

WOLFGANG SCHENKEL

Das Tübinger Konkordanz-Programm*

Im Thema meines Referats ist ein „Tübinger Konkordanz-Programm“ angekündigt oder in Erinnerung gerufen. Ich habe die Formulierung nicht selbst gewählt, kann ihr aber einen Sinn abgewinnen. Dieser Sinn ist nicht unbedingt der Sinn, den der Leser spontan erwartet.

Um es ohne Umschweife zu sagen: Es gibt kein Tübinger Computer-Programm zur Erstellung von Konkordanzen oder, vielleicht besser: Es gibt ein solches Programm nicht mehr. Was man in der Tat als ein solches Programm bezeichnen könnte, das alte Darmstädter Programmsystem M.A.A.T.¹, das ich später in Göttingen und auch noch in Tübingen benutzt, fallweise ausgebaut und an meine Computer-Peripherie angepaßt habe², existiert nicht mehr, jedenfalls nicht mehr in einer lauffähigen Form. Es hat seine Dienste getan. Ich habe mein

* Für die Schriftform leicht adaptierte Fassung eines Vortrags auf der Internationalen Arbeitstagung „Ägyptisches Wörterbuch“, Berlin 3.–5. 9. 1992.

¹ Wolfgang Schenkel, Maschinelle Analyse altägyptischer Texte, Aufbau und Grundlagen des Systems M.A.A.T. und seine lexikographische Anwendung, Schriftenreihe des Deutschen Rechenzentrums S-7, Darmstadt 1969; Rolf Gundlach/Wolfgang Schenkel, Lexikalisch-grammatische Liste zu Spruch 335 A der altägyptischen Sargtexte LL/CT.335A als Probe für die Erstellung einer Sargtextkonkordanz mit Hilfe einer elektronischen Datenverarbeitungsanlage, Schriftenreihe des Deutschen Rechenzentrums S-8, Darmstadt 1970.

² Wolfgang Schenkel, Zur Fortführung des Projektes M.A.A.T. in Göttingen, in: GM 2, 1972, S. 33–36; id., Die Göttinger und Tübinger Konkordanz zu den altägyptischen Sargtexten, in: G. Koch (ed.), Datenverarbeitung in den Geisteswissenschaften, 21. Bericht der Gesellschaft für wissenschaftliche Datenverarbeitung mbH, Göttingen 1981, S. 13–33; id., Aus der Arbeit an einer Konkordanz zu den altägyptischen Sargtexten, GOF IV 12, Wiesbaden 1983.

<i>isf</i> (<i>isf.t</i>)	1	<i>isf</i> (<i>isf.ti</i>)
<i>isf</i>¹		<i>isf</i> (<i>isf.ti</i>)
<i>isf.t</i> »Unrecht«		
Sg.		<i>isf.ti</i> »Übeltäter«
<i>i</i> {w}sf.t ^{nčs} III 317r T2L		Pl.
<i>i</i> {w}sf.t ^{nčs.P} IV 210-1a T2Be		<i>i</i> {w}sf.t ^{ts} w ₁ ^{sl.P} IV 319b T2Be
<i>i</i> {w}sf.t ^{nčs} II 139a G1T		<i>i</i> {w}sf.tw ₁ ^{sl.P} IV 300b {T2Be}. <i>i</i> {w}f[s]f.tw ₁ ^{sl.P} IV
<i>isf.t</i> VII 226h pGardII		254-5b T2Be
<i>isf.t</i> ^{wt} IV 210-1a Sq1Sq		<i>i</i> {s}f ^l .tw ₀ ^{nčs.Q} IV 300b B1Y
<i>isf.t</i> ^{mčt} II 139a G2T		<i>i</i> s ^l f ^l .tw ₀ ^{nčs.V.P} IV 319b B1Y
<i>isf.t</i> ^{nčs} I 173c B16C; II 149d G2T, S1C; III		[i]sf ^l .tw ₀ ^{sl.P} IV 254-5c Sq8Sq(N)
297f G1T; IV 210-1a T1C ^b , M1C; 254-5b		<i>isf.tw₀^{[sn].V.P} III 358c S1C^a</i>
Sq4Sq; VI 136k M22C; 153e B2Bo; 165f B1Bo;		[i]sf.tw ₀ ^{[Q].P} IV 300b Sq2C(N)
241o T1C ^a ; 267x G1T; 347e B3Bo; VII 462c		<i>isf.tw₀^{V.P} VII 100q S10C. <i>is</i>(f).tw₀^{V.P} III 358c</i>
B1Bo, B3C, B9C; 464a B1Bo, B3C, {B4C}.		S1C ^b . <i>i</i> [sf]. ^{nčs.V.P} III 358c S2C ^a . <i>i</i> ₁ f[sf.tw] ₀ ^{V.P} III
<i>isf.t</i> ^{nčs} I 173c {B13C ^b }. <i>i</i> ₁ f ^l .t ^{nčs} VI 278e TT319.		358c S2C ^b
<i>isf.t</i> (^{nčs}) VI 153e B1Bo. <i>i</i> ₁ f[sf].t ^{nčs} II 149d S2C		<i>i</i> ₁ f ^l .t ^{nčs} tw ₁ ^{thfsl.P} IV 319b T1C ^b
<i>isf.t</i> ^{nčs.mčt} I 174f {B13C ^b }, B16C		<i>isf.t</i> ^{ts} w ₁ ^{V.P} V 324f B2L
<i>isf.t</i> ^{nčs.P} IV 210-1a T1C ^a		<i>isf.tw₁^{nčs.P} IV 254-5b B1P. [i]sf.tw₁^{nčs.P} IV 254-5b</i>
<i>i</i> ₁ f ^l .t st II 139a S2C ^a		B1Y
<i>isf.t</i> ^{sth/A} II 149b pGardII		<i>isf.tw₁^{sl.P} IV 254-5b B5C</i>
<i>isf.t</i> (^Q) III 297f A1C. <i>i</i> ₁ f ^l .t ^{l[Q]} VI 165f BH3Ox.		<i>isf.tw₁^{ts.V.P} V 323j B2L</i>
<i>is</i> [f.t] ^Q IV 210-1a Sq4Sq		<i>isf.tw₁^P IV 254-5b {T1C^a}; VI 242d S14C</i>
<i>isf</i> {s}.t ^{nčs} II 139a S1C		<i>isf.tw₂^{sl.P} VII 467a B3C</i>
<i>isf.t</i> VII 2f { ⁿ L2Li ⁿ }		<i>isf.tw₂^{ts.V.P} V 323j B17C</i>
<i>isf.t</i> ^{mčt} IV 210-1a B1Y		<i>isf.tw₃^{nčs.V.P} IV 300b B3C</i>
<i>isf.t</i> ^{nčs} II 139a A1C; IV 62g BH4C; 210-1a		<i>isf.tw₃^{nčs.P} IV 300b M4C</i>
M4C, B9C ^a , B9C ^b , {L3Li ⁿ }; VI 96f B9C, B10C;		<i>isf.tw₀^[P] IV 254-5b {Sq7C}</i>
153e B4C; 373k B4C; VII 462c B1L; 464a B6C,		<i>isf.tw₀^{mwt/2.P} IV 300b T3Be, M57C</i>
B9C, B1L. <i>isf.t</i> ^{nčs} VII 462c B6C. <i>isf</i> ^l .t ^{nčs} IV		<i>isf.tw₀^{nčr.P} VI 242i {ⁿL1Liⁿ}</i>
210-1a M1NY. <i>is</i> (f).t ^{nčs} IV 210-1a BH1Br.		<i>isf.tw₀^{nčr.[.].P} VI 242e {ⁿL1Liⁿ}</i>
<i>isf</i> ^l .t ^{nčs} IV 210-1a T3L. <i>i</i> ₁ f[sf].t ^{nčs} VII 464a		<i>isf.tw₀^{sl.P} IV 254-5b M7C, M8C</i>
B1C		<i>isf.tw₀^{ts.P} IV 254-5b M57C</i>
<i>isf.t</i> ^{nčs.P} IV 49k { ⁿ L1Li ⁿ }; 210-1a T3Be, M8C,		<i>isf.tw₀^P IV 254-5b T1Be; 300b T1Be; 319b T1Be,</i>
M57C, Sq7C; 254-5b BH1Br. <i>isf.t</i> ^{nčs.P} IV 210-1a		M1NY; VI 242d { ⁿ L1Li ⁿ }; VII 467a B1L.
Sq1C. <i>isf</i> ^l .t ^{nčs.P} IV 210-1a B5C		<i>isf</i> ^l .tw ₀ ^[P] IV 300b L1NY. <i>isf</i> ^l .tw ₀ ^[P] IV 254-5b
<i>isf.t</i> ^P IV 60p { ⁿ L2Li ⁿ }; 62g { ⁿ L1Li ⁿ }; 63c		M1C. <i>i</i> ₁ f[sf.tw] ₀ ^[P] IV 254-5b L1NY
{ ⁿ L1Li ⁿ }		<i>isf.t</i> ^{nwh} w _{1is} ^P IV 300b M1NY
<i>isf.t</i> ^{nQ?} VI 165f BH2C		<i>isf.t</i> ^{nčs} w ₁ ^P IV 254-5b B9C ^a ; 300b B9C ^a .
<i>i</i> ₁ f ^l .t ^s IV 210-1a M54C		<i>i</i> ₁ f ^l .t ^{nčs} [w] ₁ ^P IV 254-5b Sq7Sq(N)
<i>is</i> {t}f.t ^{nčs} IV 210-1a B3C		<i>isf.t</i> ^[D] w ₁ ^{nčs} IV 254-5b M4C

