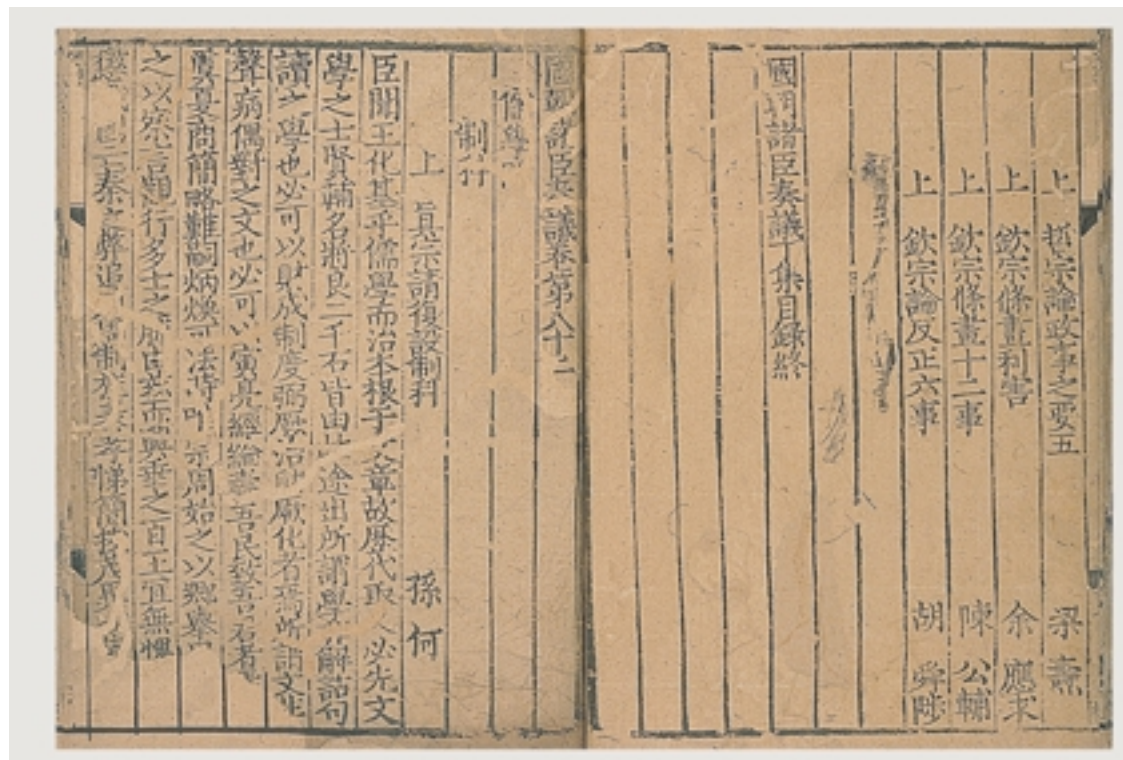


# 漢字と情報

No. 2  
2001・3



京都大学人文科学研究所 Documentation and Information Center for Chinese Studies (DICCS)  
附属漢字情報研究センター Institute for Research in Humanities, Kyoto University



情報の運不運

漢字のイメージ

UTF-2000プロジェクト

会議レポート

「全国漢籍データベース協議会第1回総会」

人文研のアーカイブス(2) 『國朝諸臣奏議』

## 情報の運不運

尾崎雄二郎

私が漢字情報研究センターの前身、東洋学文献センターに移る人事の審査で、藤枝晃さんは、尾崎はレフェリーのない雑誌にばかり書いているからとダメを出されたという。当時の教養部報『人文』などを指すだろう。ところが、レフェリーもなかったその『人文』の名は、お隣の中国には現れることがある。それに書かれた私の文章の一つ「大英博物館本蒙古字韻札記」1962は、中国人研究者が私を引用する数少ない仕事の一つになっていて、それは、世界でただ一本しかその存在が知られていない原テキストの、写本としての実質を私が初めて問題として取り上げているため、『蒙古字韻』を論ずるとき、人が私のそこでの指摘に触れずに通ることはむづかしいからである。だから孫引きかも知れないものも含め、予期しないところで私の名と『人文』の名とに出くわすことがある。

