

Biblia Hebraica *transcripta* – Das digitale Erbe

Christian Riepl, München

Anlässlich eines Festakts in der „Akademie der Halle Salomos zu Herstelle“ am 3. März 1994 legte Wolfgang Richter seinem Schülerkreis einen „Bericht über zehn Jahre Seminar für ugaritische und hebräische Sprach- und Literaturwissenschaft“¹ vor und beschrieb darin die Motivation für den Einsatz des Computers bei der linguistischen Analyse altorientalischer Texte so: „Wie immer stand ich unter dem bestimmenden Einfluß meiner Mitarbeiter. Herr Eckardt machte mir gleich am Anfang von 1984 klar: Jede Klasse und jede Regel ist als Klasse oder Regel über eine zählbare, nicht gezählte Menge formuliert. Mengen kann man berechnen, am schnellsten kann das der Computer, ein Rechner muß her. Seitdem arbeiten wir am Rechner.“² Die Vorteile einer rechnergestützten Arbeitsweise lagen für Richter auf der Hand: Vor allem ermöglichte sie Teamarbeit. So dann konnten die digitalen Daten erweitert, geändert und verbessert werden. Schon zu Beginn der Dateneingabe aber war für das gesamte spätere Projekt „Biblia Hebraica *transcripta*“³ maßgebend: Mit Hilfe von Compu-

¹ Siehe RICHTER 1994a, 1-5: Der „Festgabe“ waren Vorabdrucke von Vorträgen in Bern (siehe RICHTER 1994b, 231-235 und RICHTER 1994c, 237-245) und Münster (Zum syntaktischen Gebrauch von Substantiven im Althebräischen am Beispiel von *ʾōd*. Ein Beitrag zur Partikelforschung, Vortrag, gehalten am 26.7.1993 auf der Session „Hebrew – Semitics – Linguistics“ des SBL International Meeting in Münster), ferner Arbeitspapiere zu Stammesmodifikationen und Wortarten sowie Auszüge aus der Datenbank „bhtdb version 2“ beigefügt.

² Siehe RICHTER 1994a, 1.

³ Einen Überblick über das Projekt von 1986 bis etwa Mitte der 1990er Jahre mit einer Beschreibung der Programme, der Verfahrensweise ihrer Anwendung, Entwürfen einer Datenbank und Modellen einer Methodik rechnergestützter Sprach- und Literaturwissenschaft gibt RIEPL 1996b, 1997 und 1999. Die weitere Entwicklung des Projekts wurde von Hans Rechenmacher und Christian Riepl im Workshop „Alter Wein in neue Schläuche. Das Projekt ‚Biblia Hebraica *transcripta*‘ und die ‚Digital Humanities‘: Projekt – Datenbank – Potenzial“ am 8.8.2013 auf dem 21. Kongress der International Organization for the Study of the Old Testament, IOSOT 2013 in München vorgestellt. In jüngster Zeit weckt das Projekt auf den Gebieten Linguistik und Computerlinguistik das Interesse der sich etablierenden Forschungsrichtung „Digital Humanities“, was sich u.a. an einer Reihe von Vorträgen des Autors zeigt: Das Projekt „Biblia Hebraica *transcripta*“: Eine kollaborative Forschungsdatenbank, Vortrag, gehalten am 6.2.2014 an der Universität München auf dem XX. LIPP Symposium 2014: Linguistik 2.0 – Die Herausforderung der „Digital Humanities“; Was ich nicht weiß, ... macht mich heiß:

terprogrammen konnten in einer automatischen Analyse grammatische Regeln auf das gesamte Textkorpus angewendet werden.

Zur Genese der Daten

Richters langjährige Grundlagenforschungen zur Grammatik des Althebräischen und zur Methodik der Transkription boten zusammen mit einer dadurch vorbereiteten formalisierten Denk- und systematischen Herangehensweise ideale Voraussetzungen für die Anwendung informatischer Methoden in einer philologischen Disziplin. Von Anfang an wurde eine enge interdisziplinäre Kooperation mit der Informatik⁴ gepflegt. Große Mengen an Text- und Metadaten sowie komplexe Regeln der Grammatik interessierten Informatiker, die in den 1980er Jahren auf den Gebieten „Künstliche Intelligenz“ (Logikprogrammierung und deren Auswertungsstrategien, deduktive Datenbanken, Expertensysteme), in den 90er Jahren auf den Gebieten „Informationretrieval/Informationssysteme“ (Featurebaumstrukturen), „Multimedia“ (Multimediatdatenbanken) und „Webtechnologie“ (Webbasierte Datenbanken, Social Software) forschten.

Die „Computer-Aufnahme von Texten des Alten Testaments und des Alten Orients“⁵ erfolgte manuell in einer orthographiebezogenen, morphologisch-syntaktischen Transkription⁶ unter Beachtung einer eindeutigen und logischen Strukturierung der Textdaten und einer eindeutigen Kodierung der Transkriptionszeichen in einem definierten Betacode. So waren

Zum Mehrwert der Anwendung informatischer Methoden bei der Analyse von Textkorpora am Beispiel des Projektes „Biblia Hebraica *transcripta*“, Vortrag, gehalten am 26.3.2014 an der Universität Passau auf der ersten Jahrestagung des Verbandes DHd – „Digital Humanities im deutschsprachigen Raum“, DHd 2014 – Digital Humanities – methodischer Brückenschlag oder „feindliche Übernahme“? Chancen und Risiken der Begegnung zwischen Geisteswissenschaften und Informatik, Internetveröffentlichung der Präsentation auf der Web-Site des Verbandes DHd: <http://www.dig-hum.de/programm-präsentationen-videos-bildern> (http://dig-hum.de/sites/dig-hum.de/files/Riepl_Publikation.pdf); The Project „Biblia Hebraica transcripta“ and the „Digital Humanities“. Data, Programs, Methods and Potential, Lecture given at the CIS NLP Talk, LMU Munich, 13. Jan. 2015.

