

# 明日の 東洋

Research and Information Center for Asian Studies (RICAS)  
Institute of Oriental Culture, University of Tokyo

December 2000 no.4

ヒンドゥー儀礼研究のための  
基礎資料  
永ノ尾信悟

アラビア文字文献総合  
データベースの構築

森本一夫

Tibetan-Sanskrit 構文  
対照電子辞書 eDic

鈴木隆泰



イラン、南ホラーサーン地方のゴナーバードにて。サフラン栽培を主な生業とするこの人懐こいおじさんは、夏の午後の  
闖入者を丹精したブドウでもてなしてくれた。(撮影:森本一夫 1998夏)

# ヒンドゥー儀礼研究のための基礎資料

永ノ尾信悟

## 1. コンピュータとのつきあい

インドのサンスクリット語の文字は英語のアルファベットより数が多いが、大文字と小文字の区別がないため、大文字、小文字を使うことでほぼローマ字に転写することができる。そのため、サンスクリット文献を用いるインド学の研究者の間では、程度の差こそあれ、コンピュータの利用がかなり普及している。

前任地の国立民族学博物館はかなり早くからコンピュータを導入していたこともあって、また近くにコンピュータに詳しい研究者も多くいたので、民博の山本泰典さんなどにいろいろ教わり、私も、比較的早くからコンピュータを利用し始めた。

最初にしたのは、古代インド祭式解釈文献を祭式規定の部分や解釈のさまざまな項目に従って分類しながら行うテキストの入力であった。また京都大学の井狩教授との共同作業として祭式文献のパウダーヤナ・シュラウターストラを入力し、個人的な仕事としてはマハーバーラタを少しづつ入力していた。その後、それら3つの作業を私は中断してしまった。パウダーヤナ・シュラウターストラのテキスト入力はその後、伏見誠氏により完成され、マハーバーラタのほうは、独自ではじめられた京都大学の徳永教授により完璧に成し遂げられた。その後、数多くのサンスクリット文献が京都大学の研究者を中心に入力され、その成果は現在世界の多くの研究者によって大いに利用されている。

## 2. CARDとPMANTRA

私の研究対象は古いインドの祭式儀礼である。その研究に必要なさまざまな情報を修士時代から図書カードを用いて蓄積してきた。ある儀礼がある。そこでまずその儀礼に関する先行研究になにがあるかをメモする。ついでその儀礼を記述するテキストの箇所を記録する。それらの記述に従うと、ある儀礼はある特定の期間に、特徴的な祭場で、複数の参加者がさまざまな供物や祭具を用い、一定の所作の連鎖として、散文、韻文の祭文を唱え

つつ行われる。さらにその儀礼の解釈はどのようなテキストのどこに、いかなる神話や獨特の論理を使いいいかに行われるかを記し、神話や解釈を構成する観念世界もまたさまざまな要素に細分していく。このような項目毎に集められた情報の全体を把握することが私にとり、ある儀礼を理解することであったし、現在でもそうである。しかし、ある儀礼テキストを冒頭から入力していくと、そのように細目化された情報を組織的に集めることはできないと実感するにいたった。ブラーフマナ文献の入力の際には、一応の細目化を行い入力したのであるが、カードで見る形にするには今一度整理しなおす必要があったし、事実、入力しつつ、必要事項は從来からの図書カードに手書きで書き留めていた。

民博のある時期から、何回かインドに滞在する機会があり、以前のヴェーダ祭式のみならず、ポスト・ヴェーダ期の宗教儀礼も研究の対象とするようになると、手書きのカードでは時間がかかりすぎると感じた。そこで、東文研に来る直前頃から、ヴェーダ祭式の情報は從来通り手書きのカードで集め、ポスト・ヴェーダ期の宗教儀礼に関する情報は、コンピュータにCARDというファイルを作り、手書きのカードと同じように入力し始めた。ヴェーダ祭式の中で使用される祭文に関してはM. Bloomfield の Vedic Concordance という便利な参考文献があり、それぞれの祭文の出典箇所をすぐに調べることができるが、ポスト・ヴェーダ期の宗教儀礼に用いられる祭文に関しては、そのような参考文献がない。そこで、ポスト・ヴェーダ期の宗教儀礼を主として記述する代表的文献であるブーラーナ文献の祭文、マントラのファイルという意味で PMANTRA というファイルを作成し始めた。ヴェーダ期からポスト・ヴェーダ期の儀礼の変遷を考えはじめると、両期の間に本質的な変容はあるが、要素的に共通する部分もあり、ふたつの情報を一方では手書きのカード、他方でコンピュータのデータとしていては不便を感じるようになり、ヴェーダ祭式の情報も新たに CARD に一括して集めるようにした。

