückläufiges Wörterbuch mit 100.000 Lemmata bei Einsatz eines mittelschnellen digitalen Computers der II. Generation (etwa Robotron 300, DDR) eine Steigerung der Arbeitsproduktivität von rund 1.500 ‰, bei der bald realisierbaren automatischen Zeichenerkennung, d.h. bei vollautomatisierter Eingabe, von 20.000 ‰. Rückläufige Wörterbücher dienen zur Rekonstruktion von Wörtern, deren erster Teil beschädigt ist, sind aber auch für rezente Sprachen und unter ihnen in besonderem Maße für agglutinierende Sprachen von Bedeutung, da sie die gesamte Morphologie übersichtlich und statistisch auswertbar präsentieren. Sie dienen als heuristische Grundlage für automatische morphologische Analysen (vgl. Abschnitt 1.4) und Synthesen und sind somit ein wichtiger Beitrag für die Maschinenübersetzung, insbesondere unter dem Gesichtspunkt des ökonomischen

⁷ s. z. B. I. A. Meltschuk, Über die maschinelle Übersetzung aus dem Ungarischen ins Russische, Probleme der Kybernetik, Band I, Akademie-Verlag Berlin, 1962, S. 263–371.

misch optimalen Verhältnisses der Speicherkapazität zur Zahl der Operationen eines Übersetzungsautomaten. Bei der maschinellen Herstellung eines rückläufigen Wörterbuches kann eine Wortlängenstatistik abfallen, die ein Anhaltspunkt für die Berechnung der Inputzustände eines linguistischen Automaten ist. Auf dem Gebiet der Altaistik liegt bereits ein rückläufiges Wörterbuch vor (s. Abschnitt 3.2.), weitere werden sicherlich schnell folgen.

2.3. *Herkömmliche Wörterbücher*

Die maschinelle Erarbeitung eines normalen kanonischen Wörterbuches ist dann besonders sinnvoll und effektiv, wenn ein Wörterbuch aus einer Sprache A in eine Sprache B vorliegt, ein Wörterbuch aus der Sprache B in die Sprache A jedoch noch nicht vorhanden ist (so gibt es z.B. ein mongolisch-deutsches, aber noch kein deutsch-mongolisches Wörterbuch). In diesem Fall werden jeweils ein A-Lemma und sein B-Äquivalent auf einer Lochkarte gespeichert. Dann werden alle Lochkarten nach den 1., 2., . . . , n-ten Buchstaben der B-Lemmata sortiert und der Computer bzw. die Tabelliermaschine so programmiert, daß die B-Lemmata beim Druck ihren Platz mit den A-Lemmata vertauschen (die B-Lemmata kommen nach links, die A-Lemmata nach rechts). Es ist möglich, auch Komposita, Redewendungen und Beispielsätze in die Abarbeitung mit einzubeziehen. Natürlich bedarf ein auf diese Weise »gestürztes« Wörterbuch einer außerordentlich gründlichen Nachredaktion, da A-Lemma→B-Lemma richtig und gleichzeitig B-Lemma→A-Lemma falsch sein kann. Zudem können im A→B-Wörterbuch wichtige Wörter der Sprache B fehlen. So fehlen z.B. im mongolisch-deutschen Wörterbuch Eintragen wie Operette, Variété, Motel, Netzwerkplanung usw.

2.4. *Häufigkeitsminima*

Die maschinelle Herstellung eines Häufigkeitsminimums für den Sprachunterricht läuft darauf hinaus, daß ein größeres, sorgfältig ausgewähltes Corpus von Texten abgearbeitet wird, wobei der Computer die Häufigkeit des Vorkommens jedes einzelnen Wortes hochrechnet und die Wörter ihrer Häufigkeit nach ge-

