

漢字と文化

漢字文化の全き継承と発展のために

京都大学 21 世紀 COE 東アジア世界の人文情報学研究教育據點

第 12 号



目 次

漢字文化継承についての気懸かり	2
インターネット時代の人文学の技術	5
唐代研究ナレッジベースの構築	8
東アジアの文字に関する人文情報学的研究	12
表紙掲載写真解題	14

漢字文化繼承についての氣懸かり

高田時雄

本 COE プログラムは「漢字文化の全き繼承と發展のために」というスローガンを掲げて出發した。もとはといえばコンピュータ社會の中で却って使い勝手の悪くなった漢字まわりの環境を改善し、ひいてはコンピュータによる利便性を最大限に活用して漢字文化をさらに豊かなものにしたいという考えであった。コンピュータ上で漢字を驅使するためのさまざまな便利なツールの開發や、コードに依存しない新しいコンセプトの提示、將來の擴張性に富んだ野心的なナリッジベースの構築、さらには國際標準への寄與など、そういった方面ではこの五年間に一定の貢獻は爲し得たと考えている。ただ五年間の事業をひとまずは終えようとしている今、當初から氣懸かりであり、結局はどうすることも出来なかった事柄が一つある。今後の課題として、その氣懸かりをここで取り上げておきたいと思う。

それは他でもなく「全き繼承」の對象となるはずの日本における漢字文化の現状である。いかにも日本では漢字が使用されているし、學校でも漢字を教える。しかしその漢字を使いこなす能力は目に見えて劣化しつつある。とくに若者の漢字力の低下は著しく、年々ひどくなる一方である。このままでは日本の漢字文化は先が思いやられるという聲がしきりに聞こえてくる。教育現場でも先生方は漢字を教えるのにいろいろと苦勞をされているらしい。なぜこのような漢字力の低下が起こっているかといえば、おそらく読書に費やす時間が昔に比べて壓倒的に少なくなっているからに違いない。漢字の知識は、もちろん學校で習う部分も重要だろうが、それ以外に自主的な読書によって自然と身につくものだからである。文章を離れての漢字はあり得ないし、漢字の能力は文章のコンテクストから自ずと培われる。もちろん今日若者の手にする書物の中にどれほどの漢字が使われているかということも問題で

ある。どうも書物の版面が筆畫の多い漢字で黒っぽくなるのは嫌われるらしい。今日では書物の印刷にも漢字を避けて、できるだけ假名に「開く」というのが多くの編集者の好みのようなのだ。しかしこれは當面問わないことにしよう。昔に比べれば漢字が少なくなったとはいえ、読書に時間を費やしてくれるのであれば、状況は大きく異なるだろう。

しかし翻って考えてみれば、この國ではその氣になれば漢字を使わずに假名で書けるというのも事實である。これはなんとも有り難い條件といえそうだが、ではすべてを假名で書けるかということ、それは中々難しい相談と言える。古くから「カナモジカイ」などというものがあって、漢字の廢止が主張されているが、いっこうに實現しないのはやはりどこかに無理があるからである。漢字と假名の併用というのが効率的でしかも長年にわたって培われた、日本語に適した表記のあり方なのである。やみくもに漢字を減らそうとせず、社會全體で効果的な漢字の使用を促進するようにしたい。マスコミや出版社といった書物の作り手にも要望したいし、読者のほうでもぜひ声を上げてもらいたいと思う。

そもそも古代に漢字を用いて日本語を表記し始めた時には、萬葉假名や宣命書きなど幾つかの試行錯誤があった。漢字という古代中國語のために作られた癖のある文字では、まったく構造を異にする日本語の表記に不便な點が多々あったことは當然である。長きにわたる創意工夫が現在の漢字假名交じり文を作り上げた。しかしどの語を漢字で書き、どの語を假名で書くかというきまりは必ずしもあるわけではないから、ややもすると本來漢字語であって漢字で書くのが相應しいはずのことばも、すぐに漢字が思い浮かばない時など、ままよとばかり假名でかいてしまうということは日常多々ある。したがって日本人には一體にどうしても漢

字表記を墨守するという緊張感が欠如していることは否めない。

一方、中国ではどうかというと、こちらは何もかもすべて漢字で表記しないといけないので、大變である。そのため新しい言葉もたいていはすかさず漢字に翻譯される。ショートメールは短信、セクシャル・ハラスメントは性騷擾だし、ドーピング・テストは禁薬検査である。こういった言葉をすべて音譯していたのでは、中國語としてほとんど機能しなくなるだろう。ただし人名や地名などの固有名詞は當然音譯する他はなく、ヒラリー・クリントンは希拉里・克林頓、バラク・オバマはバラク・奥巴马などと書かれる。地名の場合はなかなかうまく作られたものが多く、紐約（ニューヨーク）や巴黎（パリ）は單純な音譯だが、牛津（オックスフォード）、劍橋（ケンブリッジ）、新西蘭（ニュージーランド）など、意味を翻譯したものや音と意味を兼ねたものなどがあるのはご存知のとおりである。地名は人名よりは固定性が強いので、こういうことが起こるのだろう。いずれにせよ中国では漢字使用の傳統は確固として揺るぎないように思われる（簡體字の問題はちょっと別なので、いまは觸れない）。ところが日本ではこういった外來語はとりあえずカタカナ表記することに決まってしまうようで、新聞など時事的な内容になるといよいよ漢字は少なくなる。日露と書けば良いものを、この頃はほとんど日ロと書かれるのは、ソ連時代の日ソが影響しているのかどうかは知らないが、それなら日米も日アにしたらどうかと思いたくなる。いずれにせよ漢字の使用頻度が少なくなるのには、假名という便利な表音文字の存在が大きく関係している。日常眠にする漢字の量が減れば減るほど、人々の漢字力は低下していくのは當然である。

