

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 Arbeitsstelle "Ägyptisches Wörterbuch" an der neugegründeten Berlin-Brandenburgischen Akademie der Wissenschaften hat gegenwärtig innerhalb des Wörterbuchprojektes drei Subprojekte.¹

1. Sammlung von ägyptischen Texten, die nach Abschluß der Wörterbucharbeiten publiziert bzw. bekannt geworden sind. Ziel ist der Aufbau eines Text-Corpus innerhalb einer elektronischen DATENBANK (DB) mit allen Informationen zu den Texten sowie den Texten selbst.
Grundlage ist dabei der Textbestand des Zettelarchivs. Jeder seinerzeit verzettelte Text erhält eine TEXTCORPUSNUMMER (TCN), wird exakt identifiziert und durch Informationen über Standort, Inventarnummer, Publikationsstand, Datierung, Textgattungsbestimmung etc. erläutert. Er findet damit Eingang in einen TEXTCORPUSVERWALTER (TCV-ALT). Eine Neuaufnahme dieser Texte selbst erfolgt vorerst allerdings nicht. Als Basisarchiv soll der TCV-ALT lediglich eine moderne Referenzgrundlage für die Belegstellen des Ägyptischen Wörterbuchs sein.
2. Aufnahme der Texte selbst, die bisher nicht für die Wörterbucharbeit erschlossen sind, mit einem geeigneten Texteingabe- und Analyseprogramm.
3. Aufbau einer lexikalischen DB. Diese soll möglichst das gesamte lexikalische Wissen aufnehmen, über das die Ägyptologie bereits verfügt. Grundlage ist das ausgedruckte Wörterbuch von Erman-Grapow, das ergänzt wird (incl. Streichungen) durch die Wörterbücher und Wortsammlungen der vergangenen Jahrzehnte wie Faulkner, Meeks, etc. Diese DB wird in ihrem ersten Arbeitsstadium in Gestalt einer Wortliste, der sog. BERLINER WORTLISTE (BWL), realisiert.

Die BWL soll zwei Funktionen erfüllen: So stellt sie einerseits den Kern der geplanten lexikalischen DB dar, wird folglich selbst schon als relationale DB aufgebaut, und ist andererseits als ein Hilfsmittel bei der Neuaufnahme von Texten konzipiert.

Indem die DB interaktiv mit einem Textanalyseprogramm gekoppelt wird²

¹ Vgl. die bereits erschienenen Beiträge zum Berliner Projekt: S. GRUNERT, W. F. REINEKE, Das Berliner ägyptische Wörterbuch, in: GM 129 (1992), 9-20; P. M. WOLF, Ägyptologie und Informatik in Berlin, in: GM 128 (1992), 43-67, S. GRUNERT, I. HAFEMANN, Zum Berliner Projekt eines rechnergestützten Wörterbuches: Problemstellung und Lösungsvarianten, in: Informatique et Egyptologie 8 (1993), 32-41, P. M. WOLF, Standards für den Datenaustausch zwischen Datenbanken in der Ägyptologie, in: Informatique et Egyptologie 8 (1993), 127-136.

² Zu diesem Ansatz jetzt auch S. J. SEIDLMAYER, Aus der Arbeit an einem philologisch-prosopographischen Texterschließungssystem, in: Informatique et Egyptologie 8 (1993), 104.

und damit beide Funktionen verbunden sind, wird gewährleistet, daß bei der Textanalyse immer das gesamte lexikalische Wissen dieser DB als Hilfsmittel genutzt werden kann. Außerdem wird durch die Analyse neuer Texte mittels der lexikalischen DB diese sowohl um Neueinträge als auch um morphologische und syntaktisch-semantische Varianten schon bestehender Einträge erweitert. Gleichzeitig wächst bei bestehenden Worteinträgen deren Belegzahl. Die konkret technische Verknüpfung beider Funktionen stellt z. Zt. noch ein Problem dar, das durch einen Programmierer zu lösen ist. Im Folgenden soll auf die Vielfältigkeit dieser lexikalischen DB eingegangen werden.

Der konzeptionelle Ansatz ist folgender: Texte, die in Abschnitte und Sätze untergliedert sind, werden in Phrasen und Wörter zerlegt, welche in der lexikalischen DB bereits erfaßt sind. Jeder sprachliche Ausdruck hat mehrere Eigenschaften, die er mit anderen Ausdrücken auf verschiedene Weise teilt. Im wesentlichen gibt es drei Analysedimensionen:

- I. Lexikalische Bedeutung,
- II. Morphologisch-syntaktische Form,
- III. Syntagmatische Funktion.