それらのデータは NEC の PC-98 の古い型のコンピュータで MIFES を使って作成している。メモリーの容量も少なく、CARD の情報量が増えるにつれ、ファイルの中である行からかなりの行をジャンプするとき、突然機械が動かなくなってしまい、何度か、数時間の作業を無駄にしてしまうこともあった。また自宅のパソコンにデータをフロッパーで運ぶために、CARD の量もフロッパーに納まる必要があり、現在では CARD も CARD 11から CARD 42 までの10個に分割して利用している。その結果 CARD はこの10年間で約550万バイト、約3万8千行のデータになり、PMANTRA も約100万バイト、約1万行のデータになった。1行は最大で1500バイトである。

## 3. CARDの中身

CARD ではまず項目名を書き、タブで12文字までとび、そこから必要事項を書き記す。例えば農耕儀礼を意味する kRSikarma の項目を見てみる。kRSikarma は農耕儀礼の総称であるので、関連する他の項目を see によって列挙していく。

kRSikarma see aagrayaNa.

といった形式で、項目ごとに行を変え、いちいち see を書き、agriculture, biijanivaapana (種蒔き), halaabhiyoga (犁つなぎ), kRSyaarambha (耕作始め), laangalayojana (犁つなぎ), siitaayajna (畝の儀礼), vanaspatizaanti (木の鎮め儀礼), zunaasiiriiya (ある儀礼名)などの関連項目があることを記す。CARD ではかっこに入れた日本語は書かれていません。次に、農耕儀礼の研究、リグヴェーダやアタルヴァヴェーダの中の農耕儀礼に関する讃歌、農耕儀礼を組み込むヴェーダ祭式アグニチャヤナへの言及、農耕儀礼に関わる願望祭の情報を載せる。農耕儀礼の具体例としてカウシカストラから9例、5種類のグリヒヤーストラから1例づつ、グリヒヤーストラへの補遺文献から3例のテキストが画面に登場する。農耕儀礼の全体を見るには、さらに上記の関連項目を検索する必要がある。aagrayaNa (収穫儀礼) や種蒔き、犁つ

なぎなどを検索することにより、古代インドの農耕儀礼の大要を知ることができる。

農耕儀礼の関連項目のひとつ agriculture を検索すると、そこでの関連事項として fertilizer, irrigation, kRSi (農耕), kRSikarma, pazupaalana (牧畜) の事例が先ず登場し、それに続いて、古代インドの農業に関する文献がリストアップされる。irrigation を検索すると、サンスクリット文献に扱われる灌漑の記述の情報の前に、関連項目として taDaaga (ため池) を見よという指示がある。そこで taDaaga の項目をみるとため池の他のサンスクリット語への参照指示に続いて taDaagaadividhi (ため池などの儀礼) という項目が29行にわたって画面に登場し、25のテキストの箇所に記述される「ため池設置儀礼」に関して、テキストの箇所または儀礼の記述を見ることができる。

agriculture の関連項目のひとつ pazupaalana を見ると、agriculture や kRSi, kRSikarma へのクロスレファランスと並んで、関連事項として19種類の牧畜儀礼の名称が登場し、さらに画面をスクロールすると50行近くにわたって pazupaalana の項目として書き込まれた牧畜儀礼の情報を得ることができ、関連事項にあげられた19種類の牧畜儀礼をも調べ合わせることによって、古代インドの牧畜儀礼の大要を把握することが可能となる。このように kRSikarma の項目、あるいは agriculture、または何かの農耕儀礼、牧畜儀礼の名称を検索することによって、つぎつぎと関連項目をたどり、古代インドの農耕儀礼、牧畜儀礼といった、生業に関する儀礼や農業、牧畜に関するサンスクリット文献における言及、さらには、それらへの参考文献についての情報を得ることができる。

儀礼の要素的行為としては、例えば、儀礼の過程で何かを見つめるというものがあるが、gaze と検索するとavekSaNa, iikSaNa, aajyaavekSaNa, ghRtaavalokana, aadarza などといった関連項目にたどることができ、儀礼における「見つめる」という行為に関してかなりの情報を集めることができる。また、ヨーロッパの行法などでも知られる「調息」という行

為は、儀礼においても重要な役割を演じ、 praaNaayaama と検索をするとヴェーダ祭式における先行形からシュラウターストラ、グリヒヤストラ、プラーナ文献などにおける記述が一覧でき、テキストをほぼ比較年代順に並べているため、儀礼文献における調息の展開、変容をもみてとることができる。

## 4. CARDはデータベースか？

古いインドの儀礼に関する情報はいくつかのレベルに分けられる。また儀礼そのものの記述もある種の構造を示す。儀礼の全体、諸要素などを抽出し、あるいは私自身が、儀礼にはこのような構造と要素があると想定し、それを項目化し、相互連関を付け、作り上げつつあるのがこの CARD という資料集である。インドの儀礼の共同研究に参加している何人かの研究者にこの CARD をコピーし、利用してもらっている。しかし、大きな難点がある。この CARD の項目の設定はある意味では、私の頭の中の虚構の反映である。従って、作成者である私以外にはどのような項目があるのかわからない。また、作成者の私にも、1万行前後のころから、このような項目も作っていたのか、こんな項目すらなかっ