ordnet und mit exakter Zahl des Vorkommens als Liste ausdrückt. Die Erarbeitung eines Häufigkeitsminimums bedarf Überlegungen bezüglich der Komposita. So müßte z.B. mong. *аж аxyй* als Wort behandelt werden, da die Kenntnis der Bedeutung von *аж* 'Leben, sein' und *axyй* 'id.' nicht zum Verstehen des Kompositums *аж аxyй* 'Wirtschaft' führt. Demgegenüber können die Komponenten des Synonymkompositums *холбоо харилцаа* 'Beziehungen' getrennt hochgerechnet werden, da sowohl die Komponenten wie auch das Kompositum im wesentlichen die gleiche Bedeutung haben. Bei Spezialhäufigkeitsminima (z.B. bei einem mathematischen Häufigkeitsminimum) müssen die Fachtermini höher bewertet werden. Man kann sie z.B. bei der Hochrechnung mit 2 multiplizieren.

Auch maschinell hergestellte exakte Häufigkeitsminima sind für die Maschinenübersetzung und automatische Dokumentation von großer Bedeutung, da aus ökonomischen Gründen die häufigsten Wörter so gespeichert werden müssen, daß sie am schnellsten abgerufen werden können.

2.5. *Nichtkanonische Wörterbücher*

Gleichzeitig mit einem Häufigkeitsminimum kann ein nichtkanonisches Wörterbuch auf der Grundlage desselben Corpus von Texten hergestellt werden. Es hat neben seiner prinzipiellen Bedeutung für die Maschinenübersetzung den Vorteil, daß bei der Hochrechnung der eingegebenen Daten zusätzlich zum Häufigkeitsminimum ersichtlich wird, welche grammatischen Erscheinungen am häufigsten sind und somit zuerst gelehrt bzw. bei der Aufstellung des Algorithmus für die Maschinenübersetzung besonders berücksichtigt werden müssen. Das trifft jedoch nur für agglutinierende Sprachen bzw. für Sprachen mit stark agglutinierendem Charakter zu.

2.6. *Graphem-, Phonem-, Morphem-, Syntagmata-, Wort-, Wortmengen- u.a. Distributionen*

Da das Prinzip aller dieser Distributionen dasselbe ist, sollen sie hier gemeinsam behandelt werden. Es geht dabei jeweils um maschinelle Verträglichkeitsuntersuchungen linguistischer Ein-

heiten mit oder ohne Statistik. Reduzieren wir zur Demonstration die Zahl der hinsichtlich ihrer Distribution zu untersuchenden Einheiten auf drei und bezeichnen sie mit a, b, c. Die Menge der Elemente {a, b, c} bezeichnen wir mit A. $F(A)$ sei die von A erzeugte freie Halbgruppe. Jetzt gibt es eine unendliche Menge von Konfigurationen $\{K_1, K_2, \dots, K_n, \dots\}$ (in der Halbgruppentheorie »Wörter« genannt), die Elemente der freien Halbgruppe über A sind, z.B.:

{a}, {ab}, {abc}, {aabac}, {bbbecabc}, {aaaacebcabcabc} usw.

Die Sprache realisiert eine endliche Teilmenge der unendlichen Menge von Konfigurationen. Seien a, b, c Morpheme, so ist z.B. die Konfiguration {aabac} für eine altaische Sprache höchst unwahrscheinlich und die Konfiguration {bbbecabc} unmöglich. Demgegenüber kommen {a} (z.B. atü. bäg), {ab} (z.B. bäglär) und {abc} (z.B. bägläri) zeifellos vor. Diese von der Sprache tatsächlich realisierte (ausgezeichnete) Menge von Konfigurationen aus der freien Halbgruppe über einem Alphabet linguistischer Einheiten ergibt dann implizite grammatische bzw. graphische, phonologische u.a. Gesetzmäßigkeiten der Sprache und ist für objektivierte Grammatiken und linguistische Programmierungsarbeiten von größter Bedeutung.