⁴ Beteiligt waren Rudolf Bayer und Günther Specht am Lehrstuhl für Datenbanksysteme und Wissensbasen an der Fakultät für Informatik der Technischen Universität München, später Ulrich Güntzer und Hans Argenton am Arbeitsbereich Datenbanken des Wilhelm-Schickard-Instituts für Informatik der Eberhard-Karls-Universität Tübingen.

⁵ So der Titel eines Interviews von Wolfgang Richter und Manfred Krebernik bei TSCHUNKE 1988, 29-32, das die ersten Erfahrungen in dem zunächst von der Ludwig-Maximilians-Universität München geförderten Pilotprojekt dokumentiert.

⁶ Siehe dazu RICHTER 1983.

nicht nur die größtmögliche Kompatibilität zwischen den Projektbeteiligten und die Voraussetzungen für die maschinelle Verarbeitung der Daten gewährleistet, sondern zugleich auch alle Grundlagen für die heute bei digitalen Projekten eingeforderte Nachhaltigkeit, Nachnutzbarkeit und Konvertierbarkeit in beliebige Strukturen und Kodierungen geschaffen.

Die Regeln der Morphologie⁷ und der Morphosyntax⁸, die Richter an einem Ausschnitt des Alten Testaments formuliert hatte, wurden mit Hilfe der beiden Expertensysteme SALOMO⁹ und AMOS¹⁰ nun erstmals auf das gesamte Textkorpus angewendet. Dies geschah jeweils in einem dialogischen Verfahren des Programms mit dem Experten, der bei grammatisch mehrdeutigen Analyseergebnissen eine Auswahl treffen musste, die wiederum unter Angabe von Kriterien und der sprachlichen Ebene, auf der sie wirken, eine Begründung erforderte. Der dynamisch, dialogisch und spiralförmig ablaufende Prozess der Regelfindung, somit auch Erkenntnisfortschritt und Regelwerk wurden damit transparent, das Regelwerk zunehmend verfeinert, ergänzt und verbessert. Die Analyse der Satzebene¹¹ sah zunächst mit einem Versuch der Analyse der Kernsätze¹² vielversprechend aus, konnte aber einerseits aufgrund der hohen Komplexität der Fakten und Regeln, andererseits aufgrund geänderter Rahmenbedingungen nicht weiter verfolgt werden.

Alle bislang erarbeiteten und in strukturiertem Format vorliegenden Daten (Transkriptionen mit Satzgrenzen, Textanmerkungen, Ergebnisse der morphologischen und morphosyntaktischen Analyse) wurden Anfang der 1990er Jahre reorganisiert und in eine relationale Datenbank (bhddb version 2) eingelesen, darin bearbeitet und erweitert.¹³ Nach der Analyse der Eigennamen wurden die Nominal- und Verbalformen sowie die Wort-

⁷ Siehe dazu RICHTER 1978.

⁸ Siehe dazu RICHTER 1979.

⁹ Entwickelt von Walter Eckardt und dokumentiert bei ECKARDT 1987.

¹⁰ Entwickelt von Günter Specht und dokumentiert bei SPECHT 1990.

¹¹ Zur Satztheorie und den Regeln der Syntax siehe RICHTER 1980.

¹² Siehe dazu HEUMEL 1997.

¹³ Die gegenüber der gedruckten Erstausgabe bei RICHTER 1991-1993 erfolgten Änderungen im Datenbestand wurden dokumentiert. Die Textdateien wurden als *Biblia Hebraica transcripta*, Release 1-4, BH^f-Software bei RICHTER 1991, 1994, 1996 und 1997 freigegeben. Die Korrekturen wurden sowohl in den Textdateien als auch in den Datenbanktabellen vorgenommen. Die Erweiterung des Datenbestandes umfasste z. B. auch die Einführung weiterer Bezugsmarker bei Pendenskonstruktionen und eines zusätzlichen Referenzsystems zur Identifikation der Sätze, siehe dazu RIEPL 1996a.

fügungen systematisch geordnet und veröffentlicht.¹⁴ Auf die Datenbank konnten folgende vier Vorhaben aufbauen:

Hans Argenton ging von den in Baumstrukturen vorliegenden Analyseergebnissen auf Wortfügungsebene aus und entwickelte VENONA¹⁵, ein sprachunabhängiges Informationssystem zur Indexierung von und zum effizienten Retrieval in Baum- und Feature-Strukturen. Das hoch performante System wurde erfolgreich zur Suche nach komplexen Formationen morphologischer und morphosyntaktischer Einheiten in Wortverbindungen eingesetzt. Es basierte auf einem Datenmodell, das attribuierte Mehrwegbäume verwendete. Diese eignen sich in hohem Maß dazu, komplexe sprachliche Strukturen unter grammatischen und semantischen Gesichtspunkten einschließlich von Phänomenen wie etwa Tilgung und Pronominalisierung abzubilden. Komplexe Regeln und Strukturen, die sich auf den Ebenen Text, Satzfügung und Satz aus dem Rück- bzw. Vorgriff auf Daten mehrerer sprachlicher Ebenen ergeben, effizient zu verarbeiten, stellt hohe Anforderungen an Retrievalsysteme und wird auch in Zukunft eine interessante und anspruchsvolle Aufgabe bleiben, zu der nun auch informatische Strategien und Methoden aus den Gebieten „big data“ und „data mining“ herangezogen werden könnten.

Ausgehend von dem zunächst offline laufenden multimedialen Datenbanksystem MultiBHT¹⁶ entstand im Rahmen des Projekts „MultiMAP/2“ unter der Leitung von Günther Specht und Martin Zirkel in der zweiten Hälfte der 1990er Jahre das webbasierte Informationssystem MultiBHT¹⁷. Erstmals war es möglich, mit Hilfe der gerade aufkommenden Webtechnologien von einer webbasierten Benutzeroberfläche aus auf multimediale Inhalte einer Datenbank zuzugreifen, wobei diese außerdem auf vielfältige Weise untereinander verlinkt werden konnten. MultiBHT war ausgestattet mit einer Benutzer- und Gruppenverwaltung. Nach erfolgreicher Anmeldung bot das System zwei Einstiegsmöglichkeiten in das Datenmaterial: Zum einen waren von den Texten aus die Ergebnisse aller bisher analysierten Ebenen zugänglich. Zum anderen erlaubte eine Recherchekompo-

¹⁴ RICHTER / RECHENMACHER / RIEPL 1996, 1998, 2000 und 2002. Bislang bei der Analyse unberücksichtigt sind die bei RICHTER / RECHENMACHER / RIEPL 1999 veröffentlichten Transkriptionen der althebräischen Inschriften sowie der ugaritischen Texte. Das Material liegt in strukturierten Textdateien vor.

