

## ÜBERLEGUNGEN ZUR XML-NOTATION ÄGYPTISCHER TEXTE

### PROBLEMSTELLUNGEN UND LÖSUNGSANSÄTZE AM BEISPIEL DER ÜBERSETZUNG DES TÜBINGER SARGTEXTKORPUS IN XML-NOTATION

WOLFGANG SCHENKEL

#### 1. Einleitung

Das Tübinger Sargtextkorpus, das den folgenden Überlegungen als Materialbasis zugrunde gelegt wird, ist eine analytische, maschinenlesbare Darstellung der Sargtexte oder, genauer gesagt, des Sargtextteilkorpus, das in der Standard-Edition von DE BUCK erfaßt ist.<sup>1</sup> Der Einfachheit halber soll das Tübinger Sargtextkorpus fernerhin mit dem Kürzel „CTTC“ angesprochen werden. Mit den Buchstaben „CT“ („Coffin Texts“) soll der zugrunde gelegten Textedition Rechnung getragen werden, mit den Buchstaben „TC“ („Text Corpus“) der vollständigen Erfassung des Wortlautes der Texte in allen ihren Textzeugen, im Gegensatz etwa zur Datenbasis des „*Coffin Texts Word Index*“, die sich auf die Erfassung der Wörter beschränkt.<sup>2</sup> „Tübinger Sargtextkorpus“ soll es auch aus der Erwägung heraus nicht weiter heißen, weil das Textkorpus, wenn es auf längere Zeit gesichert werden soll, zu gegebener Zeit an anderem Ort gesichert werden muß und dann kein Tübinger Sargtextkorpus mehr sein wird.

Im CTTC ist eine ganze Reihe von lexikalischen, grammatischen und graphematischen Aspekten der Texte in systematischer Weise erfaßt, so daß man sich anhand dieses Materials, anhand real vorliegender und philologisch erprobter Notationen also, Gedanken darüber machen kann, wie ägyptische Texte im Prinzip erfaßt werden sollen und wie sie namentlich in XML („Extensible Markup Language“) notiert werden könnten.

XML eröffnet, wenn auch nicht vollkommene, so doch vielfältige Möglichkeiten, die idiosynkratischen Notationen des CTTC in seinem gegenwärtigen Zustand in eine Standardnotation zu überführen, die allgemeineren Standards entspricht und mit industriellen Standardprogrammen manipuliert werden kann. Einige dieser Möglichkeiten sollen im folgenden erörtert wer-

<sup>1</sup> A. DE BUCK und A.H. GARDINER, *The Egyptian Coffin Texts*, Chicago 1935–1961.

<sup>2</sup> D. VAN DER PLAS und J.F. BORGHOUTS, *Coffin Texts Word Index*, Utrecht, Paris 1998.

den. Das Hauptinteresse richtet sich dabei auf philologisch-inhaltliche Fragen. Die XML-Notation selbst wird im allgemeinen nur so weit herangezogen, wie es für den Zweck unabdingbar ist. Mit anderen Worten: Die Probleme werden anhand der Notation von Textelementen erörtert, nur ausnahmsweise einmal anhand einer abstrakten DTD („Document Type Definition“). Auch werden als Schlüsselwörter mehr oder minder sprechende, fallweise langatmige Bezeichnungen benutzt, die selbstverständlich in einer realen Anwendung drastisch verkürzt werden könnten, um die Daten kompakter speichern und komplizierte Notationen überschaubarer visualisieren zu können.

Auch wenn es im folgenden nur um Vorüberlegungen geht: Es besteht durchaus die Absicht, das CTTC tatsächlich in XML-Notation zu übersetzen und zu gegebener Zeit der ägyptologischen Forschung in dieser Form bzw. *auch* in dieser Form als Ressource zur Verfügung zu stellen. Offen sind derzeit noch die Notationskonventionen im Detail, die, bevor man sich zu einer endgültigen Form entscheidet, auf ihre theoretische Basis und ihre Praktikabilität hin kritisch reflektiert werden müssen.<sup>3</sup>

## 2. Allgemeine Gesichtspunkte der Notation des Wortlauts eines Textes

### 2.1 Was sind die Grundelemente der Notation eines ägyptischen Textes?

Ägyptische Texte werden üblicherweise als eine Folge von hieroglyphischen Schriftzeichen oder als eine Folge lateinischer Transkriptionszeichen notiert. Die Notation in hieroglyphischen Schriftzeichen bedarf keiner weiteren Erörterung, da es – mehr oder minder – die originale Notation ist, die die Ägypter selbst benutzten und die dem Altertumswissenschaftler in den Primärquellen überliefert ist. Daß man sich daneben der Transkription bedient, hat nur marginal schreib- oder drucktechnische Gründe. Jeder ordentlich ausgebildete Ägyptologe kann selbstverständlich Hieroglyphen mit der Hand schreiben. In der Bleisatzzeit konnte man Hieroglyphen setzen, und heute kann dies jeder technisch auch noch so unbegabte Ägyptologe auf seinem heimischen Computer. Es gab zwar eine Zeit, in der Bleisatz zu teuer war, um ihn im wünschenswerten Umfang anzuwenden, und in der der Computersatz noch nicht oder nicht jedermann umstandslos verfügbar war. Die Transkription hat man jedoch nicht in dieser Unglückszeit als Notbehelf erfunden. Man hat seit den Anfängen der Ägyptologie transkribiert –

<sup>3</sup> Hiermit und mit der Erstellung eines Konversionsprogramms ist derzeit WOLFGANG PEMPE beschäftigt, der auch bei den im folgenden ausgebreiteten Überlegungen beratend mitwirkte.

CHAMPOLLION transkribierte in ein koptisches Alphabet – und man transkribiert heute immer noch, auch wenn man am Computer Hieroglyphen setzen kann.

Der Nutzen der Transkription liegt darin, daß man in dieser Elemente des ägyptischen Textes explizit notieren kann, die in der Hieroglyphenfolge nur implizit enthalten sind. Darüber hinaus kann man in der Transkription Elemente notieren, die man aus anderen Quellen als der Hieroglyphenfolge erschließen kann, Elemente nämlich des ägyptischen Textes, die in der Hieroglyphenfolge grundsätzlich nicht notiert sein können, üblicherweise nicht notiert sind oder im Einzelfall tatsächlich nicht notiert sind.

Welche Notation immer man zugrunde legt: Keine Notation ist explizit genug, um mit einfachen Hilfsmitteln – Programmen – die große Mehrzahl der Elemente des Textes zu extrahieren, die man im allgemeinen benötigt. Alle Notationen sind in viel zu hohem Grad mehrdeutig.

Was aber will man aus einem Text extrahieren? Zuallererst den Wortlaut, die Folge der Wortformen. Dazu müssen die einzelnen Wortformen isoliert werden. Darauf aufbauend läßt sich die Kombination der Wortformen untersuchen. Dann zweitens die Merkmale der Wortformen bzw. Wortformen mit bestimmten Merkmalen. Man möchte z.B. die Wortformen bestimmten Wörtern, Lemmata, zuweisen, man möchte den Text also lemmatisieren. Oder man möchte die Flexionsformen bestimmen, die Wortformen also in das grammatische Paradigma einordnen. Aus der Anordnung von Wortformen bestimmter Lexeme bzw. bestimmter Flexionsformen ergeben sich Einblicke in Syntax und Pragmatik etc.

Zum anderen kann man sich für die Elemente interessieren, mit denen die Wortformen geschrieben sind, für die Grapheme, ihre Funktionen und ihre Anordnung.

## *2.2 Wie bestimmt man die Grundelemente eines Textes?*

Ideal wäre, man gäbe einen Text in der Hieroglyphenzeichenfolge in den Computer ein und würde per Programm die interessierenden Informationen aus dem Text extrahiert bekommen. Bekanntlich geht das nicht. In jedem Falle gibt es Mehrdeutigkeiten, die mit den heutigen, in Programme umsetzbaren Kenntnissen nicht aufzulösen sind, letzten Endes auch Mehrdeutigkeiten, deren Auflösung umfangreiches kotextuelles, kulturelles, ja Weltwissen erfordert. Man kann dann immer noch der Illusion nachhängen, man würde zuerst ein Programm die Mehrdeutigkeiten finden lassen und diese dann sozusagen per Hand auflösen. Die Erfahrung – jedenfalls meine Erfahrung – zeigt jedoch, daß es bequemer, sicherer und effektiver ist, Texte, jedenfalls ägyptische, zuerst eindeutig zu machen und danach erst ein Programm anzusetzen. Es ist nämlich erstens unbestreitbar, daß viele ägyptische

Texte zuerst mit Hilfe eines philologischen Instrumentariums aufbereitet werden müssen, bevor man überhaupt einen Text hat. Diese Arbeit ist oft so zeitraubend, daß im Vergleich hierzu die explizite Herstellung der Eindeutigkeit überhaupt nicht ins Gewicht fällt. Zweitens sind unbestreitbar ägyptische Texte infolge der Schrift derart mit Mehrdeutigkeiten belastet, daß es ineffektiv ist, die Mehrdeutigkeiten im Einzelfall aufzulösen, so wie sie einem ein Programm vorlegen könnte. Es ist sehr viel effektiver, die Auflösung der Mehrdeutigkeiten am Text selbst vorzunehmen, da man bei zusammenhängender Lektüre den Kotext parat hat, den man sich bei der Vorgabe von Einzelstellen immer wieder neu erarbeiten muß.

Ein Text muß also, bevor man die eigentliche Bearbeitung vornimmt, wie man früher sagte, prä-ediert werden oder aber, wie man heute sagt und das bei modernen Sprachen tut, „ge-taggt“ werden.

Was das Korpus der Sargtexte angeht, so habe ich mir ab 1965 ein Verfahren ausgedacht, wie man auf der Basis einer Transkription – Hieroglyphen waren damals im Computer noch nicht manipulierbar – Texte durch Zufügung von Indizierungen vereindeutigen kann. Damals, Ende der 60er Jahre – ich füge das in Klammern ein – habe ich mir herbe Kritik bezüglich des Verfahrens von seiten eines bekannten anglistischen Linguisten gefallen lassen müssen, der von der selbstverständlichen Voraussetzung ausging, daß Texte rein maschinell analysiert werden könnten. Es war die Zeit, in der man an die rasche Lösung des Problems der maschinellen Übersetzung glaubte. Ich habe daran zurückdenken müssen, als mir in den 90er Jahren aus der Germanistik zuerst das Wort „Tagging“ zu Ohren kam und das „Tagging“ als der neueste Schrei der textkorpusbezogenen Linguistik in aller Munde war.

Doch zurück zu den Sargtexten. Mit Hilfe der Indizes habe ich per Programm für das gesamte Textkorpus die gesuchten Elemente bestimmt und diese im Klartext den Wortformen zugeordnet: Lemmata, Kategorien der Flexion, Funktionen von Schriftzeichen. Mein Textkorpus, das CTTC, besteht jetzt nicht mehr aus Wortformen und Mehrdeutigkeiten auflösenden Indizes, sondern aus Wortformen und ihren Merkmalen. Meiner Meinung nach ist dies die einzig brauchbare Darstellung eines ägyptischen Textes. Und, wenn ich mir die Dinge aus der Rückschau noch einmal überlege: Es wäre von vornherein das effektivste Verfahren gewesen, den Wortformen sofort explizit Merkmale zuzuordnen statt Indizes zur Auflösung von Mehrdeutigkeiten. Allerdings wäre dies aus Gründen der mühseligen Eingabetechniken und des beschränkten Speicherplatzes zu der Zeit, als ich mit der Bearbeitung des Textkorpus begann, nicht leicht zu realisieren gewesen. Ich kann aber auch in der Gegenwart bleiben: Andere Ägyptologen verfahren nicht anders, als ich heute verfahren würde: Lemmata werden auch von an-

deren nicht „automatisch“ per Programm bestimmt, sondern – wenn auch mit programmierten Hilfen – im Einzelfall nach Einzelentscheidungen den Wortformen zugeordnet, Kategorien der Flexion explizit zugeordnet. Ich verweise beispielsweise auf den bereits genannten „Coffin Texts Word Index“, der überhaupt nur mit Lemmata arbeitet, oder die Textdatenbank des Altägyptischen Wörterbuchs, bei der jede Wortform durch den Querverweis auf einen Thesaurus dem Lemma und der Flexionsform nach im Einzelfall bestimmt wird.

### *2.3 Wie notiert man die Grundelemente eines Textes?*

Wie man die Grundelemente eines ägyptischen Textes notiert, darüber wird man sich, solange es eine sich fortentwickelnde Wissenschaft der Ägyptologie geben wird, nie definitiv einigen können. Man wird immer Wortformen in hieroglyphischen Schriftzeichen notieren können und man wird immer Transkriptionen benutzen können und immer neue Versionen der Transkription dazuerfinden. Man wird infolge der bruchstückhaften Überlieferung – namentlich was die Vokale angeht – in sehr vielen Fällen Schwierigkeiten bei der Festsetzung der Lemmata im Wörterbuch haben und Schwierigkeiten bei der Festsetzung des Satzes der Flexionsformen in der Grammatik, wie viel mehr bei der Identifizierung von Lemmata und Flexionsformen an den einzelnen Textstellen. Man hat also auf Dauer mit Alternativlösungen zu rechnen. So betrachte ich denn auch die Lösungen, die ich im CTTC gewählt habe, lediglich als Lösungen unter anderen denkbaren Lösungen. Auch wenn es im Einzelfall nicht die denkbar besten Lösungen sind, muß man sie akzeptieren, wenn man mit dem CTTC arbeiten will, weil es schlechterdings unrealistisch ist zu glauben, man könne meine Spezifikationen rasch in andere oder bessere überführen. Noch weniger effektiv wäre es, das CTTC einfach beiseite zu lassen und die Texte von neuem mit anderen Spezifikationen aufzunehmen.

Bei der Notation von Textelementen sind mindestens drei Gesichtspunkte zu berücksichtigen, die bei der späteren Nutzung eine Rolle spielen: die Identifikation, die Sortierung und die Visualisierung. Diese drei Gesichtspunkte seien anhand des Lemmas exemplifiziert.

#### *2.3.1 Identifizierung*

Das Lemma einer Wortform eines Textes wird mit einer Zeichenfolge notiert, die entweder das Lemma selbst darstellt oder auf eine Stelle in einem Thesaurus verweist, an der das Lemma selbst dargestellt ist. Ob man die eine oder die andere Art der Darstellung wählt, ergibt sich aus praktischen, verwaltungstechnischen Gesichtspunkten. Im CTTC sind die Lemmata den

Wortformen einzeln zugeordnet. Wenn man also hier ein Lemma ändern will, muß man dies an allen Textstellen tun, an denen das Lemma vorkommt. Das ist mit einem heutigen PC vom Speicherplatz und von der Rechengeschwindigkeit her kein Problem. Diese Darstellung kann indes in die Form einer relationalen Datenbank überführt werden, wenn das notwendig oder wünschenswert ist. Was die Überführung des CTTC in XML-Notation angeht, würde man, wenn das Textkorpus in irgendeiner Datenbank gespeichert wäre, nicht von einer Datenbank ausgehen, sondern von einem Exportformat einer Datenbank. Eine Art Exportformat ist aber bereits das CTTC in der Form, wie ich es selbst benutze, so daß also auf dieser Basis direkt die Übersetzung in eine XML-Notation vorgenommen werden kann. Es spricht dann allerdings nichts dagegen, die so gewonnene Darstellung der Texte in eine XML-fähige Datenbank zu überführen.

