

The editors of the series *APPLICATIO* are

prof.dr. H. Leene, prof.dr. E. Talstra
Faculty of Theology of the Vrije Universiteit
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands

The following titles in the series *APPLICATIO* are still available:

2. Amos. Concordance and Lexical Surveys
A. van der Wal, E. Talstra (1984)
3. Amos. A Classified Bibliography
A. van der Wal (1986)
4. Trito-Isaiah. An exhaustive Concordance of Isa. 56-66
J. Bastiaans, W. Beuken, F. Postma (1984)
6. Nahum Habakkuk. A Classified Bibliography
A. van der Wal (1988)
7. Computer Assisted Analysis of Biblical Texts
E. Talstra (ed) (1989)
9. A Prophet on the Screen. Computerized Description and Literary Interpretation
of Isaianic Texts
E. Talstra, A.L.H.M. van Wieringen O.S.A. (1994)
11. Grammatica Digitalis I. The Morphological Code
in the "Werkgroep Informatica" Computer Texts of the Hebrew Bible
A.J.C. Verheij (1994)
12. Participles in Context. A Computer-Assisted Study
of Old Testament Computer Text of the Hebrew Bible
J.W. Dyk (1994)
13. A Computational Display of Joshua.
A Computer-assisted Analysis and Textual Interpretation
N. Winther-Nielsen, E. Talstra (1995)
14. In Search of Text Syntax.
Towards a Syntactic Text Segmentation Model for Biblical Hebrew
A.F. den Exter Blokland (1995)
15. Ad Fontes! Quellen erfassen - lesen - deuten
Was ist Computerphilologie?
C. Hardmeier, J. Range, W.-D. Syring, E. Talstra (2000)

APPLICATIO 15

AD FONTES!

QUELLEN ERFASSEN - LESEN - DEUTEN

WAS IST COMPUTERPHILOLOGIE?

Ansatzpunkte und Methodologie - Instrumente und Praxis

Contributions to the Conference "Computerphilologie"
November 5-8, 1998
Ernst-Moritz-Arndt-Universität Greifswald

Christof Hardmeier
Wolf-Dieter Syring
Jochen D. Range
Eep Talstra

A'dam. VU University Press 2000

pp 33-60

Genese und Kategorienentwicklung der WIVU-Datenbank

oder: ein Versuch, dem Computer Hebräisch beizubringen

Eep Talstra und Constantijn Sikkel (Amsterdam)

Einleitung

Wie kann die Aufgabe bewältigt werden, eine Text-Datenbank zu erstellen, die für Linguisten und Literaturwissenschaftler, für Exegeten und Übersetzer als Hilfestellung bei ihrer Arbeit mit Sprachen und Texten zur Verfügung stehen soll? Die hebräische Datenbank der WERKGROEP INFORMATICA an der Theologischen Fakultät der VRIJE UNIVERSITEIT in Amsterdam (WIVU) stellt einen der möglichen Ansätze dar, diese Aufgabe zu lösen.

In Kapitel 1 werden zunächst Fragen zur Methodologie der computergestützten Textanalyse formuliert, die insbesondere bei Texten in antiken Sprachen im Vordergrund stehen: Erstellt man Daten mit Hilfe einer Grammatik oder eine Grammatik mit Hilfe von Daten? In Kapitel 2 werden die Probleme beschrieben, die die Aufgabe der Textanalyse erschweren: das Problem der Unterscheidung zwischen *langue* und *parole* sowie das der Differenzierung zwischen den linguistischen Ebenen und der hierarchischen Gliederung einer konkreten Textkomposition. Als Beispiele dienen zwei sehr unterschiedliche Projekte der computergestützten Analyse hebräischer Bibeltex-te. Kapitel 3 stellt die Vorgehensweise bei der textlinguistischen Analyse zum Aufbau der WIVU-Datenbank dar. In Kapitel 4 werden schließlich die Datenkategorien erklärt, die entwickelt worden sind, um eine Vielzahl unterschiedlicher textueller und grammatischer Daten zu speichern und auszuwerten.

1. Eine Datenbank syntaktisch analysierter biblisch-hebräischer Texte - ein Resultat oder ein Instrument?

Die bisher erhältlichen Computerprogramme zur Untersuchung biblischer Texte stützen sich auf Datenbanken, die linguistische Ana-

lyseergebnisse auf Wortebene zur Verfügung stellen.¹ Exegeten, Übersetzer und Grammatiker sind dankbare Benutzer dieser neuen Werkzeuge, sind sich jedoch auch dessen bewußt, daß man für eine intensive exegetische Arbeit oder für eine sachgerechte Übersetzung der Texte wesentlich mehr sprachliche Informationen benötigt. Eine Analyse biblischer Texte kann sich nicht beschränken auf semantische und rhetorische Erscheinungen wie z.B. die Kombination oder die Wiederholung einzelner Wörter. Zum Verstehen einer Textkomposition benötigt man Kenntnisse weiterer linguistischer Texteigenschaften: Satztypen und Satzverbindungen, Verbvalenz, Einbettung direkter Rede in narrative Texte, Wiedererkennung von Aktanten in einem Textgefüge. Wegen der Notwendigkeit, über die Wortebene hinausreichende Daten zu berücksichtigen, hat man bei der Erstellung einer Textdatenbank eine zweifache Aufgabe.

- Zum einen ist die Frage zu beantworten, inwieweit eine sprachliche Beschreibung von Textebenen oberhalb der Wortebene möglich ist. Kann man linguistische Kategorien entwickeln, um Satzgefüge und andere Erscheinungen der Textebene in systematischer Weise grammatisch zu analysieren?
- Zum anderen geht es um die Frage der Vorgehensweise. Ist es notwendig, sich zunächst um eine breit angelegte linguistische Analyse und um die philologische Klärung von Einzelfragen zu bemühen, um anschließend die Ergebnisse dieser Arbeit in einer Datenbank zu speichern? Oder sollte man zuerst versuchen, eine möglichst vollständige Satz- und Textgrammatik zu definieren, die von einem Computer als Regelwerk zur Erstellung einer einwandfreien grammatischen Analyse verwendet werden könnte? Der zweite Weg geht von der Idealvorstellung einer a priori definierbaren Grammatik aus. Die Erfahrungen im Bereich der Computerlinguistik und die Debatte zur sogenannten Künstlichen Intelligenz² haben jedoch gezeigt, daß dieses Ziel derzeit noch nicht erreicht ist und möglicherweise auch nicht erreicht werden kann. Daher muß die Frage „Wie erstellt man eine Textdatenbank?“ im Sinne des erstgenannten Weges präziser formuliert werden, sowohl aus prinzipiellen Gründen wie auch aus praktischen Erwägungen: In welcher Weise können textlinguistische Daten kategorisiert und gespeichert werden, damit als Ergebnis eine Datenbank für weitere linguistische und literarische Untersuchungen zur Verfügung steht? Wenn also eine konsistente Textgrammatik als Grundlage der Textanalyse nicht verwendet werden kann,

1 Vgl. W.-D. SYRING (QUEST 2, 1998).

2 Vgl. H. DREYFUS (Critique, 1992).

dann liegt es nahe, umgekehrt vorzugehen: „Wie gewinnt man aus den Textdaten eine Grammatik?“

Die Aufgabe besteht somit nicht darin, eine Datenbank zu erstellen, die an sich bereits ein definitives Resultat textgrammatischer Analysen ist, sondern darin, eine Datenbank aufzubauen, die sich als textgrammatisches Forschungsinstrument eignet. An einem Beispiel (Kapitel 2.1) soll gezeigt werden, warum diese methodologische Entscheidung getroffen wurde, und welche praktischen Folgen für die WIVU-Datenbank sich daraus ergeben.

Wer eine Datenbank von Texten aus antiken, jedenfalls nicht mehr täglich benutzten Sprachen aufbauen will, muß eine Reihe von Vorüberlegungen anstellen. Was ist ein Text? Welche sprachlichen Mittel werden beim Textaufbau eingesetzt, insbesondere bei Texten in antiken Sprachen? Was wird mit diesen Mitteln sprachlich kodiert, so daß es von Lesern wiedererkannt und dekodiert werden kann? Kann man das Gefüge des Textaufbaus in konsistent definierten Datenkategorien beschreiben? Sollen die Mittel der linguistischen Kodierung, z.B. Morpheme wie die Verbalendung „ִּיָּ“, gespeichert werden oder die linguistischen Funktionen, die mit diesen Mitteln realisiert werden, wie z.B. die grammatische Funktion „Imperfekt“?

In diesem Zusammenhang stößt man auf drei eng zusammenhängende Problemkreise, die sich in drei Fragen formulieren lassen:

- Wie kann die Verwendung mehrerer sprachlicher Ebenen (und ggf. verschiedener Sprachen) in den Texten bei der Analyse berücksichtigt werden?
- Werden linguistische Formen und/oder linguistische Funktionen als sprachliche Daten angesehen?
- Verläuft der Analyseprozeß von den größeren Einheiten (z.B. Sätzen) zu den kleineren (z.B. Wörtern), also „top-down“, oder umgekehrt, von den kleineren zu den grösseren Einheiten, d.h. „bottom-up“?

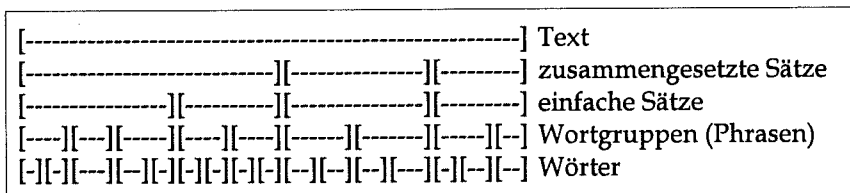
Nach unseren Erfahrungen ist der Weg, den man zur Lösung des ersten Problemkreises einschlägt, entscheidend für das analytische Verfahren, das zur Klärung der beiden anderen Problemkreise gewählt werden kann. Die Kernfrage ist daher, wie die Verwendung sprachlicher Ebenen in Texten einer Analyse zugänglich gemacht werden soll. Die Antwort auf diese Frage entscheidet auch über die Datentypen und das Analyseverfahren.