Die maschinelle Ermittlung von statistischen Graphemdistributionen kann bei einer Sprache wie der mongolischen, deren Orthographie noch nicht völlig eindeutig festgelegt ist, dazu dienen, diejenigen Formen der Rechtschreibung zu erfassen, denen die größte Häufigkeit zukommt. Das hat den Vorteil, daß eine Orthographiereform, wie sie z.B. in der MVR schon seit langem geplant ist, solche Formen der Rechtschreibung besonders berücksichtigen kann, die sich organisch herausgebildet haben, wobei deren prozentualer Anteil an allen vorkommenden Schreibweisen genau feststeht. So setzen sich z.B. die Schreibweisen явуулж, орчуулж usw. gegenüber явуулаж, орчуулаж (konverbale Formen der mong. Verben явуулах 'schicken' und орчуулах 'übersetzen') durch.

Die maschinelle Ermittlung der Distributionsstatistik der mongolischen Grapheme wäre weiterhin die Voraussetzung für die optimale Gestaltung der Tastatur einer mongolischen Schreibmaschine.

Im folgenden wollen wir andeuten, wie auch im Falle der maschinellen Erarbeitung einer Graphemdistribution die Maschinenarbeit wieder in den Dienst der Maschinenarbeit tritt. Dazu stellen wir uns die Aufgabe, die Zahl der theoretisch notwendigen Inputzustände eines Automaten zu berechnen, der Wörter einer Sprache M abarbeiten soll, die über 35 Buchstaben verfügt. Die maximale Wortlänge sei durch eine bei der Herstellung eines rückläufigen Wörterbuches abgefallene Wortlängensstatistik mit 20 Buchstaben gegeben. Käme jeder Buchstabe in jeder Position des Wortes vor, erhielten wir 35^{20} Inputzustände (direkte Potenz). Liegt eine exakte Graphemdistribution vor, sind wir nicht mehr gezwungen, für jeden der 20 Inputkanäle ein gleichmächtiges Alphabet von 35 Buchstaben anzusetzen. Ergibt die Graphemdistribution z.B., daß $\mathfrak{b}$, $\mathfrak{p}$, $\mathfrak{m}$, $\mathfrak{w}$ nicht Anlaut vorkommen, reduziert sich die Zahl der Inputzustände unseres Automaten auf $31 \cdot 35^{19}$. Kann in 2. Position kein $\mathfrak{m}$, $\mathfrak{b}$, $\mathfrak{r}$ stehen, reduziert sie sich weiter auf $31 \cdot 32 \cdot 35^{18}$. Können nach $\mathfrak{a}$, $\mathfrak{o}$, $\mathfrak{y}$ im Anlaut in 2. Position nicht $\mathfrak{e}$, $\mathfrak{n}$, $\mathfrak{o}$, $\mathfrak{c}$, $\mathfrak{w}$, $\mathfrak{b}$, $\mathfrak{m}$, $\mathfrak{a}$ folgen, verringert sich diese Zahl auf $31 \cdot 32 \cdot 35^{18-27}$ und bei der Unmöglichkeit der Verdoppelung von Konsonanten im Anlaut bei einem Alphabet, das 20 Konsonanten enthält, weiter auf $31 \cdot 32 \cdot 35^{18-27-20^2}$. Setzt man dieses Verfahren fort, was ebenfalls mit Hilfe eines Computers geschehen kann, führt die Graphemdistribution zur Zahl der theoretisch notwendigen Inputzustände eines Automaten zur Abarbeitung von Wörtern der Sprache M .

Mit Hilfe einer maschinell hergestellten statistischen Distributionsuntersuchung von Wörtern können auch fehlende Wörter in alten Texten rekonstruiert werden, wobei zusätzlich die Wahrscheinlichkeit für die Richtigkeit der Rekonstruktion angegeben werden kann. Gegeben sei z.B. ein alttürkischer Satz der Form $\#[a_1, a_2, a_3, a_4, \dots]\#$, wobei a_i ($1 \leq i$) für ein beliebiges aber feststehendes Wort und $\#$ für die Satzgrenzen steht. Kommen jetzt z.B. im gesamten alttürkischen Sprachmaterial 100 Wortfolgen der Form $[a_3, a_4, b]$ und 10 Wortfolgen der Form $[a_3, a_4, c]$ vor, ergibt die Rekonstruktion des nach a_4 folgenden Wortes mit einer Wahrscheinlichkeit von 0,9 b (0 = Unmöglichkeit, 1 = Sicherheit).