たのかと思うことが多くなった。ある友人は、あの CARD は4分の1ほどしか利用できないと語り、京都産業大学の矢野教授は、あれはデータベースではないよとおっしゃる。

インドの叙事詩、プラーナ文献の研究工具に H. v. Stietencron が中心となって1992年に出版された Epic and Purānic Bibliography (up to 1985), annotated and with indexes というものがある。メルボルンで開かれた学会で、出版されたばかりのこのビブリオグラフィーの紹介をかねた説明会が開かれた。その時、Stietencron 教授は索引の多くの項目に複数の関連事項を挙げているから、丁寧に索引を調べることによって必ずほしい情報にたどりつけるようになってあるといった内容の発言があったが、それが私にはとても示唆的であった。

たとえこの CARD が普通の意味でのデータベースではなくとも、少なくとも作成者である私には貴重な情報源である。かなりの情報は蓄積されているが、まだまだ不十分である。しかし、さらに情報を充実させ、もっと「かっこよく」して、いずれの日にか、「ヒンズー儀礼研究データベース」として、何らかの形で公開できることをめざしていくたい。

(東洋文化研究所教授)

# アラビア文字文献総合データベースの構築

森本一夫

## 1.

日本において、イスラーム地域を対象とする研究が日進月歩の勢いで進展しているという認識は、国内のみならず、国外においてもすでに広く共有されている。私の勉強している歴史学の分野においてもこの勢いは実感される。修士論文の段階から手稿本史料や文書

史料を用いた研究を行う者が増えるようになり、若手研究者の間でも早くから研究成果を外国語で発表する人が増えてきている。現在進行中の文部省科学研究費創成的基礎研究「イスラーム地域研究」は、このような動向に対しカンフル剤として働いている。

ところで、このような発展を基礎から支えるべき資料の蓄積と公開はどうなっているの

だろうか。残念ながら現状では、イスラーム地域現地語資料の国内における蓄積は、多くの分野においてきわめて不充分な状態にある。また、特定の資料が幸運にもどこかの機関に所蔵されている場合でも、個々の研究者がそれを知る術はきわめて限られていると言うしかない。今後は関係諸機関の連携のもと体系的な資料収集が実施されるとともに、効率的な公開方法が開発される必要があるだろう。

この文で紹介するアラビア文字文献総合データベース構築事業は、そのような課題に向けた重要な試みのひとつである。

## 2.

アラビア文字を用いて記される言語は、アラビア語、ペルシア語、ウルドゥー語、ウイグル語など、じつに多数にのぼる。では、このような言語で著された書物は日本国内の諸機関にどの程度所蔵されており、どのように整理されているのだろうか。ここに「イスラーム地域研究」研究班6(「イスラーム関係史料の収集」を課題とする)が1997年に行った調査の結果(『日本におけるイスラーム地域現地語資料の所蔵および整理状況の調査』)があるのでそれをもとに概観してみよう。ただし、この調査はアラビア文字文献という枠組みではなく、ラテン文字で書かれるトルコ語やインドネシア語などをも含む「イスラーム地域現地語資料」を対象として行われたものなので、数を述べるにあたっては一定の推測を交えざるをえない。

まず所蔵数であるが、これについてはおよそ100の機関に少なく見積もって10万冊程度が所蔵されていると考えられる。その分布状況に関しては、突出した所蔵数を持つごく少數の機関と、50冊以下というような小さな規模で所蔵する実に多数の機関が併存していることが見てとれる。

次に整理状況である。調査によるとイスラーム地域現地語資料を所蔵する機関の約8割がそれらの資料をなんらかの方法で整理している。ここでは問題を整理の際の使用文字種に限定して見てみよう。調査結果には、諸機

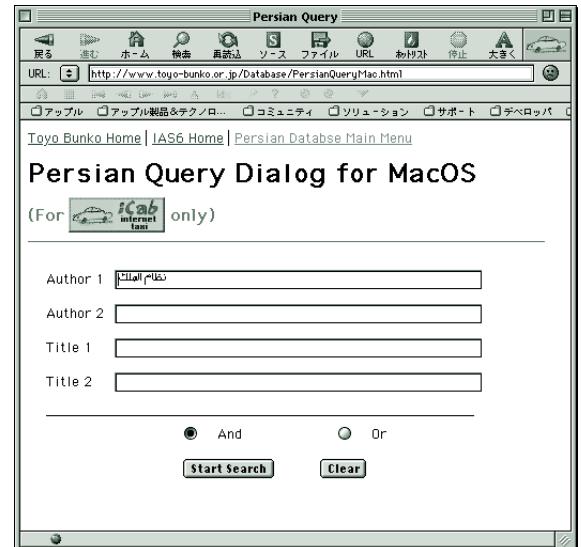
関のうち原綴を採用しているのが32%、翻字を行っているのが55.3%、「その他」が12.6%とある。アラビア文字による書物だけを考えるならば、翻字の比率はこれよりかなりあがるであろう。つまり、現在大部分のアラビア文字文献は翻字形で目録化されていることになる。その際の翻字方式は、米議会図書館方式(LC)や日本目録規則(NCR)が多いとはいえ、単なるローマ字化を行っているところもあり、雑多な方式がバラバラに行われていると言える。