『人文』を中国にまで旅させたのは、大阪外大の佐々木猛さんである。たしか北京に滞在中、相識の私のそのコピーを、中国社会科学院語言研究所の楊耐思さんに贈られたのである。書いてから日はもうずいぶん経っていたと思う。私がセンターに移ったのが75年のことだったし、佐々木さんとはセンターで知り合ったのだからである。楊氏は81年、論文の注で私の説を紹介され、それは『蒙古字韻校本』1987にも引き継がれた。

レフェリーありなら、もしや、と思える失敗も、たしかにある。小川環樹、入矢義高両先生退休記念論集のために書いた「disarticulation ということ」1974もそうであった。

それは、中国語の /ki/ などが後に /chi/ その他に取り替えられるようになる、いわゆる牙喉音と歯音との、口蓋化を経過した合流がいつごろ始まるかという、中国音韻史の古くて新しいテーマについての、中国のいわば現場、元人の「鷄稱齋」という記述を根拠として取り上げるものでも

あったのだが、その紀年、至元丙子は、六十年を隔てて二度あったのを、疎忽で、というより多分に、より古くという気持ちにも恐らくは引きずられて、前の方1276年のものだとしてしまい、誤りは私の論文集への再録の段階で見過ごされ、自ら気づいてそれを正す第二刷のときまで続く。十二年も経っていた。元人のこの記述に触れる学界最初の言及が恐らくは私のこれだと思えるだけ、この瑕疵は顧みて遺憾である。

それへの言及は私が最初と見当をつけるのは、中国学界の消息に私よりは当然詳しいはずの吉林省社会科学院・寧忌浮『古今韻会舉要及相關韻書』1997なども、私と独立に、私の挙げたと同じ記述を、他からの引用としてでなく取り上げて、その語史的意義を強調するからである。紀年の誤りはあったにしても、私の文章の初出を寧氏の自序の書かれた年と比べれば、二十一年もの隔たりがある。早くこの情報が一般のものになっていれば、これだけの歳月は、このテーマに関してもっと多くの進展を齎していた可能性がある。

かつて小川先生が私のその論文集の書評を下さったとき、批判の矛先が、だれ一人注目することのなかったこの論文の主題に主として向けられたまではよかったが、始まりに近く挙げる先の記述には、語史研究者でもあられた外ならぬ先生が、気づかれずにか素通りしてしまわれた。私の論文集「事項索引」の「事項」は、「提案」、「主張」などと言い換えてもいい。だから英訳名を Index of author's *proposal* とした。做う人あるを知らないが、ちかごろ中国の学術雑誌には、掲載論文の頭のところに、「提要」のほか「關鍵詞」すなわちキーワードを幾つかずつ挙げさせているのがある。日本でも見習っていいことだろう。私の索引でも元人のこの記述を、少なくとも牙喉音破擦音化に関する *proposal* の一つとはしていたのである。(京都大学名誉教授・元東洋学文献センター長)



## 漢字のイメージ

星野 聰

あらゆる言葉は、それがどのように使われてきたかを知らなければ、その本当の意味を理解できない。言葉の意味は、それが使われた時代によっても変化する。また、どのような環境の下で使われたかも、意味を理解する上で重要である。言葉がもつ意味を正しく知るには、辞書の解説だけでは十分でないのである。

人々が何か言葉を聞くと、過去にその言葉について見聞きした経験から、何らかの印象が引き出され、それがその言葉の本来の意味を包み込む働きをする。そして、人々は言葉の集合としての言説を解釈する際に、それを受入れたり批判したりするのである。

