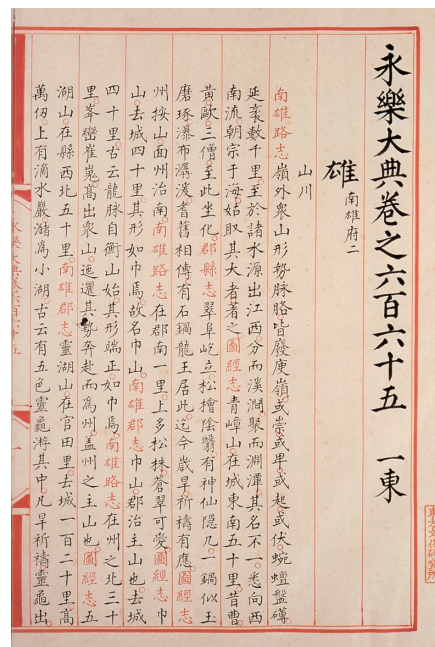


# 漢字と文化

漢字文化の全き継承と発展のために

京都大學 21 世紀 COE 東アジア世界の人文情報學研究教育據點

## 第 4 号



### 目次

「唐代史料稿」とナリッジベース .....	2
雲岡石窟通楽寺の発見 .....	4
インターネット時代の人文学の新しい技術 .....	6
人文情報學いまむかし .....	10
マークアップテキストがもたらす利便性 .....	12

大唐西域記序

攝寺

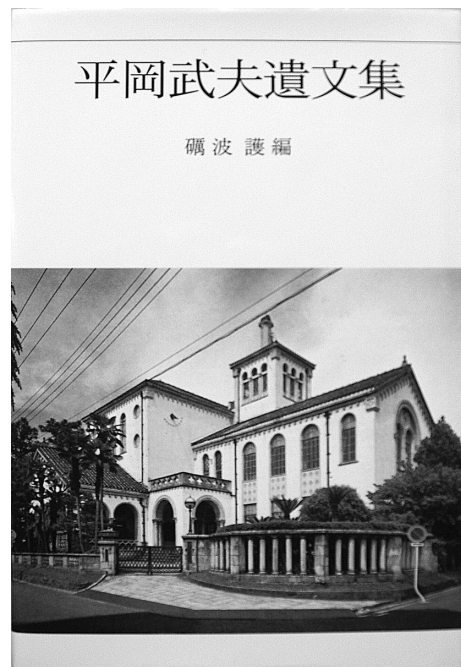
## 「唐代史料稿」とナリッジベース

井波陵一

「個人のひそやかな資料整理の方法でなし得る研究は、ほとんど限界に来ているのではあるまいか。個人の能力を越えた大きな規模において、そのような規模においてはじめて可能になるような大きさの資料の可及的に完全な整理を、現在の東洋学は必要としている。真にその名に値する共同研究も、このような史料の整理の後に、はじめて可能になるのである。個人の私室がいくつ集まっても、共通の広場にはならない。共同研究が成立するためには、共通の形式によって十分に整理された資料を、研究者のすべてが、ひとしく、同じ条件のもとにおいて、共有していることが、欠くことのできない必要条件である。この過程を経た後に、あらためて、個人研究は、新たな次元に高められてゆくのである。」

以上の文章は現在進められている「唐代研究ナリッジベース」プロジェクトの基本宣言の一節である、と申しあげても、それをいぶかしく思う人はたぶんいないだろう。「共通の形式によって十分に整理された資料を、研究者のすべてが、ひとしく、同じ条件のもとにおいて、共有していること」——我々が目指す包括的な電子アーカイブとは、まさしくそうした前提を実現するものとして構想されているからである。しかし、これは我々の宣言ではない。今を遡ること50年前、1954年8月のものである。執筆者は平岡武夫。人文科学研究所が誇る資料集、「唐代研究のしおり」の1冊である『唐代の暦』発行に際して書かれた文章に見える。「唐代研究のしおり」は内外の研究者から高い評価を与えられ、1977年には正式に復刊された。平岡氏がこだわりを持って選ばれた「朱色」の表紙は、この事業に携わった人々の熱意と苦勞を物語るかのように力強い輝きを放っている。