上記の所蔵調査は、このような状況を明らかにした上で、今後のアラビア文字文献オンライン目録作成に関する提言として、専門スタッフなど必要な資源が足りない現状では、翻字で、あるいは並列タイトルを利用してOPACに登録するのが最上の策であるとする。しかし同時に、理想としては大規模所蔵館が中心となった、原綴でのデータベース構築が望まれるとも述べている。

## 3.

利用者の観点から述べるならば、翻字OPACは利便性において遠く原綴データベースに及ばない。かりに現在バラバラの翻字の方式が統一されたとしても、文献を検索するたびに面倒な転写を行わなければならないのはあまりに億劫である。所蔵が確認されている図書をさがす場合ならまだいいだろうが、あるかないかも分からぬ図書を複雑な転写を用いながら検索する場合、結果が「0件」と出ても、本がないのだから自分の転写が間違っているのだが、場合によっては目録の転写が間違っているのだとわからない。

さらに、翻字形式の目録の作成・利用に当たっては、アラビア文字で書かれた情報を「読む」という閑門が突破されなければならない。周知のようにアラビア文字の書記法では短母音が普通省略されるから、翻字データベースがきちんと機能するためには、目録作成者と利用者とに母音のふり方に關して共通



総合データベースのもととなる東洋文庫蔵書オンライン・データベース(ペルシア語)の検索画面。著者名欄に「ニザーム・アル=ムルク」と入力してみる。

の見解が分けもたれることが必要である。これは双方にかなりの熟練を要求することになる。「正しい母音のふり方」がかなりずしもひと通りとは限らない単語の場合、ことはさらに複雑になる。それに対し原綴でゆく場合には、目録作成者も利用者も、アラビア文字の30前後(言語によって若干異なる)の字母とそれに対応するキーボード上のキーを覚えさえすれば、資料に記されたままの書誌事項を打ちこむことが可能となるのである。

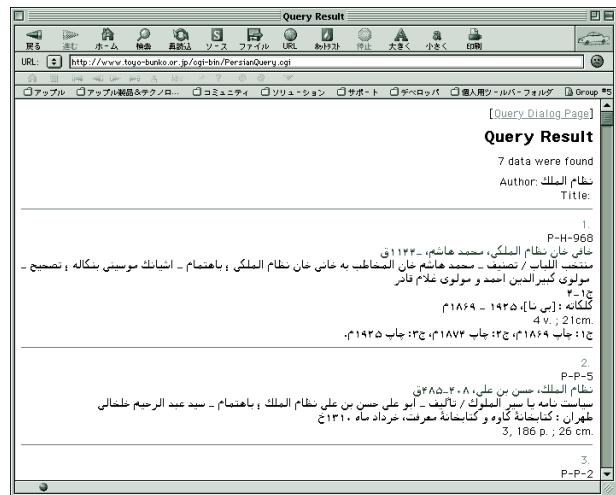
ここはぜひ、理想を追って原綴のデータベースを構築する必要がある。

## 4.

実は、上記の所蔵調査を行った「イスラーム地域研究」研究班6は、すでに理想を追って動きだしている。アラビア文字文献総合データベース構築事業が進行中なのだ。

研究班6が拠点を置く財団法人東洋文庫では、アラビア文字、チベット文字、簡体字、ハングルなど多文字種に対応するデータベースソフトの開発を行い、所蔵資料の原綴によるデータベース化事業をすすめているが、その一貫として本年、アラビア語、ペルシア語蔵書の原綴データベースのオンライン公開を試験的に開始した(Win, Macの両環境に対応<http://www.toyo-bunko.or.jp/>)。これを基礎に総合データベースを作成しようというのが事

# 明日の 東洋学



前画面の検索の結果、7件がヒットした

業の大筋である。

この構想は、なにも原綴の利便性に惹かれた研究者たちが周囲の事情を無視して突っ走ったというものではない。翻字目録との関係について言えば、国立情報学研究所や東京外国语大学附属図書館との協議の結果、原綴データベースと国情研の運営する総合図書目録(NACSIS-CAT: 翻字による)との間での双方のリンク実現が技術的に可能であるとの見通しが立っている。それどころか、将来的に総合図書目録がアラビア文字をサポートするという可能性までもが、かなり現実的な問題として議論されているのである。このような準備を経て昨年12月20日には、第1回の「アラビア文字文献データベース連絡会」が開かれ、この構想が関係の諸機関に説明された。