2. Textsyntax: die Analyse der sprachlichen Ebenen in Texten

In den neueren Grammatiken des biblischen Hebräisch wird häufig von den unterschiedlichen linguistischen Ebenen der grammatischen Analyse gesprochen. Einerseits versucht man mit den Einsichten der neueren Textlinguistik³ auch der Beschreibung des biblischen Hebräisch gerecht zu werden,⁴ andererseits wird behauptet, die Kenntnisse auf diesem Gebiet seien noch nicht so weit fortgeschritten, daß man sie bei der Beschreibung von antiken Sprachen schon verwenden könne.⁵ Unseres Erachtens sollte die „Computerphilologie“ jedoch nicht die weitere Entwicklung der textlinguistischen Forschung abwarten, sondern versuchen, die vorhandenen Kenntnisse im Bereich der Textgrammatik zu überprüfen und zu verbessern, da gerade Computerprogramme die benötigten Möglichkeiten zum Testen und Experimentieren anbieten.

2.1 Sprachsystem und Textkomposition

Theoretisch stellt man sich die linguistischen Ebenen in einem Text häufig als einfache Hierarchie vor: der Text als höchste Einheit gliedert sich in kleinere Einheiten, die wiederum in noch kleinere Einheiten segmentiert werden. Die auf diese Weise definierten linguistischen Ebenen können in einem abstrakten Modell veranschaulicht werden:



Figur 1: Idealtypische textuelle Hierarchie

Das praktische Problem besteht nun darin, daß ein „echter“ Text diesem Idealbild kaum entspricht. Sehr häufig läßt sich in einem konkreten Text eine Mischung von Ebenen und Kategorien feststellen:

3 Vgl. E. COSERIU (Textlinguistik, 1972). Einen Überblick über Möglichkeiten und Aufgaben bieten W. BODINE (Syntax, 1992) und C. VAN DER MERWE (Discourse Linguistics, 1994).

4 Vgl. W. RICHTER (Satz, 1980) und F. ANDERSEN (Sentence, 1974).

Sätze können eine Funktion auf der Wort- oder Phrasenebene haben; einzelne Wörter können als Sätze fungieren; ein Text (als Segment der direkten Rede) kann in narrativem Kontext als Objekt zu einem Verb angesehen werden, usw. Aus dieser Tatsache folgt unmittelbar, daß die Ebenen der linguistischen Textbeschreibung nicht mit den Ebenen der Textkomposition identisch sind. Die theoretischen linguistischen Ebenen können in einer strengen Hierarchie beschrieben werden, die faktischen textlichen Ebenen jedoch nicht, da sie selten in einer hierarchischen Ordnung auftreten; sie können vielmehr mehrfach ineinander eingebettet werden.

Als Testfall für das Problem der linguistischen Kategorisierung kann folgendes Beispiel dienen: wieviel Sätze enthält die Textpassage Numeri 13,32f.?

Segmentierung der Textoberfläche		Analyse-Vorschlag	
		einf. Satz	zus. Satz
ויציאו דבת הארץ (32)	Satzteil		
אשר תרו אתה	Satz=atr.		
אלבני ישראל	Satzteil	—	
לאמר	Satz=Konst.		—
הארץ	Wortgruppe	zus. Satz	zus. Satz
אשר עברנו בה	Satz=atr.		
לתור אתה	Satz=Konst.		
ארץ	Wort		
אכלת יושביה	Satz=atr.		
הוא	Wort	—	
וכלהעם	Satzteil	einf. Satz	
אשרראינו בתוכה	Satz=atr.		
אנשי מדות:	Wortgruppe	—	
ושם ראינו אדהנפלים בני ענק מןהנפלים (33)	Satz		
ונהי בעינינו כחבבים	Satz		
וכן היינו בעיניהם:	Satz		—

Figur 2a: Segmentierung eines realen Textes (Num 13,32f.)

Es zeigt sich, daß eine Hierarchie von Ebenen tatsächlich existiert, im syntaktischen Aufbau dieses konkreten Textes jedoch die Hierarchie der Elemente nicht mit der der Textebenen übereinstimmt. Die Ebenen als Elemente des grammatischen Systems lassen sich in einem Text bestimmen, ergeben aber keine eindeutige Hierarchie - etwa nach dem Bild einer Pyramide.⁶ Die sprachlichen Ebenen des Systems sind nicht identisch mit den textlichen Ebenen der Komposi-

tion. Bei der Arbeit an einer Textdatenbank besteht daher auch nach der Definition der linguistischen Beschreibungskategorien noch nicht die Möglichkeit, die Syntax konkreter Texte zu speichern. Dieses Problem bietet eine Erklärung dafür, daß die Mehrzahl der heutigen Computerinstrumente zur Textanalyse über die Wortebene nicht hinausgehen. Auf dieser Ebene ist der Unterschied zwischen *langue* und *parole* kaum wahrzunehmen. Wenn man jedoch Texte auf höheren Ebenen analysiert, nehmen die Unterschiede zwischen Sprachsystem und Texthierarchie erheblich zu.

(32) Und sie brachten ein böses Gerücht auf über das Land,
das sie erkundet hatten,
unter den Israeliten
und sprachen:
Das Land,
durch das wir gegangen sind,
um es zu erkunden,
[Land]
frißt seine Bewohner
[es],
und alles Volk,
das wir darin sahen,
sind Leute von großer Länge.
(33) Wir sahen dort auch Riesen, Anaks Söhne aus dem Geschlecht der Riesen,
und wir waren in unsern Augen wie Heuschrecken
und waren es auch in ihren Augen.

Figur 2b: Num 13,32f. nach der Übersetzung Martin Luthers (Stuttgart 1984)⁷

Daher besteht die wichtigste Frage zunächst nicht darin, wie man einen konkreten Text linguistisch analysieren und möglicherweise erklären kann. Wenn das Ziel der Datenbank-Entwicklung ein Instrument zur linguistischen und literarischen Textanalyse ist, dann sollten gerade nicht Interpretationen, sondern vielmehr Daten gespeichert werden. Vordringlich ist daher die Klärung der Frage, wie die komplexen Verhältnisse in empirischen Texten kategorial beschrieben und in einer Datenbank gespeichert werden können: Welche linguistische Segmentierung ist notwendig, welche Eigenschaften zeigen die Segmente, und welche Beziehungen haben sie zueinander?

7 Die Wörter „ein böses Gerücht auf“ wurde – in Parallele zum hebräischen – nicht unübersetzte Wörter wurden in

2.2 Experimente zur Datenanalyse: deszendent – aszendent; distributional – funktional

2.2.1 Aszendentes Verfahren

Ein Vorschlag zum Prozeß der Textanalyse sieht vor, die Datenbank „von unten“ aufzubauen: zuerst sind kleine Einheiten zu isolieren und anschließend daraus größere Einheiten zu komponieren. Diese Lösung bietet WOLFGANG RICHTER in seinen grammatischen Studien und – teilweise – in der Textbeschreibung der BIBLIA HEBRAICA TRANSCRIPTA [BH].⁸ In dieser Edition des hebräischen Alten Testaments werden bestimmte Aspekte der Textstruktur analysiert und gespeichert: die Transkription (als phonologische Analyse bzw. Wiedergabe) des Textes und die Satzabgrenzungen, nicht aber die Segmentierung in Wortgruppen oder die Markierung von Subjekt, Objekt, usw. Die Edition zeigt, daß bereits die Satzsegmentierung nicht immer einfach ist. Der oben als Beispiel angeführte Text macht dies deutlich: Wieviel Sätze enthält Numeri 13,32f.?

וּיְצִיאוּ דַבַּת הָאָרֶץ	(32) a
אֲשֶׁר תָּרוּ אֹתָהּ	aR
אֶל־בְּנֵי יִשְׂרָאֵל	a
לְאָמֹר	aI
הָאָרֶץ	bP
אֲשֶׁר עִבְרַתָּ בָּהּ	bPR
לְתוֹר אֹתָהּ	bPRI
אֶרֶץ אֲכָלֹת יוֹשְׁבֵיהָ הִיא	b
וּכְלֵהֶם	c
אֲשֶׁר־רָאִינוּ בְּתוֹכָהּ	cR
אֲנָשֵׁי מְדוּחִים:	c
(33) וְשֵׁם רָאִינוּ אֹתָהֶם בְּנֵי עֵינַי מִן־הַנְּפִלִים	a
וְזֶה־יִנְיָנוּ כְּחַגְבִּים	b
וְכֵן הָיִינוּ בְּעֵינֵיהֶם:	c
a ... c :	Sätze
R:	Relativ-Satz
I:	Infinitiv-Satz
P:	Pendens-Konstruktion

Figur 3: Datenanalyse nach Wolfgang Richter (Num 13,32f. Bh')

In RICHTERS Textanalyse werden Sätze isoliert, einige Satztypen (Nebensätze: R, I, P) gekennzeichnet und aufgrund dieser Charakte-

risierung zueinander in Beziehung gesetzt. Das geschieht aber nur teilweise: die Verbindung der Hauptsätze - wie z.B. die mit „und“ (hebr. ו) verknüpften Sätze Numeri 13,32c.33a.33b.33c - wird bei der syntaktischen Analyse nicht berücksichtigt.

Dem von RICHTER verwendeten Verfahren zur Satzsegmentierung ist zuzustimmen: zunächst ist der Text linear zu segmentieren, auch wenn dabei vorläufig unvollständige Sätze auftreten (Numeri 13,32a.32c); anschließend sind die gewonnen (Teil-) Sätze nach funktionalen Kriterien zu kombinieren.⁹ Die hinter der Edition der BH¹ stehende Datenbank stellt so ein vorläufiges Instrument dar, das Material für weitere Untersuchungen zur Verfügung stellt.¹⁰

Bei der Satzabgrenzung werden jedoch Entscheidungen bereits dort getroffen, wo diese - aus linguistischen Gründen - noch offen bleiben sollten. Ein Beispiel findet sich im letzten Satz von Numeri 13,32b und zeigt die Schwierigkeiten bei der Analyse von Sätzen, deren Prädikat ein Partizip ist (hier: אכלת יושביה הוא): Sind zwei Sätze anzunehmen („ein Land, das seine Einwohner frisst, ist es“) oder liegt - nach der Satzeinteilung RICHTERS - nur ein Satz („ein seine-Einwohner-fressendes Land ist es“) vor? Wenn das Partizip אכלת wie in diesem Fall ein eigenes Objekt (יושביה) hat,¹¹ dann hat man nach RICHTERS Vorschlag zwei Prädikate in einer Zeile (d.h. in einem Satz), was auch nach seinen eigenen Kriterien nicht stimmig ist. Daher muß diese Zeile in zwei Sätze aufgeteilt werden, was die Datenbank der BH¹ ihrer Struktur nach auch zuläßt.