本來、漢字には獨得の造語力があって、個々の漢字を組み合わせることで新しい語彙を創造していくことができる。幕末から明治にかけて、われわれ日本の先輩たちがヨーロッパの書物を翻譯する際、漢字を用いて新しい概念を鑄造することが出来たのは、それを可能とする高度な漢字の能力を備えていたからに他ならない。それらの新語の多くは漢字語としてもきわめて自然であり、したがって中国にも受け入れられた。學校、教育、銀行、病院、政治、經濟などなど、中國語と日本語で共通の言葉を用いているのは、そういう近

代日中間の語彙交流史が背景にある。いまそういった能力を日本人に見いだすことは難しいし、反對に中國で作られた漢字語を容易く受け入れるだけの素地は、カタカナ文化の榮える日本にはあまりありそうもない。こうして徐々に日中間で語彙の乖離が始まっている。

しかし實は漢字力の低下は明治の頃にすでに現れていたらしい。桑原隲藏が明治三十五年に『教育學界』という雑誌に寄せた「漢字に就きて」、明治三十八年に同じ雑誌に發表した「教育管見」などという文章を見ると、當時の中學生の漢字力がすでにかなり好い加減であったことがわかる。桑原は高等學校入學試験の答案に見える珍答の例を幾つも挙げて警鐘を鳴らしている。そのうち幾つかを挙げてみると、泰斗の泰を秦とまちがえてシンと讀んだり、緩急の緩を暖とまちがえてダンと讀み、媛とまちがえてエンと讀むものが多數あったこと、また「血あり涙ある日本男兒」と書くべきところを「血あり尿ある日本男兒」と書いたり、「大和魂」を「大和鬼」と書いたりしたものがあったというのである。それから約百年後の現在、漢字力の状況はどうであろうか。百年という時間の推移を勘案すれば、劣化の度合いはまだマシだと考えるべきか、いやいや現在の状況はそんな生やさしいものではないと考えるべきか、意見の分かれるところであろう。桑原は漢字の知識のなさが讀書力の低下をまねいてるとし、その弊害をただすための方法として、教科書の内容を輕易にすること、或いは學生の讀書力を増進することの二つを挙げ、文部省當局者の方針は主として前者によったために失敗したもので、須らく讀書力の増進を目指すべきだとする。最近もどこかの國で聞いたことのあるような話である。桑原の意見にはまったく同感で、今も昔の讀書の習慣こそが大事なものは變わらない。

ところで今日の日本では、漢字檢定が結構人氣だったり、毎年暮れになると「今年の漢字」などというのが話題になったりする。漢字が持てはやされているのではないかという向きもあるかも知れない。しかしこれらはむしろ明らかに漢字の衰退を映している以外の何ものでもない。漢字の能力が一般にたいへん低いからこそ、檢定などというものが行われるのである。普通の日本人は英語を知らないからこそ、英語檢定などというものがあるのと同じである。思えば、百數十年の

あいだずっと歐米のほうを向いてひた走りに走ってきた日本人である。今になって漢字への愛着を無理強いすることは困難であろう。

しかしこの國には文學や歴史、宗教、制度、科學など、あらゆる分野にわたる、それこそ汗牛充棟の文獻が遺産として存在し、それを受け継いで行かねばならないのも事實である。それらはみな大量の漢字を含んだテキストとして保存されてきた。漢字を知らずしてどうしてそれを保存できるだろう。保存と研究は一部の専門家に委ねて、その中の精華だけをわかりやすく書き換えてもらえばいいという意見もあるかもしれない。しかし専門家と一般人のあいだに境界を設けて、一般人は専門家によって提供されるものだけで自足するような状態が自國文化の傳承のあり方としてはたして健全であろうか。もちろん程度の差はいろいろあるに違いない。しかし誰であっても少しく努力すれば如何なる文獻にも手が出せるような条件だけは確保して

おくべきだと思う。したがって漢字を放棄することは出来ないし、漢字使用を繼續するには一定水準を保つ努力が必要である。

時代は変わる。歐米一邊倒の時代はすでに過去のものとなりつつある。やがて最新の文化が中國から發信される時代が來ないとも限らない。というよりも長期的に考えれば、その可能性はかなり高いだろう。その時、若者たちはそれこそ熱烈に漢字文化に傾倒し、漢字語が日本語の中にも滔々と流入することになるかも知れないのである。しかしその時すでに日本が漢字を手放してしまっていたとすれば、それはもう手遅れである。漢字文化の現状についてはたしかに氣懸かりな點が多い。出来るなら今のうちに日本の漢字力の挽回を圖っておきたい。もしそれが不可能だとしても、最低限何とか今の水準を維持しつつ、新しい時代を迎えたいものだ。

インターネット時代の人文学の技術

武田時昌

COE 教育部門では、東アジア人文情報学の学問的確立を目指して、人間・環境学研究科の関連分野の人文学研究スタッフとの協力のもとに高度専門カリキュラムの開発に取り組んだ。実際に推進した中心的な事業は、人間・環境学研究科の大学院生を対象として行った東アジア人文情報学サマーセミナー、東アジア人文情報学研究会の開催であった。

東アジア人文情報学サマーセミナーは、2004-06年度の3年間、夏休み期間中の9月上旬に開催した5日間のセミナーである。多言語処理やデータベースに関する情報学的手法を人文学研究に取り入れるうえで、最も基礎的であると思われる技法について、理論的な概要を講義した後で、そのパソコン実習を十分な時間をかけて行った。総合テーマは、「インターネット時代の人文学の技術」、各年度で設定したテーマは、2004年度は「TeX から XML へ」、2005年度は「XML の世界」、2006年度は「検索の理論と技法」である。

受講者は、徹底したパソコン実習指導を行う必要があるため、10名程度に限定した。実際のセミナー参加者は、およそ半数が人間・環境学研究科の大学院生で、残りの半数が中国学、仏教学を専攻する文学研究科、他大学の大学院生及び漢籍や情報ネットワークを担当する図書館職員等であった。