Im Mittelpunkt unserer Arbeit steht die lexikalische Analyse, d. h. die lexikalische Bedeutung der Wörter, auf deren Grundlage sich semantische Analysen vornehmen lassen. Semantisches Wissen kann repräsentiert werden durch semantische Netze, semantische Relationen (Sinnrelationen), Begriffssysteme oder Wissensbasen. Entscheidend ist die Art der Wissensrepräsentation, d.h. wie stellt man Wissen für die Weiterverarbeitung, konkret für maschinelle Weiterverarbeitung, dar?

In den letzten Jahren sind hierfür unterschiedliche Arten maschineller Wörterbücher erarbeitet worden.³

Die einfachste Variante ist das in maschinenlesbarer Form abgespeicherte, ursprünglich gedruckte Wörterbuch.

Weitergehend sind maschinelle Wörterbücher, deren Einträge mit codierten Angaben zur Grammatik, wie Wortart, morphologische Kategorie oder auch Valenzen u.ä. angereichert sind. Diese Arten maschineller Wörterbücher können aus verschiedenen Wörterbüchern zusammengestellt sein. Sie werden z.B. zur halbautomatischen Standardisierung (Lemmatisierung) bei literarischer Textbearbeitung oder Datenaufnahme verwendet. Ein solches Wörterbuch benutzt auch die neue Version des Programms PPTES (vgl. Anm. 2).

Eine lexikalische DB enthält strukturierte und formalisierte Informationen zu den einzelnen Einträgen auf verschiedenen Ebenen und sichern so einen wirklich effektiven Gebrauch der sich ständig entwickelnden Computer-

³Vgl. N. CALZOLARI, Computer-Aided Lexicography: Dictionaries and Word Data Bases, in: Handbuch für Sprach- und Kommunikationswissenschaft 4 (1989), 510-519 [ed. by I. S. Batori, W. Lenders, W. Putschke, Computational Linguistics].

technologie. Diese Präsentation lexikalischen Wissens kann völlig neue Ansätze und Anwendungsbereiche für die moderne Lexikographie und auch Lexikologie bieten.

Schließlich gibt es noch computerlinguistische Wörterbücher, die für Parser, für Anwendungen in der künstlichen Intelligenz oder für maschinelle Übersetzungsprogramme angelegt wurden. Sie sind äußerst spezifiziert und je nach Anwendung und Zielstellung sehr verschieden strukturiert. So können sie morphologische Informationen, Angaben zur Subkategorisierung, Strukturen der Prädikatenlogik, semantische Relationen etc. enthalten. Solcherart Wörterbücher sind ganz speziell auf den Parser bezogen, für den sie Daten zur Verfügung zu stellen haben, und sie sehen selbstredend für jede natürliche Sprache anders aus. Sie sind in starkem Maße von linguistischen Theorien abhängig und arbeiten mit entsprechend verschiedenen Formalisierungen.⁴

Tatsächlich sind die vorhandenen Wörterbücher als Resultat lexikographischer Arbeit ja nicht nur Wortlisten, sondern eine Hauptquelle unserer Kenntnisse über Weltwissen und Weltverständnis einer Kultur und ihrer Träger. Bei den Überlegungen, dieses Wissen aus den verschiedenen lexikalischen Quellen möglichst umfassend auszubeuten und dabei Wörterbücher verschiedenen Typs miteinander zu verbinden, stellt sich das Konzept einer lexikalischen DB als höchst fruchtbar dar. Diese ist multifunktional und kann größtenteils unabhängig von Systemtheorien sein.

Elementare Vorüberlegungen sollte man auf das Zusammenspiel von Zielstellung, Methoden der Speicherung und Suchroutinen richten. Man muß eine "Verbindungsmöglichkeit" zu den an sich sehr verschiedenen Typen von Informationen schaffen, um diese dann für eine Systemanalyse einer natürlichen Sprache zur Verfügung stellen zu können. Es handelt sich um morphologische, syntaktische, semantische und pragmatische Informationen, von denen jede verschiedene Anforderungen an Inhalt und Struktur des Eintrages stellt. So scheint uns der effektivste Weg, die o. g. sehr unterschiedlichen und in sich stark strukturierten Informationen mit Methoden zu verarbeiten, die typisch für *database-management-systems* (DBMS) sind. Die Verwendung von DBMS ermöglicht es, die Daten den Notwendigkeiten entsprechend zu strukturieren, ohne bestimmte Nachteile (wie z.B. Redundanz) in Kauf nehmen zu müssen. Es ist möglich, dem Nutzer Zugang zu den Daten zu verschaffen, ohne daß er mit der internen oder physischen Organisation der Daten vertraut sein muß, indem man noch ein vernünftiges Query-User-Interface auf die Datenstruktur programmiert. Es wird mit einer *query-language* gearbeitet, die der natürlichen Sprache recht ähnlich sein kann. Die *search*-Methoden ermöglichen einen direkten bzw. punktuellen Zugriff auf die Daten, im Gegensatz zum

