

**Achtung!**

Dies ist eine Internet-Sonderausgabe des Aufsatzes  
„Tocharisch mit dem Computer: Ziele und Verfahren“  
von Jost Gippert (1996).

Sie sollte nicht zitiert werden. Zitate sind der Originalausgabe in  
*Tocharian and Indo-European Studies* 7, 1997, 17-34  
zu entnehmen.

**Attention!**

This is a special internet edition of the article  
“Tocharisch mit dem Computer: Ziele und Verfahren”  
[“Applying computers to Tocharian – aims and methods”]  
by Jost Gippert (1996).

It should not be quoted as such. For quotations, please refer to the original  
edition in  
*Tocharian and Indo-European Studies* 7, 1997, 17-34.

**Alle Rechte vorbehalten / All rights reserved:**

Jost Gippert, Frankfurt 1998-2011

Jost Gippert

### **Tocharisch mit dem Computer: Ziele und Verfahren**

Obwohl inzwischen rund 100 Jahre vergangen sind, seitdem die ersten tocharischen Sprachdenkmäler entdeckt und publiziert wurden, sind wir von einer durchgreifenden Erforschung dieser Sprache noch weit entfernt. So fehlt es bisher nach wie vor an einem Wörterverzeichnis, das die Daten aus den edierten Texten zusammenfassen würde, um dann seinerseits zukünftigen Forschergenerationen bei der Bearbeitung unedierter Texte als Hilfsmittel zu dienen. Die Erstellung derartiger Hilfsmittel kann nun heute durch den Einsatz elektronischer Verfahren wesentlich erleichtert werden. Wie dieses Ziel unter Verwendung eines Computers angegangen werden kann und welche anderen Ziele der Einsatz elektronischer Medien im Zusammenhang mit dem Tocharischen haben kann, sei im folgenden kurz erörtert.

Für eine Corpußsprache wie das Tocharische, das lediglich in schriftlichen Originaldokumenten überliefert ist, bieten sich grundsätzlich zwei Bereiche an, in denen der Computer zweckmäßig eingesetzt werden kann. Obwohl beide auf dem Prinzip der Digitalisierung beruhen, sind sie doch technisch und inhaltlich klar voneinander geschieden: Zum einen geht es um eine elektronische Einspeicherung der Originalurkunden selbst, die dabei als ein "graphisches" Ganzes behandelt werden; zum anderen geht es um die Digitalisierung der in den Originalurkunden enthaltenen bzw. daraus zu entnehmenden sprachlichen Daten, d.h. eine Aufarbeitung von Textelementen.

Für die "graphische" Erfassung kommen dabei primär zwei Ziele in Betracht: Zum einen kann ein digitalisiertes Abbild für sich in Anspruch nehmen, im Gegensatz zu Papierbildern, Dias und anderen photographischen Reproduktionen ein unvergängliches, gewissermaßen "ewiges" Abbild seiner Originalvorlage darzustellen. Für die tocharischen Dokumente, deren Erhaltungszustand auch bei sachgemäßer Lagerung im Laufe der Zeit nicht besser, sondern immer nur schlechter wird, könnte somit wenigstens der heute gegebene Stand gesichert werden<sup>1</sup>. Zum anderen läßt sich aus den als graphische Dateien digitalisierten Dokumenten unmittelbar das Material für paläographische Untersuchun-

---

<sup>1</sup> Der auf der Saarbrückener Tagung diskutierte Vorschlag, mit den in verschiedenen europäischen Museen gelagerten tocharischen Sprachdenkmälern erstmalig die umfassende Digitalisierung eines Überlieferungscorpus zu unternehmen, ist inzwischen seiner Realisation einen ersten Schritt näher gekommen: Mit Unterstützung der Berlin-Brandenburgischen Akademie der Wissenschaften, der Staatsbibliothek Berlin sowie der Stiftung Tamai werden in nächster Zukunft die Schriftstücke der Berliner Turfansammlung verarbeitet. Ziel ist die Erfassung der Dokumente auf CD-ROMs, so daß sie der wissenschaftlichen Weiterverarbeitung zur Verfügung stehen.

gen entnehmen, die dann ihrerseits auf elektronischer Grundlage durchgeführt werden können, womit gegenüber manuellen Verfahren (Nachzeichnen von Hand) der Vorteil einer größeren Genauigkeit und geringeren Subjektivität einhergeht.

Auch bei der digitalen Erfassung von Textdaten sehe ich v.a. zwei Ziele: Zum einen kann sie die sprachwissenschaftliche Erschließung, z.B. im Hinblick auf die Erstellung eines Wörterbuchs, wesentlich erleichtern, indem sie eine unmittelbare, automatische Indizierung der Texte ermöglicht. Zum anderen eröffnen die digitalisierten Sprachdaten auch für textphilologische Fragestellungen neue Perspektiven, insofern sie z.B. gezielt für eine Vergleichung unterschiedlicher Textversionen innerhalb derselben Sprache oder über Sprachgrenzen hinweg eingerichtet werden können — angesichts der Tatsache, daß die meisten tocharischen Texte aus anderen Sprachen des buddhistischen Umfelds übersetzt sind, eine immer wieder in den Vordergrund tretende Aufgabe.

Das in Abb. 1 wiedergegebene Flußdiagramm soll einen Einblick vermitteln, wie die somit umrissenen Ziele einer graphischen und textualen Digitalisierung, ausgehend von den Originaldokumenten, erreicht werden können. Dabei zeigt es zugleich, daß die Digitalisierungsvorgänge auf verschiedenen Ebenen auch miteinander verknüpft sind. Die einzelnen Schritte verdienen es, unter Berücksichtigung der Besonderheiten, die die tocharische Überlieferung betreffen, näher erläutert zu werden.

Die **graphische** Digitalisierung setzt, unabhängig davon, ob sie im Hinblick auf eine "ewige" Datensicherung oder zum Zwecke paläographischer Untersuchungen durchgeführt wird, das Einlesen mittels eines optischen Scanners voraus. Dabei kommen, je nach der Vorlage, verschiedene Verfahren in Betracht. Ideal wäre eine unmittelbare Aufnahme des jeweiligen Originals mit einem sog. Flachbettscanner, die jedoch im gegebenen Fall aus konservatorischen Gründen nur selten möglich sein wird (die tocharischen Schriftstücke der Berliner Sammlung sind z.B. in Glasrahmen eingefast und sollten aus diesen nicht ohne Not entnommen werden). Die Erfassungsgrundlage wird stattdessen im Normalfall eine photographische Abbildung sein, wobei Papierabzüge oder Dias gewählt werden können. Die Wahl des Mediums bedingt dann, was für ein Scanner zu benutzen ist (Flachbettscanner für Photos vs. Diascanner); sie selbst hängt davon ab, welche Qualität beim Scannen angestrebt wird.

Hierbei kommen die unterschiedlichen Erfordernisse des avisierten Ziels zum Tragen: Für paläographische Untersuchungen reicht eine mittlere Auflösung (d.h. Dichte der Punkte, in die das digitale Bild zerlegt wird) aus, wie sie heute jeder handelsübliche Flachbettscanner zu verarbeiten imstande ist (150 bis 300

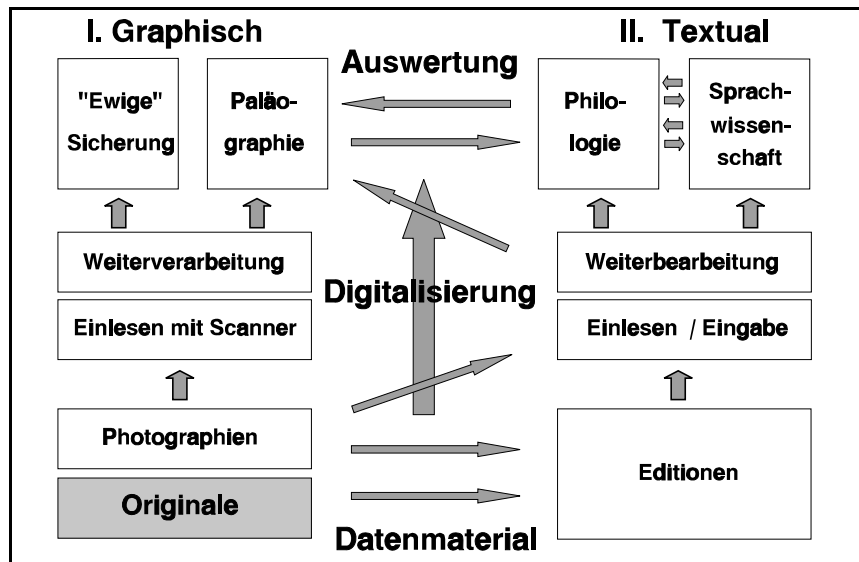


Abb. 1:

Flußdiagramm

dpi<sup>2</sup>); für eine zur Datensicherung geeignete Digitalisierung, die ein möglichst genaues Abbild des einzulesenden Originals anstrebt, sollte jedoch eine höhere Auflösung angestrebt werden (600 dpi und mehr). Als Grundregel gilt dabei zunächst, daß die beim Einlesen eingestellte Auflösung nach dem Scannen auch mit dem leistungsfähigsten Computer nicht mehr "originalgetreu" erhöht werden kann; d.h. eine Bildinformation, die beim Einlesen nicht übernommen wurde, kann später nicht mehr elektronisch hinzugefügt werden. Eine weitere Grundregel besagt, daß die beim Einlesen zu wählende Auflösung sich danach richtet, mit welcher Qualität das zu scannende Original später über ein Ausgabemedium wiedergegeben werden soll. Die untere Richtlinie sollte dabei nicht etwa die (vergleichsweise niedrige) Bildschirmauflösung vorgeben (zwischen 75 und 100 dpi), sondern diejenige von Laserdruckern, die bei der heute üblichen Auflösung von 600 dpi das dem elektronischen Bild zugrundeliegende Punktraster kaum mehr mit bloßem Auge erkennen lassen (Abb. 2a dokumentiert die

<sup>2</sup> "Dots per inch" = Punkte pro Zoll (2,54 cm); zu technischen Voraussetzungen und praktischen Aspekten der computergestützten Paläographie cf. jetzt Verf., Paläographische Untersuchungen [am Beispiel altmaledivischer Hand- und Inschriften]; Vortrag auf der IV. Arbeitstagung zum Computereinsatz in der Historisch-Vergleichenden Sprachwissenschaft, Wien, 16.9.1995 (erscheint innerhalb der Tagungsakten in SIMA 2, Prag 1996 [1997]).

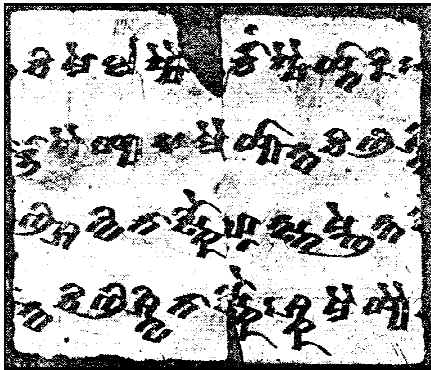


Abb. 2: Fragment, Graustufenbild, 600 dpi

Druckausgabe eines Berliner Fragments mit 600 dpi<sup>3</sup>). Mit der doppelten Druckauflösung von 1200 dpi kommt man bereits dem Standard der Druckindustrie nahe. Unter der Voraussetzung, daß kein Verkleinerungs- oder Vergrößerungsfaktor gegenüber der Vorlage zu berücksichtigen ist, sondern eine Eins-zu-eins-Wiedergabe angestrebt wird, stellt also für die Datensicherung eine Auflösung von 1200 dpi einen geeigneten Wert dar. Eine andere Berechnung ergibt sich aber, wenn mit Kleinbilddias als Zwischenmedi-

um gearbeitet wird. Da diese gegenüber dem Original normalerweise stark verkleinert sind (auf  $24 \times 36$  mm), müßte die Auflösung des zu scannenden Bildes im entsprechenden Verhältnis erhöht werden, um einen 1200-dpi-Ausdruck in Originalgröße zu ermöglichen. Dies sollte jedoch kein Problem darstellen, da die Auflösung handelsüblicher Diascanner ohnehin um ein Vielfaches höher ist als diejenige von Flachbettscannern (z.B. 2700 dpi). Da die Qualität von Dias normalerweise über derjenigen von Papierabzügen liegt, ist ihre Verwendung zum Zweck der sichernden Digitalisierung empfehlenswert.

Als zweites erhebt sich die Frage, ob die Digitalisierung in schwarz-weiß, in Graustufen oder in vielfarbiger Form erfolgen sollte. Während die verschiedenen Einstellungsvarianten für farbige Computerbilder (16 / 256 / 16 Millionen Farbtöne) neben einer unterschiedlich "feinen" Farbnuancierung lediglich einen deutlich unterschiedlichen Speicherbedarf für das Bild mit sich bringen, besteht zwischen "schwarz-weißen" und Graustufenbildern ein elementarer Unterschied: Ein schwarz-weißes Computerbild enthält in der Tat nur schwarze (belichtete) und weiße (unbelichtete) Punkte, jedoch keinerlei Zwischentöne; ein Graustufenbild verfügt hingegen über zahlreiche (meist bis zu 256) verschiedene Grautöne. Eine Schwarz-Weiß-Scheidung ist immer dann erstrebenswert, wenn eine maximale Reduktion des Bildes auf einen Vorder- und einen Hintergrund

<sup>3</sup> Als Scanvorlage diente die Photokopie einer Photographie des Fragments (offensichtlich das Textstück eines Buddhastotras); für die Bereitstellung danke ich L. SANDER, Berlin. — Da Graustufen bei Laserdruckern durch die Zusammenfassung von Punktgruppen erzeugt werden, liegt die reale Druckauflösung niedriger (im gegebenen Fall Gruppen von  $4 \cdot 4$  Punkten  $\rightarrow$  150 dpi).

gewünscht wird; dies ist z.B. bei paläographischen Untersuchungen der Fall, wo es darum geht, ein (schwarz wiederzugebendes) Zeichen möglichst deutlich von seinem (weißen) Untergrund abzuheben. Für die exakte elektronische Abbildung eines alten Originaldokuments wird es hingegen darauf ankommen, auch feinere Schattierungen mitzuerfassen; im Falle der tocharischen Dokumente, bei denen im Laufe der Jahrhunderte der Kontrast zwischen Schrift und Untergrund immer geringer geworden ist, dürfte eine vielfarbige

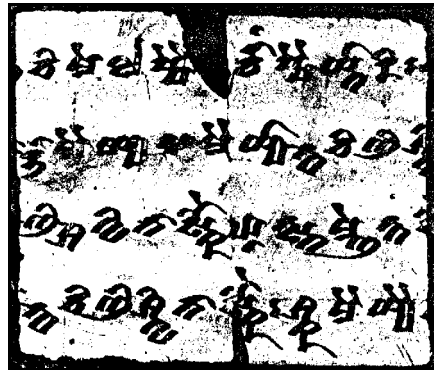


Abb. 3: Dasselbe, Schwarz-Weiß-Grafik

Erfassung sogar unumgänglich sein. Daß dabei eine riesige Datenmenge anfällt (für ein bei 2700 dpi mit 16 Millionen Farbtönen gescanntes Kleinbild-Farbdia ca. 26 MB pro Bild, das wären ca. 20 Disketten), sollte kein Problem mehr darstellen, seitdem Datenträger mit Gigabyte-Kapazitäten zur Normalausstattung gehören.

Trotz der aufgezeigten unterschiedlichen Aufgaben, vor die Datensicherung und Paläographie die elektronische Bildverarbeitung stellen, lassen sich beide Verfahren doch unmittelbar miteinander verknüpfen. Da sich eine Qualitätsreduktion im Gegensatz zu einer Verbesserung bei digitalen Bildern jederzeit durchführen läßt, können die zur Datensicherung dienenden Bilddateien ohne großen Aufwand auch als Grundlage für paläographische Analysen dienen. Prinzipiell wird es für derartige Untersuchungen zunächst immer um die Isolierung von einzelnen Zeichen, im Falle der tocharischen Brāhmī-Schrift also Aksaras, aus einem digitalisierten Original gehen. Dabei sind folgende Einzelschritte zu berücksichtigen:

- a) Identifizierung der zu bearbeitenden Aksaras durch Erstellung einer exakten Transliteration des Schriftstücks (vgl. Tfl. 1: Transliteration des oben wiedergegebenen Berliner Fragments);
- b) Erstellung eines geeigneten digitalen Abbilds des originalen Schriftträgers (hier genügt meist ein Graustufenbild wie das in

] ci me la ṣṣai[ ] rvām ṣṣai yā ri :
]ā=̄ (m)ai yya (sa) mai yyā=nt\ ci ñi (ś)w
] ñi=ś\ wi nā skau 11 tsna mīe nta [
] =nT\ ci ñi świ nā (s)kau : ska mai yyai [

Tfl. 1:

Texttransliteration

Abb. 2 dargestellte), sei es durch Einscannen, sei es durch Anpassung (Umwandlung) eines bereits vorhandenen elektronischen Abbilds (z.B. Farbgrafik);

- c) "Ausschneiden" einzelner Akşaras aus dem digitalen Bild als Einzelgrafiken mithilfe eines Grafikprogramms (vgl. Abb. 4: der Akşara *skau* aus Zeile 3 des Berliner Fragments);



Abb. 4

- d) Optimierung des Schwarz-Weiß-Kontrastes bei der Einzelgrafik, nötigenfalls Nachzeichnen des Zeichenumrisses, Korrektur von Lücken, Beseitigung von Flecken etc. und



Abb. 5

- e) Umwandlung der Einzelgrafik in das Schwarz-Weiß-Format (vgl. Abb. 5: der Akşara *skau* nach der Verarbeitung).