少数精鋭であったが、受講者の関心度が高く、熱心に取り組んだので、大いに教育効果はあったように思われる。また、その方面で活躍されている人文学者である小口雅史法政大学教授、坂内栄夫岐阜大学准教授から参加の申し出があり聴講者として加わったことが、人文情報学の講習会が稀有な企画であり、それなりの反響があったことを物語る。受講後の感想としては、ほぼ全員が XML 文書でのデータベース作成に強い関心を抱き、正規表現等の検索の手法もさらに学びたいという意欲を示した。とりわけ、岩井茂樹教授（人文

科学研究所）が共同研究班の会読テキストである『元典章』を XSLT を用いて XML から HTML へ変換したり、坂内千里准教授（大阪大学）が『説文解字繫伝』を画像の位置情報を内包させたデータベースを作成したりする実践例には大いに興味をそそられたようだ。

しかしながら、三年間を振り返ってみて、人文情報学の若手研究者を養成できたかという点に心許ない結果に終わった。というのも、大部分がパソコンをワープロでしか使用していない初心者であり、プログラミングの初歩的な基礎であってもハードルが高く、理解不足は否めなかった。情報学関連の基礎知識がほとんどない学生に、短期間のセミナーで情報学的なスキルを習熟させるには限界があることを痛感した。

もっとも、本セミナーは東アジア人文情報学の授業カリキュラムの策定や教材開発のパイロット・プロジェクトを兼ねた試みであった。その意味では、2006年度より人間・環境学研究科共生文明学専攻において正規カリキュラムに組み入れた授業を新設することができたことは、大きな成果である。開講した授業題目は「東アジア人文情報学」、前後期それぞれ二コマの開講で、授業担当者は安岡孝一漢字情報研究センター准教授、C. ウイッテルン漢字情報研究センター准教授である。なお、漢字文化の継承と発展という本 COE の課題を充足するために、人文科学研究所の宮宅潔准教授が漢字文化・書写材料についての授業を、「東アジア人文情報学」として開講したことも附言しておく。

もう一つの事業である人文情報学研究会は、人間・環境学研究科の大学院生及び COE 事業推進者を中心とする指導教員を構成員として立ち上げた研究組織である。東アジア人文学の研究に従事する学生間の研究情報交換の場を提供し、研究の向上を目的とし、2005年度からスタートした。運営の主体となる大学院生は

約20名であり、毎月1回の月例研究会で、1、2名が担当して各自の研究テーマを発表し、人間・環境学研究科の指導教員、事業推進者も交えて、3時間程度におよぶ活発な討議を繰り広げた。そして、その研究成果を『漢字文化研究年報』としてまとめ、これまでに2冊を公刊し、さらに最終年度末までに後1冊を刊行することになっている。

研究報告のテーマを列挙すれば、甲骨文字、篆書、契丹文字、唐韻、陶淵明、唐詩、唐代伝奇文学、民衆版画、古事記、満州族神話、後漢末・三国時代の益州豪族、浙江省の族譜群、日唐外交史、近世社会の女性観、近代中国の衛生学、在華ユダヤ人、在満朝鮮人、台湾の政治思想史、日本植民地時代の台湾の地図学等々である。東アジア世界の人文学の多彩な専門分野に及ぶものであったことがわかるだろう。所属する講座や研究室の枠組みを越えて研究交流を行う場であったのだ。

したがって、広い視野から各自の研究を位置づけ、再確認することができたことは、とても有意義なことであったにちがいない。ただし、研究発表者にはサマーセミナーの参加者が数多くおり、研究の補助手段として情報学的な手法を取り入れることも期待されたが、研究資料をデータベース化する試みもごく少数に限られ、その方面では目立った効果を感じることはできなかった。その点は、指導力不足として大いに反省すべきである。

ところで、最終年度には、サマーセミナーと人文情報学研究会の総括的なイベントとして、両者を合体させた東アジア人文情報学セミナー2007を開催した。開催日時は、2007年10月13日（土）10:00-17:30、場所は京都大学人文科学研究所本館大会議室において、午前の部では、中国学研究者のなかでコンピュータを導入したパイオニアである麥谷邦夫教授（人文科学研究所）、二階堂善弘教授（関西大学文学部）の両氏を招いて東アジア人文情報学の最前線の話講演してもらい、午後の部では大学院生による研究発表を行った。一般公開だったので65名前後の参加者が集まり、まずまずの盛況だった。

麥谷、二階堂両教授の特別講演は、大型計算機やNEC98シリーズの時代から現代に至るまでにどのようにコンピュータを活用してきたかの体験談を踏まえ

つ、人文学の世界においてパソコンで何ができるのかを提言するきわめて有益な話だった。

両氏の講演の要旨をかいつまんでまとめると、以下の通りである。

麥谷教授は、Emax, Perl, TeXを用いて『真誥』などの一字索引を作成し、印刷する一連のシステムを開発した経緯を概観しながら、データベース作成の技法をめぐる諸問題を様々な角度から検討した。「自分のやりたいことをコンピューターにやらせようと思ったら、プログラム言語を1つは習得しておくべきだ」と語り、「人文学はたくさん書物を読むことが基本である。読んで培った知識を有効に活用して勘を働かせるということが一番大事なことであって、あくまでもコンピューターは補助的なものと考えるべきだ」と附言する。さらに、今後の課題の一つとして、漢字を使ったテキスト処理を行ううえで漢字属性辞書データベースを作成し、運用できるようにする必要性があることを提言した。