ところで、同じ一文の中で、氏は次のように述べている。



「これらの一連の書物は、なるだけ早い時期に印刷して、すべての唐代文化の研究者に利用していただく意志をもって造られたものである。しかし、これらのものは、いずれもお未定稿であって、その定稿化を、私は、各種の資料の整理とにらみあわせて、三・四年の後に予定していたのである。いま急速にこれを定稿化することは、容易なことではない。加うるに、一方において、この時、私は、冊府元龜・唐会要、その他の唐代史料と、力いっぱいにとり組んでいた。」

一方においてとり組んでいた仕事、すなわち『冊府元龜』『唐会要』、その他の唐代史料に関する整理事業は、どのような形で結実したのだろうか。これこそ標題に記した「唐代史料稿」と呼ば

れるものなのである。我々はその仕事の目的と経過を『東方学報』に掲載されたいくつかの文章によって知ることができる。いま「唐代史料稿長慶元年」(1966年)なる一文によれば、事業内容は7項目に分けられる。

(1) 冊府元龜一千巻の宋刊本および明鈔本、唐会要一百巻の明鈔本のマイクログラフによる撮影と複製

(2) 冊府元龜の中から唐代史料の抽出

(3) 冊府元龜・唐会要の本文校勘

(4) 冊府元龜・唐会要の史料を三万五千枚のファイルに作成

(5) 史料ファイルの編年体による再編成

(6) 旧唐書・新唐書・資治通鑑・全唐文などの史料的整理

(7) それらの史料と冊府元龜および唐会要との総合整理

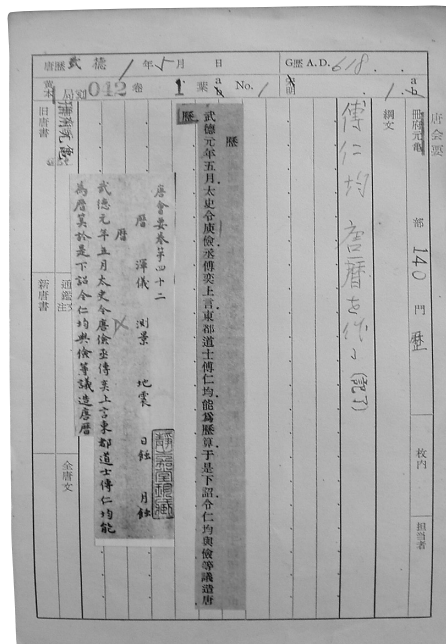
印刷に付された「唐代史料稿」とは、『東方学報』に3回にわたって掲載された(7)のはじめの部分のことを指すのであり、言うまでもなく、この事業全体から見れば、ほんの一部にすぎない。ただ、かの一文は続けて次のように述べる。



「ファイルの整理をすませているので、研究所に来れば、どなたでも利用できる。また利用者もすでにある。しかし、ファイルの体制では、利用

にも、管理にも、不便が多い。少くとも、「唐代史料稿」の形にまとめることが必要である。」

したがって、全ファイルを包括的にまとめ上げたものこそ、正式の「唐代史料稿」になるはずであった。しかし、この構想は実現に至らず、公開利用に供したファイルもいつしか顧みられなくなり、条件の悪い部屋に詰め込まれた結果、特注のファイルボックスも雨水の浸透によるカビに冒されて一部が破損し、中のファイルが散乱するという惨状を呈している(平岡氏がファイルに貼りつけた複製写真の劣化を危惧してから40年近く経過している)。



これら貴重なファイルを何らかの形で「唐代研究ナレッジベース」に活かすことは、我々の責務と言えよう。もちろん、たとえば新旧唐書の人名索引のファイルなどはすでにその役割を終えている。しかし、今では1データにしかならないファイルの一つ一つ眺めていると、かえって「共通の広場」の建設を目指した先輩がたの並々ならぬ意欲に圧倒されてしまう。我々のプロジェクトは、人文科学研究所が先端的に進めてきた共同研究の必然的な産物にほかならないのだ。そう思うと、責任の重さを改めて痛感する。

