TEXTDATENBANKEN UND/ALS VIRTUELLE WÖRTERBÜCHER

WOLFGANG SCHENKEL

1 EINLEITUNG

1.1 Lage der Lexikographie

Wir hatten uns zu diesem Kolloquium zusammengefunden, um den gegenwärtigen Stand und die künftigen Ziele der Lexikographie des Altägyptisch-Koptischen zu diskutieren. Wenn wir dies aus Anlass des 100jährigen Jubiläums des Akademieunternehmens (Alt-)Ägyptisches Wörterbuch taten, dann deshalb, weil dieses Unternehmen Jahrzehnte nach seinem ersten Abschluss und selbst 100 Jahre nach seinem Beginn immer noch einen Maßstab vorgibt, hinter den es nicht zurückzufallen gilt.

Es hat vor Adolf Erman, der das Projekt (Alt-)Ägyptisches Wörterbuch aus der Taufe hob¹, ägyptologische Lexikographie gegeben. Ich erinnere an Jean François Champollion² selbst, an Samuel Birch³ und Heinrich Brugsch⁴, um nur einige Pioniere herauszugreifen.⁵ Es hat auch neben den Berliner Arbeiten Lexikographie anderswo gegeben. Ich erinnere an den zu spät gekommenen Wallis Budge⁶, an Wolja Erichsen, der sich spät, aber noch nicht zu spät um das vernachlässigte Demotisch gekümmert hat,⁷ ja an Raymond Faulkner, dessen „Concise Dictionary“ ein Jahr vor dem letzten Band des Berliner Wörterbuchs erschien.⁸ Es hat schließlich nach der Einstellung der Arbeiten am großen Berliner Wörterbuch Lexikographie gegeben. Ich sollte hier, wo

1 Siehe Erman† — Grapow, *Das Wörterbuch der ägyptischen Sprache*.

2 Champollion, *Dictionnaire égyptien en écriture hiéroglyphique*.

3 Birch, "Dictionary of Hieroglyphics", in *Egypt's Place in Universal History*, Bd. V, 335-586.

4 Brugsch, *Hieroglyphisch-demotisches Wörterbuch*.

5 Siehe weiter Schenkel, *Studi epigrafici e linguistici sul Vicino Oriente antico* 12 (1995), 191-203.

6 Budge, *An Egyptian Hieroglyphic Dictionary*.

7 Erichsen, *Demotisches Glossar*.

8 Faulkner, *A Concise Dictionary of Middle Egyptian*.

es nicht um bibliographische Vollständigkeit geht, wenigstens die besonders nützlichen Wörterbuchergänzungen nennen, die Dimitri Meeks ein paar Jahre lang in seiner *Année lexicographique* einsammelte,⁹ auch Leonard Leskos „(A) Dictionary of Late Egyptian“ nicht vergessen¹⁰ und als jüngstes Produkt das „Große Handwörterbuch“ von Rainer Hannig anführen.¹¹

Was alle diese lexikographischen Arbeiten vor, neben und nach dem Berliner Wörterbuch von dem Akademie-Unternehmen unterscheidet — trotz gewisser Gemeinsamkeiten letztlich auch Leskos „Dictionary“ —, war der leitende Gedanke des direkten Bezugs auf das Textkorpus, die Kombination von verantworteter philologischer Textbearbeitung und Lexikographie, die Erarbeitung des Wörterbuchs aus dem Gesamtbestand des altägyptischen Textbestandes. Ich will nicht verhehlen, dass dieser globale Ansatz auch seine Schwierigkeiten hatte: Das schiere Textvolumen der Texte in hieroglyphischer und hieratischer Schrift war so groß, dass man von vornherein auf das Demotische und das Koptische weitestgehend verzichtete, obwohl doch die ägyptisch-koptische Sprache in linguistischer Hinsicht eine Einheit bildet und ohne jeden Zweifel der Philologe des Hieroglyphisch- und Hieratisch-Ägyptischen selbst noch aus dem koptischen Wörterbuch Belehrung erfährt, gar nicht zu reden vom Demotischen, dessen Absonderung vielleicht der größte Fehler war, den man in Berlin beging und doch begehen musste, um überhaupt die (damals) zentrale Aufgabe der hieroglyphischen und hieratischen Lexikographie bewältigen zu können. Schwerer noch wiegt, dass man selbst die zentrale Aufgabe der Erschließung und Auswertung der hieroglyphischen und hieratischen Texte nur näherungsweise bewältigte. Den ersten lexikographischen Auswertungsversuchen lagen Texte im Umfang von kaum mehr als einer Million laufender Wörter zugrunde. Während der Phase der lexikographischen Auswertung wurden dann noch einmal eine gute halbe Million Texte, wenn auch mit nachlassender Intensität in die Arbeit einbezogen. Im Grunde genommen war das ägyptische Textkorpus schon damals mit den vereinten Kräften der internationalen Ägyptologie in vernünftigen Zeiträumen nicht zu bewältigen. Dennoch: Trotz aller Abstriche, die der erste Realisierungsversuch zeitigte, bleibt die Koppelung von philologischer Arbeit an den Texten und lexikographischer Auswertung ein

⁹ Meeks, *Année Lexicographique*.

¹⁰ Lesko — Switalski-Lesko — Thompson (ed.), *A Dictionary of Late Egyptian*.