一方、COBOLのプログラマーであったという二階堂教授は、日本語、中国語を混在させたホームページの作成、Perlを使ってデータ検索といったことを個人的に追求してきたが、かつては難しかったことが今は簡単にできるようになったとし、Windows Vistaの最新情報に言及する。また、台湾留学中に、まだインターネットで公開していなかった台湾中央研究院の漢籍電子文献の検索システムに触れ、データベースの作成を試行錯誤を始めたことを振り返り、『四庫全書』『四部叢刊』『大蔵経』『道蔵』といった大規模なデータベースが次々と提供され、さらに『中国基本古籍庫』という統合的データベースまで出現したことを報告する。そして、「自分でデータをつくるのが非常にばかばかしくなってきた」という状況であるが、「ツールやデータが増加する一方で、中国学への応用はいまだ十分とはいえない。なるほどパソコンは当たり前に使われるようになり、台湾中央研究院の二十五史を使わない研究者は少ないといえる。しかし、ワープロ、メール、それから多少の検索以外はどうか。まだまだ情報を使いこなすというところまでには至っていないのではないか。さらに言えば、機関としての環境がなかなか整わないというところがある」と指摘する。

さらに、二階堂教授が所属する関西大学で導入した

『中国基本古籍庫』ネットワーク版が実演され、通俗小説や戯曲のデータまで検索ヒットするが披露された。参加者のほとんどが初見であり、会場に驚嘆の声が上がった。そうした中国で推進するデータベース事業の急展開に対して、両講師がともにテキストと画像の乖離が深刻であり、データと画像データは一对で考える必要があると述べたことはとても印象的であった。以上のように、二人の講師の話は、電腦中国学の歴史を総括し、現状の問題点を鋭く指摘するものであった。東アジア人文情報学の今後の課題は、そこに集約されているにちがいない。

教育班リーダーとして私が担当した主な以上の事業を振り返ってみて、率直な感想を言えば、日暮れて途遠し、である。COE プロジェクトが目指した東アジア人文情報学の学問的な確立には、多くの難題を残した。一週間のサマーセミナーは言うまでもなく、新設した講義も開講してまだ短期間であり、プログラミング技法を自在に操ることのできる若手研究者を養成するどころの話ではない。教育対象となった大学院生のパソコン活用度はきわめて低く、研究活動に情報学的手法を取り入れるレベルに到達していないことを正直に認めなくてはいけないだろう。

それは、方法論として不必要であるということではない。むしろ逆に人文学研究にパソコンを活用することは当たり前になってきている。インターネット時代の人文学研究者は、従来型の資料学と新しいデータベース学の双方に通じる必要があると、誰しもが感じているのではないだろうか。しかし、IT 技術の進歩があまりに急速であるために、教育システムがまったく時

代に追いつけないまま、アナーキーな状態にある。現在、大学の一般教養的な基盤科目として情報リテラシー関連の授業は存在するが、大学院の専門教育において情報学的手法を学ぶ場は皆無である。その観点から言えば、上記の教育プログラムはきわめて斬新なものであったと自負できるかもしれない。しかしながら、高校や大学の一般教育の過程において、文系学生に対しても情報学の基礎力を培わなければ、十分な理解は得られそうにないことも明白である。人文学と情報学の融合を実現するには、文字通りの文理横断的な教育改革が不可欠であるように思われる。

思えば、紙に書く代わりにパソコンに文字を入力し、辞書や目録を引く代わりにネット検索し、電子化した書物を読み、電子メールで情報交換を行うデジタル生活をみんなが送りはじめたのは、20年にも満たない最近のことである。にもかかわらず、学問方法論や教育体制を再編しないといけないほどに浸透した。それも、グローバルな社会変容であり、最早後戻りすることはできない。

人文学が情報産業と手を組んで中途半端な文献学が横行し、従来の文献考証の方法論が衰退しようとしているのではないか、マシンの記憶でしか思考できない人間ばかりになってしまうと危惧するのは、年寄りの杞憂かもしれない。そうであっても、行く先は遠く帰り道もないのは徒労感がつのるばかりである。ネット王子やケータイ姫（と香山リカ、森健両氏が中公新書ラクレで命名した若者）が思いも寄らないアプローチの論考を発信してくれるようになるまで、このまま仮想空間を漂流し続けることになるのだろうか。

唐代研究ナリッジベースの構築

井波陵一

■ なぜ唐代か

21世紀を迎える頃、大量のデジタル・データの利用が可能になることで、中国研究に新しい可能性が開けるのではないかと期待がにわかに現実味を帯びてきた。実際、『四庫全書』をはじめ、多くのテキストがデータベース化され、その一部についてはインターネットを介して無料で利用できるようになり、その環境は順調に整いつつあるかに見えたのである。しかし現実にはたちまち厄介な問題が持ち上がった。こうしたデータベースは文字列検索こそお手の物だが、検索結果が大量になればなるほど、その後の処理に困難を来したからである。

唐代に関する情報を包括的な電子アーカイブとして提供することにより、こうした状況の打開を図ろうとしたのが「唐代研究ナリッジベース」である。歴代王朝の中から唐を選んだのは、それが東アジア全体における制度的・文化的基盤を提供した点で突出した存在であるからにほかならない。また、宋代以降になると文献が爆発的に増加するという事情もある。扱うに値する一定程度の文献が現存し、なおかつそれを全体として把握することが可能な点においても、唐代は格好の対象であった。

■ 『唐代研究のしおり』

「唐代研究ナリッジベース」開発にあたって極めて大きな示唆を与えてくれたのが、平岡武夫編『唐代研究のしおり』全12冊（京都大学人文科学研究所、1954年～1965年）である。同シリーズ復刊（同朋舎、1977年）の内容見本に付された平岡氏の「唐代研究のしおり 復刊のことば」にいう。

時間と場所とそして人と、この三本の柱が歴史を構

成する。私が唐代に関心を持ち始めた時、これらの柱を明確にした書物があれば、研究にどれほど便利なことかと思った。

この方針のもと、『唐代研究のしおり』プロジェクトは唐代、ひいては中国の古典研究に従事する者にとって不可欠であるところの、当時の詩文の出典、その時代に用いられていた暦法、行政地理の優れたインデックス、コンコーダンスを編纂して世に送り出し、今なお学界に多大の恩恵を施している。「唐代研究ナリッジベース」もこういった索引ないし資料集成を誰もが利用できる形で提供することを目標とした。

