

Wolfgang Schenkel :

## Texterschließung mit Hilfe des Systems M. A. A. T. Überblick über die verfügbaren Materialien

### 1 Einleitung: Stufen und Ablauf der Texterschließung

Das System M. A. A. T. (Maschinelle Analyse Altägyptischer Texte) ist ein System der integrierten Texterschließung: zum Zweck einer intensiven und vielseitigen Erfassung der Merkmale von Texten ist eine Reihe von Frageebenen - so etwa die lexikalische, die grammatische und die semantische - derart ineinander verzahnt, daß die Ergebnisse aus einer Ebene als Grundlage für die Bearbeitung anderer Ebenen benutzt werden können, zum Beispiel die morphologischen Analyseergebnisse zur Bestimmung der Lemmata im Rahmen der lexikographischen Auswertung. Zur Zeit sind die Stufen der lexikalischen und morphologischen Texterschließung ausgebaut, die der syntaktischen oder inhaltlichen Texterschließung in Grundzügen angelegt.

Der Ablauf einer Texterschließung sei am Beispiel der lexikalisch-grammatischen Texterschließung für eine lexikalisch-grammatische Konkordanz veranschaulicht (siehe Fig. 25). Nach der Erstellung der als Grundlage für die Analyse dienenden Flexionstabellen, Glossare usw. (siehe Abschnitt 2) werden die auf Lochkarten stehenden Texte wortformweise analysiert und die lexikographischen Einheiten, das heißt die zu einem Wortformbeleg gehörigen lexikalischen und grammatischen Merkmale, zusammengestellt. Anschließend werden die lexikographischen Einheiten nach mehreren hierarchisch geordneten Merkmalen sortiert, so nach den Wurzeln, nach den Zitierformen der von den Wurzeln abgeleiteten Wörter, nach den Wortarten, nach den Flexionsformen und anderem mehr. Zuletzt werden die Belege in allen Stufen der Hierarchie ausgezählt, gleiche Angaben zu einer einzigen vereinigt, Übersetzungen und Literaturverweise eingefügt und die so gewonnene Konkordanz in übersichtlicher Form ausgedruckt.

Die Verarbeitung wird durch Programme gesteuert, die in einer maschinenverständlichen Programmiersprache geschrieben sind (siehe Abschnitt 3).

### 2 Der Thesaurus des Systems M. A. A. T.

Als Grundlage der Texterschließung dienen im System M. A. A. T. lexikalische, grammatische, chronologische und geographische Listen, in denen die in der maschinellen Analyse zu behandelnden Merkmale der Texte festgelegt sind. Die Listen werden unter dem Namen Thesaurus zusammengefaßt, wobei dieser Begriff einerseits auf den Charakter der Listen als Schatzkammer für die verschiedensten maschinellen und konventionellen Zwecke, andererseits auf ihren systematischen Aufbau hinweisen soll.

Die größte Liste des Thesaurus ist die Wortsammlung für größere Glossare. Sie enthält zur Zeit zu ca. 10 000 altägyptischen Wurzeln und Wörtern die festgelegten Lautformen, Indizes zur Unterscheidung von Homographen und von semantischen Gruppen, Angaben zum Wurzeltyp, pauschale Übersetzungen, Literaturverweise und Wurzelgleichungen (zum Beispiel  $s\dot{a}m_{51} = s\dot{a}m_j$  "hören"). Die Wortsammlung wird für den Handgebrauch in drei Listen ausgedruckt: einem nach Wurzeln und Wörtern geordneten Verzeichnis (siehe den - im einzelnen überholten - Ausschnitt in [1], Abb. 2), einem Umwandlungsglossar mit den Wurzelgleichungen und einem Verzeichnis nach Literaturverweisen. Für die maschinelle Anwendung werden aus der Wortsammlung ein Wurzelvariantenglossar und ein Übersetzungs- und Literaturverweisglossar angefertigt.

Daneben gibt es als selbständige Sammlungen kleinere Spezialglossare, zum Beispiel ein Glossar mit allen "Prospektiv"-stämmen auf  $t$  ( $jnt$  "bringen",  $jwt$  "kommen"). Zur Zeit sind ca. 25 solche Glossare vorhanden.