Speicherungstechnische Fragen sind nicht mein Thema. Es geht vielmehr um die Frage, wie die Zeichenfolge aufgebaut ist, mit deren Hilfe man im Text oder in einem Thesaurus ein Lemma beschreibt. Im CTTC sind die Lemmata durch ein Quadrupel definiert: die Wurzel, einen homonyme Wurzeln unterscheidenden Index – genannt Wurzelindex –, das von dieser Wurzel abgeleitete Wort und den homonyme Wörter unterscheidenden Index – genannt Wortindex. Das ist nicht besser und nicht schlechter als die Identifizierung eines Lemmas mithilfe einer Nummer, es ist aber anschaulicher und leichter memorierbar. Es ist auch nicht besser oder schlechter als die Beschreibung durch das Paar Wort plus homonyme Wörter unterscheidender Index, vielleicht aber leichter zu handhaben, weil die Homonymenunterscheidung auf zwei Ebenen – die Wurzelebene und die Wortebene – verteilt und damit auf der einzelnen Ebene leichter zu handhaben ist. Die beiden Darstellungsweisen sind, was die Identifizierung der Lemmata angeht, einander äquivalent und lassen sich im Prinzip ineinander überführen, völlig problemlos in der Richtung Quadrupel – Paar, in der umgekehrten Richtung Paar – Quadrupel dann, wenn man irgendwoher die Information erhält, zu welcher Wurzel das Wort gehört. Die CTTC-Darstellung ist, wenn nicht die von jedermann gewünschte, so doch die aussagestärkere.

Letztlich ist die Form der Identifikation eine Marginalie gegenüber der inhaltlichen Identifikation der Lemmata. Infolge der Überlieferung bereitet es nicht selten ernsthafte Schwierigkeiten zu entscheiden, wie viele Lemmata man ansetzt und wie man sie voneinander unterscheidet. Handelt es sich bei *nh* „leben“ und *nh* „binden“ – Wurzel oder Wort – um zwei Lemmata oder doch um nur eines? „leben“ könnte „(sozial) eingebunden“ sein, im Gegensatz zum Zustand nach dem Tod, wo man außerhalb der Gemeinschaft der wirklich „Lebenden“ steht. Es gibt andere Beispiele, die jedem geläufig sind und die jedermann als zutreffend akzeptieren würde. Es geht hier nur um den

prinzipiell in Rechnung zu stellenden Fall. In anderen Fällen hat man das Problem der Unterscheidung von Wörtern mit und ohne schwache Endung – z.B. *ḥkʾ* (*\*ḥakāʾ*) „Zauber“ und *ḥkʾ.w* (*\*ḥíkʾaw*) „Zauber“ – oder Wörter mit gleichem Konsonantenbestand, aber unterschiedlicher Vokalisation – z.B. die Adjektive oder Partizipien *\*šūrīr* „klein“ und *\*šīrar* „klein“. Für praktische Zwecke kann man auf viele Unterscheidungen verzichten, es gibt aber keine Unterscheidung, die nicht irgendwann einmal von Belang sein könnte, wenn sie in der Sprache selbst real existierte, z.B. kann die genaue Lautform bei der stilistischen Beurteilung von Textstellen eine Rolle spielen, und für einen solchen Fall sollte z.B. ein Wörterbuch schon alle linguistisch möglichen Formen bereit halten, auch wenn nicht alles für den philologischen Alltag von Belang ist.

### 2.3.2 Sortierung

In Anbetracht der gewaltigen Menge an Lemmata benötigt man eine Vorschrift, die die Reihenfolge der Lemmata in einem Verzeichnis der Lemmata regelt, damit man ein Lemma im Verzeichnis, z.B. in einem Wörterbuch, umstandslos wiederfindet. Üblicherweise ordnet man die Lemmata des Ägyptischen – von Sonderfällen abgesehen – nach ihrer Konsonantenfolge und nach Maßgabe eines Alphabets. Bezüglich der Anordnung von Homonymen herrscht keine Klarheit.

Was die Lemmata des CTTC angeht, so suggeriert das Beschreibungsquadrupel eine Sortierung zuerst nach Wurzeln und dann innerhalb der Wurzeln nach Wörtern. Das ist aber keineswegs zwingend. Man kann genauso gut zuerst nach Wörtern sortieren. Tut man dies, sind die drei weiteren Komponenten des Quadrupels Indizes zur Homonymenscheidung. Lassen wir, um uns auf den Kern des Problems zu konzentrieren, Wurzelindex und Wortindex beiseite: Die Angabe Wurzel – Wort kann man auch lesen als Wort mit Wurzel als Index, als ein Wort, das sich von anderen homonymen Wörtern durch die Zugehörigkeit zu einer bestimmten Wurzel unterscheidet. Grundsätzlich kann man dann auch die Indizes insgesamt, Wurzel, Wurzelindex und Wortindex in einen einzigen Index zusammenfassen. Zum Beispiel kann man, um ein Beispiel wieder aufzugreifen, *ḥḥ* „das Leben“ und *ḥḥ* „Gebinde“, Blumenstrauß“ als *ḥḥ<sup>1</sup>* „das Leben“ und *ḥḥ<sup>2</sup>* „Gebinde“, Blumenstrauß“ voneinander unterscheiden, statt als *ḥḥ* von der Wurzel *ḥḥ<sup>1</sup>* „leben“ und *ḥḥ* von der Wurzel *ḥḥ<sup>2</sup>* „binden“. Da aber von jeder der beiden Wurzeln mehr als ein Wort *ḥḥ* abgeleitet ist, muß man diese Wörter alle paarweise voneinander durch den Wortindex unterscheiden, z.B. so: *ḥḥ<sup>1</sup>* „leben“ (als Verb), *ḥḥ<sup>2</sup>* „das Leben“, *ḥḥ<sup>3</sup>* „der Lebende“ (wenn es nicht richtiger *ḥḥ.w* heißt), *ḥḥ<sup>4</sup>* „Spiegel“, *ḥḥ<sup>5</sup>* „binden“, *ḥḥ<sup>6</sup>* „Blumenstrauß“, *ḥḥ<sup>7</sup>* „Eid“, *ḥḥ<sup>8</sup>* „Ohr“. Es werden in einer Zählreihe dann natürlich mehr

Indizes benötigt, als wenn man sich zweier Zählreihen, Wurzel und Wort, bedient.

Das Verhältnis zwischen Wurzel und Wort ist hier nicht deshalb als Beispiel gewählt worden, um die Frage zu entscheiden, ob in einem Wörterbuch bei der Anordnung der Lemmata die Wurzel eine Rolle spielen soll oder nicht. Es wurde der Fall nur deshalb zur Sprache gebracht, weil man sich daran vor Augen führen kann, daß beim Ansatz von Lemmata neben der Identifizierung auch die Anordnung der Lemmata – namentlich im Wörterbuch – zu bedenken ist und daß sich aus einer identifizierenden Beschreibung noch nicht zwangsweise eine Sortieranweisung ergibt.

### 2.3.3. Visualisierung

Angenommen, ein ägyptischer Text soll als die Folge der hieroglyphischen Schriftzeichen notiert werden. In diesem Fall erfüllen ganz unterschiedliche Formen der Notation den Zweck: Man kann die hieroglyphischen Schriftzeichen als Bilder, d.h. als Graphiken speichern, man kann sie als eine Zeichennummer speichern, z.B. die Gardiner-Zeichennummer A 1 für den hockenden Mann, oder als einen Lautwert, z.B. den hockenden Mann, Zeichennummer A 1, als den Logogrammwert *s.i* (oder was auch immer) „Mann“. Selbstverständlich ist ein Text nicht in jeder beliebigen Art von Notation gleich gut lesbar. Man wird also zum Vorteil des Lesers eine Form der Darstellung, eine Visualisierung wählen, die ihm das Lesen möglichst erleichtert. Zum Beispiel wird man, wenn der Text als Folge der hieroglyphischen Schriftzeichen notiert ist, im allgemeinen die Abbildung in Bildzeichen liefern, ganz unabhängig davon, wie diese Bildzeichen computerintern gespeichert sind. Dafür muß allerdings klargestellt sein, welches hieroglyphische Bildzeichen dem internen Code, der Gardiner-Zeichennummer bzw. dem Logogrammwert, entspricht. Das ist trivial. Auf dieser Basis arbeitet z.B. das Programmsystem *Glyph*.

Nicht anders verhält es sich, was man vielleicht weniger gerne hört, im Falle der Wiedergabe eines ägyptischen Textes in Transkription. Ein Transkriptionsalphabet besteht aus einer bestimmten Menge von paarweise verschiedenen Zeichen, von denen jedes einzelne eine bestimmte Bedeutung hat und die in einer bestimmten Reihenfolge aufeinander folgen. Ziemlich gleichgültig ist jedoch, wie die Zeichen eines Alphabets notiert werden. Man kann sich genau so gut darauf verständigen, daß das sog. Aleph, wie traditionell üblich, als Doppelhaken *ʾ* dargestellt wird, wie darauf, daß man es, wie in der *Glyph*-Darstellung, als großes „A“ notiert, oder, wie ich das seit Jahrzehnten – namentlich im CTTC – zu tun beliebe, als schließende runde Klammer „)“ oder, wie ich für andere Zwecke eine Weile lang gerne tat, als die Zeichenkombination Nummernzeichen plus die Ziffer 3, „#3“. Wenn

man allerdings dem Leser das Lesen erleichtern will, wird man die Transkription nach Möglichkeit den Gewohnheiten des Lesers anpassen und Aleph also als Doppelhaken ʾ drucken. Mit der internen Darstellung hat das nur insoweit etwas zu tun, als klar sein muß, welches Druckzeichen dem internen Code entspricht. Der lange Rede kurzer Sinn: Die Visualisierung ist keine Frage der computerinternen Darstellung, sie stellt ein Problem sui generis dar.

### *3. Elemente der Beschreibung des Wortlauts eines Textes im CTTC und ihre Übersetzung in eine XML-Notation*

#### *3.1 Allgemeines*

Nun zur Hauptsache: der Notation des CTTC in XML-Notation. Wie schon gesagt: Ich bin nicht im Besitz einer definitiven Lösung. Ich kann nur über Lösungsansätze und Probleme berichten. Ich halte mich, wie bereits angekündigt, nicht mit einer allgemeinen Beschreibung der XML-Notation auf, sondern führe die wenigen Notationselemente von XML, die für den gegenwärtigen Zweck gebraucht werden, am Beispiel ein. Das ist für den Ägyptologen anschaulicher und für den Zweck auch ausreichend.

*Was* in XML-Notation dargestellt werden kann, hängt davon ab, was im CTTC enthalten ist: Man kann nicht mehr notieren als im CTTC enthalten ist, allenfalls weniger. In der Tat gibt es – abgesehen von gelegentlichen Redundanzen in der CTTC-Notation – einzelne Elemente, die man möglicherweise besser nicht übernimmt. Auf solche Fälle wird noch die Sprache kommen.

*Wie gut* etwas in XML-Notation dargestellt werden kann, hängt nicht nur von den CTTC-Daten ab. Es hängt auch ab von den Darstellungsmöglichkeiten, die XML bietet, auch davon also, wie gut die CTTC-Datenstrukturen in XML abgebildet werden können. Grundsätzlich lassen sich hierarchisch geordnete Textelemente hervorragend in XML abbilden. Nicht hierarchisch geordnete Textelemente sind zwar grundsätzlich darstellbar, man hat aber keine Garantie dafür, daß jeder Browser damit zurechtkommt und daß das, was ein Browser aus den Datenstrukturen macht, für den Ägyptologen brauchbar ist. Das Augenmerk wird also zunächst auf hierarchisch geordnete Elemente gerichtet. Die Problemfälle kommen bei passender Gelegenheit.

*Wie leicht oder schwierig* sich die *Übersetzung* des CTTC in XML-Notation gestaltet, hängt von der Art der Kennzeichnung ab, die im CTTC zur Abgrenzung der Elemente im Einzelfall gewählt wurde. Finden sich Anfangs- und Endkennzeichnungen oder solche Markierungen, die man als Anfangs- und Endkennzeichnungen interpretieren kann, so ist die Übersetzung in XML im allgemeinen ziemlich einfach. Zum Beispiel könnte man das Element Wurzel, das in der CTTC-Notation der Textwörter durch expli-

zite Markierungen vor und hinter den die Wurzel darstellenden Transkriptionszeichen begrenzt ist, mühelos in die XML-Spitzklammer-Notation mit Start Tag und End Tag übersetzen. So könnte man die Wurzel *šĉm*, die im CTTC – in den Details nur aus der langen Geschichte der Erstellung des Textkorpus erklärbar – in einer mit „&[“, beginnenden und mit „!“ endenden Zeichenfolge „&[SJM!“ enthalten ist, in XML übersetzen als

```
<wurzel>šĉm</wurzel>
```

Dagegen ist es schwieriger, ein hinten komplementiertes Zweikonsonantenzeichen in der Graphemnotation wie z.B. das Zweikonsonantenzeichen *mn* mit hinterem Komplement *n* zu übersetzen, was im CTTC – ich gebe hier nicht die interne, sondern die für die Visualisierung im Druck vorgesehene Form – als *mn<sub>1</sub>*, notiert ist, d.h. als die Zeichenfolge *mn* plus ein Index „1“ (die Ziffer 1), der den Komplementierungstyp beinhaltet.<sup>4</sup> Die Zweikonsonantengruppe ist zwar hinten durch eine einstellige Ziffer, die „1“, abgegrenzt, nach vorn aber gibt es keine explizite Abgrenzung, man muß die zugehörigen Zeichen durch Abzählen bestimmen. Dies getan, kann man dann eine XML-Notation wie etwa die folgende erzeugen:

```
<komplementierte.gruppe>
  <2-kons.zeichen>mn</2-kons.zeichen>
  <komplement>n</komplement>
</komplementierte.gruppe>
```

(Die buchstabenreiche Notation wird unten §3.7 gestraft werden.)

Es gibt Fälle, in denen das Abzählen wesentlich schwieriger ist. Schwieriger ist es vor allem dann, wenn unterschiedliche einseitige Abgrenzungen zusammentreffen, und am schwierigsten, wenn sich die Länge nicht aus der hinteren Abgrenzung ergibt. Letzteres ist z.B. bei der Notation der Phonetischen Determinative der Fall, deren Lautwert sich aus vorangehenden Konsonanten ergibt, ohne daß explizit notiert ist, welche der vorangehenden Konsonanten dazugehören. Für ein Phonetisches Determinativ steht in der Transkription ein Index „g“. Z.B. steht *ʾw<sub>g</sub>* für eine Transkriptionszeichen-

<sup>4</sup> Zu den Prinzipien der Graphemnotation s. W. SCHENKEL, *Aus der Arbeit an einer Konkordanz zu den altägyptischen Sargtexten*, GOF IV 12, Wiesbaden 1983. 1–170; man beachte, daß inzwischen der oben (§2.3.2) angesprochene Gesichtspunkt der Sortierfolge zu einer Vertauschung der Komplementierungsindizes „1“ und „2“ geführt hat (eine zwischenzeitlich erwogene und leider gelegentlich angewandte Vertauschung von „2“ und „3“ beruht auf einem Denkfehler).

folge mit der Bedeutung „erben“, gefolgt von einem Phonetischen Determinativ  $\dot{w}^c$  „Fleischstück“ (Gardiner-Zeichenummer F 44). Der versierte Philologe erfaßt intuitiv und mit hinreichender Sicherheit, daß es sich bei „g“ um ein Phonetisches Determinativ  $\dot{w}^c$  und nicht etwa um ein Phonetisches Determinativ  $w^c$  handelt, da es entweder kein Phonetisches Determinativ  $w^c$  gibt oder, falls es ein solches doch gibt, dieses bei  $\dot{w}^c$  „erben“ denkbar unwahrscheinlich wäre. Noch schwieriger gestaltet sich die mechanische Bestimmung des Phonetischen Determinativs, wenn dem „g“ noch ein „c“ zugefügt ist. In diesem Fall bezieht sich „g“ nicht auf die unmittelbar vorangehende Transkriptionszeichenfolge, sondern auf einen vorangehenden „centralen“ Wortbestandteil, beispielsweise und vorzugsweise auf den Wortstamm. Zum Beispiel handelt es sich auch bei  $\dot{w}^c.t_{gc}$  der vorigen Transkriptionszeichenfolge plus Femininendung -t, um das Phonetische Determinativ  $\dot{w}^c$  „Fleischstück“, das jetzt jedoch durch die Femininendung von der zugehörigen Konsonantenfolge getrennt ist. Im einfachsten Fall muß man also wissen, was als Phonetisches Determinativ von der Konsonantenfolge her überhaupt in Betracht kommt, bzw. man muß dies anhand einer vollständigen Liste der Phonetischen Determinative entscheiden. In komplizierteren Fällen muß man darüber hinaus noch wissen, was als „centrales“ Element in Frage kommt, bzw. Regeln finden, nach denen man dieses bestimmen kann.