⁴ Vgl. ausführlich dazu P. HELLWIG, Parsing natürlicher Sprachen: Grundlagen, sowie ders., Parsing natürlicher Sprachen: Realisierungen, in: Handbuch für Sprach- und Kommunikationswissenschaft 8 (1989), 348-432.

sequential-reading anderer Systeme. Allerdings sind bisherige kommerzielle DBMS nicht so gut für unsere oben umrissene Aufgabenstellung geeignet.⁵

Zur Entwicklung einer solchen DB, also zur logischen Organisierung der Daten ist eine Art Modell der lexikalischen Welt zu entwerfen. Dazu muß folgendes festgelegt werden:

1. Objekt-Typ oder *entity-set* (*entity* ist ein Element, das durch einen Namen und bestimmte Eigenschaften charakterisiert ist).
2. Mögliche Beziehungen und Assoziationen zwischen den Eigenschaften des gleichen *sets* oder den Entitäten verschiedener *sets*.

Hierbei ist das relationale Datenbankmodell den Eigenschaften der lexikalischen Welt angemessen, wobei allerdings zukünftige Lösungen wohl durch objektorientierte Datenbankmodelle bestimmt sein werden.⁶

Das Design einer relationalen DB besteht aus der Bestimmung eines *sets* von Relationen, die Fakten beschreiben, mit denen wir umgehen. In unserem Fall ist das *set* der Lemmata die grundlegende Entität, die mit allen anderen Relationen in Beziehung steht. Mit der Vergabe von *primary* und *secondary keys* (Kupplungsglieder zwischen unterschiedliche Tabellen, deren Werte übereinstimmen) werden potentiell die bestehenden physischen und virtuellen Verknüpfungen hergestellt. Im Moment der Anfrage können die in sich unterschiedlichen Daten nicht nur einzeln abgerufen, sondern auch entsprechend ihren Strukturen beliebig zusammengestellt werden, und zwar jeweils nach den definierten Attributen einer Entität. Dabei ergibt sich die Möglichkeit sowohl von 'eins-zu-eins'-Relationen, wie auch 'eins-zu-viele', 'viele-zu-eins' oder 'viele-zu-viele'. So können beliebige Verbindungen hergestellt werden, die bei einem alphabetisch aufgearbeiteten Wörterbuch selbst dann nicht möglich sind, wenn die Daten im Rechner abgespeichert wurden.

Die lexikalische DB sollte so angelegt sein, daß sie verschiedene linguistisch-lexikalische Fragestellungen befriedigen und klären helfen kann, d.h. verschiedene Nutzer der gleichen Daten auch verschiedenen Zugang zu ihnen bekommen können. Die unterschiedlichen *views* einer DB existieren dabei aber nur virtuell; tatsächlich sind sie definiert in Begriffen der *database-retrieval-functions*, d. h. es erfolgt die Projektion der Daten zu einem *view* bei der Abfrage. Daher können alle Daten ohne physische Änderungen mit großer Freizügigkeit automatisch so umgeordnet werden, wie sie für einen Report entsprechend einer ausgewählten Fragestellung erwünscht sind. Diese Ordnung kann z. B. identisch sein mit der Wortart- oder Satzglied-Bestimmung, die zum Such-Kriterium definiert wurde. So könnte man z. B. alle Adjektive zusammenstellen bzw. alle intransitiven Verben oder alle Substantive, die in Lokalbestimmungen verwendet werden.

⁵ Dazu vgl. jetzt auch P. M. WOLF, a. a. O. (1993).

⁶ Ders., a. a. O., 128.