Für die Codierung chronologischer und geographischer Angaben dienen das chronologische und das geographische Glossar (zum Prinzip der Codierung siehe das Referat von A. Schwab-Schlott).

Die Flexionsformanalyse stützt sich auf 40 Flexionstabellen, die neben dem Code für die Flexionsform im allgemeinen die Textform, die als Norm angesetzte Form und eine Klartextbeschreibung der Form (Flexionsangabe) enthalten (siehe die Flexionsformen der Suffixkonjugation in Fig. 26).

### 3 Die grammatischen Algorithmen des Systems M. A. A. T

Für die morphologische Analyse der Wortformen sind präzise Verarbeitungsregeln, Algorithmen, ausgearbeitet. Als Beispiel ist in Fig. 27 ein Algorithmus zur Bestimmung der Formen der Suffixkonjugation wiedergegeben. Er wird in einem Stadium des Analyseprozesses angewandt, in dem bereits unter anderem folgende Merkmale der Wortform bekannt sind:

- die Endung;
- der Wurzeltyp (L = leer, merkmallos; I = ult.  $j$ ; w = ult.  $w$ ; G = ult. gem.; D = denominal);
- der Stammtyp (L = leer, merkmallos; K = Kurzform, ohne letzten schwachen Radikal; G = geminiert; P, T, N = Spezialformen);
- fallweise das Verbalgenus (Q = Passiv).

Der Algorithmus arbeitet mit den in Fig. 26 wiedergegebenen Flexionstabellen (siehe Abschnitt 2).

#### 4 Die philologischen Textaufnahmen

Nach den Vorschriften des Systems M. A. A. T. ist bis jetzt hauptsächlich der Spruch 335 a der altägyptischen Sargtexte mit allen Varianten [3] aufgenommen worden (siehe das Beispiel in [2], Abb. 1). Eine Konkordanz zu diesem Text ist hergestellt und wird nach einer abschließenden Revision demnächst publiziert. Der Text ist auf ca. 22 000 Lochkarten abgelocht und stellt ca. 4 % des Sargtextmaterials dar. Weiter befinden sich historische und biographische Texte der Spätzeit in Arbeit.

#### 5 Ausblick: Pläne für die Fortsetzung der Arbeit

Das System M. A. A. T. ist für die lexikalische Texterschließung einsatzbereit. Es wird angestrebt, die Aufnahme der Sargtexte auf das gesamte Korpus der Sargtexte auszuweiten. Da eine solche Konkordanz wegen der ständigen Entwicklung in der elektronischen Datenverarbeitung und aus dem Wunsch heraus, bald das Gesamtmaterial leicht benützen zu können, rasch erarbeitet werden sollte, könnte das Vorhaben am besten von einer gut koordinierten Arbeitsgruppe realisiert werden. Konkordanzen zu anderen Textgruppen (zum Beispiel zu den Pyramidentexten) sollten eigentlich gleichzeitig erstellt werden.

Das System M. A. A. T. wird nach dem vorläufigen Abschluß der Entwicklungsarbeiten am lexikographischen Teil für die speziell morphologische Texterschließung voll ausgebaut. Für die syntaktische Erschließung sind Vorarbeiten geleistet.

#### 6 Literatur

- [1] GUNDLACH, R. und SCHENKEL, W.:  
M. A. A. T., Ein System zur lexikalischen und grammatischen Erschließung altägyptischer Texte mit Hilfe einer elektronischen Datenverarbeitungsanlage (Projektbeschreibung),  
Chonique d'Égypte XLII, 83 (1967), 41 - 64.
- [2] SCHENKEL, W.:  
Der Computer als Hilfsmittel für die lexikalische und grammatische Beschreibung des Altägyptischen. Möglichkeiten und Grenzen,  
Zeitschrift der Deutschen Morgenländischen Gesellschaft, Supplementa I: XVII. Deutscher Orientalistentag vom 21. bis 27. Juli 1968 in Würzburg (1969), 97 - 105.
- [3] DE BUCK, A.:  
The Egyptian Coffin Texts,  
The University of Chicago, Oriental Institute Publications, Chicago  
1935 - 1961, 7 Bde..