Je nachdem, auf welche Weise die so gewonnenen Daten weiterverarbeitet werden sollen, können die Einzelgrafiken nun als selbständige Dateien gespeichert, in einer Datenbank abgelegt oder auch in einen Computerzeichensatz für Drucker und Bildschirm ("Font") übertragen werden<sup>4</sup>.

Während paläographische Untersuchungen dieser Art noch ein vergleichsweise junges Einsatzgebiet des Computers darstellen, ist die Verarbeitung **textualer** Daten innerhalb geisteswissenschaftlicher Disziplinen schon seit längerer Zeit eine seiner Domänen. Dabei geht es nicht nur um die alltägliche Erfassung wissenschaftlichen Schrifttums (Briefe, Aufsätze, Bücher), sondern es tritt mehr und mehr auch der Einsatz als sprachwissenschaftlich-philologisches Untersuchungsinstrument in den Vordergrund. Nachdem immer leistungsfähigere Programme für die Textanalyse zur Verfügung stehen, zeichnet sich ab, daß der Computer in zunehmendem Maße die auf Zettelkästen basierende Auswertung von Textdaten übernehmen wird.

Eine vorrangige Aufgabe in diesem Bereich stellt die Indizierung der in Texten enthaltenen Wortformen dar, die als Grundlage für Wörterbücher dienen kann. Das Tocharische als eine Corpussprache stellt hierfür geradezu den idealen Anwendungsfall dar. Voraussetzung für eine Indizierung des in tocharischen Texten enthaltenen Wortmaterials ist natürlich die Erfassung der Texte selbst. Diese kann zwar prinzipiell von Hand erfolgen — und sie muß sogar manuell geschehen, wenn es sich um Fragmente wie das oben wiedergegebene handelt,

---

<sup>4</sup> Für Anwendungsbeispiele vgl. den o.g. Beitrag in SIMA 2.

das noch nicht ediert ist. Wo immer gedruckte Editionen vorliegen, kann jedoch wiederum auf optische Scanner zurückgegriffen werden. Dabei kommt freilich ein ganz anderes Verfahren zum Tragen, als wenn die graphische Verarbeitung eines Originals beabsichtigt ist. Der Scanner erzeugt zwar auf jeden Fall zunächst das graphische Abbild einer Vorlage; bei der Erfassung von Textdaten aus einer gedruckten Edition geht es jedoch darum, aus diesem Abbild heraus die den Text bildenden Einzelbuchstaben automatisch herauszufiltern und als solche zu "erkennen". Dieser Vorgang setzt den Einsatz sog. OCR-Programme voraus ("Optical Character Recognition")<sup>5</sup>.

Im Falle des Tocharischen wird man bei der Zeichenerkennung mit einem Problem konfrontiert, das für die meisten der handelsüblichen OCR-Programme nicht lösbar ist, nämlich daß die Editionen neben den alltäglichen Lateinbuchstaben eine Vielzahl von Diakritika verwenden, die den Programmen nicht bekannt und deshalb nicht ohne weiteres "erkennbar" sind. Hinzu kommt, etwa in den "klassischen" Ausgaben von SIEG-SIEGLING, die Verwendung von Unterstreichungen zur Kennzeichnung der sog. Fremdbuchstaben; auch diese erschwert den Erkennungsprozeß bei einfachen OCR-Programmen wesentlich. Erforderlich ist für das Einlesen edierter tocharischer Texte deshalb auf jeden Fall ein OCR-Programm, das "lernfähig" ist in dem Sinne, daß ihm neue, über die Basisbuchstaben hinausgehende Zeichenformen und -kombinationen mitgeteilt werden können (sog. "Teaching"). Dabei kommt es anfänglich überhaupt nicht darauf an, daß eine gegebene Zeichenform als solche gespeichert und wieder ausgegeben wird; entscheidend ist vielmehr eine umkehrbar-eindeutige Codierung, durch die jedem Zeichen im Computer ein eigener Wert zugewiesen wird. Auf diese Weise läßt sich mit einem leistungsfähigen, "lernfähigen" OCR-Programm z.B. der Text 217a aus der Edition der A-Texte wie in Tfl. 2 dargestellt einlesen.

Mithilfe eines ersten Konvertierungsschritts, der im wesentlichen aus Vorgängen des Suchens und Ersetzens besteht, läßt sich hieraus bereits ein lesbarer Text erzeugen, der dem Original mit Ausnahme der unvermeidbaren Verlesungen ähnlicher Zeichen bereits sehr nahe kommt (kein OCR-Programm kann für sich eine 100%ige Lesegenauigkeit in Anspruch nehmen; eine scheinbar hohe Genau-

---

<sup>5</sup> Cf. dazu allgemein bereits Verf., Perspektiven des Computereinsatzes in der Orientalistik, in: Forschungsforum, Universität Bamberg, Ht. 2, 1990, 133-136. — Über die rechtliche Situation, in der man sich beim Einlesen gedruckter Werke befindet, herrscht vielerorts Unsicherheit. Im Rahmen des deutschen Rechts ist diese Frage im Wissenschaftsbereich jedoch eindeutig geregelt: Niemand kann es einem Wissenschaftler, sei es als Einzelperson, sei es als Mitglied einer Arbeitsgruppe, verwehren, die für seine wissenschaftliche Betätigung notwendigen Materialien mit einem Computer aufzuarbeiten. Die Wahl der Mittel (Scanner, Eintippen) ist dabei völlig unerheblich.



```

2 //// kñ[a]ññā [TA] - - - [p.] sIie y .y wāryāñc [ṣṣ ] - - -ptā ////
4 //// e skākā wāIpāt\ p(tā)ñKAT\ ārkIṣossIs ~ant\ MArkanIpaL\ āk-
sis(s)i - - - - -
[p]ûK\ knānmām TMAṣ\ ~ram poñcām wāltsa -
~1 - - - sn - k(ā)ckeyo : Paklyoṣas wrasañĀ pūk kācke p~sācĀ pūk
~ānmām - k. ñKat\
Parko parnoIIt\ nIāgat sI. m ypeyam : wārpā
5 - ks[I]ssI.~ krañcām markampal mā~u ~a~tam ok~~ ñāktas napenas Sam\
o~I.acI.: 8
Palskāt\ pūk\ knāImām ke mal~ āksIsaM\ lyāklyām k~pā
6 - - ~ma)rkampal\ n ktI.m walunt\ nd[r ]akem : Palkāt kārunik\ tam
nekk aI.ādem ṣaptakoñim
walnnt\ Palkāt\ katkont\ ynālek\ cmoll eIItsntt : pañcake~s\]
7 - - [sa]kk atsek naslyes\ dūskacāI. s(I) klopam ṣkam ñI. yāmar spak-
tām tñqyo cem\ :
cesm~~ āksI.ññam kI.aIIt markampal ~ Inetas ūr~ilwā yāṣ ~~I.āna-
syacĀ Pa
3 - - ram ūpage : (-) ptāñkt(e) araIn~a ~Tt\ t asI. na~ ] lkālam sal-
pIIIām okI. pūttI.ṣpar siIn
kāwāaltenam pañIyo : sne Ptsak IIpape Palkāt\ ptāñāktacĀ\ sa -

```

Tfl. 2:

A 217a nach dem Einlesen mittels optischem Scanner

igkeit von 99% bedeutet immerhin bereits einen Lesefehler in jeder zweiten Zeile!). Bei dieser Gelegenheit ist es natürlich zugleich möglich, die Codierungsgepflogenheiten der zugrundeliegenden Edition zu überdenken und gegebenenfalls durch andere, für die Weiterverarbeitung mit dem Computer besser geeignete zu ersetzen. So erscheint es nicht unbedingt zweckmäßig, die von SIEG-SIEGLING — vermutlich aus rein drucktechnischen Gründen — eingeführte Unterstreichung der "Fremdbuchstaben" beizubehalten. Obwohl die Erstellung eines Fonts (Zeichensatzes für Bildschirm und Drucker), der unterstrichene Buchstaben als selbständige, eindeutig codierte Zeichenformen enthält, heute ohne weiteres möglich ist, sollte doch zumindest dann, wenn eine Übertragbarkeit des bearbeiteten Textes zwischen verschiedenen Systemen angestrebt wird, eine möglichst systemunabhängige Lösung gewählt werden (die Verwendung von zeichenunabhängigen Unterstreichungsfunktionen, die nahezu jedes Textverarbeitungsprogramm kennt, scheidet aus diesem Grund von vornherein aus). Man könnte so z.B. eine Codierung erwägen, bei der die "Fremdbuchstaben" statt durch Unterstreichung durch Majuskeln dargestellt sind (z.B. *kā* durch KA), darüber hinaus das <sup>ā</sup> durch Ā, das <sup>u</sup> durch U, der Virāma durch <sup>ˆ</sup> usw. Damit ergäbe sich für das o.a. Textstück eine Wiedergabe wie in Tfl. 3.