Anmerkung: Die zuletzt genannte Steigerung der Abgrenzungsproblematik ist übrigens nicht die Folge einer ungeschickten Notation im CTTC. Im Gegenteil: Die CTTC-Notation bildet den Befund der hieroglyphischen Graphien in durchaus adäquater Weise ab. Die Frage, ob man bei der Übersetzung in XML eine Präzisierung der Notation vornehmen kann und soll, wird unten (§3.7) anzudiskutieren sein.

Das CTTC besteht in der Hauptsache aus einer Folge von Wortformen. Die Beschreibung einer Wortform enthält im Kern

- die Transkription der Wortform, genauer gesagt: die analytische Darstellung der hieroglyphischen Graphie der Wortform.

Der Transkription sind

- vorgeschaltet Angaben zum Lemma (wovon oben in §2.3.1 schon die Rede war) und Angaben zur Flexionsform sowie
- nachgeschaltet Angaben zur Position der Wortform innerhalb des Textkorpus.

Anderes kann an dieser Stelle übergangen werden.

Vor- und Nachschaltung hängen mit der Reihenfolge der Sortiermerkmale bei der Erstellung einer lexikalischen Konkordanz, meinem nächsten Ziel, zusammen. An sich ist die Reihenfolge belanglos. Die Bedenkzeit mit

eingeschlossen, die man sich vor Umstellungsoperationen in mehr als einer halben Million Sätzen nehmen sollte, wäre auf einem durchschnittlichen PC binnen einer Viertelstunde jede andere Reihenfolge herstellbar.

Der CTTC-„Text“ besteht im wesentlichen nicht aus Text, d.h. aus den Wortformen des Textes – das sind genau genommen nur die im Zentrum stehenden Graphien –, sondern aus vielerlei Kommentaren zum Text, den Angaben zu Lemma und Flexionsform sowie zur Position der Wortform im Textkorpus (und ein paar ungenannte Kommentare mehr, die weniger zu Buche schlagen).

Ich bespreche nun die wichtigsten Teile der Wortform-Notation in großen Zügen. Ich beginne mit den Informationen zum Lemma, weil davon schon die Rede war, schließe dann die Behandlung der vorgeschalteten Informationen mit der Behandlung der zweiten Gruppe von Informationen, den Flexionsformangaben, ab. Danach bespreche ich die nachgeschalteten Positionsangaben, um dann zuletzt die Kerngruppe, die Notation der Graphie, vorzunehmen, die die komplizierteste ist und die in unserem Zusammenhang auch gar nicht in allen Einzelaspekten behandelt werden kann.

### 3.2 Die Lemma-Notation

#### 3.2.1 *Simplicia*

Das Lemma der Wortform (genauer: der *Simplicia*) ist im CTTC als eine Folge von mehreren Spezifikationen in der angegebenen Reihenfolge notiert:

- Wurzel; Wurzelindex; Wurzeltyp;
- Wort; Wortindex; Semantischer Wortindex;
- Namensindex;

und weitere Angaben, die später zur Sprache kommen.

Von Wurzel / Wurzelindex und Wort / Wortindex war bereits die Rede. Man könnte präzisierend noch hinzufügen, daß bei unabgeleiteten Verben als Wort die Wurzel gilt, das Wort also mit der Wurzel identisch ist.

Die Angabe des Wurzeltyps, z.B. „ult.inf.“, steht an dieser Stelle, weil sie zunächst als ein Index (neben dem Wurzelindex) zur Indizierung der Wurzeln konzipiert war. Der Wurzeltyp wird zu diesem Zweck nicht mehr genutzt. Die Angabe ist indes weiterhin nützlich, weil sich in Sonderfällen der Wurzeltyp nicht aus der Form der Wurzel ableiten läßt. Zum Beispiel wäre eine durch Lautwandel aus *wtč* entstandene Wurzel *wtt* „zeugen, erzeugen“ nicht notwendig ult.gem., obwohl die letzten beiden Konsonanten gleich sind. Das Beispiel ist insofern fiktiv, als in diesem Falle die Wurzel im CTTC mit der älteren Lautform *wtč* angesetzt ist. Wenn aber die ältere Lautform etwa nicht überliefert wäre, würde man möglicherweise die Wurzel

mangels genauerer Information als *wtt* ansetzen müssen, auch wenn die Wurzel nicht ult.gem. ist.

Der semantische Wortindex ist im Prinzip nützlich, in der Praxis jedoch schwer zu handhaben. Mit seiner Hilfe können Bedeutungen eines und desselben Wortes voneinander unterschieden werden. Zum Beispiel kann man auf diese Weise Belege für *nʿt* „Stadt“ von Belegen desselben Wortes im Sinne von „Stadt als ihre Einwohner, Stadtbevölkerung“ unterscheiden. Auch die Ägypter haben, wenn gewiß auch nicht konsequent, dennoch vielfach diesem Unterschied durch die Determinierung Rechnung getragen. Mit einem solchen Index arbeiten implizit, wenn auch von Wort zu Wort mit immer wieder anderen Abgrenzungen, die Wörterbücher, wenn sie Wortartikel nach semantischen Kriterien unterteilen. Bei der Erfassung eines Textkorpus ergeben sich praktische Probleme, auf die unten bei der Besprechung der Übersetzung die Rede kommen wird.

Der Namensindex kennzeichnet Namen unterschiedlicher Art. Diese Kennzeichnung ist indes problematischer Natur, da sich vielfach nicht entscheiden läßt, ob ein Wort Name oder Bezeichnung ist und der Übergang zwischen Objekt und seiner Personifikation fließend ist. Zum Beispiel kann man für ein X in „X ist sein (eines göttlichen Wesens) ‚Name‘“ oft nicht entscheiden, ob X der Name des göttlichen Wesens ist oder nur eine charakterisierende Bezeichnung. Oder es ist eine Grenzziehung zwischen dem „Westen“ als einem Ort und seiner Personifikation als Westgöttin schwer durchzuhalten. Den letzteren Fall beleuchtet sehr schön die schwankende Determinierung, derzufolge die Ägypter selbst schwankten, ob sie ein Personendeterminativ gebrauchen sollten oder das Fremdlandzeichen oder beides zusammen. Verzichtet man auf die Kennzeichnung der Namen, entfällt auch die korrespondierende, derzeit im CTTC noch vorhandene Großschreibung der Wörter, sofern es sich um Namen handelt. (Die ägyptologischen Linguisten sind sich uneins, ob man Namen groß schreibt: GARDINER im vorzüglichen Wortverzeichnis seiner *Egyptian Grammar* tut es, das Berliner Wörterbuch tut es nicht.)

So weit könnte man für die Lemma-Informationen etwa folgende XML-Notation wählen:

```
<lemma.simplex>
  <wurzel wurzel.index=„WURZELINDEX“
    wurzel.typ=„WURZELTYP“>WURZEL</wurzel>
  <wort wort.index=„WORTINDEX“
    semant.wort.index=„SEMANT.WORT.INDEX“>WORT</wort>
</lemma.simplex>
```

D.h. das Element Lemma enthält zwei Elemente: das Element Wurzel und das Element Wort. Der Wurzelindex und der Wurzeltyp sind Attribute zum Element Wurzel, der Wortindex und der semantische Wortindex Attribute zum Element Wort.

Anmerkung: Die Indizes unterscheiden von „l“ (Wurzelindex, Wortindex) bzw. „a“ (semantischer Wortindex) an lückenlos aufsteigend die Homonyme, die tatsächlich im CTTC vorkommen. Meine ursprüngliche Vergabe von Indizes anhand einer präfabrizierten, einen größeren Ausschnitt des Lexikons erfassenden „Indexliste“<sup>5</sup> ist obsolet, wenn auch noch weitgehend im CTTC intakt erhalten.

Zwei weitere auf das Wort bezogene Informationseinheiten bedürfen genauerer Erörterung: die Übersetzung und der Wortbildungstyp.

Wie der Wurzeltyp ist auch die Übersetzung eine Angabe, die keine distinguierende Funktion besitzt. Im Gegensatz zum Wurzeltyp, der wenigstens noch ein der Wurzel inhärierendes Merkmal bezeichnet, handelt es sich bei der Übersetzung nur um den unvollkommenen Versuch, die Bedeutung eines Wortes in Begriffen einer Übersetzungssprache näherungsweise zu beschreiben. Schon aus diesem Grund sollte die Übersetzung nicht in die Beschreibung des Wortes integriert werden. Hinzu kommt, daß die Einbeziehung der Übersetzung in die Wortbeschreibung als Attribut diese wegen der Länge der Übersetzungen unübersichtlich macht (und u.U. wegen der für Attribute geltenden Beschränkungen bezüglich des Zeichensatzes notationstechnische Probleme aufwerfen könnte). Es empfiehlt sich also, in die Lemma-Notation neben Wurzel und Wort noch als drittes Element das Element „Übersetzung“ einzufügen:

```
<lemma.simplex>
  <wurzel ...>...</wurzel>
  <wort ...>...</wort>
  <übersetzung>ÜBERSETZUNG</übersetzung>
</lemma.simplex>
```

Die Übersetzungen im CTTC sind Lemma-Übersetzungen. Sie sind also korrekt in der Lemma-Beschreibung untergebracht. Es fragt sich allerdings, ob es nicht besser gewesen wäre, statt Lemma-Übersetzungen textstellenbezogene Übersetzungen zu notieren bzw. Lemma- und textstellenbezogene Übersetzungen beide unabhängig voneinander. (Wären textstellenbezogene

<sup>5</sup> W. SCHENKEL, *Maschinelle Analyse altägyptischer Texte*, Darmstadt 1969, 177–183.

Übersetzungen notiert worden, wären diese selbstverständlich nicht in der Lemma-Beschreibung unterzubringen, von der im Augenblick die Rede ist, sondern in einer Wortform-Beschreibung, von der später die Rede sein wird.) Wie bereits gesagt, wurden bei den Wörtern mit Hilfe eines semantischen Wortindex Differenzierungen der Wortbedeutung vorgenommen. Die Übersetzung bezieht sich folglich nicht auf das Lemma als ganzes, sondern auf die durch den semantischen Wortindex bezeichnete Teilmenge der Belege. Zum Beispiel lautet die Übersetzung zu *n'.t* im Sinne von „Stadt“ mit dem semantischen Index „a“ „Stadt“, zu demselben Wort *n'.t* im Sinne von „Stadt als ihre Einwohner, Stadtbevölkerung“ mit dem semantischen Index „b“ dagegen „Stadtbewohner (koll[ektiv])“.

Dagegen ist im Prinzip nichts einzuwenden. Ein Problem ergibt sich jedoch daraus, daß bei der Aufnahme der Texte noch gar nicht klar ist, welche Wortbedeutungen in einem Textkorpus zweckmäßigerweise unterschieden werden. Das läßt sich allenfalls bestimmen, wenn man einen Überblick über die Gesamtmenge der Vorkommen eines Wortes im Textkorpus hat. Es wäre also eine Endredaktion erforderlich, wie man sie üblicherweise bei einem Wörterbuch durchführt. Andererseits könnte man sich natürlich auch an den semantischen Gliederungen der Wortartikel in vorhandenen Wörterbüchern orientieren und alle dort getroffenen semantischen Differenzierungen bereits bei der Aufnahme eines Textkorpus berücksichtigen. Ginge man so vor, würde man möglicherweise unnötige Arbeit investieren, weil die im Textkorpus vorhandenen Belege für die einzelnen Bedeutungen so gering an Zahl sein können, daß man, um nicht zu kleinteilig zu differenzieren, letztendlich doch wieder am besten die Differenzierungen aufheben würde. Aber vor allem: Bieten denn die vorhandenen Wörterbücher überhaupt die richtigen und alle erforderlichen semantischen Differenzierungen? Will man nicht auch überhaupt erst solche Fragen anhand des Textkorpus klären, an dem man arbeitet?

Im Falle des CTTC habe ich semantische Differenzierungen mit Hilfe semantischer Indizes nach meinem Kenntnisstand vor Erarbeitung des Textkorpus vorgenommen. Bei einer späteren Kollation der Textaufnahmen anhand der Edition wurde überprüft, ob die Übersetzungen in etwa die Bedeutung eines Wortes an der einzelnen Textstelle abdecken. War dies nicht der Fall, wurde die Lemma-Übersetzung um weitere Übersetzungsäquivalente erweitert. Zum Beispiel wurde ein ursprüngliches *wz* „öffnen“ um die Übersetzungsalternative „sich öffnen“ erweitert. Dabei stellte sich zunehmend die Einsicht ein, man könne Gruppen von Bedeutungen unterscheiden, die eigentlich mit Hilfe des semantischen Index zu unterscheiden wären und nicht bloß und nicht erst durch die Zurwahlstellung von Übersetzungsäquivalenten. Um beim Beispiel zu bleiben: Man könnte sich überlegen, ob *wz* „sich

öffnen“ = „aufgehen“ und *wn* „öffnen“ = „aufgehen lassen“ auf der Basis eines innerägyptisches Merkmals [ $\pm$  KAUSATIVITÄT] durch semantische Indizes voneinander zu unterscheiden wären. (Die Sachfrage kann hier nicht weiter verfolgt werden. Es geht im gegenwärtigen Zusammenhang lediglich um die Veranschaulichung des Problemtyps.) Wollte man eine solche nachträgliche Differenzierung vornehmen, müßte man für alle einzelnen Vorkommen eines Wortes, also z.B. des Wortes *wn* „öffnen / sich öffnen“, im Textkorpus eine Entscheidung fällen, welcher der neu eingeführten semantischen Indizes im Einzelfall zutrifft. Wenn Wörter nicht allzu oft im Textkorpus vorkommen, ist dies leicht zu bewerkstelligen. Aber gerade dann, wenn die Überarbeitung wünschenswert wäre, wenn nämlich viele Belege vorliegen und die semantische Differenzierung ordnende Wirkung hätte, wäre so viel Nacharbeit erforderlich, daß sie die Arbeitskraft – oder die Programmierbarkeit – übersteigen kann. Im Falle des CTTC ist der Stand der Dinge der, daß eine gründliche Überarbeitung namentlich der Lemmata anstünde, bei denen sich semantisch leicht zu differenzierende Übersetzungsäquivalente angehäuft haben. Das ist selbstverständlich kein Problem, das im Zuge der Übersetzung der CTTC Notation in XML-Notation zu lösen ist.

Die Relation zwischen Wurzel und Wort spielt namentlich bei den Nomina eine Rolle. Dementsprechend finden sich im CTTC Angaben zum Verfahren der Ableitung des Wortes aus der Wurzel, also zum Nominalbildungstyp. Zweckmäßigerweise fügt man den Nominalbildungstyp dem Element Wort als Attribut zu, etwa analog zur Angabe des Wurzeltyps bei der Wurzel:

Wurzel:

```
<wurzel wurzel.index=„WURZELINDEX“
      wurzel.typ=„WURZELTYP“>WURZEL</wurzel>
```

Entsprechend Wort:

```
<wort wort.index=„WORTINDEX“
      wortbild.typ=„WORTBILDUNGSTYP“
      semant.wort.index=„SEMANT.WORT.INDEX“>WORT</wort>
```

Ein Problem ergibt sich jedoch daraus, daß ich zu verschiedenen Zeiten zwei unterschiedliche Beschreibungen der Wortbildung in das CTTC aufgenommen habe. Es fragt sich also, welche der beiden Beschreibungen man aufnimmt oder ob man etwa beide aufnimmt.

