

FORMALISIERUNG DER ÄGYPTOLOGISCHEN LEXIKOGRAPHIE
ALS VORAUSSETZUNG UND ALS ERGEBNIS
DES EINSATZES VON EDV-ANLAGEN

von Wolfgang Schenkel

1. Formalisierung und EDV-Anlage: Methode versus Verfahren

Die Notwendigkeit der Formalisierung linguistischer Beschreibungen ergibt sich aus wissenschaftstheoretischen und methodologischen Ansätzen; sie besteht unabhängig von den arbeitstechnischen Verfahren, die für die Lösung einzelner Probleme zur Verfügung stehen und unter denen die EDV-Verfahren unter Umständen besonders effektiv sind. Es besteht zwischen der Forderung nach Formalisierung und dem Einsatz von EDV-Anlagen allerdings insofern ein Zusammenhang, als EDV-Anlagen ein speziell geeignetes Instrument für die Lösung formalisierter Problemstellungen sind und in manchen Fällen überhaupt erst die praktische Lösung einer Problemstellung gestatten. Man darf noch einen Schritt weiter gehen: manche schwierige Problemstellung wird erst dadurch angestoßen, daß EDV-Anlagen zu ihrer Lösung bereitstehen. Trotzdem bleibt der grundsätzliche Vorrang der formalisierten Problemstellung vor den arbeitstechnischen Verfahren bestehen.

EDV-Anlagen sind geeignet für algorithmische Arbeitsweise; sie sind ungeeignet für assoziative Arbeitsweise, die höchstens mit Aufwand simuliert werden kann. Ihr Einsatz lohnt sich besonders bei solchen linguistischen und philologischen Problemstellungen, die entweder mit einem komplizierten Regelwerk gelöst werden und/oder mit der Bewältigung großer Datenmengen verbunden sind.

1.1 Historischer Exkurs

Die Fakten des Darmstädter Projektes M.A.A.T., in dem bis jetzt mit dem größten Aufwand die Anwendung von EDV-Verfahren auf die ägyptologische Lexikographie vorbereitet und erprobt wurde [4, 6, 1], ebenso die Fakten über die Göttinger Anschlußarbeiten [5] sind weitestgehend zugänglich gemacht worden. Wenn hier trotzdem noch einmal aus der Rückschau auf das Darmstädter Unternehmen eingegangen wird, so deshalb, weil man nach Abschluß der ersten Etappe im zeitlichen Abstand heute vielleicht schon abschätzen kann, in welches Verhältnis dort EDV und Ägyptologie traten, und damit Verfahren und Methode.

Vorarbeiten des Projektes M.A.A.T. waren um das Jahr 1963 entstandene Versuche, unter disparaten strukturalistischen Einflüssen, teilweise auch solcher aus der heute weithin verfemten Weisgerber-Schule, die Beschreibung der altägyptischen Syntax zu formalisieren. Es wurde sehr schnell auch die Möglichkeit realisiert, zum Zwecke solcher Untersuchungen technische Hilfsmittel wie Randlochkarten oder konventionelle Hollerithlochkartenverfahren zu benutzen. Der Einsatz von EDV-Anlagen in dem Umfang, wie er kommen sollte, wäre von der Ägyptologie aus damals nicht in den Bereich des Möglichen gerechnet worden, eher sogar - wenn das machbar gewesen wäre - verhindert worden.

Daß EDV-Anlagen dann doch im Rahmen der Untersuchungen eine große Rolle zu spielen begannen, lag an einmaligen, kaum wiederholbaren Umständen: Die EDV war in Deutschland im Bereich der nichtnumerischen Anwendungen um das Jahr 1965 gerade in eine Entwicklungsstufe eingetreten, auf der es möglich schien, philologische und linguistische Untersuchungen größeren Umfangs mit Erfolg durchzuführen. Es gab in den EDV-Anlagen der IBM 7000-Familie eine für Großversuche vorzüglich geeignete Hardware, und man hatte damals eine nichtnumerische Software so weit entwickelt, daß praktische Problemlösungen in greifbarer Nähe lagen. Die EDV, die ja ihre Bewährungsprobe in diesem Bereich erst noch zu bestehen hatte, war infolgedessen stark