Zu diesem Problembereich der maschinellen Ermittlung von Distributionen gehört auch die im Allunionsinstitut für wissen-

schaftliche und technische Information der Akademie der Wissenschaften der UdSSR ausgearbeitete und auf Vorschlag von Y. V. Knorozov für die Entzifferung der Kitanschrift verwendete Methode der Positionsstatistik, die in Abschnitt 3.1. beschrieben wird.

2.7. *Kombinationen von automatischen linguistischen Abarbeitungsverfahren*

Die aufgezählten linguistischen Automatenarbeiten, die natürlich nur einen Teil der Möglichkeiten für den Computereinsatz in der Altaistik umfaßten, werden in einem noch wesentlich höheren Maße effektiv, wenn ein und dieselben auf Maschinenmedien gespeicherten Daten nach verschiedenen Programmen abgearbeitet werden. Derartige mögliche Kombinationen sind z.B.:

- a. Normales kanonisches Wörterbuch – Graphemdistribution – Morphemdistribution – Syntagmadistribution (wobei der Wert der letzten drei Distributionen allerdings eingeschränkt ist, da z.B. im Türkischen und Mongolischen keine Kasus- u.a. syntaktisch relevanten Formen berücksichtigt werden können).
- b. Allgemeines Häufigkeitsminimum – spezielles Häufigkeitsminimum – nichtkanonisches Wörterbuch – Graphem-, Morphem-, Syntagma-, Wort-, Wortmengendistribution u.a.

3. Auf dem Gebiet der Altaistik durchgeführte Automatenarbeiten

3.1. *Die maschinelle Entzifferung der Kitanschrift*

Bei der maschinellen Entzifferung der Kitanschrift handelt es sich um die erste altaistische Automatenarbeit. Sie begann 1962 mit der Aufstellung einer Liste aller bekannten und rekonstruierten Grapheme und Allographen der Kitanschrift, die 1966 von V. S. Starikov veröffentlicht wurde.⁸ Die Liste enthält 755 Grapheme und Allographen, denen für die maschinelle Abarbeitung jeweils eine dreistellige Zahl zugeordnet wurde. Graphisch aus-

⁸ V. S. Starikov, Catalogue of Graphemes of the Kitan Script (First variant), Nauka Publishing House, Moscow 1966 (anlässlich des XI. Pacific Science Congress, Tokyo, August-September 1966).

gezeichnete Graphemmen (Wortformen) wurden als Blöcke bezeichnet. Man erhielt z.B. folgende Blöcke: 110–182–277–004–084, 151–277–175, 115, 074–066 u.a.

Die numerische Transkription der Grapheme und Allographe erlaubte nun die Herstellung von einem Wörterbuch mit vorlaufender Anordnung und von 5 rückläufigen Wörterbüchern (der längste Block besteht aus 5 Graphemen [= dreistelligen Ziffern], so daß ein rückläufiges Wörterbuch für die fünften, eines für die vierten usw. bis ersten Grapheme hergestellt wurde). Die dann untersuchte Graphemdistribution innerhalb eines Blockes (von den Autoren Mikroparadigma genannt) gab bereits Aufschluß über die agglutinierende Struktur der Kitansprache. Danach wurde die Distribution der Blöcke untersucht, wobei bei einer Folge der Form $[B_1, B_2, \dots, B_i, \dots, B_n]$ ($1 \leq i \leq n$) von Blöcken bei Heraushebung eines Blockes B_i ($1 \leq i \leq n$) maximal eine Blockfolge $[B_{i-4}, B_{i-3}, B_{i-2}, B_{i-1}, B_i, B_{i+1}, B_{i+2}, B_{i+3}, B_{i+4}]$ bearbeitet wurde (Paradigma von B_i). Sind jetzt z. B. B_{i-1} und B_{i+1} in einer hinreichend großen Zahl von Fällen Substantive und ist B_i als Menge von Graphemen einelementig, so kann mit einiger Sicherheit geschlossen werden, daß B_i ein Genitivsuffix ist.