筆者は、戦争中に小学校教育を受けた。教科書には、日本を「豊葦原瑞穂の国」と記していた。学校は民家が建て込んだ大阪市内にあったが、それでも見渡す限り稲穂が揺れる状景を脳裏に浮かべたものである。これは、豊葦原や瑞穂という漢字が引き起こす効果によるのである。戦中戦後の食料難を振り返ると、言葉がもつカリスマ性というべきものに気が付く。これは、何も戦争中に限ったことではない。言葉には常に何らかの情報が陰影のように付随して、人々にその言葉が持つカリスマ性、アイマイさ、陳腐さなどを印象付けるのである。

そこで、ある民族のことを正確に記すには、その民族の言葉を使わなければならないことがわかる。日本のことは、日本語で記さないと、その心情を記すのが難しい。現在、我が国では、ITの重要性が叫ばれている。我々は、ITを活用して、真意を相手に文字で正しく伝えなければならない時代に生きている。そこで、漢字の情報化が、このIT時代に重要な課題になっていると考えている。

リストラクチャリングという言葉が略してリストラという。しかし、我々が理解するリストラは必ずしもリストラクチャリングと一致していない。リストラは日本語である。誰かがリストラされたというと、会社のリストラクチャリングとは関係なく人員整理されたのかも知れないのである。これは、日本語のアイマイ性の例である。また国語審議会は、イノベーションを革新という日本語で置き換えた方が分かりやすいという。しかし、革新という言葉は、色々な場面ですでに使い古された言葉である。革新政党というのはその一例であるが、言葉がどのように用いられているかは、インターネットで用例を調べるのが手っ取り早い。用例が分かれば、その言葉が人々にどのように印象づけられてきたかを知る手係りになるのである。

イノベーションという言葉には、未知のものを創造する力といった、何かカリスマ的なものが感じられる。国語審議会が例示しているアカウントビリティという言葉についても同様である。そこで、これらの言葉を簡単に日本語に当てはめるのは危険でないかと思われる。たとえ分かりやすくも陳腐な日本語に置き換えてしまうと、もとの言葉のカリスマ性が消えてしまって、人々に安易に受け取られてしまう。その結果として、社会の改革にも悪い影響を与えるかも知れないのである。もし日本語にしたいのであれば、イノベーションは、例えば「念創」という風に、シガラミのない新語を造るのもよいだろう。

それにしても、我々は日本語の言葉に、よいイメージを残して行きたいものである。(京都大学名誉教授)

## UTF-2000プロジェクト 守岡知彦

電子化文書は作成・加工・転送などが容易であり、非常に低コストで情報を共有することが可能である。また、検索やハイパーリンクなどにより、情報を以前の文書とは異なった視点で加工し提示することを可能にしている。このような点で、文書の電子化技術は情報化社会の基盤技術であるだけでなく、これまでに蓄積されて来た知を新たな視点で広く人類全体で共有可能にする可能性を秘めており、実際、そのような動きが社会のさまざまな局面で起こって来ている。このような素晴らしい利便性・可能性は計算機で文字を扱えるようにしたことによって起こったことであるが、見方を変えれば、計算機で扱うことが不可能な文字ではこのような利便性は享受できないし、同様に、計算機で扱いにくい文字ではその効果は減少するということになる。実際、(この問題だけが原因ではないと思われるものの) 現在もっとも扱うことが容易な ASCII で表現可能な英文を使う社会・分野ではこのような動きがもっとも進んでおり、それに比べれば和文圏・漢字圏は遅れている。

英文に比べて和文が、ラテン文字に比べて漢字が計算機で扱いづらい理由の1つは、現在の計算機で文字を扱う仕組みにあると考えられる。現在、多くの文字処理システムでは、文字そのものを扱う代わりに、文字に固有の番号を振って、その番号で文字を表す手法を用いている。ここで、文字と番号の対応規則を「符号化文字集合」(Coded Character Set; CCS) もしくは「文字符号」(character code) と呼び、文字に振られた番号を「符号位置」(code point) と呼ぶ。また、このように、有限の文字の集合を定め、各文字に固有の番号を振りその番号で文字を表現する手法のことを、ここでは『符号化文字モデル』と呼ぶことにする。図1に示すように、符号化文字モデルでは文字列は文字に振られた番号の列で表現される。文字に関する知識は符号化文字集合の定義の中に