Ein älterer Eintrag findet sich im CTTC in einem morphologischen Code, in dem morphologische Kategorien unterschiedlicher Art notiert sind, flexivische Kategorien, von denen später noch die Rede sein wird, und wortbezo-

gene Kategorien wie der Ableitungstyp, aber auch, worauf ebenfalls später noch die Sprache kommt, die Wortart. Die hier enthaltenen Angaben zur Wortbildung beziehen sich auf das konsonantische Skelett, fallweise aber auch auf Differenzierungen, die sich aus Bedeutung und zu vermutender Vokalisation ergeben. Solche Angaben finden sich grundsätzlich bei jedem Wort, das abgeleitet ist.

Ein neuerer Eintrag spezifiziert die Zugehörigkeit zu einer Nominalbildungsklasse nach OSINGs Nominalbildungslehre<sup>6</sup>, unter fallweiser Hinzufügung von Korrekturen, die ich selbst an OSINGs Aufstellungen machen zu müssen glaubte.<sup>7</sup> Solche Angaben finden sich naturgemäß bei den Wörtern, bei denen man sich eine gewisse Vorstellung von der Vokalisation machen kann. Das sind die wenigsten. Ich tendiere dahin, die älteren, weniger informativen Angaben zur Wortbildung zurückzu behalten und nur die informativeren, wenn auch nur sporadisch angebbaren genaueren Nominalbildungstypen OSINGs bzw. deren Modifikationen anzugeben. Es wäre dann für das Element Wort als Attribut der Nominalbildungstyp vorzusehen, der beispielsweise für das Wort *čwḏ.wt* „,sich Aufrichtende“, Kobra“, bei dem man über den Nominalbildungstyp streiten kann, etwa so aussähe:

wortbild.typ=„A II 6 (RDNBÄ: III 1 b, NBÄ: III 1)“

mit „A II 6“ für die im CTTC gewählte Lösung, „A III 1 b“ für die in meiner „Rekonstruktion der deverbale Nominalbildung“ bestimmte und mit „A III 1“ für die von OSING in seiner „Nominalbildung des Ägyptischen“ gegebene Lösung.

Anmerkung: Die Verbalbildung ist, worauf hier nicht näher eingegangen zu werden braucht, durch den Wurzeltyp abgedeckt.

Noch eine weitere wortbezogene Angabe läßt sich aus dem morphologischen Code des CTTC extrahieren: die Wortart. Das ist eine nicht unproblematische Kategorie, und zwar deshalb, weil durch diese teils syntaktische Funktionen eines Wortes, teils seine Flexionsart charakterisiert sind. Das ist ein Grundproblem, das nicht in diesem Zusammenhang geklärt werden kann. Immerhin ist die Wortartangabe ein nützlicher, weil einfach zu handhaben-

<sup>6</sup> J. OSING, *Die Nominalbildung des Ägyptischen*, Mainz 1976.

<sup>7</sup> W. SCHENKEL, *Zur Rekonstruktion der deverbale Nominalbildung des Ägyptischen*, Wiesbaden 1983.

der „Tag“ für ein eventuelles syntaktisches Parsing. Man könnte demnach in die Wortbeschreibung ein weiteres Attribut wie folgt einfügen:

```
<wort wort.index=„WORTINDEX“
      wortbild.typ=„WORTBILDUNGSTYP“
      wort.art=„WORTART“
      semant.wort.index=„SEMANT.WORT.INDEX“>wort</wort>
```

Ein konkretes Beispiel einer Wortartangabe wäre etwa die Wortart „Substantiv“, die man wie folgt übersetzen könnte:

```
wort.art=„Subst.“
```

Man muß sich darüber im klaren sein, daß es keine kanonische Wortarteneinteilung gibt. Wenn man mit den Wortarten des CTTC arbeiten will, muß man den Katalog dieser Wortarten vor sich liegen haben.<sup>8</sup> In XML kann dieser Katalog in der DTD, der „Document Type Definition“, definiert werden. Da ich, wie bereits gesagt, auf die an sich wünschenswerte, wenn nicht unentbehrliche DTD nicht näher eingehe, möchte ich die Gelegenheit nutzen, wenigstens einmal eine DTD-Definition als Beispiel zu geben. Eine solche enthielte für die Wortarten etwa die folgende Definition:

```
<!ATTLIST wort
      wort.art (Subst. | Verb | ... | unbekannt) „unbekannt“>
```

D.h. als Wortart-Attribute zu Wort wären „Subst.“, „Verb“ etc. zulässig, im Falle der Unentscheidbarkeit das Attribut „unbekannt“. Fehlt die explizite Angabe einer Wortart, wird „unbekannt“ angenommen.

<sup>8</sup> W. SCHENKEL, *Maschinelle Analyse altägyptischer Texte*, Darmstadt 1969, 67–70.

### 3.2.2 Komposita

Bislang war von der Lemmatisierung der einfachen Wörter, der *Simplicia*, die Rede. Es gibt aber im Ägyptischen auch Komposita und feste Wortverbindungen, die man zweckmäßigerweise als solche verfügbar macht. Unter Komposita fallen zum mindesten Namen. Was man als feste Wortverbindungen ansetzen will, darüber kann man gut streiten. Ich fasse beide Fälle, durchaus angreifbar, unter dem Begriff „Kompositum“ zusammen. Es geht mir hier nicht um den Begriff, sondern um die Sache: die pragmatische Behandlung der angesprochenen Fälle.

Komposita müssen erstens als solche explizit in der Folge der einzelnen Wortformen abgegrenzt werden und sie müssen teilweise lemmatisiert werden. Ein Name wird – wie natürlich auch der nichtzusammengesetzte Name – nicht lemmatisiert, er hat nur eine einzige Flexionsform, nämlich die Folge der in ihm enthaltenen Wortformen. Zum Beispiel wäre ein Satzname wie *nčr.w-ḥtp.wi* o.ä. „Die-Götter-sind-zufrieden“ bereits als solcher, mit unveränderten pluralischen Wortformen, das Kompositum-Lemma. Bei anderen Komposita ist zum mindesten die erste Wortform zu lemmatisieren. Zum Beispiel ist die pluralische Wortform *im.(i)w-ḥt* „dahinter befindliche“ zu lemmatisieren in den Singular *im.(i)-ḥt* „dahinter befindlicher“. Leider weiß man gelegentlich nicht, ob ein Kompositum ein Name ist oder nicht, ob es also zu lemmatisieren ist oder nicht. Typischerweise kann man, wie schon bei der Behandlung der *Simplicia* zur Sprache kam, im Korpus der Sargtexte vielfach nicht unterscheiden, ob ein Kompositum, das als der *m* „Name“, wie man zu übersetzen pflegt, eines göttlichen Wesens bezeichnet ist, wirklich ein Name ist und nicht doch bloß eine Bezeichnung eines göttlichen Wesens nach seinen Eigenschaften. Unter den Nicht-Namen gibt es selten auch Fälle, in denen eine weitere enthaltene Wortform zu lemmatisieren ist, typischerweise ein attributives Adjektiv / Partizip. Zum Beispiel sind *t'.(w)-ḥç.(w)* „die Weißbrote“ zu lemmatisieren in *t'-ḥç* „das Weißbrot“. Schließlich gibt es leider auch Fälle, in denen man nicht entscheiden kann, ob ein nicht-erstes Element zu lemmatisieren ist oder nicht. Zum Beispiel könnte ein *imn.(w)-m.(w)* „die (vielen) mit (vielen) verborgenen Namen“ im Singular entweder *jmn-m.(w)* „der mit (vielen) verborgenen Namen“ oder *jmn-m* „der (einzelne) mit (einem einzelnen) verborgenen Namen“ sein. Die im CTTC gewählte Lösung für solche Fälle ist die, den Plural beim zweiten Element stehen zu lassen und – behelfsweise – der Übersetzung des Kompositums den Zusatz „im Plural“ zuzufügen: *jmn-m.(w)* „der mit verborgenen Namen (im Plural)“. Genau genommen ist natürlich nicht die Übersetzung der Ort, an dem eine solche Aussage zu machen wäre.

Die explizite Lemmatisierung der Komposita ist auch aus einem weiteren Grund erforderlich. Es kommt vor, daß in der Folge der Bestandteile eines Kompositums lemmatisierbare Wortformen auftreten, die nicht zum Kompositum gehören, sei es, daß sie in den sonst fehlenden Bestandteil eines Kompositums emendiert werden müssen, sei es, daß sie ersatzlos zu eliminieren sind. Zum Beispiel kann statt der Partikel  $\acute{ir}=\check{c}$  infolge mangelhafter Textadaptation auf dem Sarg einer Frau das auf einen Mann bezogene  $\acute{ir}=k$  stehen geblieben sein. Es lohnt in solchen trivialen Fällen nicht unbedingt, die Partikel als ganze zu emendieren, den Sachverhalt somit als  $\{\acute{ir}=k\} \langle \acute{ir}=\check{c} \rangle$  zu notieren. Man kann auch auf das fehlerhafte Kompositum verzichten und nur das durch Emendation hergestellte nachweisen wollen. In diesem Fall würde man  $\acute{ir}\{=k\} \langle =\check{c} \rangle$  notieren. Dabei bleibt das nicht zum Kompositum gehörige  $\{=k\}$ , für sich genommen, immer noch ein korrektes Wort (vorausgesetzt, man versteht Suffixpronomina, wie im CTTC der Fall, als Wörter).

Anmerkung: Während im Beispielfall  $\{=\check{c}\}$  aus dem Lemma eliminiert wird, kann es in der Notation der Graphie (s. unten §3.7) stehen bleiben (z.B. die Standard-Graphie  $\acute{ir}\{=k\} \langle =\check{c} \rangle$ ), so daß also über die Graphien von  $\acute{ir}=\check{c}$  notfalls auch die verworfenen Belege für  $\acute{ir}=k$  aufgefunden werden können.

Ob man im Beispielfall so verfahren will oder nicht (im CTTC wird so verfahren), tut nichts zur Sache. Es gibt andere, kompliziertere Fälle, in denen man so verfahren muß, dann nämlich, wenn ein für sich genommen sinnvolles Wort unter keinen Umständen Bestandteil des Kompositums sein kann. Ein solcher Fall liegt z.B. vor, wenn ein göttliches Wesen  $\acute{hnt}(\acute{i})-m(\acute{i})n(\acute{i}).wt=f$  „Der auf seinen Pflöcken ist“ heißt, der Schreiber aber statt  $m(\acute{i})n(\acute{i}).wt$  fälschlich  $(\acute{i}:)mn.yt$  „dauerndes Opfer“ verstand und schrieb.<sup>9</sup> Der philologisch bereinigte Befund lautet  $\acute{hnt}.\acute{i}\{-(\acute{i}:)mn.yt\} \langle -m\acute{i}n\acute{i}.wt \rangle =f$ . Als Lemmata sind anzusetzen sowohl das Kompositum  $\acute{hnt}(\acute{i})-m(\acute{i})n(\acute{i}).wt=f$  als auch das Simplex  $(\acute{i}:)mn.yt$ .

Im CTTC ist das Lemma der Komposita beim Kopfwort untergebracht. Derzeit ist es dort noch hierarchisch dem semantischen Wortindex nachgeordnet. Letzteres hat sich, abgesehen von der Problematik des semantischen Wortindex, von der die Rede war, für die Anordnung der Wörter in lexikalischer Folge als ungünstig erwiesen. Es hätte dies nämlich zur Folge, daß etwa mit der Präposition  $m$  zusammengesetzte Präpositionen unter den  $m$ 's mit verschiedenen semantischen Indizes (lokales  $m$ , modales  $m$ , ...) aufträten, so daß man bisweilen zweifeln könnte, unter welches  $m$  das Komposi-

<sup>9</sup> CTI 290a (T1C).

tum gehört bzw. unter welchem *m* man suchen soll, ganz abgesehen von dem unangenehmen Fall, daß eine formal gleiche Wortverbindung auch auf mehr als eines der *m*'s verteilt sein kann. Zum Beispiel könnte es geschehen, daß *m-sʃ* „im Rücken von, hinter“ unter lokalem *m* „an (einem Ort)“ zu finden wäre, *m-sʃ* „\*im Danach?, nach“ dagegen unter temporalem *m* „zu (einer Zeit)“.

Ob eine solche Aufteilung sinnvoll wäre, ist eine schwer zu beantwortende Sachfrage. Geht man davon aus, daß, wenn ein Wort lokale und temporale Bedeutung hat, die lokale die ursprüngliche ist, so stellt sich die Frage, ob die temporale Bedeutung von *m-sʃ* von lokalem *m-sʃ* herzuleiten ist oder ob zuerst unabhängig voneinander die Präposition *m* und das Substantiv *sʃ* in temporaler Bedeutung gebraucht und temporales *m-sʃ* aus temporalem *m* und temporalem *sʃ* zusammengesetzt worden wären. Ob man die Sachfrage entscheiden kann oder nicht: Bevor man sich in diesem und anderen solchen Fällen in Spitzfindigkeiten verliert, sollte man die Frage der (z.B.) lokalen vs. temporalen Einordnung der Bestandteile auf sich beruhen lassen, nur ein einziges Kompositum ansetzen und erst bei diesem mit Hilfe eines semantischen Index die Unterscheidung nach (z.B.) lokalem vs. temporalem Gebrauch treffen. Im Beispielfall gäbe es also neben der Präposition *m* mit lokaler bzw. temporaler Bedeutung ein semantisch unspezifiziertes *m* als (erstem) Bestandteil von Komposita.

Ein semantischer Index für Komposita wird im übrigen ohnehin benötigt. Zum Beispiel läßt sich analog zu dem oben beispielsweise angeführten *n.t* „Stadt“ vs. „Stadt als ihre Einwohner, Stadtbevölkerung“ unterscheiden zwischen *im.ʃ-wr.t-ʃ* „Steuerbord (vorn)“ und *im.ʃ-wr.t-ʃ* „(vorn) Steuerbordseitige (Personen)“. Ich werde also voraussichtlich, zum mindesten in einem Teil der Fälle, noch vor der Übersetzung des Textkorpus in eine XML-Notation die Indizierung so ändern, daß eine semantische Differenzierung auf der Ebene der Bestandteile unterbleibt und bei Bedarf erst auf Kompositum-Ebene vorgenommen wird.