## Diskussion

(1) P. DERCHAIN stellt die Frage, ob phraseologische Wendungen durch das System M. A. A. T. erkannt und zusammengeordnet werden können.

Die Möglichkeit besteht im Prinzip, und Vorarbeiten in dieser Richtung sind bereits vorhanden. Eine praktikable Lösung ist allerdings frühestens dann zu erwarten, wenn die maschinelle syntaktische Analyse ausgearbeitet ist und das nicht ägyptologische Problem, das als Zitierverifizierung bekannt ist, eine grundsätzliche Klärung gefunden hat.

(2) W. WESTENDORF wirft die Frage nach der Brauchbarkeit des M. A. A. T. - Transkriptionsalphabets (siehe [1], 61) auf, das gegenwärtig durch den Zeichenvorrat des Kartenlochers IBM 026 und die Standard-Druckkette des Schnelldruckers IBM 1403 bestimmt ist.

Bei der Höhe der Kosten und wegen der Gefahr des schnellen Veraltens durch die rasche Entwicklung auf dem Gebiet der elektronischen Datenverarbeitung wäre die Anfertigung spezieller Geräte - zum Beispiel der Guß einer besonderen Druckkette mit den üblichen Transkriptionszeichen - unwirtschaftlich. Es besteht aber Bereitschaft, eine Druckkette zu beschaffen und zu benutzen, wenn die Mittel dafür zur Verfügung gestellt werden.

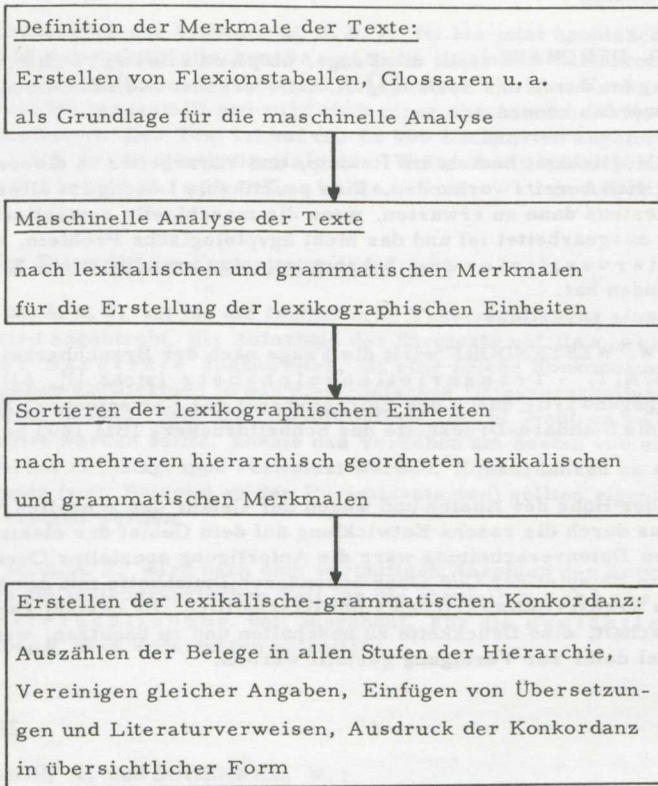


Fig. 25: Der Ablauf der lexikalisch-grammatischen Texterschließung

FLEXIONSTABELLE F-1

SUFFIXKONJUGATION (SK)  
MEHRDEUTIGE ENDUNGEN

CODE	FORM	NORMFORM	FLEXIONSANGABE
-10001			ØL-FORM, AKT.
-30001	TI	TI	ØL-FORM, TW-PASS.
-30003	TW	TI	ØL-FORM, TW-PASS.
+10001			ØK-FORM, AKT.
+30001	TI	TI	ØK-FORM, TW-PASS.
+30003	TW	TI	ØK-FORM, TW-PASS.
210001			ØG-FORM, AKT.
220001			ØG-FORM, PASS.
230001	TI	TI	ØG-FORM, TW-PASS.
230003	TW	TI	ØG-FORM, TW-PASS.
310001			ØS-FORM, AKT.
330001	TI	TI	ØS-FORM, TW-PASS.
330003	TW	TI	ØS-FORM, TW-PASS.
410001	W	W	W-FORM, AKT.
410003	I	W	W-FORM, AKT.
420001	W	W	W-FORM, PASS.
420003	I	W	W-FORM, PASS.