■ 「唐代研究ナリッジベース」の概念編成

「唐代研究ナリッジベース」に収録される情報の概念編成は、次のような情報軸に沿って行われている。

- ・ 唐代の人々の姓名、履歴
- ・ 地名および経緯度、地方行政機関、デジタル地図
- ・ 唐代に作られた作品（テキスト、美術工芸品、建築物）
- ・ 暦年および時間
- ・ 国家的重要事件

ナリッジベースの情報は、これらの情報軸の複数にまたがるものがかなりの量に上っており、それらが内部的に相互リンクされてウェブのような構造を成している。さらに、これらの情報は階層的オントロジーにより編成されており、そこから階層内の位置や他の情報との関係をもとにアクセスすることができる。

このような階層を構成する項目は、都市など地理的な位置を示す情報の場合はその都市を管轄する上位の行政機関であるし、人名の場合は家系や、出身地、所属する学派、また、僧侶の場合はこれに授戒や嗣法の系統が加わることになる。

■ 階層の活用

こうした階層を縦横に活用することで情報へのアクセスや検索は充実したものとなる。たとえば、則天武后が国号を周と改めて皇帝に即位した時期に権力を握った官吏について概要を得たり、唐朝復興後のその行く末を調べたりすることは、従来の研究ツールではきわめて困難である。そのためには何百人もの伝記を調べ上げ、一人ひとりについてその運命を辿った上で、総合的な概観を得ることが必要だからだ。これに対して、「唐代研究ナリッジベース」を使えば、すべての関連情報をたちどころに融合・咀嚼し、しかも結果を分かりやすく視覚的に表示することもできる。

このナリッジベースを活用できる分野やテーマはそれ以外にも数多くある。ナリッジベースの構成はデータに対して特定の解釈を押しつけたりしない。むしろ様々な異説や相矛盾する見解すら成立可能になるように配慮されている。一枚の紙がそこに書きつけられる内容について何ら制限を加えないのと同様である。

■ ナリッジベースの三本柱

技術的な観点から見ると、このナリッジベースは3つの柱から構成されている。

1. 唐代文化に関する第1次テキストのコレクション
2. これらのテキストが伝える事物の情報やその他の多くの情報（いわゆるメタ・データ）
3. ナリッジベースの操作と、テキスト、メタ・データその他すべての項目を結合するためのシステム

ナリッジベースの構築に当たっては、その内容のみならず、適切なインターフェースをも構築することが必要であり、最善の結果を得るにはその両方を完全に統合しなければならない。

■ マークアップの作業

マークアップの作業については、次のような新機軸を打ち出した。

1. 情報レイヤー（メタデータ）には人物や集団などの要素に対する識別情報を付与する。
2. テキストレイヤー（原典テキスト群）側でも、

たとえば「李世民」に言及しているすべての箇所に対して参照キーが付けられ、この参照キーによって情報レイヤーへの参照ができるようになる。

この結果として外見は全く異なる「唐太宗」、「李世民」、あるいは「文武聖皇帝」という情報が同等に扱われるようになる。また「秦王」という情報の中から「李世民」と関連づけられるもののみ選択され、「唐太宗」などが検索された際、その結果に含まれるはずである。このような場合、有効な検索結果を出すためには、情報レイヤーとリンクするのに曖昧さを解消しておく必要がある。

■ テキスト・コレクション

テキスト・コレクションの作業はおおむね既存の電子化テキストを利用した（中央研究院歴史語言研究所のテキスト・コレクションなど）。ナリッジベースではこれらのテキストに基本的な構造マークアップを施し、データベース内でのアクセスやリンクを実現した。また、一部のテキストを選んでその内容情報をさらに処理・マーケティングした。

■ コンテンツ・マークアップ

コンテンツ・マークアップは両唐書から開始した。マークアップの対象としては、テキストの内的構造のほかに、年号、人名、地名、および作品・著作の標題などを計画するのが適切だと判断した。

従前のデータベースにおいては、検索するよう求められた語彙は、テキストの中から単なる文字の羅列として何の区別もすることなく利用者に提供されていたが、このマークアップの作業により、年号、人名、地名、作品・著作の標題が全て別箇に認識されることになる。たとえば利用者が年号に関する情報のみ必要とする場合は、それのみが検索結果として表示される。

■ メタ・データ

ナリッジベース内のメタ・データはすべてトピック・マップにより管理される。トピック・マップのものと情報モデルはきわめて単純かつ抽象的なもので、任意の情報とその関係を記述することができる。柔軟で幅広い応用が可能なため、メタ・データ管理にはきわめて適している。メタ・データの収集と処理は

原則として次のように行われた。

1. メタ・データは既存の電子化第1次文献（両唐書）、文献目録、伝記に現れる固有名詞から収集する。
2. 得られた情報を処理・高度化する。両唐書から採った地名に、それらの地誌情報を重ね合わせることで、位置や上位の行政機関、名称変更、規模変更に関する情報提供が可能になる。また人名も同様に、伝記に記載される重要な事実とリンクされる。
3. 他の資料（紙、電子データ）から得られた情報もトピック・マップに適宜組み入れる。
4. 地理的情報を加える（地図からのアクセスや地名による検索実現のため）
5. 年号、日付や関連情報をマーク・正規化する（対照年表や日付による検索実現のため）

トピック・マップの情報はいずれの段階でも利用できる。これによりテキストに含まれる他の情報を探したり、その情報をトピック・マップにフィードバックすることができる。

■ ナリッジベース管理システム

ナリッジベースのシステムは、既存のコンポーネントやソフトウェア・アプリケーションをもとに開発される。それらをつなぎ合わせて一貫した使いやすいインターフェースを実現することが主な仕事となった。

第一段階では、XML テキスト・ベースに対する Cocon ベースの検索インターフェースを開発した。これはシステムのその後の開発への足掛かりとなるものである。次の段階では、トピック・マップ・エンジンおよびトピック・マップの情報をテキストと統合した。