¹¹ Hannig, *Großes Handwörterbuch*

hohes Ideal. Denn Texte mit all ihren philologischen Schwierigkeiten, in philologischer Arbeit zu bewältigenden Unklarheiten sind praktisch die einzigen Quellen, über die man bei einer toten Sprache verfügt und aus denen man das Lexikon erschließen muss. Die Frage bleibt allerdings: Lässt sich ein solches Ideal, das Erman und eine Generation der bedeutendsten Philologen nicht recht bewältigen konnten, überhaupt bewältigen, heute bewältigen, wo sich das Textmaterial gegenüber damals gewaltig vermehrt hat und die Ägyptologie in ganz anderem Ausmaß als damals auch ganz andere Interessen verfolgt als die philologische Kärnerarbeit?

1.2 Andere Konzepte und das Berliner Konzept

Es kann keinen Streit darüber geben, dass es auf dem Gebiet der ägyptischen Lexikographie unterschiedliche Marschrichtungen gibt und geben muss. Ich habe selbst in einer früheren Diskussionsrunde am selben Ort, als es um die Zukunft des (Alt-)Ägyptischen Wörterbuchs nach der Wende ging, sehr dezidiert dafür plädiert, die Aufgabe der ägyptischen Lexikographie in Bereiche aufzugliedern, die nebeneinander bestehen können und sich wechselseitig stützen sollen.¹² Ein Handwörterbuch ist ebenso notwendig wie ein Thesaurus, Spezialwörterbücher sind ebenso wünschenswert wie Gesamtübersichten. Ich brauche das nicht auszuführen. Dafür steht das Programm des Kolloquiums, über das dieser Band berichtet. Jeder der hier vorgetragenen Ansätze verdient erwogen zu werden. Dass man dann vielleicht das eine besser findet als das andere, versteht sich in einer lebendigen Wissenschaft von selbst.

Was das aktuelle Berliner Projekt „Altägyptisches Wörterbuch“ angeht, so sieht es sich in der Tradition des alten Berliner Wörterbuchs, insoweit jedenfalls, als es das Lexikon in Beziehung zu den Texten selbst in ihrem vollen Umfang stellt. Das ist natürlich nur dann sinnvoll, wenn sich die alten Probleme bei zusätzlich noch bedeutend vermehrtem Textmaterial und bei zusätzlich noch gesteigerten Ansprüchen lösen lassen. Selbstverständlich ist die heutige Ausgangssituation nicht mehr die von 1897. Wir haben es im ausgehenden 20. und beginnenden 21. Jh. nach der harten Arbeit mehrerer Generationen von Philologen heute in vielerlei Hinsicht leichter als die Heroen des ausgehenden 19. und beginnenden 20. Jh. Was aber bleibt, ist das schiere Textvolumen, das zu bewältigen heute eher noch schwieriger, um nicht zu sagen: hoffnungsloser ist als vor 100 Jahren.

¹² Schenkel, *ZÄS* 121 (1994), 154-59.

Was unsere Situation von der von 1897 radikal unterscheidet, ist dies: dass wir anders als die Altvorderen den Computer haben und das Internet. Der Akzent ist dabei auf Internet zu legen. Gewiss hat es Vorteile, Textdaten mit dem Computer zu erfassen und zu verwalten. Entscheidend wichtig aber ist in unserem Zusammenhang die Möglichkeit des Datenaustauschs von Computer zu Computer. Auf diese Weise wird es nämlich möglich, statt Fertigprodukte nach langen Vorbereitungszeiten — als gedrucktes Buch — irgendwann in ferner Zukunft der wissenschaftlichen Öffentlichkeit zu übergeben, sofort ohne größere Umstände, bereits bei laufender Arbeit, Einblicke in das bereits Erarbeitete zu gewähren. Wann und ob ein Abschluss erreicht wird, ist keine zentrale Frage mehr — es sei denn für den Geldgeber. Es gibt andere Techniken des Datenaustauschs, namentlich CD-ROM, allen ist der Austausch im Internet vorzuziehen, weil er der Aktualität und der Spontaneität die größten Spielräume eröffnet.

Diesen neuen Möglichkeiten Rechnung tragend wurde die lexikographische Arbeit in der Arbeitsstelle Ägyptisches Wörterbuch in folgender Weise neu begonnen bzw. werden interimsmäßig die Ergebnisse der früheren Arbeiten neu organisiert:

- *Textdatenbank im Internet-Zugriff als Ersatz für das gedruckte Großwörterbuch:* Die Arbeitsstelle hat als Ziel die Erstellung einer möglichst umfassenden ägyptischen Textdatenbank. Sie arbeitet nicht an einer Neuauflage des Erman/Gradow. Vielmehr tritt die Textdatenbank künftig an die Stelle des gedruckten Großwörterbuchs. Der Zugang erfolgt über das Internet (bzw. interimsmäßig u.U. über CD-ROM). Vgl. hierzu Stefan Grunert und Walter F. Reineke in diesem Band.
- *Datenbankverbund:* Die neue Berliner Datenbank umfasst derzeit Texte im Umfang von ca. 100.000 laufenden Wörtern. Der derzeit zu erwartende jährliche Zuwachs beträgt ca. 40.000-50.000 laufende Wörter Text. Um rascher dem Ziel entgegenzukommen, den gesamten ägyptischen Textbestand zu erschließen, wird ein Verbund mit anderen Datenbanken angestrebt. Der Bestand aller miteinander verbundenen Datenbestände soll über eine einheitliche Bildschirmoberfläche zugänglich sein. Vgl. hierzu Ingelore Hafemann in diesem Band.
- *Digitalisierung des alten Zettelarchivs:* Das alte Zettelarchiv im Umfang von ca. 1,7 Mio. laufenden Wörtern Text wird in den wesentlichen Teilen — ca. 1,4 Mio. Zettel — als eine (indizierte) Bilddaten-

- bank in den Datenbankverbund integriert. Vgl. hierzu Adelheid Burkhardt in diesem Band.
- *Textdatenbank als virtuelle Wörterbücher*: Die Textdatenbank wird so entwickelt, dass aus ihr bei Bedarf und nach den besonderen Bedürfnissen des Benutzers (Teil-)Wörterbücher abgeleitet und evtl. auf dem eigenen Drucker des Benutzers ausgedruckt werden können. Zu letzterem Punkt nun meine eigenen Ausführungen über das Thema „Textdatenbanken und/als virtuelle Wörterbücher“.