Abb. 1. Ausschnitt aus der Sargtext-Konkordanz (Darstellung im Druck)

ʔsf (*ʔsf*)

2

ʔss (*ʔss.w*)

ʔsf.tw₁²/₂ = ₁.P IV 254–5b {L3Li}̃
 ʔsf.tw₁^{mwt/2.P} I 405d M6C
 ʔsf.tw₁^{nčš.s₁^m.P} IV 300b Sq7Sq(N)
 [ʔsf.tw]₁^{nčš.V^m.P} IV 300b M54C
 ʔsf.tw₁^{s^m.P} VII 467a B6C
 ʔsf.tw₁^{O.P} IV 328n B1L
 ʔsf.tw₁^P IV 254–5b T3Be, Sq1C. ʔsf.t¹w₁^P IV
 254–5b M1NY. <ʔ>sf.tw₁^P IV 254–5b M54C
 ʔsf.tw₂^P V 326i B2L
 ʔsf.tw₂^{V.P} V 324f B17C
 ʔsf.t^Dw₃^{nčš.s₁.P} IV 319b B3C
 [ʔsf].t¹w₃^{mwt/2¹.P} VII 467a {B4C}̃
 ʔsf.tw₃^{nčš.s₁.P} IV 254–5b B3C
 ʔsf.tw₃^{O¹.P} VII 467a B9C
 ʔsf.tw₃^P IV 300b B9C^b
 ʔsf.t¹w₃^{Q¹} IV 397c {M39C}̃
 {s}ʔ(sf).tw₀^{V.P} V 326i B4C
 {s}ʔ(sf).tw₂^{V.P} V 324f B4C
 {s}ʔ(sf).tw₃^{nčš.P} IV 254–5b B9C^b

ʔsf ?*ʔsf* »(Substantiv)«

Sg.

ʔsf^{mčt.pr^r} VII 473e B1Pʔsf^{mčt.Q} VII 473e B1Be*ʔsr**ʔsr*_a »Tamariske«

Sg.

ʔsr^{ʔsm} VII 197f pGardIIIʔsr^{ʔsr} VI 309c B1Boʔsr^{ht} VI 2j B2L. ʔsr^{ht} VI 2j B1P

Pl.

ʔsr.w^{ʔsm.P} III 245g B1L*ʔsr*_b »Tamariskenhholz(szepter)«

Sg.

ʔsr^{ʔsm} VII 59g TT240ʔsr^{ht} VII 59g Sq3C^a, Sq3C^bʔsr<r>^{mtw}{r}^{ht} VII 59g Sq1Sqʔsr<r>^{ht} VII 59g Sq10Cʔsr<r>^{ht} VII 59g Sq2Sq*ʔsr.t* »Tamariskengott«

Pl.