最も重要な部分は、利用者がナリッジベースによりデータを検索したり、情報をシステムにフィードバックするためのユーザー・インターフェースである。これによって、ナリッジベース管理システムは利用者のニーズに対応したものとなり、その内容を進化させることが可能になるからにはほかならない。

■ 期待される効果

両唐書に現れる人々がどのように分布していたかを地図上に示すことができる。しかも、一定の地位を得

た者や官吏となった者などを異なる色で表示することが可能である。また、この同じ情報を使ってデータベース（あるいはその一部、たとえば『全唐文』）に対する検索範囲を狭めることもできる（たとえば北方出身者だけを検索し、その結果を南方出身者の検索結果と比較するなど）。比較のために検索範囲を詩の行末の文字だけに絞り込むことも簡単である。メタ・データの重ね合わせを行うことで、テキストの中から当面の研究テーマにとって必要な部分だけを的確に抽出することが可能となり、何千件もの検索結果の中から苦勞して適当なものを探し求めるといった手間が省ける。

■ 3つの知識ベース

『唐代研究のしおり』の構想を承けて、「唐代研究ナリッジベース」を人物、官職、地理の各分野を対象とした3つのより小規模な知識ベースに区分した。

（1）唐代人物知識ベース

両唐書及び『資治通鑑』のデータ、ならびに周祖譔主編『中國文学大辞典』唐五代卷（中華書局、1992年）など最新の研究成果を生かして、唐五代の文学者約4000人の人物情報を知識ベースとして構築した。平成17年度には内部公開、18年度の半ばには一般公開を実現した。

（2）唐代官職知識ベース

各官職に誰が任官していたかという「静的」な側面はかなり解明されたが、どの官職からどの官職へ遷るのが通例であったのか、またいかなる傾向が見られるかという「動的」な側面については、意外と研究が手薄である。「唐代官職知識ベース」を構想するに当たって、このような「動的」側面を少しでも明らかにし、今後の官制研究に役立ち得るものを作ることを目指した。これまで別個に存在してきたこれらの諸成果を包括的に把握できる視座を獲得できたことは、大きな意味をもつであろう。

（3）唐代地理知識ベース

『唐代研究のしおり』シリーズの一冊、『唐代の行政地理』のデータ化から着手した。『唐代の行政地理』には、唐代に存在した行政区約3500件について、各一次資料ではどこにそれが見えるかを明示している。同書の問題点はこれら行政区の上下関係や統廃合・改称に関する時期情報が必ずしも明確になっていないこ

とにある。このため、『唐代の行政地理』をデータ化する一方で、唐代の地理に関する一次資料についても一定程度の情報（地名・従属関係・統廃合・改称）をデータ化した。資料相互の矛盾点はきわめて多いが、一つの説を採用するのではなく、ゆるやかに全ての情報を包含することを目指した。

■ 知識ベースの統合と唐代研究の可能性

「唐代人物知識ベース」、 「唐代官職知識ベース」、 「唐代地理知識ベース」 はあくまで「唐代研究ナリッ

ジベース」の一部である。三者の統合が現実のものとなれば、たとえば「唐代人物知識ベース」から某人が某州の刺史（知事）になったという情報を得ると同時に、「唐代官職知識ベース」からは某人の前任者・後任者の、また「唐代地理知識ベース」からはその州に関する情報が総合的、かつ視覚的に得られることになる。これらが今後の唐代研究の可能性を大きく広げるであろうことはいうまでもない。平成19年度末までに全体として「唐代研究ナリッジベース」を完成させることを目指している。

東アジアの文字に関する人文情報学的研究

安岡孝一

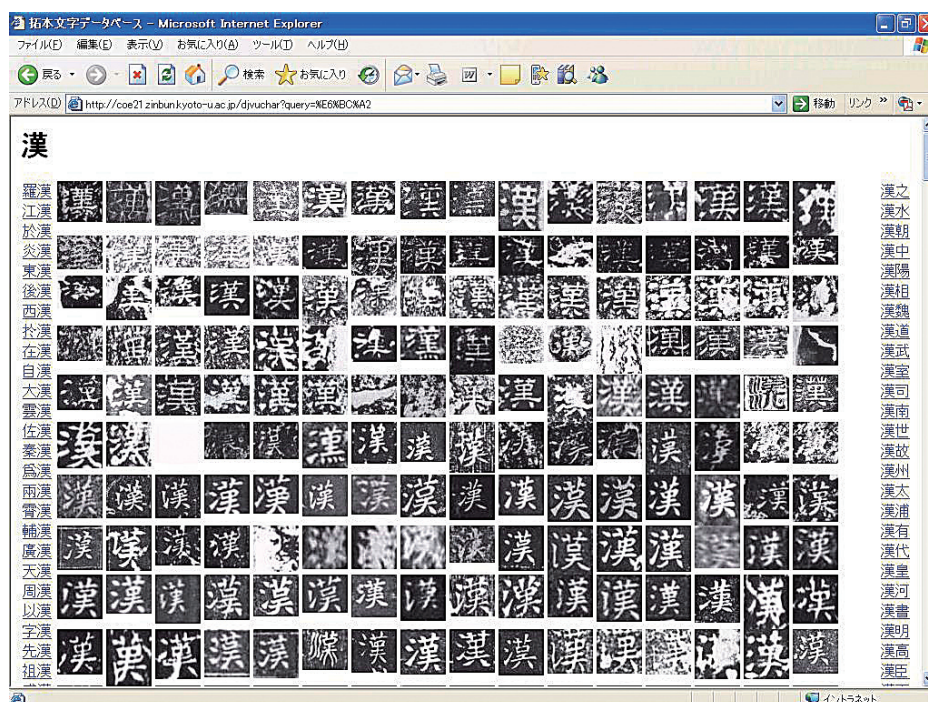
2003年7月の21世紀 COE 発足以来、「東アジアの文字に関する人文情報学的研究」グループは、漢字文化圏における漢字使用の現状とその変遷、漢字から派生した周辺の文字との有機的な結合、および、それら諸種の文字を扱うための手法、に関する人文学的・情報学的研究をおこなってきた。以下では、その成果について述べる。