2.2.2 Deszendentes Verfahren

Ein völlig andersgeartetes Vorgehen hat JOHN SAILHAMER vorgeschlagen,¹² dessen Ausgangspunkt bei den Fragen der textuellen Organisation liegt, d.h. am anderen Ende des Spektrums textanalytischer Methodik. Er führt daher eine „von oben nach unten“ fortschreitende Analyse durch, beginnend mit der Identifizierung von narrativen und von diskursiven Textteilen, Satztypen und Aktanten.

9 Diese Vorgehensweise empfiehlt sich auch für die Kombination von Wörtern zu Wortgruppen.

10 W. ECKARDT (Analyse, 1987).

11 Vgl. unten Kapitel 4.2.

12 Vgl. J. SAILHAMER (Approach, 1990); J. SAILHAMER (2 Samuel 13, 1992); E. Gü-

	Ref. Form	CT	DT	ML	ATP
ויוציאו דבת הארץ אשר תרו אתה אל-בני ישראל לאמר	a1 WAYYIQTOL	V	Nar	1	@+=
הארץ אשר עברת בה לתור אחיה	b1 NOM ABS		Dis	2	
ארץ אכלת יושביה הוא	b2 0 + NC	N	Dis	2	
וכלהעם אשר-דיאני בתוכה אנשי מדות:	b3 W + NC	N	Dis	2	
ושם ראיתי אתהנפילים בני ענק מןהנפלים	a1 W+X+QATAL	V	Dis	2	
ונהי בעינינו כחנבים	b1 WAYYIQTOL	V	Dis	2	@+=
וכן היינו בעיניהם:	b2 W+X+QATAL	V	Dis	2	
a1, b1 ... b3: Halbvers und Satznummer					
CT: Clause Type (V = Verbalsatz; N = Nominalsatz)					
DT: Discourse Type (Nar = narrativ; Dis = diskursiv)					
ML: Metacomunicative Level					
ATP: Actant Time Place					

Figur 4: Datenanalyse nach John Sailhamer (Num 13,32f.)

SAILHAMERS Analyse kombiniert auf diese Weise linguistische und literarische Daten: Sätze und Kommunikationsebenen, „Spieler im Text“, lokale und temporale Verweisungen.¹³ Diese Kombination textlinguistischer und literarischer Interessen hat weitreichende Auswirkungen für die Datenanalyse.

Erstens bedeutet es, daß als Daten nur die Eigenschaften der Hauptsätze in Betracht kommen, da nur sie für die Textstruktur konstitutiv sind. Infinitiv-Sätze und Partizipial-Sätze werden in diesem System nicht als separate Sätze behandelt, weil sie an der Stelle von Satzkonstituenten stehen und daher nach SAILHAMERS Auffassung als unselbständige Größen nichts zur Textstruktur beitragen.¹⁴ Daraus folgt, daß nur Sätze als sprachliche Zeichen mit spezifischen linguistischen Merkmalen betrachtet werden. Es muß gefragt werden, ob diese Annahme sachgemäß ist. Wird eine Textstruktur nur dadurch bestimmt, daß Satztypen in einer bestimmten Folge auftreten, daß z.B. ein W-X-Qatal-Satz einem Wayyiqtol-Satz¹⁵ folgt? Muß man nicht auch auf bestimmte lexikalische Wiederholungen oder gram-

13 Vgl. die Beschreibung der Datenbank von A. DEN EXTER BLOKLAND (Clause-Analysis, 1990) 75: „A Combination of Linguistic and Literary Data.“

14 Vgl. A. DEN EXTER BLOKLAND (Clause-Analysis, 1990) 79: „Participial and infinitival phrases, except the infinitive absolute ..., are excluded: on the surface level they are not clauses.“

Die Wörter des Beispieltextes haben nicht nur linguistische Eigenschaften, sondern auch eine bestimmte Position im Text als einem schriftlichen Dokument: sie gehören zur zweiten Vershälfte von Numeri 13,32. Es muß daher prinzipiell zwischen zwei Betrachtungsweisen eines Textes bzw. den entsprechenden Textstrukturen unterschieden werden: der Text stellt sich einerseits als (historisches) Dokument und andererseits als (linguistische) Komposition dar.

- Der Text als *Dokument* repräsentiert die Gestalt des Textes nach der masoretischen Überlieferung bzw. gemäß der Edition der ältesten, vollständig erhaltenen Handschrift der hebräischen Bibel.¹⁸ Die Datenbank sollte daher zumindest die zur Gliederung des Dokuments verwendeten Kategorien unterscheiden: Bücher, Kapitel, Verse, Halbverse und einzelne graphisch unterscheidbare Wörter (inkl. der - Vokale und Akzente repräsentierenden - Punktierung). Diese Kategorien sind ähnlich der oben (in Figur 1) gezeigten idealtypischen Hierarchie streng hierarchisch geordnet. Jede kleinere Einheit gehört - meist mit weiteren benachbarten Einheiten - eindeutig zu einer Einheit höherer Ordnung. So ist z.B. ein Vers stets *einem* Kapitel zugeordnet und alle Kapitel eines Buches stehen in *linearer* Folge hintereinander. Daher sind die Daten auf der Dokumentenebene nur von geringer Komplexität.
- Der Text als *Komposition* wird mit den Kategorien der linguistischen Ebenen beschrieben, die bereits in den vorausgehenden Abschnitten genannt wurden: Wörter und Wortgruppen (Phrasen), einfache Sätze und zusammengesetzte Sätze. Es wurde bereits deutlich, daß diese Kategorien in einem realen Text nicht einer strengen Hierarchie folgen, sondern daß die Ebenen vermischt werden können. So zeigt Figur 2 (s. oben), daß man während des Lesens der Textoberfläche immer wieder die linguistische Ebene wechselt: Wort, Satz, Phrase, zusammengesetzter Satz, usw. Komplexer noch wird der Lesevorgang dadurch, daß Elemente unterschiedlicher Ebenen in Elemente anderer Ebenen eingebettet sein können und so deren lineare Abfolge unterbrechen.

Das sich daraus ergebende Problem der Datenstrukturen wird in der WIVU-Datenbank gelöst durch eine präzise Unterscheidung zwischen Einheiten der linguistischen Distribution (als Beschreibung ihres Aufbaus und ihrer Verteilung) und Einheiten der linguistischen Funktion. Dabei steht der Gedanke im Vordergrund, daß auf allen Ebenen die Ermittlung linguistischer Funktionen (z.B. „Attribut“,

18 Der sog. Leningradensis liegt der Biblia Hebraica Stuttgartensia (BHS, 1967/1977) und der im Erscheinen begriffenen Biblia Hebraica Quinta

„Subjekt“, „Nebensatz“) erst *nach* der Segmentierung des Textes nach distributionalen Kriterien erfolgen kann. Denn eine Abgrenzung von Wortfolgen oder Wortgruppen ist möglich, auch ohne deren Funktion (z.B. als Satzkonstituente oder als Satz) festzulegen. Im folgenden (3.1 - 3.4) wird diese Art von Textanalyse schrittweise dargestellt.

3.1 Analyse der Wortgruppenabgrenzungen

Auf der Grundlage eines bereits morphologisch analysierten Textes¹⁹ kann man mit Hilfe eines Computerprogramms nachprüfen, ob in diesem Text z.B. eine Wortfolge mit folgenden Eigenschaften existiert:

Nomen
Artikel
Nomen

Wenn eine solche Wortfolge auftritt (z.B. דַבַּת הָאָרֶץ = „böses Gerücht von dem Land“ in Numeri 13,32), ist vom Benutzer des Programms zu entscheiden, ob die drei Wörter eine grammatisch akzeptable Konstruktion darstellen. Im positiven Fall muß das Programm die gefundene Kombination nicht nur im Text als Wortgruppe markieren, sondern auch festhalten, welche zur Bildung der Wortgruppe relevanten morphologischen Eigenschaften die Bestandteile haben. Was als relevant anzusehen ist, wird nach einem - vom Benutzer entworfenen - vorläufigen morphologischen Paradigma entschieden.

So bedeuten die Feststellungen „דַבַּת hat die Nominalendung ם und kein pronominales Suffix“ und „אָרֶץ hat als Nominalendung ein Nullmorphem und kein pronominales Suffix“ folgendes: der Numerus von דַבַּת ist 'singular', der Status ist 'unbekannt'; der Numerus von אָרֶץ ist 'singular', der Status ist 'unbekannt'; beide Wörter tragen 'kein Suffix'. Auf diese Weise kann der Benutzer eine erste Regel zum Aufbau einer Nominalphrase formulieren lassen:

Nomen (Status: unbekannt; Numerus: singular; Suffix: 0)
Artikel
Nomen (Status: unbekannt; Numerus: singular; Suffix: 0)
→ NP (determiniert)

19 Die Kategorien des von FERENC POSTMA und ARIAN J. C. VERHEIJ erstellten, morphologisch analysierten Textes der „Werkgroep Informatica“ werden

Durch die Verwendung einer „Rückkopplung“ kann das Programm nun auf weitere linguistische Werte schließen: der 'unbekannte' Status des ersten Nomens kann als 'status constructus', der des zweiten Nomens als 'status absolutus' qualifiziert werden; eine Nominalphrase mit Artikel muß als 'determiniert' gelten. Streng genommen hat dieses Ergebnis nur hypothetischen Wert. Es stellt noch keine allgemeine grammatische Regel dar, sondern eine vorläufige Hypothese zu den zugrunde liegenden Daten, die sich nur anhand weiterer Textdaten bestätigen oder korrigieren läßt.

Zur Tatsache, daß kein Suffix vorhanden ist, würde ein Hebraist feststellen, daß man bei הארץ wegen des (proklitischen) Artikels ה auch kein Suffix erwarten dürfe. Das ist natürlich richtig, aber eben diese Regel soll der Computer aus den Daten „lernen“ und nicht als Teil eines bereits feststehenden grammatischen Regelwerks vorfinden. Dieses Prinzip gehört wesentlich zum Verfahren der distributionalen Segmentierung: ausgehend von den Textdaten bildet man unter Ausnutzung intuitiven Wissens Hypothesen, aus denen ein Computerprogramm analytische Vorschläge zusammenstellt; ein bereits vorliegendes, mutmaßlich vollständiges grammatisches Regelsystem findet keine Verwendung. Die Grammatik soll nicht vorausgesetzt werden, sondern im Zuge der Datenanalyse Gestalt annehmen, also Ergebnis der Arbeit sein.