2 VIRTUELLES WÖRTERBUCH

2.1 Einleitung

Eine Textdatenbank ist kein Wörterbuch. Sie ist auch *per se* kein Ersatz für ein solches. Ein Wörterbuch *selektiert* aus einem *vor-klassifizierten*, vorzugsweise alphabetisch nach Lexemen geordneten Roh-Material aussagekräftige Vorkommen von Lexemen, und es bietet eine *Feinklassifikation* der Vorkommen der einzelnen Lexeme. Soll eine Textdatenbank an die Stelle des Großwörterbuchs in der Art des gedruckten Berliner Wörterbuchs der ägyptischen Sprache treten, so müssen erstens die Texte in der Datenbank auf die eine oder andere Art so aufbereitet werden, dass ein selektierender Zugriff Ansatzpunkte findet, und es muss zweitens zwischen den Benutzer und die Textdatenbank ein Zwischenglied eingeschaltet werden, das auf die eine oder andere Weise die Selektion und Feinklassifikation erledigt bzw. zum mindesten den Benutzer bei der Selektion und Neuklassifikation in eigener Regie und Verantwortung nachhaltig unterstützt.

Solange man es nur mit Textdatenbanken geringen Umfangs und geringer Detaillierung der Textbefunde zu tun hat, stellt sich das Problem nicht unbedingt. Man kann, ohne über Abfragestrategien allzu sehr ins Grübeln zu geraten, Fragen an die Datenbank richten und kann aus einfach strukturierten, überschaubaren Ergebnislisten ohne allzu große Beschwernis das heraussuchen, was man tatsächlich braucht. Große und anspruchsvoll erfasste Textkorpora lassen sich dagegen nicht leicht und sicher ohne zusätzliche Maßnahmen benutzen.

Zweck des Wörterbuchs ist nicht, einen Gegenstandsbereich, den ägyptischen Wortschatz *darzustellen*, sondern gesuchte Ausschnitte aus dem Gegenstandsbereich, z.B. ein einzelnes ägyptisches Wort/Lexem, seine Verwendungsweisen *aufzufinden*. Man liest ein Wörterbuch nicht als Bildungslektüre vom einen zum anderen Ende durch, sondern schlägt

in ihm nach. Ein gedrucktes Wörterbuch unterstützt den Suchvorgang durch mancherlei Maßnahmen: durch seine Kompaktheit, die die Übersicht erleichtert; durch die Möglichkeit der Annäherung an das Gesuchte durch rasches Blättern; durch gezielte Querverweise zu nicht ohne weiteres auffindbaren Einträgen. Einer üblichen Datenbank dagegen steht man ziemlich blind gegenüber: Sie ist unüberschaubar groß; sie verlangt den umweglosen Zugriff; sie setzt die genaue Kenntnis ihrer Notations-Konventionen voraus.

2.2 Aufbereitung der Texte für die Selektion

2.2.1 Die Suche nach indirekt Bezeugtem/Erschlossenem

Der Philologe sucht manchmal nach einzelnen Wörtern, die im Text positiv bezeugt sind, er sucht oft aber auch nach Aussagen, die nur irgendwie an einem Wort hängen. In letzterem Fall muss das betreffende Wort nicht positiv bezeugt sein, es kann sich aus dem Kontext ergeben.

Fall 1: Nicht vorhandene, weil verderbte Wörter

Beispiel 1:

it „(zweizeilige) Gerste“ und *bt.t* „Emmer“, die beiden Hauptgetreidearten, werden nicht ganz selten spielerisch durch im Verhältnis zueinander invertierte Zeichenfolgen dargestellt: $\overline{\text{it}}$ ist *it* „Gerste“, $\overline{\text{bt.t}}$ ist *bt.t* „Emmer“. *it* „Gerste“ ist phonographisch mit *t* geschrieben und als zweizeilige Frucht mit zwei Strichen für zwei Körner sowie der Getreideähre determiniert. *bt.t* ist logographisch mit der Getreideähre und der Femininendung *t* geschrieben und als zweizeilige Frucht ebenfalls mit zwei Strichen für zwei Körner determiniert. In CT VI 14d schreibt der Textzeuge B1Y *it* „Gerste“, meint aber, wie das folgende Attribut *tšr.t* „rot“ grammatisch und sachlich zeigt, *bt.t* „Emmer“. Nun könnte sich der Philologe z.B. ein Bild von solchen Verwechslungen machen wollen. Dann muss aber die Textdatenbank in irgendeiner Weise Zugang zur positiv bezeugten Lesung und zur durch Emendation hergestellten eröffnen, am besten zu beiden auch über einen gemeinsamen Suchweg.

