

Die Lexikographie der Gāndhārī-Sprache

Unter Benutzung moderner Technologien und Standards entsteht das erste Wörterbuch der buddhistischen Literatursprache Gāndhārī.

VON STEFAN BAUMS UND ANDREW GLASS

DIE MITTELINDISCHE Sprache, die heute allgemein als „Gāndhārī“ bezeichnet wird, erhielt diesen Namen im Jahre 1946 durch den australisch-englischen Gelehrten H. W. Bailey. Der Begriff sollte den nordwestindischen Dialekt bezeichnen, der sein Kerngebiet in der alten Region von Gandhāra (dem Peschawar-Tal in Pakistan) hatte und in den Inschriften des indischen Kaisers Aśoka in Shahbazgarhi und Mansehra (3. Jhdt. v. Chr.) sowie einer Dharmapada-Handschrift aus Khotan und zahlreichen Holz-, Leder- und Seidendokumenten aus Niya an der südlichen Seidenstraße (3. Jhdt. n. Chr.) bezeugt war. Die frühesten chinesischen Übersetzungen buddhistischer Texte (ab dem 2. Jhdt. n. Chr.) wurden wenigstens teilweise auf Grundlage von Gāndhārī-Originalen angefertigt und stellen zusammen mit Lehnwörtern in der mitteliranischen Sprache Khotan-Sakisch und im Tocharischen eine wichtige Nebenüberlieferung für die Gāndhārī-Sprache dar.

Forschungsstand

Seit dem Erscheinen von Baileys Artikel ist die Materialgrundlage für die Gāndhārī-Lexikographie durch umfangreiche Neufunde von Handschriften – aber auch Inschriften, Verwaltungsdokumenten und Münzen – um ein Vielfaches angewachsen: Der von uns erstellte Catalog of Gāndhārī Texts (<http://gandhari.org/catalog>) verzeichnet derzeit 77 umfangreiche Schriftrollen, 330 Handschriftenfragmente, 834 Inschriften, 792 Niya-Dokumente und 335 unterschiedliche Münzlegenden mit einem geschätzten Textbestand von insgesamt 120.000 Wortbelegen.

Für dieses Material liegt noch kein umfassendes Wörterbuch vor, und man musste sich bislang mit einer unhandlichen Anzahl von Wortlisten in den Anhängen einzelner Textausgaben behelfen. Die Ausgaben der Inschriften (Hultsch, Konow),

Niya-Dokumente (Boyer, Rapson & Senart) und des Khotan-Dharmapada (Brough) verzeichnen nach unserer Schätzung insgesamt etwa 11.900 Wortformen, allerdings nur in reinen Listen ohne morphologische Analyse oder Bedeutungsangabe. Für die neuen Handschriftenfunde, die seit dem Jahr 2000 in der Reihe „Gandhāran Buddhist Texts“ herausgegeben werden, liegen gründlichere Behandlungen mit morphologischen und semantischen Angaben sowie Sanskrit- und Pali-Parallelen vor, aber wieder nur als Anhänge zu einzelnen Textausgaben. Die Glossare der bereits veröffentlichten Ausgaben (Salomon, Allon, Lenz, Glass) beinhalten etwa 2.700 Wortformen, und zwei in Vorbereitung befindliche Ausgaben (Baums, Lee) fügen weitere 2.200 Wortformen hinzu. Wenn man Wortformduplikate zwischen den Einzelausgaben berücksichtigt und annimmt, dass jede Wortform im Korpus durchschnittlich zweimal belegt ist, dann wird bisher insgesamt nur etwa ein Fünftel des gesamten Wortmaterials in Ausgaben mit Wortindizes aufgelistet, und nur etwa ein Zwanzigstel wird dort morphologisch und semantisch näher bestimmt.

Die lexikographische Erfassung wird nicht nur durch die große Materialfülle, sondern auch durch eine Reihe weiterer Faktoren erschwert.

𑀧𑀲𑀭𑀯𑀢𑀺𑀓 *spadoaṭṭhaṇa-satipaṭṭhāna-*. The form *spadoaṭṭhaṇa-*, *śpaḍova* [spaḍot^h:a:na]; *spado* Mindfulness.

nom. sg. *svadova*

nom. pl. *śpaḍova*

gen. pl. *spaḍoṭha*

unclear [*śpaḍova*]

In *spadoaṭṭhaṇasama* *svadovaṭṭhaṇasañi-*.