Um den Text für eine elektronische Analyse vorzubereiten, ist es dann erforderlich, die allfälligen Lesefehler zu beseitigen. Sofern nicht ein bestimmter Fehler in gleichem Kontext regelmäßig auftritt, kann dies nur von Hand erfolgen. Darüber hinaus muß der Text mit Strukturangaben versehen werden, die eine eindeutige Lokalisierung von Wortformen etc. gestatten. In welcher Form derartige Strukturangaben (z.B. Text- und Zeilennummern) zu gestalten sind,

```

2 //// kñ[a]ññā [TA] - - - [p.] sne y .y wāryāñc [śś ] - - -ptā ////
3 //// e skākā wāipāt' p(tā)ñKAT' ārkisōssis ~anT' MArkampal' āksis(s)i - - - - -
   [p]jūK' knānmām TMAŠ' ~ram poñcām wāłtsa -
4 - - - sn - k(ā)ckeyo : PaklyoŠas wrasañĀ pūk käckē p~sācĀ pūk ~ānmām - k. ñKat'
   Parko parñont' māgat si m ypeyam : wārpā
5 - ks[i]ssi~ krañcām markampal mā~u ~a~tam ok~~ ñāktas napenas Sam' o~iaci:8
   Palskāt' pūk' knānmām ke mal~ āksisaM' lyāklyām k~pā
6 ---ma)rkampal' n ktim walunt' nd[r]akem: Palkāt kārunik' tam nekk aiādem Šaptakoñim
   walñnt' Palkāt' katkont' ynālek' cmoll entsñnt' : pañcake~s]
7 - - [sa]kk atsek naslyes' dūskacāi s(i) klopam śkam ñi yāmar spaktām tñqkyo cem' :
   cesm~~ āksiññām kiant markampal ~ inetas ūr~ilwā yāŠ ~~~iānasyacĀ Pa
8 - - ram ūpage : (-) ptāñkt(e) arain~a ~Tt' asi na~ ] lkālam salpñiām oki pūttišpar siin
   kāwāłtenam pañiyo : sne Ptsak npage Palkāt' ptāñāktacĀ' sa -

```

**Tfl. 3:** Dasselbe nach erster Überarbeitung (Suchen und Ersetzen)

hängt wiederum von den Vorgaben des Programms ab, mit dem man die Analyse durchzuführen gedenkt. Im Rahmen des "Thesaurus Indogermanischer Text- und Sprachmaterialien" (TITUS), der in naher Zukunft alle für die Indogermanistik relevanten Textmaterialien in elektronischer Form umfassen soll und für den die tocharischen Texte derzeit bearbeitet werden<sup>6</sup>, besteht aufgrund positiver Erfahrungen z.B. eine Präferenz für das von der Brigham Young University entwickelte Retrievalsystem "Wordcruncher", das eine wenig aufwendige Kennzeichnung derartiger Strukturebenen ermöglicht<sup>7</sup>. Damit würde das behandelte Textstück die in Tfl. 4 wiedergegebene Form annehmen (es bezeichnen: |c217a: Blattnummer; |p1 bis |p8: Zeilennummern).

Für bestimmte Fragestellungen können zusätzliche Markierungen erforderlich sein. So wäre es z.B. sinnvoll, zur Vorbereitung paläographischer Untersuchungen innerhalb der Textstücke die Aksaragrenzen zu markieren, so daß auch ein Zeichenindex automatisch aus dem eingegebenen Text extrahiert werden kann. Durch die Verwendung von Bindestrichen etc. lassen sich gleichzeitig Wortzusammenhänge darstellen. Im Falle des oben behandelten Berliner Fragments ergäben sich damit etwa Gestaltungen wie in Tfl. 5 und 6 dargestellt.

<sup>6</sup> Cf. den Bericht "TITUS – Das Projekt eines indogermanistischen Thesaurus" in: LDV-Forum 12/2, 1995, 35-47; über den derzeitigen Stand informiert "TITUS – Alte und neue Perspektiven eines indogermanistischen Thesaurus" (erscheint in SIMA 2, 1996 [1997]). — An den tocharischen Textcorpora arbeiten Chr. SCHAEFER und O. HACKSTEIN.

<sup>7</sup> Gegenüber der älteren DOS-Version, die nur drei Ebenen zu verwalten gestattete (z.B. Buch, Kapitel, Vers), kennt die jüngst auf den Markt gekommene Windows-Version bis zu 10 Ebenen, was auch für komplexeste Textstrukturen ausreichen wird.

c217a
p1
p2 //// kñ[ā]ññā [TA] --- [p · ] sne y · · y wāryāñc · [śś·] -- -ptā ////
p3 //// · · e skākā wārpāT' p(tā)ñKAT' ārkīsoṣṣis kranT' MArkampaL' āksis(s)i --- ----- [p]ūK' knānmām TMAŠ' bram poñcām wāltsa -
p4 --- sn · · · k(ā)ckeyo : PAKlyoŠAs wrasañĀ' pūk kācke parsācĀ' pūk' knānmām - k · ñKAT' PARko parnonT' māgat šim ypeyam : wārpā
p5 - ks[i]ssi krañcām MArkampaL' māryu praṣṭam okñāš ñāktas napeNAs SAM' oñkraci: 8 PAIskāt' pūk' knānmām ke maltw āksisaM' lyāklyām kUpā
p6 -- (MA)rkampaL' n · ktim walunT' ud[r]akem : PAlkāt kārūñiK' TAm nekk arā- dem ŠAptakoñim walunT' PAlkāt' KATkonT' ynāleK' cmoll entsunT' : pañcake[S']
p7 -- [ša]kk atsek naslyeS' dūškacār š(i) klopaṃ škam ñi yāMAR spaktām tuñkyo ceM' : cesmy ākṣiññam kranT' MArkampal meTAs ūrbilwā yāŠ' bārānasyacĀ' PA
p8 -- ram ūpage : (9) ptāñkt(e) aRAM[pā]T(·) asiNA[T'] lkĀLAM SAIPmām oki pūttīspar šim kāwālteNAM pañiyo : sne PtSAk upage PAlkāt' ptāñāktacĀ' sa -

Tfl. 4:

A 217a nach Einrichtung für "Wordcruncher"

] ci me la ṣṣai[ ] rvām ṣṣai ytā ri : [
](ā)= <sub>u</sub> (m)ai yya (sa) mai yyā=nt\ ci ñi (ś)[w
] ñi=ś\ wi nā skau 11 tsna mñe nta [
] =nT\ ci ñi świ nā (s)kau : śka mai yyai [

Tfl. 5:

Zeichenweise Transliteration

] ci-me-la-ṣṣai [ne-]rvām-ṣṣai ytā-ri : [
](ā)= <sub>u</sub> (m)ai-yya-(sa) mai-yyā=nt\ ci ñi(ś)[w
] ñi=ś\ wi-nā-skau 11 tsna-mñe-nta [
] =nT\ ci ñi-ś_wi-nā-(s)kau : śka-mai-yyai [

Tfl. 6:

Dasselbe mit wortweiser Gruppierung

Derartige Zusatzmarkierungen sollten jedoch ebenso wie Klammern, mit denen etwa unsichere Lesungen oder Ergänzungen angezeigt werden können, die Auswertung nach anderen Kriterien, etwa im Hinblick auf einen Wortformenindex, nicht beeinträchtigen. Dies setzt voraus, daß das verwendete Analyseprogramm ("Retrievalsystem") in der Lage ist, sie bei Bedarf zu ignorieren. Da die elektronische Analyse von Textdaten letztlich immer auf der Auswertung von Zeichenketten beruht, hängt ihre Effizienz in entscheidendem Maße davon ab, wie flexibel die Funktion der einzelnen Zeichen bestimmt werden kann. So muß es z.B. möglich sein, zu wählen, ob man den Unterschied zwischen der "Normal-schreibung" *mā* und der "Fremdgraphie" *MA* ignorieren will oder nicht. Einfache Suchalgorithmen, über die heute jedes Textverarbeitungsprogramm ("Schreibprogramm") verfügt, stoßen hier schnell an ihre Grenzen.