Dieses Detailproblem zunächst außer acht gelassen, sollten in der XML-Notation in die Lemmagruppe das Lemma des Kompositums und ein Kompositum-Index aufgenommen werden, was etwa so aussehen könnte (ich füge stillschweigend die Wortart des Kompositums hinzu und vorläufig auch noch einen semantischen Kompositum-Index):

```
<lemma.kompos>
  <wurzel wurzel.index=„WURZELINDEX“
    wurzel.typ=„WURZELTYP“>WURZEL</wurzel>
  <wort wort.index=„WORTINDEX“
    wortbild.typ=„WORTBILDUNGSTYP“
```

```

wort.art=„WORTART“
semant.index=„SEMANT.INDEX“ >WORT</wort>
<kompositum kompos.index=„KOMPOS.INDEX“
semant.kompos.index=„SEMANT.KOMPOS.INDEX“
kompos.wort.art=„WORTART“
semant.kompos.index=„SEMANT.KOMPOS.INDEX“>
KOMPOSITUM</kompositum>
</lemma.kompos>

```

Konkretes Beispiel: *im.ĭ-wr.t-ŝ* „(vorn) Steuerbordseitige“ (die in geschweifte Klammern eingeschlossene Angabe evtl. als der merkmallöse Fall, die in eckige Klammern eingeschlossene Angabe als der Fall ohne Alternativen innerhalb des Textkorpus wegläßbar):

```

<lemma.kompos>
  <wurzel [wurzel.index=„1“]
    {wurzel.typ=„firm“}>im</wurzel>
  <wort [wort.index=„1“]
    wortbild.typ=„B III 1 A“
    wort.art=„Nisba“
    semant.index=„a“>im.ĭ</wort>
  <kompositum [kompos.index=„1“]
    kompos.wort.art=„Subst.“
    semant.kompos.index=„b“>im.ĭ-wr.t-ŝ>
  </kompositum>
</lemma.kompos>

```

Will man jetzt, wie oben ins Auge gefaßt, eine doppelte semantische Differenzierung, auf Simplex- und auf Kompositum-Ebene, vermeiden, bieten sich zwei alternative Lösungen an: Man kann, wie oben erwogen, den semantischen Index auf Simplex-Ebene als unspezifiziert kennzeichnen. Würde man z.B. lokalem *m* den semantischen Wortindex „a“ zusprechen, temporalem *m* den semantischen Wortindex „b“, so wäre neben beiden noch ein weiterer semantischer Wortindex anzusetzen, etwa „x“, der für „nicht spezifiziert“ steht, oder es müßte der Index in letzterem Fall einfach weggelassen werden. Im Falle des Kompositums *im.ĭ-wr.t-ŝ* „(vorn) Steuerbordseitige“ stünde im Element „Wort“ dann entweder ein semantischer Index „x“:

```

<wort [wort.index=„1“]
  wortbild.typ=„B III 1 A“

```

```
wort.art=„Nisba“semant.wort.index=„x“>im.í</wort>
```

oder der semantische Index entfiel:

```
<wort [wort.index=„1“]
  wortbild.typ=„B III 1 A“
  wort.art=„Nisba“>im.í</wort>
```

Eine andere Lösung wäre die, die semantische Untergliederung der Lemmata in einem Unter-Element zum Element „Lemma“ zu beschreiben. In diesem Fall könnte man die semantische Untergliederung fallweise für die Simplicia nutzen, fallweise für die Komposita. Man sollte dann jedoch definitiv darauf verzichten, einen ersten Bestandteil eines Kompositums semantisch zu spezifizieren. Allgemein sähe diese Lösung folgendermaßen aus:

```
<lemma.kompos>
  <wurzel wurzel.index=„WURZELINDEX“
    wurzel.typ=„WURZELTYP“>WURZEL</wurzel>
  <wort wort.index=„WORTINDEX“
    wortbild.typ=„WORTBILDUNGSTYP“
    wort.art=„WORTART“>WORT</wort>
  <kompositum kompos.index=„KOMPOS.INDEX“
    kompos.wort.art=„WORTART“>KOMPOSITUM
  </kompositum>
  <semantik>SEMANT.INDEX</semantik>
</lemma.kompos>
```

Im konkreten Beispiel:

```
<lemma.kompos>
  <wurzel [wurzel.index=„1“]
    {wurzel.typ=„firm“}>im</wurzel>
  <wort [wort.index=„1“] wortbild.typ=„B III 1 A“
    wortart=„Nisba“>im.í</wort>
  <kompositum kompos.index=„1“wort.art=„Subst“>
    im.í-wr.t-ꜥ</kompositum>
  <semantik>b</semantik>
</lemma.kompos>
```

Die Wortart des Kompositums kann mit der Wortart des Kopfwortes identisch sein. Zum Beispiel hat *im.(i)-ht* „dahinter befindlicher“ dieselben syn-

taktischen Funktionen und dieselbe Art der Flexion wie das Simplex *ím.(i)* „darin befindlicher“. Es handelt sich beim Kompositum also ebenso wie beim Kopfwort um eine Nisba. Es kann aber auch anders sein. Im Falle des Beispiels *ím.í-wr.t-ḡ* ist das Kopfwort ebenfalls eine Nisba, das Kompositum dagegen ein (nicht-genusvariables) Substantiv. (In der derzeitigen Fassung des CTTC ist diesem Tatbestand nur so weit Rechnung getragen, wie er Folgen für die lexikalische Konkordanz hat. Ansonsten befinden sich die Angaben – infolge arbeitspraktischer Entscheidungen – in einem undefinierten Zustand, müßten also für eine korrekte Darstellung in XML noch überarbeitet werden.)

Eine Komplikation ergibt sich daraus, daß Komposita in Komposita enthalten sein können und dabei ein Wort das Kopfwort zu mehr als einem Kompositum sein kann. So könnte man z.B. *ím.(i)w-ht-hr.(w)* „die Gefolgsleute des Horus“ als Kompositum verstehen, zugleich aber das darin enthaltene *ím.(i)-ht* als ein Kompositum führen wollen. Das ist im CTTC so gelöst, daß für beide Komposita je ein kompletter Datensatz angelegt ist, von denen der für das längere Kompositum zuständige zuerst steht, gefolgt vom Datensatz für das enthaltene Kompositum. Beschränkt man sich auf die bislang allein behandelte Lemmagruppe und reduziert diese der Übersichtlichkeit halber auf einen Kern, sieht das Beispiel in XML-Notation etwa so aus:

```
<lemma.kompos>
  <wurzel>ím</wurzel>
  <wort>ím.í</wort>
  <kompositum>ím.(i)w-ht-hr.(w)</kompositum>
</lemma.kompos>
<lemma.kompos>
  <wurzel>ím</wurzel>
  <wort>ím.í</wort>
  <kompositum>ím.(i)-ht</kompositum>
</lemma.kompos>
```

(Die Duplizierung des Kopfeintrags entspricht dem generellen Verfahren, alle Bestandteile eines Kompositums, auch die darin enthaltenen Simplicia, noch einmal gesondert zu notieren; hierzu unten §3.4.)

Zum Schluß noch eine Anmerkung: In der CTTC-Notation ist auch an trennbare und Distanzkomposita gedacht, z.B. an *tp.í-b'* „auf Erden Befindlicher, Hinterbliebener“, wozu *tp.í=f-b'* „sein auf Erden Befindlicher, sein Hinterbliebener“, oder an *m<sup>3c</sup>-hr-w* „gerechtfertigt sein“, wozu *hr-w* NN. *pn m<sup>3c</sup>.w* „dieser NN. ist gerechtfertigt“. Man kann sich im Einzelfall natürlich darüber streiten, ob man derartige Wendungen als standardmäßige Wortver-

bindungen, „Komposita“, notiert. Meine praktische Erfahrung bei der Arbeit an der lexikalischen Konkordanz zu den Sargtexten ist die, daß man eher zu viel des Guten tun sollte als zu wenig. Der Ansatz von Komposita erleichtert ganz erheblich die Benutzung eines lexikalischen Verzeichnisses, weil er den Belegbestand der *Simplicia* entlastet.

### 3.3. *Wortform / Flexionsform*

Im CTTC sind die (realen) Wortformen / Flexionsformen in einer analytischen Transkription der Hieroglyphenzeichenfolge notiert (Genaueres hierzu unten §3.7). Diese Notation ist kompliziert und verklausuliert und wird auch in einer XML-Notation kompliziert bleiben, auch wenn man eine weniger verklausulierte Darstellung wählt. Es dürfte sich indes empfehlen, zusätzlich eine Einfach-Transkription vom Typ der Transkription bereit zu stellen, die JAMES P. ALLEN als „written forms“ bezeichnet.<sup>10</sup> Diese Transkription sollte sich im wesentlichen auf die real bezugten Konsonanten und, der besseren Lesbarkeit halber, linguistische Strukturzeichen beschränken. Eine solche Transkription ist der analytischen Transkription des CTTC ableitbar. Einem menschlichen Leser würde die Umformung nicht allzu schwer fallen, so daß er sie ad hoc vornehmen könnte. Die Umsetzung ist indes nicht so trivial, daß man ad hoc die Regeln formulieren könnte, die für eine automatische Umsetzung per Programm notwendig sind. Aus diesem Grund sollte man in eine XML-Notation die „written forms“ explizit einsetzen.

Was die Details der Notation von „written forms“ angeht, so sind drei Punkte zu besprechen:

Ein Logogramm kann nicht grundsätzlich als eine Lautfolge interpretiert werden. Bei logographischen Schreibungen sind nur die Konsonanten gewährleistet, die als Komplemente (oder auch als phonetische Determinative) zugefügt sind. Im Prinzip wäre der Genauigkeit Rechnung getragen, wenn man die nicht-komplementierten Teile der konsonantischen Transkription markierte. Ich schlage vor, statt dessen die zum Logogramm gehörigen Konsonanten insgesamt zu markieren, unabhängig davon also, ob sie komplementiert sind oder nicht. Man erkennt auf diese Weise das Logogramm. Das ist informativer als die konsonantischen Details. Es genügt völlig, daß mit dieser Kennzeichnung ein Ausrufezeichen gesetzt ist, das einen davor warnt, sich auf den konsonantischen Wortlaut blindlings zu verlassen. Wenn man sich ausnahmsweise für den Befund interessiert, kann man die analytische Darstellung der Graphie zu Rate ziehen oder schlicht und einfach die hiero-

<sup>10</sup> J.P. ALLEN, *The Inflection of the Verb in the Pyramid Texts*, Malibu 1984.

glyphische Textedition aufschlagen. Die vorgeschlagene Lösung, ob nun zweckmäßig oder nicht, hat im Falle des CTTC immerhin den Vorzug, daß sie konzeptionell gut zur Notation textkritisch aufbereiteter Folgen von konsonantischen Transkriptionszeichen paßt, bei denen auch Global-Information vor Detail-Information der Vorzug gegeben wurde. Die textkritischen Zeichen – auf deren Notation unten (§3.7) einzugehen sein wird – beziehen sich nämlich in der analytischen Transkription des CTTC bei komplementierten Zeichengruppen primär auf die Gruppe als ganze. Infolge dessen können textkritische Zeichen fallweise nicht exakt auf die einzelnen Konsonanten der „written forms“, die nur Einzelkonsonanten, aber keine Gruppen von Zeichen notieren, verteilt werden. Das hat zur Folge, daß ein einzelner Konsonant besser erhalten sein kann als angegeben, d.h. besser als die Gruppe, in der ihn die analytische Transkription bewertet.

Problematisch bleibt grundsätzlich, wie in jeder üblichen Transkription, daß gewisse Informationen über den Lautbestand, die durch Interpretation aus der analytischen Transkription gewonnen werden können, grundsätzlich ignoriert werden, z.B. Pluralstriche, die fallweise als die Wiedergabe einer Endung *.w* interpretiert werden können.

In XML-Notation wäre somit in die Notation des Textwortes neben der Lemmagruppe eine Wortformgruppe / Flexionsformgruppe aufzunehmen („lemma“ für entweder „lemma.simplex“ oder „lemma.kompos“):

```
<textwort>
  <lemma>...</lemma>
  <wortform>WRITTEN.FORM</wortform>
</textwort>
```

Bei Komposita ist als „written form“ die „written form“ des Kompositums zu notieren (und zwar möglicherweise bei tatsächlich getrennten Komposita unter expliziter Angabe der Trennstelle):

```
<textwort>
  <lemma.kompos>...</lemma.kompos>
  <wortform>WRITTEN.FORM1-TRENNZEICHEN-
  WRITTEN.FORM2...</wortform>
</textwort>
```

*Auch zur Wortform sind Attribute hinzufüßbar. An erster Stelle sind hier die kategorialen Spezifikationen zu nennen, die im morphologischen Code des CTTC enthalten sind. Diese sind von Wortart zu Wortart teilweise verschieden. Es sollen hier zur Veranschaulichung nur einige dieser Spezifikationen*

genannt werden. Die Notation sähe in XML z.B. bei einer Verbalform  $\acute{s}\check{c}m.n=f$  etwa so aus:

```
<wordform paradigm=„Suff.-Konj.“diathese=„Akt.“
    tempus=„ $\acute{s}\check{c}m.n=f$ “numerus=„Sg.“person=„3.“
    genus=„M.“>WRITTEN.FORM</wordform>
```

bei einem aktivischen geminierenden Partizip  $prr.(i)$  etwa so:

```
<wordform paradigm=„Partizip“diathese=„Akt.“
    merkmalfhaft=„Gem.“numerus=„Sg.“
    genus=„M.“>WRITTEN.FORM</wordform>
```

bei einer Präposition etwa so:

```
<wordform status=„St.pr.“>WRITTEN.FORM</wordform>
```

(NB: Ich habe den Status der Präpositionen nur in einem Teil der Fälle aus bestimmter Veranlassung heraus explizit codiert, der Status läßt sich jedoch grundsätzlich per Programm eruieren.)

Eine letzte Angabe im CTTC, die man der Wortform zuordnen könnte, ist die Wurzel- oder Stammvarianten-Spezifikation. Damit kann man Verweise von Wurzeln bzw. Stämmen, die nicht als die Standardform angenommen werden, auf die im CTTC geltende Standardform erzeugen, z.B. von  $kw$  „du, dich“ auf  $\acute{c}w$ , von  $m$  „in“ auf  $\acute{i}m$ , von  $\acute{s}tm$  „hören“ auf  $\acute{s}\check{c}m$ . Ich nutze diese Notate in meiner auf der Basis des CTTC erstellten lexikalischen Konkordanz zu den Sargtexten. Man kann die Verweise jedoch auch aus den „written forms“ gewinnen, so daß, wenn man solche explizit notiert, die Verweise der genannten Arten nicht explizit notiert zu werden brauchen. Für den Benutzer bequemer ist jedoch fraglos, die Stammvarianten explizit anzugeben, da er sich dann über die Regeln der Extraktion aus den „written forms“ keine Gedanken zu machen braucht.

### 3.4 Das Kompositum und seine Bestandteile

Alle Bestandteile eines Kompositums sind im CTTC, wie oben (§3.2.2) bereits gesagt, noch einmal gesondert notiert: die in einem Kompositum enthaltenen Komposita und – in letzter Instanz – alle Simplicia, seien sie Bestandteile eines selbständigen Kompositums oder eines enthaltenen Kompositums. Auf diese Weise sind nicht nur die Komposita als ganze nachweisbar, sondern umstandslos, nach allen ihren Merkmalen, auch ihre Bestandteile.

Die Bestandteile des Kompositums folgen im allgemeinen unmittelbar auf das Kompositum in der Reihenfolge, in der sie im Kompositum aufeinander folgen. Um Bestandteile als solche kenntlich zu machen und sie zum Kompositum in Beziehung zu setzen, sind den Bestandteilen lexikalische Verweise zugefügt, die auf das Kompositum verweisen. Sind Komposita ineinander geschachtelt, wird stufenweise verwiesen. Die lexikalischen Verweise verweisen auf das Lemma des Kompositums, nicht auf das real zugehörige Kompositum. Letzteres ergibt sich implizit als das nächste vorausgehende Kompositum mit dem betreffenden Lemma.