Einerseits ist der Zugang zu einem Wort das Wichtigste, und andererseits muß die Computer-Organisation die verschiedenen Ebenen der linguistischen Informationen widerspiegeln. Durch diese Verknüpfung der verschiedenen Ebenen wird ein Wörterbuch, das in Form einer DB organisiert ist, die Grenzen eines normalen, sequentiell alphabetisch geordneten Wörterbuchs sprengen. Voraussetzung ist die exakte Analyse eines lexikalischen Eintrages und die Festlegung seiner Basis-Elemente, d.h. für die Formalisierung eines lexikalischen Eintrages nach den relevanten Kategorien in der DB muß vorher eine Arbeitshypothese erarbeitet werden. Einige solcher Kategorien findet man bereits in ausgedruckten Wörterbüchern codiert wie Wortartbestimmung (Subst., Verb, Adj.), Grammatikangaben (mask., fem., pl., dat., akk. etc.), oder Sprachgebrauchsebene (vulg., dial., poet. etc.) und Sachgebietsangaben (zool., techn., astron. etc.). Die interessante Beziehung zwischen lexikalischen und syntaktischen Aspekten sowie der Semantik und Pragmatik der Wörter könnte in Zukunft effektiver in einer lexikalischen DB abgefragt werden.

Die von uns konzipierte lexikalische DB für das Ägyptische soll vorerst folgende Datenfelder erhalten, die in verschiedenen Listen zusammengestellt werden können :

- | | |
|---------------------------------|--------------------------------|
| 1. Wortnummer, | 2. Numerische Ordnungsangabe, |
| 3. Derivationsangabe, | 4. Transkription, |
| 5. Neue Transkription, | 6. Hieroglyphische Schreibung, |
| 7. Wortensprechung/Übersetzung, | 8. Äquivalenzzählung |
| 9. Sachgebiet, | 10. Grammatik, |
| 11. Belegstelle, | 12. Literatur |
| 13. Variantennummer | 14. Datierung/Belegzeitraum, |
| 15. Kommentar. | |

1. Die Wortnummer ist der nur einmal vorhandene, eindeutige "Name" einer Entität. Sie stellt als rein numerischer Ersatzcode für den einzelnen Worteintrag eine mathematisch ganze Zahl vom Datentyp *Integer* (Bereich zwischen - 2147483648 bis + 2147483647) dar, die zur Festlegung referentieller Integritätsbedingungen zwischen den verschiedenen Tabellen (Datenlisten) der Datenbank durch unterschiedliche Schlüsselvergabe (*primary und foreign key*) notwendig ist. Da sich alle Informationen in den verschiedenen Tabellen jeweils auf einen ganz konkreten Worteintrag beziehen, muß dieses Datenfeld in den Datenlisten, die einen direkten Bezug zur Wortnummer haben, als Verbindungsglied auftreten.
2. Die numerische Ordnungsangabe ist ein vorläufiges Hilfsmittel zur Ordnung nach der Lesung eines Worteintrags und wird als *numerus currens* vergeben. Der Datentyp entspricht dem der Wortnummer.
3. Die numerische Derivationsangabe, der sog. Owner, verweist auf die eingetragene Wortnummer eines Lexems, von dem das durch eine eigene Wort-

nummer benannte Lexem abgeleitet ist. Dadurch werden die im Grammatikfeld (s. u. 10.) codierten allgemeinen Ableitungstypen für jeden Worteintrag konkretisiert: Als deverbale Nomina werden beispielsweise die Einträge *jt3 Dieb* zu *jt3 fortnehmen* oder *z3 Schutz* zu *z3w schützen* gestellt. Hier können wichtige Forschungen der letzten Jahre in die DB einfließen, so die zur Nominalbildung im Ägyptischen, und auf effektive Art erreichbar und verarbeitbar gemacht werden. Der Datentyp entspricht dem der Wortnummer.