Auf der Grundlage eines eingespeicherten elektronischen Textes eröffnen sich verschiedene sprachwissenschaftlich-philologische Untersuchungsmöglichkeiten, die über die einfache Indizierung von Texten im Hinblick auf ihre Elemente (Wortformen, Akşaras etc.) hinausgehen. So dürfte es im Falle der tocharischen Texte in mancherlei Hinsicht, z.B. im Hinblick auf eine Wortkonkordanz, zweckdienlicher sein, statt der durch das Originaldokument gegebenen Textgestaltung (Textblätter, Seiten, Zeilen) eine auf syntaktischen oder metrischen Einheiten beruhende Einteilung zugrunde zu legen. Diese sollte die Originalstruktur nicht einfach ersetzen, sondern vielmehr ergänzend hinzugefügt werden, so daß eine Wahlmöglichkeit zwischen beiden Ebenen bestehen bleibt. So könnten z.B. in dem bereits behandelten Textstück A 217a zusätzliche Markierungen wie in Tfl. 7 dargestellt eingegeben werden (es bezeichnen: |bUA den Buchtitel [UA= Udānavarga-Alamkāra], |s21,7 bis 13 Strophennummern innerhalb des 21. Kapitels, |va bis |vd jeweils Verse 1-4 innerhalb der Strophen).

bUA
s21,7 c217a p1
vd  p2 /// kñ[ā]ñnā [TA] --- [p ·] sne y · · y wāryāñc · [śś·] ---ptā ///
s21,8
va  p3 /// · · e skākā wārpāt· p(tā)ñKAT· ārkišoşşis kranT· MArkampaL· äksis(s)i ----
vb ---- [p]üK· knānmām TMAŞ· bram poñcām wāłtsa -  p4 --- sn · - k(ā)ckeyo :
vc PAklyoŞAs wrasañĀ· pük kække parsācĀ· pük· knānmām - k ·ñKAT· PArko parnonT· māgat şim ypeyañ :
vd wārpā  p5 - ks[i]ssi krañcām MArkampaL· māryu praştam okñās nāktas napeNAs SAM· onkraci: 8
s21,9
va PAłskāt· pük· knānmām ke maltw äksisaM· lyāklyām kUpā  p6 -- (MA)rkampaL· n ·ktiñ walunT· ud[r]akem :
vb PAłkāt kārūniK· TAm nekk arādem ŞAptakoñim walunT· PAłkāt· KATkonT· ynāleK· cmoll entsunT· :
vc pañcake[S·]  p7 -- [sa]kk atsek naslyeS· düşkacār ş(i) klopañ škam ñi yāMAR spaktām tuñkyo ceM· :
vd cesmy äkşinñiam kranT· MArkampal meTAs ūrbilwā yāŞ· bārānasyacĀ· PA  p8 -- rañ ūpage : (9)
s21,10
va ptāñkt(e) aRAm[pā]T(·) asiNA[T·] lkālAm SAłpmām oki pütüspar şim kāwālteNAM pañiyo :
vb sne PTAk upage PAłkāt· ptāñaktacĀ· sa -  c217b  p1 ---- [i]sk · [a]tS· kus SAs tāk · - śnis su[K·](:)
vc wtsy [o]ky akriñri laP· tsem atSats śāku śāk ŞAK pi şi mañKAtt (o)ki ypicĀ· akmal şi wārKAntyo :
vd tsem yo -  p2 -- [ā]tSats PATkrūyo ś(i)şk[i]ss oki w(ā)kmtsam şamwem ci aşuk malañ kāsweñi : 10
s21,11
va RAtram lymem cinçram āryamśĀ· āñkari lyāklyām wrok şi sark oki sa[M·] kamañ pa -  p3 -- --(:)
vb [ā]len ·cākKār sopiñ [püK·] [p]rārwañ SAtkār tsā yāŞ· TPAr ākāşam yTāŞTRAm yon nü tkaniS· mraçĀ· :
vc we SAM· ptāññaktacĀ· nanmü ślāñcālyi kus · ·  p4 nāTAK· ñ[KA ·]tje n · - TAprem lkātsi pañitswātS· :
vd ponsām tāŞ· tkanā mā cw oky ālyAKAm waŞTAŞ lañtunT· PAłko şem nāŞ· ūdūmpar şi [ş ·]ā  p5 -- --(:) [11]
s21,12
va -- [tñ ·] indriñĀ· āştr akmal şy aRAm wsoK· wsā yok yatS· IUKAşnū worpü k[aŞ·] swāñceno sük· lkātsi :
vb kUCac waŞTAŞ lañtu ky artār [M]A  p6 ----- Ş PArklune - ·k · şi tñi kārūniK· :
vc piññKAT KAssi we pük· ly[utā]r · · ----- nasaM· pükyo sne loplune wraMAmny[o] (:)
vd  p7 /// l(·)o p · s[k]i · ä - [MA]tt(a)k nü yn(e)ş [KArso] -- am ŞArpñiM· şlok we
s21,13
va nuna ---- ŞAr[p]i ///  p8

Tfl. 7: A 217a mit Markierungen entsprechend metrischen und Sinnabschnitten

Auf dieser Grundlage kann man weiter erwägen, auch die Entschlüsselung metrischer Strukturen mit elektronischer Unterstützung anzugehen. Während dies für Texte wie die homerischen Epen, Plautusstücke oder die Rgveda-Samhitā bereits mit interessanten Ergebnissen gelungen ist<sup>8</sup>, wird sich im Falle des Tocharischen hierbei aus der lückenhaften oder sogar fragmentarischen Überlieferung ein größeres Problem ergeben. Wie groß das Problem tatsächlich ist, kann jedoch erst dann beantwortet werden, wenn Versuche in dieser Richtung einmal unternommen worden sind.

Weniger als problematisch denn als zeitaufwendig dürfte sich demgegenüber ein weiterer Überarbeitungsschritt darstellen, der im Hinblick auf eine umfassende elektronischsprachwissenschaftlich-philologische Auswertung der tocharischen Texte erforderlich wäre, nämlich eine durchgehende morphologisch-syntaktische Bestimmung der Wortformen (sog. "Tagging"), die es z.B. erlauben würde, Fragestellungen wie diejenige nach der Verbindung eines bestimmten Prädikatsverbs mit einer Adverbialbestimmung im Perlativ "abrufbar" zu machen. Obwohl derartige Verfahren ebenfalls bereits für zahlreiche Sprachen erarbeitet worden sind, lassen sie sich doch nicht ohne weiteres auf das Tocharische übertragen, da sie in erheblichem Maße von der jeweiligen morphologischen

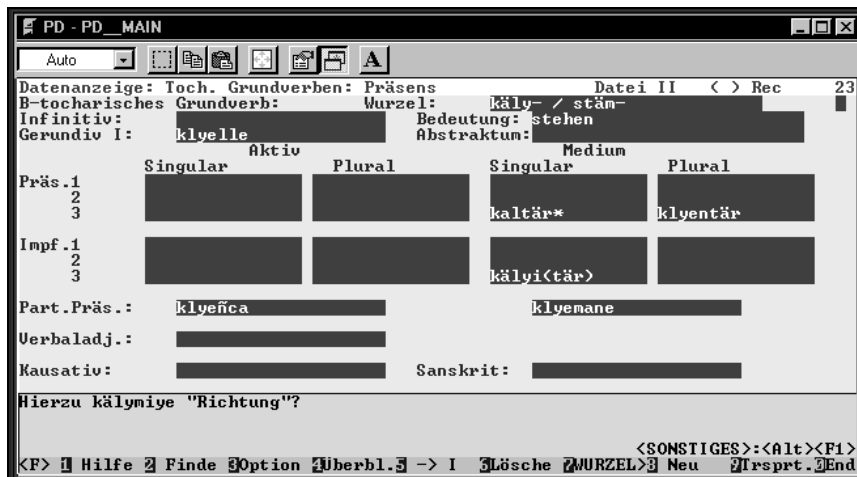


Abb. 6: Bildschirmmaske einer Datenbank zum tocharischen Verb

<sup>8</sup> Cf. z.B. jetzt Verf., "Neue Wege zur sprachwissenschaftlichen Analyse der vedischen Metrik" (erscheint in: Stefanoma [Gedenkschrift Schindler], 1996 [1997]).

Struktur abhängen. Zweckmäßig wird hier zunächst die Erstellung einer Datenbank sein, die die überlieferten Wortformen zusammen mit den zugehörigen morphologischen Angaben erfaßt. Eine solche Datenbank könnte etwa wie in Abb. 6 dargestellt aussehen.