ʔsr^{ʔsm}.tw₀^{V.P} VI 242e S14Cʔsr.{w}^{ʔsm}tw₀^{V.P} VI 242i S14C*ʔss*

~ »in der Vogelfalle fangen«

ščmf gem.

– ohne Suffix

ʔss{w}^{nwh} VI 33f S1C

– mit Suffix Pl. 2.

ʔss^{tw} VI 17f B1Bo

Imperativ

– allg.

ʔs₁ VI 3h B6Bo, B9C, B1Y; 4b B6Bo, B9C. ʔs₁¹

VI 3h B1C; 4b B1Y

– Pl.

ʔs₁^P VI 4b B1C

Negativkomplement

ʔss VI 37n pGardII

Relativform gem.

– Sg./Pl. f.

– – mit Suffix Sg. 2. m.

ʔss.t VI 37n pGardII

– – mit Suffix Pl. 2.

ʔss.t^{tw} VI 17f B1Bo; 34h B1P; 34k B1Pʔss.t^{ʔssmw¹} VI 30a S2C*ʔss.w*¹ »(in der Vogelfalle) Gefangener«

Pl.

ʔss.w^{ʔs^m.P} VI 29g S1Cʔss{w}.t^{ʔv.P¹} VI 29g S2C*ʔss.w*² »Vogelfänger«

Sg.

ʔss¹.w^v VI 34h B1P

Abb. 2. Ausschnitt aus der Sargtext-Konkordanz (Darstellung am Bildschirm)

Textkorpus, das der Sargtexte, vollständig mit seiner Hilfe erfaßt und verarbeitet. Ich arbeite jetzt nur noch auf der Basis der mit Hilfe dieses Programms erzeugten Ausgabedaten. Und: Sollte ich, was ich hoffe, die Zeit finden, noch einmal ein anderes Textkorpus aufzunehmen, würde ich, dem heutigen Stand der Computertechnik entsprechend, andere Programme wählen oder schreiben, vor allem aber würde ich mir die im Sargtext-Projekt erworbene lexikalische Basis zunutze machen, d.h. lexikonbasierte Texterfassung betreiben, die zuvor, mangels geeigneter maschinenlesbarer lexikalischer Sammlungen, nicht möglich war.

Nun, ganz ohne Programme kann auch ich nicht arbeiten. Was ich verwende, ist jedoch vor allem ein fertiges Programmpaket für philologische Anwendungen, das „Tübinger System von Textverarbeitungs-Programmen“, kurz „TUSTEP“ genannt, das für meine gegenwärtigen Bedürfnisse bestens geeignet ist³. Nur wenn dieses Programmsystem nicht ausreicht, schreibe ich mir ad hoc auch einmal ein Zusatzprogramm. Das einzige, das ich aktuell in Gebrauch habe – ein Programm zur Vorbereitung von Sortierschlüsseln –, werde ich noch zur Sprache bringen.

Ich will mein Licht nicht unter den Scheffel stellen: Ich habe zwar kein besonderes Computer-Programm, aber ich habe in verschiedener Hinsicht aus langer Erfahrung gewisse Vorstellungen davon, wie ägyptische Textdaten strukturiert werden können. Und, denke ich, wenn ich mir über die Strukturierung der Daten im klaren bin, ist der Weg nicht weit, solche Daten mit Standard-Programmen der einen oder anderen Art zu manipulieren oder aber auch einmal ein Programm selbst zu schreiben.

Also: In diesem Sinne besitze ich kein „Konkordanz-Programm“, das ich Ihnen anpreisen könnte.

In ganz anderem Sinn kann ich mir das Thema „Konkordanz-Programm“ zu eigen machen: im Sinne eines Plans zur Konkordanzarbeit an einem Textkorpus, wie im konkreten Fall dem der Sargtexte. Einen solchen Plan will ich Ihnen im folgenden vorführen oder vielmehr: wie der Mensch einen Plan macht und noch einen Plan und wie sie beide nicht gehen.

Ab 1964 habe ich – auf der Basis von Erfahrungen mit Handlochkarten – ein Grundkonzept für die Eingabe und Analyse von Textdaten unter lexikalischen und grammatischen (realisiert: morphologischen) Gesichtspunkten entwickelt, das dann 1969 im Druck erschien⁴. Als Probe aufs Exempel wurde 1970 in Zusammenarbeit mit Rolf Gundlach eine Probe-Konkordanz über einen kleinen Ausschnitt aus den Sargtexten produziert⁵. Es handelt sich hier um eine Konkordanz im traditionellen Sinn, insofern als hier jeder Wortbeleg im Ko-Text, im Satz-Zusammenhang, vorgeführt wird. 1979 dann, als ich in Grenoble über den Fortgang der Konkordanzarbeit berichtete, war in einer Probeseite der Stoff bereits anders gegliedert⁶: Es wurden jetzt vor den Satz-Belegen (Buchstabe B: Kontexte [nach heutigem Sprachgebrauch: Ko-Texte]) die Graphien der Wortformen aufgelistet (Buchstabe A). 1992, in diesem Augenblick, sieht die Konkordanz – keine Probe, keine Simulation, sondern das Material der vollständigen Konkordanz in einer fast schon druckfertigen Version – so aus. (s. Abb. 1): Es

³ TUSTEP, Tübinger System von Textverarbeitungs-Programmen, Zentrum für Datenverarbeitung der Universität Tübingen, Tübingen 1989 (das von Wilhelm Ott und Kuno Schälkle entwickelte Programm-System läuft unter einer ganzen Reihe von Betriebssystemen [Großrechner, aber auch DOS] und wird an zahlreichen Forschungseinrichtungen des In- und Auslandes betrieben).

⁴ S. oben Anm. 1.

⁵ S. oben Anm. 1 (eine leichter lesbare Probe aus dieser Konkordanz findet sich in der in der folgenden Fußnote zitierten Arbeit auf S. 46).