4. Im Datenfeld "Transkription" erfolgt die Dateneingabe nach den Prinzipien des Wörterbuchs von Erman/Gradow und den Codierungsvorschriften von WINGLYPH. Der Datentyp ist *Varchar*, es handelt sich also um Zeichendaten (ASCII-Daten) mit variabler Länge, die bei der Definition des Datenfeldes allerdings auf 50 Zeichen begrenzt wurde. Die Umsetzung der ASCII-Daten in den Transkriptionszeichensatz von WINGLYPH erfolgt gegenwärtig nur außerhalb der Datenbank.
5. Das optionale Datenfeld "Neue Transkription" ist vorgesehen, um einem - im Rahmen einer Arbeitstagung 1994 noch zu erarbeitenden - moderneren Transkriptionssystem den nötigen Platz einzuräumen. Vorläufig wurde als Datentyp der der Transkription definiert.
6. Die hieroglyphische Schreibung wird durch einen Code, der dem "Leitfaden zur Verschlüsselung hieroglyphischer Texte" folgt, erfasst.⁷ Im multitasking-Betrieb (unter Nutzung der Clipboard-Technik) ermöglicht die Verwendung von WINGLYPH eine sofortige Sichtbarmachung der Hieroglyphen am Bildschirm; was die Überprüfung von Schreibungen in einem aktuellen Text mit den bereits in der Datenbank gespeicherten Varianten wesentlich beschleunigt. Alle Schreibungsvarianten tragen einen numerischen Code, der mit der Wortnummer verbunden den Bezug auf einen Worteintrag ermöglicht (s. u. 13.). Der alphanumerische Datencharakter bedingt *Varchar* als Datentyp; es handelt sich also wieder um ASCII-Daten in einem Datenfeldes, das auf 250 Zeichen begrenzt wurde. Die Umsetzung der ASCII-Daten in den hieroglyphischen Zeichensatz von WINGLYPH erfolgt gegenwärtig nur außerhalb der Datenbank.
7. Die Übersetzung wird, soweit vorhanden, dem Wörterbuch oder bei neuen Wörtern aus der Sprache entnommen, in der ein Wort oder Text erstmals publiziert wurde, vor allem, um nicht durch die Übertragung in eine dritte

⁷ Für die Codierung der hieroglyphischen Schreibung unter WINGLYPH, das jetzt immerhin über 5000 Zeichen zur Verfügung stellt, kann in den Fenstern der Zeichenliste (Gardiner-Zeichensatz) und zusätzlich einem Handbuch geblättert werden. Als Eingabehilfe wurde auch ein unter WINGLYPH arbeitendes Graphiktablett mit frei definierten Makros entwickelt, das zwar weniger Hieroglyphen, diese aber weit übersichtlicher angeordnet bietet (maximal 1320 auf einer Ebene liegende Zeichen). Vgl. H. VAN DEN BERG, Glyph for Windows, in: *Informatique et Egyptologie* 8 (1993), 113-120; I. HAFEMANN, Eine Eingabehilfe bei der Hieroglyphencodierung, in: *GM* 129 (1992), 21-25.

Sprache Bedeutungsnuancen zu verwischen. Der rein alphabetische Datencharakter bedingt *Varchar* als Datentyp; es handelt sich also wieder um ASCII-Daten auf der Basis der Zeichenumsetzungstabellen 437 und 850 in einem Datenfeld, das auf 250 Zeichen begrenzt wurde. Dadurch können Wortentsprechungen der ägyptologischen Standardsprachen (Englisch, Französisch, Deutsch und Italienisch) problemlos übernommen werden. Es ist geplant, innerhalb der Wortentsprechungen separate Datenfelder für diese Sprachen unter Mithilfe von Muttersprachlern aufzubauen.

8. Der diachrone Charakter der Wortdatenbank führt zur Aufnahme semantischer (aber auch morphologischer und grammatischer; s.u.) Varianten, einschließlich des bildlichen Gebrauchs eines Worteintrages. Bei den Wortentsprechungen werden diese Einträge durch die Äquivalenzzählung erfasst und mit dem Ausgangseintrag (Wortnummer) verbunden. Sie wird als *numeris currens* vergeben und stellt eine mathematisch ganze Zahl vom Datentyp *Smallint* (Bereich zwischen - 32768 bis + 32767) dar.
9. Für die Zuordnung eines Lexems zu einem Sachgebietes kann aus Begriffen eines verbindlichen Thesaurus gewählt werden. Dem Begriffsthesaurus können im Bedarfsfall anderssprachige Synonymlisten zugefügt werden. Als Verbindungsglied dient ein numerischer Code vom Datentyp der Äquivalenzzählung. Nur der numerische Code, der durch einen *foreign key* mit dem Begriffsthesaurus verbunden ist und daher jederzeit durch den eigentlichen Begriff ersetzt werden kann, wird in der DB gespeichert.

Die Erarbeitung des Begriffsthesaurus folgt der modernen lexikographischen Klassifikation und ist durch Termini der altägyptischen Begriffswelt ergänzt. Hierbei kann auf die Klassifizierung des 6. Bandes von Erman/Grapow (S. 195-221) zurückgegriffen werden.⁸

Da jedes Wort Teil eines Textes ist wird die Sprachgebrauchsebene, die durch einen Textgattungs-Thesaurus klassifiziert und innerhalb des TCV bei der Textbeschreibung angegeben ist, ebenfalls für jedes Wort feststellbar sein.⁹ Das gilt allerdings nur für neu eingegebene Texte, da für das ausgedruckte Wörterbuch nicht alle einzelnen im WbZ vorhandenen, sondern nur die dort indizierten Lemma-Belege notiert werden.

