

Der Thesaurus indogermanischer Text- und Sprachmaterialien (TITUS) – ein Pionierprojekt der EDV in der Historisch-Vergleichenden Sprachwissenschaft

1 Einleitung

1.1 Computereinsatz in der Historisch-Vergleichenden Sprachwissenschaft

Der Einsatz des Computers in der Historisch-Vergleichenden Sprachwissenschaft hat noch keine lange Geschichte und noch keine breite Akzeptanz im Wissenschaftsbetrieb. Die Gründe sind nicht allein in den Beharrungskräften zu sehen, die der Umstellung von tradierten und bei Generationen von Wissenschaftlern bewährten Arbeitsmethoden auf ein gänzlich neues Instrument widerstreben, sondern auch und vor allem in der Beschaffenheit der bislang verfügbaren Betriebssysteme und Anwenderprogramme. Bei den hier besprochenen Computeranwendungen geht es nicht in erster Linie um den Einsatz des Rechners als Schreibmaschine. Selbstverständlich hat das Desktop-Publishing gerade auch in der Historisch-Vergleichenden Sprachwissenschaft einen hohen Stellenwert, wobei freilich die rechnergestützte Produktion wissenschaftlicher Texte dieses Faches nicht nur Erleichterungen, sondern vielmehr neue, durchaus gravierende Probleme mit sich bringt¹. Entscheidend ist jedoch, daß das Desktop-Publishing nicht eigentlich denjenigen Arbeitsbereich darstellt, in dem der Computer die Ausführung des wissenschaftlichen Auftrages der Historisch-Vergleichenden Sprachwissenschaft unterstützt.

Vornehmliches Ziel einer DV-Strategie in der Historisch-Vergleichenden Sprachwissenschaft muß vielmehr die Bereitstellung des Primärmaterials in elektronischer Form sein. Dies sind, wie unten im einzelnen zu erläutern sein wird, Texte in jeweils verschiedenen Formaten sowie Bilder der Textträger, ferner Analysen der Texte.

Es versteht sich von selbst, daß im Zuge der raschen Entwicklung des World Wide Web und der damit verbundenen Technologien (portable, systemunabhängige Text- und Graphikformate) derartige Anstrengungen stets in einem kommunikativen Rahmen geschehen sollen. Nur so kann wissenschaftliche Diskussion von der Digitalisierung ihres Gegenstandes wirklich profitieren. Hypertextualisierung ist also nicht allein eine – bei jeder wissenschaftlichen Arbeit notwendige – Strukturvorgabe in der geschlossenen Umgebung eines bestimmten Projektes bzw. Wissenschaftsstandortes (Institut, Fachbereich, Fakultät), sondern in der welt-

1 Vgl. dazu Bunz 1997a (in Einzelheiten überholt) und Bunz 1998.

weiten wissenschaftlichen Gemeinschaft. Denn die Grenzen einzelner Institutionen, Disziplinen und wissenschaftstheoretischer Kategorien wie die Trennung von Geistes- und Naturwissenschaften werden im digitalen Netz viel leichter überwunden als im traditionellen Raster der Wissenschaftsorganisation, wo bestenfalls gemeinsame Kolloquien und Tagungen unmittelbare Kontakte zwischen Instituten gleicher Fachrichtung, geschweige denn unterschiedlicher Fakultäten herstellen. Die gedruckte Publikation erübrigt sich damit keinesfalls, nur kann sie, bedingt allein schon durch die zeitliche Verzögerung, niemals die Unmittelbarkeit digitaler Daten erreichen.

Beim derzeitigen Stand der EDV-Technik bringt der Computereinsatz in der Wissenschaft überhaupt und hier speziell in der Historisch-Vergleichenden Sprachwissenschaft demnach nicht allein eine technische Perfektionierung bei der Verwaltung und Aufarbeitung des Primärmaterials mit sich, sondern erzeugt gleichzeitig eine kommunikative Dimension im Wissenschaftsbetrieb, die bislang nicht existierte.

Der vorliegende Beitrag zum Workshop „Geisteswissenschaftliche Hypermedia-Anwendungen“ möchte ein Pionierprojekt in der Historisch-Vergleichenden Sprachwissenschaft vorstellen, welches durchaus auch für andere verwandte Disziplinen motivierend und stimulierend wirken sollte. Dabei spielt die in anderen Fächern normalerweise nicht erreichte Diversifizierung der Historisch-Vergleichenden Sprachwissenschaft eine entscheidende Rolle. Wissenschaftspraktisch betrachtet hat dies den Effekt, daß dieses an den Hochschulen wenig etablierte und heutzutage oft in Frage gestellte Fach in seiner ganzen Tragweite und Leistungsfähigkeit präsentiert wird.

1.2 Die Idee des TITUS-Projektes und ihre Realisierung

Das Projekt des „Thesaurus indogermanischer Text- und Sprachmaterialien“ – im Akronym: TITUS – wurde in informellem Rahmen ins Leben gerufen. Auf der 8. Fachtagung der Indogermanischen Gesellschaft in Leiden, 31.8.-4.9.1987, berieten einige Teilnehmer über die Möglichkeit, die Eingabe für Sprachvergleichung und Rekonstruktion relevanter Texte altindogermanischer Sprachen in den Computer zu koordinieren. So sollte unnötige Mehrarbeit rechtzeitig verhindert werden, die dadurch entstünde, daß an verschiedenen Orten dieselben Texte eingegeben werden. Diese Idee eines Gemeinschaftsprojektes wurde kurze Zeit später in die Tat umgesetzt.

Der erste Titel des Projektes lautete „Thesaurus altindogermanischer Texte auf Datenträgern“. Ein Aufruf zur Mitarbeit erfolgte in einer Anzeige in der Zeitschrift ‚Die Sprache‘, Bd. 32/2, S. 151t, unter der Leitung von Jost Gippert (damals Berlin, jetzt Frankfurt a.M.) begann die Sammlung von Informationen über entsprechende laufende Eingabeprojekte bzw. bereits vorliegende elektronische Texte, und im Laufe der folgenden Jahre entstand eine umfangreiche Textdatenbank.

Ein wichtiger Punkt in der Geschichte des Projektes war die 3. Arbeitstagung für Computereinsatz in der Historisch-Vergleichenden Sprachwissenschaft, Dresden 1994. Dort konstituierte sich die Gruppe von Fachvertretern, die auf dem Gebiet der Computeranwendung

tätig sind, als Arbeitskreis „Historisch-Vergleichende Sprachwissenschaft“ (Leitung: Jost Gippert) der Gesellschaft für linguistische Datenverarbeitung (GLDV), womit diese Dresdener Tagung gleichzeitig die erste Arbeitstagung des GLDV-AK darstellte und die Textdatenbank als Projekt des AK gelten konnte. Bei dieser Gelegenheit wurde der neue Projektname „Thesaurus indogermanischer Text- und Sprachmaterialien – TITUS“ eingeführt, der die Zielsetzung erweitert und präzisiert. Seit Ende der achtziger Jahre hatte sich die Kapazität von Hard- und Software stetig vergrößert, so daß sich auch neue Möglichkeiten eröffneten, elektronische Texte mit entsprechenden Programmen weiterzubearbeiten. Daher die Präzisierung ‚Textmaterialien‘ und der Zusatz ‚Sprachmaterialien‘: Der Thesaurus bot jetzt auch bereits analysiertes Material bzw. elektronische Hilfsmittel zur Analyse².



Abb. 1: TITUS-Datenpool

Ebenfalls seit 1994 ist das Projekt mit einer eigenen Web-Site im Internet präsent: <http://titus.uni-frankfurt.de/>. Neben den Informationen über die Textdatenbank, deren Inhalt aus urheberrechtlichen Gründen selbstverständlich nur zum Teil frei zugänglich gemacht werden kann, finden sich auf den TITUS-Webseiten eine laufend aktualisierte ‚interaktive‘ Fachbibliographie mit Eingabeformular für bibliographische Daten, eine Download-Area mit elektronischen Arbeitsmitteln (Zeichensätze, Datenverwaltungsprogramme) sowie aktuelle Informationen zum Studienfach (Lehrveranstaltungen, Tagungen, Stellen). Daher ist es gerechtfertigt, vom TITUS-Datenpool zu sprechen.

2 Vgl. den Statusbericht zum Anlaß der Konstitution des GLDV-Arbeitskreises: Gippert 1995a. Eine gut illustrierte Projektpräsentation ist ferner Gippert 1995b.

Im Herbst 1997 konnte TITUS mit der „6th International Conference about the Use of Computers in Historical and Comparative Linguistics“ (Frankfurt a.M., 21.-24. Oktober 1997) sein zehnjähriges Bestehen feiern (vgl. <http://titus.uni-frankfurt.de/curric/frankf-97b.html>). Im Bereich der Textdatenbank war das Ziel des Projektes schon fast erreicht: Im Februar 1998 wurde dann die erste CD-ROM gebrannt, die alle für die Vergleichende Indogermanische Sprachwissenschaft relevanten Texte im WordCruncher-Format (s. u. S.17ff.) enthält.