⁶ Wolfgang Schenkel, Eine Konkordanz zu den altägyptischen Sargtexten, in: *L'égyptologie en 1979*, Paris 1982, I, S. 45–53, speziell S. 48.

gibt keine Satz-Belege mehr, sondern nur noch die Graphien. Es handelt sich bei diesem Produkt also gar nicht mehr um eine Konkordanz im traditionellen Sinne. Nun ist es keineswegs so, daß ich eine Konkordanz nicht mehr als Ziel vor Augen hätte. Ganz im Gegenteil: Das Ziel steht unverrückbar fest. Ich habe nur im Laufe von zwei Jahrzehnten gelernt, daß dieses Ziel in zufriedenstellender Weise nicht so zu erreichen ist, wie ich mir das einmal vorstellte. Ich gebe allerdings zu, daß ich der Konkordanz vielleicht doch etwas näher gekommen wäre, wenn ich meine Zeit nicht auf das Problem der Aufnahme und Manipulation der jetzt dominierenden Graphien verschwendet hätte⁷.

So sehen meine Sargtextdaten aktuell aus: Für jede im Textkorpus vorkommende Wortform gibt es in der Regel einen Satz (Record) (kompliziertere Fälle will ich hier nicht vorführen). Solche Sätze haben praktisch beliebige Länge. Der längste Satz dürfte um die 600 Zeichen lang sein (NB: Ein Satz notiert eine Wortform). Da solche Sätze fallweise nicht schnell zu überblicken sind, zerlege ich sie zum Bearbeiten am Bildschirm in Abschnitte, die – man mag es glauben oder nicht – rasch zu überblicken und ziemlich gut zu lesen sind (s. Abb. 2): Der Satz ist nach inhaltlichen Kriterien auf mehrere Zeilen verteilt und durch einen horizontalen Trennstrich vom folgenden Satz abgesetzt.

Solche Sätze korrekt niederzuschreiben, ist fast schon ein Ding der Unmöglichkeit. Tatsächlich wurden die Sätze so auch nie niedergeschrieben. Sie wurden stufenweise erzeugt:

Stufe I

Was ursprünglich in einem Arbeitsgang niedergeschrieben wurde, steht in der Zeile über dem Strich zwischen den Markierungen @“ und @@: Dies ist das Aufnahme-Format des Programms M.A.A.T., das zu Lochkartenzeiten entwickelt wurde (bei den Angaben wie z. B. „E“ und „0“ über dem obersten Strich handelt es sich um inzwischen obsolet gewordene arbeitstechnische Vermerke aus einer späteren Bearbeitungsphase). Genauer genommen wurden selbst diese Zeilen nur in jedem vierten Fall wirklich so niedergeschrieben. Da jede Sargtextstelle (de Buck-Zitierabschnitt) durchschnittlich in 3,8 Textzeugen überliefert ist, konnten alle nicht-ersten Textzeugen – teilweise bis zu ca. 30 – durch Duplizieren und Überkorrigieren eines bereits notierten Textzeugen gewonnen werden. In Klammern: Wie arbeitssparend diese Prozedur wirklich war, ist schwer zu sagen, jedenfalls nicht in dem Ausmaß, wie man erwartete. Man kann überspitzt sagen: Wenn der Wortlaut von parallelen Textzeugen wirklich einmal vollständig gleich ist, dann ist zum mindesten noch der Erhaltungszustand verschieden. Es müssen also, ein langwieriges und mühsames Geschäft, die textkritischen Zeichen korrigiert werden. Zurück zum Endprodukt, der Zeile über dem Strich: An sich ist diese Zeile – jedenfalls im Prinzip – heute vollständig redundant, weil alle Angaben, die hier stehen, noch irgendwie aus den Zeilen darüber zu entnehmen sind. Ich schleppe die alten Notationen jedoch einstweilen weiter mit, aus Sicherheitsgründen, vor allem aber, weil sie – man mag es glauben oder nicht – so bequem und rasch gelesen werden können (bequemer und rascher als die aktuellen Codierungen). Dagegen ist das Niederschreiben selbst derart einfacher Notationen ziemlich fehleranfällig. Rolf Gundlach, mit dem ich die ersten Niederschriften besorgte, steckt diese Arbeit heute noch in den Knochen. Trotzdem: Das ist lange ausgestanden.

⁷ Zur Darstellung der Graphien allgemein Wolfgang Schenkel, Aus der Arbeit (s. oben Anm. 2).

Stufe II

Mitte der 70er Jahre – ich kann nicht mehr genau sagen: wann – habe ich mich dann zu etwas hinreißen lassen, wovor ich mich gehütet hätte, wenn ich geahnt hätte, was ich mir damit auflud: die Aufnahme der Graphien in einer im Computer verarbeitbaren analytischen Notation. In Darmstadt waren – dies bereits eine Erweiterung des Ausgangskonzepts – semantisch relevante Hieroglyphenzeichen, Semogramme, d.h. Logogramme und Determinative, notiert worden (Abb. 2, Zeilen über dem Strich, vor der Markierung @ @). Ein Grund für diese Entscheidung waren die Kritiker der Darmstädter Probe-Konkordanz, die statt Transkriptionen Hieroglyphen sehen wollten. Meine Reaktion auf diese Kritik war dann allerdings gerade das Gegenteil von dem, was man erwartet hätte: Ich wollte keine Hieroglyphenzeichen reproduzieren, sondern die Struktur der hieroglyphischen Graphien sichtbar machen. Man mag das akzeptieren oder auch nicht. Ich kann aber versichern, daß auf diese Weise Graphien im Computer sachgemäß, wenn auch nicht in jeder beliebigen Hinsicht, manipuliert werden können, während dies in diesem Ausmaß mit einzelnen Hieroglyphenzeichen bzw. ihren Codes schwerlich möglich ist. Ich komme darauf zurück. Realisiert wurde die graphematische Anreicherung der bereits vorhandenen Grund-Aufnahme – Zeile über dem Strich – auf folgende Art und Weise: Es wurde ein Graphienlexikon aufgebaut, das zu jeder der verschiedenen Wortformen, die im Textkorpus vorkommen, all die graphematische Information enthielt, die in den Textaufnahmen noch nicht notiert war. Im Bearbeitungsverfahren wurden zu jeder laufenden Wortform im Text alle bereits im Lexikon vorhandenen Möglichkeiten eingespielt, der Bearbeiter akzeptierte die passende bzw. brachte sie, wenn diese im Lexikon noch nicht vorgesehen war, bei dieser Gelegenheit in das Lexikon. Bei Abschluß der Arbeit enthielt das Lexikon zwischen 30 000 und 40 000 Einträge – ich weiß die genaue Zahl nicht mehr –, d.h. es trat im Textkorpus auch etwa diese horrende Zahl von verschiedenen Graphien der verschiedenen Wortformen auf, und dies, obwohl für Logogramme und Determinative nur Platzhalter im Lexikon standen und im Falle der Determinative die Platzhalter sogar nicht einmal für einzelne Determinative, sondern für eine beliebig lange Determinativsequenz standen (z.B. stand für die häufige Folge Determinativ + Plural-Determinativ nur ein Platzhalter). Die Menge der verschiedenen Graphien ist angesichts eines Textkorpus von rund viermal parallel überlieferten ca. 150 000 Wörtern Text geradezu alptraumhaft. Entsprechend unrealistisch waren die Vorausberechnungen der Arbeitszeit, die man für die Anreicherung der Textaufnahmen in Anschlag brachte. Hinzu kam schließlich ein letztes: Die Montage der diversen Teile der Graphemnotation – Phonogramme + Semogramm-Position aus dem Lexikon, Semogramme selbst (anhand einer Tabelle aus den ursprünglichen Gardiner-Nummern in Lautwerte meines Transkriptionssystems übersetzt) aus der Textaufnahme, textkritische Zeichen aus der Textaufnahme – ist so kompliziert, daß ich sie nicht in jedem Detail bewältigen konnte, da – nach ca. 20jähriger Laufzeit der M.A.A.T.-Programme – an der Universität Tübingen der letzte FORTRAN-Compiler außer Dienst gestellt wurde, der die Programme noch übersetzen konnte. Da eine weitere Adaptation der Programme, die ich schon mehrfach hatte vornehmen müssen, nicht mehr mit sinnvollem Aufwand möglich war, mußte ich die abschließende Produktion, so gut es eben zu dem Zeitpunkt ging, über die Bühne bringen. Ich schätze, daß dieser Umstand allein das Projekt ein Jahr zurückgeworfen hat. Auch das ist praktisch ausgestanden.

