

Berlin-Brandenburgische Akademie der Wissenschaften

Mitteilungen

aus der Arbeit am

WÖRTERBUCH DER ÄGYPTISCHEN SPRACHE

1

Herausgegeben von

Walter F. Reineke

unter Mitarbeit von

Adelheid Burkhardt, Elke Freier, Stefan Grunert und Ingelore Hafemann
(Mitarbeiter am Akademienvorhaben Altägyptisches Wörterbuch)

Berlin 1993

Die Abkürzungen der ägyptologischen Literatur folgen allgemein den Regeln des *Lexikons der Ägyptologie*. Darauf basiert ein Abkürzungsverzeichnis zur numerischen Codierung aller Daten, das demnächst in der Reihe *Egyptology* des Internationalen Ägyptologen-Verbandes erscheint.

Als Abkürzung der *Mitteilungen aus der Arbeit am Wörterbuch der Ägyptischen Sprache* wird **MittWb** vorgeschlagen.

© 1993 by Altägyptisches Wörterbuch • Berlin
Printed in Germany

Gedruckt mit Unterstützung des Internationalen Ägyptologen-Verbandes

FRITZ HINTZE

Im Andenken
an seine ständige
Förderung und sachkundige Begleitung
der Arbeiten für die Weiterführung des
Wörterbuches
der
Ägyptischen
Sprache



Editorial

Die ersten zwei vorliegenden *Mitteilungen aus der Arbeit am Wörterbuch der Ägyptischen Sprache* sind der Beginn einer in loser Folge erscheinenden Reihe von Berichten und Resultaten der real seit 1992 wiederaufgenommenen Arbeiten am neuen Wörterbuch-Projekt. Diese und auch die weiteren Veröffentlichungen verstehen sich nicht als eine weitere ägyptologische Zeitschrift: Als Verwalter eines sehr umfangreichen Textarchives ägyptischen Sprachgutes sehen wir es als eine für alle lohnende Aufgabe an, die zum Teil einzigartigen Materialien sowohl für die Zukunft international effektiver nutzbar zu machen, als auch als Basis bei der Fortführung der Sammlung ägyptischer Textzeugnisse anzuwenden.

Die gesamte Arbeit wird durch den Einsatz moderner Computertechnik unterstützt. Dabei erfordert der Aufbau einer lexikalischen Datenbank, die mit einer Textdatenbank und einem Textcorpusverwalter verbunden ist, eine Vielzahl von vorbereitenden Arbeiten, die als solche nicht nur für die Berliner Arbeiten an einem rechnergestützten Wörterbuch geeignete Hilfsmittel sind. Sie sollen für eine Veröffentlichung in den *Mitteilungen* aufbereitet und auf diesem Weg einer möglichst großen Zahl von Fachkollegen zur Nutzung angeboten werden. Der Charakter dieser gedruckten Dokumente wird ganz wesentlich durch deren ursprüngliche Funktion bestimmt - es sind tools für die Computerarbeit, die auf Grund ihrer Struktur und Informationsvielfalt ständiger Veränderung und Erweiterung unterworfen sind. Ihrer Menge nach sind sie nur bedingt in Papierformat umsetzbar. Daher können - über die ausgedruckten Hilfsmittel hinaus - auf Anfrage auch die primären Dateien in computergemäßer Form zur Verfügung gestellt werden.

Inhalt

Editorial	5
J. Assmann, S. Grunert, W. F. Reineke Bericht über die Arbeitstagung "Die Zukunft des Altägyptischen Wörterbuchs" vom 3.-5. 9. 1992 in Berlin	7
Jan Assmann Folgerungen für die Zukunft	21
Nigel Strudwick The next generation of the Wörterbuch	25
Walter F. Reineke Das Wörterbuch der Ägyptischen Sprache. - Stand, Probleme, Perspektiven -	31
Silvia Köpstein Das Abklatscharchiv beim "Wörterbuch der ägyptischen Sprache" (Arbeitsbericht)	37
Ingelore Hafemann Lexikalische Datenbank contra Wörterbuch	43
Stefan Grunert Zur Berliner Wortliste	54