daran interessiert, praktische Probleme aus den verschiedensten Fachdisziplinen zum Nachweis ihrer Leistungsfähigkeit und zur Exploration weiterer Anwendungsgebiete zu lösen. Durch einige Zufälle kamen damals die ägyptologischen Formalisierungsversuche in Kontakt mit der EDV, und so wurde als ein Projekt der nichtnumerischen Datenverarbeitung unter zahlreichen anderen Versuchen das Unternehmen M.A.A.T. gefördert. Daß unter diesen Umständen der Akzent auf die EDV-Verfahren verlegt werden mußte, versteht sich von selbst.

Ohne die Anstöße aus der EDV wäre die Entwicklung der Formalisierungsversuche wohl anders verlaufen. Da während der langwierigen Programmierungsarbeiten nicht ständig das Konzept geändert werden konnte, mußte die Formalisierung schon sehr bald, etwa im Jahre 1966, abgeschlossen werden. Unter anderen Umständen hätte sie konsequent fortgesetzt werden können. Nachdem nun die Entwicklung der EDV-Verfahren etwa 1969 oder spätestens 1970 zu einem ersten vorläufigen Abschluß gekommen ist, steht jetzt um so drängender das Formalisierungsproblem an, das heute nach den Wandlungen der Linguistik im letzten Jahrzehnt weit höhere Ansprüche an den Bearbeiter stellt als noch zu Beginn der 60er Jahre. Die EDV-Verfahren könnten nach den Investitionen der letzten Jahre dabei ein nützliches Instrumentarium sein.

2. EDV-anwendbares und EDV-erstelltes Lexikon

2.1 Linguistische Datenverarbeitung und Ägyptologie

Die Lage, in der sich die Ägyptologie hinsichtlich der Erstellung und Anwendung von Lexika befindet, ist verschieden vom Standardfall der "linguistischen Datenverarbeitung" (LDV; der entsprechende englische Ausdruck, "Computational Linguistics", drückt das Verhältnis Linguistik-EDV wohl zutreffender aus). Der Unterschied ergibt sich daraus, daß sich die LDV in der Regel mit relativ gut aufgearbeiteten, meist modernen Sprachen beschäftigt, während das Altägyptische in die lange Reihe der

schlecht erschlossenen Sprachen gehört. Worin sich die Lage der Ägyptologie unterscheidet, zeigt sich am besten im Kontrast zu einem elementaren Beispiel aus der LDV.

2.2 Ein Beispiel aus der LDV

Die LDV beschäftigt sich unter anderem mit der Analyse von Texten gemäß den Methoden der verschiedenen taxonomischen Grammatikmodelle. Die Aufgabe besteht darin, den laufenden Text in linguistisch relevante Abschnitte zu segmentieren (z. B. in Phoneme, Morpheme, Lexeme, Sätze) und die so gewonnenen Segmente zu klassifizieren (z. B. die Verben gemäß ihrer "Valenz" oder die Sätze nach ihrer Phrasenstruktur).

Betrachten wir näher die Zuordnung einer Strukturbeschreibung zu einem Satz, ein Arbeitsgang, der als "Parsing" bezeichnet wird! Im Falle einer kontextfreien und projektiven Grammatik kann dies mit Hilfe eines Parsing-Algorithmus erreicht werden, der folgendes enthält [2, 3]:

- (1) ein Lexikon, das für jedes im Satz vorkommende Segment einen Eintrag enthält, in dem seine Merkmale beschrieben sind (z. B. bei "Wörtern" die Wortart, das Genus, die Valenz);
- (2) eine Grammatik, die angibt, welche Merkmale zwei benachbarte Segmente (oder bereits früher während des Parsing-Vorgangs bestimmte Syntagmen) haben müssen oder dürfen bzw. nicht haben müssen oder dürfen, damit sie zu einem Syntagma mit wiederum ganz bestimmten Merkmalen zusammengefaßt werden können;
- (3) eine Parsing-Strategie, die festlegt, in welcher Reihenfolge die Kombinationsfähigkeit benachbarter Segmente geprüft werden soll. Beispielsweise kann man zunächst die gültigen Kombinationen je zweier benachbarter Segmente suchen, dann die Kombinationen der gefundenen Zweierkombinationen mit einem dritten, benachbarten Segment, usw., bis man zuletzt die Kombinationen sucht, die sämtliche Segmente des Satzes umfassen.