In Analogie zu dem bereits behandelten Beispiel kann das Programm die Gruppe אל-בני ישראל (= „bei den Söhnen Israels“) in Numeri 13,32 als neue Präpositionalphrase in die Liste einfügen:

Präposition (Suffix: 0)
Nomen (Status: constructus; Numerus: plural; Suffix: 0)
Eigenname
→ PP (determiniert)

Ein weiteres Beispiel in Numeri 13,33 (אדהנפילים/בני ענק/מן-הנפלים) = „die Riesen, die Söhne Enaks von den Riesen“ zeigt die rekursive Vorgehensweise. Bei der distributionalen Segmentierung wird noch nicht festgelegt, daß die acht Wörter dieser Gruppe einen Satzkonstituenten bilden, in diesem Falle ein Objekt (אוד); vielmehr werden zunächst drei Gruppen unterschieden und zwar nach zwei Strukturmustern:

Präposition (Suffix: 0)
Artikel
Nomen (Status: absolutus; Numerus: plural; Suffix: 0)
→ PP (determiniert)
und:
Nomen (Status: constructus; Numerus: plural; Suffix: 0)
Eigenname
→ NP (determiniert)

Erst im Anschluß daran werden - unter Anwendung desselben Verfahrens - die Gruppen mit weiteren Mustern verglichen. Im genannten Beispiel kann man ein neues Strukturmuster definieren, das aus einer Kombination der beiden verwendeten Muster besteht. Von diesem neuen, erweiterten Muster ausgehend kann das Programm die zweite Wortgruppe als Apposition qualifizieren:

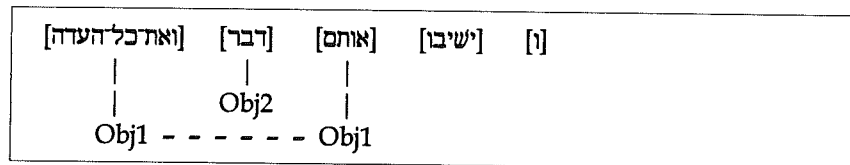
Präposition (Suffix: 0)
Artikel
Nomen (Status: absolutus; Numerus: plural; Suffix: 0)
Nomen (Status: constructus; Numerus: plural; Suffix: 0)
Eigenname
→ PP (determiniert) + NP (determiniert; Apposition)

Dieses Beispiel zeigt, daß in bestimmten Fällen mit dem distributionalen Verfahren - aufgrund der Informationen über Wortart und morphologische Eigenschaften - nicht nur eine Segmentierung des Textes durchgeführt, sondern auch spezifische Beziehungen zwischen den neu gebildeten Einheiten festgestellt werden können. In den meisten Fällen gelingt das nicht vollständig. So ist die Beziehung zur dritten Gruppe (מן הנפלים) auf dieser Ebene der Analyse noch nicht entscheidbar, weil man dazu die Information benötigt, ob diese מן-Gruppe auf Satzebene mit dem Prädikat (d.h. als präpositionales Objekt zu ראה) oder auf Phrasenebene mit der Appositionsgruppe verbunden ist (d.h. als Spezifizierung der בני ענק). Diese Entscheidung kann erst im Zuge der Analyse der Satzkonstituenten (Prädikat, Subjekt, Objekt, Komplement, usw.) getroffen werden.

Die bei der distributionalen Segmentierung gewonnenen Einheiten sind somit grammatisch durchaus korrekt, stellen aber nicht immer vollständige linguistische Größen dar. Daher werden diese Segmente in der WIVU-Datenbank als „Atome“ bezeichnet.²⁰ Mit diesem

Begriff werden zwei Sachverhalte gekennzeichnet: zum einen sind Atome auf ihrer linguistischen Ebene nicht weiter teilbar; zum anderen sind sie – einzeln oder in Kombination – Bausteine höherer funktionaler Einheiten. Die funktionalen Phrasen (Konstituenten) der Satzebene bestehen aus „phrase atoms“: so fungieren beispielsweise die drei phrase atoms in Numeri 13,33 gemeinsam als Objekt-Phrase. Analog werden Sätze als funktionale Einheiten in Satzgefügen aus „clause atoms“ gebildet: so bilden z.B. die drei clause atoms in Numeri 13,32 zwei Sätze.

Dieses Verfahren erlaubt die korrekte Beschreibung eingeschobener Elemente (embedding) und dadurch entstehender Lücken (gapping) in funktionalen Einheiten: während Atome stets lineare, ununterbrochene Wortfolgen sind, lassen die aus ihnen zusammengesetzten funktionalen Einheiten Lücken bzw. Einschübe zu, wie das Beispiel einer der geteilten Wortgruppe in Numeri 13,26 zeigt:



3.2 Analyse der Satzabgrenzungen

Das oben angeführte Beispiel aus Numeri 13,32 (vgl. Figur 2) zeigt, daß die Analyse auf Satzebene nach demselben Prinzip durchgeführt wird wie die Analyse auf Phrasenebene: die mit der distributionalen Beschreibung der Textoberfläche gewonnenen Daten werden zu Satzmustern kombiniert, die dem Analyseprogramm zur Generierung von Vorschlägen im weiteren Verlauf der Textsegmentierung dienen. Dennoch ist es auf der Satzebene für den Benutzer schwieriger, durch sprachliche Intuition eine erste Hypothese zu bilden. Auch ist man durch die konsequente maschinelle Konsistenzüberprüfung häufiger dazu gezwungen, bereits formulierte Hypothesen zu korrigieren bzw. zu präzisieren. Gemeinsam mit dem Computer zu „lernen“, ist daher der einzige gangbare Weg, da „native speakers“ als korrigierende Instanz bei Texten in antiken Sprachen nicht zur Verfügung stehen.

Die auf Satzebene entstehenden Probleme zeigen sich z.B. in der Frage, wie der attributive Partizipialsatz in *חזא אכלת יושביה* in Numeri 13,32 zu behandeln ist. Das Strukturmuster der Textoberfläche besteht aus folgenden Elementen: einem Substantiv folgen ein Partizip, eine Nominalphrase mit pronominalem Suffix und ein Personalpronomen. Handelt es sich bei dieser Wortfolge um einen Satz oder um zwei Sätze? Liegt – unter philologischen Gesichtspunkten – zwischen dem Partizip *אכלת* und der Nominalphrase *יושביה* eine regens-rectum-Verbindung oder Prädikat-Objekt-Beziehung vor, besitzt also das Partizip eigene, abhängige Konstituenten („verzehrend seine Einwohner“) oder nicht („Verzehrer seiner Einwohner“)?

Bei einem Partizip sind prinzipiell alle – im Rahmen der jeweiligen Verbvalenz zugelassenen – Verbindungen möglich, wegen der nominalen Form des Partizips kommen die Möglichkeiten der constructus-Verbindung hinzu. Partizipien eröffnen daher Konstruktionsmöglichkeiten auf zwei Ebenen. Im Fall einer verbalen Valenz-Verbindung resultieren Formen auf der Ebene der Satzsyntax; bei einer nominalen Verbindung ergeben sich Formen auf Phrasenebene.

Ein Problem entsteht insbesondere dann, wenn aufgrund der Nominalendungen nicht über den Status (absolutus oder constructus) eines Partizips entschieden werden kann. Wie kann man in diesem Fall eine Hypothese zum Satzbau bzw. zum Phrasenbau begründen? In der Anfangsphase der Textkodierung kann man nur aufgrund einer naheliegenden Interpretation eine erste Hypothese formulieren. Liegen bereits mehrere vergleichbare Fälle in den Textdaten vor, kann der Benutzer diese für seine Hypothesenbildung heranziehen.

Das aus der WIVU-Datenbank hervorgegangene Programm QUEST²¹ bietet zwar noch nicht die Möglichkeit, nach Subjekt und Prädikat zu suchen, aber wenn man nach einem einfachen Muster (wie z.B. „Nomen + Partizip mit femininer Endung + Nomen mit Suffix“) fragt, findet man genügend Stellen, die zur weiteren linguistischen Analyse hilfreich sein können:

Einerseits finden sich Belege, die eine Interpretation auf Satzebene nahelegen, bei der eine Partizipial-Konstruktion als Attributiv- oder Adverbial-Satz verwendet wird und mit Subjekt, Objekt oder auch mit anderen Konstituenten verbunden sein kann.

21 E. TALSTRA/C. HARDMEIER/A. GROVES (Ouest. 1992): vgl. E. TALSTRA (Demon-

Numeri 13,32-33		T 21	TxtT	ClsT	PNG
[<Ob> דבת הארץ [<Pr> יוציאו [<Cj> ו]		7	.. N	Way0 << Q0tl	3plM
[<Ob> אותה [<Pr> ותרו [<Re> אשר		9	.e N	XQtl [attrib]	3pl -
[<Co> אל בני ישראל לאמר		8	d. N	Defc << Way0	- - -
[<Pr> לאמר]		9	.. N	Infrc [adjunct]	- - -
===== +					
[<Fr> הארץ		10	cq NQ	CPen << Infrc	- - -
[<Co> ביה [<Pr> ועברנו [<Re> אשר		14	.. NQ	XQtl [attrib]	1pl -
[<Ob> אותה [<Pr> לתור		15	.. NQ	Infrc [adjunct]	- - -
[ארץ		13	.. NQ	NmCl [resumpt]	- - -
[<Ob> יושביה [<PC> אכלת		15	.e NQ	Ptc [attrib]	-sgF
[הוא [<Su>		14	d. NQ	Defc << NmCl	- - -
[<Su> כל העם [<Cj> ו]		12	d. NQ	Defc [resumpt]	- - -
[<Lo> בתוכה [<PO> ראינו [<Re> אשר		14	.e NQ	XQtl [attrib]	1pl -
[אנשי מדות [<PC>		13	.. NQ	NmCl << Defc	- - -
[<Lo> אנשי מדות / אתהנפלים /		11	.. NQ	XQtl [resumpt]	1pl -
[<Ob> ענק / מן הנפלים <ap>					
===== +					
[<PC> כחגבים [<Co> בעינינו [<Pr> נהי [<Cj> ו]		12	.. NQN	Way0 << Xqtl	1pl -
[<PC> בעיניהם [<Pr> דיינו [<Mo> כן [<Cj> ו]		13	.. NQN	XQtl << Way0	1pl -