Die Arbeiten, die noch nicht völlig abgeschlossen sind, bewiesen, daß die Kitansprache – wie allerdings auch schon vorher angenommen wurde – zu den agglutinierenden Sprachen gehört. Die Entzifferung einiger Blöcke erlaubt zudem die einigermaßen sichere Aussage über die Zugehörigkeit der Kitansprache zur mongolischen Gruppe. Eingehend berichten über diese Arbeiten V. S. Starikov und V. M. Nadeljaev (Предварительное сообщение о дешифровке киданьского письма) und M. A. Probst (О машинной обработке киданьских текстов) in: Предварительное сообщение о дешифровке киданьского письма, Всесоюзный Институт Научной и Технической Информации АН СССР, Институт Этнографии АН СССР, Москва 1964, S. 5–26 bzw. 27–36.

3.2. *John R. Krueger, Mongolian Epigraphical Dictionary in Reverse Listing, Indiana University Publications, Uralic and Altaic Series, Volume 88, Mouton & Co., The Hague 1967.*

Dieses mit Hilfe eines mittelschnellen CD-Computers von John R. Krueger hergestellte rückläufige Wörterbuch enthält 24.561

Eintragungen, von denen ca. 99 % aus Lessings Mongolian-English Dictionary, California, 1960, die restlichen aus weiteren 8 Wörterbüchern und Wortlisten zusammengetragen wurden. Die Eintragungen sind mit einer genauen Quellenangabe versehen, längenverschoben und übersichtlich angeordnet. Dieses rückläufige Wörterbuch hat neben seiner epigraphischen Bedeutung zweifellos den Charakter einer Fundgrube für jede morphologisch orientierte Arbeit über die mongolische Sprache.

3.3. *Igor de Rachewiltz, Index Verborum zur Geheimen Geschichte der Mongolen*

Bei diesem Index Verborum handelt es sich um ein nicht-kanonisches Wörterbuch mit vorlaufender Anordnung der Lemmata und genauen Quellenangaben. Zugrunde lag der von Pelliot rekonstruierte Text der Geheimen Geschichte, in den die neueren Forschungsergebnisse von Cleaves, Mostaert, Poppe, Ligeti und Igor de Rachewiltz selbst eingearbeitet wurden. Der Index Verborum wurde mit einem mittelschellen Computer vom Typ IBM 360/50 hergestellt. Er wird gegenwärtig zusammen mit dem neu überarbeiteten Text der Geheimen Geschichte für den Druck vorbereitet.

3.4. *Vietze, Deutsch-mongolisches Vokabelverzeichnis zum Lehrbuch der mongolischen Sprache*

Das Vokabelverzeichnis enthält 2700 Eintragungen, wobei entsprechend dem Charakter des Lehrbuches dem ökonomischen und geologischen Wortschatz besondere Aufmerksamkeit zukam. Es wurde mit BULL- und Soemtron-Maschinen hergestellt und erschien als Anhang des Lehrbuches 1970 beim VEB Verlag Enzyklopädie Leipzig.

3.5. *Vietze, Zeuner, Mater, Rückläufiges Wörterbuch zu Mangᡤol un Niuca Tobca'an (Geheime Geschichte der Mongolen)*

Dieses rückläufige Wörterbuch wurde im Juni 1967 mit Bull-Maschinen fertiggestellt. Es erschien 1969 beim VEB Verlag Enzyklopädie Leipzig. Zugrunde lag Haenischs Wörterbuch zu Mangᡤol un Niuca Tobca'an, Wiesbaden 1962. Das rückläufige Wörterbuch enthält 5096 längenverschobene Eintragungen, eine

zusätzliche Liste der Eigennamen und geographische Bezeichnungen, eine Wortlängenstatistik und Hochrechnungen der in letzter und vorletzter Position gleichen Buchstaben.