Die Berliner Wortliste versteht sich als ein Hilfswerkzeug bei der Eingabe und Analyse neu aufzunehmender Texte in das Textcorpus am Berliner Wörterbuch der ägyptischen Sprache; sie will und kann keine überarbeitete oder erweiterte Fassung des von Adolf Erman und Hermann Grapow im Auftrag der deutschen Akademien herausgegebenen und in den Jahren von 1926 bis 1963 gedruckten Exemplares sein. Allen, die dessen zunehmende Unzulänglichkeiten beklagen, ist bewußt, daß die von Erman und Grapow geleitete Arbeit von mehr als 50 Ägyptologen bisher nicht ersetzt werden konnte; nach wie vor ist es als Standardwerk in der Ägyptologie unabdingbar.

Unter den neuen technischen Bedingungen ist es möglich geworden, den ursprünglichen Gedanken der Schaffung eines zentralen Gesamtkorpus altägyptischer Texte nach mehr als 40 Jahren wieder aufzunehmen und umzusetzen. In der Vergangenheit erfolgte die Sammlung der Texte nach dem Muster des "Thesaurus Linguae Latinae", und in einem über fast zwei Jahrzehnte dauernden Prozeß wurde - das Material auswertend - auf dieser Basis die Lexik der altägyptischen Sprache erforscht. Durch den Einsatz der Rechentechnik muß vom Prinzip des gewohnten, in zigfacher Ausfertigung vorliegenden Wörterbuchzettels abgewichen werden.

Durch das handschriftliche Abschreiben der Hieroglyphen erfolgte früher die Erfassung der Texte relativ originalgetreu - hieratischen Texten, die man in Hieroglyphenschrift umsetzte, wurde in Zweifelsfällen das originale Schriftbild als Kopie zur späteren Entscheidung beigegeben. Eine solche Form der Textspeicherung ist heute qualitativ besser möglich: Die neu zu erfassenden Texte werden - abhängig von ihrer Publikationsform - in ihrem Erscheinungsbild mittels Scanner erfaßt und in dieser, dem Original adäquateren Form in einer separaten Bild-Datenbank unter ihrer Textcorpusnummer abgespeichert.

Zur eigentlichen Erfassung durchlaufen die Texte einen umfangreichen Analyseprozeß. Bisher wurde dieser Ablauf unter Nutzung von zwei unterschiedlichen sog. fullscreen Programmen (PPTES und THOT) getestet, die aus speziellen, nicht mit dem Wörterbuch verbundenen Aufgabenanforderungen entstanden. Das Analyseprogramm THOT verlangt eine durchgängige programmspezifische alphanumerische Hieroglyphencodierung sowie zusätzlich die Angabe der Transkription. Beide Bereiche müssen sich in der Lexemzahl entsprechen. Eine Übersetzung kann angemerkt werden. Umfangreichere Texte können in einzelne Records, vergleichbar den früheren Wörterbuchzetteln, untergliedert werden. In der grammatischen Textanalyse arbeitet THOT mit einem starren System, daß auf der mittellägyptischen Sprachstufe basiert. Jeder Eintrag muß zwangsweise einer durch das System vorgegebenen gram-

matischen Erscheinung zugeordnet werden. Bezüglich der Lexeme verfügt THOT über kein Grundwissen. Erst aus dem verarbeiteten Material wird eine Wortliste erstellt. Dadurch bietet es die Möglichkeit, aus einer bestimmten Textgruppe einen Spezialthesaurus zu erstellen.