T: Tabulation (= Hierarchie der Satzverbindungen)
 2: Satzmerkmal 2 (d = „defective“; c = „casus pendens“)
 1: Satzmerkmal 1 (e = „embedded“; q = „quotation“)
 TxtT: Texttyp (N = Narrativ; Q = direkte Rede)
 ClsT: Satztyp (z.B. Way0 = Wayyiqtol ohne Subjekts-NP;
 Xqatal = Element + Qatal; Defc = unvollst. Teil eines Satzes)
 PNG: Verbum: Person, Numerus, Genus
 adjunct: adjunktiver Konstituent-Satz
 attrib: attributiver Satz
 resumpt: Wiederaufnahme nach Casus Pendens
 <Co>: Komplement
 <Cj>: Konjunktion
 <Fr>: („fronted“) Casus Pendens
 <Lo>: Lokative Ergänzung
 <Mo>: Modifier
 <Ob>: Objekt
 <PC>: Prädikats-Komplement (Nominales Prädikat)
 <PO>: Prädikat mit suffigiertem Objekt
 <Pr>: Prädikat
 <Re>: Relativum
 <Su>: Subjekt
 <ap>: Apposition
 <sp>: Spezifikation (Erweiterung)

Verwendung sprachlicher Ebenen in einem konkreten Text: Einheiten unterschiedlicher Ebenen können miteinander verbunden oder ineinander eingebettet sein. Verschiedene Ebenen der Hierarchie können sich so zu einer Einheit höherer Ordnung zusammenfügen.

Aufgrund der oben ausgeführten Beispiele können nun einige vorläufige Ergebnisse und die sich daraus ergebenden Aufgaben formuliert werden.

Es ist deutlich, daß die erste Phase der Erstellung einer Textdatenbank an sich noch keine für Theologen oder Literaturwissenschaftler ertragreiche Forschung darstellt. Sie ist jedoch insofern eine grundlegende Arbeit, als in der Kombination von distributionalen und funktionalen Verfahren ein Modell geschaffen wird, das die Kategorien zur Beschreibung und Speicherung der unterschiedlichen linguistischen Daten überhaupt erst bereitstellt.

Anschließend werden auf der Basis der so entstandenen Datenbank weitere Analysen durchgeführt, um die Daten zu präzisieren bzw. zu korrigieren und auf diesem Wege die Konsistenz der analysierten Daten und der dazugehörigen Grammatik zu erhöhen. Erst auf dieser Grundlage einer konsistenten Datenanalyse kann die Diskussion mit Exegeten und Übersetzern über die Interpretation der Texte beginnen.²⁴

Die Anwendung der geschilderten Verfahren bedeutet nicht, daß die Erstellung einer Textdatenbank im direkten Gang zu einem fertiges Forschungsinstrument führt. Es handelt sich eher um Schritte einer computergestützten Textanalyse, bei der auch linguistische Experimente erforderlich sind. Die Anfertigung der Datenbank entspricht daher nicht nur dem Bau eines Instruments oder eines Labors zu literarischer und linguistischer Forschung; vielmehr ist der Entwicklungsprozeß bereits Teil der Forschung, wie auch die Nutzung der Datenbank im Rahmen der Textforschung geschieht. Weil es kein „neutrales“, d.h. theoretisch unreflektiertes Eingeben linguistischer Daten geben kann, muß diese Tatsache immer wieder betont werden: die Erstellung einer Textdatenbank stellt in erster Linie selbst eine Forschungsaufgabe dar und schafft nicht nur die instrumentellen Rahmenbedingungen fundierter Textforschung.

24 Vgl. die Bemerkung N. LOHFINKS (Bewegung, 1995) 97 über die „moralische Verpflichtung“ mit den neuen Instrumenten die Daten zu den klas-

4. Die Kategorien der WIVU Datenbank

Anhand einiger Verse aus Numeri 13 werden im folgenden die in der WIVU-Datenbank verwendeten Kategorien beschrieben und erläutert, aus welchen Gründen bzw. mit welchem Ziel sie entwickelt wurden. Aus Kap. 3 ging bereits hervor, daß Kategorien aus drei Beschreibungsebenen kombiniert werden; es gibt

- Kategorien, die einen Text als *Dokument* (hierarchisch) gliedern,
- Kategorien, die einen Text als (lineare) Folge *distributionaler* linguistischer Einheiten („Atome“) darstellen, und schließlich
- Kategorien, die einen Text als ein (komplexes) Gefüge *funktionaler* linguistischer Einheiten beschreiben.

Das Beispiel einer komplexen Wortgruppe in Numeri 13,26 zeigt, daß es sich bei der linguistischen Beschreibung nicht nur um die Segmentierung des Textes in linguistische Einheiten geht, sondern auch um die Bestimmung der inneren und äußeren Beziehungen dieser Einheiten:

Numeri 13,26 ... ויבאו אל-משה ואל-אהרן ואל-כל-עדת בני-ישראל ...

Diese Wortfolge besteht - gemäß der zunächst durchzuführenden distributionalen Analyse - aus sechs Segmenten („phrase atoms“):

... [ויבאו] [אל-משה] [ו] [אל-אהרן] [ו] [אל-כל-עדת בני-ישראל] ...

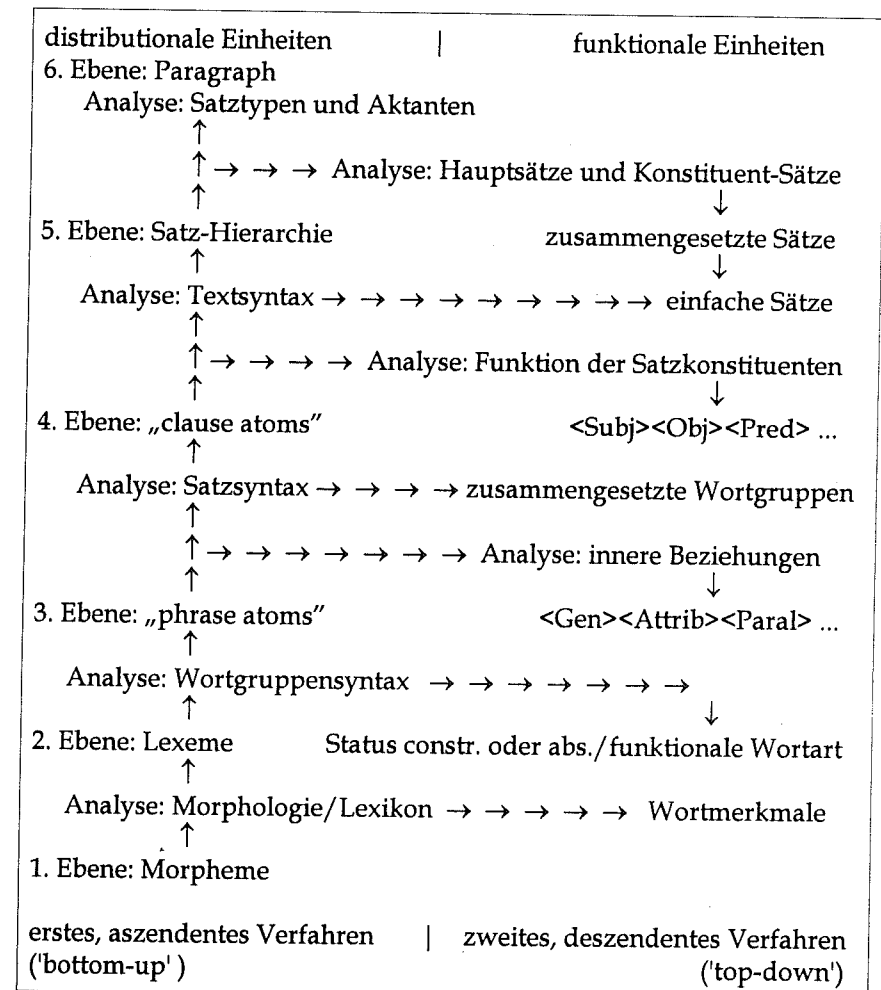
In einem weiteren Bearbeitungsgang dienen zwei weiteren Analyseschritte der Bestimmung der linguistischen Beziehungen und Funktionen: jedes Segment kann innere (z.B. Genitiv- oder Attribut-) Beziehungen und äußere Verbindungen (z.B. zu parallelen Einheiten oder zum Prädikat des Satzes) realisieren:

Innere Beziehungen (reg = regens; rec = rectum):
 ... [ישראל] {rec→} {←reg} בני {rec→} {←reg} עדת {rec→} {←reg} [אל-כל] ...
 Äußere Verbindungen (p = parallel):
 ... [אל-משה] [ו] {←p} [אל-אהרן] [ו] {←p} [אל-כל-עדת בני-ישראל] [p→] ...
 Äußere Verbindungen (cmpl = Komplement; pred = Prädikat):
 ... <cmpl> [ויבאו] <pred> [אל-משה ואל-אהרן ואל-כל-עדת בני-ישראל] ...

Die Beschreibung der Textdaten erfolgt also einerseits mit Dokument-orientierten Kategorien, die die Textgliederung - z.B. nach Kapiteln und Abschnitten bzw. Versen - festhalten, und andererseits mit linguistischen Kategorien, die auf jeder linguistischen Ebene die

Einheiten und deren Funktionen und Beziehungen erfassen. Zur Beschreibung aller in der Datenbank verwendeten Kategorien ergeben sich daher drei Fragestellungen bzw. Aufgaben, die im folgenden besprochen werden:

- In welcher Reihenfolge gewinnt man linguistische Textdaten? (Kap. 4.1.: Das Verhältnis distributionaler und funktionaler Daten)
- Welche Kategorien sind zur Textdaten-Speicherung nötig? (Kap. 4.2.: Ein Beispiel aus der WIVU-Datenbank)
- Wie können Textdaten für Abfragen genutzt werden? (Kap. 4.3.: Möglichkeiten für und Erwartungen an ein Retrieval-Programm wie QUEST 2)



Figur 7: Übersicht über den Analyse-Prozeß

4.1 In welcher Reihenfolge gewinnt man linguistische Daten?
Das Verhältnis distributionaler und funktionaler Daten

In diesem Abschnitt werden alle Schritte des analytischen Verfahrens kurz zusammengefaßt (vgl. Figur 7). Die primäre Analyse ist distributional und ascendend („bottom up“): aus kleineren Elementen werden größere Einheiten gebildet. Die sekundäre Analyse ist funktional und descendend und wird auf jeder Ebene durchgeführt: aus größeren Einheiten ergeben sich bestimmte Eigenschaften der kleineren Einheiten. Dazu gehören beispielsweise die Änderung der Wortart in einer Wortgruppe: ein Verb kann (als Partizip) syntaktisch die Funktion eines Adjektivs wahrnehmen.