Es ist ohne weiteres einsichtig, daß eine Grammatik und ein Lexikon für eine solche Parsing-Routine eine sehr hohe Anzahl von Grammatikregeln bzw. ein sehr umfangreiches Wörterbuch mit einer beträchtlichen Anzahl von grammatischen und semantischen Merkmalen erfordern.

2.3 Verfahren M.A.A.T.

Es ist klar, daß die Ägyptologie bei der Bruchstückhaftigkeit der Sprachkenntnisse, besonders beim derzeitigen Stand der Kenntnis des Wortschatzes, nicht in der Lage ist, ein entsprechendes umfassendes Lexikon zu erstellen. Selbst wenn man die immense Arbeit investieren sollte, die ein solches Lexikon kostet, wäre über kurz oder lang ein guter Teil der Arbeit verfallen, weil höchst wahrscheinlich im Fortgang der Arbeit die Grammatik geändert würde und dadurch ganz andere Merkmale beim größten Teil der Lexikoneinträge eingesetzt werden müßten.

Um der Mißlichkeit der Erstellung eines umfassenden Wörterbuchs zu entgehen, arbeitet das Verfahren M.A.A.T. weitgehend ohne Lexikon. Ein Lexikon wird benutzt für die Zurückführung von Lautvarianten auf eine Normform (z. B. wird sdm "hören" in s_dm normiert); hier reicht für das klassische Altägyptisch ein Lexikon von höchstens wenigen Tausend Einträgen einfachsten Aufbaus (die Anzahl der Gleichungen liegt im System M.A.A.T. derzeit immer noch unter tausend); ein solches Lexikon herzustellen ist auch nicht riskant, weil man in diesem Elementarbereich der Lautformen ausreichend Bescheid weiß.

Dagegen werden Merkmale, besonders solche zur Auflösung von Mehrdeutigkeiten, den Wortformen bzw. Syntagmen explizit zugeordnet, in einem Verfahren, das man Präedition nennt. Der Vorteil des Verfahrens liegt darin, daß man nicht in einem Lexikon und einer Grammatik alle Fälle der Kollokation verschiedener Wortformen bzw. Syntagmen im voraus bedacht haben muß. Der Nachteil ist der, daß man immer wieder, besonders bei häufigen Wortformen und syntaktischen Relationen, die gleichen präeditorischen Vermerke dem Text zufügen muß.

2.4 Erstellung eines Lexikons auf der Grundlage von prä- edierten Texten

Ein Lexikon von der Art der in der LDV manuell erstellten Lexika als Vorarbeit für die maschinelle Bearbeitung altägyptischer Texte liegt nicht im Bereich des gegenwärtig Möglichen. Es besteht jedoch umgekehrt die Möglichkeit, nach und nach ein solches Lexikon durch Auswertung der explizit zugeordneten Merkmale präedierter Texte aufzubauen. Das Lexikon könnte als Nebenprodukt der maschinellen Konkordanz-Erstellung entstehen. Sobald genügend viele präediierte Texte für den Aufbau des Lexikons zur Verfügung stehen, kann man dann auch versuchen, dieses Lexikon bei der maschinellen Texterschließung selbst wieder in der Art eines LDV-Lexikons einzusetzen. Texterschließung und Lexikonerstellung würden also rückgekoppelt und würden sich gegenseitig "aufschaukeln". Dabei ist von vornherein klar, daß - wenigstens in absehbarer Zeit - kein vollständiges Lexikon erreicht werden kann und daß daher das Verfahren der Texterschließung halbautomatisch bleiben muß: neu auftretende Fälle müssen manuell, durch Präedition oder durch Eingriff in den Ablauf der maschinellen Texterschließung, behandelt werden.