Stufe III

Zur Zeit stehe ich mitten in der Endredaktion einer ersten, lexikalischen Konkordanz, deren ungefähres Druckbild ich bereits vorgeführt habe (Abb. 1). Ein solcher endredaktioneller Zwischenschritt war in dem in Darmstadt einmal erprobten Verfahren überhaupt nicht vorgesehen und wurde auch bei Erstellung der Darmstädter Probe-Konkordanz nicht durchgeführt. Damals wurden die Ergebnisse der Textanalyse einfach sortiert und per Programm auf Druckseiten angeordnet. Der Unterschied zwischen damals und heute liegt in folgendem: Es ist ein qualitativer, nicht nur ein quantitativer Unterschied, ob man eine Textmenge im Umfang von kaum 25000 laufenden Wortformen aus einem eher einfachen Sargtextspruch erschließen soll – so in Darmstadt – oder eine Textmenge im Umfang von mehr als 1/2 Million laufenden Wortformen aus Texten teilweise doch schwieriger Art. Es häufen sich die Problemfälle, deren Entscheidung man bei der Textaufnahme, z.B. in Erwartung von Parallelmaterial, vertagt. Im übrigen: Vertagt man bei kleinem Textmaterial, so vertagt man um Stunden, Tage oder Wochen; vertagt man bei umfangreichem Material, so vertagt man um Monate und Jahre, und kommt dann endlich die Parallele, hat man das alte Problem oft nicht mehr recht präsent. Zweites Problem: Belegmaterial geringeren Umfangs läßt sich in relativ anspruchsloser Weise sortieren. Die in sich ungeordneten Beleggruppen bleiben so klein, daß der Benutzer, wenn er Bestimmtes sucht, notfalls einfach die ganze Gruppe durchgeht oder sogar die Belegstellen nacheinander in der Edition aufschlägt. Wie aber, wenn ihm tausende und zehntausende Belege vorgesetzt werden? In solchem Fall dürfte es schon fast angenehmer sein, die Konkordanz beiseite zu legen und statt dessen die Edition durchzugehen. Ich gebe zu: Wenn die Textaufnahmen mehr Merkmale notieren als bei der Sortierung benutzt, kann man computergespeichertes Material recherchieren. Ich bezweifle jedoch, daß man aus dem Stand heraus ohne weiteres auf all die Gesichtspunkte kommt, nach denen man fragen könnte. Was ich z.B. bei den Pronomina der 1. Person, Trivialformen, bei der Endredaktion gesehen habe, wird niemandem einfallen, der das Material nicht Beleg für Beleg gesehen hat. (Ich berichte darüber bei Gelegenheit der de Buck-Konferenz im Dezember 1992 in Leiden.) Was letztlich eine Endredaktion (und auch eine Recherche) schwierig macht, sind die Extreme: die Mini-Beleggruppen und die Maxi-Beleggruppen. Die Gruppen mittlerer Größe, die einem *prima vista* ins Auge fallen, sind relativ einfach zu bewältigen. Die Mini-Beleggruppen sind deshalb so zeitraubend, weil einzelne oder wenige Belege gesonderte Überlegungen erfordern, die Maxi-Beleggruppen, weil man Mittel und Wege finden muß, sie dem Benutzer überschaubar zu machen (Am Rande: Keine Maxi-Beleggruppe ist wie die andere!). Was aber das Allerschlimmste ist, sind die unentscheidbaren Restfälle, im allgemeinen Mini-Beleggruppen, mit allen beliebigen Arten und Graden der Nichtentscheidbarkeit.

Vereinfacht gesagt besteht die aktuelle Aufnahme der Sargtexte aus Datensätzen, von denen jeder, wie bereits gezeigt, eine Wortform enthält (s. Abb. 2). Diese Datensätze können in unterschiedliche Reihenfolgen sortiert werden, z.B. in die Wort-Folge des Textes oder in die lexikographische Folge der lexikalischen Konkordanz. Datensätze können ferner selektiert werden. Und, ein letztes: Datensätze können geändert werden, was die Sortierfolge und die Selektion beeinflussen kann. Mehr passiert bei mir nicht. Als Programme benötige ich derzeit einen allgemeinen Text-Editor – ich benutze den leistungsadäquaten TUSTEP-Editor – und Sortierprogramme – wozu gleich noch etwas mehr. Man könnte die Aufgabe u. U. bequemer mit Hilfe eines Datenbank-Programms lösen – es käme auf einen Versuch an –, sicher bin ich mir dessen nicht, mein Code ist höchst verwickelt.