Wie in Kap. 3 gezeigt wurde, basiert das Verfahren auf dem Prinzip der Musterwiedererkennung. Auf jeder Ebene bietet das Analyseprogramm Vorschläge an, die aus einer Systematisierung der bereits analysierten Daten gewonnen wurden. Dieses Vorgehen wird erst sekundär und nur dann, wenn genügend Daten für eine Konsistenzkontrolle vorliegen, durch daraus ableitbare grammatische Regeln oder lexikalische Informationen ergänzt, z.B. durch ein „Valenzwörterbuch“ im Fall der Satzanalyse.

Die Daten werden in fünf Arbeitsgänge - von der Morphemebene bis zur Textebene aufsteigend - produziert. Der Schlußsatz in Numeri

5.↑ Satzhierarchie/Texttyp: /[אודפרי הארץ] [יראום] [ן]/ Verbalsatz/Narrativer Text					
4.↓ Satzanalyse: /[<Obj> אודפרי הארץ] [<Pred+Obj> יראום] [<Conj> ן]/					
4.↑ Satzabgrenzung: /[אודפרי הארץ] [יראום] [ן]/					
3.↓ Phrasenanalyse: [ן] [אודפרי הארץ] [יראום] [ן] ↓ ↓ status abs status constr					
3.↑ Phrasenabgrenzung: [אודפרי הארץ] [יראום] [ן] (PrepP) (VerbP) (ConjP)					
2.↑ Lexikalische Analyse: ארץ ה פרי אוד יראום ן (Noun) (Art) (Noun) (Prep) (Verb) (Conj)					
1.↑ Morphologische Analyse: ארץ ה פרי אוד יראום ן 0/ - 0/ - Suffix+„W“ Prefix „J“					

Figur 8: Beispiel der „bottom-up“-Analyse von Ebene 1-5

13,26 illustriert einige der Kategorien, die in jedem Arbeitsgang hinzugefügt werden (Stufe 1-5 in Figur 8, vgl. auch Figur 7).

4.2 Welche Kategorien sind zur Textdaten-Speicherung nötig?
Ein Beispiel aus der WIVU-Datenbank

Die folgenden Figuren 9a-d zeigen anhand von vier Wörtern aus Numeri 13,32 die in der WIVU-Datenbank enthaltenen Daten auf Wort-, Phrasen und Satz-Ebene.

Beispiel Numeri 13,32: ארץ אכלה יושביה הוא

I. Dokument:

Buch	Numeri	Numeri	Numeri	Numeri
Kapitel	13	13	13	13
Vers	32	32	32	32
Halbvers	B	B	B	B
Wortform	ארץ	אכלה	יושביה	הוא
Wort-Nummer (absolut)	8765	8766	8767	8768

Figur 9a / 9b: Dokument-Eigenschaften / linguist. Eigenschaften auf Wortebene

(Wortform	ארץ	אכלה	יושביה	הוא)
II.1. Wort				
II.1.1. Morphologie (p = paradigmatische Form; r = realisierte Form)				
Präformativ (p)	.	empty	empty	.
Präformativ (r)	.	empty	empty	.
Stammmorphem (p)	.	empty	empty	.
Stammmorphem (r)	.	empty	empty	.
Verbalendung (p)	.	empty	empty	.
Verbalendung (r)	.	empty	empty	.
Nominalendung (p)	empty	ת	.	.
Nominalendung (r)	empty	ת	.	.
Suffix (p)	empty	empty	ה	.
Suffix (r)	empty	empty	ה	.
Locativsuffix (p)	empty	.	.	.
Locativsuffix (r)	empty	.	.	.
II.1.2. Lexikon:				
Lexem (p)	ארץ	אכל	ישב	היא
Lexem (r)	ארץ	אכל	ישב	הוא
Wortart (lexikalisch)	Nomen	Verb	Verb	Pers.Pron.
II.2. Wortfunktion:				
Verbfunktion	.	Ptz.	Ptz.	.
Verbalstamm	.	Qal	Qal	.
Person	.	empty	empty	3.
Numerus	sing.	sing.	plur.	sing.
Geschlecht	empty	Fem.	Masc.	Fem.
Status	st.abs.	st.abs.	st.constr.	.
Suffix	empty	empty	3.sg.F.	.
Wortart (grammatisch)	Nomen	Verb	Nomen	Pers.Pron.

(Wortform	ארץ	אכלת	יושביה	הוא
II.3.1. Phrase (distributional):				
Phrasenatom-Nummer (abs.)	4622	4623	4624	4625
Phrasenatom-Typ	nominal	verbal	nominal	pronominal
Determination	indet	indet	det	det
Phrasenatom-Verbindung	empty	empty	empty	empty
Abstand (Anzahl)	empty	empty	empty	empty
Abstand (Einheit)	empty	empty	empty	empty
II.3.2. Phrase (funktional):				
Interne Beziehung	empty	empty	empty	empty
Beziehungs-Typ	empty	empty	empty	empty
Bezugswort	empty	empty	empty	empty
Abstand (Anzahl)	empty	empty	empty	empty
Phrasen-Nummer (im Satz)	1	1	2	2
Phrasen-Typ	NP	VP	NP	NP(pron)
Determination	indet	.	det	det
Phrasen-Funktion	Kompl.	Kompl.	Obj.	Subj.

Figur 9c: Linguistische Eigenschaften auf Phrasenebene

(Wortform	ארץ	אכלת	יושביה	הוא
II.4.1. Einfacher Satz (distributional):				
Satzatom-Nummer (abs.)	1446	1447	1447	1448
Satzatom-Typ	empty	→	Partic.Act.	Defectiv
Abstand	-3	→	-1	-2
Kodierung Satzrelation	100	→	160	223
II.4.2. Einfacher Satz (funktional):				
Satz-Nummer (im Satzgefüge)	4	5	5	4
Satz-Typ	→	→	NomSatz	NomSatz
Konstituenten-Satz	Wiederaufnahme	→	AttrSatz	←
Abstand	-3	→	-2	←
Abstand (gezählt in:)	Satzatom	→	Phr	←
Satzfunktion	→	→	[Experimentaldaten]	
II.5.1. Zusammengesetzter Satz (distributional):				
Satzgefüge-Nummer (abs.)	822	822	822	822
II.5.2. Zusammengesetzter Satz (funktional):				
Satzgefüge-Nummer (im Kap.)	59	59	59	59
II.6. Kommunikationsebene:				
Text-Typ	→	→	NQ	NQ
	(direkte Rede in narrativem Kontext)			
Paragraph	→	→	[Experimentaldaten]	
Aktant	→	→	[Experimentaldaten]	

Figur 9d: Linguistische Eigenschaften Satzebene und darüber

4.3 Wie können Textdaten für Abfragen genutzt werden? Möglichkeiten für und Erwartungen an ein Retrieval-Programm

Zur Beantwortung der dritten Frage nach der Nutzung der Daten kann eine vereinfachte Präsentation von Textdaten oberhalb der Wortebene dienen (vgl. Figur 12 im Anhang). Sie zeigt, wie der Text in „Atome“ segmentiert wird und wie diese „Atome“ zu funktionalen sprachlichen Einheiten gruppiert werden. Man gewinnt auf diese Weise die Möglichkeit, beide Datentypen aus der Datenbank wieder zu erfragen: die distributionalen Einheiten (ungeteilte Phrasen und Sätze oder Teile unterbrochener Phrasen und Sätze) und die funktionalen Einheiten (Phrasen auf Konstituentenebene, oder Sätze auf Textebene).

4.3.1 Beispiele zur Nutzung paradigmatischer Daten

a) Wie findet man Sätze?

Für einen Hebraisten oder für einen Exegeten ist die Suche nach Sätzen eine - vielleicht doch nur scheinbar - schlichte Frage, für eine einfache Maschine jedoch eine komplizierte Angelegenheit. Ein Retrieval-Programm muß bei dieser Frage bereits die Vorentscheidung treffen, daß im Bereich der linguistischen Daten nach *funktionalen* sprachlichen Einheiten zu suchen ist und unterbrechende und unterbrochene Einheiten bei der Suche einkalkuliert werden müssen. Bei den in Figur 12 präsentierten Daten treten Sätze nicht als selbständige Größen auf, sondern nur als Teile größerer linguistischer Einheiten, z.B. als Konstituenten zusammengesetzter Sätze.²⁵

Für einen Benutzer der Datenbank behält die Suchaufgabe die Gestalt einer einfachen Abfrage, z.B.: „Suche die in Numeri 13 enthaltenen Sätze“. Die Abarbeitung dieser Suchaufgabe geschieht auf der Basis der WIVU-Daten folgendermaßen:

Zur „Treffermenge“ werden alle Wortfolgen gezählt, bei denen die Satz-Nummer und die Satzgefüge-Nummer übereinstimmen, wie z.B. in Numeri 13,32: Satz = 1 und zusammengesetzter Satz = 1, u.s.w. So erhält der Benutzer auch die Sätze, die durch Nebensätze in zwei oder mehr Abschnitte unterteilt worden sind. Die Suche nach spezifischen Satzformen oder Satzteilen ergibt sich als qualifizierte Teilmenge der so gefundenen Sätze:

²⁵ Auf der Ebene der zusammengesetzten Sätze...