Abb. 1: Artikelentwurf für das Dictionary of Gāndhārī.

- n. (Skt. *smṛtyupasthāna-*, P
m [*spado*]aṭaṇa- under central MIA
vadovaṭhaṇa- [smaḍo:vat^h:a:na];
aṭhaṇa- [spaḍo:vat^h:a:na]; *spadoṭhaṇa-*
aṭaṇa- [spaḍo:vaṭ:a:na].

ṭhaṇa Akṣs-G 214.

[ṭha]ṇa SaṅCm-G 126, 129, 133.

ṇa SĀ-G^{S5} 33–34, 38.

[ṭhaṇa] SaṅCm-G 243.

gi- (*spadoaṭaṇasamagi-*),

Der Erhaltungszustand des Quellenmaterials (besonders der Handschriften und Inschriften) ist oft ungemein fragmentarisch, so dass viele Wortbelege unvollständig und ihres Kontextes beraubt sind. Die phonetische Mehrdeutigkeit der Kharoṣṭhī-Schrift, die zudem in *scriptio continua* mit zumeist minimaler Interpunktion verwendet wird, erschwert die Identifizierung von Lexemen und Formen. Der Wettbewerb mehrerer orthographischer Systeme für die Kharoṣṭhī-Schrift sowie das Nebeneinander von Lehnwort- und echten Gāndhārī-Formen (siehe den Beispielartikel *spadoaṭhaṇa*, Abb. 1) stellen die Lemmatisierung des Wortbestandes vor große Herausforderungen. Ganz besonders aber erfordert die Tatsache, dass ein Großteil der Gāndhārī-Texte noch gar nicht in wissenschaftlichen Ausgaben verfügbar ist und einem fortlaufenden Interpretationsprozess unterliegt, ein

offenes lexikographisches System, das Hand in Hand mit der Editions-tätigkeit ausgebaut wird.

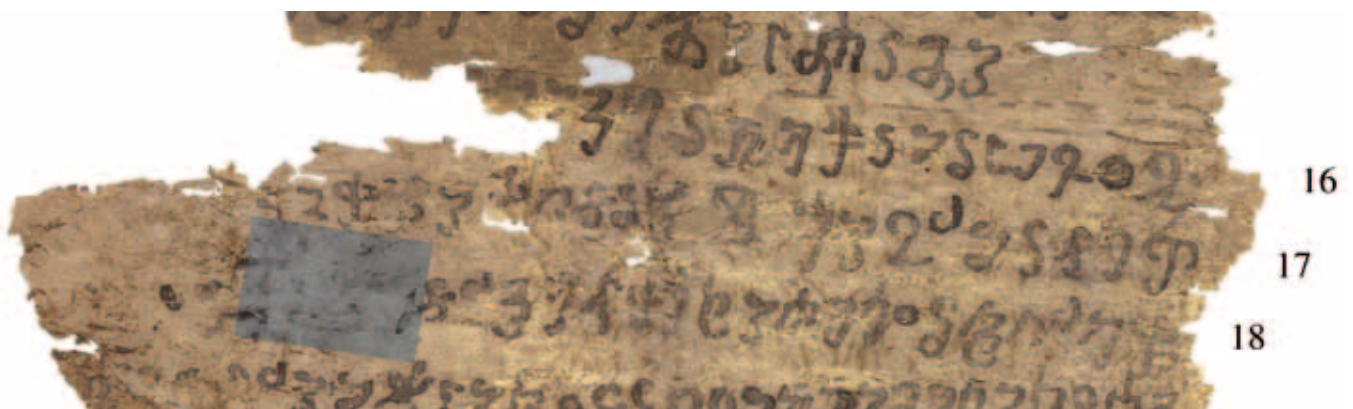
Das Dictionary of Gāndhārī

Die Anfänge des Dictionary of Gāndhārī liegen im Jahr 1997 (<http://gandhari.org/dictionary>), als Andrew Glass sich dem Early Buddhist Manuscripts Project der University of Washington in Seattle anschloss. Um die Bearbeitung der neugefundenen Gāndhārī-Handschriften zu erleichtern, erstellte er digitale Texte der einzigen zuvor bekannten Gāndhārī-Handschrift (des Khotan-Dharmapada) sowie der Gāndhārī-Inschriften Aśokas und des Lokalfürsten Senavarma. Aufgrund der technischen Beschränkungen der Zeit kam zur Darstellung der Sonderzeichen dieser Texte zunächst eine Privatkodierung zum Einsatz,

die aber bald durch den weltweit verbreiteten Unicode-Standard ersetzt wurde. Diese digitalen Texte repräsentativer Gāndhārī-Quellen konnten zunächst nur als Ganzes durchsucht werden.