この構想によって構築される総合データベースが広く多くの機関に受け入れられてゆくひとつの重要な条件として、データベースに含まれる個々のデータが規範として用いられるほどに正確であり、件数も豊富であることが挙げられよう。したがって、事業の初期に、所蔵資料数が多くアラビア文字資料に関する専門的知識を持ったスタッフの動員が可能な諸機関が積極的に参画し、データを出しあってゆくことが重要になる。東洋文化研究所は、そのような機関のひとつとしてこの事業の重要性を認識し、早くから協力してきた。アラビア語単行本、ペルシア語単行本、オスマン・トルコ語逐次刊行物に関しては、すで

に所蔵目録の、東洋文庫データベース・フォーマットに依拠した入力を完了している。現在それらのデータは研究班6にフィードバックされ、東洋文庫・東洋文化研究所蔵書の本邦初の「アラビア文字文献原綴合併データベース」作成作業が進んでいるのである。

## 5.

もちろん、総合データベース完成への道のりはまだ

まだ遠い。他の大規模所蔵機関の参加を得ることは必須であるし、「イスラーム地域研究」が終了する2002年3月以降の事業主体を定めることも必要になる。しかしたとえば、多数にわたる小規模所蔵機関の協力を募るために希望機関に人材・機材・技術を提供する「出張データベース作成サービス」が企画されるなど、建設的な試みがたゆむことなく続けられている(連絡先: 研究班6 <IAS6@toyobunko.or.jp>)。「ミミズの這い跡」と不当に馬鹿にされることの多いアラビア文字が方々のコンピュータ端末で自在に踊りこねる日を心に描きながら、この事業は根気強く進められているのである。(東洋文化研究所助手)

# Tibetan-Sanskrit構文対照 電子辞書 eDic

鈴木 隆泰

## 1. 序言

「Tibetan-Sanskrit構文対照電子辞書」(以下、eDic)<sup>1)</sup>に関する構想・システムは、かつて1999年度東京大学東洋文化研究所班研究「インターネット利用技術」研究会(2000年1月7日)にて提唱、報告し、その資料をWebサイトで公開している<sup>2)</sup>。ただし研究会の性格上、その際の発表及び資料は、テクスト操作の詳細ならびに情報理論に基づく専門的な内容を含むものであったため、当該分野の研究者以外には少々理解しづらい面があったようである。今回、『明日の東洋学』に執筆機会をえていただいたこともあり、研究会以降の成果も織り交ぜながら、できるだけ一般的な立場からeDicについて述べさせていただくことにする。

## 2. eDicの背景

インド仏教を文献学の立場から研究しようとする場合、インド語、とりわけ広義のサンスクリットの資料が第一次資料となる。しかし、特に大乗仏教の典籍に関して言えば、原典であるサンスクリット資料の多くは散逸し、結果、チベット語や漢語を代表とする翻訳資料(チベット訳・漢訳)を用いた、異訳対照比較に基づく研究が不可欠となる。しかも翻訳資料を用いながら原典レヴェルでの考察を行うためには、チベット語・漢語をサンスクリットの次元で捉える必要が生じてくる。この、翻訳語を原語であるサンスクリットの次元で捉えるという観点に立つとき、チベット訳資料が最も適していることについて異論はないだろう。チベット語とサンスクリ

ツト語の間には、『翻訳名義大集 *Mahāvyutpatti*』という訳語集が存在しているのみならず、そもそも古典チベット語自体が、サンスクリット仏教文献を翻訳するために整理されたという歴史を抱えているからである。

もっとも、チベット語とサンスクリット語との間に上記のような工具や歴史があるとは言え、それだけで翻訳されたチベット語の全てがサンスクリット語に対応させられるわけではない。そのため、サンスクリット・チベットの両資料が現存しているテクストに関する独自のTibetan-Sanskritインデックスが各種必要とされ、さらに、それらを統合するかたちで、いわゆる“藏梵辞典”も数種類編纂されてきた。

以上のような経緯、そして先学たちの大変な努力にも関わらず、チベット資料をサンスクリットの次元で捉えるための工具は、残念ながらいまだに不十分である。その原因を考えるとき、大きくは以下の二点に集約されることになる。

(1) サンスクリットの現存している資料が少ないため、Tibetan-Sanskritインデックス・辞書も、必然的に限られた資料を用いたものに留まらざるえないこと。

(2) 現在利用しうる、ただでさえ少数のインデックス・辞書のほとんどが、単語対単語の対応関係を示すことに終始し、用例を併記していないこと。

まず(1)については、散逸していたサンスクリット原典が大量に発見されるということでも起こらない限り、今後劇的な改善は望めそうもない。しかも、新たな原典が大量に発見される可能性は非常に低いというのが、大方の一一致した見解である。次に(2)についてであるが、インデックスは置くとして、そもそも「用例のない辞書」というものが果たして「辞書」として正しく機能しうるのかという自明の問題がある。その点、用例を併記しているJ. S. Negiの辞書は評価されるものの、資料の総体から見て、その用例と収録語彙の不足は否めない。それに関わらず、Negiに