Durch die so mögliche Kopplung von Sachgebiet und Sprachgebrauchsebene kann die Beziehung Semantik-Pragmatik erfasst werden. Problematisch ist die Zuordnung eines Textes zu einer Gattung bei den Fällen, in denen ein Text aus mehreren Teilen unterschiedlicher Gattungen besteht, also z. B. Grabinschriften mit Opferformeln, Hymnen, Listen etc.

⁸ Der Sachgebietsthesaurus ist noch nicht abgeschlossen.

⁹ Zu den unterschiedlichen Auffassungen über die Anforderungen an einen Textgattungsthesaurus vgl. die Standpunkte von D. VAN DER PLAS und S. GRUNERT auf der letzten Arbeitstagung der Gruppe Informatique et Egyptologie 1993 in Genf (erscheint in: Informatique et Egyptologie 9, 1994).

10. Für die Grammatikangaben bei einem Lexems kann ebenfalls aus Begriffen eines verbindlichen Thesaurus gewählt werden. Dem Begriffsthesaurus können im Bedarfsfall anderssprachige Synonymlisten zugefügt werden. Als Verbindungsglied dient ein numerischer Code vom Datentyp der Äquivalenzzählung. Nur der numerische Code, der durch einen *foreign key* mit dem Begriffsthesaurus verbunden ist und daher jederzeit durch den eigentlichen Begriff ersetzt werden kann, wird in der DB gespeichert. Die Begriffe beziehen sich gegenwärtig ausschließlich auf die Ebene der Wortartbestimmung und beinhalten bedingt auch Genus und Verbkategorie (transitiv/intransitiv). Konkrete Derivationsangaben zu einem speziellen Eintrag finden sich als Nummerncode in einem eigenen Datenfeld (s.o. 3.). Als Ergebnis der Arbeit mit dem Textanalyseprogramm PPTES ist ferner geplant, bei Verben u.a. auch die Verbalklasse und bei anderen Wortarten auch Bildungsart und Ableitungstyp (Substantive als Nisbildungen, verbale Substantive von Partizipien, Präpositionaladverbien etc.) zu codieren.

Da die grammatische Bestimmung eines Beleges eindeutig sein muß, werden auch bestimmte semantische Varianten oder einige häufige Transpositionen auf der Ebene des Lexikons definiert, m.a.W. lexikalisiert. Somit werden Verben, von denen transitive Belege neben intransitiven bekannt sind, zweimal aufgenommen, desgleichen werden Eigenschaftsverben, als Verb und Adjektiv aufgenommen oder Zeitsubstantive auch als Adverbien - dies jeweils unter Angabe ihres Ableitungswortes durch den numerische Derivationscode.¹⁰

11. Einerseits beinhaltet die Belegstelle bei neu eingegebenen Texten die für ein durch die Wortnummer spezifiziertes Lexem jeweils gültige Textcorpusnummer (TCN).¹¹ Bei dieser handelt es sich um einen rein numerischen Ersatzcode für jeden einzelnen Text, der eine mathematisch ganze Zahl vom Datentyp *Integer* (s.o. 1.) darstellt. Andererseits werden der TCN in eignen Datenfelder interne Zählungen (Zeilennummer+Wortnummer) hinzugefügt, die vom Datentyp her mit der Äquivalenzzählung identisch sind.
12. Innerhalb der Literatur-Angaben wird bei der Ausarbeitung der BWL im allgemeinen auf die Datenquellen verwiesen, die für die Berliner Wortliste ausgewertet wurden (vgl. Beitrag Reineke). Im Interesse einer sofortigen Überprüfbarkeit durch den Nutzer wurde hier auf die numerische Codie-

¹⁰ Bei der Entscheidung, welche Transpositionen lexikalisiert werden, lassen wir uns von pragmatischen Gesichtspunkten im ägyptischen Sprachgebrauch leiten; hier sind aber jederzeit Einschübe möglich durch die Nutzung der numerischen Ordnungsangabe und der Wortnummer als Owner.

¹¹ Über die TCN ist der Zugriff zum TCV mit allen interessanten Informationen zu einem Text möglich, z. B. zu einer vollständigen Bibliographie. Ein bibliographisches Nummerncodesystem ist bereits erarbeitet und findet bei dem Aufbau der Literaturdatenbank Verwendung.

rung entsprechend dem vorliegenden Nummerncodesystem (vgl. Anm. 11) verzichtet und auf die gängigen ägyptologischen Abkürzungen zurückgegriffen. Der in aller Regel alphanumerische Datencharakter bedingt *Varchar* als Datentyp; es handelt sich also wieder um ASCII-Daten in einem Datenfeld, das auf 50 Zeichen begrenzt wurde.