## 雲岡石窟通楽寺の発見

向井佑介／岡村秀典

人文科学研究所の前身である東方文化学院京都研究所は、1938～1944年に山西省大同市の雲岡石窟を調査し、水野清一・長広敏雄『雲岡石窟 西暦五世紀における中国北部仏教窟院の考古学的調査報告』全16巻32冊（1951～1956年）として大部の報告書が刊行された。調査では石窟の実測と撮影のほか、石窟前面や台地上などが部分的に発掘され、石窟に付属する寺院の遺構が検出されたのであるが、出土遺物についてはほとんど未整理のまま人文研の収蔵庫に眠ってきた。

そうした考古資料は、文字通り瓦礫のようなものだろうか。そんな疑問から、とりあえず文学研究科の大学院生らとともに瓦の整理からはじめることにした。本誌第2号に紹介したように、文字瓦について考古学の目で詳しくみると、これまで知られていなかった事実がいろいろわかってきた。その勢いで2004年には雲岡石窟から出土した陶磁器の調査をはじめたのである。

ここで紹介するのは、第20洞前から出土した遼金代の白磁である。第20洞は沙門統の曇曜の発議によって最初に開鑿された5窟のうちの1つで、



雲岡石窟第20洞

石窟の中心となる大仏が安座している。白磁は壺底部の破片で、底部中央に「寿□」の2文字、高台内面に手慣れた楷書体で「通楽館置」と墨書している。前者の不明の1字は他とちがう記号風の筆致だが、「昌」の草書体で遼の年号「寿昌（1095-1100）」を記したものかもしれない。後者の墨書はこの白磁の設置場所が「通楽館」であることを示したもの。この「通楽」が積年の問題を解決する重要な手がかりとなった。

第2次世界大戦の後に北京大学の宿白氏が北京大学図書館で発見した元の熊夢祥『析津志』に引く「大金西京武州山重修大石窟寺碑」によれば、北魏の雲岡石窟寺は「通楽」「靈巖」「鯨崇」「鎮国」「護国」「天宮」「崇教」「童子」「華巖」「兜率」の十寺からなり、唐の貞観十五年（641）に重建された後、遼の重熙十八年（1049）に欽愛皇太后により再修、遼末まで数次の小規模な改修が行われ、遼の滅亡時に灰燼に帰すものの、金の皇統三～六年（1143-46）にふたたび大規模な造修がなされたという。この北魏十寺の筆頭にあげられる「通楽」こそ、『統高僧伝』巻1 釈曇曜伝に雲岡石窟を開いた曇曜が「恒安の石窟通楽寺に住む、即ち魏帝の造る所なり」という通楽寺にほかならない。

通楽寺の所在地について「金碑」を考証した宿白氏は曇曜五窟（第16～第20洞）付近と推定していたが、今回の墨書の発見はそれを裏づけることになった。もっとも、曇曜の通楽寺にちなむ「通楽館」なる寺院が遼金代の第20洞前に実在したことが明らかになり、北魏代の建築遺構や大量の瓦がその下層から出土しているとしても、その場所が本来の曇曜の通楽寺であったとは断定できない。とはいえ北魏の和平年間（460-465）に開かれた

石窟は第20洞を含む曇曜五窟のみであるから、曇曜の住んだという通楽寺がその近辺にあったことはまちがいないだろう。

第20洞前からは同時に「□統四年」の文字を刻んだ丸瓦片が出土している。水野清一らはこれを遼の乾統四年（1104）に比定したが、「金碑」によって金の皇統四年（1144）の可能性もありうるだろう。いずれにせよ第20洞前から遼金代の瓦が大量に出土し、「金碑」の記述と合わせると、北魏から遼金にいたるまで石窟の前面に木造の仏殿

が繰り返し修築されていた様子が明らかにできた。また、ここから出土した遼金代の陶磁器はほとんどが日常雑器の白磁碗であるため、寺院内には多数の僧侶が生活していたこともうかがえる。ブタやウマの骨は、僧侶が肉食していた証かもしれない。そうした生活の一端がみえてくるのも、考古学ならではの研究成果である。