Die Verfügbarkeit morphologisch-syntaktischer Angaben zu den Wortformen ist zugleich die Voraussetzung für eine Ausweitung der Untersuchungen über die Sprachgrenzen hinaus, d.h. für eine Einbeziehung von (inner- und außertocharischen) Parallelversionen der Texte zum Zwecke der Feststellung von Abhängigkeiten. Erste Bedingung für eine solche Ausweitung ist natürlich die elektronische Verfügbarkeit der Paralleltexte selbst (Buddhistisches Sanskrit, Pāli, Prakrit, Tibetisch, Chinesisch, Ujgurisch u.a.), wozu auf einschlägige Bemühungen verschiedener Institutionen hingewiesen werden kann<sup>9</sup>; auch im Rahmen des indogermanistischen Thesaurus sind bereits einige Sanskrit- und Prakrit-Texte aufbereitet worden. Allerdings reicht in diesem Zusammenhang eine einfache Parallelanordnung der Texte, wie sie Tfl. 8 ff. exemplarisch zeigen, noch nicht für eine automatische Analyse im Hinblick auf Abhängigkeiten aus (hier wurden lediglich bestimmte übereinstimmende Stichwörter durch Unterstreichung markiert; es bedeuten: LV: Lalitavistara; CPS: Caṣṣ-pariṣatsūtra; Vin.: Vinayapīṭaka; MV: Mahāvagga; Uv: Udānavarga. Unter SS FW ist der tocharische Text zusammen mit den von E. SIEG und W. SIEGLING in der Fs. Moriz Winternitz, Leipzig 1933, 167 ff. vorgeschlagenen, hier fett gedruckten Ergänzungen in einer vereinheitlichenden Transkription erfaßt). Erforderlich wäre statt dessen die Erarbeitung von Verfahren, mit denen sich der Entsprechungsbezug zwischen lexikalischen Elementen der einzelnen Versionen eindeutig markieren ließe. Ob sich ein derartiges "Linking" zweckmäßig auf der Basis eines sog. "Hypertext"-Verfahrens realisieren läßt, wie es sich z.B. innerhalb von Internet-Anwendungen in jüngster Zeit immer größerer Beliebtheit erfreut, oder ob eine andere Grundlage besser geeignet ist, muß sich erst noch erweisen.<sup>10</sup>

Eine durchgreifende sprachwissenschaftliche Auswertung, die die aus der Textanalyse zu gewinnenden Erkenntnisse systematisch in einer Datenbank erfaßt, um somit als Grundlage für die Erstellung von Wörterbüchern, Grammatiken usw. zu dienen, wäre also das Ziel, das sich mit dem Einsatz elektronischer

---

<sup>9</sup> So z.B. die "Buddhist Text Encoding Initiative"; ein CD-ROM mit Pāli-Texten hat die Mahidol-Universität Bangkok unter dem Titel "BUDSIR" veröffentlicht.

<sup>10</sup> Zur Problematik der automatischen Verknüpfung von Übersetzungstexten und Originalen vgl. Verf., "Towards an automatical analysis of a translated text and its original", in: SIMA 1, 1994 [1995], 21-60; zu den Grundlagen von "Hypertext" K. FENDT, *Interactive Reading: Hypertext and Literary Studies*, ib., 161-166.

c217a s21,8	//// · · · e skākā wārpāt' p(tā)ñKAT' ārkisōssis kranT' MArkampal' āksis(s)i - - - -
va p3	//// · · · e skākā wārpāt p(tā)ñkāt ārkisōssis kranT' mārkaṃpal āksis(s)i - - - -)
(SS FW 168	- - - - [p]ūK' knānmām TMAŠ' bram poñcām wāltsa -  p4 - - - - sn · · · k(ā)ckeyo:
vb	- - - - [p]ūK' knānmām tmās bram poñcām wāltsa -  p4 - - - - sn · · · k(ā)ckeyo:)
(SS FW 168	[[LV400,20-22
[[CPS8,18	atha khalu śikhī mahābrahmā tathāgatasyādihivāsanām viditvā tuṣṭa udagra ātamanāḥ prasudhātāḥ prīṭisauṃnasyajātas tathāgatasya pādau śirasā vanditvā tatraivāntaradhā.]]
[[CPS8,18	atha brahmanāḥ sabhāmpater etad abha(vat / deśayisyati bhagavān dharmam deśa)jisyati sugato dharmam iti viditvā hr̥ṣṭas tuṣṭa pramudita udagra prīṭisauṃnasyajāto bhagavatpādau śirasā vanditvā bhagavantam iriḥpradaksinīkṛtya tatraivāntaradhātā.]]
⟨Vin.I,5,12	atha kho brahmā sahaṃpati katavakāso kho 'mhi bhagavatā dhammedāsanayā 'ti bhagavantam abhivādetvā padakkhinam katvā tathā, ev' antaradhāyi. )]
vc	PAklyōSAs wrasañĀ' pūk kācke parsācĀ' pūk' knānmām - k · ñKAT' PARko parnoNT' māgatsim ypeyam :
(SS FW 168	pāklyōsās wrasañ pūk kācke parsāc pūk knānmām k(om)ñkāt pārko parnoNT' māgatsim ypeyam:)
[[LV401,1-3/MV319	atha khalu bhikṣavo bhaumo devās tasyām velāyām antarīkṣebhyo devebhyo ghoṣam udīra-yanti sma / śabdām anuśrāvayanti sma / adya māṛṣā tathāgatenārhatā samyak saṃbuddhena dharmacakrapravartanāyai pratiśrutam ...]]
[[LV400,19	praviśantī śraddhā navihetanjñāḥ śṛvanti dharmam magadhesu sattvāḥ]]
[[CPS8,9	(prādūr) babhūva magadhesu pūrvam dh(armo hy a)śuddhaḥ samalānubuddhaḥ]]
⟨Vin.I,5,7	pāturahosi magadhesu pubbe dhammo asuddho samalehi cintito))
vd	wārpā  p5 - ks i ssi krañcām MArkampal' māryu praṣṭam okñās nāktas napeNAs SAM' onkraci: 8
(SS FW 168	wārpā(t ā)ksissi krañcām mārkaṃpal māryu praṣṭam okñās nāktas napeñās sām onkraci: 8
[LV400,18	apāvṛtās teṣām amṛtasya dvārā brahmanti satatam ye śrotavataḥ]]
[[CPS8,9	avavṛṇṣva amṛtasya dvāram vadasya dharmam virajānubuddham //]]
⟨Vin.I,5,7	apāpur' etaṃ amatassa dvāram sunantu dhammam vimalenānubuddham //))
s21,9  va	PAkṣāT' pūk' knānmām ke maltw āksisam lyāklyām kUpā  p6 - - - (MA)rkampal' n · ktim walunT' ud[r]akem :
(SS FW 168	pāḥskāt pūk knānmām ke maltw āksisam lyāklyām kupā(rām kranT' mā)rkampal n(o)ktim walunT' ud[r]akem :))
[[CPS9,6-8	6 (atha bhagavata etad abhavat / kasya nv ahaṃ prathamato dharmam deśayeyam / atha bhagavata etad abhavat / yanv ahaṃ udrakasva rāma)putrasya pr(atha)mato dharmam deśayeya(m) .. 7 (deva)jā (bhagavata) ā(roca)yanti abhidosaḥkālagato māṛṣa udrako rāmaputr(ah /) 8 (bhagavato 'pi samanvāhṛtya jñānadarśanam pravartate / abhidosaḥkālagato udrako rāmaputraḥ]]
⟨Vin.I,6,3-4	atha kho bhagavato etad ahoṣi: kassa nu kho ahaṃ pathamam dhammam deśeyyam, ko imaṃ dhammam khippam eva ājānissatīti. atha kho bhagavato etad ahoṣi: ayaṃ kho Uddako Rāmaputto paṇḍito vyattome dhāvī dīgharattam apparajakkhajātiko. yaṃ nūnāhaṃ Uddakassa Rāmaputtassapathamam dhammam deśeyyam, so imaṃ dhammam khippam eva ājānissatīti. //3// atha kho antarāhitā devatā bhagavato ārocesi: abhidosaḥkālamkato bhante Uddako Rāma putto 'ti. bhagavato pi kho nāṇam udapādi abhidosaḥkālamkato Uddako Rāmaputto 'ti. ...))
vb	PAkāt kārūnik' TAm nekk arādem sāptakoṇim walunT' PAkāt' KATkonT' ynālek' cmoll entsunT' :
(SS FW 168	pāḥkāt kārūnik tām nekk arādem sāptakoṇim walunT' pāḥkāt kātkont ynālek cmoll entsunt :))
[[CPS9,1-3	1 (a)jha bhāga(vata e)jād abhavat / kasya nv ahaṃ pratha(mato dharmam deśayeyam /) 2 (atha bhagavata etad abhavat yanv ahaṃ arādakālāmasya prathamato) dh(a)rmaṃ deśayeya(m) .. 3 (devatā bhagavata) āro(cayanti) sāptāhakālāgata ār(ā)ḍakālāmah (/) ]]
⟨Vin.I,6,1-2	atha kho bhagavato etad ahoṣi: kassa nu kho ahaṃ pathamam dhammam deśeyyam, ko imaṃ dhammam khippam eva ājānissatīti. atha kho bhagavato etad ahoṣi: ayaṃ kho Ālāro Kālāmo paṇḍito vyatto medhāvī dīgharattam apparajakkhajātiko. yaṃ nūnāhaṃ Ālārassa Kālāmassa pathamam dhammam deśeyyam, so imaṃ dhammam khippam eva ājānissatīti. //1// atha kho antarāhitā devatā bhagavato ārocesi: sattāhakālāmkato bhante Ālāro Kālāmo 'ti. bhagavato pi kho nāṇam udapādi sattāhakālāmkato Ālāro Kālāmo 'ti. atha kho bhagavato etad ahoṣi: mahājāniyo kho Ālāro Kālāmo, sace hi so imaṃ dhammam suṇeyya, khippam eva ājāneyyā 'ti.))