続き用例を豊富に併記した新たなな「藏梵辞典」を編纂しようという動きは寡聞にして未詳である<sup>3)</sup>。

### 3. 電子データの普及

2に掲げた背景とはある意味で別の脈絡の状況として、近年ではチベット資料やサンスクリット資料が電子データ化され、テクストデータベースとして普及しつつある。検索する上で便利なことこの上ないことは事実であるし、ロケーション情報も入っていれば、そのデータ内でのインデックスを作成することも即座に可能である。ただ、惜しむらくは、多大な労苦を費やして作成されたデータが、「検索」や「データ内部でのインデックス作成」あるいは「論文作成時に貼り付ける」以外の目的に用いられることはほとんど聞かれず、他のデータとのリンクも行われてはいないようである。筆者は以前より、「テクストデータベースは作ったが、その後どうする」という問題に対して、何らかの解答を提示することはできないだろうかと考えていた。

### 4. eDic初期構想

前述の2, 3を総合すると、「同一テクストのサンスクリット・チベット双方が利用可能である電子データを利用して、構文ごとに対照させる」という道筋が浮かび上がってくる。「既存のデータを利用する」という点で3に対する解答が、そして「構文ごとに対照させて用例を明示する」という点で2の(2)に対する解答が導かれるからである。以下に手続きと例を挙げてみよう。

#### [eDic初期構想の手続き]

- (1)まず、同一テクストのサンスクリット・チベット双方の電子データを用意する。
- (2)次に、双方を構文ごとに切り分ける。
- (3)チベット語を検索キーとして入力すると、その文字列を含むチベット語



eDic トップページ

の構文と、対応するサンスクリット語の構文を出力する。

#### [実例]

用意されたデータを『法華經』(*Saddharma-puṇḍarīka*)の第十章「法師品」(*Dharma-bhāṣyakāparivarta*)検索キーを「bcom Idan 'das」<sup>4)</sup>とした場合、最初にヒットする構文は以下のようになる<sup>5)</sup>。

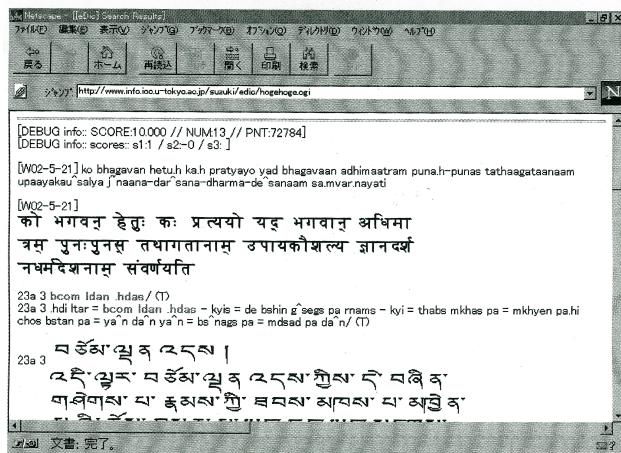
[Tibetan] de nas bcom Idan 'das kyis bya^n  
chub sems dpa' sman gyi rgyal po la  
sogs pa bya^n chub sems dpa' brgyad  
khri po de dag la bka' tsal pa/  
[Sanskrit] atha khalu bhagavaan bhai.  
sajyaraaja.m bodhisattva.m mahaasattvam  
aarabhya taany a'siiti. m bodhisattvasahasraa.  
ny aamantrayate sma/

極めて単純な発想ながら、チベット語「bcom Idan 'das」に対応するサンスクリット原語「bhagavān (bhagavatのnom. sg. m)」が、用例を伴った上で抽出された。上記の手続きに従う限り、データさえ増やしていくれば労力を大幅に節約しながら、用例を伴った「藏梵辞典」を作成することが可能となる。

### 5. 散逸した原典の論理的再構成へ向けて

さて、4で提示した方法論によって、「藏梵辞典」の作成という目的が達成されること

# 明日の 東洋学



bcom ldn das をキーとした検索結果の一例

が確認された。豊富な用例を伴った「蔵梵辞典」の作成は、筆者自身を含め、インド仏教学研究者にとって大きな利益をもたらすものであるため、本来ならばこの段階での作業を行っただけでもよかつたのかも知れない。しかし、結局われわれは、さらに次の段階を目指して作業に着手することになった。その理由の一つには、メンバーの一員であり情報科学研究者である相場氏にとって4の手続きは魅力が少なかったことがあり、そして一つには、2の(1)に対する解答が依然として与えられないということが挙げられる。では、その「さらに次の段階」とは何かと言えば、章題に示した「散逸した原典の論理的再構成」である。

インド大乗仏教の典籍のうち、原典が残存しているものは相対的に少数で、大多数の原典が散逸していることはすでに述べたとおりである。したがって、後者に属する典籍を扱うとき、研究者は必然的に原語を想定しながら翻訳資料に向い合うことになる。その際、原語を想定する工具がいまだに不十分であるというのが、本研究の出発点であった。そして、その工具を飛躍的に進化させる方法自体はすでに手に入れることができた。しかしながら、その新しい工具を使ったからと言って、対応させられた構文同士は同一テクストの原典と翻訳であるため、翻訳のみ存在する異種テクストとその散逸した原典との距離は依然