考古資料も捨てたものではない——なかば捨てられたような「お蔵入り」の遺物を前に、そんな思いを強くしたのである。



雲岡石窟第20洞前の「通楽館置」白磁壺

サマーセミナー実施報告（1）

## インターネット時代の人文学の新しい技術

武田時昌

2004年9月6日（月）から10日（金）までの5日間、人間・環境学研究科棟433・434演習室において、東アジア人文情報学サマーセミナーを実施した。テーマは「インターネット時代の人文学の新しい技術」。その目的、カリキュラム及び反省等の実施報告を、来年度以降の展望を兼ねて以下で略述したい。

我々の21世紀COEプロジェクトでは、漢字文化の全き継承と発展のために、人文学に情報学的手法を導入し、漢字文化圏に関する人文情報学を新たに構築しようとしている。したがって、従来の文献学的方法論の習得とともに、新たに情報学の技能を身につけた人材育成の教育プログラムを立ち上げる必要があるように思われる。

パソコン、インターネットが急速に普及し、社会の情報化が進展した今日において、コンピュータや情報通信ネットワークなどの情報手段を適切に活用し、主体的に情報を選択・処理・発信できる能力は、必須の教養である。情報活用の実践力を培う「情報教育」に関して、近年の学習指導要領の改訂によって、中学では技術・家庭科で「情報とコンピュータ」が必修となり、高校では普通科目「情報」の必修授業が新設されるに至った。大学教育の現場においても、情報を活用するための基礎的な知識・技能＝情報リテラシーの向上をめざす授業を、共通教育科目に設けようとする気運が高まっている。しかしながら、大学院研究科の専門教育カリキュラムに組み入れられているほどではなく、まして漢字文化の継承と発展のための授業やセミナーは皆無と言ってもいいだろう。

そうした状況を打破し、人文学と情報学を横断する教育の場を設けるために、人文科学研究所と人間・環境学研究科の事業推進担当者を中心とし

て協議を行い、人間・環境学研究科の正規カリキュラムにおいて「東アジア人文情報学」の新設を内定させた。ただし、その開講は、人間・環境学研究科が改組後5年を経っていないため、カリキュラム変更が可能となる平成18年度以降になる予定である。そこで、そのパイロット・プロジェクトを兼ねて、博士後期課程の学生を対象とするサマーセミナーを企画することにした。

サマーセミナーの究極の目的には、COEプロジェクトの2大柱、東アジア言語のコンピュータ処理、東アジア文献のナレッジベース化を推進する若手研究者の育成がある。しかし1週間という短い期間であるから、多言語処理やデータベースのための諸技法のなかで、最も基礎的であると思われるものに対象を絞り、パソコン実習を行うことにした。



対象となる大学院生の予備調査では、パソコンをWord、一太郎等のワープロソフトで使用している者が大半であり、研究論文や資料を汎用性のあるデータベースにする工夫を行っている者はごく少数に限られていた。しかし、ホームページ作成には大いに関心があり、すでに立ち上げている

学生も少なからず存在した。したがって、ウェブページを書くためのHTML (Hyper Text Markup Language) やその基礎知識について、もっと学びたいという希望が最も強かった。

そこで、本年度のサマーセミナーのテーマとして「インターネット時代の人文学の技術」を掲げ、TeX と XML という二つの手法を取り上げることにした。

TeX とは、スタンフォード大学の Donald E. Knuth 教授が開発した文書組版ソフトであり、TeX の機能を強化した LaTeX, 日本語が使えるようにした pTeX 等がさらに考案されている。それらは、自分の思ったとおりに印刷用の出力が機種、OS の環境に依存しないで得られるフリーソフトであるために、数式、化学式等を多用する理系研究者を中心に広く用いられている。