Derzeit ist das Institut für Vergleichende Sprachwissenschaft an der Johann Wolfgang Goethe-Universität Frankfurt a.M. Träger des Thesaurus, in Kooperation mit dem Ústav starého předního východu an der Karls-Universität Prag, dem Institut for Almen og Anvendt Sprogvidenskab der Universität Kopenhagen und dem Departamento de Filología Griega y Lingüística Indoeuropea de la Universidad Complutense de Madrid. TITUS bemüht sich um weitere Kooperationen.

2 Historisch-Vergleichende Sprachwissenschaft und Vergleichende Indogermanische Sprachwissenschaft

2.1 Was ist Historisch-Vergleichende Sprachwissenschaft?

Historisch-Vergleichende Sprachwissenschaft kann ausschließlich in einem interdisziplinären Rahmen existieren. Der ‚Input‘, die historischen Textdokumente, durchläuft einen Filter, der je nach Herkunft und Schrift bzw. Sprache des Textes sowie nach Art des Textträgers aufgebaut ist (vgl. hierzu und zum folgenden das Schaubild auf S. 9): Bestandteile des Filters sind linguistische, philologische, historische und archäologische Disziplinen. Ist die Sprache eines Textdokumentes bekannt, beanspruchen für die Bearbeitung von Form und Inhalt des Textes zunächst diejenigen Disziplinen ihre Zuständigkeit, die auf die Erforschung des betreffenden Sprachbereiches spezialisiert sind. Nationale und ethnische Interessen kommen hier zum Tragen („Nationalphilologien“ und „-archäologien“) ebenso wie international etablierte Traditionen in der Wissenschaftspraxis (z.B. die Festlegung der Klassischen Philologie/Classical Studies auf die Beschäftigung mit der griechisch-römischen Antike). Andere Spezialfächer definieren ihr Untersuchungsobjekt über das beschriftete Material: Die Epigraphik bearbeitet Inschriften auf Stein oder Metall, die Papyrologie Texte auf Papyrus, etc. Die Paläographie hat, unabhängig vom Beschreibstoff, die synchrone und diachrone Klassifikation der Schriftzeichenformen zum Gegenstand. Spezialisierungen hinsichtlich bestimmter Schriften und/oder Sprachen können in den letztgenannten Disziplinen zu einer weiteren Auffächerung führen, ohne daß sich die betreffenden Bereiche ausgliedern würden. Daß alle Fächer des ‚Filters‘ per definitionem untereinander vernetzt sind, sowohl über ihre Forschungsgegenstände als auch über die angewandten Methoden, braucht nicht eigens betont zu werden.

Die spezifisch linguistisch-komparatistische Arbeit der Historisch-Vergleichenden Sprachwissenschaft beginnt nach der Aufbereitung der Textdokumente, d.h. der möglichst genauen Beschreibung des historischen Kontextes, welchem das Monument (Text und Textträger)

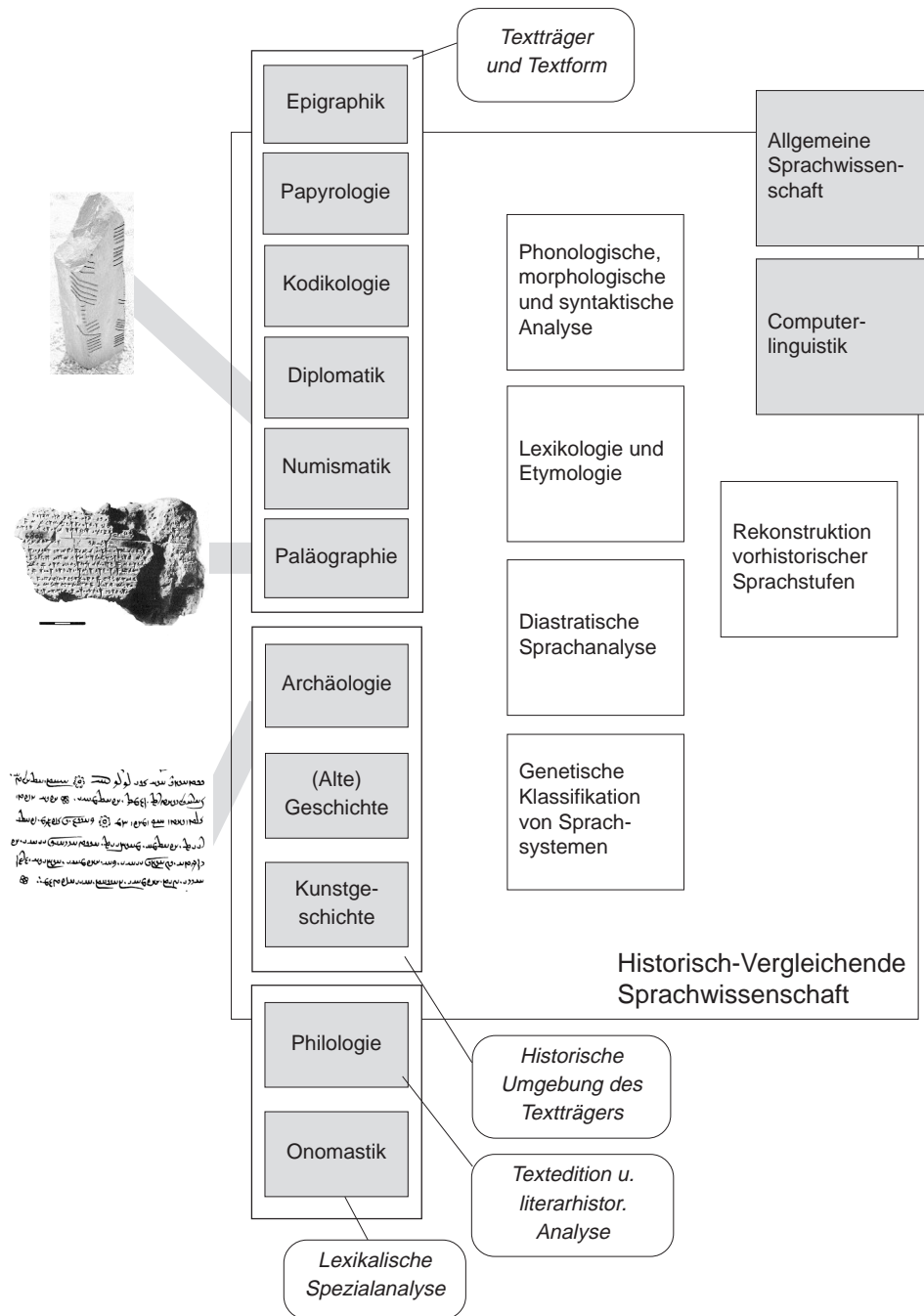


Abb. 2: Stellung der Historisch-Vergleichenden Sprachwissenschaft

entstammt, sowie der Schrift des Textes. Demnach gehört die Entzifferung bislang unbekannter Schriften noch ins Vorfeld historisch-vergleichender Analyse und sollte auch methodisch vom Einfluß bereits erzielter Analyseergebnisse möglichst frei bleiben, da nur der Nachweis einer internen Funktionalität die Entzifferung eines Schriftsystems sichern kann.

Die Inhalte und Ziele des historischen Sprachvergleichs lassen sich durchaus unabhängig von bestimmten Sprachen beschreiben. Die phonologische, morphologische und syntaktische Analyse der überlieferten Texte erlaubt es, je nach Art und Umfang des Materials, die sprachlichen Phänomene verschiedener Texte bestimmten Phasen der Sprachentwicklung zuzuweisen. Die somit erzielte Periodisierung ermöglicht ihrerseits eine diachrone Untersuchung des Wortschatzes – die Bedeutungsgeschichte der Lexeme ist seit jeher ein Arbeitsfeld der Historisch-Vergleichenden Sprachwissenschaft. Über die Bearbeitung einzelnsprachlichen Materials hinaus führt dann die Vergleichende Methode, welche es erlaubt, durch die Analyse der Übereinstimmungen und Unterschiede der lautlichen und morphologischen Gestalt einen voreinzelsprachlichen Zustand zu rekonstruieren, wobei die Präzision der Formulierung der konstatierten Regeln des spontanen sowie des kombinatorischen Lautwandels den Aussagewert eines Rekonstruktives ausmacht. Mit der Rückführung lautlicher und morphologischer Einheiten verschiedener lebender und/oder historischer Sprachen auf eine gemeinsame Vorstufe kann die genetische Verwandtschaft dieser Sprachen beschrieben und somit erwiesen werden. Im voreinzelsprachlichen Bereich läßt sich dann ebenfalls eine chronologische Schichtung und genetische Abhängigkeit einzelner Phasen feststellen. Die Rekonstruktion führt schließlich hinauf zu einem (nicht notwendigerweise homogenen) Zustand, welcher der ‚Ursprache‘ der betreffenden Sprachfamilie zugeordnet wird. Damit ist die Grenze linguistischer Rekonstruktion erreicht, zumal der Bezug zur Materialbasis, d.h. zu den jeweils ältesten Belegen einer schriftlichen Tradition, stets erhalten bleiben muß.