- Die Suche nach *Nominalsätzen* in Numeri beschränkt die Abfrage auf Sätze mit der Kenzeichnung „NS“ (Nominalsatz), wobei wieder die Nummer des Satzes mit der des zusammengesetzten Satzes übereinstimmen muß.
- Die Suche nach den *Objekten in Verbalsätzen* in Numeri ermittelt die Wortfolgen mit dem Label „Obj“ und registriert die Nummer der Phrase innerhalb des Satzes, die des Satzes und die des zusammengesetzten Satzes. Zum Ergebnis gehören alle Phrasen mit derselben Satznummerierung. Schließlich als zu prüfen, ob die letzte Phrase eines Satzes als „VS“ (Verbalsatz) gekennzeichnet ist. Diese Aktion dient eigentlich nur zur Kontrolle, da man nur in Verbalsätzen Objekte erwarten kann.

b) Beispiele der Phrasen-Struktur

Die Analyseergebnis der phrasen-internen Beziehungen wird als Zahlenkombination gespeichert. Ein Benutzer kann anhand dieser Daten sowohl die innere Struktur von Phrasen wie auch deren Teile untersuchen. So kann beispielsweise die vollständige Wortgruppe *ממי המרים המארים האלה* aus Numeri 5,19 dem Benutzer auf folgende Weise präsentiert werden, wobei die intern verwendeten Zahlenkombinationen als „Label“ dargestellt werden:

[[מ {מי dem+} {reg} {+atr} {מרים rec-} {+atr} {מארים dem-} {האלה}]]

Weiterhin ergibt sich die Möglichkeit, nach Phrasen mit einer bestimmten inneren Beziehung zu suchen, z.B. nach Wortgruppen mit einem demonstrativen Element. Zum Ergebnis gehört dann die folgende Wortgruppe:

[[מ {מי dem+} {מרים המארים} {האלה}]]

Die Daten erlauben auch, die Struktur der Wortgruppenbildung im Hebräischen zu untersuchen, indem man z.B. eine Folge von Wortarten definiert und die diesem Muster entsprechenden und im Suchbereich vorhandenen Phrasen als Ergebnis erhält. Einige Beispiele:

Nomen - Nomen	
Numeri 35,14	{ערי reg+} {מקלט rec-} תהיינה
Numeri 28,6	{עלת reg+} {תמיד rec-}
Numeri 28,3 (ajt = „adjunctive“)	{עלת ajt+} {תמיד ajt-}
Nomen - Adjektiv	

Nomen - Nomen - Nomen	
Numeri 15,10	{אשה reg+} {ריחד rec-} {+reg} {ליחה rec-} {+reg} {ניחה} ליהיה
Nomen - Adjektiv -Konjunktion - Adjektiv	
Numeri 6,3	{ענבים atr+} {+atr} {לחים par-} {+par} {לא יאכל} {+atr}
Nomen - Artikel - Nomen - Konjunktion - Nomen - Artikel - Nomen	
Numeri 31,48	{שרי reg+} {האלפים rec-} {+reg} {+par} {+par} {שרי par-} {+reg} {המאות} {+reg}
Präposition - Nomen - Artikel - Nomen - Artikel - Adjektiv	
Numeri 5,24	{השקה אתהאשה את reg+} {מי reg+} {+reg} {מרים rec-} {+atr} {מארים atr-}
Präposition - Nomen - Art. - Nomen - Art. - Adjektiv - Art. - Dem.Pron.	
Num. 5,19	{הנקי ממי dem+} {+dem} {+reg} {מרים rec-} {+atr} {מארים atr-} {האלה dem-}

4.3.2 Beispiele zur Nutzung syntagmatischer Daten

a) Syntax und komplexe Strukturen

Für einige Untersuchungen werden nicht nur Sammlungen paralleler Fälle benötigt, sondern eher ein Überblick über bestimmte Kombinationen linguistischer Daten in einem Text, z.B. unerwartete Übergänge von narrativen Sätzen zu direkter oder indirekter Rede wie in Numeri 5,2 („Gebiete den Israeliten, daß sie ... hinausschicken.“) oder 9,1 („JHWH redete ... und sagte, daß die Israeliten ... feiern sollen.“).

[<Co>] [אל משה] [<Su>] [יהוה] [<Pr>] [ידבר] [<C>] [ו] [<C>] [לדבר] [<Pr>] [לאמר]	1 N	WayX	3sgM	1
.....	2 N	InfC	- - -	1
[<Ob>] [את בני ישראל] [<Pr>] [צו] [<Pr>] [מנהלמחנה] [<Pr>] [ישלחו] [<C>] [ו] [<Ob>] [כל צרוע וכל זב וכל] [<Co>] [טמא] [<PC>] [לנפש] [<Co>]	3 NQ	ImpV	2sgM	2
.....	4 NQ	Way0	3plM	2
.....	5 NQ	AjCl	- sg -	2
[<PC>] [מזכר] [עד נקבה] [<Colsp>] [תשלחו] [<PC>] [אל מחוץ למחנה] [<Co>] [תשלחום] [<PO>] [ו] [<C>] [לא] [<Ng>] [יטמאו] [<Pr>] [את מחניהם] [<Ob>] [אשר] [<Re>] [אני] [<Su>] [שכן] [<PC>] [בתוכם] [<Co>]	6 NQQ	xYqt	2plM	3
.....	7 NQQ	xYqt	2plM	3
.....	8 NQQ	WxYq	3plM	3
.....	9 NQQ	PtcA	-sgM	3
.....	10 N	WayX	3plM	4
[<Su>] [יעשו] [<Pr>] [בן] [<Mo>] [בני ישראל] [<Co>] [אל מחוץ למחנה] [<Co>] [אשר] [<Re>] [ידבר] [<Pr>] [יהוה] [<Su>] [ואל משה] [<Co>]	11 N	Way0	3plM	4
.....	12 N	xQtl	3sgM	4
[<C>] [בן] [<Mo>] [יעשו] [<Pr>] [בני ישראל] [<C>] [ו] [<C>] [ידבר] [<Pr>] [יהוה] [<Su>] [ואל משה] [<Co>] [ו] [<C>] [ידבר] [<Pr>] [יהוה] [<Su>] [ואל משה] [<Co>]	13 N	OQtI	3pl -	4
.....	14 N	WayX	3plM	5

Zum Auffinden solcher Übergänge sind Daten zur Satzhierarchie erforderlich, z.B. auf die Satztyp-Folge 'Imperativ + WeYiqtol' bzw. 'Infinitiv „sagen“ + WeYiqtol'. Als Ergebnis einer Abfrage zur Satz-hierarchie kann die grammatische Struktur eines Textabschnittes auf dem Bildschirm gezeigt werden (vgl. Figur 10).

b) Syntax und Texttypkombinationen

Das Programm zur Erstellung der Texthierarchie ordnet beim Aufbau der Datenbank allen Sätzen in einem Text Texttypmarkierungen zu.²⁶

- ein Wayyiqtol-Satz erhält die Texttypmarkierung „N“ (für „narrative“);
- eine von אָמַר oder דָּבַר eingeleitete direkte Rede wird mit „Q“ (für „quotation“),
- ein Imperativ- oder ein Yiqtol-Satz ohne Redeeinleitung mit „D“ (für „discourse“) gekennzeichnet;
- fehlen grammatische Hinweise, wird die Kennzeichnung des nächsthöheren Satzes übernommen.

Bei der Untersuchung des Textes können diese Daten dazu verwendet werden, Textstellen mit unerwarteten Kombinationen ausfindig zu machen, wie z.B. unerwartete diskursive Texte in einem narrativen Text.

So bedeutet die Kennzeichnung „ND“, daß in einen - durch einen Wayyiqtol-Satz eingeleiteten - narrativen Text Abschnitte mit Yiqtol-Sätzen auftreten. Beispiele sind Numeri 3,23-26.29-31.35-37, wo gesagt wird, daß die Söhne Aarons hinten, südlich und nordlich des Heiligtums „zelteten“, oder „zelteten sollen“? Viele moderne Übersetzungen bieten hier die narrative Form „zelteten“, ohne die besondere diskursive Verbform des Hebräischen zu berücksichtigen. Zu diesen und anderen Stellen kann man sich aber fragen ob es nicht besser sei sie als Instruktion, Regel oder als Gewohnheit zu übersetzen (So z.B. Buber). Vgl. Numeri 9,15-23; 10,36; 11,20,33; 13,33; 15,34; 17,5.

Die Kennzeichnung „QN“ verweist auf das ebenso unerwartete Auftreten (narrativer) Wayyiqtol-Sätze innerhalb einer direkte Rede. Beispiele finden sich u.a. in Numeri 5,20,27; 8,17-19; 14,16,22,24; 16,14; 20,15.

²⁶ Die hier verwendete Klassifizierung folgt den textgrammatischen Anweisungen von H. WEINRICH (Tempus, 1994⁹) und W. SCHNEIDER (Grammatik,

c) Syntax und Sätze: funktionale Analyse und Übersetzung.

Die in der Satzkonstituenten-Analyse verwendeten Kategorien (Prädikat, Subjekt, Objekt, Komplement, usw.) ermöglichen systematische Vergleiche auf dem Gebiet der verbalen Valenz. Das ist insofern wichtig, als in den alten Sprachen manchmal kaum nachgeprüft werden kann, welche semantischen Effekte bestimmte Kombinationen von Verben mit Objekten oder Komplementen haben können. Systematische Informationen zur Verbvalenz können hier ebenso weiterhelfen wie bei der Konsistenzüberprüfung von Übersetzungen. Ohne damit zu fordern, daß ein Übersetzer das hebräische Idiom einfach imitieren sollte, ist dennoch eine gute Verständlichkeit der hebräischen Spracheigenarten nötig. Grammatische und semantische Textdaten entscheiden natürlich nicht über die Theorie der Übersetzung, aber sie ermöglichen umfangreiche Vergleiche und eine erweiterte Konsistenzprüfung und können daher auch der konzisen Theoriebildung dienen.

Verbvalenz עַם הַלֵךְ	Buber gehen mit	Luther Bibel 1985 gehen mit 22,12; ziehen mit 10,32; 22,13.14.21.35.35.39
אֵל הַלֵךְ	gehen in 22,13 gehen nach/zu 10,30 gehen zu 16,25; 22,16 gehen an (Ort) 23,13	gehen in ziehen in/zu gehen zu / ziehen zu kommen an (Ort)
אִתּוֹ הַלֵךְ בְּ הַלֵךְ	gehen mit 10,29; 22,20 gehen auf den Weg 21,22	kommen mit / ziehen mit Obj + ziehen (=20,17 [ohne בּ !])
	gehen ins Feld 22,23	gehen auf

Figur 11: Beispiel mit הֵלֵךְ (gehen) in Numeri

5. Ergebnis

Eine brauchbare Datenbank zur Analyse von Texten der antiken Literatur sollte mindestens drei Typen sprachlicher Daten verwenden, deren Kategorien zur Darstellung und Speicherung der Textstrukturen dienen, der Dokumentstruktur einerseits und der distributionalen sowie der funktionalen linguistischen Struktur andererseits.