¹⁵ Dokumentiert bei ARGENTON 1998.

¹⁶ Vorgestellt bei SPECHT 1995; vgl. dazu auch den Bericht von MÜLLER-HÄRLIN 1995.

¹⁷ Zu Projektbeschreibung, Entwicklung des Basissystems und Dokumentation der Anwendungen MultiLIB und MultiBHT siehe SPECHT / ZIRKEL 1999.

nente eine Volltextsuche sowie eine Suche nach bestimmten Kriterien und auf verschiedenen Ebenen. Auch eine frei formulierbare Datenbankabfrage war über ein SQL-Interface möglich. Darüber hinaus ließ ein Annotationssystem die Eingabe von personalisierten Kommentaren zu. MultiBHT war von 1997 bis 2010 im Einsatz und wurde zuletzt von einer Gruppe von etwa 40 Forschern vor allem zur Einsicht in die syntaktische Transkription (Satzgrenzen) und die Analyseergebnisse sowie der Erstellung von Satzkonkordanzen benutzt.

Eine kleine Arbeitsgruppe¹⁸ um Wolfgang Richter nahm 1999 die manuelle Analyse der Satzebene auf. Aus der Datenbank wurden die Textdaten mit Stellenangabe, reserialisierten Tokens¹⁹ und Intervallen sowie die für die Satzebene relevanten Daten der Morphologie und Morphosyntax satzweise exportiert und in Dateien mit einem definierten Notationsschema²⁰ bereitgestellt, um eine strukturierte Eingabe der Analysen zu ermöglichen. Die Analysen wurden im Sommer 2015 abgeschlossen. Die Daten liegen nun zum Import in die Datenbank bereit.

Auf der Grundlage der in der Datenbank gespeicherten Erstanalyse aller Eigennamen zielt das „Datenbankprojekt ‚Biblisch-Hebräische Personennamen‘“²¹ darauf ab, die morphologischen und syntaktisch-semantischen Daten der Personennamen unter Berücksichtigung des Gesamtkorpus für die Hebraistik, Judaistik, Semitistik, Altorientalistik, die alttestamentlichen sowie die Geschichts- und Religionswissenschaften online zu erschließen. Die Analyseinträge umfassen die Personennamen in ihrer masoretischen und außermasoretischen Gestalt mit deren Transliteration und Transkription. Dazu kommen strukturierte Informationen zu den ver-

¹⁸ Dazu gehörten u. a. an der Julius-Maximilians-Universität Würzburg Theo Seidl, Lehrstuhl für Altes Testament und biblisch-orientalische Sprachen, Hans Rechenmacher, Professur für biblische Einleitung und biblische Hilfswissenschaften, und Monika Berwanger.

¹⁹ Ein Token ist in diesem Zusammenhang eine Zeichenkette zwischen Separatoren. Als Separatoren sind hier definiert die Codes für „Leerzeichen“, „=“ und „Neue Zeile“. Beispiel: Das Wort bzw. die graphische Einheit *ha=malk* besteht aus den Tokens *ha* und *malk*. In der Datenbanktabelle „beleg“ ist jedem Token ein Datensatz zugeordnet. Zum Begriffspaar Token (Vorkommnis) und Type (Vorkommistyp) siehe https://de.wikipedia.org/wiki/Token_und_Type.

²⁰ Das Notationsschema wird anhand eines Beispiels vorgestellt bei RECHENMACHER / VAN DER MERWE 2005, 72-74.

²¹ An dem seit 2014 von der DFG geförderten Projekt wirken mit: Hans Rechenmacher, Professur für biblische Einleitung und biblische Hilfswissenschaften an der Julius-Maximilians-Universität Würzburg, Viktor Golinets, Lehrstuhl für Hebräische Sprachwissenschaft an der Hochschule für Jüdische Studien Heidelberg, und Christian Riepl, IT-Gruppe Geisteswissenschaften der Ludwig-Maximilians-Universität München.

balen und nominalen Bildelementen nach Bauformen, Basen und Wurzeln, ferner syntaktisch-semantische Angaben (Wortverbindungen, Satzarten) und jeweils eine Übersetzungsphrase. Die Beiträge enthalten ferner eine Diskussion von Einzelproblemen, Verweise auf die althebräisch-epigraphische und nordwestsemitische Onomastik sowie Literaturangaben. Die Datenbank wird laufend erweitert.

Zum Datenbestand

Nachdem die Online-Plattform MultiBHT im Jahr 2010 hinsichtlich der Software und Hardware veraltet war, wurde eine Neukonzeption des Systems im Sinne einer kollaborativen Forschungsdatenbank ins Auge gefasst. Diese wird seit 2011 an der IT-Gruppe Geisteswissenschaften der Ludwig-Maximilians-Universität München nach und nach umgesetzt. Ausgehend von der bestehenden Datenbank „bhtdb version 2“ wurde zuerst deren Struktur in der Datenbank „bhtdb version 3“ differenziert, erweitert und optimiert. Die Daten der sprachwissenschaftlichen Analyse werden dabei weiterhin in Anlehnung an das ebenenspezifische Grammatikmodell abgebildet. Die folgende Übersicht verdeutlicht die Zuordnung von methodischen Ebenen und Datenbanktabellen mit Angabe der jeweiligen Einheiten und ihres Umfangs in Datensätzen:

<i>Ebene</i>	<i>Tabelle</i>	<i>Einheit</i>	<i>Datensätze</i>
Wort	beleg	Tokens	489.437
		Types	31.288
	wort	Wörter: BH ^t , BHS	316.054
	anm	Textanmerkungen	11.938
	eigennamen	Eigennamen	12.702
Wortfügung	wv	Wortverbindungen	120.828
	wvvar	Varianten	179.477
Satz	satz	Zeilen	88.537
		Sätze	70.653
	satzanalyse	Satzanalysen	