Tfl. 8: Tocharischer Text (A 217 ab) in synoptischer Darstellung mit Parallelversionen

vc	<b>pañcake[S*]  p7 -- [ʃa]kk atsek naslyeS* dūskacārs(i) klopaṃ śkaṃ nī yāMaR spaktām tuñkyo ceM' :</b>
((SS FW 168	<b>pañcake[s] (pālkāt)</b> [ʃa]kk atsek naslyes dūskacārs(i) klopaṃ śkaṃ nī yāmār spaktām tuñkyo cem :))
[[CPS9,9	9 (atha bhagavata etad abhavat (/) kasya nv ah(aṃ pratha)mat(o) dha(r)m(aṃ) d(e)śay(e)y(aṃ) (/) atha (bhagavata) etad abhavat (/) yanv ahaṃ pañcakānām bhikṣūnām prathamato dharmam deśayeyam ye māṃ pū(r)v(aṃ) duḥkha)prahānāvogam anu(yuktaṃ viharantaṃ) satkrty(opa)ta-sthu(h) pre(mnā ca gauraveṇa ca /))
⟨Vin.I,6,5	atha kho bhagavato etad ahoṣi: kassa nu kho ahaṃ paṭhamam dhammam deseyyam, ko imaṃ dhammam khippam eva ājānissatīti. atha kho bhagavato etad ahoṣi: bahūpakārā kho 'me pañcavaggiyā bhikkhū, ye maṃ padhanapahitattam upatthahimsu.))
vd	<b>cesmy āksiññāṃ kraṇT* MArkampal meTAs ūrbilwā yās* bārānasyacĀ* PA  p8</b> -- <b>ram ūpage : (9)</b>
((SS FW 169	cesmy āksiññāṃ kraṇt mārkampal metās ūrbilwā yās bārānasyac pā(lkāt vtā)ram ūpage : (9)))
[[CPS9,11-10,1	11 (drṣṭvā ca punar) yathābhiraṃyaṃ bodhimūle vihrtya yena bār(ā)ṃ(asi) kāsīnām nigamaṃ tena caryām prakrāntah) \ 10.1 (tena khalu sama)yena upaga ājīvikas ta(smīn eva mārge 'dhvaprati-panno) 'bhūt (/)
⟨Vin.I,6,5-6	yaṃ nūnāhaṃ pañcavaggiyānaṃ bhikkhūnaṃ paṭhamam dhammam deseyyān ti. atha kho bhagavato etad ahoṣi: kahaṃ nu kho etarahi pañcavaggiyā bhikkhū viharantīti. addasa kho bhagavā dibbena cakkhunā visuddhena atikkantaṃanusakena pañcavaggiye bhikkhū Bārānasiyaṃ viharante Isipatane migadāye. atha kho bhagavā Uruvelāyaṃ yathābhiraṃtaṃ viharitvā yena Bārānasī tena cārīkaṃ pakkāmi. addasa kho Upako ājīviko bhagavantaṃ antarā ca Gayam antarā ca bodhim addhānamaggaṃpatipannaṃ))
s21,10	
va	<b>ptāñkt(e) aRAm[pā]T(*) asiNA[T*] lkālAm sālpmām oki pūttīsparsim kāwālteNAm pañiyo:</b>
((SS FW 169	ptāñkt(e) arām[pā]t asinā[t] lkālām sālpmām oki pūttīsparsim kāwāltenām pañiyo :))
vb	<b>sne pISak upage PAlkāt* ptāñāktacĀ* sa -  c217b  p1 - - - - [I]sk · [a]tS* kus SAs tāk - - śnis su[K*] (:)</b>
((SS FW 169	sne ptsāk upage pālkāt ptāñāktac sa -  c217b  p1 -- (pā)[I]sk(am) [a]ts kus sās tāk(iṣṣ a)śnis su[k] (:))
vc	<b>wtsy [o]ky akritri laP* tsem ātSats śaku śak ŚAk pi ṣi mañKAtt (o)ki ypicĀ* akmaḷṣi wārKAntyo :</b>
((SS FW 169	wtsy [o]ky akritri lap tsem ātsāts śaku śak śak pi ṣi mañkātt (o)ki ypic akmaḷṣi wārkāntyo :))
vd	<b>tsem yo -  p2 - - [ā]tSats PAtkrūyo ś(i)šk[i]ss oki w(ā)kmtsam śamwem ci aṣuk malañ kāsweñi : 10</b>
((SS FW 169	tsem yo(kāñ aśām) [ā]tsāts pātkrūyo ś(i)šk[i]ss oki w(ā)kmtsam śamwem ci aṣuk malañ kāsweñi : 10
s21,11	
va	<b>RAtram lymem ciñcraṃ ārkyamśĀ* ānkari lyāklyām wroksi sarkk oki sa[M*] kamañ pa -  p3 - - - (:)</b>
((SS FW 169	rātram lymem ciñcraṃ ārkyamś ānkari lyāklyām wroksi sarkk oki sa[m] kamañ pa -  p3 - - - (:))
vb	<b>[ā]len · cākkār sopiñ [pūK*] [p]rārwaṃ SATkār tsā yāS* TParr ākāśam yTāSTRAM yon nū tkaniS* mrācĀ* :</b>
((SS FW 169	[ā]len(am) cākkār sopiñ [pūk] [p]rārwaṃ sātkaṛtsā yās tpār ākāśam ytāstrām yon nū tkanis mrāc :))
vc	<b>we SAM* ptāñāktacĀ* nanmū ślāncālyi kus · ·  p4 nāTAK* n[KA · t]e n · · TApreṃ lkātsi pañitswāts* :</b>
((SS FW 169	we sām ptāñāktac nanmū ślāncālyi kus (tu) nātāk n[kā](t)[t]e n(apem) tāpreṃ lkātsi pañitswāts :))
⟨Vin.I,6,7	divāna bhagavantaṃ etad avoca:))
vd	<b>pontsām tāS* tkanā mā cw oky ālyakAm waṢTAṣ lamtunT* PAlko ṣem nāS* udūmpar ṣi [ṣ · jā  p5 - - - (:) [11]</b>
((SS FW 169	pontsām tās tkanā mā cw oky ālyakām waṣtāṣ lamtunt pālko ṣem nās udūmparṣi [ṣ](t)(ā)(mm oki) (:) [11]))