Beispiel 2:

In CT II 4a hat der Textzeuge B1C *Tfn.wt*, die Schwester des Schu, von der alle anderen Textzeugen sprechen, sinnwidrig in die Himmelsgöttin *Nw.t* verderbt. Selbstverständlich soll das in Rede stehende Wort unter *Tfn.wt* nachgewiesen werden, um die Tefnut betreffende Aussage nicht verloren gehen zu lassen. Aber der positive Befund ist nun einmal *Nw.t*, sinnlos oder — theoretisch nie definitiv ausschließbar — auch einmal

eine besondere Überlegung des Abschreibers. Im übrigen ist die Verschreibung als solche gewiss für den textkritisch arbeitenden Philologen nicht ganz ohne Interesse.

Beispiel 3:

In CT VI 12g hat der Textzeuge B1Y den Gott „Osiris“ in die Bezeichnung des Verstorbenen als „Osiris NN.“ verderbt. Vermutlich wird man sich mehr für die im Archetyp anzusetzende Aussage über den Gott Osiris als die daraus verderbte über Osiris NN. interessieren. Setzt man, was zweckmäßig ist, den Gottesnamen „Osiris“ und „Osiris“ als Bezeichnung des Verstorbenen als zwei Lemmata an, genügt es nicht, in „Osiris NN.“ einfach „NN.“ zu ignorieren, man muss „Osiris NN.“ als ganzes in den Gottesnamen „Osiris“ emendieren.

Fall 2: Nicht vorhandene, weil zerstörte (oder versehentlich ausgelassene) Wörter

Analog zu Fall 1 sind zerstörte (und versehentlich ausgelassene) Wörter zu behandeln: Sie sind zu ergänzen, um ihren Kotext finden zu können.

Beispiel:

CT II 105b ist von einer Kopflaus („Laus des Kopfes“) die Rede, auch wenn das Wort „Laus“ nicht erhalten ist. Es ergibt sich jedoch die Ergänzung „Laus“ ohne weiteres aus der inhaltlich parallelen Stelle CT II 94a, wo *kt.t n.t tp* „Laus des Kopfes“ einwandfrei zu lesen steht. Würde man nicht ergänzen bzw. das ergänzte Wort als nicht positiv bezeugt bei der Abfrage der Datenbank nicht nachweisen, wäre man um eine Information über die Kopflaus ärmer.

Fall 3: Pronominale und deiktische Bezüge sind aufzulösen

Pronominalbezüge und deiktische Bezüge sind aufzulösen, durch direkte Zuordnung der Auflösung oder durch Verweis auf den Ausdruck, auf den sich Pronomen oder Deixis beziehen.

Beispiel:

In den seltensten Fällen wird der Philologe nach geläufigen Götternamen um ihrer selbst willen suchen. Er sucht die Kotexte, in denen sie stehen. Kotexte sind aber auch die, in denen der betreffende Gott nur pronominal oder deiktisch — etwa: „dieser (in Rede stehende) Gott“ — vertreten ist. Ich brauche dies nicht mit konkreten Beispielen zu belegen.

Fall 4: Alternativlesungen sind alle zu notieren

Ist mehr als eine Lesung sinnvoll und eine Entscheidung zwischen den Lesungen objektiv schwierig, so wird nicht diejenige ausgewählt, der der Bearbeiter des Textes am ehesten den Vorzug gibt, sondern es werden

alle festgehalten, die er nicht plausibel und definitiv verwerfen kann. Ich brauche auch dies nicht konkret zu exemplifizieren.

Fall 5: Dummy-Wörter und Wortklassen

Ist die Lesung eines Wortes nicht möglich, so kann doch vielleicht festgestellt werden, dass ein Wort vorliegt und evtl. sogar noch, in welche Bedeutungsklasse es gehört.

Beispiel:

CT V 59b, wo Textzeuge S2C „some animal“ bezeugt. Da die anderen Textzeugen sich nicht über die Art des Tieres einig sind — T1C zeigt *rw* „Löwe“, Y1C zeigt das *iw*-Tier (De Buck anmerkungsweise: „perhaps ~~an~~ was meant“), B2Bo und B4Bo lesen ohne jede Tierhieroglyphe *Inp.w* „Anubis“, S10C vielleicht *Šth* „Seth“ (De Buck anmerkungsweise: „an animal, species uncertain“) — da also die anderen Textzeugen sich nicht einig sind, ist es schwierig, sich bei S2C für irgendeine der positiv bezeugten Lesarten zu entscheiden. Am besten notiert man in der Textdatenbank ein Dummy-Wort etwa in dem Sinne, wie dies De Bucks Fußnote indirekt tut, im Sinne also von „some animal“.

Hierzu ganz allgemein: Man interessiert sich nicht nur für Wörter, sondern auch für Wortklassen, besonders Bedeutungsklassen, z.B. für irgendeinen Gott in einem bestimmten Zusammenhang. Es ist die Aufgabe von Lemmalisten, die Zuordnung von Wörtern zu Klassen vorzunehmen.

In unserem Zusammenhang ist von Belang, dass an einer Textstelle zwar nicht das Wort selbst rekonstruiert werden kann, aber dessen Klassenzugehörigkeit. In solchen Fällen sind anstelle der konkreten Wörter Repräsentanten der Wortklasse einzusetzen.

Fall 6: Falsche Kotexte

Zum Thema „Wort und Kotext“ lässt sich noch Unangenehmeres berichten: Der Kotext kann auseinandergerissen sein.