Auch der Datenbestand wurde in verschiedener Hinsicht ergänzt: So enthält die Tabelle wvvar die Varianten der Wortverbindungen, die durch die Analyse mit AMOS zustande gekommen waren und die bislang in

Dateien aufbewahrt waren. Die Tabelle satzanalyse²² dient der Aufnahme der aktuell in Dateien vorliegenden syntaktischen Analyse des gesamten Korpus. Je Datensatz sind zehn Felder²³ in folgender Struktur erfasst:

Feld	Beschreibung
1	Stelle [Stamm-Basis] (von bis) Tokens mit Angabe der Intervalle
2	Satzart, Satzbauplan, Kernsaterweiterungen / Partikel (Art / Anzahl, ausgedrückt / getilgt)
3	Syntagmen und Position mit Angabe von Tilgungen
4	Syntagmrelationen, semantische Funktionen und Kernseme mit Auflösung von Tilgungen und Pronominalisierungen
5	Tiefenstruktur zu 3 (bei eingebetteten Sätzen)
6	Tiefenstruktur zu 4 (bei eingebetteten Sätzen)
7	Satzhafte, den Kernsatz erweiternde Elemente (Vokativ, Interjektion)
8	Partikel: Subklassen und Position
9	Partikel: Funktion
10	Kommentar

Nach dem vollständigen Datenimport sind weitere Normalisierungsschritte in Struktur und Daten dieser Tabelle sowie eine umfangreiche Konsistenzprüfung des gesamten Datenbestandes erforderlich.

Daneben dokumentieren Codetabellen das gesamte Zeicheninventar, das für die Transkriptionen der BH^t (codetable) sowie für die Transliterationen der BHS²⁴ (codetable_bhs) und griechischer Buchstaben (codetable_lxx) verwendet werden. Diese ordnen zugleich die Zeichen im Beta-code den entsprechenden Zeichen in der Unicode- bzw. der TeX-Kodierung zu. Damit ist eine bequeme Ergänzung um weitere Kodierungssysteme und eine dynamische Konvertierung auch nach dem „on-the-fly-Prinzip“ in beliebige Kodierungssysteme gewährleistet. Andere Tabellen dokumentieren sämtliche Änderungen des transkribierten Textes seit dem

²² Zu Demonstrationszwecken auf der IOSOT 2013 wurde die Analyse der Satzebene des Buches Rut in die Datenbank eingelesen. Die Satzanalysen der übrigen Bücher werden in Kürze folgen.

²³ Zum Notationsschema siehe RECHENMACHER / VAN DER MERWE 2005, 72-74.

²⁴ Die Tabelle wort enthält die Wörter von BHS nach PARUNAK / WHITAKER / GROVES 1987 in einer 1:1 Transliteration der Konsonanten, Vokale und Akzente.

Erscheinen der Erstausgabe (release²⁵), ordnen die alttestamentlichen Bücher und die Quellen von Sirach nach Buchgruppen (buch), oder enthalten den jeweiligen Stand von Klassifizierungen der Wortarten (wortarten), Wortverbindungsarten (wvarnten) und Satzarten (satzarten_vs, satzarten_ns). Datenbank und Daten lassen sich zu jeder Zeit dynamisch verändern oder erweitern und flexibel allen Anforderungen anpassen.

Aus der nun fast dreißigjährigen Projektgeschichte können heute digitale Projekte im Kontext der sich etablierenden Digital Humanities lernen: Den primären Kern und Mittelpunkt eines digitalen Projektes bilden die eindeutig strukturierten Daten. Diese gilt es, nachhaltig verfügbar und zugänglich zu bewahren. Dazu bedarf es einer dauerhaften personellen und technischen Infrastruktur.²⁶ Sekundär und ersetzbar sind alle Programme und Systeme zur Analyse, Recherche und Präsentation. Diese sind jeweils projektspezifisch, beruhen auf bestimmten Theorien und Fragestellungen, sind meist übertragbar auf andere Problemstellungen, somit auch synergiefördernd. Sie sind zudem dokumentiert, in Softwarepaketen konserviert²⁷ und können prinzipiell jederzeit durch Portierung oder Reengineering (auch in anderen Projekten) wiederverwendet werden. Ein entscheidender Punkt aber wird in Zukunft sein: Durch die Herausbildung und den Einsatz von Standards, den rasenden Fortschritt digitaler Technologien und nicht zuletzt das wachsende IT-Wissen, das in den digitalen Geisteswissenschaften ein vertieftes Verständnis um informatische Methoden fördert, reduziert sich der Aufwand für projektspezifische Entwicklungen zunehmend. Solche Aufgaben lassen sich im Rahmen befristeter Projekte auch mit interdisziplinärem Horizont verwirklichen. Somit werden sich Einzelprojekte in den digitalen Geisteswissenschaften immer als Teil eines grö-

²⁵ Die Tabelle enthält bereits die Dokumentation der Änderungen, die sich aus der Analyse der Satzebene ergeben haben und die 2016 als Release 5 freigegeben werden soll; die Änderungen sind noch in den Bestand der Texte und der Datenbank zu übernehmen.

²⁶ Solche Infrastrukturen existieren an der IT-Gruppe Geisteswissenschaften der Ludwig-Maximilians-Universität München. Die zukünftige Entwicklung wird auch die Universitätsbibliothek und das Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften bei der Bildung von Repositorien für Forschungsdaten und Software einbeziehen, um insbesondere auch auf Dauer zitierfähige Versionierungen von Datenbeständen zu ermöglichen. Repositorien wie TextGrid, CLARIN oder DARIAH tragen zusätzlich dazu bei, die Sichtbarkeit eines Projektes zu erhöhen und einen Austausch auf nationaler und internationaler Ebene vorzubereiten.

²⁷ Plattformen wie z.B. GitHub (<http://github.com>) erlauben eine kollaborative Softwareentwicklung und bieten Repositorien zur Speicherung von Softwarepaketen.

ßeren Ganzen verstehen. Einen Beitrag dazu können integrative und interaktive Forschungsinfrastrukturen leisten.