Im Jahr 2001 entschlossen wir uns, die vorliegenden Arbeiten zu einem umfassenden Wörterbuch der Gāndhārī auszubauen. Glass entwickelte einen Algorithmus zur Alphabetisierung von Gāndhārī-Wörtern und automatisierte die Erstellung von Wortindizes der bis dahin digitalisierten Texte. Im folgenden Jahr wurden mit Unterstützung des Programmierers Tho To die vorliegenden Volltexte und Wortindizes in eine relationale Datenbank importiert. Wir integrierten sie mit dem Catalog of Gāndhārī Texts und unserer Bibliography of Gāndhārī Studies (<http://gandhari.org/bibliography>) und machten die Datenbank über das Internet frei zugänglich.

Abb. 2: Ein Ausschnitt aus einer Gāndhārī-Handschrift. Fragment 1 der British Library Collection.



```

<lb n="16" type="worddiv"/>ho ° <lg met="anustubh"></lg>eva ṇadu tada
thero </bhrad<unclear reason="damage">a</unclear> budha<unclear
reason="damage">sa tadi</unclear> <supplied reason="edge" extent="2"
unit="aksara">ṇo °</supplied> </><supplied reason="edge" extent="7"
unit="aksara">śpaya karmu viagha</supplied>

<lb n="17" type="worddiv"/>ṣe </>aṇodatu mahasare ☒</lg> <lg
met="anustubh"></lg>caduḍīśami saghami </><unclear
reason="damage">kuḍa</unclear>

<lb n="18" type="worddiv"/>gharo maya kridu ° </>vivaśisa praveṇo
</>badhumadi<unclear reason="damage">raya</unclear> <supplied
reason="lost" extent="1" unit="aksara">dha</supplied> <unclear
reason="damage">ṇ</unclear><supplied reason="lost" extent="1"
unit="portion">i</supplied><unclear reason="damage">e</unclear> °</lg>

```

16. ho ° eva ṇadu tada thero bhrad[a] budha[sa tadi] /// + + + + + + + + +
 17. ṣe aṇodatu mahasare ☒ caduḍīśami saghami [kuḍa]
 18. gharo maya kridu ° vivaśisa praveṇo badhumadi[raya] + [ṇ].[e] °

Abb. 3 oben: Die Zeilen der Handschrift aus Abb. 2 in der Datenbank mit Textmarkierung unter Anwendung des TEI-Standards.

Abb. 3 unten: Vom Datenbanksystem automatisch erzeugte Darstellung der diplomatischen Transliteration dieser Zeilen.

Eine neue Herausforderung ergab sich, als Mark Allon, der für die Bearbeitung der Senior Collection zuständig war, im Jahr 2002 Seattle verließ, um eine Stelle an der University of Sydney anzutreten, und als im Jahr darauf Glass einen Forschungsaufenthalt an der Bukkyo-Universität in Kyoto begann. Die Datenbank der veröffentlichten Gāndhāri-Texte war zwar nun über das Internet weltweit verfügbar, aber der Austausch und die Indizierung der noch in Bearbeitung befindlichen Texte war bislang nur im Rahmen der wöchentlichen Projekttreffen in Seattle möglich. Aus diesem Grund entwickelte Glass in Japan ein von Grund auf neues Datenbanksystem, in dem die Eingabe, der Austausch, die Indizierung und die Veröffentlichung neuen Textmaterials seither vollständig online funktionieren, während Baums die Erstellung des Dictionary in Seattle weiter vorantrieb. Jede Bearbeitung eines Textes führt nun zu einer unmittelbaren automatischen Einfügung der Änderung in den Wortindex des Dictionary, und den Bearbeitern neuer Texte steht somit bei ihrer Arbeit ein immer aktuelles lexikographisches Hilfsmittel zur Seite. Das neue System ermöglicht es den Bearbeitern von Texten außerdem, ihre Transliterationen gezielt anderen Benutzern zugänglich zu machen, und schafft damit die Voraussetzung für editorische Gruppenprojekte.