Die Allgemeine Sprachwissenschaft stellt der Historisch-Vergleichenden Sprachwissenschaft ihre durch immer neue Betrachtungsmodelle gewonnenen Erkenntnisse über strukturelle Phänomene auf allen Ebenen der Hierarchie sprachlicher Zeichen zur Verfügung. Die Beurteilung bestimmter Strukturen in historischen und vorhistorischen Stadien profitiert von diesen allgemein-linguistischen Ansätzen. Umgekehrt haben Ergebnisse historischer Sprachforschung immer auch Einfluß auf eine synchrone Betrachtung, da bei der Analyse eines Sprachsystems Synchronie grundsätzlich als abstrakte, adynamische Schnittebene zu gelten hat, während Sprachwandel jedem System inhärent ist.

Die auf dem geschilderten Wege erzielte genetische Klassifikation der Sprachen der Erde bringt es mit sich, daß sich eine Reihe Historisch-Vergleichender Sprachwissenschaften institutionalisiert haben: Altaistik, Amerindianistik, Finnougristik, Indogermanistik, Kaukasiistik, Semitistik, Turkologie, Sino-Tibetistik etc., wobei natürlich sämtliche Disziplinen, die Sprach-Subfamilien bearbeiten, ebenso als Historisch-Vergleichende Sprachwissenschaften betrachtet werden müssen (Romanistik, Slavistik, Iranistik, Kartvelologie u.v.a.m.).

Zu beachten ist, daß der im Rahmen der Historisch-Vergleichenden Sprachwissenschaft praktizierte typologische Sprachvergleich über eine Kontrastive Linguistik, wie sie in der Allgemeinen Sprachwissenschaft betrieben wird, hinausgeht, indem die Historisch-Verglei-

chende Sprachwissenschaft in der Lage ist, die beträchtlichen Differenzen zwischen genetischer und typologischer Affinität von Sprachen zu untersuchen.

Ihren umfassenden Charakter erhält die Historisch-Vergleichende Sprachwissenschaft dadurch, daß sie in den verschiedenen Phasen der Bearbeitung historischer Textdokumente immer dann einspringt, wenn keine Spezialdisziplin für eine bestimmte Aufgabe institutionalisiert ist, auch im Vorfeld linguistischer Analyse. Natürlich können sich aus einer solchen Situation heraus jederzeit gesonderte Wissenschaftszweige etablieren, wie z.B. die Epigraphik ägäischer Schriften des 2. Jahrtausends v.Chr., die Paläographie zentralasiatischer Brāhmī-Schriften o.ä. In der Wissenschaftsorganisation und -praxis nimmt die Historisch-Vergleichende Sprachwissenschaft demnach die Funktion eines Netzbauers und -verwalters wahr, da sie ja durch ihren Forschungsgegenstand selbst eine Vernetzung der beteiligten Disziplinen definiert und etabliert. Es liegt auf der Hand, daß ein solches Fach mit naturgegebener interdisziplinärer Arbeitsweise bei der elektronischen Vernetzung unverzichtbar ist.

2. 2 *Indogermanische Sprachen und ihre älteste Textüberlieferung*

Unter den Historisch-Vergleichenden Sprachwissenschaften nimmt die Indogermanistik eine besondere Stellung ein, da sie eine Sprachfamilie untersucht, die sehr alte (Indoarisch, Hethitisch und verwandte altanatolische Sprachen, Griechisch: seit dem 2. Jahrtausend v. Chr.) und kontinuierliche (Indoarisch und Griechisch über drei, Iranisch und Romanisch über zweieinhalb Jahrtausende) Sprachtraditionen aufweist. Eine vergleichbare Konstellation, verbunden mit einem außerordentlichen Reichtum des überlieferten Textmaterials hinsichtlich diatopischer Schichtung und Textsorte findet sich bei keiner anderen Sprachfamilie der Erde. An dieser Stelle ist nicht der nötige Raum, um die Ausdehnung der indogermanischen Sprachfamilie zu illustrieren. Nur eine Serie von vier oder fünf Karten Eurasiens, wobei jeweils eine Zeitspanne von nicht mehr als tausend Jahren dargestellt ist, könnte ein angemessenes Bild vermitteln.

In Abb. 3 habe ich die ältesten Textcorpora altindogermanischer Sprachen benannt und das Einsetzen der Überlieferung, d.h. das Alter der frühesten erhaltenen Texte angegeben.

3 **TITUS: Textdatenbank und Datenpool als hypermediales System**

3. 1 *Wie können historische Textdokumente elektronisch erfaßt werden im Hinblick auf eine effiziente rechnergestützte linguistische Analyse sowie eine Hypertextualisierung des Materials?*

Im folgenden sollen die einzelnen Arbeitsschritte dargestellt werden, die notwendig sind, um historische Textmaterialien in digitalisierter Form verfügbar zu machen, gleichzeitig eindeutig zu codieren bzw. für eine solche Codierung vorzubereiten und schließlich das gesamte Material einschließlich der Analysen in Hypertext einzubinden. Abb. 4 visualisiert den Ablauf.

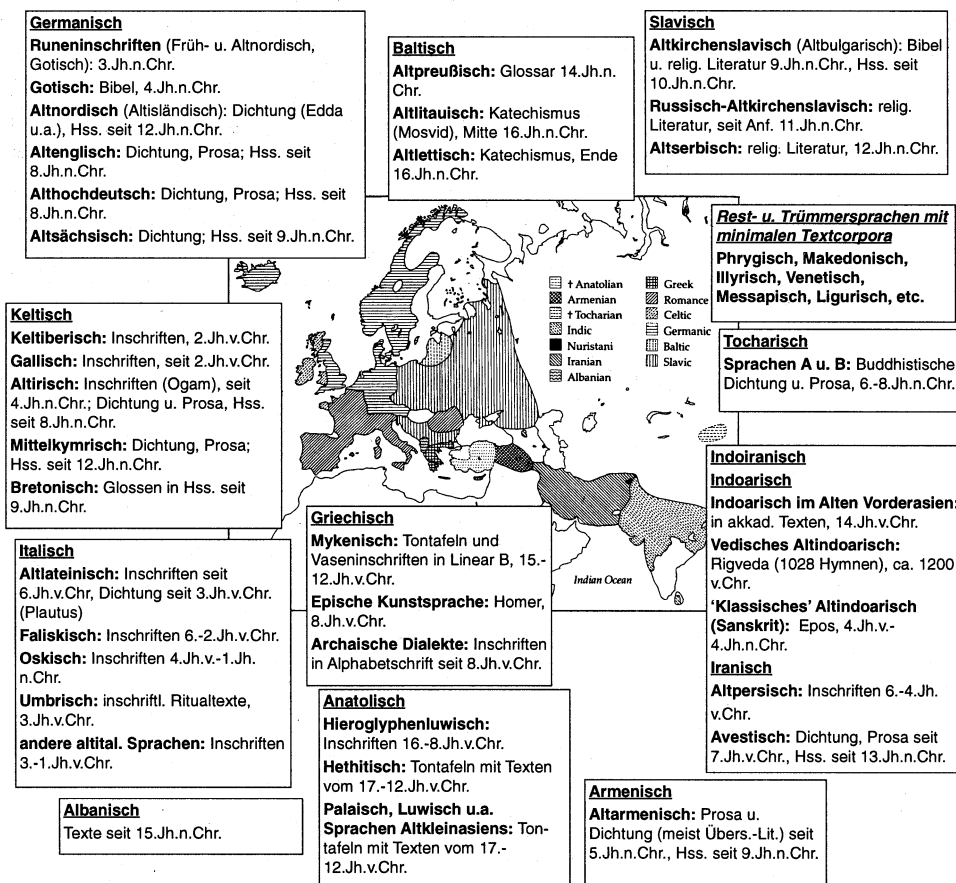


Abb. 3: Die indogermanischen Sprachen und ihre ältesten Textüberlieferungen (Karte aus: Merritt Ruhlen, *A Guide to the World's Languages. Vol. 1: Classification*, Stanford 1987, 36)

(a) Bild des originalen Textträgers (soweit vorhanden):

Bei Inschriften auf Stein, Holz, Ton sollte der Textträger zunächst fotografiert werden. Dabei geht es nicht allein um ein Bild des beschrifteten Bereiches, sondern vielmehr um eine archäologische Bestandsaufnahme, da die Lagerung und der Zustand des Textträgers als ganzem für die Deutung der Inschrift von entscheidendem Wert sein kann (Entzifferung, Inhalt des Textes). Eine Fotoserie mit Ansichten von mehreren Seiten oder eine Videosequenz (in TITUS noch nicht verwendet) genügt für eine Dokumentation; mit Rücksicht auf die spätere Digitalisierung ist ein hochauflösender Farbdiafilm zu empfehlen. In bestimmten Fällen kann auch schon eine Digitalkamera eingesetzt werden³, meist ist jedoch die Auflösung zu gering.