Biblia Hebraica transcripta - Forschungsdatenbank

Basierend auf der Datenbank „bhtdb version 3“ ist nach dem Vorbild von MultiBHT in den letzten Jahren die Online-Plattform „Biblia Hebraica transcripta – Forschungsdatenbank“ entstanden. Auf der IOSOT 2013 konnte ein erster Prototyp²⁸, der unter Verwendung moderner Web- und Datenbanktechnologien den transkribierten Text, die Analyseergebnisse auf Wort- und Wortfügungsebene, experimentell auch der Satzebene online zugänglich machte, vorgestellt werden. Eine seit Sommer 2015 auf der Basis des Contentmanagementsystems WordPress²⁹ weiterentwickelte Version³⁰ beruht wesentlich auf dem Konzept einer kollaborativen Forschungsumgebung. Sie befindet sich derzeit in der fortgeschrittenen Testphase und wird 2016 unter der URL <http://www.bht.gwi.uni-muenchen.de> freigegeben. Der gesamte Datenbestand soll unter geeigneten Lizenzen der Organisation Creative Commons³¹ online und in verschiedenen Formaten zugänglich sein bzw. für die Aufnahme in verschiedene Repositorien transformiert werden.

Die Online-Plattform „Biblia Hebraica transcripta - Forschungsdatenbank“ bietet vorläufig folgende Navigationspunkte: Home, Bücher, Suche,

²⁸ Mein Dank gebührt hier den Herren Ansgar Zeblin, Lukas Sandmeir und Samuel Teuber. Ansgar Zeblin entwickelte die erste lauffähige Version des Prototyps. Lukas Sandmeir programmierte Funktionen, die Dokumentationen, Release Notes und Analysen der Personennamen integrieren ließen. Samuel Teuber fügte schließlich Suchfunktionen sowie eine Benutzer- und Gruppenverwaltung hinzu.

²⁹ Siehe <https://de.wordpress.org/>. Ich danke Frau Susanne Grandmontagne, auf deren Initiative hin das von einer weltweiten Community getragene OpenSource-Produkt WordPress an der IT-Gruppe Geisteswissenschaften erprobt und eingeführt wurde. Es hat sich mittlerweile in zahlreichen Projekten bestens bewährt, u.a. auf Grund der Vielzahl und Vielfalt an vorhandenen Softwaremodulen (sog. Plugins), der modularen Erweiterbarkeit sowie der Möglichkeit, auch externe Datenbanken zu integrieren und Software kollaborativ zu entwickeln.

³⁰ Ich danke Herrn Samuel Teuber, der in genialer Weise alle Funktionen des Prototyps in eine WordPress-Umgebung überführte und PlugIns für die Kommunikation der Online-Plattform mit der Datenbank „bhtdb version 3“ sowie für die Suchfunktionen entwickelte.

³¹ Siehe <http://de.creativecommons.org>. Die genauen Modalitäten der Lizenzierung sind noch zu klären.

Dokumentation, Release Notes, Anmelden und Registrieren. Diese führen jeweils auf bestimmte Seiten, die im folgenden kurz zu erläutern sind³²:

Ein Einstieg in das Datenmaterial kann zunächst über den Navigationspunkt „Bücher“ erfolgen. Nach Auswahl eines Buches und Kapitels gelangt man zum transkribierten Text. Dort sind diejenigen Wörter oder Wortgruppen, die Textanmerkungen aufweisen, farbig hinterlegt. Die Textanmerkungen erscheinen in Form von sog. Tooltips, wenn der Mauszeiger darüber bewegt wird; sie sind auch als Liste kapitelweise einsehbar durch einen Klick auf den Punkt „Textanmerkungen“, der sich jeweils unter der Buch-/Kapitelüberschrift der Texte befindet. Von einem Kapitel eines Buches aus kann man sich horizontal kapitelweise vor und zurück bewegen. Ein vertikaler Einstieg in die analysierte Tiefenstruktur der Ebenen Wort, Wortfügung und Satz ist von den Texten aus durch einen Klick auf ein Token möglich. Daraufhin zeigt die Seite „Wortebene“ die vollständige morphologische Analyse eines Tokens an. Von hier aus kann man sich entweder horizontal von Token zu Token bewegen, wieder zum Text zurückkehren oder auf die Seiten „Wortfügungsebene“ bzw. „Satzebene“ wechseln, die die morphosyntaktische bzw. die syntaktische Analyse anzeigen. Auf allen Seiten sind Navigationspunkte eingebaut, die eine Bewegung in alle Richtungen zulassen.

Der Navigationspunkt „Suche“ führt auf eine Seite, die eine nach Art und Anzahl beliebige Auswahl von Suchkriterien auf Wortebene³³ ermöglicht. Aus einem Dropdownmenü wählt man zunächst ein Suchkriterium wie z.B. „Basis“ aus. Ein Klick in das Suchfenster zeigt dann zum Teil hierarchisch (z.B. bei „Wortart“) die Werte an, die in der Datenbank enthalten sind und nach denen gesucht werden kann. Ein Klick auf den Plus-Button erweitert die Liste der Suchkriterien, ein Klick auf den Minus-Button reduziert sie. Ein Klick auf den suchen-Button führt die Suche aus. Das Ergebnis wird in Form einer tabellarischen Satzkonkordanz geboten. Von dort aus führt ein Klick auf das jeweilige Token auf die Seite der morphologischen Analyse, ein Klick auf die jeweilige Stellenangabe zum genauen Ort des Gesuchten im Text. Die Anzahl der gefundenen Belege wird unten auf der Seite des Suchergebnisses angezeigt.

³² Beschrieben wird hier die Funktionalität zum Entwicklungsstand Dezember 2015.

³³ Von dieser aus sind in der aktuellen Version alle anderen Analyseebenen zugänglich. Eine Erweiterung um Suchfunktionen für die anderen Ebenen samt deren Verknüpfungen ist vorgesehen.