Das Analyseprogramm PPTES wurde zur lexikographischen, philologisch-phraseologischen sowie prosopographischen Erschließung von Texten entwickelt. Es dient also primär der Erstellung von Wordindices, Wortkonkordanzen sowie Indices von Spezialworteinträgen wie Personennamen, Götternamen, Geographica. Der Textanalyse liegt eine interne Wortliste in Transkription zugrunde, die aus den Einträgen des Wörterbuches von Erman&Grapow entstand. Anhand dieser linearen Wortliste werden alle Wörter eines Textes lemmatisiert, d.h. standardisiert, indem sie einem Lexem der Wortliste zugeordnet und damit identifiziert werden. Der zu analysierende Text wird nur in einer systemspezifischen Transkription erfaßt, auf der auch die Analyse beruht. Durch eine interne Umsetzung kann dabei zwischen einer englischen und zwei deutschen Transkriptionen gewählt werden. Besondere Schreibungen (wie z.B. von Flexionsendungen) werden in einigen Fällen vom Programm selbst erkannt und dem richtigen Lexem zugeordnet. Meist aber erfolgt die Zuordnung der speziellen Wortform zum Lexem interaktiv. Jedoch kann in den Konkordanzen der Wortkontext mit der aktuellen Wortformschreibung ausgedrückt werden. Allen Einträgen der Wortliste sind bestimmte morphologisch-syntaktische Eigenschaften zugeordnet. Hieroglyphische Schreibungen in alphanumerischer Codierung werden z.Zt. nicht aufgenommen, können aber in einem optionalen Feld ergänzt werden.

Bereits in der ersten Phase der Vorbereitungen zur Wiederaufnahme der Wörterbucharbeit in Berlin war davon ausgegangen worden, das bisherige Wissen zur ägyptischen Lexik im weiteren Arbeitsprozeß produktiv in Form eines sog. Computerinternen Wörterbuches zu nutzen. Die Erfahrungen beim Aufbau einer Datenbank unter dem Betriebssystem OS/2 sowie die späteren bei der Textbearbeitung mittels der erwähnten Analyseprogramme führten zu der Entscheidung, eine lineare Liste der altägyptischen Lexeme aufzubauen, die relational mit einer Vielzahl anderer Informationen zum einzelnen Lexem verbunden ist. Die Diskussionen auf der Berliner Wörterbuchkonferenz (3.-5. September 1992), auf der eine deutliche Polarisierung zur Frage der Notwendigkeit einer alphanumerischen Codierung des hieroglyphischen Erscheinungsbildes vorherrschte, führten schließlich zu dem theoretisch angedachten Kompromiß der Berliner Wortliste. Ausgangspunkt hierfür waren drei grundsätzliche Überlegungen: Einerseits sollte der aufwendige und in der Mehrzahl wenig ertragreiche Prozeß einer vollständigen Codierung der hieroglyphischen Schreibung aller neu zu erfassenden Texte umgangen werden, um entsprechend der mehrheitlich geäußerten Forderung möglichst rasch die Texte erfassen und auswerten zu können. Andererseits sollte wie bisher über die

Transkription stets der Zugang zu den einzig originären Informationen, der hieroglyphischen Schreibung, gewährleistet sein - eine Erfahrung, die sich aus der jahrzehntelangen Praxis in der Arbeit mit dem Wörterbuch-Archiv und der Beantwortung Hunderte von Anfragen ergab. Schließlich blieb ein technisches Problem zu lösen, daß auch in der Archivarbeit bereits eine Rolle gespielt hat. Bei einer Vielzahl von Lexemen kam es zu Änderungen bei der Lesung, jedoch konnte und wurde das Archiv-Material nicht permanent umsortiert. Ähnlich problematisch würde es werden, wenn auch im Bereich eines rechnergestützten Wörterbuches oder einer Textdatenbank im Computer ausschließlich die Transkription als Primärinformation zu verarbeiten wäre. Änderungen in der Lesung hieroglyphischer Lexeme würden - bei unterschiedlichen Zeitpunkten ihrer Eingabe - zu einer Nicht-Identität gleicher Einträge führen bzw. das Textcorpus müßte permanent aktualisiert werden, was ab einer gewissen Textmenge kaum mehr realisierbar ist.

Da die Entwicklung der ägyptischen Sprache historisch abgeschlossen ist, kann deren ehemaliger Wortschatz insgesamt als eine endliche Menge betrachtet werden. Dabei hat das einzelne Lexem einen historischen Entwicklungsprozeß durchlaufen, kann also in dieser Menge mehrfach in unterschiedlicher Form und mit verschiedenen Inhalten vertreten sein. Daneben gibt es einen weiteren dynamischen Charakterzug dieser toten Sprache, der sich aus der Verteilung der einzelnen Textzeugnisse über drei Jahrtausende ergibt.