本研究グループの成果は、大きく『拓本文字データベース』『京都大学人文科学研究所所蔵甲骨文字データベース』『新人名用漢字983字フォント』『CHISE』の4つの形で、WWW上に公開している。もちろん、これ以外にも細かな成果はあるが、以下では、これら4つを中心にそれぞれについて述べる。

『拓本文字データベース』は、京都大学人文科学研究所所蔵の拓本のうち、釈文が確定している約7,500枚

について、釈文付きの画像データベースを製作する試みである。その際、全ての拓本を文字単位の画像に分解してデータ化し、文字あるいは文字列ごとにまとめて検索可能とし、漢字の字形が漢代から清代にかけて歴史的にどう変化したかを通覧できるシステムとしている。また、検索システムに異体字処理を導入することで、日本の常用漢字は言うまでもなく、中国の簡体字でも、台湾の常用國字でも、韓国・北朝鮮のハングルでも、検索可能としている。

『京都大学人文科学研究所所蔵甲骨文字データベース』は、京都大学人文科学研究所所蔵の甲骨のうち、3,246片を画像データベースとしたものである。その際に、人文科学研究所で付与した甲骨番号で検索可能とし、インターネットに掲載された論文等からハイパーリンクでアクセス可能としている。



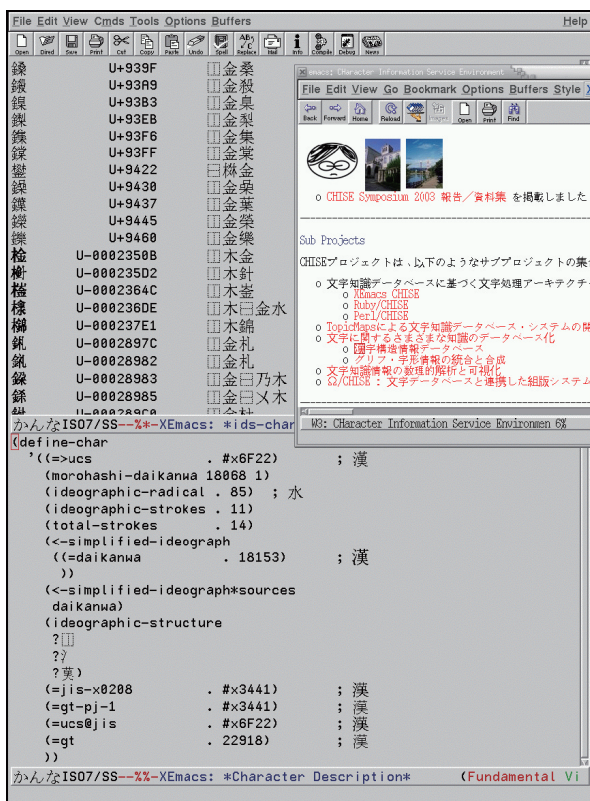
『新人名用漢字983字フォント』は、OpenType アウトラインフォントを、画像から直接合成するシステムの応用例として作成されたものである。実際には、2004年9月27日官報をスキャンした画像をもとに、同日施行された新しい人名用漢字の OpenType フォントを合成し、WWW で公開したものである。もちろん、これはあくまで応用例であり、誰もがこのようなフォントを簡単に作成できるよう、製作システムそのものも WWW で公開している。

『CHISE』(Character Information Service Environment) は、漢字に関するさまざまな知識の集合体であり、漢字ナリッジベースとも呼べるものである。『CHISE』の中核をなすのは、「漢字構造情報データベース」「異体字・類字データベース」「グリフ・字形情報統合処理エンジン」の3つのシステムである。「漢字構造情報データベース」は、漢字の字体の構造を、へんやつくりなどの部品を基に IDS と呼ばれる手法で

記述したものである。この結果、大漢和収録の全ての漢字(約50,000字)や Unicode 収録の全ての漢字(約70,000字)を、その部品の組み合わせで検索可能としている。「異体字・類字データベース」は、漢字の異体字関係をデータベース化したものであり、『拓本文字データベース』で構築してきた異体字処理を、他の一般のデータベースにも移植できるよう、汎用のソフトウェア部品としたものである。「グリフ・字形情報統合処理エンジン」は、いかなる漢字の OpenType フォントも、漢字の部品の組み合わせで合成できるシステムである。これら3つのシステムにより、文字コードを持たないような漢字を含め、全ての漢字を、入力・検索・出力できるようにしている。

本研究グループの最大の目標は、21世紀 COE の目標と同じく「漢字文化の全き継承と発展」であり、それに関してはほぼ達成できたと考えられる。「継承」に関しては、『京都大学人文科学研究所蔵甲骨文字データベース』から始めて『拓本文字データベース』を通覧することで、殷代から清代に至る漢字字体の変遷を、一望のもとに把握することが可能となった。「発展」に関しては、『CHISE』によって、過去・現在そして未来の漢字を、情報処理手法の下に全て知識集積可能としており、甲骨文字であろうが、2004年に現れた人名用漢字であろうが、あるいは今後あらたに作られるだろう漢字であろうが、全て入力・検索・出力可能となった。

なお、『拓本文字データベース』と『CHISE』に関しては、21世紀 COE 終了後も継続的にデータをメンテナンスし、さらなる拡充をおこなうために、科学研究費補助金の申請を予定している。また、本研究グループで得られた知見と、その要素技術の多くは、2008年11月に開館予定の京都大学デジタル・アーカイブに採用されており、より多くの文献および文字史料のデジタル化に資する予定である。これらにより、本研究グループの成果は、21世紀 COE 終了後も、漢字文化の全き継承とさらなる発展を支えていくことになるであろう。