4. Begonnene und geplante Automatenarbeiten auf dem Gebiet der Altaistik

4.1. *Entzifferung der Kitanschrift*

Den zitierten Veröffentlichungen von Starikov, Nadeljaev und Probst sowie einem Vortrag von Starikov, den er 1967 in Berlin hielt, war zu entnehmen, daß die Arbeiten an der Entzifferung der Kitanschrift fortgesetzt werden. Über den gegenwärtigen Stand der Arbeiten ist mir jedoch nichts bekannt.

4.2. *Oiratisches Datenverarbeitungsprojekt*

Dieses Projekt wurde von John R. Krueger in dem oben zitierten Aufsatz angekündigt. Den Kern wird ein oiratisches Wörterbuch bilden, wobei die für dieses Wörterbuch zu speichernden Daten im folgenden nach anderen Programmen weiter abgearbeitet werden können. So können z.B. auf dieser Grundlage ein mongolisch-oiratisches, ein englisch-oiratisches Glossar, verschiedene Distributionsuntersuchungen u.a. Arbeiten maschinell hergestellt werden.

4.3. *Datenverarbeitungsprojekt zur modernen mongolischen Sprache*

Dieses Projekt wird von einem Kollektiv an der Sektion Asienwissenschaften der Humboldt-Universität zu Berlin durchgeführt. Es ist ebenfalls effektiv, da drei umfangreiche Arbeiten mit Hilfe ein und desselben Lochkartenmaterials hergestellt werden. Im einzelnen handelt es sich dabei um

- a. ein deutsch-mongolisches Wörterbuch, das wir durch Stürzen des vorhandenen mongolisch-deutschen Wörterbuches⁹ und unter Verwendung des deutsch-mongolischen Gesprächsbuches¹⁰ erhalten (vgl. Abschnitt 2.3.),

⁹ Zebek, Schubert, Mongolisch-deutsches Wörterbuch, VEB Verlag Enzyklopädie Leipzig, 1961.

¹⁰ Vietze, Deutsch-mongolisches Gesprächsbuch, VEB Verlag Enzyklopädie Leipzig, 1963.

- b. ein rückläufiges Wörterbuch der modernen mongolischen Sprache,
- c. eine Untersuchung der mongolischen Graphemdistribution.

Stand der Arbeit: Speicherung der Daten auf Lochstreifen wurde begonnen.

4.4. *Klassifizierung der Tabu-Register des Osmanischen Reiches (Mufaşşal und İğmāl) und später der ğizye defterleri*

Dieses Projekt wird von Prof. Gustav Bayerle, Bloomington, geplant und vorbereitet. Es hat einen detaillierten Überblick über alle existierenden o.a. Register und verschiedene Formen der Abarbeitung nach ökonomischen und demographischen Gesichtspunkten zum Ziel.

5. Grundsätzliches

- a. In allen Wissenschaftsbereichen, auch in den Gesellschaftswissenschaften ist schon seit einigen Jahren die ausgeprägte Tendenz von Integration, Mathematisierung und Verwendung theoretischer und praktischer kybernetischer Verfahren vorhanden. Die Altaistik hat begonnen, sich in diesen Prozeß einzufügen.
- b. Da praktische Automatenarbeiten sehr schnell und mit einem in kurzer Zeit anfallenden relativ großen finanziellen Aufwand durchgeführt werden, ist es erforderlich, alle geplanten derartigen Arbeiten rechtzeitig anzukündigen und auch über den Stand ihrer Durchführung bzw. über den Entschluß, sie abzuberechnen, rechtzeitig Mitteilung zu machen. Am besten sind dafür die PIAC-Newsletters und die Zeitschrift Computers and the Humanities (veröffentlicht vom Queens College of the City University of New York), die laufend Listen über praktische Automatenarbeiten auf dem Gebiet der Gesellschaftswissenschaften bringt, geeignet.
- c. Aus rein pragmatischen Gründen ist es im Hinblick auf kybernetische Arbeitsmethoden und praktische Automatenarbeiten angebracht, für diese Zwecke (eventuell nur für diese Zwecke) alle Sprachen mit agglutinierender Struktur (finnisch-ugrische, altaische, japanisch, koreanisch) zu einer Gruppe zusammenzufassen.