Ich komme wieder auf die bereits vorgestellten Datensätze zurück (Abb. 2), die man mit der Zeichenfolge konfrontieren mag, die diese Datensätze in der Konkordanz (Abb. 1, Anfang) bewirken:

Alles, was im Datensatz mit geschweiften Klammern gekennzeichnet ist – und manches mehr – sind endredaktionelle Eingriffe in die Sortierfolge, Fälle, in denen ich für die Zwecke der Sortierung eine andere Codierung wähle als für die ursprüngliche Aufnahme, die primär die Identifizierung zum Ziel hatte. Identifizierung ist eines, Sortierung/Klassifizierung ist ein anderes, auch wenn beides miteinander zu tun hat. Z.B. ergibt sich aus einer wie auch immer gearteten distinguierenden Indizierung homonymer Lexeme nicht unbedingt eine für Wörterbuch oder Konkordanz geeignete Reihenfolge der Lexeme.

Im Beispielfall (Abb. 2 und 1) ist der Index „1“ („01“) zum Wort *isf.t* durch geschweifte Klammern bei der Endredaktion als irrelevant gekennzeichnet worden, da es im Sargtextmaterial kein anderes von der Wurzel *isf*¹ abgeleitetes homonymes Wort *isf.t* gibt, das durch einen Index davon zu unterscheiden wäre. Gäbe es ein solches Wort, wäre bei der Endredaktion zu entscheiden gewesen, welchen Index man welchem Wort zuordnet, und damit, in welcher Reihenfolge die Wörter in der Konkordanz auftreten sollen (NB: Die ursprünglichen Indizes dienen der Homonymenscheidung, die endgültigen legen darüber hinaus die Sortierfolge fest und sollten so vergeben werden, daß sie von „1“ an lückenlos aufsteigen, d.h. daß keine Indizes für im Belegmaterial nicht vorkommende Wörter ausgespart bleiben). – Demgegenüber wird der Index „1“ („01“) zur Wurzel *isf* nicht durch geschweifte Klammern eliminiert, da es eine zweite, weniger wichtige bzw. unklare Wurzel *isf*² im Belegmaterial gibt, der der Index „2“ zugeordnet wird (unabhängig davon, ob bei der Homonymentrennung diese Wurzel bereits mit „2“ oder mit einem anderen Index wie z.B. „2“ von *isf*¹ unterschieden war).

Meine Codierung der Datensätze ist – ich darf das in aller Bescheidenheit sagen – ein Wunderwerk. Sie sucht einen Kompromiß zwischen drei Anforderungen: Sie soll identifizieren und sortierfähig sein, sie soll in Lichtsatz-Steuerzeichen umsetzbar sein, sie soll schließlich am Bildschirm lesbar und korrigierbar sein. Eine Lösung, die allen drei Anforderungen in idealer Weise Rechnung trägt, gibt es nicht. Also muß einerseits für die Sortierung und für den Druck noch ein komplizierter Zwischenschritt angehängt werden, andererseits bedarf der Leser am Bildschirm einiger Gewöhnung. Was die Umsetzung in Lichtsatz-Steuerzeichen angeht, mögen Sie aus der Gegenüberstellung von interner Darstellung und Druckbild erahnen, was da alles geschehen muß (dafür stehen mein Mitarbeiter Knut Buroh und das Programmsystem TUSTEP gerade). Was die Sortierung angeht, verwende ich so weit wie möglich Standard-Software, konkret die philologisch inspirierten TUSTEP-Programme. Was TUSTEP nicht leisten kann, ist in einem besonderen Codierungsprogramm dazwischen geschoben. In diesem Zwischenschritt werden namentlich die Graphien sortierfähig gemacht, und, was hier geschieht, will ich jetzt, da ich auf diese Weise etwas Neues vorführen kann, in den Grundzügen schildern. (Ich kann hier natürlich nicht auf die programmiertechnische Seite des Problems eingehen, die in der Tat etwas kompliziert ist.)

Lexikalisches Belegmaterial wird üblicherweise in eindimensionaler Sequenz dargeboten und muß im großen und ganzen wohl auch im Normalfall so dargeboten werden. Lexem folgt auf Lexem, Flexionsform auf Flexionsform, Graphie auf Graphie, Zitatstelle auf Zitatstelle. Man hat allerdings die Möglichkeit, die Sortiermerkmale in unterschiedlicher Reihenfolge zu berücksichtigen. Z.B. kann man die Belege eines Verbuns zuerst nach der Rektion oder Valenz sortieren, dann innerhalb der Rektionen/Valenzen nach Flexionsformen oder aber genau umgekehrt zuerst nach Flexionsformen, dann innerhalb der Flexionsformen nach Rektionen.

nen/Valenzen. Leider kann es hier kein Patentrezept geben, das für alle Lexeme in gleicher Weise zweckmäßig wäre, in dem Sinne nämlich, dem Benutzer eines Verzeichnisses die jeweils auffälligsten Klassenbildungen vorrangig vor Augen zu führen.

Die Frage der Sortierhierarchie stellt sich natürlich nicht, wenn für ein Lexem 10 Belege vorzuführen sind. Selbst wenn man die Belege in irgendeiner zufälligen Folge vorführt, wird sich der Leser zu helfen wissen: Er kann im Notfall eine Liste mit 10 Belegen einfach durchlesen. Ebensovienig würde ein Leser Probleme haben, wenn er 10000 Belege vorgeführt bekommt, diese aber alle, abgesehen von der Belegstelle, gleich wären, z. B. 10000 Belege des Suffixpronomens $=k$ aus einem Textkorpus, das $=k$ immer mit dem Henkelkorb schreibt. Das aber sind Ausnahmen. Im Normalfall gilt es die Sortiermerkmale so zu hierarchisieren, daß sie ein Maximum an Transparenz in das Belegmaterial hineinbringen. Wie man dann im einzelnen verfährt, läßt sich leider nicht generell sagen. Es hängt vom Fall ab.