13. Die Variantenummer - vom Datentyp her mit der Äquivalenzzählung identisch - dient der internen Unterscheidung zwischen unterschiedlichen Varianten in der hieroglyphischen Schreibung eines Lexems.
14. Die Datierung, also der Belegzeitpunkt bzw. Belegzeitraum der einzelnen spezifischen Schreibung eines Lexems folgt ebenfalls wieder den Prinzipien der Eingabe mittels Thesaurus (s.o. 9.), wobei ein verbindlicher Datierungsthesaurus vorliegt, der auch im TCV verwendet wird. Die dabei vorhandene eins-zu-viele-Beziehung (gleichartige Schreibungen eines Lexems können zu den unterschiedlichsten Zeiten auftreten) wurden durch den n-dimensionalen Charakter der aus insgesamt 4 numerischen Datenfelder (*Smallintr*) bestehenden Datierungscodierung erreicht (WCN-Variantennummer-interne Datierungszählung-Datierungscode).

Da für die Verarbeitung der Lexeme neben semantischen Aspekten auch grammatische und morphologische berücksichtigt werden, ergab sich, wie bereits ausgeführt, zwangsläufig die Aufspaltung eines Worteintrages in zwei oder in mehrere Einträge, wenn beispielsweise neben einer maskulinen auch eine feminine, neben einer intransitiven auch eine transitive Form bekannt ist und sich gegebenenfalls daraus Derivate entwickelten. Ausgangspunkt hierfür war die Überlegung, den Lexemen der BWL jeweils eindeutig Worte aus einzelnen Texten zuordnen zu können. Gerade für diesen Bereich scheint für die Zukunft ein Datenfeld interessant, das lemmaspezifische lexikalische Ergänzungen und ihre morpho-syntaktische Charakterisierung enthält.

Für jede der genannten Ebenen mußte eine eigene Datenstruktur entwickelt werden, d.h. das Datenfeld 'Hieroglyphische Schreibungen' ist anders aufgebaut als das Datenfeld 'Sachgebietsangabe' oder 'Datierung', letztere sind z.B. thesaurusabhängig. Die Ausarbeitung der entsprechenden Thesauri stellte und stellt eine nicht zu unterschätzende Vorarbeit dar, bei der Fachwissen eingebracht und mit praktischen Erwägungen verbunden werden muß. So darf ein Textgattungs-Thesaurus neuen Forschungen in der ägyptologischen Literaturwissenschaft nicht gerade widersprechen, kann aber nicht alle Feinheiten der dort vorgenommenen Klassifizierung bieten, zudem diese auch in der Diskussion sind.

Abhängig von den Fragestellungen kann eine Reorganisation der Daten vorgenommen werden, d.h. der Computer kann bei solcherart Ausnutzung den Pluralismus und das Dynamische der Daten deutlicher machen als jede herkömmliche Art Wörterbuch. Eine echte lexikalische DB kann als ein multifunktionales Objekt charakterisiert werden und ist in gewisser Hinsicht auch

ein multidimensionales Wörterbuch, in dem nach verschiedenen Aspekten eines Wortes gesucht werden kann, wenn man verschiedene Pfade innerhalb der DB verfolgt. Eine DB, die originäre Daten unter verschiedenen Aspekten speichert, hat die Eigenschaft, dieselben Informationen zu multiplizieren. Der Multiplikationsfaktor wird durch die verschiedenen *views* oder Suchstrategien gegeben. Neben der Nutzung einer solchen DB für die unterschiedlichsten Abfragen besteht auch die Möglichkeit daraus Sublexika zu erstellen, die ausgewählte Teile des gesamten Lexikons beinhalten, wie terminologische Fachlexika oder Sammlungen von Ortsnamen, Personennamen u.ä. In unserem Fall könnte man durch die Derivations- und Owner-Angaben z.B. Vorarbeiten für ein Wurzelwörterbuch oder ein etymologisches Wörterbuch leisten, Untersuchungen zu Bildungstypen von Wortarten unter Berücksichtigung semantischer oder historischer Aspekte, oder zur Verwendung bestimmter Termini oder hieroglyphischer Schreibungen in bestimmten Textgattungen zu bestimmten Zeiten etc. anstellen. Ergebnisse können ausgedruckt, aber auch auf die neuen Speichermedien übertragen und als dynamische Teilwörterbücher verwendet werden. Damit würde z. B. auch die Trennung zwischen den verschiedenen Typen von Wörter- und Spezialwörterbüchern aufgehoben werden.