Beispiel:

In der spaltenweise organisierten Schiffsteilliste von Spruch 398 der Sargtexte (CT V 125ff.) ist in einem Teil der Textzeugen eine ganze Reihe von Schiffsteilen gegenüber dem damit identifizierten göttlichen Wesen um eine Spalte verrutscht. Hat man De Bucks Edition der Liste vor Augen, sieht man das. Z.B. (CT V 135) assoziiert ein Teil der Textzeugen korrekt die Spalten *m: spr.wt* mit den Lippen von Isis und Osiris, ein anderer Teil der Textzeugen dagegen bringt fälschlich die Lippen der Götter mit dem Schiffsteil *wgy* zusammen, die Spalten dagegen mit dem göttlichen Wesen *Hčč.ti*. Arbeitet man in Zettelkastenmanier

— so Brigitte Altenmüller in ihrer Arbeit über den Synkretismus in den Sargtexten¹³— oder — unser Problem — mit einer Textdatenbank, kann einem die Textverderbnis leicht entgehen, und kann man in folgedessen auf Schiffsteilvergottungen kommen, auf die selbst kein Ägypter spekuliert hätte. Solche Fälle in einer Textdatenbank zu behandeln, ist nicht ganz einfach, aber absolut notwendig, wenn man sich nicht zum Ziel gesetzt hat, den Benutzer irrezuleiten.

Das Thema „Wort und Kotext“ abschließend: Nicht zuletzt für das sehr ergiebige Verfahren der kombinierten Suche („combined search“), der Suche nach benachbarten Vorkommen von zwei (oder mehr) Textelementen, sollten die geforderten Vorarbeiten am Text die Voraussetzungen schaffen.

2.2.2 Kotext als lineare oder strukturelle Nachbarschaft

Für die Textrekonstruktion und für die Textinterpretation ist es von eminenter Wichtigkeit, Parallelen mit gleichartigen Wort-Kollokationen zu finden. Als erfahrenen Philologen ist dies bereits den Bearbeitern des Berliner Wörterbuchs aufgefallen, die im Fortgang der Arbeit in zunehmendem Maß sich mit den Kollokationen befasst haben. Zwei Typen solcher Kollokationen sind Abhängigkeit und Parallelismus. Abhängigkeit liegt z.B. vor in der Verbindung von Verben mit bestimmter Präposition, Parallelismus z.B. in koordinierten Wörtern oder in semantischen Entsprechungen zwischen Sätzen; z.B. Synonymen und Antonymen im Parallelismus membrorum.

Beispiel:

Die Bearbeiter des Berliner Wörterbuchs haben sich für solche Kollokationen interessiert. So haben sie etwa die Sequenz *p.t* „Himmel“, *t3* „Erde“, *mw* „Wasser“, *čw.w* „Berge“ beobachtet. Auch haben sie beobachtet, dass solche Elemente der Welt in parallelen Phrasen stehen können. Vergleicht man jedoch das Belegmaterial des Zettelarchivs mit den Belegnachweisen des gedruckten Wörterbuchs, kann man sich wundern. Eigentlich würde man erwarten, die Kollokationen unter jedem Element in gleicher Weise nachgewiesen zu finden. Ich exemplifiziere den Befund an der Sequenz *p.t* „Himmel“, *t3* „Erde“, *tʷ3.t* „Unterwelt“, *mw* „Wasser“, *čw.w* „Berge“. In Band I des Wörterbuchs, bei *p.t* „Himmel“, wäre die erste Gelegenheit gewesen, die Kollokation zu belegen. Tatsächlich wird die Kollokation aber zum ersten Mal in Band II, bei *mw* „Wasser“, als nachweiswürdig erkannt, um in Band

¹³ Altenmüller, *Synkretismus in den Sargtexten*.

V unter *tʷʒ.t* „Unterwelt“ die intensivste Belegung zu erfahren. Danach war die Kollokation für die Wörterbuchbearbeiter nicht mehr interessant. Unter *čw* „Berg“ findet sich die frühere Belegdürre. Ein Glücksvogel der Philologe, der sich der „Unterwelt“ zuwendet und nicht dem „Himmel“.

Noch eine andere Beobachtung am Berliner Wörterbuch: Ist ein Wort insgesamt häufig belegt, wird die Kollokation schlecht wahrgenommen. Ist ein Wort insgesamt weniger häufig belegt, wird die Kollokation gut wahrgenommen. Z.B. wird in Band I bzw. II beim häufigeren *p.t* „Himmel“ die Kollokation übersehen, beim selteren *mw* „Wasser“ dagegen beobachtet. Ähnlich wird in Band V beim sehr häufigen *tʒ* „Erde“ die Kollokation weit weniger gut dokumentiert als beim vergleichsweise sehr viel selteneren *tʷʒ.t* „Unterwelt“. (Bei *čw* „Berg“ war, wie gesagt, das Interesse erlahmt.)

Der langen Rede kurzer Sinn: Die konventionelle Lexikographie ist mit der Forderung nach egalitärem Zugang zu Kollokationen von jedem der gleichwertigen daran beteiligten Elemente aus nicht gewachsen. Eine Textdatenbank, die ihre Auskünfte nicht vorkonfektioniert, sollte in diesem Punkt mehr leisten können.

Wie gut oder schlecht man interessante Kollokationen in einer Textdatenbank aufspürt, hängt u.a. und vor allem davon ab, wie man die Texte bei der Aufnahme strukturiert. Was in einem Dependenzverhältnis stehende Elemente angeht, setzt eine gezielte Suche die Notation der Abhängigkeitsverhältnisse voraus, also eine syntaktische Notation. Was Kollokationen parallel stehender Elemente angeht, ist darüber hinaus noch einem Standard-Strukturprinzip ägyptischer Texte Rechnung zu tragen, dem Parallelismus membrorum, der „metrischen“ Struktur der Texte, dem Vers- und Strophenbau.