[s21,12			
[va	-- [tñ ] indriñ <sup>Ä</sup> āstr akmalsy aRAm wsoK <sup>4</sup> wsā yok yatS <sup>4</sup> IUKAśnū worpū k[as <sup>4</sup> ]		
	swāñcenyō sūK <sup>4</sup> lkātsi :		
((SS FW 169	-- [tñ](i) indriñ āstr akmalsy arām wsok wsā yok yats lūkāśnū worpū k[as] swāñcenyō sūk		
	lkātsi :))		
[[CPS10,2	(viprasannāni te āyu)sm(an) gautama i(ndri)yāni <sup>23</sup> pari(suddho mukhavarna <sup>24</sup> paryavadātas		
	chavivarnah <sup>25</sup> kas te āyu)sm(an) gautama sā(stā))		
⟨Vin.I,6,7	vippasannāni kho te āvuso indriyāni, parisuddho chavivanno pariyodāto.⟩		
[vb	kUcac waSTAṣ lamtu ky artār [M]A [p6 - - - - - Ṣ PArkLune - - k · si tñi kārūnik <sup>4</sup> :		
((SS FW 170	kucaç waštās lamtu ky artār [m]ä[rkampal] - - - (ca)ṣ pārklune (tu l)k(ā)ṣit nī kārūnik:))		
[[CPS10,2	k(am) vāsy uddi(s)ya (pravrajit)ah kasya v(ā) dharmam r(o)c(a)yasi (v) ]]		
⟨Vin.I,6,7	kam 'si tvam āvuso uddissa pabbajito, ko vā te sathā, kassa vā tvam dhammam rocesīti.⟩		
[vc	ptāñKAT KAssi we pūK <sup>4</sup> ly[utā]r · · · · · nasaM <sup>4</sup>		
	pūkyo sne loplune wraMAmny[o] (:)		
((SS FW 170	ptāñkāt kāsī we pūk lyutār [nāntsu] (pūk knānmām/kārso) škam nasam		
	pūkyo sne loplune wramānyo (:))		
[[Uv.21,1ab	sarvābhibhūh sarvavid eva cāsmi sarvāś ca dharmāih satatam na liptah]]		
⟨Vin.I,6,8	evaṃ vutte bhagavā Upakam ājivikam gāthāhi ajjhabhāsi:		
	sabbābhibhū sabbavidu 'ham asmi sabbesu dhammesu anupalitto))		
[vdp7	//// l(·)o p · s[k]i · ä -		
	[MA]tt(a)k nū yn(e)ś [KArso] -- ām SArpñiM <sup>4</sup> slok we		
((SS FW 170	(pūk lek tārkorās tsā)l(po) p(r)a)skī[n](tw)ās		
	[mā]tt(a)k nū yn(e)ś kārso(rās) ām sārpnīm (slok we)		
[[Uv.21,1cd	sarvamjahah sarvabhayād vimuktah svayam hy abhijñāya kam uddiseyam // 1]]		
⟨Vin.I,6,8	sabbajjaho tanhakkhaye vimutto, sayam abhiññāya kam uddiseyyam.⟩		

Tfl. 10:

Dasselbe, Fortsetzung

Verfahren im Hinblick auf die Textanalyse verknüpft. Eine derartige Datenbank würde letztlich auch einen indogermanistischen Ausblick eröffnen: Sie könnte als Basis dienen für einen mit elektronischen Mitteln durchzuführenden linguistischen Vergleich der tocharischen Dialektdaten untereinander oder im indogermanistischen Zusammenhang. Die Aufgabe würde darin bestehen, mit programmtechnischen Verfahren einerseits die innertocharischen lautgesetzlichen (und analogischen) Entwicklungen unter Berücksichtigung ihrer relativen Chronologie abzubilden, wobei das zu rekonstruierende Urtocharische als Vergleichsbasis anzusetzen wäre; zum anderen sollte es die zwischen diesem und der urindogermanischen Grundsprache anzunehmenden Entwicklungslinien widerspiegeln. In Abb. 7-9 ist das avisierte Verfahren an verschiedenen A- und B-tocharischen Verbalformen zur Wurzel *kārś-* "wissen" exemplifiziert. Theoretische Grundlagen und praktische Ausgestaltung eines derartigen Computerprogramms sollen in nächster Zukunft im Rahmen des von der DFG geförderten Projekts "AUREA" ("Avesta und Rgveda: Elektronische Auswertung") am Beispiel des Indoiranischen ausgearbeitet werden. Angesichts der Tatsache, daß die Divergenzen zwischen dem ältesten Altindischen und Altiranischen den Unterschieden zwischen den tocharischen Dialekten weitgehend entsprechen, sollten sich die dort zu gewinnenden Erkenntnisse ohne weiteres auf das Tocharische übertragen lassen.

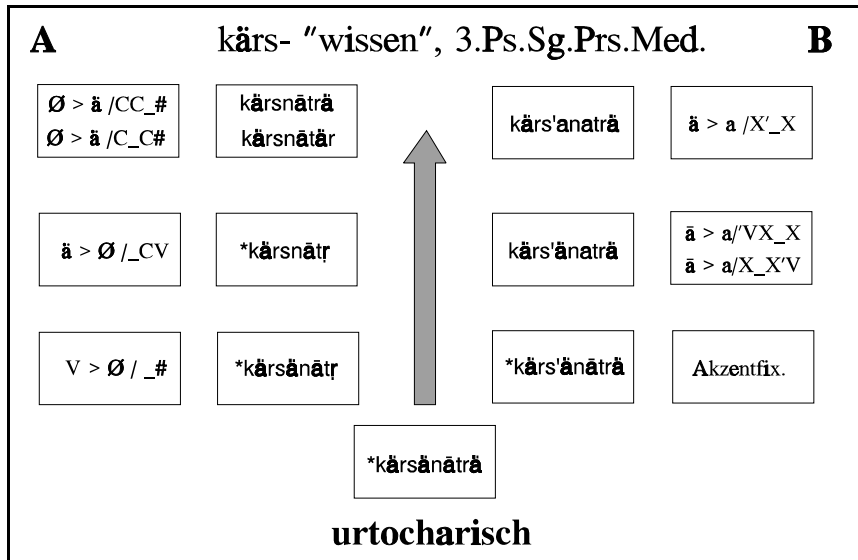


Abb. 7: Entwicklung der 3.Ps.Sg.Präs.Ind.Med. von  $\sqrt{\text{kärs}}$

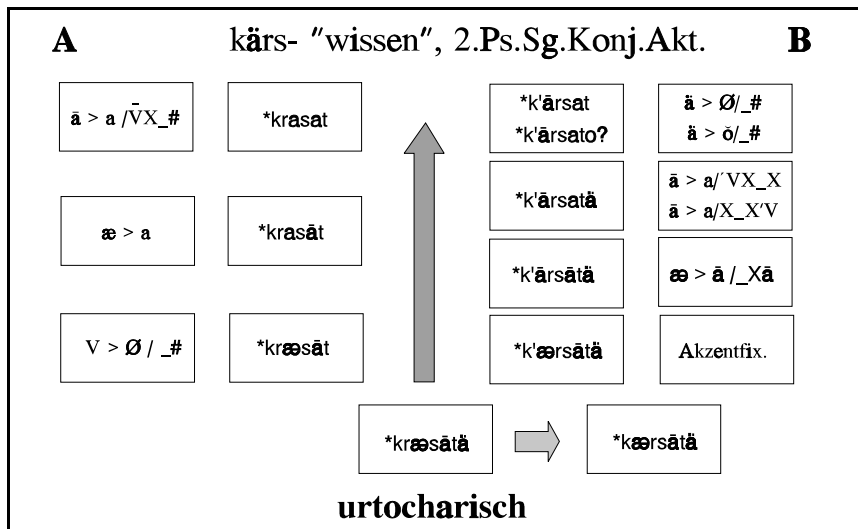


Abb. 8: Entwicklung der 3.Ps.Sg.Aor.Ind.Med. von  $\sqrt{\text{kärs}}$

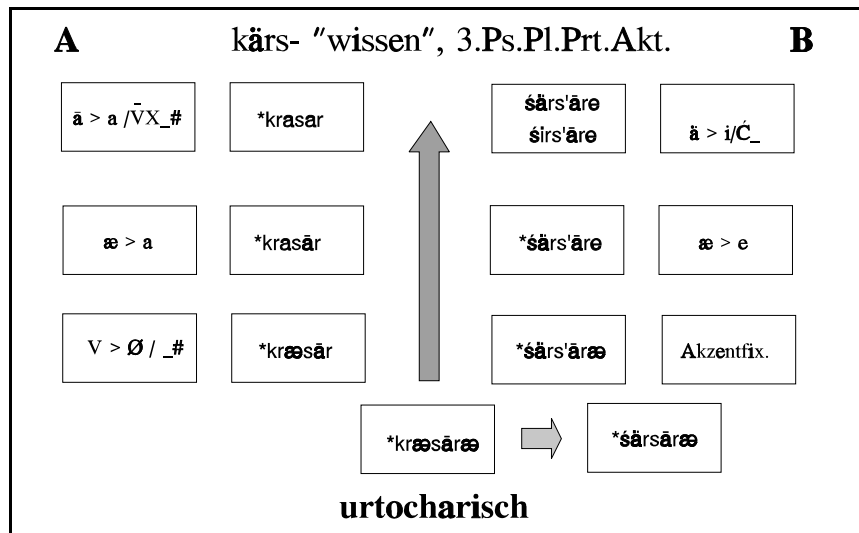


Abb. 9:

Entwicklung der 3.Ps.Pl.Aor.Ind.Med. von  $\sqrt{kärs}$