COE プロジェクトが構想するデータベースという見地において、TeX の文書整形技法で注目されるのは、タグづけによって文章の構造や修飾情報を記述していくマークアップ言語であるという側面である。SGML (Standard Generalized Markup Language) から発展した HTML (Hyper Text Markup Language) もマークアップ言語の一つであるが、その言語によって記述された文書はテキストファイル形式であるため、文系の研究者にとっても、汎用性の高い論文、資料を作成することのできる利点がある。

一方、XML (eXtensible Markup Language) は、SGML をインターネットで活用しやすくするために1998年2月に開発されたもので、タグが固定的で表示に特化したHTMLの制約を克服しており、XML 文書は汎用性のあるデータ交換の標準フォーマットとして現在最も期待されている。

以上のような現状を踏まえて、セミナーでは、TeX, LaTeX の文書整形技術の訓練を導入とし、TeX から XML へという流れにおいて、文章の論理構造、意味構造を記述するマークアップ手法に理解を深め、XML の基礎を実習することにした。そして、さらに応用としてプログラム言語の Perl, XML の HTML 変換用プログラム XSLT の初歩を概説してみることにした。

5 日間にわたる講義テーマと講師は、以下の通りである。

第1講 漢字情報学序説

「入力、出力、そして検索」 安岡孝一

第2講 TeX 入門

「人文系研究者のための文書整形術」真下裕之

第3講 LaTeX 応用

「効率的な文書作成のために」 守岡知彦

第4講 多言語化概論

「Ω, TeX と OTF パッケージ」 宮崎 泉

第5講 マークアップ概論

「マークアップ概念の導入— LaTeX を例に」  
守岡知彦

第6講 XML 入門

「The Basics — Understanding and Using XML」  
C. ウイッテルン

第7講 XML 応用

「Applications of XML— XML for Text Encoding」  
C. ウイッテルン

第8講 XML プログラミング入門

「正規表現と XML 文書の操作」 師 茂樹

第9講 XSLT 入門

「XML で XML を処理する？」 岩井茂樹

第10講 多言語情報処理概論

「漢字政策と漢字情報処理」 安岡孝一

なお、パソコンによる実習指導を徹底するために、受講生は10名程度に限定せざるを得なかった。選考方法は、人間・環境学研究科の指導教員に依頼して博士課程学生を推薦してもらうとともに、ホームページで受講希望者を公募し、7月中旬に選考委員会を開き、書類審査によって選定した。

受講した者10名を列記すると、人間・環境学研究科 D4の鍾翀、崔水晶、D2の陳捷、桂由起、D1の閻淑珍、文学研究科 D2の鄭宰相、高永聡、石野一晴、他大学からは立命館大学文学研究科研究生の秋山陽一郎、関西大学大学院文学研究科 D3の大井和也、以上の10名の諸君である。また坂内榮夫岐阜大学教育学部助教授も聴講者としてセミナーに参加した。実習指導員には、セミナー講師とともに COE 研究員の山本一登氏に加わってもらった。また、京都ノートルダム女子大学人

間文化学科3年の車愛順さんが、ゼミの指導教員吉田智子助教授を通じてセミナーの見学を申し出たので、ビデオ記録係を務めてもらうことにした。

さて、サマーセミナーを終えてまず感じたことは、短期間の小規模な企画にもかかわらず、思った以上に大変であり、人文情報学あるいは漢字情報学の立ち上げを口で言うのは簡単だが、相当な覚悟がいるということである。参加者も普通の集中講義なら馴れているが、コンピュータ実習を兼ねたものであったので、かなりハードな一週間だったようである。ところが、しばらくして感想レポートを提出してもらったら、とても疲れたがその分充実感もあり、大いに勉強になったという意見が大半だった。たとえそれがお世辞半分としても、そこでようやく責任者としてホッと胸をなでおろすことができた次第である。

全体的に振り返ってみて、初めての試みということもあり、反省すべき点は多々ある。まず受講生の感想レポートに異口同音に述べられているように、情報量が多すぎて消化不良であった印象は拭えない。少人数とはいえコンピュータへの習熟度にかかなりのばらつきがある受講生を相手に、TeXやXMLの実習を行うのは、思った以上に困難さが伴うということである。基本的な教え方として、インストールからすべて自力で行えるようにするというにしたので、インストールや基本操作に馴れるのに少し時間がかかるのはいたしかたがない。しかし、パソコン操作に気を取られすぎると、何をしようとしているかの理解が疎かになってしまう。