Ein praktisches Problem ist: Nach welchen Kriterien begrenzt man die Textwortfolge, innerhalb derer man Kollokationen sucht. Man kann z.B. innerhalb von Sätzen (Hauptsätzen) suchen. Man kann aber auch innerhalb von Verspaaren, Verstripeln, Versquadrupeln suchen wollen, vor allem abhängig davon, ob man sich mehr für syntaktische oder mehr für semantische Aspekte des Textes interessiert. Satz- und Versgruppenstruktur überlagern sich: Eine Versgruppe kann aus einem oder aus mehreren Sätzen bestehen. Ein Satz kann Teil eines Verses sein, ein ganzer Vers, eine Versgruppe oder eine Folge von Versgruppen oder kann Versgruppen in sich enthalten (typisch: Reihen von Epitheta als Attribute zu einem Substantiv).

Schlussfolgerung: Idealerweise wird ein Text in einer Textdatenbank doppelt strukturiert: syntaktisch und „metrisch“.

2.2.3 Mehrfachbezeugung: *Realer Textzeuge vs. Archetyp*

Ist ein Text durch mehrere Textzeugen belegt, so ist nach den Gegebenheiten der Textüberlieferung und nach dem Stand der ägyptologischen Forschung nur ein Verfahren praktikabel: Jeder Text wird als gesonderter Text in die Textdatenbank aufgenommen. Es genügt im allgemeinen nicht, irgendeinen (und sei es den besten) Textzeugen aufzunehmen. Und es ist weder sinnvoll noch machbar, sich auf einen rekonstruierten Archetyp zu beschränken. Die Textzeugen variieren dazu in ihrem Wortlaut und — vor allem! — in ihrem Erhaltungszustand zu sehr. Es muss umstandslos erkennbar sein, was positiv bezeugt ist.

Auf der einen Seite muss also der Zugriff auf alle Besonderheiten der einzelnen Textzeugen möglich sein. Auf der anderen Seite sollte der Fragesteller vor redundanten Antworten bewahrt werden für den Fall — den Normalfall —, dass Textzeugen miteinander übereinstimmen. Dieses Problem ist nur dadurch lösbar, dass man Gleichlautendes Wort für Wort in Beziehung zueinander setzt, desgleichen das, was als Variante zueinander gelten kann, explizit als solche kennzeichnet.

Ich brauche das Problem der Mehrfachbezeugung nicht zu exemplifizieren. Ich möchte aber doch noch auf ein Randproblem hinweisen, mit dem die beste Parallel-Ausgabe von Texten ihre Schwierigkeiten hat: Es gibt Textfassungen, die so weitgehend voneinander abweichen, dass man sie als ganze nicht recht parallelisieren kann. Z.B. kommt es bei De Bucks mustergültiger Sargtextausgabe vor, dass er einzelne Textfassungen der „Standard“-Textfassung als eigenen Text (Spruch) folgen lässt. Dabei ist es durchaus möglich, viele gleichlautende Teil-Sequenzen miteinander zu identifizieren. Fallweise sind Querverweise ausgesprochen hilfreich, z.B. wenn die „Standard“-Textfassung und singuläre Fassungen sich wechselseitig erhellen können.

Beispiel:

Der singuläre Spruch 143 ist eine Art Variante zu Spruch 137. Erkennbar sind 175a und 166f in irgendeiner Weise gemeinsamen Ursprungs. Keine der Textfassungen ist ohne weiteres klar, sie erhellen sich aber womöglich wechselseitig. Es kann z.B. nützlich sein zu wissen, dass anstelle eines Wortes, das *prima vista* als *rt.t* erscheint, die andere Textfassung *prima vista* *hrꜥ.t* zeigt.

In einer herkömmlichen Textausgabe, d.h. auf dem zweidimensionalen Papier, sind solche Zusammenhänge — die dem aufmerksamen

Herausgeber vielfach aufgefallen sind — schwer darstellbar, allenfalls durch fußnotenartige Querverweise. Im Computer sind die kompliziertesten Zusammenhänge darstellbar. Selbstverständlich ist auch am Bildschirm nicht beliebige Komplexität erträglich. Man hat aber die Option, komplexe Daten bei und nach Bedarf zu zerlegen und auszugsweise vorzuführen. So wäre es im Beispielfall durchaus praktikabel, die Zitierabschnitte De Bucks aus „Standard“-Textfassung und „Sonder“- Fassungen aus den Texten herauszulösen und am Bildschirm nebeneinander zu montieren.

2.3 *Der Mensch-Maschine-Dialog*

2.3.1 *Übersetzung zwischen Normen*

Der Benutzer sucht linguistische Einheiten, nicht Zeichenketten. Linguistische Einheiten können auf unterschiedliche Weise in Zeichenketten abgebildet werden und werden auch sinnvollerweise von Fall zu Fall unterschiedlich in Zeichenketten abgebildet. Ich lasse das Problem der Abbildung der hieroglyphischen Graphien beiseite und beschränke mich auf die einfachere Frage der Abbildung von linguistischen Einheiten in eine lateinische Transkription.

Beispiel 1:

Es sollte belanglos sein, welches der existierenden oder noch zu erfindenden *Transkriptionsalphabete* der Benutzer wählt. Sagt er dem Computer, welches er wählt, kann computerintern die Anfrage in dasjenige Transkriptionsalphabet umcodiert werden, das die Datenbank benutzt, und die Antwort aus dem Transkriptionsalphabet der Datenbank in dasjenige des Benutzers umcodiert werden.