Unter dem Navigationspunkt „Dokumentation“ sind bislang einsehbar: Die Einleitung von RICHTER 1991, 1-11, die Tabellen für die Kodierung der Transkriptionszeichen und die Systematik der Wortarten. Eine Erweiterung dieser Seite um die Erläuterung des Zeicheninventars bei RICHTER 1991, 12-14, der Systematik der Satzarten, einen Überblick über den jeweils aktuellen Datenbestand, eine Dokumentation des Datenbankschemas mit Relationen, Attributen und deren Datentypen sowie eine Beschreibung der Systemarchitektur sind in Vorbereitung. Weitere Ergänzungen sind - auch in kollaborativer Weise - möglich und vorgesehen.

Die Seite „Release Notes“ erschließt die Änderungen in Textkorpus und Datenbestand gegenüber der gedruckten Erstveröffentlichung bei RICHTER 1991-1993.

Die Navigationspunkte „Anmelden“ und „Registrieren“ dienen der Verwaltung von Benutzern. Eine Registrierung ist nur dann erforderlich, wenn ein Benutzer wünscht, sich am System anzumelden, um sich an der wissenschaftlichen Diskussion, der Ergänzung der Dokumentationen oder der Weiterentwicklung der Software zu beteiligen. Nach erfolgter Registrierung erhält der Benutzer einen Account für die Anmeldung. Durch die Administratoren wird er einer Benutzergruppe zugeteilt, die wiederum bestimmte Zugriffsrechte besitzt. Auf einfache Weise können damit verschiedene Arbeitsgruppen³⁴ gebildet werden.

Nach erfolgter Anmeldung steht dem Benutzer auch eine interaktive Kommentarfunktion zur Verfügung, um sowohl Kommentare als auch Antworten zu Kommentaren einzugeben. Dies ist vorläufig an folgenden Stellen vorgesehen: Auf den Seiten „Wortebene“ und „Wortfügungsebene“ können z.B. die morphologische und morphosyntaktische Analyse, auf den Textseiten jeweils durch einen Klick auf den Vers³⁵ z.B. die satzsyntaktische Analyse oder Phänomene der Satzfügungsebene diskutiert werden. So kann die wissenschaftliche Diskussion direkt am Gegenstand erfolgen. Damit Kommentare oder auch alternative Analysevorschlüsse strukturiert eingegeben und ohne zusätzlichen Aufwand weiterverarbeitet werden können, wird es hilfreich sein, standardisierte Eingabemasken vorzubereiten. Alle Beiträge können auf der Online-Plattform je nach Wunsch

³⁴ Vorläufig eingeführt sind die Gruppen `bht_admin`, `bht_expert`, `bht_user`, `bht_guest` und `bht_names`, wobei letztere als Beispiel für eine projektbezogene Arbeitsgruppe im oben erwähnten „Datenbankprojekt ‚Biblisch-Hebräische Personennamen‘“ genannt sei.

³⁵ Bei den Werten des Attributes „vers“ handelt es sich um feste Größen, die sich im Gegensatz zu den Werten des Attributes „satz“ nicht ändern.

und Bedarf nach verschiedenen Themen verlinkt und unter Angabe der Autoren auch veröffentlicht werden.

Schon die nähere Zukunft wird nicht nur hier, sondern in vielen digitalen Projekten zeigen, wie ausbau-, aufnahme- und austauschfähig solche Forschungsumgebungen werden können.³⁶

Potenzial

Am Ende des eingangs erwähnten Berichts resümiert Wolfgang Richter: „Es schmerzt, sehen zu müssen, wieviel noch zu tun bleibt. Nicht nur fehlt der Aufstieg über Satzformen und Satzfügungen zum Text, die gesamte Analyse der Inhaltsseite, sondern auch das Lösen der vielen Fragen, die sich aus der bisherigen Arbeit ergeben haben und sich noch ergeben werden. Es könnte freudig stimmen, wenn das erarbeitete Instrumentarium zu fruchtbarer Weiterarbeit helfen könnte.“³⁷ Heute, nach etwa 30 Jahren Projektgeschichte, lässt sich sagen: Das im Projekt „Biblia Hebraica *transcripta*“ überlieferte digitale Erbe von Wolfgang Richter birgt ein reichhaltiges und vielfältiges Potenzial für viele Forschergenerationen verschiedenster Disziplinen. Der gesamte Datenbestand – ein vollständiges Textkorpus, tokenisiert und annotiert – ist nachhaltig und interdisziplinär zu „fruchtbarer Weiterarbeit“³⁸ nutzbar.

Durch die Analyse der Satzebene, die 2015 abgeschlossen wurde und bei der auch semantische Merkmale sowie die Auflösung von Tilgungen und pronominalen Verweisen berücksichtigt wurden, ist die Analyse der Satzfügungsebene und damit ein regelbasiertes Erkennen von Satzformen vorbereitet. Der Datenbestand ist darüber hinaus im Prinzip offen für eine Neuberechnung aller Ebenen der Grammatik und Semantik, auch mit anderen Theorien und Methoden. Von einem regelbasierten Ansatz aus sind

³⁶ Im Prinzip ist auch eine Einbindung in die akademische Lehre möglich. An der IT-Gruppe Geisteswissenschaften der Ludwig-Maximilians-Universität München soll im Rahmen des Projekts „IT for all – Digitale Datenanalyse in den Geschichts- und Kunstwissenschaften“ eine integrative und kollaborative Online-Plattform zur fachbezogenen Vermittlung von IT-Wissen in der akademischen Lehre entwickelt und erprobt werden; die Online-Plattform umfasst eine interaktive Lehr-/Lernumgebung, eine virtuelle Rechenumgebung, ein kollaboratives Datenrepository und eine Publikationsumgebung. Mit Hilfe einer ähnlichen Plattform versucht das Projekt „Mehrsprach-O-Mat“ im Bereich der Didaktik der Deutschen Sprache und Literatur mehrsprachige Online-Materialien dynamisch und flexibel gestaltbar zugänglich zu machen. Beide Projekte sind aufgeführt unter <http://www.itg.lmu.de/projekte>.

³⁷ Siehe RICHTER 1994a, 4-5.