として縮まっていないことに気づくだろう。したがって、その距離を縮めるためには、翻訳資料しかないテクストをもとに、それと「似通った」構文を原典の存在する別テクストの中から抽出してこなければならぬのである。

ここで言う「似通った」とは、単なる単語同士の共通性のみを指すものではなく、「文章の構造」を含めた共通性、相似性も含まれる。これを実現するために

は、切り分けた構文にさらに「単語情報」を追加する必要があり、われわれの作業も、この単語情報の追加に多くの時間をかけている。

単語情報の加え方や、この方法論に基づく再構成過程の原理自体は先に紹介したWebドキュメントに記してあるので、ここでは実例のみを示すことにしよう。まず、(a)のような原典未詳のチベット文があったとする。

(a) nam ^zig ru sbal spu rnams las// gos su legs par btags gyur te// dgun cha dag tu gyon gyur pa// de tshe ri^n bsrel yod par 'gyur//

この(a)を検索キーに指定した場合、予め作成しておいた『金光明経』のデータ(サンスクリット・チベットとともに現存)に存する以下のチベット文(b)が、「nam ^zig」「spu - rnams - las」「gos - su」「legs par btags」「dgun」「gyur - te」「de tshe」「ri^n bsrel」「yod par 'gyur」の九要素、及びその順序がマッチしているためヒットし、同時に(b)に対応する原典(c)も出力される。

(b) nam ^zig = rus sbal = spu - rnams - las// gos - su = legs par btags = gyur - te// dgun - gyi = gra^n ba = sel byed pa// de tshe = ri^n bsrel = yod par 'gyur// (SuvT, 11: 14.12-15)

(c) yadaa kacchapa-lomaanaa.m praavaara.h suk.rto bhavet/ hemante ^siita-hara.nas tadaa dhaatur bhavi.syati// (SuvS, 11: 15.5-6)

さらに(c)に基づいて追加検索が行われ、

パーリ文(d)が出力される。

(d) yadaa kacchapa-lomaana.m pavaaro tividho siyaa/ hemantika.m paapura.na.m atha nuuna tadaa siyaa// (Jaataka 425, 78: 477.16-17)

そして研究者はこれらの検索結果を総合して、(a)に対応するサンスクリット文を(e)として論理的に再構成することが可能となる。

(e) yadaa kacchpalomaanaa.m praavaara.h suk.rto bhavet/ hemantaka.m praavara.na.m tadaa dhaatur bhavi.syati//

上記の実例を見ても分かるように、eDicの精度を上げるためににはデータの蓄積が不可欠となっている。幸い、データ量は増加しているものの、まだまだ人手が足りないのが偽らざる実状である。一人でも多くの方にこのプロジェクトに参加していただけるよう、ご協力を願う次第である。

(東洋学研究情報センター助手)

1)<http://www.info.ioc.u-tokyo.ac.jp/suzuki/edic/>

2)<http://www.vacia.is.tohoku.ac.jp/member/aiba/project/edic/pr/document/20001.ioc/genko.html>  
執筆者 鈴木隆泰(東京大学東洋文化研究所)  
相場徹(東北大学大学院情報科学研究科)  
松本峰哲(東北大学大学院文学研究科)

3)ここまで、インド仏教学専攻の研究者にとって周知のことがらであろう。しかし、これからeDicについて述べていくに際して現状を再確認しておくことは、前提条件を提示するための重要な手続きとなっている。

4)サンスクリット語のbhagavatに対応するチベット語である。漢訳では一般に「世尊」とされ、仏陀に対する尊称の一つである。

5)実際はロケーション情報を付帯して出力される。なお、転写方式には「相場式」を使用した。転写方式の諸相については、拙文「インド語・チベット語の処理とデータの互換性」(『全国文献・情報センター人文社会科学研究情報セミナーシリーズ』#9, 1999所収)を参照されたい。

#### 所外委員

|                      |                             |
|----------------------|-----------------------------|
| 落合 隆四郎               | 附属図書館長、大学院数理科学<br>研究科・理学部教授 |
| Chen, Paul Heng-Chao | 大学院法学政治学研究科・<br>法学部教授       |
| 河原 秀城                | 大学院人文社会系研究科・<br>文学部教授       |
| 泉田 洋一                | 大学院農学生命科学研究科・<br>農学部教授      |
| 中兼 和津次               | 大学院経済学研究科・<br>経済学部教授        |
| 黒住 真                 | 大学院総合文化研究科・<br>教養学部教授       |
| 田嶋 俊雄                | 社会科学研究所教授                   |
| 小林 宏一                | 社会情報研究所教授                   |
| 鶴田 啓                 | 史料編さん所助教授                   |