Manuskripte aus verschiedenen Materialien (Papier, Papyrus, Pergament etc.) werden ebenfalls fotografiert. Bei beschädigten Blättern ist darauf zu achten, daß ein günstiger Un-

3 Vgl. Vavroušek 1998.

tergrund gewählt wird, damit sich an beschädigten Stellen die durchscheinende Unterlage von der Farbe, gegebenenfalls auch von der Struktur tieferer Materialschichten (bei Fissuren, Rasuren etc.) des Beschreibstoffes ausreichend abhebt.

(b) Digitalisierung der Bilder des originalen Textträgers:

Die Diapositive werden mit Hilfe eines Diascanners digitalisiert. Vorzugsweise werden zwei oder mehr Scans in verschiedener Auflösung angefertigt, je nach Beschaffenheit des Objektes. Entscheidend ist jedoch die Handhabbarkeit der Bilddatei unter den heutigen technischen Voraussetzungen⁴. Die rasch voranschreitende Entwicklung der Speichermedien wird es in absehbarer Zeit möglich machen, auch umfangreiche Bilddateien ohne Schwierigkeiten zu verwalten, zumal sie niemals den Umfang bewegter Bilder erreichen.

Selbstverständlich gehören zu einer Bilddokumentation des Textträgers auch Fotografien, Abklatsche, Pausen und Abzeichnungen, die im Rahmen bisheriger wissenschaftlicher Arbeit angefertigt wurden. Sie werden ebenfalls eingescannt.

(c) Bearbeitung der digitalisierten Bilder:

Durch den Einsatz handelsüblicher Bildbearbeitungsprogramme lassen sich die Bilddateien weiterbearbeiten mit dem Ziel, die Beschriftung zu untersuchen. Änderungen der Einstellungen hinsichtlich Kontrast, Ausleuchtung, Farbe, Sättigung, die Anwendung unterschiedlicher Filter (z.B. Weich- und Scharfzeichner) machen undeutliche Konturen der Schriftzeichen besser sichtbar und können an beschädigten Stellen über das Ausmaß des Schadens und über vorhandene Spuren Aufschluß geben. Hier vollzieht sich die paläographische Arbeit im digitalen Zeitalter⁵. Teile der Dateien werden ausgeschnitten, separat gespeichert und mit anderen Ausschnitten mit verschiedenen Techniken verglichen (Überblendung, Merging etc.). Mit Hilfe eines Graphiktablets und eines Stifts mit Induktionsspule als Eingabegeräte bringt der Paläograph im digitalisierten Bild Markierungen an, gegebenenfalls in unterschiedlichen Farben, um eine bestimmte Schriftzeichenform bzw. deren Ergänzung anzuzeigen.

An dieser Stelle teilt sich die Bearbeitung der Textmaterialien in einen sprachwissenschaftlichen und einen philologischen Pfad. Beide Bearbeitungswege sind jedoch auf einen ständigen gegenseitigen Austausch angewiesen.

(d) Entzifferung, Analyse und Codierung der Schrift:

Im Bereich des TITUS-Projektes befinden sich durchaus noch Schriftsysteme, die bislang noch nicht (z.B. Linear A) oder unvollständig bzw. nicht befriedigend (z.B. Keltiberisch) entziffert wurden. Was jedoch erst für wenige historische Schriftsysteme geleistet wurde, ist eine Codierung der verwendeten Zeicheninventare, die eine eindeutige Zuweisung von Codepositionen in einer digitalen Tabelle ermöglicht. Voraussetzung dafür ist die Standardisierung des Zeicheninventars sowie ein Codierungsraum ausreichender Größe. Letztere Bedingung war bis vor wenigen Jahren bei kommerziellen Betriebssystemen nicht gegeben,

4 Zu den technischen Einzelheiten vgl. Jost Gippert, <http://titus.uni-frankfurt.de/texte/tocharic/index.html> (mit Bildbeispielen tocharischer Handschriften) sowie Gippert 1998a.

5 Vgl. Gippert 1997a.

Schriftzeichen wurden mit 7 oder maximal 8 Bit verwaltet, wodurch 128 bzw. 256 Codepositionen zur Verfügung stehen. Auswege aus dieser Beschränkung waren die Definition von ASCII-Zeichensequenzen für jeweils einzelne Zeichen, auf kommerzieller Ebene programminterne 16-Bit-Verwaltungen, länderspezifische Double-Byte-Codierungen und – zum Leidwesen der Textcodierung – Formatierungen mit Hilfe verschiedener Fonts, wobei beliebige Zeichen auf dieselbe Codeposition projiziert wurden. Zur Erfassung historischer Schriften kamen demnach nur spezifische Lösungen in Frage, auf keinen Fall jedoch konnten systemunabhängige Codierungsstandards entwickelt werden. In jüngster Zeit etabliert sich ISO 10646/Unicode® als internationale 16-Bit-Norm. Das TITUS-Projekt hält Kontakt zum Unicode Consortium und wurde aufgefordert, bei der Aufnahme historischer Schriften in die Codierung bzw. beim Ausbau der Codierung lebender Schriften mitzuwirken⁶.

(e) Ein wichtiges Hilfsmittel für die Schriftsystemanalyse ist die Transliteration, d.h. die Übertragung der graphischen Einheiten der Originalschrift in andere, neutrale Symbole (z.B. Ziffern), ohne Rücksicht auf den funktionalen Wert der Einheiten (graphematisch, phonologisch oder morphologisch). Die im Falle entzifferter Schriften in der Wissenschaftspraxis verwendeten Lateinschrift-basierten Transliterationssysteme implizieren bereits eine lautliche Interpretation. Für eine elektronische Speicherung von Textdaten in Transliteration werden wiederum Codierungen benötigt, die mindestens 16-Bit-weite digitale Zeichen voraussetzen, wenn eine parallele Verarbeitung verschiedener Transliterations- und Originalschriftsysteme möglich sein soll.

(f) Nach der Schriftsystemanalyse kann ein Text ediert werden, selbst ohne eine vollständige Entzifferung und lautliche Deutung der Schriftzeichen. Die in der Klassischen Philologie entwickelte Methode der Kritischen Edition durch Kollation, d.h. Vergleich der vorhandenen Exemplare des Textes läßt sich freilich umso effizienter anwenden, je genauer Schriftsystem und Sprache verstanden werden. Im Zuge der elektronischen Verfügbarkeit der Textzeugen in einem Hypertext wird man bei der Kritischen Edition ein apodikthisches Verfahren, wie es in der humanistischen Tradition oft angewandt wurde, vermeiden: Anhand eines ausführlichen Variantenapparates, der mit den digitalisierten Bildern der Textträger „verlinkt“ ist, kann der Benutzer einer hypermedialen Textedition die Kollation in vollem Umfang nachvollziehen und die Entscheidung des Herausgebers für eine bestimmte Lesart aufgrund des gesamten Materials beurteilen.

(g) Die phonologische, morphologische und syntaktische Analyse historischer Sprachstufen durch die Historisch-Vergleichende Sprachwissenschaft geschieht aufgrund des Textmaterials, das durch die Schriftsystemanalyse zugänglich gemacht wurde, und erschließt wiederum den Inhalt des Textes für die philologische, d.h. literarhistorische, historische etc. Bearbeitung.

6 Die bisherigen Aktivitäten des TITUS-Projektes in diesem Bereich sind dokumentiert in Bunz/Gippert 1997 und Bunz 1997b. Zum aktuellen Stand der Bemühungen und zur Problematik Unicode-konformer Codierung historischen und ‚linguistischen‘ Sprachmaterials überhaupt s. Bunz 1998. Zu Unicode allgemein s. <http://www.unicode.org>.