Selbstredend sind aber für die Bereitstellung einer solchen DB eine Reihe von Vorarbeiten nötig. Jegliche Operationen sind abhängig von den vorher definierten Strukturen. Da es hier vorrangig um lexikographische Arbeiten geht, stehen beispielsweise die syntaktischen Strukturen des Ägyptischen nicht im Vordergrund. So kann z. B. nicht die konkrete Subjekt-Objekt-Beziehung in einem Satz eines konkreten Textes abgefragt werden, oder die hieroglyphische Schreibung bestimmter Flexionsformen in einem Text oder einer Gruppe von Texten. Hierfür müssen spezielle Programme erarbeitet werden, und das für die verschiedenen Sprachstufen des Ägyptischen.¹²

Dieses Projekt folgt konzeptionell den neuen Ansätzen in der Computerlexikographie. Selbstverständlich wird es kaum mehr möglich sein, die Datenfülle einer solchen lexikalischen DB in einem Buch, also einem neuen Wörterbuch unterzubringen. Dennoch ist geplant, Teilergebnisse zu publizieren, vorerst in Form einer sich ständig erweiternden Wortliste.

Vielsprechend scheint uns der Ansatz zu sein, diese computerisierte Wortbasis als Lexikon zu benutzen und ihre strukturierten Informationen auf

¹² Vgl. das Textanalyseprogramm THOT, welches die drei Ebenen Transkription - Übersetzung - hieroglyphische Schreibung verknüpft analysiert - dazu J. HALLOF, THOT - Zum Konzept eines Programms zur Analyse ägyptischer Texte, in: *Informatique et Egyptologie* 8 (1993), 42-46.

Um auch syntaktische Erscheinungen in der oben beschriebenen DB besser zu erfassen, könnte man bspw. bestimmte Flexionsformen, die als Affixe oder Infixe vom Grundwort abzutrennen und separat abzulegen wären, lexikalisieren, oder aber eine Liste der Wortform-Schreibungen anlegen. Die Diskussionen hierüber sind nicht abgeschlossen.

einen laufenden, aktuell aufzunehmenden Text zu spiegeln, der dann vorher nicht indiziert werden muß. Das würde, natürlich immer entsprechend den vorher ausgewählten Kriterien, eine syntaktisch-semantische Lemmatisierung gestatten. Man kann so verschiedene Bedeutungsschattierungen ermitteln und Vergleiche mit anderen Begriffen des gleichen semantischen Feldes anstellen, Kollokationen oder idiomatischen Gebrauch sowie Häufigkeiten etc. feststellen. Andererseits kann man durch Befragung der Textdatei Wörter in ihrem Kontext studieren, neue Entscheidungen in der lexikalischen DB speichern und so für lexikologische Untersuchungen bzw. als Übersetzungshilfen bei der Bearbeitung weiterer Texte konsultieren.¹³

Abschließend muß darauf verwiesen werden, daß das Problem der Verknüpfung von lexikalischer DB und Text-DB durch ein integriertes Textanalyseprogramm noch nicht gelöst ist, da dies die Zusammenarbeit mit einem professionellen Programmierer erfordert. Gegenwärtig konzentrieren sich alle Vorarbeiten auf die Fertigstellung der "Berliner Wortliste" mit den oben beschriebenen verschiedenen Datenfeldern. Die Ausführungen sollten sowohl den Umfang der notwendigen Vorarbeiten verdeutlichen als auch die Notwendigkeit weiterer Diskussionen wie zu Fragen der Thesauri, zur Transkription u. a.m. aufzeigen.

¹³Zu diesem Ansatz vgl. H. H. ZIMMERMANN, Multifunctional Dictionaries, *Linguistica Computazionale* III (1983), 279-288, bes. 286. In der Ägyptologie jetzt auch S. J. SEIDL-MAYER, a. a. O., 104, der auf Transkriptionsbasis mit einer Wortliste von standardisierten Primäreinträgen arbeitet, indem er die in den Texten vorkommenden Wortformen diesen Primäreinträgen zuordnet, also die Textwörter standardisiert bzw. lemmatisiert. Den Primärwerten der Wortliste sind morphologisch-syntaktische Eigenschaften zugeordnet.