Beispiel 2:

Es muss erlaubt sein, linguistische Einheiten in unterschiedlicher Lautform abzufragen, schon deshalb, weil sich die Lautformen im Laufe der Sprachgeschichte ändern. Es ist fallweise zweckmäßig, *s* und *ś* voneinander zu unterscheiden, und fallweise nicht. Es ist sinnvoll, ein Wort in seiner AR-Form abzufragen — *s* und *ś* sind dann zu unterscheiden — oder in seiner NR-Form — *s* und *ś* sind nicht zu unterscheiden. Man kann auf die Unterscheidung verzichten wollen, wenn einmal — nicht selten der Fall — nicht klar ist, welches der genaue Lautansatz ist, oder man kann umgekehrt auf der Unterscheidung beharren wollen, wenn man dadurch Material mit dem nicht zutreffenden *s*-Laut ausschließen und damit das Ergebnis der Anfrage von unnützem Ballast freihalten kann.

Beispiel 3:

Es ist zweckmäßig, fallweise mehr oder weniger morphematische Strukturzeichen zu verwenden. Es kann fallweise günstig sein, ein Kausativpräfix *s*- oder ein *i*-Präfix abzutrennen und fallweise nicht.

In Fällen wie den in Beispiel 2 und 3 genannten — *s*-Laute, Präfixe — hat ein *Regelwerk*, ein *Programm* die Koordination zu leisten.

Beispiel 4:

Es ist sinnvoll, ein hieroglyphisches Wort in seiner demotischen Lautgestalt zu erfragen wie ein demotisches in seiner hieroglyphischen Gestalt. Es ist sinnvoll, ein Wort mit seinen tatsächlich belegten Konsonanten abzufragen wie in einer um Endungen wie *w* oder *i* erweiterten urkoptischen Rekonstruktion.

In solchen Fällen erfolgt die Koordination über Listen, Wortlisten, Lemmalisten bzw. über Links zwischen Listen. Hierzu der Beitrag von Ingelore Hafemann in diesem Band.

*2.3.2 Die Formulierung einer ergiebigen Frage**2.3.2.1 Die Begrenzung der Antworten durch Präzisierung der Fragestellung*

Zuviel Information ist keine Information.

Ein Teil meiner Forderungen gilt nur für den Fall, dass als Antwort auf eine Datenbankanfrage viele Antworten kommen. Ist ein Wort nur zehnmal im Textkorpus enthalten oder ist eine Kollokation nur zehnmal belegt oder ist ein Wort nur an einer einzigen Textstelle in zehn Textzeugen überliefert, so lese ich das Ergebnis einfach durch und segregiere mit den Augen bei Bedarf Nachweise, die ich infolge zu allgemeiner Fragestellung unnützerweise erhalten habe. Aber was, wenn ich 10.000 Einzelwortbelege erhalte oder Tausende von Kollokationen oder durch Dutzende von Textzeugen aufgeschwemmte Ergebnislisten? Nun ist es ja keineswegs so, dass sich der Philologe für häufig belegte Wörter grundsätzlich weniger interessiert als für selten belegte. Selbstverständlich hat der Nachweis seltener Wörter erhebliches Interesse für ihren ersten Nachweis und für die Bestimmung ihrer Bedeutung. Aber der Philologe interessiert sich genauso für geläufige Wörter, nicht weil er sie als solche nachweisen und erfassen will, sondern weil er sich für ihre Verwendung interessiert, für ihre Verwendungsbedingungen, ihren Kontext.

Das gedruckte (Alt-)Ägyptische Wörterbuch löst das Problem durch zwei Maßnahmen: durch die Feinsortierung der Wortbelege nach Formenbildung, nach syntaktischen Rektionen, nach semantischen

Kollokationen u.dgl. und durch — eher willkürliche — Reduktion der Belege auf eine „repräsentative“ Auswahl. Entsprechend der eindimensionalen Abfolge der Belege im gedruckten Wörterbuch hat im allgemeinen jeder Beleg eine ganz bestimmte systematische Stelle.

Beispiel:

Entweder werden die Belege zu einem Substantiv zuerst nach dem Numerus geordnet und danach innerhalb der Belege eines jeden Numerus nach „Verwendungen“ oder aber man sortiert sofort nach „Verwendungen“ und lässt die Numeri bei Gelegenheit einfließen. Z.B. verfährt das gedruckte Berliner Wörterbuch beim Lemma *p.t* „Himmel“ anders als beim Lemma *rn* „Name“, obwohl beide Substantive Numeri unterscheiden und diverse standardmäßige „Verwendungen“ haben. Bei *p.t* „Himmel“ wird zuerst nach dem Numerus geschieden, danach kommen die charakteristischen „Verbindungen“ wie etwa die Verbindung mit einem Attribut (*p.t tn* „dieser Himmel“). Bei *rn* „Name“ dagegen wird die Verbindung mit einem Attribut bzw. das Fehlen eines Attributs zum obersten Ordnungsprinzip bestimmt, Numeri kommen dann beiläufig zur Sprache (§§ *rn.w* „mit vielen Namen“). Man hat sich dabei etwas gedacht.

Aus einer Textdatenbank kann man ebenfalls solche Sortierungen gewinnen, sofern die entsprechenden Merkmale entweder kodiert sind (z.B. Numerus) oder aus dem Kotext gewonnen werden können (Kollokationen). Der Unterschied liegt aber darin, dass bei der Textdatenbank-Anfrage die Merkmale und die Hierarchie ihrer Berücksichtigung nicht für jedes Wort ein für alle Male festgelegt sind, sondern zum Zeitpunkt der Anfrage bestimmt werden können bzw. müssen, unter den möglichen bzw. denkbaren ausgewählt werden können bzw. müssen.