Um diese Aspekte in ihrer Gesamtheit zu berücksichtigen, wurde beschlossen, jedes einzelne Lexem durch einen numerischen Code, die **WORTCORPUS-NUMMER**, zu repräsentieren. So wie über die Textcorpusnummer (TCN) sowohl die bildhafte Information zum Text als auch die zu dessen Publikationen aus den Datenbanken abrufbar sind, können zur WCN die unterschiedlichsten Informationen separat abgespeichert und gleichfalls separat oder aber auch komplex ausgeworfen werden. Die Abruf-Form wird bestimmt durch den Nutzer; da die einzelnen Informationen in ihrer Gesamtheit relational miteinander verknüpft sind, ergeben sich die vielfältigsten Möglichkeiten.

Texte, die unter Nutzung der WCN und einer durch das konkrete Erscheinungsbild spezifizierten Zusatzinformation numerisch codiert wurden, bleiben auch dann in ihrer physischen (ausschließlich numerischen) Existenz unverändert, wenn sich zu einem Lexem einzelne Informationsinhalte wie beispielsweise Transkription, grammatische Zuordnung oder Übersetzung ändern. Änderungen des einzelnen hieroglyphischen Erscheinungsbildes eines Lexems in einem einzelnen Text, der letztlich einzig originären Information, sind bedingt möglich; allerdings kann sich hier die Notwendigkeit ergeben, bei späteren Änderungen den gesamten Textcorpus auf das Erscheinen eines solchen Beleges überprüfen zu müssen, da er theoretisch mehrfach auftreten kann. Es kann jedoch davon ausgegangen werden, daß ein einmal festgestelltes Erscheinungsbild eine relativ hohe Stabilität besitzt; ohnehin sind interpretative Ver-

änderungen wie beispielsweise Emendationen in der Regel textbezogen und werden separat erfaßt.

Für die Eingabe der Texte sowie deren gleichzeitige Analyse existiert bisher keine aufgabenspezifische Software; grundsätzlich sind die dazu notwendigen Arbeitsprozesse aber im Multitasking unter Window-Technik simulierbar. Da allerdings gegenwärtig die Berliner Wortliste ohnehin noch nicht vollständig ist und zudem bei einer derartigen Eingabe eine 100%ige Datenintegrität nicht gewährleistet wäre, wird vorrangig die Fertigstellung der Berliner Wortliste sowie die Software-Entwicklung betrieben. Nach Fertigstellung beider Komponenten wird die Texteingabe in verschiedenen Arbeitsschritten erfolgen. Die von THOT und PPTES vorgenommene direkte Einbindung aller Daten zur Textbeschreibung in den jeweiligen Text wird am Berliner Wörterbuch in der separaten TCN-Datei getrennt vom Text erfaßt. Der eigentliche Text wird unter seiner TCN in einem interaktiven Prozeß eingegeben. Dabei sind mittels Window-Technik auf dem Bildschirm mehrere Bereiche sichtbar: Im Sinne eines Computer-Thesaurus (d.h. eine Liste feststehender Begriffe, die mit einer Liste von Synonymen verbunden ist) wird die Berliner Wortliste in den dafür notwendigen Bestandteilen (Transkription, Hieroglyphencode, Grammatik, Übersetzung) bei deren optischer Präsenz über sensitive Eingabemasken konsultiert. Der Bearbeiter hat die Möglichkeit, sich anhand der aufbereiteten Vorlage bei der Eingabe für den Bereich der Transkription oder aber den der hieroglyphischen Schreibung zu entscheiden, wobei der Eingabeprozess bei frühzeitiger Übereinstimmung zwischen einzugebendem Lexem und dem vorhandenen Eintrag der BWL durch dessen Übernahme abgebrochen werden kann. Nach einer derartigen Zuordnung wird das einzelne Lexem intern in den entsprechenden Nummerncode umgesetzt. Gleichzeitig erfolgt im Bereich der Wortformen-Datenbank, einer modifizierten Form der Berliner Wortliste, die Hinzufügung der jeweiligen TCN. Lexeme, die noch nicht oder nicht in der entsprechenden Variante vorhanden sind, müssen vollständig eingetragen werden, um die notwendigen internen Übernahmen in die Wortformen-Datenbank und in die BWL zu ermöglichen. Damit wird deren stete Vollständigkeit und Aktualität gewährleistet. Vor allem durch den hohen Variationsgrad in der hieroglyphischen Schreibung ergeben sich hier anfangs umfangreichere Codierungsarbeiten, die sich jedoch im Laufe der Zeit zunehmend reduzieren. Gerade dieser Umstand hat das hier vorgelegte Erscheinungsbild der Listen-Einträge unter Aleph entscheidend geprägt.