Beispiel 1

St.abs. and st.cs. lassen sich anhand der Graphien in der Regel nicht unterscheiden. Es ist also im Normalfall wenig ergiebig, Belege nach st.abs. und st.cs. zu sortieren. Aber es gibt einzelne Lexeme, bei denen der Fall anders liegt, bei denen man dem Leser einen interessanten Befund verschleiern würde, wenn man die Belege gleich nach kompletten Graphien und nicht zuerst nach Flexionsformen sortieren würde.

Beispiel 2

In meiner lexikalischen Konkordanz wird grundsätzlich zuerst nach Flexionsformen, dann innerhalb der einzelnen Flexionsformen nach Graphien sortiert. Den Unterschied zwischen einem passivischen $\acute{s}\check{c}m=f$ und einem passivischen $\acute{s}\check{c}m.w=f$ kann ich sowohl aus dem Code der Flexionsform entnehmen als auch fallweise aus der Graphie. Sortiere ich zuerst nach der Flexionsform, suggeriere ich dem Leser, ich wüßte, wie $\acute{s}\check{c}m=f$ und $\acute{s}\check{c}m.w=f$ voneinander abzugrenzen sind; lasse ich es sein, überlasse ich dem Leser, die Graphien selbst daraufhin zu interpretieren. (De facto: Ich überlasse es dem Leser.)

Beispiel 3

Ursprünglich habe ich bei Verben noch vor der Flexionsform nach der Rektion sortiert. Die Belege eines Verbs zerfallen also zunächst in Belege mit Objekt und Belege ohne Objekt. Diese hohe Bewertung der Rektion ist ungünstig. Leider weiß ich keine hierarchische Positionierung der Rektion, die generell zweckmäßig wäre. Trans. und Intr. sind für das Ägyptische problematische Kategorien, was vor allem beim Pseudopartizip zu praktischen Problemen führt. Also verzichte ich notgedrungen gänzlich darauf, nach der Rektion zu sortieren, eine auch nicht unbedingt zufriedenstellende Lösung.

Geradezu verheerend können sich die nicht entscheidbaren Fälle auswirken. Auch diese müssen irgendwo einsortiert werden, müssen in irgendeiner Weise durch Indizierung sortierfähig gemacht werden. Im Einzelfall kann man sich dazu durchringen, nach Probabilität eine Zuordnung zu klaren Beleggruppen zu treffen, in anderen Fällen bleibt nichts anderes übrig, als einzelne Belege oder Beleggruppen als isolierten Sonderfall stehen zu lassen.

Die optimale Lösung gibt es nicht, aber es gibt bessere und schlechtere Lösungen, manchmal Generallösungen, manchmal nur Lösungen für den Einzelfall.

Regeln für die Sortierung von Graphien gibt es bislang in der Ägyptologie nicht (wenn ich einmal von den frühen Versuchen – Champollion – absehe, das ägyptische Wörterbuch in der Art einer Zeichenliste zu organisieren, mithin die Zeichenklassenzugehörigkeit der Grapheme – so gut es eben geht – als Sortierkriterium zu benutzen). Ich möchte mich hier kurz mit diesem Problem beschäftigen, mich allerdings nicht in allgemeinen Erörterungen darüber ergehen, wie man das Problem grundsätzlich anpacken kann – es gibt nicht nur eine denkbare Lösung –, sondern vorführen, wie die Lösung aussieht, die ich benutze. Es hängt diese, wie man nicht anders erwarten wird, mit der Art meiner Codierung der Graphien zusammen, ist aber an deren spezifische Details nicht notwendig gebunden. Hier also die wichtigsten Regeln:

1. Grundsätzlich werden die Graphemnotationen, wenn im folgenden nichts anderes angegeben wird, elementweise von links nach rechts abgearbeitet – niemand würde etwas anderes erwarten.
2. Die Elemente der Graphemnotation werden auf sechs Ebenen verteilt. Es werden – von links nach rechts – zuerst alle Elemente der 1. Ebene abgearbeitet, dann wieder von vorn beginnend bis zum Ende alle Elemente der 2. Ebene usw.

Die sechs Ebenen der Graphemnotation – zwei Ebenen Lautwerte, vier Ebenen Textkritik – sind diese:

1. **Ebene:** alle Elemente, die – in meinem Notationssystem – durch einen Lautwert dargestellt sind, also Phonogramme, Logogramme und Determinative (nicht aber sogenannte Phonetische Determinative; zu den Komplementen, die auf Unterebenen zu dieser Ebene behandelt werden, später);
2. **Ebene:** Determinative (die in meiner Notation alle durch Indizes dargestellt sind);
3. **Ebene:** die textkritischen Zeichen, die über die Bonität der Überlieferung Auskunft geben, also Angaben über Zerstörung, Teilzerstörung, Fraglichkeit u. a. m.;
4. **Ebene:** die textkritischen Zeichen, die – NB: für die Zwecke der ersten, lexikalischen Sargtext-Konkordanz – zweitrangige, aber immer noch wichtige Angaben zu den Graphemen liefern, z. B. die durch den Philologen vorgenommene Duplizierung von Graphemen bei partiell gespaltenen Kolumnenschreibung;
5. **Ebene:** die textkritischen Zeichen, die – für die Zwecke der ersten, lexikalischen Sargtext-Konkordanz – weniger wichtig sind, z. B. das Faktum, daß der Herausgeber der Edition den Text nicht selbst gesehen hat, man also auf das Zeugnis älterer, im allgemeinen weniger kompetenter Abschreiber angewiesen ist;
6. **Ebene:** die textkritischen Zeichen, die – für die Zwecke der ersten, lexikalischen Sargtext-Konkordanz – entbehrlich erscheinen, z. B. die Angabe von Rubra (die in anderem Zusammenhang evtl. hoch interessant sind, deshalb auch notiert wurden).

Soweit der allgemeine Überblick über die zwei plus vier Ebenen.