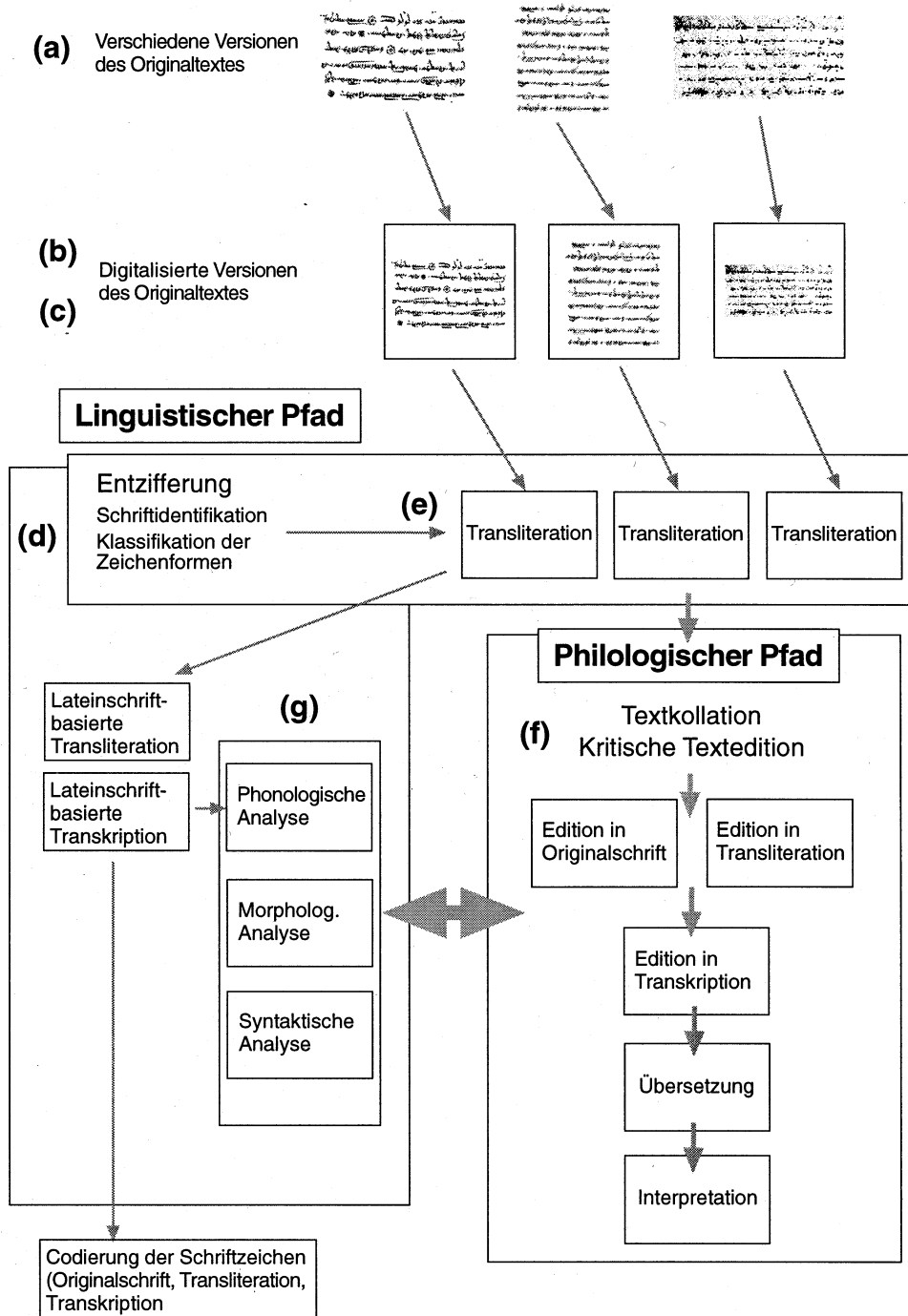


Abb. 4: Aufbereitung und Analyse von Texten in der Historisch-Vergleichenden Sprachwissenschaft

Ein linguistisches Werkzeug zur Segmentierung und zum Vergleich sprachlicher Zeichen ist die Transkription, d.h. die Wiedergabe der originalschriftlichen Notation nach bzw. aufgrund einer graphematischen *und* phonologisch-morphologischen Analyse der Sprachdaten. Die lateinschriftliche Transkription dominiert in der internationalen Wissenschaftspraxis (Lateinbuchstaben mit diakritischen Zeichen). Die Codierungsproblematik tritt hier ebenfalls in ihrer ganzen Komplexität auf.

Das TITUS-Projekt bietet die Text- und Sprachmaterialien in Umschrift (Transliteration oder Transkription je nach wissenschaftlicher Konvention in den betreffenden Gebieten). Die Texte sind mit 8 und 16 Bit codiert: Im erweiterten ASCII-Format beziehen sich die Codezuweisungen auf einen bestimmten Zeichensatz, die exakte Zuweisung kann unter jedem beliebigen System durch Konvertierung wiederhergestellt werden. Das Format WordPerfect 5.1 (DOS) nutzt die programminterne 16-Bit-Verwaltung, das WordCruncher-Format (8 Bit) enthält die notwendigen Markierungen, so daß der Text als Indexdatei im WordCruncher-Viewer verwendet werden kann (s. u. S.17ff.).

In einer Hypertext-Umgebung ist die weitere analytische Aufbereitung der historischen Texte gefordert. Morphologische Markierungen (Tagging) im Zusammenhang mit der syntaktischen Struktur können mit einer hypertextuellen Grammatik der betreffenden Sprache verbunden werden, ebenso besteht die Möglichkeit, elektronische Wörterbücher mit einzu- beziehen. Für die historisch-linguistische Forschung von herausragender Bedeutung ist die „Verlinkung“ der genetisch verwandten Textmaterialien untereinander und dies nicht nur auf Wortebene (etwa über ein elektronisches etymologisches Wörterbuch), sondern auch auf der Ebene der Syntagmen. Da die überwiegende Mehrzahl der ältesten Textcorpora indogermanischer Sprachen poetische Texte enthalten, läßt eine rechnergestützte poetologisch-phraseologische Analyse wichtige Erkenntnisse erwarten, welche wiederum unabdingbare Voraussetzung sind für Untersuchungen zur Entwicklung der Syntax⁷.

Das TITUS-Projekt hat in all diesen Bereichen bereits Ergebnisse vorzuweisen; einige davon sollen nun kurz präsentiert werden.

3.2 *Internet-Edition von epigraphischem Material: Der Ogam-Server*

Ogam ist eine frühe keltische Schriftentwicklung. Aufgrund historisch-linguistischer Kriterien läßt sich als *Terminus ante quem* der Erfindung das Ende 8. Jahrhunderts n. Chr. bestimmen; einige Belege sind möglicherweise auf das Ende des 3. Jahrhunderts datierbar, jedoch ist dies nicht gesichert. Die Ogam-Schrift ist auch in Handschriften überliefert: Wir besitzen Ogam-Dokumentationen irischer Mönche. Die Herkunft der Zeichen konnte bisher nicht endgültig geklärt werden.

⁷ S. u. S. 22. Ferner hat das ebenfalls in Frankfurt durchgeführte Projekt AUREA (Avesta und Rigveda: Elektronische Analyse) die elektronische Abbildung lautgesetzlicher Entwicklungen zum Gegenstand. Es wird im April 1998 abgeschlossen sein, die Publikation der Ergebnisse ist im Laufe des Jahres zu erwarten.

TITUS Ogam-Server

<http://titus.uni-frankfurt.de/ogam>

Ogam inscriptions: Single inscriptions
2nd edition: May 1997
Inscriptions documented so far (30/04/97)



Standardisierung der Ogam-Schrift
Irish Standard 434:1997
Codierungsvorschlag der ISO/IEC für ISO/IEC 10646/Unicode

	365 A	169 B
0		
1	T	
2	II	
3	III	
4	IIII	
5	IIII	Y
6	I	◊
7	II	◻
8	III	✕
9	IIII	■
A	IIII	—
B	f	>
C	#	—
D	##	<
E	###	—
F	####	—



Thesaurus Indogermanischer Text- und Sprachmaterialien
Ogam-Inschrift: CIIC-Nr. 104
Ogam Inscription: CIIC no. 104

Original location: Coolineagh
County: Cork
Surroundings: Churchyard
Year of discovery: 1817
Actual location: - (Aghabullogue)

Illustrations:

Actual reading:
Latin Transcription: [C] [RR]E[A M(A)Q(V)] [D(D)] [M]E[ATT]
Ogam Transcription: [IIII] [###] [(-)IIII] [II] [Y] [X] [IIII]
Ogam Transliteration: [IIII] [####] [#####] [(-)IIII] [II] [(-)] [Y] [X] [IIII]
Interpretation: *[ANM] CORR[E]A MAQ V[VE]D[D] [PI.] ME[ATT]
Direction of reading: dn

Other readings, history, comments etc.:

Location and history:
The monument stands inside the churchyard of the ruined (protestant) parish church of Aghabullogue (the townland is called "Coolineagh" by Macalister in [191a] and [191c]). It is remarkable because of the stone attached to its top, according to Brash, OIM 109, this had been called *Cogean* (or *Cogean*: 130) *olan* "from time immemorial", while the monument itself was called "St. Olan's stone" (130). According to Ferguson, OI 92 (144), it was "regarded as St. Olan's pillar stone" (the church bears the name of that saint, "the old cap-stone has disappeared, and a modern one has been recently substituted. The belief regarding it is, that to whatever distance it may be removed it will be found next day in its old place" (similarly [Brash, OIM 130 sq. according to whom it had "the gift of locomotion"; Brash describes some superstitious practices concerning the monument). From the records it is not clear when the cap stone was cemented in the way it is today; in Brash's time "the present representative of the cogean" was "a coarse unshapely lump".
Another stone from the St. Olan's church is no. [105] which was removed [when?] to the Cork U.I.C. The stone from "Mount Rivers" (no. 1123) was only secondarily placed beside St. Olan's well which is not more than 5 m. from Aghabullogue. According to Brash (109), "St. Olan does not appear in any of our calendars, his name is not on the stone".
The entire length of the monument is according to Brash (130), 6' 9". When visited in 1981, its lower

Reading Gippert (1981):

Angle up:

[C] [RR]E[A M(A)Q(V)] [D(D)] [M]E[ATT]
[IIII] [###] [(-)IIII] [II] [Y] [X] [IIII]
[IIII] [####] [#####] [(-)IIII] [II] [(-)] [Y] [X] [IIII]
The beginning of the inscription was below ground and not visible. There are hardly any traces of letters between 2M and 2m, except a slanted stroke on the B-surface immediately below the M.

Copyright: Jan Gippert, Frankfurt a.M 1996. No parts of this document may be republished in any form without prior permission by the copyright holder.



Textausgabe in HTML-Format

Abb. 5: TITUS-Ogam-Server

Die Textdokumente in Ogam-Schrift haben trotz ihrer Kürze und stereotypen Form (Grab- und Stifterinschriften, ca. 450 bekannt) große Bedeutung für die keltische Sprach- und Kulturgeschichte. Die Keltistik gewinnt aus diesem Material Aufschluß über eine frühe Form des Inselkeltischen (Frühirisch) und ist dadurch in der Lage, die Wirkung charakteristischer Lautgesetze zu verfolgen. Eminenten kulturgeschichtliches Interesse beanspruchen die in den Inschriften überlieferten Namen.

Die Codierung der Ogam-Schrift steht kurz vor der Standardisierung. Die Work-Group 2 der zuständigen ISO-Kommission (ISO/IEC JTC1/SC2/WG2) hat dem aus Irland stammenden und auf dem WWW zur Diskussion gestellten Codierungsvorschlag zugestimmt, das Unicode-Consortium muß noch über die Aufnahme in die Norm entscheiden⁸. Das TITUS-Projekt hofft, bis dahin noch Korrekturen einbringen zu können.

Jost Gippert gibt mit dem Ogam-Server ein mustergültiges Beispiel dafür, wie epigraphisches Material in einer hypermedialen Umgebung im Internet ediert werden kann. Abb.5 gibt eine Übersicht über die Anlage des Servers. Der Benutzer der Edition vollzieht die oben unter 3.1 dargestellten Bearbeitungsschritte nach: Aus einer „Picture Gallery“ (sog. Thumbnail Images, Kopien der Bilder in Miniatur) wählt er durch Mausclick einen Stein aus und gelangt auf die Editionsseite der betreffenden Inschrift. Dort befinden sich wiederum Thumbnail Images sämtlicher Fotografien, aber auch anderer Abbildungen (Abzeichnungen) des Objektes, die jeweils mit Bilddateien hoher Auflösung verlinkt sind. Nachbearbeitete, d. h. mit Markierungen versehene Bilder dienen der epigraphischen Argumentation: Die nach Auffassung des Herausgebers sichtbaren Spuren von Schriftzeichen sind farbig gekennzeichnet. Auf der Editionsseite folgen Transliteration und Transkription in Ogam-Schrift (imitierende Wiedergabe in Originalschrift und Wiedergabe mit standardisierten Zeichen) sowie Transkription in Lateinschrift. Eine ausführliche Dokumentation der bisherigen Lesungen, Kommentare und Bearbeitungen der Inschrift schließt sich an, am Ende steht die Edition des Herausgebers selbst, wobei er seine Argumente mit digitalen Bildern und Markierungen in diesen Bildern stützt. Bibliographische Verweise auf der gesamten Editionsseite sind mit der TITUS-Bibliographie verlinkt.

3.3 *Hypertext in WordCruncher I: Komplexe Textedition des Avesta*

Im Bereich Textretrieval arbeitet das TITUS-Projekt schon seit Jahren mit einer integrierten Software-Lösung, die auf einer Vorindizierung der Texte beruht: Es handelt sich um das Programm WordCruncher⁹. Die 1996 erschienene Windows-Version nutzt die Möglichkeiten des Betriebssystems erschöpfend aus. Im System installierte TrueType-Fonts können bei der Vorindizierung in der Text Style-Definition bestimmten Sprachen zugewiesen werden. Durch dieses Font-Mapping wird multilinguales Retrieval, selbst mit linksläufigen Schrif-

8 Vgl. <http://www.indigo.ie/egt/standards/og/ogfreagra.html> (daraus die Zeichentabelle in Abb. 5) sowie <http://www.unicode.org/pending/pending.html>.

9 Entwickelt in der Brigham Young University, jetzt in eigener Firma weiterentwickelt und vermarktet: <http://www.wordcruncher.com>.

ten, möglich, freilich ohne eine eindeutige Codierung der Zeichen, da ja nach wie vor nur 8 Bit für die Codierung zur Verfügung stehen¹⁰.

Das altiranische Avesta-Corpus beinhaltet die heiligen Texte der Zarathustrier, heute repräsentiert durch die Religionsgemeinschaft der Parsen; nur ungefähr ein Viertel des ursprünglichen Umfangs des Corpus ist erhalten geblieben. Die Götterhymnen und die Ritual- und Rechtsliteratur des Avesta, zwischen 1000 und 500 v.Chr. entstanden, ist in zwei Sprachstufen des Alt-Nordost-Iranischen verfaßt (Alt- bzw. Jungavestisch).

Ein Charakteristikum der Avesta-Überlieferung in einer Vielzahl von Handschriften (die älteste auf 1288 n.Chr. datierbar) ist die Multilingualität. Da die Tradition des Avestischen als Sprache des Rituals von Sprechern bzw. Benutzern des Mittelpersischen und des Sanskrit getragen wurde, nimmt es nicht wunder, daß in den Handschriften häufig mittelpersische Übersetzungen und Kommentare in die Texte eingebettet sind. Aus dem 11. Jahrhundert n.Chr. stammt eine auf der Grundlage der mittelpersischen Version erstellte Sanskrit-Übersetzung, die ebenfalls in einige Manuskripte integriert ist. Angesichts der enigmatischen Natur der Texte haben all diese nicht-avestischen Textzeugen für die Avesta-Philologie große Bedeutung.

Die elektronische Bearbeitung des Avesta-Corpus repräsentiert den ‚historischen Kern‘ der TITUS-Textdatenbank. Die Eingabe erfolgte im Rahmen des Projektes eines Avesta-Wörterbuchs bereits in den Jahren 1985 bis 1988¹¹. Dank der Weiterentwicklung des Programms WordCruncher kann der elektronische Avesta-Text heute in der vollen Komplexität seiner Überlieferung in einer Hypertext-Umgebung benutzt werden. Abb.6 stellt die maßgebende gedruckte Edition von Karl Geldner (in Originalschrift) und die Edition in WordCruncher einander gegenüber. An der Integration einer morphologischen Analyse in den Hypertext wird gegenwärtig gearbeitet, ebenso wird die Originalschrift-Wiedergabe zu einer künftigen Version gehören. Die beiden WordCruncher-Textfenster zeigen jeweils oben die Ansicht des Lesetextes, unten die erste Textzeile mit den Handschriftenvarianten. Die kursiv gesetzten Einträge sind jungavestische Varianten. Durch Mausklick auf ein beliebiges Wort können sämtliche Belegstellen in einem Referenzfenster sichtbar gemacht werden. Eingebildet ist hier das Search-Manager-Fenster: Dort wird im Ausklappenmenü die Sprache ausgewählt, im WordWheel-Feld erscheinen dann die Wortformen der betreffenden Sprache und die Zahl ihrer Belege. In der z. Z. in Arbeit befindlichen Version ist der Variantenapparat mit digitalisierten Bildern der Handschriften verlinkt.