Im Blick auf die Dokumentstruktur müssen die klassischen Gliederungs-Kategorien (bei biblischen Texten: Buch, Kapitel, Vers, Halbvers) stets zugänglich sein.

Hinsichtlich der linguistischen Struktur liefert eine erste, distributionale Analyse der Texte linguistische Daten in einer vorläufigen Fassung: die kleinsten, auf ihrer jeweiligen sprachlichen Ebene nicht weiter teilbaren Einheiten, die sogenannten „Atome“. Diese ermöglichen es, einen Text auch im Zuge der linguistischen Untersuchung in einer vollständigen Hierarchie zu beschreiben, ohne Lücken (gapping) und eingebettete Einheiten (embedding). Die Atome dienen somit als Elemente einer Meta-Sprache, mit deren Hilfe die funktionalen sprachlichen Einheiten eines Textes definiert werden können.

Dieser Weg linguistischer Textkodierung bietet ein offenes System in dem Sinne, daß die so erstellte Textdatenbank für viele Zwecke, von phonologischen und morphologischen Untersuchungen angefangen bis zur Analyse textlicher Makrostrukturen, eingesetzt und daher von Linguisten und Literaturwissenschaftlern, von Übersetzern und Exegeten genutzt werden kann.

Nach der Datenproduktion mit Hilfe von Musterwiedererkennungsprogrammen ist die Weiterentwicklung der Retrievalsoftware notwendig. Denn diese muß einerseits die Suche nach paradigmatische Daten (die Sammlung von möglichst vielen Parallelstellen zu dem untersuchten Phänomen), sondern auch die Suche nach syntagmatischen Textstrukturen (die Untersuchung möglichst vieler Phänomene in einer Textkomposition) ermöglichen. Auch die Programme zur Datenanalyse werden weiter entwickelt, insbesondere mit dem Ziel der Textsegmentierung auf höheren Ebenen (z.B. die linguistische Definition von Textparagrafen) und Verfahren zur Wiedererkennung von Aktanten in einem Text (participant tracking).

Literatur

- Biblia Hebraica Stuttgartensia [BHS] (Stuttgart 1967-1977)
 Biblia Hebraica Quinta [BHQ] (Stuttgart 1999ff.)
 FRANCIS I. ANDERSEN, *The Sentence in Biblical Hebrew*, Janua Linguarum, Series Practica 231 (The Hague 1974)
 WALTER R. BODINE, *How Linguists Study Syntax*, in: WALTER R. BODINE (Hg.), *Linguistics and Biblical Hebrew* (Winona Lake, Indiana 1992) 89-107
 EUGENIO COSPRILI, *Textlinguistik*. Eine Einführung (Tübingen 1972)

- CHRIST-JAN DOEDENS, *Text Databases*. One Database Model and Several Retrieval Languages, *Language and Computers*. Studies in Practical Linguistics 14 (Amsterdam 1994)
 HUBERT L. DREYFUS, *What computers still can't do. A Critique of Artificial Reason* (Cambridge, Mass. 1992)
 WALTER ECKARDT, *Computergestützte Analyse althebräischer Texte*. Algorithmische Erkennung der Morphologie, *Arbeiten zu Text und Sprache im Alten Testament [ATS]* 29 (St. Ottilien 1987)
 A. F. DEN EXTER BLOKLAND, *Clause-Analysis in Biblical Hebrew Narrative - an Explanation and a Manual for Compilation*, *Trinity Journal* 11 (Chicago 1990) 73-102
 ELISABETH GÜLICH/WOLFGANG RAIBLE, *Überlegungen zu einer makrostrukturellen Textanalyse*. [Am Beispiel von] J. Thurber, *The Lover and his Lass*, in: ELISABETH GÜLICH/KLAUS HEGER/WOLFGANG RAIBLE, *Linguistische Textanalyse*. Überlegungen zur Gliederung von Texten, *Papiere zur Textlinguistik* 8 (Hamburg 1974 1979²) 73-126, auch in: ELISABETH GÜLICH/KLAUS HEGER/WOLFGANG RAIBLE, *Untersuchungen in Texttheorie* (Göttingen 1977) 132-175
 CHRISTOF HARDMEIER/EEP TALSTRA, *Sprachgestalt und Sinngehalt*. Wege zu neuen Instrumenten der computergestützten Textwahrnehmung, in: *Zeitschrift für die alttestamentliche Wissenschaft [ZAW]* 101 (Berlin 1989) 408-428
 NORBERT LOHFINK, *Gab es eine deuteronomistische Bewegung?*, in: WALTER GROSS (Hg.), *Jeremia und die „deuteronomistische Bewegung“*, *Bonner biblische Beiträge [BBB]* 98 (Weinheim 1995) 313-382, auch in: NORBERT LOHFINK, *Studien zum Deuteronomium und zur deuteronomistischen Literatur III*, *Stuttgarter Biblische Aufsatzbände [SBAB]* 20 (Stuttgart 1995) 65-142
 C. H. J. VAN DER MERWE, *Discourse Linguistics and Biblical Hebrew Grammar*, in: ROBERT D. BERGEN (Hg.), *Biblical Hebrew and Discourse Linguistics*, (Winona Lake, Indiana 1994) 13-49.
 WOLFGANG RICHTER, *Grundlagen einer althebräischen Grammatik*. B. Die Beschreibungsebenen: III. *Der Satz* (Satztheorie), *Arbeiten zu Text und Sprache im Alten Testament [ATS]* 13 (St. Ottilien 1980)
 WOLFGANG RICHTER, *Biblia Hebraica transcripta [BH^T]*, 3. *Numeri*, *Deuteronomium*, *Arbeiten zu Text und Sprache im Alten Testament [ATS]* 33.3 (St. Ottilien 1991) [Rezension von E. Talstra in: *Journal of Semitic Studies [JSS]* 39 (Leiden 1994) 290-329]
 JOHN H. SAILHAMER, *A Database approach to the analysis of Hebrew Narrative*, *Maarav* 5-6 (Santa Monica, Calif. 1990) 319-335
 JOHN H. SAILHAMER, *2 Samuel 13:1-4 and a Data Base Approach to the Analysis of Hebrew Narrative*, in: *Association Internationale Bible et Informatique* (Hg.), *Actes du Troisième Colloque International, Bible et Informatique: Interprétation, Herméneutiques, Compétence Informatique*, Tübingen 26-30 août 1991 (Paris/Genève 1992) 99-122
 WOLFGANG SCHNEIDER, *Grammatik des biblischen Hebräisch*. Ein Lehrbuch (München 1993⁸)

WOLF-DIETER SYRING, *QUEST 2 - Computergestützte Philologie und Exegese, Zeitschrift für Althebraistik [ZAH] 11* (Stuttgart 1998) 85-89.

EEP TALSTRA, *Hebrew Syntax: Clause Types and Clause Hierarchy*, in: K. JONGELING/H. L. MURRE - VAN DEN BERG/L. VAN ROMPAY (Hg.), *Studies in Hebrew and Aramaic Syntax presented to Professor J. Hoftijzer* (Leiden 1991) 180-193

EEP TALSTRA/CHRISTOF HARDMEIER/ALAN GROVES (Hg.), *Quest. Electronic Concordance Applications for the Hebrew Bible [data base and retrieval software]* (Haarlem 1992)

EEP TALSTRA, *Demonstration ECA database and retrieval software. A preliminary Report*, in: *Association Internationale Bible et Informatique (Hg.), Actes du Troisième Colloque International, Bible et Informatique: Interprétation, Herméneutiques, Compétence Informatique*, Tübingen 26-30 août 1991 (Paris/Genève 1992) 605-611

EEP TALSTRA, *Text Grammar and Biblical Hebrew: The Viewpoint of Wolfgang Schneider*, *Journal of Translation and Textlinguistics 5* (Dallas, Texas 1992) 269-297

EEP TALSTRA, *Reconstructing the Menorah on Disk. Some Syntactic Remarks*, in: MARC VERVENNE (Hg.), *Studies in the Book of Exodus, Bibliotheca ephemeridum theologicarum Lovaniensium [BETL] 126* (Leuven 1996) 523-533

EEP TALSTRA, *A Hierarchy of Clauses in Biblical Hebrew Narrative*, in: E. J. VAN WOLDE (Hg.), *Narrative Syntax and the Hebrew Bible. Papers of the Tilburg Conference 1996*, *Biblical Interpretation Series 29* (Leiden, 1997) 85-118

ARIAN J. C. VERHEIJ, *Grammatica Digitalis I. The Morphological Code in the „Werkgroep Informatica“ Computer Text of the Hebrew Bible, Applicatio 11* (Amsterdam 1994)

B. WALTKE/M. O'CONNOR, *An Introduction to Biblical Hebrew Syntax* (Winona Lake, Indiana 1991)

HARALD WEINRICH, *Tempus. Besprochene und erzählte Welt, Sprache und Literatur 16* (Stuttgart [u.a.] 1994⁵)

Anhang

Die Struktur der Textdaten (vgl. die beiden folgenden Seiten)

Struktur der Textdaten (vereinfachte Wiedergabe)

p	Zusg.Satz	Funktion Satz		↔	Form 'clause atom'		↔	Funktion Wortgruppe		↔	Form 'phrase atom'		↔	Wort Textnum	
		Idf	Bez		Typ	Idf		Fnkt	Bez		Typ	Idf		Fnkt	Bez
1	1	1						1	Cj					1	Num 13,32
1	1	1						2	Pred					2	Num 13,32
1	1	1						3	Obj	reg				3	Num 13,32
1	1	1						1	Rel	rec				1	Num 13,32
1	1	1						2	Pred					2	Num 13,32
1	1	1						3	Obj	reg				3	Num 13,32
1	1	1						3	Cmpl	rec				3	Num 13,32
1	1	1						1	Pred					1	Num 13,32
2	2	2						1	Front					1	Num 13,32
2	2	2						1	Rel					1	Num 13,32
2	2	2						2	Pred					2	Num 13,32
2	2	2						3	Loc					3	Num 13,32
2	2	3						1	Pred					1	Num 13,32
2	2	3						2	Obj					2	Num 13,32
2	2	4						1	PCmp					1	Num 13,32
2	2	5						1	PCmp					1	Num 13,32
2	2	5						2	Obj					2	Num 13,32
2	2	4						2	Subj					2	Num 13,32