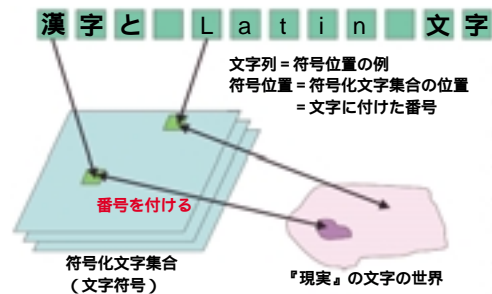


図1：符号化文字モデルの概念図

存在し、通常、計算機の中には存在しない。このため、計算機は少ない記憶量で文字を表現できる半面、計算機が扱うことができる文字の種類や文字の概念は符号化文字集合の定義に束縛される。英文で用いるラテン文字のように文字数が少なく個々の文字の同定が容易な用字系 (script) の場合、このモデルは適切であり容易に実現できる。しかし、漢字は文字の数が非常に多く、その同定は容易ではない。字形のデザイン差と字体の差と文字の違いは連続的であり、その区別は恣意的なものにならざるを得ない。そのため、さまざまな要求を概ね満たす高品質の符号化文字集合を作るためには非常な労力を要する上、それでも全ての要望を満たすことは事実上不可能になる。

このような問題を抜本的に解決するためには、符号化文字集合における文字の定義に依存しない文字の表現が必要である。このためには、文字を扱うシステム自身が表現する文字の性質に関する知識を持ちそれを利用するような新しい文書処理の枠組が必要となる。これは即ち、各文字や文字体系の性質を記述可能なメタな文字表現系とそれを処理する文字処理モデルを構築し、さまざまな文字の性質に適した処理を利用可能にすることであると考えられる。例えば、漢字であれば漢字の持つ性質や各文字の性質に関する知識を記述したデータベースを作成し、そのデータベースを参照することによって文字を処理することになる。このような考えに立ち、著者は“UTF-2000”と呼ぶ文字表現モデルを提案している。

図2に示すように、UTF-2000モデルでは文字列を文字オブジェクト(への参照)の列で表現する。文字オブジェクトは文字に関するさまざまな属性の集合で表現されるようなもので、文字に関する知識を表現したものである。文字オブジェクトの属性としては、文字の種類、音価・部首・画数などの属性や、その他のさまざまなものが考えられる。既存の符号化文字の世界の文書やフォントなどの資産を利用したり、既存のシステムを利用している人々と情報交換することは需要であるが、これも符号化文字集合の種類と符号位置を文字の属性の一種として捉えることによって可能となる。

UTF-2000モデルの実現可能性や妥当性を検証するための試験実装として、著者は XEmacs

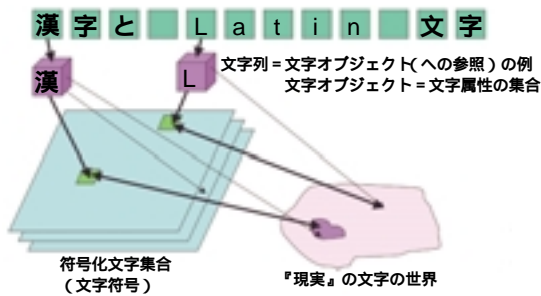


図2：UTF2000モデルの概念図



図3：XEmacs UTF-2000

UTF-2000 と呼ぶ文書編集系を開発している。XEmacs UTF-2000 は XEmacs と呼ばれる対話型統合環境を基にしている。XEmacs は GNU Emacs と呼ばれる拡張可能なエディタを中心とした対話型統合環境を整理・拡張し、絵なども扱えるようにしたものである。一方、Mule (MULTilingual enhancement of GNU Emacsen) は GNU Emacs の多言語拡張で、現在は GNU Emacs に統合されている。この Mule 機能を XEmacs に統合した XEmacs-Mule の開発も進められており、XEmacs の一部として配布されている。XEmacs UTF-2000 はこの XEmacs-Mule を基に大幅に改変したものである。