3.4 *Hypertext in WordCruncher II: Edition und metrische Analyse der Ṛgveda-Saṃhitā*

Ebenfalls zum ‚alten Bestand‘ der TITUS-Datenbank gehört das Corpus der Ṛgveda-Saṃhitā. Der Ṛgveda ist das älteste Textcorpus in altindoarischer Sprache (Entstehungszeit ca. 1200 v. Chr.), heute der älteste Teil der heiligen Schriften der Hindus. Es handelt sich um eine

10 Ausführlich zur Anwendung von WordCruncher für Windows Gippert 1998b.

11 Vgl. <http://titus.uni-frankfurt.de/texte/avesta.htm>

Rgveda-Saṃhitā: Gedruckte Edition in Originalschrift: Saṃhitā-Text	॥३०॥ १-२२ विन्नामिषः ॥ ईन्द्रः ॥ चिहुम् ॥ ॥३०॥ इच्छन्ति वा सोम्यासः सखायः सुन्वंति सोमं दधति प्रयासि । तितिक्षन्ति अभिऽशस्त्रिं जनानांमिद्रु तदा कश्चन हि प्रकेतः ॥१॥	
Padapāṭha-Text	॥३०॥ इच्छन्ति वा सोम्यासः सखायः सुन्वंति सोमं दधति प्रयासि तितिक्षन्ति अभिऽशस्त्रिं जनानां ईद्रु तदा कश्चन हि प्रकेतः ॥१॥	
Linguistische Bearbeitung: Transliteration Saṃhitā-Text	ic ^h ānti tvā somyāsaḥ sāk ^h āyaḥ sunvānti sōmaṃ dād ^h ati prāyāmsi / tītikṣante ab ^h iśastim jānānām indra tvād ā kās canā hī praketaḥ //1//	
Transliteration Padapāṭha-Text	ic ^h ānti tvā somyāsaḥ sāk ^h āyaḥ sunvānti sōmam dād ^h ati prāyāmsi / tītikṣante ab ^h iśastim jānānām indra tvāt ā kāḥ canā hī praketaḥ //1//	
Transliteration u. metrische Restitution	ic ^h ānti tvā somyāsaḥ sāk ^h āyaḥ sunvānti sōmaṃ dād ^h ati prāyāmsi / tītikṣante ab ^h iśastim jānānām indra tvād ā kās canā hī praketaḥ //1//	A
Transkription (phonologisch)	/i ^h c ^h : ^h anti tva: so:mi ^h 'ya:sas 'sak ^h 'a:jas/ /sun ^h 'wanti 'so:mam 'dad ^h 'ati 'praja:msi/ /'tītikṣante a ^h 'b ^h iṣastim 'jana:na:m/ /'indra 'tvat 'a: 'kas cana 'hi prake'tas/	
Übersetzung	<i>Es verlangen nach dir die Somafreunde, sie pressen Soma, bereiten Genüsse, sie erdulden die üble Rede der Leute, Indra, es gibt wirklich keinen Glanz außer dir.</i>	

Metrische Analyse (vgl. Gippert 1998c): Legende zum ASCII-Symbolinventar	B, C, D
U C ₀ Ŷ.	Ŷ ← V V
- C ₀ Ŷ.	/ Wortgrenze an der Zäsur im Trimeter
= C ₀ VC ₁ C _x	\ Wortgrenze vor der Kadenz im Trimeter
- C ₀ VT.RVC ₀	Ü „Zerdehnungen“ geschriebener Langvokale
J C ₀ Ŷ.(#)hVC ₀	: einsilbige „Lücke“ in „unterzähligen“
Y C ₀ Ŷ.(#)T ^h VC ₀	Trimeterversen
~ C ₀ Ŷ.c ^h VC ₀ , C ₀ Ŷ.l ^h VC ₀	
W C ₀ Ŷ.C#VC ₀ (? < C ₀ Ŷ.C#HVC ₀)	
V C ₀ Ŷ#VC ₀	

Metrische Analyse: Wortstruktur	B
va ic ^h ānti tvā somiyāsaḥ sāk ^h āyaḥ ic ^h ānti°11:3-1°-== tvā°11:4-4°-/ somyāsaḥ°11:8-5°-U==°somyāsaḥ\ sāk ^h āyaḥ°11:11-9°Y-X	
vb sunvānti sōmaṃ dād ^h ati prāyāmsi / sunvānti°11:3-1°==U sōmam°11:5-4°-='sōmaṃ/ dād ^h ati°11:8-6°YU_\ prāyāmsi°11:11-9°U-X	
vc tītikṣante ab ^h iśastim jānānām tītikṣante°11:4-1°U==-/ ab ^h iśastim°11:8-5°YU==°ab ^h iśastim\ jānānām°11:11-U-X	
vd indra tvād ā kās canā hī praketaḥ // indra°11:2-1°== tvāt°11:3-3°Wtvād ā°11:4-4°-/ kāḥ°11:5-5°='kās canā°11:7-6°UU hí°11:8-8°_\ praketaḥ°11:11-9°U-X	

Abb. 7: Bearbeitung des Rigveda-Textes:
Edition und metrischer Analyse durch Indizierung mit WordCruncher

Sammlung von 1028 Götterhymnen (10.462 Strophen, insg. 165.007 Wörter) in einem quantifizierenden Versmaß (vorwiegend acht-, elf- und zwölfsilbige Verszeilen).

Die orale brahmanische Texttradition ist von bemerkenswerter Präzision und Geschlossenheit, die verhältnismäßig jungen Handschriften zeigen nur minimale Abweichungen untereinander, jedoch bedarf die handschriftliche Veda-Überlieferung noch einer genauen philologischen und linguistischen Untersuchung. Bestimmte metrische und phonotaktische Strukturen wurden allerdings bereits durch eine frühe Redaktion des Textes, die sog. Diaskeuase, verwischt, so daß in einer Edition die ursprüngliche Gestalt wiederhergestellt werden muß. Ein besonderes Merkmal der R̥gveda-Textgeschichte ist die parallele Überlieferung einer Version ohne Junkturscheinungen an den Wortgrenzen, der sog. Padapāṭha („wortweise Rezitation“). Für die linguistische und philologische Bearbeitung des R̥gveda hat der Padapāṭha den Stellenwert eines ersten sprachwissenschaftlichen Kommentars (vgl. Abb. 7).

Der elektronische R̥gveda-Text basiert auf Eingaben, die schon in den siebziger Jahren in den USA und in Indien geleistet wurden, und konnte im Jahre 1988 in die TITUS-Datenbank integriert und in der Folgezeit revidiert, korrigiert und für WordCruncher formatiert werden. Im Jahre 1994 kam ein elektronischer Padapāṭha-Text, 1996 als dritte Version der metrisch restaurierte Text der Edition von van Nooten/Holland hinzu¹². Das WordCruncher-



Abb. 8: Hypertext-Edition des R̥gveda in WordCruncher für Windows: A Interlinear-Edition des Textes; metrisch analysierte Texte: B Wortstruktur, C Silbenstruktur, D Versstruktur

12 Zur Geschichte des elektronischen Textes vgl. <http://titus.uni-frankfurt.de/texte/rv-inf-e.htm>.

Fenster in Abb. 8 oben links zeigt die drei Textversionen in interlinearer Anordnung; die lange Zeile gibt jeweils den Saṃhitā-Text mit modifizierter Silben- und Zeilenstruktur wieder.

Versifikationstechnische Analyse: Positionierung morphologischer Typen;
Beispiel: Wortausgang -a/ánti im elfsilbigen Vers mit früher und später Zäsur (Form a und b)

11a/b	☑	☐	☑	☐☐	☑☐	☐	☐☐	☐	☐	-	×	
	1	2	3	4	5	6	7	8	9	10	11	
	0/5*a/ánti 11:2-1											
		20/23*a/ánti 11:3-2										
			14/5*a/ánti 11:4-3									
				89/12*a/ánti 11:5-4								
					1/2*a/ánti 11:6-5							
						0/1*a/ánti 11:7-6						
							8/0*a/ánti 11:8-7					
								101/37*a/ánti 11:9-8				
									0/0*a/ánti 11:10-9			
										161/31*a/ánti 11:11-10		

Legende: Zahl der Belege (nicht akzentuierter/akzentuierter Ausgang) *a/ánti° Position im Vers

Abb. 9: Untersuchung der Position morphologischer Typen im Vers mit Hilfe der metrischen Analyse des R̥gveda-Textes durch Indizierung mit WordCruncher

Seit 1995 arbeitet Jost Gippert an einer elektronischen metrischen Analyse des Textes (vgl. Abb. 7 unten). Dabei wurden bestimmte Silbentypen zunächst automatisch durch Symbole ersetzt; bei der anschließenden Indizierung mit WordCruncher konnten die Symbolketten dann im Umfang von Silben, Wörtern oder ganzen Verszeilen als Referenzeinheit gewählt werden. Die so entstandenen metrischen Texte des R̥gveda bilden eine Grundlage von bislang nie erreichter Aussagefähigkeit für sprachhistorische und versifikationstechnische Untersuchungen¹³. Da WordCruncher die Synchronisierung von Textfenstern ermöglicht, lassen sich die verschiedenen Textversionen auf dem Bildschirm simultan darstellen und scrollen (s. Abb. 8).