ANZAHL DER FLEXIONSFORMEN 17

Fig. 26: Flexionstabellen für die Suffixkonjugation (Anfang)

FLEXIONSTABELLE F-2

SUFFIXKONJUGATION (SK)  
EINDEUTIGE ENDUNGEN

CODE	FORM	NORMFORM	FLEXIONSANGABE
430001	WTI	WTI	W-FORM, TW-PASS.
430002	ITI	WTI	W-FORM, TW-PASS.
430003	WTW	WTI	W-FORM, TW-PASS.
430004	ITW	WTI	W-FORM, TW-PASS.
510001	N	N	N-FORM, AKT.
510001	NI	NI	N-FORM, AKT., ADV.
510003	NW	NI	N-FORM, AKT., ADV.
530001	NTI	NTI	N-FORM, TW-PASS.
530003	NTW	NTI	N-FORM, TW-PASS.
610001	IN	IN	IN-FORM, AKT.
630001	INTI	INTI	IN-FORM, TW-PASS.
630003	INTW	INTI	IN-FORM, TW-PASS.
710001	XR	XR	XR-FORM, AKT.
730001	XRTI	XRTI	XR-FORM, TW-PASS.
730003	XRTW	XRTI	XR-FORM, TW-PASS.
810001	K)	K)	K)-FORM, AKT.
830001	K)TI	K)TI	K)-FORM, TW-PASS.
830003	K)TW	K)TI	K)-FORM, TW-PASS.
910001	T	T	T-FORM, AKT.

ANZAHL DER FLEXIONSFORMEN 19

Fig. 26: (Ende)