#### 所内委員

|       |                    |
|-------|--------------------|
| 岡本 サエ | 教授 汎アジア部門          |
| 松井 健  | 教授 汎アジア部門          |
| 平勢 隆郎 | 教授 東アジア研究部門(第一)    |
| 甘 懐真  | 教授 東アジア研究部門(第一)    |
| 小川 裕充 | 教授 東アジア研究部門(第二)    |
| 永ノ尾信悟 | 教授 南アジア研究部門<br>委員長 |
| 鎌田 繁  | 教授 西アジア研究部門        |
| 中里 成章 | 教授 センター造形分野        |
| 宮嶌 博史 | 教授 センター文献分野        |
| 板倉 聖哲 | 助教授 センター造形分野       |

#### センター長

|       |         |
|-------|---------|
| 原 洋之介 | 教授、研究所長 |
|-------|---------|

#### センターのスタッフ

|                   |                                |
|-------------------|--------------------------------|
| 原 洋之介 (はら ようのすけ)  | センター長・<br>東洋文化研究所長。東南アジア経済。    |
| 中里 成章 (なかざと なりあき) | センター主任・<br>造形資料学分野教授。南アジア近現代史。 |
| 宮嶌 博史 (みやじま ひろし)  | 比較文献資料学分野教授。朝鮮近代史。             |
| 板倉 聖哲 (いたくら まさあき) | 造形資料学分野助教授。東洋絵画史。              |
| 鈴木 隆泰 (すずき たかやす)  | 比較文献資料学分野助手。仏教学。               |
| 深見奈緒子 (ふかみ なおこ)   | 客員教授。イスラーム建築史。                 |
| 佐々木郁子 (ささき いくこ)   | 業務掛長。                          |

#### 明日の東洋学

東京大学東洋文化研究所附属東洋学  
研究情報センター報 第4号

発行日 2000年12月25日  
編集・発行 東京大学東洋文化研究所  
附属東洋学研究情報センター  
〒113-0033 東京都文京区本郷7丁目3番1号  
電話 03-5841-5839(直通)  
FAX 03-5841-5898  
ホームページ  
<http://www.info.ioc.u-tokyo.ac.jp/>

## センター便り

### 平成12年度漢籍整理長期研修

今年も漢籍整理長期研修が6月26日から7月7日と、9月26日から10月6日の2期に分けて実施された。受講者数は6名であった。本年度は新しい試みとして、南アジア文献学概説および西アジア文献学概説という講義が設けられた。講師をお引き受けいただいたのは、東洋文化研究所スタッフ8名の他、山城喜憲(慶應義塾大学) 小島毅(東京大学文学部) 井波陵一(京都大学) 大塚秀高(埼玉大学) 高橋孝信(東京大学文学部) 中善寺慎(東洋文庫) 長澤孝三(国立公文書館) 横山謙次(宮内庁書陵部) 持井康孝(金沢大学) 藤本幸夫(富山大学) の諸氏である。今年の講習会が受講生に好評だったのは、これらの方々のご努力によるところが大きい。ここに記して謝意を表したい。

### 第36回全国文献・情報センター長会議

全国文献・情報センター長会議は、東京大学法学部附属外国法文献センター、一橋大学経済研究所附属日本経済統計情報センター、神戸大学経済経営研究所附属経営分析文献センター、京都大学人文科学研究所附属漢字情報研究センターに、当センターを加えた5つのセンターの長から構成される。その36回目の会議が、来賓に文部省学術国際局学術情報課の濱田幸夫大学図書館係長を迎えて、11月7日、東京大学法学部で開催された。事業報告に続き、将来構想や競争的資金の獲得等の問題について活発に意見が交換された。

### 平成12年度全国文献・情報センター人文社会科学学術情報セミナー

平成12年度全国文献・情報センター人文社会科学学術情報セミナーが、11月20日から22日の3日間、東京神田の学術総合センターで開催された。このセミナーは平成7年度から毎年5センターが共催しているものである。今年度は「アジア情報学のフロンティア」という全体テーマの下に、「インターネットの法律情報」(外国法文献センター担当) 「漢字と情報」(漢字情報研究センター担当) 「造形資料学のフロンティア」(東洋学研究情報センター担当) 「企業情報のフロンティア」(経営分析文献センター担当) 「経済統計について」(日本経済統計情報センター担当) の5つのセッションが設けられた。参加者数は20日96名、21日66名、22日28名であった。本センターは当番校として、セミナーの運営を担当した。

### アッシャー博士講演会

11月18日午後2時から東洋文化研究所大会議室で、キャサリン・アッシャー(Catherine B. Asher) ミネソタ大学美術史学科助教授の講演会が行なわれた。この講演会は、現在センターが進めている、インド・イスラーム史跡調査団の写真資料のデータベース化プロジェクトに関する行われたものである。アッシャー助教授はインド・イスラーム建築史の専門家で、「Building Authority: Architectural Programs and Changing Royal Ideology in Late Sultanate and Mughal India」と題して1時間余り講演した。参加者数はおよそ30名であった。