Elemente mit Lautwert – Ebene 1 – werden nach folgenden Kriterien sortiert (vgl. Abb. 3):












1.	$R^c W$	
2.	$r^c w$	
1.	$mr_{ 1}$	
2.	$mr_{ 2}$	
1.	$m+r$	
2.	mr	
1.	p^3_0	
2.1	p^3_1	
2.2	p^3_2	
3.	p^3_3	

Abb. 3. Sortierhierarchien graphematischer Elemente mit Lautwert

- a) Graphien mit Logogrammen stehen vor solchen ohne Logogramme;
- b) homophone Grapheme werden nach den Indizes sortiert, die sie in meinem Transkriptionssystem haben;
- c) Grapheme mit kürzeren Lautfolgen stehen vor längeren, z. B. solchen mit Einkonsonantenzeichen *m* vor Zweikonsonantenzeichen *mr*;
- d) Komplemente werden, unabhängig davon, ob sie vor oder nach dem komplementierten Zeichen stehen, erst nach dem komplementierten Zeichen berücksichtigt, und zwar wird zuerst nach der Anzahl der komplementierten Laute des Mehrkonsonantenzeichens geordnet, dann nach der Position der Komplemente vor bzw. nach dem Mehrkonsonantenzeichen.
- Eine besondere Raffinesse habe ich mir für die Sortierung der textkritischen Angaben der Ebene 3, die Bonität der Überlieferung betreffend, ausgedacht: Es wird hier zuerst einmal nach der Bonität der Gesamt-Graphie (die aus dem Umfang und der Art der lädierten Lautzeichen berechnet wird) und dann nach dem Schwerpunkt der Läsion (weiter rechts oder weiter links in der Graphie) sortiert, so daß die Belege in der Reihenfolge ihrer Bonität und in der Reihenfolge des Schwerpunktes ihrer Läsion angeordnet werden. Danach werden die Belege, wie auch die textkritischen Zeichen auf den nachfolgenden Ebenen 4 und 5, nach der Position der einzelnen textkritischen Zeichen, bezogen auf die Position der Lautwerte der 1. Ebene, sortiert.
- Die textkritischen Zeichen der Ebene 5, die letzten überhaupt noch (für die Zwecke der lexikalischen Sargtext-Konkordanz) berücksichtigten, werden in der Notation eliminiert und bei der Graphemsortierung nicht berücksichtigt, führen aber zu einem „Flag“ bei der Stellenangabe, das besagt, daß in der Graphie irgendwo noch eine bestimmte Besonderheit vorliegt, ohne zu sagen, an welcher Stelle genau (z. B. ist in Abb. 1 die Textsigle mit  und

umklammert, wenn der Textedition an der betreffenden Stelle – vollständig oder partiell – eine ältere, weniger zuverlässige Textabschrift zugrunde liegt). Soviel auszugsweise zum Sortieren.

Wie geht es nun weiter mit der Konkordanz, dem „Konkordanz-Programm“? Zunächst muß natürlich die lexikalische Konkordanz fertiggestellt werden. Es sind aber dann unmittelbar, auf der Grundlage der jetzt vorliegenden Datensätze, zwei weitere Konkordanzen herstellbar: eine morphologische Konkordanz und eine graphematische Konkordanz. Ebenso gut sind ein paar Register herstellbar, z. B. Häufigkeitsregister.

Sollte dies alles einmal bewältigt sein, käme ich zu meinem Ausgangspunkt der Computer-Anwendung zurück, der Syntax, die mich 1964 zum Computer führte und die mich bereits Ende der 60er Jahre schon einmal bis zu den Problemen des Parsing vorpreschen ließ. Ich will mich hier nicht allzusehr in Plänen ergehen. Ich will nur soviel sagen: Erst wenn dieser Schritt bewältigt ist, kann man an die Produktion einer echten Konkordanz, einer Kontext- oder, wenn man lieber will, einer Ko-Text-Konkordanz denken. Denn erst anhand syntaktischer Strukturbäume läßt sich das Belegmaterial in einer Weise sortieren und selektieren, daß man inhaltlich vergleichbare Belege aus einem so umfangreichen Material unter einer Vielzahl von Gesichtspunkten zusammenfindet.

Ein Durchgang durch das Textkorpus anhand der Edition dauerte jeweils mindestens ein Jahr. Um die jetzt angehäuften Notationen zuwege zu bringen, waren viele Durchgänge nötig. Ein früherer Durchgang durch das Material in der lexikalischen Reihenfolge, zu einem Zeitpunkt, als nur jeweils ein Textzeuge aufgenommen und die Anreicherung der Graphien noch nicht durchgeführt war, dauerte mehr als ein Jahr (ich habe vergessen, wie lange genau), die aktuelle Endredaktion der lexikalischen Konkordanz dauert allermindestens drei Jahre. Insgesamt ist das Projekt nach dem Darmstädter Probelauf jetzt 20 Jahre ununterbrochen im Gange. Ich wurde ständig durch Partner und/oder Hilfskräfte unterstützt, in der Mehrzahl ausgezeichnete Kräfte, zuletzt und am längsten und besonders effektiv durch Doris Topmann.

In dieser Zeit haben sich die Vorstellungen, jedenfalls meine Vorstellungen von ägyptischer Morphologie und – das ist die Schuld von Jürgen Osing – von ägyptischer Wortbildung u. a. m. gewandelt. Ursprünglich, bis etwa zwei Jahre nach 1965, gab es bei mir keine „emphatischen“ Formen, danach die „emphatischen“ Formen von 1965 plus weitere solche auf der Linie der Jungeschen Verbalsyntax. Polotsky 1976 habe ich nicht mehr einarbeiten können. Die Transpositionenlehre wurde also nie konsequent angewandt (nicht zuletzt nicht außerhalb der Verbalsyntax, wo sie mit Nutzen hätte angewandt werden können). Inzwischen streiche ich, bei der Endredaktion, alle „emphatischen“ Formen als solche wieder heraus. Wie immer man das beurteilen mag: Der Rückzug auf die – wie James P. Allen das nennt – „written forms“ bringt jedenfalls eine größere Homogenität der Befundbeschreibung zuwege. – In den 60er Jahren habe ich mich bei den Partizipien an Edels „Altägyptischer Grammatik“ orientiert. Später habe ich versucht, per Programm ein besseres Konzept zu realisieren. Das ist mir allerdings nur sehr partiell gelungen, weil die präeditorischen Kennzeichnungen der Partizipien teilweise in den Textaufnahmen geändert werden mußten. Auch hier habe ich mich einstweilen mit reduzierten Ansprüchen zu begnügen, muß mich mehr, wenn auch nicht ausschließlich, an den „written forms“ orientieren. – Wer will mir heute sagen, wie man in 10, 20, 30 Jahren die Dinge sehen wird?