XEmacs UTF-2000 は文字データベースを取り換えることにより、容易に異なる文字表現や文字概念を利用可能であり、符号化文字モデルに基づく実装に比べて格段の柔軟性を持っているといえる。XEmacs UTF-2000 は符号化文字モデルに基づく従来の XEmacs-Mule に比べればより多くの記憶量を要求するが、現在の豊富な記憶量を鑑みればこのことはそれほど問題ではないと考えられる。また、現在は文字定義を全て XEmacs UTF-2000 が管理する主記憶上に格納しているが、これを外部のデータベースに持ち、必要な時に必要な分だけ主記憶に取り込むようにすることで、必要とする記憶量を削減することが考えられる。

ところで、実際にさまざまな文字の扱いを利用するにはその基礎となる標準的な文字データベースの存在が不可欠である。現在、利用できる大規模文字データベースは Unicode 用のものが大部分であり、Unicode とは異なる文字統合・分離規則を用いるのは困難である。また、Unicode に含まれない文字に関する情報を系統的に集めた自由に利用可能な文字データベースは現状ではほとんど存在しない。このため、このような標準的なデータベースを作成し、自由に利用できる形で配布することは重要であると考えられる。また、こうし

\*1 GNU Emacs の次の版 (21.1) でも絵も表示することができる。

たものを拓本や印刷文字などの一次資料に基づいて作成すればその利用価値ははかり知れないといえる。そこで、このような字形の出典に関する情報も含めた文字データベースを作成することは、符号化文字モデルの問題点の克服するために大変重要なことだと考えられる。(センター助手)

## [ 会議レポート ]

# 全国漢籍データベース協議会第1回総会

2001年3月9日午後、学術総合センター内一橋記念講堂を会場として「全国漢籍データベース協議会」の第1回総会が開催され、全国の漢籍所蔵機関の職員、研究者ら、約90名が参加した。この協議会は、国立情報学研究所、東京大学東洋文化研究所附属東洋学研究情報センター、京都大学人文科学研究所附属漢字情報研究センターを幹事機関として発足したもので、事務局は当面漢字情報研究センターにおかれ、全国的な漢籍データベース実現のための諸問題を討議し、連携をはかっていく。総会では、事務局からの趣旨説明、フォーマットの提案、技術的な展望が提示され、それらに関して活発な議論が行われた。また先行して進んでいる東洋文化研究所の漢籍データベースの現状報告もあわせ行われた。三幹事機関の連名で出された協議会の趣意書は以下のごとくである。

1. 日本全国の国公立図書館に所蔵される「漢籍」が、我が国の東洋学のみならず関係諸学の研究上、きわめて重要な資料であることはいうまでもない。しかしながら、全国規模での漢籍目録が存在しない現状では、必要とする「漢籍」の所在を知ることすら大きな困難を伴っている。
2. 国書に関しては、国文学研究資料館の提供する「国書基本データベース」が実験運用中である。

これと同様の全国漢籍データベースを構築し、ウェブ上で公開することが、当面緊急の課題として要請されている。

3. 一方、図書館での目録としては国立情報学研究所が運用するNACSIS-CAT総合目録データベースが存在し、漢籍を含めて我が国の教育・研究図書館の有する資料の所在情報を幅広く収録している。システムで蔵書管理が必要な図書館等では、この総合目録データベースに登録するという方法をとっている。