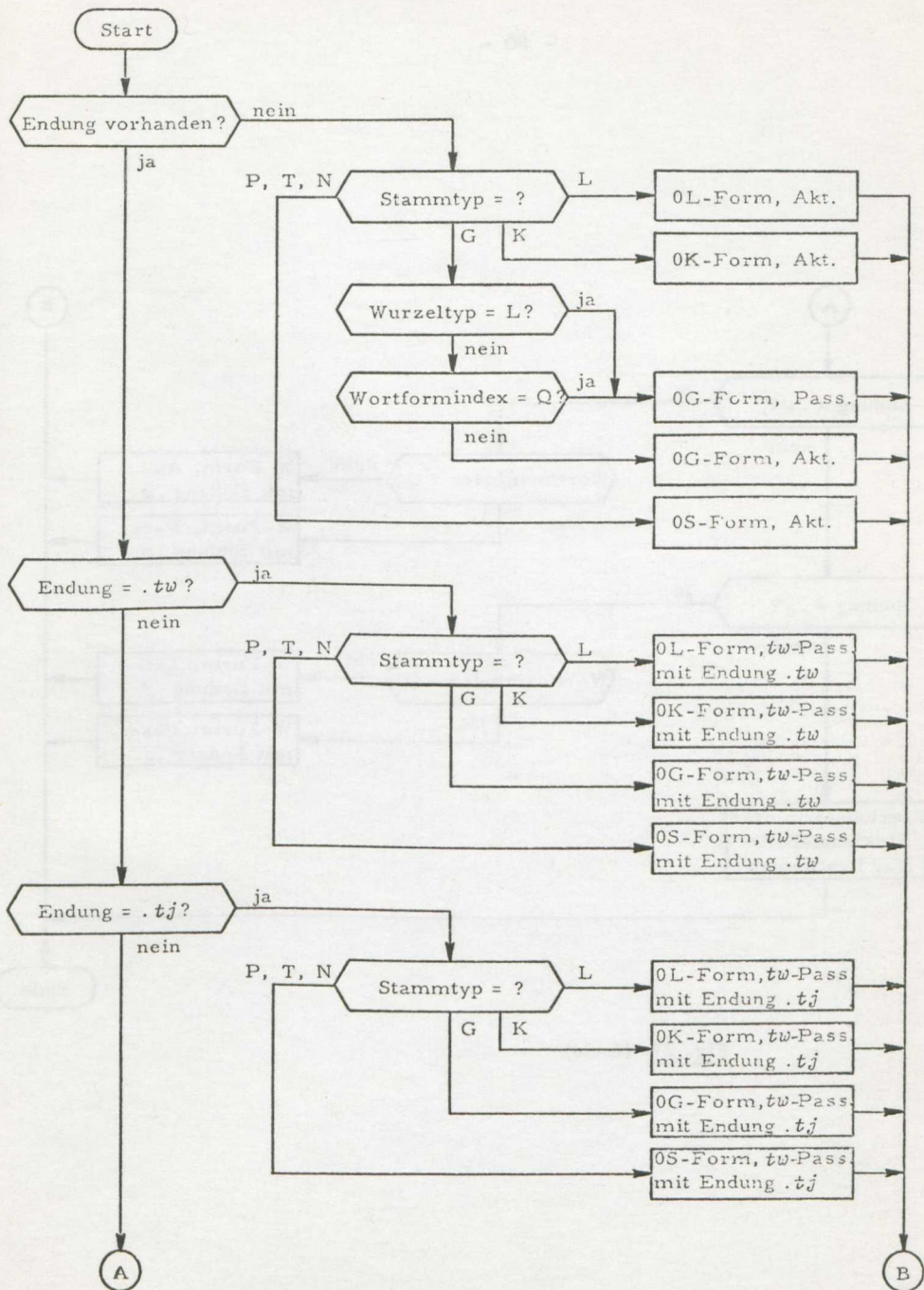


Fig. 27: Algorithmus zur Bestimmung der Formen der Suffixkonjugation (Anfang)



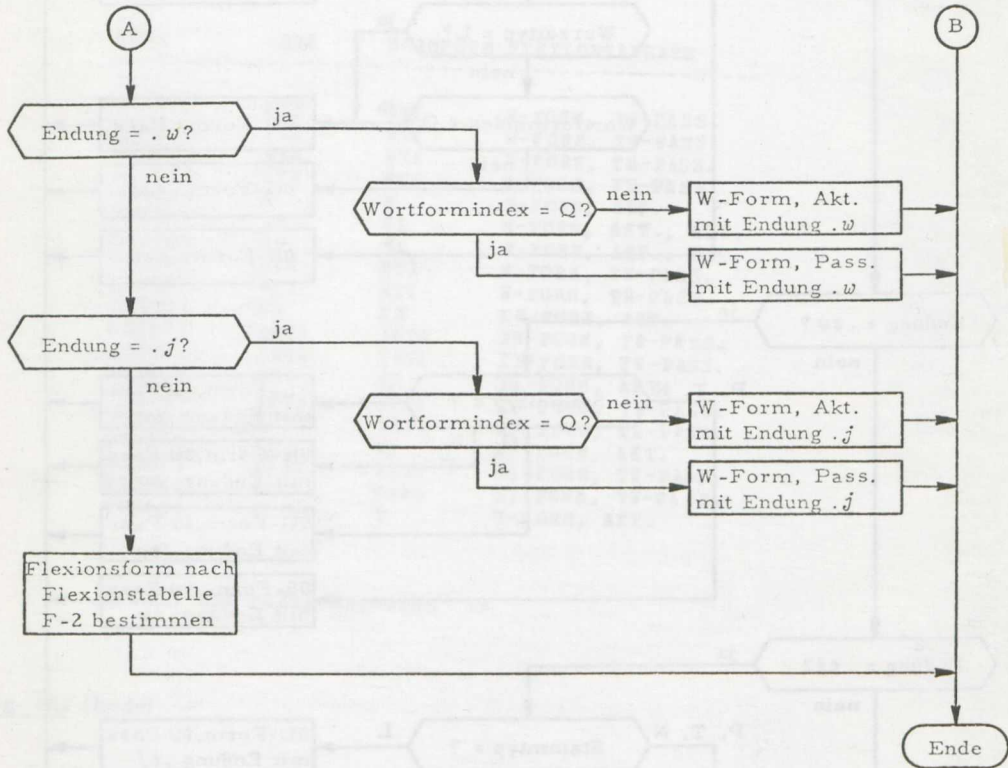


Fig. 27: (Ende)