3. Eigenschaften des Lexikons

Wie kann ein solches EDV-erstelltes bzw. EDV-anwendbares Lexikon aussehen? Es können hier nicht die Einzelheiten des Aufbaus spezifiziert werden, weil dafür einerseits kein fertiges Konzept vorliegt und weil es zum anderen gerade Aufgabe dieser Tagung ist, überhaupt die Möglichkeiten der ägyptologischen Lexikographie durch Aufzeigen der verschiedenartigsten Aspekte ansatzweise zu formulieren. Es dürfte jedoch nützlich sein, die Möglichkeiten für den Aufbau eines Lexikons, wie sie im bisherigen Konzept des Projektes M.A.A.T. angelegt sind, an der lexikalischen Komponente des Sprachmodells der generativen Transformationsgrammatik zu messen.

Im System M.A.A.T. sind lexikalische Informationen an folgenden Stellen greifbar:

- (1) in Hilfslisten, von denen an erster Stelle die Indexlisten zu nennen sind,
- (2) in den präeditorischen Zusätzen zu systemgerecht aufgenommenen Texten,
- (3) im Nachweis der Distribution der Elemente in den durch M.A.A.T. erstellten Konkordanzen.

Im einzelnen verhalten sich die Leistungen der in M.A.A.T. bereitgestellten Angaben zu den linguistischen Anforderungen wie folgt:

(1) Phonologische Komponente:

Jedes Wort des altägyptischen Lexikons wird im System M.A.A.T. durch eine Folge von Phonemen, Strukturzeichen und Indizes eindeutig bezeichnet (Wurzel, Wurzelindex, Wurzeltyp, Wort, Wortindex). Der Ansatz ist in den Indexlisten kodifiziert. Homokonsonantische Wurzeln bzw. Wörter sind also explizit durch Indizes verschiedener Art von einander unterschieden.

(2) Syntaktische Komponente:

(2.1) Kategorien:

Grammatische Kategorien (Wortarten) sind über die präeditorischen Zusätze für das Lexikon abrufbar.

(2.2) Merkmale (Subkategorisierung):

Einige syntaktische Merkmale sind über die präeditorischen Zusätze für das Lexikon abrufbar (Eigennamen, Numerus, Verbalgenus) oder zusätzlich aus den Indexlisten zu entnehmen (semantischer Index), andere ergeben sich aus der maschinellen Analyse in den Konkordanzen (z. B. die vollständigen morphologischen Merkmale). Merkmale, wie sie in der generativen Transformationsgrammatik benutzt werden wie [[±]belebt], [[±]abzählbar], [[±]menschlich], [[±]abstrakt] usw. dürften sich anhand der in den Konkordanzen nachgewiesenen Distributionen manuell bestimmen lassen.

(2.3) Selektionsbeschränkungen:

Selektionsbeschränkungen, z. B. die Valenz der Verben oder die Kombinierbarkeit von Nomina in Nominalphrasen, können weitgehend aus den Distributionslisten gewonnen werden, wie sie vor allem für die - noch nicht realisierte - syntaktische Stufe des Verfahrens M.A.A.T. vorgesehen sind.

4. Zweck des Lexikons

Der Zweck eines derart konstruierten Lexikons ist ein dreifacher:

(1) EDV-Anwendung:

Von geringstem Interesse ist die Möglichkeit, damit das Verfahren der maschinellen Texterschließung stärker zu automatisieren, da hier einerseits nicht sehr rasch entscheidende und langanhaltende Fortschritte erzielt werden können und weil die Frage der Anwendung von EDV-Anlagen eine Frage der Arbeitstechnik und nicht eine der wissenschaftlichen Methodik und der wissenschaftlichen Ergebnisse ist.

(2) Philologie:

Es dient als Grundlage für die philologische, sachorientierte Arbeit an den Texten. Es könnten auf der Grundlage eines solchen umfänglichen Lexikons leichter Mehrdeutigkeiten erkannt und aufgelöst werden, weil das Lexikon beispielsweise im Idealfall sämtliche Selektionsbeschränkungen angibt und daher Mittel zur Verfügung stellt, auf Grund linguistischer Argumentation aus dem Kontext heraus scheinbar mögliche Lösungen wieder zu eliminieren.