³⁸ Siehe RICHTER 1994a, 5.

ferner Experimente denkbar, indem eine Regel formuliert und auf den gesamten Datenbestand angewendet werden kann. Konkurrierende Analysen könnten damit auf die Genauigkeit der Regeln und die Tauglichkeit einer Theorie hin überprüft werden.

Damit wäre eine Reihe von Anknüpfungspunkten für die Computerlinguistik genannt. Mit Hilfe regelbasierter Ansätze ließen sich komplexe Regeln der Grammatik und Semantik formulieren, um sie in Analyseprogramme z.B. für Experimente zu überführen.³⁹ Aber auch die Statistik könnte das Datenmaterial mit ihren Methoden modellieren, berechnen und visualisieren, um damit wertvolle Daten für semantische oder formkritische Untersuchungen zu liefern.⁴⁰ Ebenfalls im Bereich der Datenanalyse könnte die Informatik bei der Abbildung komplexer Regeln helfen. Interessant ist hier aktuell ein Versuch, Algorithmen für die intrinsische Plagiatserkennung auf alttestamentliche Texte anzuwenden, um sie für literarkritische Untersuchungen zu nutzen.⁴¹

Die oben beschriebene Online-Plattform lässt sich als eine integrative und kollaborative Forschungsumgebung weiter ausbauen⁴². Für die alt-hebräische Sprach- und Literaturwissenschaft kann sie zugleich Analyse-

³⁹ Dazu würde sich wohl die Programmiersprache „Grammatical Framework“ gut eignen, siehe <http://www.grammaticalframework.org>. Erfahrungen damit liegen am Centrum für Informations- und Sprachverarbeitung (CIS) der Ludwig-Maximilians-Universität München vor.

⁴⁰ Unter der Leitung von Helmut Küchenhoff führt Stefanie Schneider im Rahmen ihrer Bachelorarbeit am Institut für Statistik der Ludwig-Maximilians-Universität München zur Zeit umfangreiche statistische Analysen am morphologischen Datenmaterial der *Biblia Hebraica transcripta* durch; anhand sog. Kollokationsanalysen lässt sich z. B. bei ausgewählten Basen von Handlungs- und Fortbewegungsverben der Grad der Anziehung bestimmter nominaler Basen bestimmen; siehe dazu SCHNEIDER 2016.

⁴¹ Zum Projekt „Intrinsic Plagiarism Detection and Author Analysis“ siehe <https://dbis-informatik.uibk.ac.at/254-0-Intrinsic-Plagiarism-Detection.html>. Die Experimente werden in der Forschergruppe Datenbanken und Informationssysteme (DBIS) des Instituts für Informatik der Universität Innsbruck von Michael Tschugnall und Günther Specht durchgeführt und scheinen insofern erfolgversprechend zu werden, als diese Algorithmen nicht nur literarkritisch relevante Einheiten isolieren, sondern auch literarische Strukturen erkennen lassen (vgl. in diesem Band, S. 415-418).

⁴² Im Sinne des Konzepts der sog. Citizen-Science könnte eine Kollaboration von einer Expertencrowd aus auch auf eine Laiencrowd ausgeweitet werden, um interessierte Laien einzubinden, siehe dazu z.B.: <http://www.citizen-science-germany.de>. Nach dem Konzept des Crowdsourcing wird z.B. auch in den Projekten „play4science“ (siehe dazu: <http://www.play4science.uni-muenchen.de> und insb. <http://artigo.org>) und VerbaAlpina (<http://www.verba-alpina.gwi.uni-muenchen.de>) verfahren.

instrument, Lexikon, Konkordanz, Grammatik und Semantik, auch Publikationsumgebung⁴³ werden.

Eine spannende Herausforderung ist schließlich die Verknüpfung mit Datenbanken anderer geisteswissenschaftlicher Disziplinen, wie z.B. der Kunstgeschichte, der Musikwissenschaft oder der Ägyptologie. Mit solchen fachübergreifenden Datenbankabfragen⁴⁴ könnte z.B. die Rezeption literarischer Motive, Themen und Werke des Alten Testaments in Kunst und Musik untersucht werden. Annotierte Bilddatenbanken der Ägyptologie könnten z.B. helfen, den semantischen Gehalt fester Wendungen genauer zu erfassen.

Gerade im Hinblick auf digitale Projekte muss es nicht schmerzen, wenn vieles „noch zu tun bleibt“⁴⁵. Andere Forschergenerationen werden weitermachen können. Digitale Projekte ermöglichen interdisziplinäre Kooperation. Sie rücken den digitalen Forschungsgegenstand in den Mittelpunkt, eröffnen eine plurale Sicht auf ihn, entfalten eine je eigene Dynamik. Digitale Projekte sind nie abgeschlossen.

⁴³ Die Publikationsplattform „Korpus im Text“ versucht, innovatives Publizieren im Umfeld der Korpuslinguistik zu ermöglichen (siehe <http://www.kit.gwi.uni-muenchen.de/>), indem sie ausgehend von den Bedürfnissen von Promovierenden der Class of Language der Graduate School Language & Literature Munich (ehemals LIPP) eine Verknüpfung von wissenschaftlichen Publikationen mit der Datenbasis in Korpusdatenbanken herstellt.

⁴⁴ An der fach- und fakultätsübergreifend organisierten IT-Gruppe Geisteswissenschaften der Ludwig-Maximilians-Universität München sind in zahlreichen Datenbanken eine Reihe von Datensammlungen der verschiedensten Fachgebiete entstanden. Dazu gehören z.B. Datenbanken mit Bildern, Personen und Werken aus der Ägyptologie (siehe <http://mudira.gwi.uni-muenchen.de/>), Kunstgeschichte (siehe <http://artigo.org>) und Musikwissenschaft (siehe <http://bmlo.de/>). Deren Verknüpfung untereinander wäre technisch schon vorbereitet.

⁴⁵ Siehe RICHTER 1994a, 5.