Abb. 4: Vom Datenbanksystem automatisch erzeugte Darstellung der Textrekonstruktion dieser Zeilen unter Berücksichtigung der Verseinteilung.

Mit Forschungsmitteln des National Endowment for the Humanities (NEH) konnten wir von 2007 bis 2010 wieder an der University of Washington an unserem Wörterbuchprojekt zusammen-

arbeiten. In diesem Zeitraum erfuhr die Dictionary-Datenbank einen erheblichen Textzuwachs durch die Einspeisung von Transliterationen aller veröffentlichten Niya-Dokumente (bereitgestellt von Mariner Padwa) sowie der noch unveröffentlichten Transliterationen der Senior Collection (Mark Allon) und der Bajaur Collection (Ingo Strauch). Gleichzeitig trieb Baums die Erfassung der Inschriften und Münzlegenden voran, so dass die Datenbank des Dictionary of Gāndhāri nunmehr vollständige Transliterationen fast aller bekannten Gāndhāri-Texte enthält, allerdings meist noch ohne lexikographische Markierung und beim unveröffentlichten Material in noch sehr vorläufiger Fassung.

Gegen Ende dieser Arbeitsphase wurde deutlich, dass die technische Einrichtung der Datenbank auf den international üblichen Kodierungsstandard der Text Encoding Initiative (TEI) umgestellt werden sollte – sowohl, um den Austausch von Erfahrung und Werkzeugen mit verwandten Projekten zu erleichtern, als auch, um die größtmögliche Nachhaltigkeit der Forschungsergebnisse zu gewährleisten –, und der Digital Humanist Donald Craig begann, nach unseren Formatvorgaben eine TEI-Exportfunktion für die Dictionary-Datenbank zu entwickeln. Gleichzeitig entstand bei uns der Wunsch, auch Photographien der Quelldokumente in die Datenbank einzubinden und mit den Transliterationen zu verknüpfen, und wir entwickelten einen Prototyp dieser Funktionalität. Anfang 2010 (zum

```

eva ṇadu tada thero
bhrada budhasa tadi(*ṇo °
śpaya karmu viagha)[17]ṣe
aṇodatu mahasare ☒

caduḍīśami saghami
kuḍa[18]gharo maya kridu °
vivaśisa praveṇo
badhumadiraya(*dha)ṇ(*i)e °

```

Ende des Finanzierungszeitraums durch das NEH) lagen somit Vorarbeiten für eine umfassende Formatkonvertierung und Funktionserweiterung der Datenbank vor, deren Umsetzung aber einen längerfristigen Planungsrahmen erforderte. Wir verlagerten unseren Arbeitsschwerpunkt deswegen zunächst auf die Vervollständigung der Quelltext-Sammlung des Dictionary.

Zukunftspläne und Akademievorhaben

Mit der Einrichtung des Münchner Akademievorhabens im Jahr 2012 bestehen nun die nötigen Voraussetzungen für die geplante durchgreifende Verbesserung der Datenbank. In Henriette Roued-Cunliffe steht uns eine Mitarbeiterin mit langjähriger Programmiererfahrung zur Seite, die nun die Konvertierung des Datenbankinhalts sowie der Datenbank-Software zum TEI-Standard schrittweise vornimmt. Als nächstes inhaltliches Zwischenziel wird die vollständige lexikographische Markierung der bereits veröffentlichten

Vorarbeiten von Glass) dienen und auch bei der digitalen Veröffentlichung der Handschriften sowie ihrer Vermittlung im universitären Unterricht eine wichtige Rolle spielen. Das Dictionary wird weiterhin frei zugänglich online veröffentlicht, soll aber zusätzlich in einer Druckversion erscheinen. Dank des innovativen Arbeitsflusses wird es nicht in alphabetischer Reihenfolge der Einträge, sondern schrittweise mit der Einarbeitung je neuen Quellmaterials komplettiert.