Der Verweis vom Bestandteil auf das Kompositum hat zwei mögliche Formen: Der Verweis vom ersten Bestandteil eines Kompositums auf das Kompositum braucht nur auf das Kompositum selbst (und evtl. auf den semantischen Wortindex) zu verweisen, weil alle übrigen Lemma-Angaben des Bestandteils mit denen des Kompositums identisch sind:

```
<kompos.verweis>
  <kompositum kompos.index=„KOMPOS.INDEX“>KOMPOSITUM
    </kompositum>
  <semantik>SEMANT.WORT.INDEX</semantik>
</kompos.verweis>
```

Dagegen muß bei den nicht-ersten Bestandteilen die komplette Lemma-Angabe im Verweis stehen:

```
<kompos.verweis>
  <wurzel wurzel.index=„WURZELINDEX“
    wurzel.typ=„WURZELTYP“>WURZEL</wurzel>
  <wort wort.index=„WORTINDEX“>WORT</wort>
  <kompositum kompos.index=„KOMPOS.INDEX“>KOMPOSITUM
    </kompositum>
  <semantik>SEMANT.WORT.INDEX</semantik>
</kompos.verweis>
```

Konkretes Beispiel: *īm.(i)w-ht-ḥr.(w)* „die Gefolgsleute des Horus“ (einschließlich Kopf-Lemmata):

```
<textwort>
  <lemma>
    <wurzel>īm</wurzel>
    <wort>īm.ī</wort>
    <kompositum>īm.(i)w-ht-ḥr.(w)</kompositum>
```

```

        <semantik>b</semantik>
    </lemma>
    <wortform>...</wortform>
</textwort>

<textwort>
    <lemma>
        <wurzel>im</wurzel>
        <wort>im.i</wort>
        <kompositum>im.(i)-ht</kompositum>
    </lemma>
    <wortform>...</wortform>
    <kompos.verweis>
        <kompositum>im.(i)-ht</kompositum>
        <semantik>b</semantik>
    </kompos.verweis>
</textwort>

<textwort>
    <lemma>
        <wurzel>ht</wurzel>
        <wort>ht</wort>
    </lemma>
    <wortform>...</wortform>
    <kompos.verweis>
        <wurzel>im</wurzel>
        <wort>im.i</wort>
        <kompositum>im.i-ht</kompositum>
        <semantik>b</semantik>
    </kompos.verweis>
</textwort>

<textwort>
    <lemma>
        <wurzel>hri</wurzel>
        <wort>hr.w</wort>
    </lemma>
    <wortform>...</wortform>
    <kompos.verweis>
        <wurzel>im</wurzel>
        <wort>im.i</wort>
        <kompositum>im.i-ht-hr.w</kompositum>
        <semantik>b</semantik>

```

```

    </kompos.verweis>
</textwort>

```

Die Bestandteile des Kompositums sind ferner, sofern sie, wie angenommen, aufeinander folgen, durch Bindezeichen miteinander verknüpft, die man ebenfalls in die XML-Notation übernehmen sollte. Diese Bindezeichen sind im CTTC beim vorangehenden Bestandteil notiert, gehören aber, genau genommen, weder zu diesem noch zum folgenden Bestandteil. Sie sollten also als Element zwischen den Textwort-Elementen notiert werden. Beispiel *im.í-ht* „dahinter befindlicher“:

```

<textwort>
  <lemma>
    <wurzel>im</wurzel>
    <wort>im.í</wort>
  </lemma>
  <wortform>...</wortform>
  <kompos.verweis>...</kompos.verweis>
</textwort>
<bindezeichen>-</bindezeichen>
<textwort>
  <lemma>
    <wurzel>ht</wurzel>
    <wort>ht</wort>
  </lemma>
  <wortform>...</wortform>
  <kompos.verweis>...</kompos.verweis>
</textwort>

```

Anmerkung: Bindezeichen werden im CTTC, wie üblich, auch vor Suffixpronomina gesetzt.

Durch Bindezeichen verbundene Textwörter könnten in XML-Notation explizit als Element Textwortgruppe gekennzeichnet werden. Im Beispiel *im.í-ht* „dahinter befindlicher“ sähe dies, vereinfacht, so aus:

```

<textwortgruppe>
  <textwort>...</textwort>
  <bindezeichen>-</bindezeichen>
  <textwort>...</textwort>
</textwortgruppe>

```

Würde man eine solche Notation wählen, könnte man die Angaben zum Kompositum selbst in die Textwortgruppe aufnehmen:

```
<textwortgruppe>
  <kompositum>...</kompositum>
  <textwort>...</textwort>
  <bindezeichen>-</bindezeichen>
  <textwort>...</textwort>
</textwortgruppe>
```

Bei geschachtelten Komposita würden Textwortgruppen in Textwortgruppen stehen.

Einen Problemfall stellen die Distanzkomposita dar. Es sei erinnert an das Beispiel *hr·w* NN. *pn m<sup>3c</sup>·w* „dieser NN. ist gerechtfertigt“ zu *m<sup>3c</sup>·hr·w* „gerechtfertigt sein“. Im Falle einer Inversion der Bestandteile eines Kompositums wie in *hr·w* NN. *pn m<sup>3c</sup>·w* „dieser NN. ist gerechtfertigt“ steht die Information zum Kompositum selbstverständlich beim Kopfwort des Kompositums, d.h. bei *m<sup>3c</sup>*; bei *hr·w* steht ein Verweis. Problematisch ist jedoch, daß es weder bei *m<sup>3c</sup>* bzw. *m<sup>3c</sup>·hr·w* einen Hinweis darauf gibt, wo in der Textfolge das zugehörige *hr·w* steht, und ebensowenig bei *hr·w* einen Hinweis darauf, wo das zugehörige *m<sup>3c</sup>* bzw. *m<sup>3c</sup>·hr·w* im Text zu finden ist. (Der undefinierte Zustand resultiert daraus, daß die Distanzkomposita, die es in den Grammatiken nicht gibt, erst spät in das Blickfeld gerieten und bislang ihre Notation nur so weit geklärt wurde, wie es für die Zwecke einer lexikalischen Konkordanz erforderlich ist.) Diesem Manko kann bei der Übersetzung in XML-Notation abgeholfen werden, mutmaßlich sogar per Programm. Denn es handelt sich nur um ganz wenige Wörter (ich erinnere mich nicht genau, aber vielleicht handelt es sich überhaupt nur um *m<sup>3c</sup>·hr·w*), und vermutlich immer ist der gesuchte Bestandteil das nächste vorausgehende Vorkommen des fraglichen Bestandteils (jedenfalls gilt dies für *hr·w*). Ich benutze jedenfalls bei der formalen Überprüfung der Komposita solche Vermutungen mit Erfolg.

Im Hinblick auf die Distanzkomposita bietet sich unter XML eine andere Notation der Komposita an, die der zunächst ins Auge gefaßten und im CTTC realisierten überlegen ist: die Relationierung des Kompositums und seiner Bestandteile mit Hilfe eines Identitäts-Attributs („id“) und von Referenz-Attributen. Das Identitäts-Attribut kann z.B. die Position des Kompositums innerhalb des Textkorpus sein. Die Referenzattribute verweisen auf das Identitäts-Attribut des folgenden („s“, sequens) bzw. des vorausgehenden („a“, antecedens) Wortes. Mit Hilfe solcher Attribute können selbstver-

ständig „normale“ Komposita, aber genauso gut auch unterbrochene und Distanzkomposita dargestellt werden, also Komposita vom Typ  $tp.i-b'$  „auf Erden Befindlicher, Hinterbliebener“, wozu  $tp.i=f-b'$  „sein auf Erden Befindlicher, sein Hinterbliebener“, und  $m^c-hr.w$  „gerechtfertigt sein“, wozu  $hr.w$  NN.  $pn m^c.w$  „dieser NN. ist gerechtfertigt“. Beispiele:

1.  $tp.i=f-b'$  „sein auf Erden Befindlicher, sein Hinterbliebener“:

```

<textwort id=„pos1“s=„pos3“>
  <lemma>
    <wurzel>tp</wurzel>
    <wort>tp.i</wort>
    <kompositum>tp.(i)-b'</kompositum>
  </lemma>
  ...
</textwort>
<textwort>
  <lemma>
    <wurzel>f</wurzel>
    <wort>f</wort>
  </lemma>
  ...
</textwort>
<textwort id=„pos3“a=„pos1“>
  <lemma>
    <wurzel>b'</wurzel>
    <wort>b'</wort>
  </lemma>
  ...
</textwort>

```

2.  $hr.w$  NN.  $pn m^c.w$  „dieser NN. ist gerechtfertigt“:

```

<textwort id=„pos1“a=„pos4“>
  <lemma>
    <wurzel>hrw</wurzel>
    <wort>hr.w</wort>
  </lemma>
  ...
</textwort>
<textwort>
  <lemma>
    <wurzel>NN.</wurzel>

```

```

        <wort>NN.</wort>
    </lemma>
    ...
</textwort>

<textwort>
    <lemma>
        <wurzel>pn</wurzel>
        <wort>pn</wort>
    </lemma>
    ...
</textwort>
<textwort id=„pos4“s=„pos1“>
    <lemma>
        <wurzel>m3c</wurzel>
        <wort>m3c.w</wort>
        <kompositum>m3c-hr.w</kompositum>
    </lemma>
    ...
</textwort>

```

Eine etwas einfachere Lösung wäre die, statt rückwärts und vorwärts zu verweisen, nur ein einziges Referenz-Attribut („target“) zu benutzen und mit diesem nur von den Bestandteilen auf das Kompositum zu verweisen. Die beiden Beispiele läsen sich – etwas verkürzt – dann etwa so (das in geschweifte Klammern eingeschlossene Identitätsattribut könnte wegbleiben):

1. *tp.i=f-b* „sein auf Erden Befindlicher, sein Hinterliebener“:

```

<textwort id=„pos1“>
    ... <kompositum>tp.(i)-b</kompositum>
</textwort>
<textwort>
    ... <wort>f</wort>
</textwort>
<textwort {id=„pos3“} target=„pos1“>
    ... <wort>b</wort>
</textwort>

```

2. *hr-w NN. pn m<sup>3</sup>c.w* „dieser NN. ist gerechtfertigt“:

```

<textwort {id=„pos1“} target=„pos4“>
    ... <wort>hr-w</wort>

```

```

</textwort>
<textwort>
  <lemma>
    <wort>NN.</wort>
  </lemma>
  ...
</textwort>
<textwort>
  <lemma>
    <wurzel>pn</wurzel>
    <wort>pn</wort>
  </lemma>
  ...
</textwort>
<textwort {id=„pos4“} target=„pos1“>
  ... <wort>mc.w</wort>
    <kompositum>mc-hr.w</kompositum>
</textwort>

```

Entscheidet man sich für eine so geartete Notation in XML, kann man auf die zuvor benutzten Verweise („kompos.verweis“) von Bestandteilen eines Kompositums auf den Kopf des Kompositums verzichten. Andererseits kann man dem lexikalisch interessierten Anwender die Arbeit dadurch erleichtern, daß man ihm die fertig konstruierten Verweise liefert.

### 3.4 Syntaktische Struktur (Satzzeichen)

Die Satzzeichen des CTTC könnten in XML als Textzeichen (#PCDATA) zwischen die Wortformbeschreibungen eingefügt werden, fallweise, um jeder denkbaren Verwechslung mit Strukturzeichen der Transkription vorzubeugen, als „entity references“; z.B. Satzzeichen Punkt:

```
&punkt ;
```

Genauer betrachtet sind im CTTC nicht Satzzeichen notiert, sondern syntaktische und semantische Einheiten umgrenzt. Ohne dies an dieser Stelle im Detail behandeln zu wollen oder zu können: Es wird eine Syntax in der Art derjenigen angenommen, die FRIEDRICH JUNGE in seiner *Syntax der mittelägypti-*

*schen Literatursprache* beschreibt, also Matrixsatz und eingebettete Sätze.<sup>11</sup> Z.B. sieht eine Emphatische Konstruktion mit Umstandssatz als Vedette im Prinzip so aus:

((EMPHATISCHE FORM ...)(VEDETTE))

In der Notation entfallen alle äußeren Klammern, notiert ist also tatsächlich:

EMPHATISCHE FORM ...)(VEDETTE

Sätze sind durch ein Satzzeichen Punkt voneinander getrennt, so daß also die vollständige Notation im Prinzip so aussieht:

EMPHATISCHE FORM ...)(VEDETTE.

Die solcherart beschriebene syntaktische Struktur könnte man in eine XML-Notation etwa folgendermaßen übersetzen:

```
<matrixsatz>
  <eingebetteter.satz>EMPHATISCHE.FORM ...
  </eingebetteter.satz>
  <eingebetteter.satz>VEDETTE</eingebetteter.satz>
</matrixsatz>
```

Solche Strukturen kann man selbstverständlich zum guten Teil auch als Strukturen anderer Syntaxmodelle interpretieren, wenn man andere Syntaxmodelle bevorzugt. Es gibt allerdings bei der Uminterpretation fallweise auch Probleme. Zum Beispiel kann man die Satzerweiterungen meines aktuellen Syntaxmodells<sup>12</sup> nicht ohne weiteres identifizieren. (Bevor ich mein älteres, Darmstädter, Modell verwarf, um mich dem Jungeschen Modell anzuschließen, war ich meinem heutigen Modell mit dem traditionellen Ansatz antizipierter Satzteile schon einmal näher.<sup>13</sup>)

<sup>11</sup> F. JUNGE, *Syntax der mittelägyptischen Literatursprache*, Mainz 1978, bes. 29f.

<sup>12</sup> W. SCHENKEL, *Standardtheorie und invertierte Standardtheorie*, ZÄS 125, 1998, 140–160, bes. 145–147.

<sup>13</sup> W. SCHENKEL, *Maschinelle Analyse altägyptischer Texte*, Darmstadt 1969, 162 (Prolepse abgeschlossen durch „/“).

### 3.6 Positionsangaben (Stellenangaben)

Positionsangaben bestimmen die Position einer Wortform im Textkorpus, im Teil-Textkorpus der Sargtexte, aber auch im Gesamt-Textkorpus der ägyptischen Texte. Zweckmäßigerweise bestehen solche Positionsangaben aus einer hierarchisch geordneten Menge von Spezifikationen oder aus mehreren hierarchisch geordneten Mengen von Spezifikationen. Solche Spezifikationen sind z.B. die Bezeichnung des Textes, die Textzeile, die Bezeichnung der Publikation, die Seite der Publikation, die Zeile auf der Seite der Publikation u.ä. Für jede solche Angabe muß beschrieben sein

- der Inhalt (z.B. Angabe der Seite, Zeile, ...), damit man den Inhalt bei Bedarf explizit machen kann (z.B. eine Seite mit Hilfe des Zusatzes „S.“, eine Zeile mit Hilfe des Zusatzes „Z.“),
- die Form (z.B. arabische Ziffern, römische Ziffern, lateinische Buchstaben), damit man Positionsangaben einer bestimmten Art korrekt sortieren kann.

NB: Es handelt sich also nicht um eine bloße Formatierungsangabe, die erst in einem style sheet, konkret: in XSL, der XML Stylesheet Language, zu behandeln wäre.

Im CTTC bzw. für das CTTC stehen die folgenden Angaben zur Verfügung und könnten in XML-Darstellung in folgender Weise übersetzt werden:

#### a) publikationsbezogene Angaben:

- Textedition („editio“) De Bucks:  
<ed>CT</ed>  
(NB: CT ist dann also nicht die Bezeichnung des Textkorpus der Sargtexte, sondern die Edition desjenigen Teils des Textkorpus, das DE BUCK erfaßt hat.)
- Band („volumen“) der Edition in römischen Zahlen (Attribut „numerus“ mit der Spezifikation „romanus“):  
<vol num=„rom“>V</vol>
- Seite („pagina“) des Bandes:  
<pag>100</pag>
- Abschnitt („sectio“) innerhalb der Seite:  
<sect>b</sect>

- Weitere Spezifikationen („annotatio“): Der Text steht in der Edition in einem Appendix („a“, appendix) (hinter der Paralleledition des Spruches oder am Bandende), in einer Anmerkung (nota, „n“), steht nur auszugsweise in einer Fußnote oder ist in der Edition von DE BUCK nicht enthalten („s“, supplementum); z.B.:

`<annot>a</annot>`

b) Textkorpusbezogene Angaben:

Nummer des Spruches („caput“, Kapitel) im Korpus der Sargtexte:

`<cap>335</cap>`

- c) Ein Problem stellen die Textzeugen („testis“) dar. Einerseits bezeichnen sie archäologische Objekte, nämlich bestimmte Särge an bestimmten Aufbewahrungsorten (z.B. B3Bo, ein Sarg aus al-Biršā, der in Boston aufbewahrt wird), dann aber auch Bezeugungen von Einzeltexten im Textkorpus der Sargtexte (z.B. werden mit den Siglen B3Bo<sup>a</sup> und B3Bo<sup>b</sup> zwei Teilbezeugungen des Spruches 575 auf ein und demselben Sarg voneinander unterschieden), schließlich die Textzeugen der Parallel-Edition von DE BUCK (z.B. B3Bo, B3Bo<sup>a</sup>, B3Bo<sup>b</sup>, S2C, S2Chass. und B4C als Versionen des genannten Spruches 575). Die XML-Notation könnte in den beiden letztgenannten Fällen, die für den Philologen im Vordergrund stehen, dieselbe sein; z.B.:

`<test>B3Boa</test>.`

Man kann dann alternativ editionsbezogen sortieren nach Band, Seite, Abschnitt und Textzeugen (die Stellung der weiteren Spezifikationen wäre zu klären) oder textkorpusbezogen nach Spruchnummer und Textzeugen.