4. しかしながら、このNACSIS-CAT総合目録は、全国の所在目録を形成するとともに図書館資料の蔵書管理を想定しているために、出版によって成立する書誌単位で本の外的記録を作成しており、本の内容に即した単位で記録を作成する伝統的な漢籍目録とは、書誌事項の見方や分類の方法などで大きく異なっている。

5. したがって、蔵書管理を想定したNACSIS-CAT総合目録に加えて、本格的な古典中国学研究のためには、中国固有の学術体系に則り、伝統的な四部分類を採用して編纂してきた従来の漢籍目録と整合性をもつ全国規模のデータベースを構築する必要がある。

6. この漢籍データベースの構築により、著作としての漢籍の書誌データが整備されることになり、NACSIS-CAT総合目録データベースを始めとした様々なデータベースでは、漢籍データベースを書誌的な典拠として活用することが可能となる。

7. また、図書館の蔵書管理が不要なためNACSIS-CAT総合目録データベースに登録されないような資料をこの漢籍データベースが収録することにより、全体として我が国の漢籍を網羅することができるようになる。

8. よって、全国漢籍データベース実現のための諸問題を討議し、連携をはかる組織として全国漢籍データベース協議会を置くことを提案するものである。

## 人文研のアーカイブス(2)

## 國朝諸臣奏議 残二十六卷

宋趙汝愚輯

梶浦 晋

本所は漢籍を多く所蔵することで知られてはいるが、前身たる東方文化学院京都研究所の開設時より、清朝の学問を基調に校訂の行き届いた本を中心に蒐集につとめてきた。このため所蔵の典籍の多くは清代の版本で、漢籍の収蔵に富む他の同規模の図書館に比し宋元明の古版や古鈔の典籍は必ずしも豊富とは言えない。所蔵の宋元版は60数点あるが、その多くは松本文庫の仏書で、これ以外の宋元版は数点にとどまる。ここにあげた本は

そのうちの1点である。

『國朝諸臣奏議』百五十巻は、南宋の宰相であった趙汝愚が淳熙十三年(1186)に編纂したもので、北宋時代の名臣241人の奏議1630篇を、君道・帝繫・天道・百官・儒学・礼学・賞刑・財賦・兵・方域・边防・總義の十二門に分類して収めている。現存の版本には淳祐十年(1250)刊本、弘治中無錫華氏會通館活字印本、四庫全書本の3種がある。

本所蔵本は巻第八十二至第八十八第九十七至第一百一百十二至第一百十四第一百十七至第一百二十二第一百四十五至第一百五十のわずか26巻と目録2巻の残本である。本の大きさは27.8×23.3cm、襷装で金箔散らしの後補表紙が付され16冊に分冊されている。版式は左右双辺(21.6×15.7)、有界、每半葉11行、每行23字、注文小字双行、白口、双黒魚尾で、上象鼻に大小字数、下象鼻に丁次と刻工名を刻する。〈宗廟〉〈陛下〉〈聖恩〉等の語や宋帝の廟號など宋朝に関わる語は上一字分を空格としている。

刻工名等により静嘉堂文庫や台北故宮博物院蔵本などと同版であることが判明する。この本は巻第一を欠くため、直接には確認できないが、故宮本などには巻首に淳祐十年(1250)の趙希滌の跋と同年の史季温の跋があり、これにより淳祐十年に福州路提学の史季温が刊刻したものであることがわかる。元代にいたり大徳元統の間に補刻が行われ、巻第八十二の末には「大徳四年九月補刊/福州路儒學教授劉直内命工刊補」と、また巻第一百十四第十七葉の下象鼻には「元統二年刊」などがある。残念ながら伝来をしめす印記や墨書等はない。

北京図書館や台北故宮のほか美国国会図書館などに完本・零本取り混せて約20点が確認されているが、国内では静嘉堂文庫、お茶の水図書館成篁堂文庫、京都大学文学部(巻第六十七)と本所の所蔵が知られるのみである。本所蔵本は全体の二割にもみたくない残本で、かつ元代の補刻をまじえるものではあるが、宋代撰述書の宋代刊本として貴重なものといえる。(センター助手)