Das Münchner Akademievorhaben leistet einen zentralen Beitrag zur Erstellung des Dictionary of Gāndhārī durch die Edition und begleitende Einarbeitung der Mehrzahl der noch unveröffentlichten Handschriften, durch die lexikographische Erfassung der Niya-Dokumente, durch die technische Weiterentwicklung der Datenbank des Wörterbuches und nicht zuletzt durch die Integration der Gāndhārī-Daten mit der Erforschung der Sanskrit-Literatur Gandhāras, die in München eine lange Tradition besitzt. ■

DIE AUTOREN

Dr. Stefan Baums ist wissenschaftlicher Mitarbeiter und Arbeitsstellenleiter des Projektes „Die frühbuddhistischen Handschriften aus Gandhāra“. Seine Hauptarbeitsgebiete sind Lexikographie und Grammatik der Gāndhārī sowie die Edition der scholastischen Werke.

Dr. Andrew Glass wurde an der University of Washington mit der Edition einer Gāndhārī-Handschrift promoviert und arbeitet als Program Manager bei der Microsoft Corporation. Seine Hauptarbeitsgebiete sind Lexikographie der Gāndhārī und Paläographie der Kharoṣṭhī-Schrift.

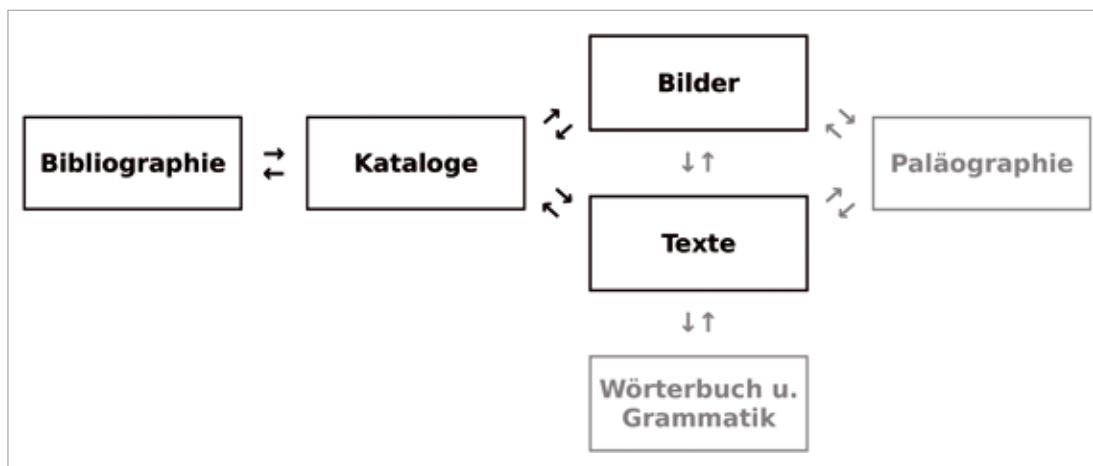


Abb. 5: Die Komponenten der Datenbank des Dictionary of Gāndhārī.

Quelltexte angestrebt, auf deren Grundlage dann mit Hilfe des neuen Datenbanksystems eine Arbeitsgrammatik der Gāndhārī erstellt werden soll, die der weiteren Gāndhārī-Forschung innerhalb und außerhalb des Vorhabens als wichtiges Werkzeug und Referenzpunkt dienen wird. Die enge Verknüpfung der Transliterationen mit den zugehörigen Photographien der Quelltexte wird (im Gegensatz zu unseren ersten Experimenten) nun auch auf Grundlage des TEI-Standards durchgeführt und soll den Bearbeitern der Quelltexte schon bei der Herstellung der Erstaussagen als praktische Hilfe dienen. Längerfristig wird die Text-Bild-Verknüpfung als Werkzeug für die Erarbeitung einer vollständigen Kharoṣṭhī-Paläographie (nach

Literatur und WWW

H. W. Bailey, Gāndhārī. In: Bulletin of the School of Oriental and African Studies 11, 764–797, 1946

S. Baums und A. Glass, A Dictionary of Gāndhārī, 2002 ff., <http://gandhari.org/dictionary>

A. Glass, A Preliminary Study of Gāndhārī Lexicography. In: B. Tikkanen and H. Hettrich (eds.), Themes and Tasks in Old and Middle Indo-Aryan Linguistics. Papers of the 12th World Sanskrit Conference, Helsinki, vol. 5, 273–303, Motilal Banarsidass Publishers, Delhi 2006