(3) Linguistik:

Den hauptsächlichlichen Gewinn zöge die Linguistik. Eine linguistisch akzeptable Beschreibung des altägyptischen Lexikons ist absolute Voraussetzung dafür, daß man im Sinne einer interdisziplinären Zusammenarbeit die Linguisten für das Altägyptische interessieren könnte. Damit könnte man das reiche Material der Ägyptologie insbesondere für die

diachronische Linguistik fruchtbar machen, zu der die Ägyptologie in besonderem Maß beitragen können sollte, da sie eine der längsten, vielleicht die längste an Textzeugnissen verfolgbare Sprachgeschichte überblickt, die es bislang in der Geschichte der Menschheit überhaupt gibt, eine Sprachgeschichte, die zudem zum mindesten streckenweise weitgehend unbeeinflusst von Nachbarsprachen verläuft und damit leichter als anderswo die Frage nach dem Zusammenhang von Sprache und Gesellschaft bzw. Sprachgeschichte und Geschichte der Gesellschaft beantworten kann. Dies ist aber nur unter zwei Voraussetzungen möglich:

- (3.1) daß die Ägyptologie durch Übernahme der Methoden der heutigen Linguistik zuerst einmal den Boden für eine interdisziplinäre Verständigung bereitet,
- (3.2) daß die Ägyptologie ihr Material sehr viel genauer aufbereitet, als es ihr bisher möglich war - wozu sie sich heute die Hilfsmittel der EDV zunutze machen kann.

5. Literatur

- [1] R. Gundlach und W. Schenkel, Lexikalisch-grammatische Liste zu Spruch 335a der altägyptischen Sargtexte LL/CT.335A, Schriftenreihe des Deutschen Rechenzentrums, Heft S-8, Darmstadt 1970.
- [2] D. G. Hays, Introduction to Computational Linguistics, New York 1967.
- [3] W. Klein, Parsing, Linguistische Forschungen 2, Frankfurt 1971.
- [4] W. Schenkel, Maschinelle Analyse Altägyptischer Texte, Schriftenreihe des Deutschen Rechenzentrums, Heft S-7, Darmstadt 1969.

[5] W. Schenkel, "Zur Fortführung des Projektes M.A.A.T. in Göttingen", GM 2 (1972), 33 - 36.

[6] W. Schenkel, "Texterschließung mit Hilfe des Systems M.A.A.T.", in: A. Schwab-Schlott (Hrsg.), Dokumentation ägyptischer Altertümer, Darmstadt 1970, 82 - 90.

4. Sprache und Gesellschaft bzw. Sprachgeschichte und Ge-
 schichte der Gesellschaft beantwortet kann. Dies ist aber
 nur unter zwei Voraussetzungen möglich:
 (2.1) daß die Ägyptologie durch Übernahme der Methoden der
 heutigen Linguistik zuerst einmal den Boden für eine
 interdisziplinäre Verständigung bereitet
 (2.2) daß die Ägyptologie ihr Material sehr viel besser
 aufbereitet als es ihr bisher möglich war - was als
 ein Hauptziel der BZV angesehen werden kann.
 5. Literatur
 1) R. Gundlach und W. Schenkel, Lexikalisch-grammatische Liste
 zu Buch 325 der ägyptischen Papyri, Teil 3-8,
 Schöpfung des Deutschen Researchrates, Heft 3-8,
 Darmstadt 1970.
 2) J. A. Hays, Introduction to Computational Linguistics, New
 York 1967.
 3) W. Klein, Parallele linguistische Forschungen 2, Frankfurt
 1971.
 4) W. Schenkel, Maschinelle Analyse Ägyptischer Texte,
 Schöpfung des Deutschen Researchrates, Heft 3-1,
 Darmstadt 1969.
 5) W. Schenkel, Die ägyptische Hieroglyphen, Teil 1, 1-10,
 Göttingen 1972.