Im Rahmen meines eigenen Dissertationsprojektes¹⁴ nutze ich die metrischen Texte des R̥gveda, um die Positionierung morphologischer Typen im Vers festzustellen und daraus Schlüsse auf das versifikationstechnische Verfahren der Dichter zu ziehen bzw. versifikationstechnische Modelle zu verifizieren (vgl. Abb. 9). Ich gehe von morphologischen Typen, d.h. Wortformen mit gleichen Endungen, ungeachtet der Wortart, aus und berücksichtige erst in einem zweiten Schritt Wortart und Form. Die Formenbestimmung wird in der neuesten Version des TITUS-R̥gveda-Textes implementiert sein. Auf diese Weise kann die Syntax im Vers in ihrem Verhältnis zur rhythmischen Gestaltung beschrieben und somit erstmals eine verlässliche Basis für die vergleichende Syntaxforschung der altindogermanischen Sprachen geschaffen werden.

13 Gippert 1997b und 1998c.

14 Archilochos' Dichtersprache und ihre Vorgeschichte. Versuch eines indogermanistisch-poetologischen Kommentars.

4 Schluß

Aus dem oben Gezeigten wird die eminente Bedeutung einer in eine hypermediale Umgebung eingebundenen elektronischen Textdatenbank sowohl für die genuin historisch-linguistisch-komparatistische Arbeit als auch für die philologische Auswertung der Materialien offensichtlich. Durch Koordination und Kooperation einzelner Projekte kann die Digitalisierung des gesamten Materials im Laufe der kommenden Jahre geleistet werden. War die Arbeit an einem bestimmten Wissenschaftsstandort im ‚prädigitalen‘ Zeitalter abhängig von der Ausstattung der erreichbaren Bibliotheken, so kann jetzt nicht nur das Sekundärmaterial, d.h. Texteditionen, sondern auch das Primärmaterial, Faksimile-Abbildungen der Textträger, auf dem Bildschirm betrachtet und, bei nicht-lokalem Retrieval nach Herunterladen der Dateien, bearbeitet werden, gleichgültig an welchem Ort der Welt sich der Datenserver befindet. Das Tertiärmaterial, die phonologisch, morphologisch und syntaktisch analysierten elektronischen Texte, werden künftig ebenso über das Internet verfügbar sein. Das TITUS-Projekt wird in Kürze einen WordCruncher-Server aufbauen, so daß die vorindizierten Texte und damit auch ihre bearbeiteten Versionen (z.B. die metrischen Texte des Ṛgveda) mit dem WordCruncher-Viewer als Remote Libraries angesteuert werden können.

Als größtes Hindernis bleibt derzeit die Codierung der Textmaterialien bestehen. Obwohl Unicode® als internationale 16-Bit-Norm in naher Zukunft zweifellos spürbare Erleichterung bringen wird, darf nicht übersehen werden, daß die Architektur dieses Standards nicht auf linguistische, sondern auf industrielle Bedürfnisse zugeschnitten ist. Unicode codiert Zeicheninventare: Die Kennzeichnung von Unicode-Character-Strings nach Sprachen bleibt Protokollen höherer Ebene vorbehalten (Language Tagging). Ist also Unicode-Plain-Text schon in kommerzieller Anwendung auf die lebenden Sprachen der Erde ambig, so wird es umso schwieriger sein, Lateinschrift-basierte Transliterations- und Transkriptionssysteme zur Darstellung historischer Sprachen und Sprachstufen, geschweige denn die Codierung rekonstruierter Sprachschichten über Unicode effizient zu verwalten. Außerdem sind noch einige komplexe Schriftsysteme (Keilschrift, Hieroglyphen) für die Codierung vorzubereiten. Die Aufnahme historischer Schriften in den 16-Bit-Raum ist nur dann möglich, wenn sie von einer lebenden Kultur benutzt oder zumindest als unmittelbares kulturelles Erbe beansprucht werden (z.B. Ogam). Andere historische Schriften werden auf das nächsthöhere Plateau des 31-Bit-Codierungsraumes der ISO 10646 gelegt und können innerhalb Unicode daher nur in einem Transformationsformat (UTF-16, Surrogate Pairs) dargestellt werden¹⁵.

Das TITUS-Projekt testet z.Z. den Umgang mit dem Unicode-8-Bit-Transformationsformat (UTF-8)¹⁶. Es bleibt zu hoffen, daß wir mit einer Typeware-Firma zu einer Vereinbarung kommen, damit ein 16-Bit-Font guter Qualität nach wissenschaftlichen Bedürfnissen modifiziert und im Netz frei verfügbar gemacht werden kann.

[Manuskript abgeschlossen am 28.2.98]

15 Ausführlich dazu Bunz 1998.

16 Vgl. <http://titus.uni-frankfurt.de/unicode/unitest.htm>.

Referenzen

- Bunz, Carl-Martin (1997a): Überlegungen zu einer indogermanistischen 16-Bit-Zeichenverwaltung. IdgFonts für den Macintosh. In: *Studia Iranica, Mesopotamica et Anatolica* 2, 1996[1997], 5-48.
- Bunz, Carl-Martin (1997b): Browsing the Memory of the World. In: 11th International Unicode Conference, San Jose, California, 2.-5.9.1997, Proceedings, 1, A 7 (vgl. <http://titus.uni-frankfurt.de/unicode/papiuc11.htm>).
- Bunz, Carl-Martin (1998): Unicode and Historical Linguistics. In: *Studia Iranica, Mesopotamica et Anatolica* 3, 1997[1998] (*noch nicht erschienen*).
- Bunz, Carl-Martin / Gippert, Jost (1997): Unicode, Ancient Languages and the WWW. In: 10th International Unicode Conference, Mainz, 10.-12.3.1997, Proceedings, 2, C 11 (vgl. <http://titus.uni-frankfurt.de/personalia/jg/unicode/unicode.htm>).
- Gippert, Jost (1990): Präliminarien zu einer Neuausgabe der Ogaminschriften. In: Bammesberger, Alfred / Wollmann, Alfred: *Britain 400-600: Language and History*, Heidelberg 1990, 291-306 (vgl. <http://titus.uni-frankfurt.de/personalia/jg/jg1988b.htm>).
- Gippert, Jost (1993): Ogam – Eine frühe keltische Schrifterfindung. In: *Lectiones eruditorum extraneorum in facultate philosophica Universitatis Carolinae Pragensis factae* 1, Praha 1992 [1993] (vgl. <http://titus.uni-frankfurt.de/personalia/jg/jg1992e.htm>).
- Gippert, Jost (1995a): TITUS. Das Projekt eines indogermanischen Thesaurus, LDV-Forum 12/2, 1995, 35-47 (vgl. <http://titus.uni-frankfurt.de/texte/titusldv.htm>).
- Gippert, Jost (1995b): Von der Keilschrifttafel zur Textdatenbank. *Forschung Frankfurt – Wissenschaftsmagazin der Johann Wolfgang Goethe-Universität Frankfurt am Main*, 1995/4, 46-56 (vgl. <http://titus.uni-frankfurt.de/texte/titusff.htm>).
- Gippert, Jost (1997a): Paläographische Untersuchungen mit dem Computer. In: *Studia Iranica, Mesopotamica et Anatolica* 2, 1996 [1997], 77-100.
- Gippert, Jost (1997b): Laryngeals and Vedic metre. In: *Sound Law and Analogy. Papers in honor of Robert S.P. Beekes on occasion of his 60th birthday*, ed. by Alexander Lubotsky, Amsterdam 1997, 63-79.
- Gippert, Jost (1998a): Digitization of Tocharian Manuscripts from the Berlin Turfan Collection. In: „*Manuscripta Orientalia*“, *International Journal for Oriental Manuscript Research*, St. Petersburg 1998 (*im Druck* – vgl. <http://titus.uni-frankfurt.de/personalia/jg/jg1998a.pdf>).
- Gippert, Jost (1998b): Multilingual text retrieval: Requirements and solutions. In: *Studia Iranica, Mesopotamica et Anatolica* 3, 1997[1998] (*noch nicht erschienen*).
- Gippert, Jost (1998c): Neue Wege zur sprachwissenschaftlichen Analyse der vedischen Metrik. In: *Compositiones Indogermanicae in memoriam Jochem Schindler*, Wien [1998] (*noch nicht erschienen*).
- Vavroušek, Petr (1998): Digitalisierung von Keilschrifttafeln, in: *Studia Iranica, Mesopotamica et Anatolica* 3, 1997[1998] (*noch nicht erschienen*).