Ein weiteres Problem, das sich bei den Textzeugen stellt, ist die Frage der Sortierfolge. Im CTTC ist den Textsiglen ein Code zugeordnet, der eine Sortierung nach den Herkunftsorten von Süden nach Norden, innerhalb eines Ortes nach Aufbewahrungsorten (danach weiter alphabetisch bzw. numerisch) erlaubt. Dieser Code könnte in die XML-Notation als Attribut („loc“) übernommen werden; z.B für den Textzeugen B3Bo:

`<test loc=„4515“>B3Bo</test>.`

Die allgemeine Form der Positionsangaben könnte demnach sein:

```

<ed>EDITION</ed>
<vol num=„NUMERIERUNGSSYSTEM“>BAND</vol>
<pag>SEITE</pag>
<sect>ABSCHNITT</sect>
<annot>WEITERE.SPEZIFIKATION</annot>
<cap>SPRUCH.NUMMER</cap>
<test loc=„GEOGRAPH.SORTIERSCHLÜSSEL“>
    TEXTZEUGE</test>

```

### 3.7 Graphie

Der komplizierteste Bereich der CTTC-Notation ist die analytische Transkription der hieroglyphischen Graphien. Es können deshalb im folgenden unmöglich alle Gesichtspunkte erörtert werden, die hierbei eine Rolle spielen. Ich beschränke mich notgedrungen auf die mehr grundsätzlichen Aspekte der Notation, ohne die überhaupt kein Einstieg in eine Diskussion möglich ist. Grundlegend sind die folgenden vier Punkte:

1. Grundsätzlich werden die Zeichen des hieroglyphischen Textes in der Lesefolge notiert, nicht also immer in der Reihenfolge, in der die Hieroglyphenzeichen aufeinander folgen. Das heißt: Die Reihenfolge entspricht der in der „*written form*“ angenommenen. Sind Hieroglyphenzeichen im originalen Hieroglyphentext invertiert, so wird dies explizit durch einen Inversionsindex angezeigt.

Anmerkung: Will man auf dieser Basis den Hieroglyphensatz steuern, muß man also vor dem Drucken die notierten Zeichen invertieren.

2. Grundsätzlich wird jedes Hieroglyphenzeichen (bzw. jede Hieroglyphenzeichengruppe, die eine bestimmte Funktion hat) ihrer Funktion nach gekennzeichnet. Die Hauptfunktionsklassen sind Semogramm, Phonogramm, Phonetisches Determinativ und Füllzeichen. Homonyme Zeichen werden innerhalb dieser Funktionsklassen durch Indizes unterschieden. Das heißt, ein Semogramm mit dem unterscheidenden Index „2“ ist im Prinzip nicht dasselbe Hieroglyphenzeichen wie das homonyme Phonogramm mit dem unterscheidenden Index „2“ (obwohl es sich im Einzelfall durchaus um dasselbe Hieroglyphenzeichen handeln kann, oft sogar sich so verhält, wenn es innerhalb der beiden Funktionsklassen nur wenige Homonyme gibt, und immer so, wenn es innerhalb der beiden Funktionsklassen keine Homonyme gibt). Nach ihrer Verwendung zerfallen Semogramme und Phonogramme jeweils weiter in zwei Teilklassen: Notationen und Kennzeichnungen. Das heißt Semo-

gramme haben entweder die Funktion von Logogrammen oder die Funktion von Determinativen, Phonogramme haben entweder die Funktion von Phonogrammen im engeren Sinn oder die Funktion von phonetischen Komplementen. Die Differenzierung der Semogramme bzw. Phonogramme nach ihrer Funktion hat keinen Einfluß auf die Indizes zur Unterscheidung der Homonyme. Das heißt, ein Semogramm bzw. ein Phonogramm hat in der Funktion einer Notation keinen anderen Index als in der Funktion als Kennzeichnung.

3. Die hieroglyphischen Schriftzeichen werden nach den Prinzipien notiert, die ich in *GOF IV 12* dargelegt habe.<sup>14</sup>
4. Die Notation der Graphien in meinem Textkorpus setzt wie die Hieroglyphenschrift selbst immer noch auf einen intelligenten menschlichen Leser, der sozusagen zwischen den Zeilen zu lesen vermag. Das heißt meine computer-interne Notation ist zwar selbst schwer lesbar, aber sie ist druckorientiert. Sie unterstützt eine möglichst komprimierte Wiedergabe der graphematischen Transkription im Druck. Ebenso ist meine Notation für Sortierzwecke geeignet. Man kann in der Tat auf der Basis dieser Notation Graphien in eine Reihenfolge bringen, die dem intelligenten Leser einsichtig ist. Diese meine Notation umgeht dadurch, daß sie die intellektuellen Fähigkeiten des Lesers nutzt, trickreich gewisse Probleme, die sich bei der Umsetzung der hieroglyphischen Zeichenfolgen in Transkription ergeben, zumal bei der Übersetzung in XML-Notation. Und zwar liegt das Hauptproblem darin, daß Elemente der Hieroglyphenschrift bzw. ihrer Transkription teilweise, wie oben (§3.1) exemplifiziert, *nicht* hierarchisch geordnet sind.

Hier einige Beispiele dafür, wie man Grapheme in XML-Notation übersetzen könnte:

Phonogramme werden als Element Phonogramm („p“) notiert, Komplemente als Element Komplement („c“, complementum). Homonyme Phonogramme bzw. Komplemente werden durch das Attribut Index („i“) voneinander unterschieden. Komplementierte Phonogramme werden als Element Phonogrammgruppe („pg“) zusammengefaßt. Es stehen in der Phonogrammgruppe die Komplemente vor bzw. hinter dem Phonogramm in der Reihenfolge der Konsonanten des Phonogramms. Nicht-Komplementierung wird als leere Komplementierung notiert. Das heißt auch, nicht-komplementierte Phonogramme werden in eine Phonogrammgruppe einge-

<sup>14</sup> W. SCHENKEL, *Aus der Arbeit an einer Konkordanz zu den altägyptischen Sargtexten*, *GOF IV 12*, Wiesbaden 1983, 1–170.

geschlossen. (Auf diese Weise läßt sich die Komplementlosigkeit auf bequeme Weise feststellen.) Beispiele:

Zweikonsonantenzeichen *mn* (Spielbrett), unkomplementiert:

```
<p i="1">mn</p>
```

Zweikonsonantenzeichen *mn* (Spielbrett) mit hinterem Komplement *n* (Wasserwelle):

```
<pg>
```

```
  <p i="1">mn</p>
```

```
  <c i="1">n</c>
```

```
</pg>
```

Zweikonsonantenzeichen *mr/2* (Meißel) mit vorderem Komplement *m* (Eule) und hinterem Komplement *r* (Mund), in der Reihenfolge *m*, *mr/2*, *r*.

```
<pg>
```

```
  <c i="1">m</c>
```

```
  <p i="2">mr</p>
```

```
  <c i="1">r</c>
```

```
</pg>
```

Analog werden Logogramme als Element Logogramm („l“) notiert bzw. in ein Element Logogrammgruppe („lg“) eingeschlossen. Bei diesen (und bei den noch zu besprechenden Determinativen) enthält mein Textkorpus außer der Notation des Logogramms mit seinem Lautwert zusätzlich die GARDINER-Zeichennummer bzw. Zeichennummern, die den Vorrat der GARDINER-Zeichenliste erweitern („n“). Diese Nummern könnten in der XML-Notation beibehalten werden. (Man könnte per Programm die Zeichennummern selbstverständlich auch bei den Phonogrammen erzeugen; bei diesen sind sie aber weniger informativ als bei den Semogrammen.) Beispiel: Logogramm *ič* „nehmen“ (Phonogramm *č* plus Beinchen, GARDINER-Zeichennummer V 15), hinten komplementiert mit Phonogramm *t*, genauer anzusetzen somit als *ičt*:

```
<lg>
```

```
  <l n="V15"i="1">ičt</l>
```

```
  <c i="1">t</c>
```

```
</lg>
```

Determinative („d“) werden analog den Logogrammen notiert, es erübrigt sich jedoch der Ansatz einer Determinativgruppe, da Determinative nicht komplementiert werden, sondern selbst eine Art Komplement zur phonographischen / logographischen Notation sind. Beispiel:

Determinativ sitzender Mann mit Zeremonialbart (GARDINER-Zeichennummer A40), in meiner Transkription als „V“ („venerabilis“) notiert:

<d n=„A40“i=„1“>V</d>

Entsprechend als Wortgruppendedeterminativ („dg“):

<dg n=„A40“i=„1“>V</dg>

Logogramme können teilweise in Komponenten zerlegt werden: eine Kern-Komponente, die (auch) für den Lautwert des Logogramms steht, und eine oder mehrere periphere Komponenten, die als semantische Ergänzungen aufgefaßt werden können. Zum Beispiel *ič*+*nht* „nehmen“ (Hilfs-Zeichennummer V§ 3), bestehend aus der Kern-Komponente *ič* (GARDINER-Zeichennummer V 15) mit am vorderen Fuß integrierten schlagenden Arm. Auch in diesen Fällen sollte die Kern-Komponente notiert werden wie angegeben. Die Ergänzungen würden dann innerhalb der Logogrammgruppe („lg“) in unmittelbarer Nachbarschaft neben der Kern-Komponente (nach bzw. gelegentlich auch vor dieser) als ergänzende Logogramm-Komponente („lc“) notiert. Beispiel (Lautwert der Kern-Komponente wieder durch das Komplement in *ič* präzisiert):

<lg>

<l n=„V§3“i=„1“>ič</l>

<lc>+nht</lc>

<c i=„1“>ič</c>

</lg>

Anmerkung 1: Das zuvor zitierte Logogramm *ič* (Gardiner-Zeichennummer V 15), ein Monogramm aus dem Phonogramm *č* und dem Semogramm Beinchen, wird nicht als zusammengesetzt notiert, da der Lautwert für die klassische Hieroglyphenschrift als *ič* angesetzt wird und dieser nur dem Logogramm als ganzem zugehört, nicht einem seiner Bestandteile.

Anmerkung 2: Das „+“ in der Notation der Komponente bezeichnet die Relation der Komponente zur benachbarten Komponente, hier also die Relation von *nht* zu *ič* (es gibt

auch andere Relationen). Es stellt sich die Frage, ob es zweckmäßig ist, den Relator innerhalb der Komponente zu notieren.

Dieser Lösung haftet ein Mangel an: die Position der Zeichennummer (V§ 3) und des Indexes, die sich auf beide, die Kern-Komponente plus die ergänzende Komponente, beziehen, bei der Kern-Komponente. Diesem Mangel könnte man durch eine der folgenden Notationen begegnen:

Alternativlösung 1: Die ergänzende Komponente („lc“) wird innerhalb der Kern-Komponente („l“) notiert:

```
<lg>
  <l n=„V§3“i=„1“>
    it
    <lc>+nht</lc>
  </l>
  <c i=„1“>t</c>
</lg>
```

Alternativlösung 2 (wohl die geeignetste): Kern-Komponente („lp“ = zugleich phonographische Komponente) und ergänzende Komponente („ls“ = nur semantische Komponente) werden beide als gesonderte Elemente notiert:

```
<lg>
  <l n=„V§3“i=„1“>
    <lp>it</lp>
    <ls>+nht</ls>
  </l>
  <c i=„1“>t</c>
</lg>
```

In letzterem Fall wären auch alle ergänzungslosen Logogramme mit „lp“ zu kennzeichnen, was als ziemlich aufwendig erscheint in Anbetracht dessen, daß die Mehrzahl der Logogramme aus nur einer Komponente besteht; z.B. im Falle des oben bereits behandelten *ič* ohne *nht*:

```
<lg>
  <l n=„V15“i=„1“>
    <lp>ič</lp>
  </l>
  <c i=„1“>t</c>
</lg>
```

Nun zu einem der Problemfälle einer Übersetzung in XML-Notation, und zwar gleich zum schwierigsten: zu dem oben (§3.1) bereits einmal angesprochenen Problem der Notation der Phonetischen Determinative. Phonetische Determinative – man darf sich an der Bezeichnung orientieren – „determinieren“ vorangehende Hieroglyphenzeichen oder Hieroglyphenzeichengruppen mit phonetischer, moderner gesagt: phonematischer Information. Man könnte sich also zunächst vorstellen, man könne „determinierte“ und „phonetisch determinierende“ Elemente („pd“) zu einer „phonetisch determinierten“ Gruppe („pdg“) zusammenfassen. Zum Beispiel Phonogramm *i* plus Phonogramm *b* plus Phonetisches Determinativ *ib* „Böckchen“:

```
<pdg>
  <pg>
    <p i=„1“>i</p>
  </pg>
  <pg>
    <p i=„1“>b</p>
  </pg>
  <pd n=„E8“ i=„1“>ib</pd>
</pdg>
```

Es gibt indes ein gravierendes Problem, das eine solche Notation praktisch verunmöglicht: Das Phonetische Determinativ muß nicht unmittelbar auf die „determinierte“ Information folgen, sondern kann von dieser durch andere Elemente getrennt sein. Zum Beispiel *ib.t* „Durst“, geschrieben mit Phonogramm *i* plus Phonogramm *b* (Stamm) plus Phonogramm *t* (Endung) plus Phonetisches Determinativ *ib* „Böckchen“. Dieser Sachverhalt ist zwar grundsätzlich mit den Mitteln von XML darstellbar. (Man könnte die phonetisch determinierte Gruppe als „spanning structure“ notieren). Aber man kann sich keineswegs darauf verlassen, daß ein kommerzieller Browser eine solche nicht-hierarchische Struktur brauchbar visualisiert. Im übrigen bleibt ungeklärt, was man mit solchen Strukturen tut, wenn man sie mit Hilfe eines Spezialprogramms weiterverarbeitet.

Nun könnte man auf die phonetisch determinierte Gruppe einfach verzichten wollen und determinierte Information und Phonetisches Determinativ je für sich an der Stelle notieren, an der sie in der hieroglyphischen Zeichenfolge stehen. Das Beispiel *ib.t* „Durst“ stellte sich dann so dar:

```
<pg>
  <p i=„1“>i</p>
```

```

</pg>
<pg>
    <p i="1">b<p>
</pg>
.      (Strukturzeichen)
<pg>
    <p i="1">t<p>
</pg>
<pd n="E8"i="1">ib</pd>

```

Diese Notation wäre konform mit der Notation der (echten) Determinative, die man auch an der Stelle notiert, an der sie in der hieroglyphischen Zeichenfolge stehen, und bei denen man auch keine Auskunft darüber gibt, worauf sich die Determinierung bezieht, ob etwa auf den Stamm, auf das Wort oder auf die Flexionsform.