Zunächst war davon ausgegangen worden, anhand des archivierten Berliner Materials die dort vorhandene gesamte Variationsbreite in den Erscheinungsbildern eines jeden Lexem-Eintrages wiederzugeben (bis WCN 47), was neben der strikten Überprüfung aller Einträge auf deren Tragfähigkeit zu einem auf Dauer nicht vertretbaren Zeitverzug führte. Für den Rest der Einträge wurde ein selektives und interpretatives Verfahren gewählt, bei dem einerseits

die häufigsten, andererseits aber auch die untypischsten Vertreter für die Aufnahme in die BWL ausgesondert wurden. Während bei den Einträgen bis WCN 47 und ab 217 bis 358 grundsätzlich auch die Stellung der Hieroglyphen untereinander notiert wurde, ist diese für die Einträge WCN 48 bis 216 normiert worden, d.h. es wurde lediglich auf das Vorhandensein einzelner Hieroglyphen in einer bestimmten Reihenfolge geachtet, um die jeweilige Schreibung einem konkreten Eintrag zuordnen zu können. Dieses Verfahren hat insofern seine Berechtigung, da ein großer Teil des Archivmaterials aus hieratischen Textzeugnissen durch die damaligen Bearbeiter unterschiedlich in Hieroglyphen umgesetzt worden war. Von Bedeutung war ferner der Umstand, daß die Wortform das konkrete Erscheinungsbild eines Beleges bestimmt. Während der Numerus bei Substantiva im allgemeinen in seinen differenzierten Erscheinungsbildern ohne besondere Hervorhebung übernommen wurde, ist beim Genus strikt unterschieden worden, was beim Vorliegen unterschiedlicher Genera zu unterschiedlichen Haupteinträgen in der BWL führte (vgl. WCN 74 und 75). Außer wortbildenden Präfixen werden Affixe und Infixe für die Haupteinträge in der BWL abgetrennt. Die konkreten Erscheinungsbilder hierzu werden in einer separaten Datei erfaßt werden und sind später über einen in Arbeit befindlichen grammatischen Thesaurus ansteuerbar.

Bei allen Einträgen ist die Möglichkeit einer Angabe zur Derivation vorhanden. Ausgehend von Untersuchungen zum Gebrauch des Verbs *ʒwj* und seiner Komposita (WCN 49-54) wurde beschlossen, bei Vorhandensein von transitivem und intransitivem Gebrauch separate Einträge vorzunehmen und diese auf eine sog. Stammform, die im eigentlichen Sinne nicht belegbar ist, zurückzuführen. Dieses Verfahren soll klarere Angaben zu weiteren Derivationen ermöglichen. Beispielsweise entwickelt sich aus dem intransitiven Gebrauch von *ʒwj* in Verbindung mit dem Nomen *ib* ("weit sein an Herz") sowohl eine intransitive als auch transitive Verwendung: "froh sein" und "erfreuen" (WCN 53). Aus dem sog. Stammeintrag können sich - ohne Möglichkeiten einer differenzierteren Zuordnung - weitere Derivationen ergeben (vgl. WCN 41: ein Raum, der weit ist bzw. weit gebaut wurde). Für den vorliegenden Ausdruck der Aleph-Einträge werden die Derivations-Angaben aus drucktechnischen Gründen separat gegeben. Insgesamt ergeben sich Schwierigkeiten, alle zur Berliner Wortliste in Form einer Datenbank erfaßten Informationen in ausgedruckter Form präsentieren zu können. Um dies zu verdeutlichen, soll zumindest deren Struktur hier kurz vorgestellt werden:

Struktur der Berliner Wortliste

<p>1. Wortnummer (BWLWCN; LI; Data required)</p> <p>"238" - Die Wortnummer wird als <i>numerus currentis</i> vergeben.</p>	<p>"1" - Worteinträge, die nicht-synonyme, unterschiedliche Entsprechungen besitzen, werden durch Variantenanzahl unterschieden.</p>
<p>2. Numerische Ordnungsangabe (SEQUENCE; LI; Data required)</p> <p>"23800" - Die numerische Ordnungsangabe ist als manuelles Hilfsmittel zur Ordnung nach der Transkription der einzelnen Worteinträge zu verstehen.</p>	<p>9. Sachgebiets-Code (SUBJCODE; SI => THS)</p> <p>"7329" - Der Sachgebiets-Code folgt den Prinzipien der Eingabe mittels Thesaurus (im Stadium der Ausarbeitung).</p>
<p>3. Numerische Derivationsangabe (OWNER, LI)</p> <p>"228" - Die numerische Derivationsangabe ist als optionaler Eintrag zu verstehen und verweist auf die unter BWLWCN eingetragene Wortnummer.</p>	<p>10. Grammatik-Code (GRAMCODE; SI => THS; Data required)</p> <p>"1122" - Der Grammatik-Code folgt den Prinzipien der Eingabe mittels Thesaurus (im Stadium der Ausarbeitung).</p>
<p>4. Traditionelle Transkription (TTRANSCR; 50 CHAR; Data required)</p> <p>"Ax.tj" (= <i>sh.tj</i>) - Die traditionelle Transkription folgt den Prinzipien des Wörterbuches von Erman/Grapow.</p>	<p>11. Belegstelle I (TCN; LI; Data required)</p> <p>"1000000" - Die Belegstelle ist in der Anfangsphase identisch mit einem Eintrag aus den Datenquellen (Wörterbuch u.ä.); erst im späteren Verlauf werden die Belege mit der realen TCN gekoppelt.</p>
<p>5. Neue Transkription (NTRANSCR; 50 CHAR)</p> <p>"(noch zu bestimmen)" - Die neue Transkription folgt Prinzipien, die sich aus noch zu fixierenden Festlegungen am Berliner Wörterbuch ergeben.</p>	<p>12. Belegstelle II (REF; 50 CHAR)</p> <p>"Wb 1,18.11" - Referenz-Angabe für Einträge, die nicht aus realen Texten übernommen sind.</p>
<p>6. Hieroglyphen-Code (HIEROCOD; 250 CHAR; Data required)</p> <p>"H2:1-t:Z4-M2:Z2" - Der Hieroglyphen-Code folgt den Prinzipien des Leitfadens zur Verschlüsselung hieroglyphischer Texte.</p>	<p>13. Variantennummer (WVN; SI; Data required)</p> <p>"1" - Die Wortvariantennummer wird als <i>numerus currentis</i> vergeben und dient der numerischen Codierung unterschiedlicher Schreibungen eines Wortes.</p>
<p>7. Wortentsprechung (TRANSLAT; 250 CHAR)</p> <p>"Ackerpflanzen" - Die optionale Angabe der Wortentsprechung wird aus der Sprache übernommen, in der der Text publiziert wurde; es ist beabsichtigt, mehrere gleichartige Datenfelder für unterschiedliche Zielsprachen (Englisch, Französisch, Italienisch, Arabisch) aufzubauen.</p>	<p>14. Datierungs-Code (DATCODE; SI => THS; Data required)</p> <p>"19" - Der Datierungs-Code folgt den Prinzipien der Eingabe mittels Thesaurus; ein verbindliches neues Codierungssystem zur Datierung liegt vor.</p>
<p>8. Äquivalenzzählung (AEQVN; SI; Data required)</p>	<p>15. Kommentarfeld (BWLCOMM; 2500 CHAR)</p> <p>"keine konkrete Nutzpflanze sondern allgemeiner Sammelbegriff" - Das Kommentarfeld ist inhaltlich offen gestaltet.</p>