## HP・TOPICS

漢字情報研究センターのホームページ (<http://www.kanji.zinbun.kyoto-u.ac.jp/>) で CHANA 3 (「東洋学文献類目」1981年以降のデータベース) の検索が利用できるようになりました。これまでは、京都大学の大型計算機センターのMSPシステム上で有料の検索しかできませんでしたでしたが、本センターの中心事業の一つとして、無料公開することにしました。奮ってご利用下さい。

### CHINA3 for WWW

Last modified: Mon Nov 13 18:31:53 JST 2000

東洋学文献類目(京都大学人文科学研究所附属東洋学文献センター発行)のタイトルをWWWから検索します。

検索文字列

文字符号 \* ISO-2022-JP ↑ Shift\_JIS  
↑ EUC-JP

検索開始

[database]  
Powered by XEmacs UTF-2000.

検索結果

36 件見付かりました。

ID	A96141400008
題名	伏羲與歐亞奈斯生物 (伏羲與歐亞奈斯生物)
著者	關靜芬 (關靜芬)
ジャーナル	遠東大學學報(哲學社會科學) 2 号 (總第138期) pp.67-71 (遠東大學學報(哲學社會科學)) (總第138期)[296200]
発行年(月)	1996
分類	通論 (1410)
地域	中国 [110000]
時代	古代 [A2]
C	11A2XX1
C3	00
C4	1
N3	00001
N4	11840
Z	2

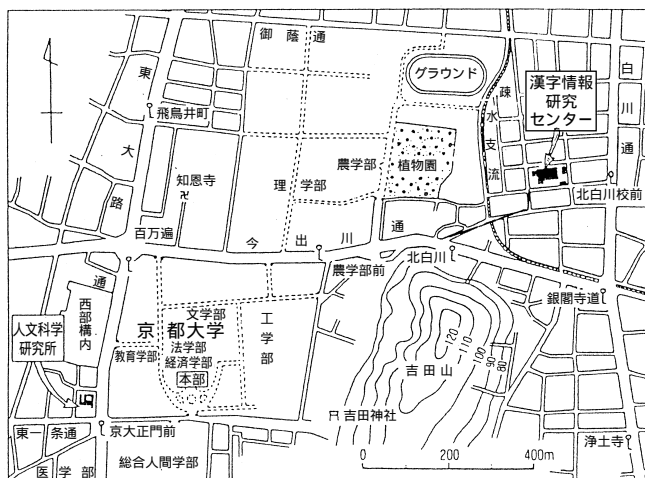
## 【DICCS NEWS】

・東洋学文献類目について、従来の作成作業を見直し、Web上でスムーズに公開できるような新方式を現在検討中である。海外でのデータ入力を実現させるために、2000年10月31日から11月4日にかけて、森時彦東洋学研究所主任・高田時雄センター主任・安岡孝一センター助教授の3名が、台湾に出張し、東洋学文献類目入力フォーマットの開発について現地関係者と打ち合わせを行った。なお、冊子での配布はこれまで通りに行う予定である。

・2001年3月9日に、全国漢籍データベース協議会第1回総会が学術総合センターで開催された。詳細は、本号6頁の記事を参照のこと。



・本年度の補正予算で、大型サーバー(富士通 GP7000s モデル45)を購入し、さらにギガビット・イーサネット化を実現した。これによって、本センターが構築中の漢字情報データベースは、インターネットを介して高速に検索することができるようになるだろう。また、館内に無線LANの情報ネットワークシステムを構築し、研究室や書庫のどこからでもノートパソコンで自在にアクセスすることが可能になった。



発行日 2001年3月15日

発行所 京都大学人文科学研究所附属  
漢字情報研究センター

〒606-8265 京都市左京区北白川東小倉町47

電話 075-753-6997 FAX 075-753-6999

<http://www.kanji.zinbun.kyoto-u.ac.jp/>