Anmerkung: Der Übergang zwischen Phonetischem Determinativ und Determinativ ist fließend. Phonetische Determinative, die sich auf eine Wortform beziehen, werden üblicherweise nicht unter den Phonetischen Determinativen aufgeführt: Dual- und Pluralstriche, die anzeigen, daß eine Wortform, die nicht Dual bzw. Plural ist noch semantisch Dualität bzw. Pluralität beinhaltet, lautlich einer Dual- bzw. Pluralform gleich oder ähnlich klingt. Solche Phonetischen Determinative habe ich in meinem Textkorpus, um unausweichliche Abgrenzungsprobleme zu vermeiden, grundsätzlich als Determinative notiert.

Im CTTC sehen die beiden Beispiele so aus (für den Druck vorgesehene Darstellung):

- *ib<sub>g</sub>* und
- *ib.t<sub>gc</sub>*

„g“ bezeichnet die Position des Phonetischen Determinativs, der Zusatz „c“, daß sich „g“ nicht auf die unmittelbar vorausgehende Konsonantenfolge bezieht, sondern auf die Konsonantenfolge eines vorausgehenden „centralen“ Elements, z.B. des Stamms. So weit ließe sich die als zweite Alternative eingeführte XML-Notation realisieren. Der Lautwert des Phonetischen Determinativs, der ja nicht explizit dasteht, läßt sich mit Hilfe einer Liste der möglichen Phonetischen Determinative als *ib* bestimmen, auch in dem Fall, in dem das Phonetische Determinativ nicht unmittelbar auf die zugehörige Konsonantenfolge folgt. Ein Problem ergibt sich jedoch bei der Textkritik. Textkritische Zeichen beziehen sich grundsätzlich auf die Gruppe aus pho-

netisch determinierter Information und Phonetischem Determinativ. Zum Beispiel besagen in

*ʾib*<sub>g</sub>

die Halbklammern, daß die phonetisch determinierte Gruppe als ganze teilzerstört ist, was daraus resultieren kann, daß entweder die Phonogramme *ʾ* und *b* beide teilzerstört sind oder das Phonetische Determinativ teilzerstört ist oder, wenn einer dieser beiden Fälle vorliegt, zusätzlich das bzw. eines der anderen Zeichen teilzerstört oder auch vollkommen zerstört ist. Es gibt keinerlei Möglichkeit, die textkritische Information sachgerecht auf die Phonogramme und auf das Phonetische Determinativ zu verteilen. Infolgedessen muß auch in der XML-Notation eine Gruppennotation realisiert werden. Evtl. könnte man so verfahren, daß man

- eine phonetisch determinierte Gruppe („pdg“) notiert, die auch die Textkritik enthält, das Phonetische Determinativ selbst jedoch nicht in diese einbezieht und
- in seiner tatsächlichen Position das Phonetische Determinativ notiert, unter der ausdrücklichen Vereinbarung, daß die zugehörige Textkritik durch die vorangehende phonetisch determinierte Gruppe („pdg“) abgedeckt ist.

Die Beispiele *ʾib.t<sub>gc</sub>* und *ʾib.t<sub>gc</sub>* sähen mit textkritischer Kennzeichnung der Gruppe *ʾib* mit Halbklammern so aus:

```
<pdg>
  r
  <pg>
    <p i=„1“>ʾ</p>
  </pg>
  <pg>
    <p i=„1“>b<p>
  </pg>
  ,
</pdg>
<pd n=„E8“i=„1“>ʾib</pd>
```

bzw.:

```

<pdg>
  r
  <pg>
    <p i="1">i</p>
  </pg>
  <pg>
    <p i="1">b</p>
  </pg>
</pdg>
.      (Strukturzeichen)
<pg>
  <p i="1">t<p>
</pg>
<pd n="E8"i="1">ib</pd>

```

Einfacher wieder ist die Übersetzung der letzten Art von Zeichen, der Füllzeichen.

Einzelne Füllzeichen werden als Textzeichen (#PCDATA) mit Hilfe einer entity reference notiert, etwa:

```
&f;
```

Füllzeichen, die über einen Bereich streuen durch Anfangs- und Ende-Kennzeichnung:

```
<fanf> bzw. </fend>.
```

Schließlich könnten Lücken im Text (nicht ergänzbare Zerstörungen, vom Schreiber freigelassene Stellen) durch Angaben der folgenden Art notiert werden:

```
<lücke länge="NUMERISCHE.ANGABE">
```

bzw.

```
<lücke länge="UNDEFINIERT">
```

Nachdem die Hauptfunktionsklassen besprochen sind, steht jetzt die Klärung der Probleme an, die sich aus diversen Kombinationen der bereits definierten Elemente ergeben. Vorausgenommen wurde bereits das Problem des wech-

selweisen Bezugs von Phonetischem Determinativ und dadurch determinierter Information. Als nächste Problemgruppe sollen sich überlappende Elemente besprochen werden.

Hier sind sich überlappende Phonogrammgruppen zu nennen. Beispiel: *wnw*, geschrieben mit Zweikonsonantengruppe *wn* und Zweikonsonantengruppe *nw*, konkret etwa:

```
<pg>
  <p i="1">wn</p>
  <c i="1">n</c>
</pg>
<pg>
  <p i="1">nw</p>
  <c i="1">w</c>
</pg>
```

Dieses Problem kann evtl. auf der Ebene der Textkritik aufgefangen werden, nämlich dadurch, daß man *n* in der zweiten Gruppe als redundant kennzeichnet (behelfsmäßig mit doppelten geschweiften Klammern notiert; selbstverständlich handelt es sich nicht um die geschweiften Klammern des Leidener Systems):

```
<pg>
  <p i="1">wn</p>
  <c i="1">n</c>
</pg>
<pg>
  <p i="1">{{n}}w</p>
  <c i="1">w</c>
</pg>
```

Eine andere Möglichkeit bestünde darin, das redundante *n* in der Art, wie Komplemente notiert werden, explizit als redundantes Element („r“) innerhalb der Phonogrammgruppe zu notieren:

```
<pg>
  <p i="1">wn</p>
  <c i="1">n</c>
</pg>
<pg>
  <r>n</r>
```

```

<p i="1">nw</p>
<c i="1">w</c>
</pg>

```

Ein ähnliches Problem stellt sich bei den Logogrammen, zu deren Definition eine Endung gehört, wenn diese Endung nach dem Logogramm noch einmal explizit phonographisch geschrieben ist. Beispiel: *ś.t* „Thronsitz“, geschrieben mit Logogramm Thronsitz *ś.t* plus Phonogramm *t* (NB: Das *t* ist – hier nicht näher zu begründen – kein Komplement, sondern genau die Endung). In meinem Textkorpus gilt die Endung des Logogramms als explizit phonographisch geschrieben. Ist die Endung nicht explizit geschrieben, wird sie in textkritische runde Klammern gesetzt: *ś.(t)*. In der XML-Notation würde man dagegen besser eine Darstellung in der Art derer wählen, die eben für die sich überlappenden Phonogrammgruppen erörtert wurden, allerdings mit Kennzeichnung beim ersten der beiden Elemente, dem Logogramm, und nicht beim zweiten Element, dem Phonogramm. Also, Endung nicht geschrieben:

```

<lg>
  <l n="Q1"i="1">ś{{t}}</l>
</lg>

```

oder:

```

<lg>
  <l n="Q1"i="1">ś.t</l>
  <r>.t</r>
</lg>

```

Mit expliziter Schreibung der Endung:

```

<lg>
  <l n="Q1"i="1">ś{{t}}</l>
</lg>
. (Strukturzeichen)
<pg>
  <p i="1">t</p>
</pg>

```

Oder:

```
<lg>
  <l n="Q1" i="1">ś.t</l>
  <r>.t</r>
</lg>
. (Strukturzeichen)
<pg>
  <p i="1">t</p>
</pg>
```

Noch ungelöst ist das Problem der Notation von Monogrammen und Inversionen. Zwar könnte man zunächst an folgende sehr einfache Lösungen denken:

Bestandteile von Monogrammen werden in ein Monogramm-Element („m“) eingeschlossen. Zum Beispiel Phonogrammgruppe mit Phonogramm *ś* und Phonogrammgruppe mit Phonogramm *f* als Monogramm mit der Lesefolge *śf*:

```
<m>
  <pg>
    <p i="1">ś</p>
  <pg>
  <pg>
    <p i="1">f</p>
  <pg>
</m>
```

Invertierte Elemente werden in ein Inversions-Element („i“) eingeschlossen. Zum Beispiel hinteres Komplement *t*, gefolgt von Zweikonsonantenzeichen *īt*, eine übliche Schreibung von *īt(i)* „Vater“, mit der Lesefolge Zweikonsonantenzeichen *īt* plus hinteres Komplement *t*:

```
<pg>
  <i>
    <p i="1">īt</p>
    <c i="1">t</c>
  </i>
</pg>
```

Mit einer solchen Notation können aber nicht ohne weiteres Monogramme und Inversionen aufgefangen werden, an denen Teile von Elementen beteiligt sind, weil dann die bereits problematisierten „spanned structures“ entstünden. Zum Beispiel (CT VII 232j) *swʿ* „vorbegehen“, geschrieben mit

Einkonsonantenzeichen  $s$ , Zweikonsonantenzeichen  $w$ , Komplement  $\acute{s}$  und Determinativ „Weg“ unter Inversion des Determinativs und des Komplements. In diesem Fall überschneidet sich die Komplementgruppe mit der Inversionsgruppe. Die Komplementgruppe umfaßt das Zweikonsonantenzeichen  $w$  und das Komplement  $\acute{s}$ , die Inversionsgruppe umfaßt das Komplement  $\acute{s}$  und das Determinativ „Weg“.

Morphematische Strukturzeichen werden als Transkriptionszeichen ohne jede Kennzeichnung zusammen mit konsonantischen Transkriptionszeichen verwendet oder zwischen Elemente der Graphemnotation eingefügt, technisch gesprochen: sie werden als #PCDATA behandelt. Die Setzung zwischen die Elemente der Graphemnotation hat zur Folge, daß die übergeordneten Elemente vom Typ „mixed“ sind, d.h. sowohl Text als auch Elemente enthalten können. Beispiele für den Gebrauch zusammen mit Konsonantenzeichen wurden bereits zuvor gegeben. Ein Beispiel für die Verwendung zwischen Elementen wäre die mit Einkonsonantenzeichen geschriebene Partikel *gr.t*:

```
<...>
  <pg>
    <p i="1">g</p>
  </pg>
  <pg>
    <p i="1">r</p>
  </pg>
  (Strukturzeichen)
  <pg>
    <p i="1">t</p>
  </pg>
</...>
```

Ebenso werden im Prinzip textkritische Zeichen behandelt. Beispiel:

```
<p i="1">[m]</p>
```

Anmerkung 1: Die als Beispiel gewählten eckigen Klammern stehen nicht für „eckige Klammer“, sondern für „zerstört“, was üblicherweise mit eckigen Klammern wiedergegeben wird (Leidener System). In der XML-Notation wird über die Wiedergabe im Druck keine Aussage gemacht, die eckigen Klammern sind nur aus mnemotechnischen Gründen gewählt worden. Theoretisch könnte man die eckigen Klammern im Druck auch mit Schraffur über dem betreffenden Zeichen darstellen oder mit eckigen Klammern und Schraffur gleichzeitig.

Anmerkung 2: Der Zeichenvorrat der #PCDATA wird durch entity references erweitert, z.B.:

&hklauf;

für „öffnende Halbklammer“.

Die textkritische Kennzeichnung erfolgt wortformweise, d.h. ohne Berücksichtigung der textkritischen Kennzeichnung der vorausgehenden und der folgenden Wortform. Folgen z.B. zwei zerstörte Wortformen aufeinander, so wird jede für sich als zerstört gekennzeichnet:

[wortform1] [wortform2].

Auf diese Weise wird der textkritische Befund überschaubar gehalten, der infolge des Umfangs der Notationen leicht aus dem Auge geraten kann.

Geklärt werden muß noch, ob textkritische Zeichen so weit wie möglich innen oder so weit wie möglich außen stehen oder ob nach noch zu findenden Regeln teils innen, teils außen. Soll man z.B.

<p i="1">[m]</p>

oder

[<p i="1">m</p>]

notieren?

#### 4. Epilog: Datenaustauschformat, Architekturen

Die hier diskutierten XML-Notationen sind weniger idiosynkratisch als die CTTC-Notationen, aber immer noch idiosynkratisch. Sie unterliegen zwar einer standardisierten Syntax, verwenden aber im wesentlichen nicht-standardisierte Schlüsselwörter und implizieren ein nicht-standardisiertes Modell zur Beschreibung ägyptischer Textdaten. Immerhin ist eine solche XML-Notation als Datenaustauschformat bereits wesentlich geeigneter als das CTTC-Format.

Was die nicht-standardisierten Schlüsselwörter angeht, ist Abhilfe relativ leicht zu schaffen. Zum einen kann man die Schlüsselwörter leicht per Programm durch andere Schlüsselwörter ersetzen. Schwererwiegend ist das Problem des zugrunde gelegten Beschreibungsmodells. Wie sich aus den vorangehenden Diskussionen der Detaillösungen ergibt, kann selbst im en-

gen Rahmen der Gegebenheiten des CTTC fraglich sein, wie man Textelemente beschreibt. Es ist jedoch mit Modellen der Beschreibung ägyptischer Texte zu rechnen, die die ägyptischen Texte nach teilweise anderen Kriterien erfassen als im CTTC geschehen und die Texte, unter Verwendung selbst der XML-Syntax, anders strukturieren. Sofern man sich jedoch der XML-Syntax bedient, können unterschiedliche Modelle, zum mindesten weitgehend, in Beziehung zueinander gesetzt werden, nämlich mit Hilfe von Meta-DTD's, Meta-„Document Type Definitions“, über die die DTD's unterschiedlicher Beschreibungsmodelle in Bezug zueinander gesetzt werden können. Zwar wäre es wünschenswert, wenn man sich für die Beschreibung ägyptischer Texte auf wenige Standard-Modelle verständigen könnte. Die Forderung nach einer Einigung auf ein einziges Modell wäre jedoch weder sinnvoll noch realistisch. Die Erfassung ägyptischer Texte kann heute und in aller Zukunft durchaus unterschiedliche Zielsetzungen haben, man kann Texte also in unterschiedlicher Intensität und in unterschiedlichen Merkmalen erfassen wollen. Zum anderen erscheint es unausweichlich, daß man nach den Erfahrungen mit Beschreibungsmodellen künftig neue und bessere Beschreibungsmodelle entwickeln wird. Nicht zuletzt aber wird sich der einzelne die Freiheit nehmen, für seine eigenen Zwecke Modelle zu verwenden, die seinen eigenen Vorstellungen von ägyptischen Textstrukturen entsprechen, die von den Vorstellungen jedes anderen Ägyptologen verschieden sein können. Es liegt also daran, die Beschreibungsmodelle zueinander in Bezug zu setzen. Als ein Schritt in diese Richtung ist die Übersetzung der CTTC-Daten in eine XML-Notation gedacht, die Gegenstand der hier ausgebreiteten Überlegungen war.

### Postscriptum

Die in § 3.2.2. erörterte und zum Programm erhobene Änderung der Indizierung der Komposita in den CTTC-Daten wird, zum mindesten einstweilen, nicht realisiert. Es hat sich nämlich beim genaueren Studium der Einzelfälle ergeben, daß das Problem auch auf andere Weise gelöst werden kann und in Anbetracht der verhältnismäßig geringen Anzahl der Fälle auch anders gelöst werden darf, zum mindesten für die Zwecke der dem Abschluß entgegengehenden lexikalischen Konkordanz zu den Sargtexten. Grundsätzlich wäre in der Konkordanz das Problem praktikabel mit Querverweisen zwischen den homonymen Komposita lösbar oder aber mit Übersichten über die unter einem Lemma behandelten semantischen Unter-Lemmata und Komposita unmittelbar unterhalb der Lemma-Überschrift. Letztere Lösung ist für die gedruckte Konkordanz jetzt vorgesehen.