Literatur

- ARGENTON, Hans, Indexierung und Retrieval von Feature-Bäumen am Beispiel der linguistischen Analyse von Textkorpora, DISDBIS 40, St. Augustin 1998.
- ECKARDT, Walter, Computergestützte Analyse althebräischer Texte, ATSAT 29, St. Ottilien 1987.
- HEUMEL, Peter, Erstellung einer deduktiven Datenbank zur semantischen Analyse der althebräischen Sprache mittels Montague-Grammatik, Diplomarbeit, Technische Universität München, Institut für Informatik, München 1997.
- MÜLLER-HÄRLIN, Franziska, Frage an die Wissenschaft: Was tut eine Bibelhandschrift aus dem Jahr 1009 in einer multimedialen Datenbank?, in: Münchner UNI Magazin Nr. 5, Ausgabe Juli/August, hg. v. der Leitung der Ludwig-Maximilians-Universität München, München 1995, 11-12.
- PARUNAK, H. van Dyke / WHITAKER, Richard E. / GROVES, J. Alan, Michigan-Claremont BHS, Philadelphia 1987.
- RECHENMACHER, Hans / VAN DER MERWE, Christo H.J., The contribution of Wolfgang Richter to current developments in the study of biblical Hebrew, in: Journal of Semitic Studies, Vol. L, No. 1, Manchester 2005, 59-82.
- RICHTER, Wolfgang, Grundlagen einer althebräischen Grammatik. A. Grundfragen einer sprachwissenschaftlichen Grammatik. B. Die Beschreibungsebenen: I. Das Wort (Morphologie), ATSAT 8, St. Ottilien 1978.
- , Grundlagen einer althebräischen Grammatik. B. Die Beschreibungsebenen: II. Die Wortfügung (Morphosyntax), ATSAT 10, St. Ottilien 1979.
- , Grundlagen einer althebräischen Grammatik. B. Die Beschreibungsebenen: III. Der Satz (Satztheorie), ATSAT 13, St. Ottilien 1980.
- , Transliteration und Transkription. Objekt- und metasprachliche Metazeichensysteme zur Wiedergabe hebräischer Texte, ATSAT 19, St. Ottilien 1983.
- , Biblia Hebraica transcripta, ATSAT 33.1-16, St. Ottilien, 1991-1993.
- , Biblia Hebraica transcripta, Release 1-4, BH¹-Software, München 1991, 1994, 1996, 1997.

- , Bericht über zehn Jahre Seminar für ugaritische und hebräische Sprach- und Literaturwissenschaft, unveröffentlichte Broschüre, München 1994a, 1-4.
 - , AMOS and its environment - our experiences, Referat, gehalten am 29.10.1992 in der Universität Bern, in: *Studia Iranica, Mesopotamica et Anatolica* 1, hg. v. GIPPERT, Jost / VAVROUŠEK, Petr, Prag 1994b, 231-235.
 - , Text, Sprache und Computer – oder: Hilft der Rechner auch einem Geisteswissenschaftler?, Vortrag, gehalten am 30.10.1992 in der Universität Bern, in: *Studia Iranica, Mesopotamica et Anatolica* 1, hg. v. GIPPERT, Jost / VAVROUŠEK, Petr, Prag 1994c, 237-245.
- RICHTER, Wolfgang / RECHENMACHER, Hans / RIEPL, Christian, Materialien einer althebräischen Datenbank. Die bibelhebräischen und – aramäischen Eigennamen morphologisch und syntaktisch analysiert. *ATSAT* 47. St. Ottilien 1996.
- , Materialien einer althebräischen Datenbank. Nominalformen, *ATSAT* 51, St. Ottilien 1998.
 - , Althebräische Inschriften transkribiert. *ATSAT* 52, St. Ottilien 1999.
 - , Materialien einer althebräischen Datenbank. Wortfügungen: *ATSAT* 53, St. Ottilien 2000.
 - , Materialien einer althebräischen Datenbank. Verbalformen, in: „Wer darf hinaufsteigen zum Berg Jahwes?“. Beiträge aus Prophetie und Psalmen: FS S. Ö. Steingrimsson, *ATSAT* 72, St. Ottilien 2002.
- RIEPL, Christian, Satz- und Metasatzbezeichnung in BHt. Probleme, Lösungen und Änderungen, in: *RB* 1996a, T.103-4, 561-580.
- , Literatur – Sprache – Computer. Zur rechnergestützten Analyse des Althebräischen demonstriert am Projekt BHt, Vortrag, gehalten am 7.3.1995 an der RWTH Aachen, in: *ZAH*, 9. Band, 1996b, Heft 1, S.24-41.
 - , Wie wird Literatur berechenbar? Ein Modell zur rechnergestützten Analyse althebräischer Texte, Internetveröffentlichung in: *Zeitschrift für Computerphilologie* 1 (1997), hg. v. Volker Deubel / Karl Eibl / Fotis Jannidis, München 1997:
<http://computerphilologie.uni-muenchen.de/jahrbuch/jb1-content.html>,
 Buchveröffentlichung in: *Jahrbuch für Computerphilologie* 1 (1999), Hg. v. Volker Deubel / Karl Eibl / Fotis Jannidis, Paderborn 1999, S.107-134.

- SCHNEIDER, Stefanie, Statistisch-Linguistische Analyse alttestamentlicher Textkorpora, Bachelorarbeit, Ludwig-Maximilians-Universität München, Institut für Statistik, München 2016:
<https://epub.ub.uni-muenchen.de/27646/>
- SPECHT, Günther, Wissensbasierte Analyse althebräischer Morphosyntax. Das Expertensystem Amos: ATSAT 35, St. Ottilien 1990.
- , MultiBHT – Ein multimediales Datenbanksystem zur Sprachanalyse; vorgestellt am 29. März 1995 an der LMU München vor der Fakultät für Altertumskunde und Kulturwissenschaften. Technische Universität München, Institut für Informatik, München 1995.
- SPECHT, Günther / ZIRKEL, Martin, MultiMAP/2. Netzzugang und Netzbetrieb für das multimediale Datenbanksystem MultiMAP: TUM-I9920, Technische Universität München, Institut für Informatik, München 1999.
- TSCHUNKE, Lilli, Computer-Aufnahme von Texten des Alten Testaments und des Alten Orients, Interview mit Wolfgang Richter und Manfred Krebernik, in: Berichte aus der Forschung 82, hg. v. Präsidialkollegium der Ludwig-Maximilians-Universität München, München 1988, 29-32.