Vorausgesetzt, in der Textdatenbank sind den Textwörtern vielerlei Merkmale zugeordnet oder es lassen sich per Programm solche Merkmale generieren: Wie kann der Benutzer wissen, mit Hilfe welcher Merkmale er seine Anfrage präzisieren kann, um den Output zu vermindern, auf seine tatsächlichen Bedürfnisse zu reduzieren? Und wie kann der Benutzer wissen, welche Merkmalkombination für seine Zwecke am ergiebigsten ist. Wie also kann er sich vor nutzlosen Materialmassen schützen?

Viererlei ist notwendig:

1. Der Benutzer muss am Bildschirm darüber informiert werden, welche Möglichkeiten er hat, das Material zu unterteilen.
2. Der Benutzer muss vor Augen geführt bekommen, wie umfangreich die Teilmengen sind, die er mit den von ihm ausgewählten (im Ma-

terial explizit codierten) Kriterien selektiert.

3. Der Computer muss dem Benutzer vorschlagen können, welche (im Material explizit codierten) Kriterien er wählen kann, um überschaubare Teilmengen zu gewinnen.
4. Der Computer sollte nach mengenmäßig geeigneten Kollokationen Ausschau halten.

Punkt 4 ist extrem schwierig zu realisieren, so dass man einstweilen am besten zunächst noch darauf verzichtet. Die Punkte 1-3 dagegen sind ein unbedingtes Muss: Auskunft über mögliche Selektionskriterien, Berechnung des Umfangs selektierbarer Datenmengen, Vorschlag effektiver Selektionskriterien.

2.3.2.2 Die Vermehrung der Antworten durch Ausweitung der Fragestellung

Umgekehrt können Anfragen, weil zu eng formuliert, auch weniger Antworten erbringen als wünschenswert.

Beispiel:

In CT VI 5d zeigen die meisten Textzeugen ein Partizip *ššn.t*, gefolgt von einem enklitischen Pronomen *ší*, also *ššn.t ší* „die (Frau), die sie (d.i. einen Schiffsteil) zwirnt“. Der Textzeuge B6Bo dagegen versteht *ššn.t* als Substantiv, folglich =*š* als Suffixpronomen, also: *ššn.t=š* „ihre (d.i. eines Schiffsteils) Zwirnerin“. (Jedenfalls kann man die Graphien so interpretieren.) In solchem Fall wäre es für den Datenbankbenutzer, wenn er sich für Textinhalte und nicht bloß für Vokabeln interessiert, sehr viel ergiebiger, statt nach dem Verb *ššn* „zwirnen“ oder nach dem Substantiv *ššn.t* „Zwirnerin“ nach der beiden zugrunde liegenden Wurzel *ššn* zu fragen (NB: nicht nach der sinnlosen Zeichenfolge / „regular expression“ *ššn**!). Der Computer kann ihn auch auf diese Möglichkeit hinweisen, evtl. als Antwort auf die Anfrage, ob der Computer eine sinnvolle und mengenmäßig zu bewältigende Möglichkeit zur Erweiterung der Fragestellung sieht. Grundsätzlich ist das möglich durch den Rückgriff auf die Lemmalisten, in denen den Lemmata Merkmale zugeordnet werden können, z.B. die Zugehörigkeit zu einer Wurzel.

3 DIE BERLINER TRADITION

Das virtuelle Wörterbuch ist das Berliner Großwörterbuch mit modernen Mitteln: Die Wortbelege werden nach wie vor sortiert, nicht jedoch mehr in eine einzige starre Sequenz, sondern fallweise und nach den jeweiligen Bedürfnissen in unterschiedliche Sequenzen. Die Belege werden nach wie vor selektiert, nicht jedoch definitiv und ein für alle

Male, sondern fallweise und nach den jeweiligen Bedürfnissen, nach wechselnden Kriterien. Die Frage- und Antwortmöglichkeiten sollten dadurch vielfältiger werden. Leider wird die notwendige Vorarbeit bei der Aufnahme des Wortlautes und bei der Zuordnung von abfragbaren Merkmalen zum Wortlaut dadurch nicht geringer.

Ich will dies denen, die die Wörterbuchzettel kennen, noch einmal anders sagen: An die Stelle der blauen Kreuze, mit denen die Altvorderen interessante Belegzettel kennzeichneten, treten bunte Kreuze, Kreuze in vielen Farben, mit deren Hilfe man nicht Material als in irgendeiner Weise interessant abrufbar macht, sondern mit denen man Material unter vielerlei wechselnden und kombinierten Gesichtspunkten verfügbar macht.

4 SCHLUSS

Textdatenbanken *und* virtuelle Wörterbücher — Textdatenbanken *als* virtuelle Wörterbücher. Die „als“-Version des Titels meines Beitrags ist das Ziel: „Textdatenbanken *als* virtuelle Wörterbücher“, die Textdatenbanken *sind* virtuelle Wörterbücher. Die „und“-Version steht für das in überschaubarer Zukunft Erreichbare: „Textdatenbanken *und* virtuelle Wörterbücher“, die Textdatenbanken sind noch nicht das voll durchstrukturierte Instrument der virtuellen Wörterbücher, zeigen aber bereits handfeste Ansatzpunkte für die Weiterentwicklung zu virtuellen Wörterbüchern. Um das Machbare zu realisieren muss man das Utopische fordern: „Textdatenbanken *als* virtuelle Wörterbücher“.