表紙掲載写真解題

稲葉 穰

5年間にわたって継続したCOEプロジェクトの終了とともに、本ニューズレターも最終号を迎えた。通算刊号としては12号であるが、2004年10月に中国宗教文献研究国際シンポジウム (<http://coe21.zinbun.kyoto-u.ac.jp/is-religion-2004.html> ja 参照) にあわせた特集号を組んだので、実質は13冊目にあたる。特集号を除いて、執筆者の数はのべ51人、そのうち事業推進担当者やCOE研究員など内部の者が33人、人文科学研究所所員が9人、それ以外が9人という内訳になっているが、もちろん複数回執筆している人もいる。筆者も実はこれが3回目の寄稿となる。

それぞれの原稿は学術的論文やフィールドワーク報告、学会動向といった学術資源から、もう少しだけ読み物まで多岐にわたっており、硬すぎず軟らかすぎず、ニューズレターという性格にうまく合致するものになったのではないかと自負している。貴重な学識や練達の文章を提供していただいた執筆者各位に、この場を借りてあらためて御礼を申し上げる次第である。編集担当としては実は執筆していただいた内容について追加のお願いをする必要が全くなかった、という点でも大変ありがたかった。それぞれがご自身の原稿に責任を持ち、欠くところなき内容をあらかじめ準備してくださったお陰であると、この点も非常に感謝している。

というわけで非常に楽な編集作業であったのだが、実は一番苦労したのは表紙の写真であった。当初から漢字情報研究センターの建物や所蔵物の写真を少しずつ載せていこうと方針を立てていたのだが、その選定はそうそう容易ではなかった。号によって写真にキャプションがついていたりついていなかったり、あるいは途中から執筆者名が目次から消えたり、といった不統一や不手際についてこの場を借りて深くお詫びするとともに、ここであらためて各号の表紙写真について

少しく説明しておきたい。お手元に全号が揃っているなら、それらを参照しつつ以下をお読みいただければと思う。

創刊号の表紙には、1930年に建てられた旧東方文化学院京都研究所、現京都大学人文科学研究所附属漢字情報研究センターの外観写真を掲載した。この建物は京都大学の多くの建物を設計し、同時に多くの優れた弟子を育成したことで知られる武田五一京都大学工学部教授に学んだ東畑謙三氏の設計にかかるもので、現在国の登録有形文化財に指定されており、創刊号を飾るに相応しいと考えた。

続いて第2号表紙は漢字情報研究センター2階閲覧室の写真。かつてはここで人文科学研究所の定例公開講演会が行われたこともあり、現在でも毎年秋に開催される全国漢籍担当者講習会の会場となっている。第3号の写真は、長年に渉って蒐集された貴重な漢籍コレクションを有するセンター書庫内の様子。

第4号からは研究所やセンターが所蔵する各種資料の写真を掲載することとし、第4号には『永楽大典』（明隆慶年間重写本）残巻の写真を選んだ。第5号は松本文庫所収の真諦三蔵訳『随相論』（金刊本／趙城廣勝寺金版大藏經之一）。第6号は魏の黄初元年（220年）の「孔子廟碑」拓本の写真である。

第7号には甲骨文字コレクションから一点（B0848／貝塚茂樹著『京都大学人文科学研究所蔵甲骨文字』図版51）を、第8号には1938年から44年にかけて水野清一、長廣敏雄両教授によって行われた雲崗石窟調査の写真資料から一点を、それぞれ掲載した。

第9号表紙は1904年から1939年にかけて上海で発行された新聞『時報』の貴重なマイクロフィルム（人文科学研究所は全号のマイクロフィルムを所蔵する）の一齣。第10号には『利瑪竇題寶像圖』（民國十六年 武進陶氏涉園 用程氏墨苑本景印）から当時の漢字音を

ローマナイズした一頁を掲載した。

第11号は特集号として、金坂清則京都大学人間・環境学研究科教授によるマラッカでのフィールドワーク報告を掲載したが、表紙写真も同教授撮影の牌匾（看板）である。

上記のうち『永楽大典』、『随相論』は漢字情報研究センター「東方学デジタル図書館」(<http://kanji.zinbun.kyoto-u.ac.jp/db-machine/toho/html/top.html>)にて、また「魏孔子廟碑」は同「京都大学人文科学研究所所蔵石刻拓本資料」データベース(<http://kanji.zinbun.kyoto-u.ac.jp/db-machine/imgsrv/takuhon/>)にて、さらに甲骨文については同じく「京都大学人文科学研究所所蔵甲骨文字」データベー

ス(<http://mousai.kanji.zinbun.kyoto-u.ac.jp/koukotsu/>)にて、それぞれ参照可能である。

これら表紙写真を蒐集する過程で様々な方に教えを乞い、筆者自身、学ぶことが多かった。先人たちがつくりあげてきた貴重な研究の場と資源とがどのようなものであるのか、その雰囲気の一端なりとも示し得ていれればと願う。

ちなみに最終号の表紙には、本 COE プロジェクトメンバーがデータベース構築やプログラム作成作業を行ってきたプロジェクト事務所（通称まつお館）の写真を掲載した。伝統を受け継ぎ、斯学の新たな発展を担う研究者たちが若き日を過ごした場の思い出となるかも知れない、と期待してのことである。

Chinese Characters
and Culture



発行日 2008年 3 月31日
発行者 文部科学省21世紀 COE プログラム
「東アジアにおける人文情報学研究教育拠点—漢字文化の全き継承と発展のために—」
住 所 〒606-8265 京都市左京区北白川東小倉町47 京都大学人文科学研究所
電 話 075-753-6997 FAX 075-753-6999
e-mail coe@zinbun.kyoto-u.ac.jp • Web Site <http://coe21.zinbun.kyoto-u.ac.jp/>

竊以宮儀方載之廣漢識棟靈之異談夫以無於其極