教える側も今回が初めての試みという側面もあるが、教える内容と手順を十分に吟味する必要がある。また、事前学習の必要性も検討すべきであるが、そうした準備段階での工夫で、かなり解消できるように思われる。

最も大きな問題は、やはり講義、演習のどのようなカリキュラムを組めばいいのか、そしてどの程度のレベルで教えればいいのか、ということである。正規の授業ではなく、一週間で集中的に行うセミナーであるだけに、学問的な基本概念を教えることなく、パソコン操作の実習に終始すると、まますると初心者向けのパソコン教室的なものになってしまう。また、学生の関心を向けるために、データベースの横断検索やn-gram処理の応用例を紹介するのも悪くはないが、プログラミングやコンピュータ処理の基礎知識が理解できていないと安直な考え方を植え付けてしまいかねない。そうした危惧を回避するために、概論と実習をワンセットとし、1日目に入門、2日目に応用という段階的なメニューを組んだのであるが、教える側と受講生のレベルの差がありすぎて、果たして狙い通りにいったかどうかは疑問である。

受講生の感想レポートによれば、1日目のEmEditorとM's TeX Helper 2を使ったTeXの講義と演習は、比較的やさしかったために理解しやすかったらしく、組版ソフトであるTeX、LaTeXによって論文を作成することに強い関心を示した。実習後も勉強を重ね、TeXによるレポートを提出してきた者も複数いたくらいである。

ところが、その応用となると全体像がつかめずに中途半端な理解に終わってしまった印象が強い。XEmacs CHISEのインストールや基本操作の説明で時間的なロスがあつて話の途中で終わってしまい、論理構造の記述するという特性を把握したり、多言語処理の手法という側面は、十分に教えることができなかった点は残念であった。

3日目以降のXMLに関しては、XML自体に馴染みがないために、さらに面食らったようである。基本的な知識が乏しいために、何のためのパソコン処理であるのかがわかりにくかったにちがいない。実習では、マークアップされたXML文



書を<oXygen/>を使って組版し、PDFに出力する仕方を取り上げ、XMLをパソコン上で走らせることで親しみを持たせようとした。しかしながら、初心者向けの操作法を教える工夫がやや不足していたので、かえって逆効果で、難解であるという印象を持たせてしまったかもしれない。マークアップ概念をもう少しわかりやすく概説したり、受講生が普段取り扱っている文献によってダグ付けする具体的な実習を取り入れてもよかったように思われる。今後の課題としてマークアップ手法の簡便な教材作りは、欠かせないだろう。



ただし、受講生の意欲として、テキストファイルによる組版のテクニックよりも高度なデータベース化を学びたいという気持ちがあって、関心の度合いとしては、むしろXMLのほうが高かったことも見逃してはならない。

というのも、レベルの高い話にまったく興味を

示さなかったかということ、そうではないからである。とりわけ師茂樹氏の講義で、正規表現を用いた検索の実例を示すことでXML文書の構造的な本質を論じたこと、岩井茂樹氏の講義で、共同研究会の会読テキストとして用いている『元典章』を具体例として、マークアップテキストのプログラミング処理の実例を示したこと等に、受講生は知的な驚きを伴って強い関心を抱いたようである。プログラミング学習について、参加する前は、自分は文系人間だからと及び腰の受講生もいたが、これからの人文学にはデータベース作成が不可欠になってくることを痛感したにちがいない。だから、そのようなスキルを身につけることができれば是非とも活用してみたいと参加者全員が感じたことも確かである。そこに、サマーセミナーの役割を見いだすことができるだろう。

以上の反省を踏まえて、来年度以降も人文情報学サマーセミナーを継続して実施する予定である。セミナー受講生には、今回の実習体験を通して得られた技能と知見をさらに深め、今後の研究活動に大いに活用してもらいたいし、アフターサポートの協力も惜しまないつもりである。

近い将来において、サマーセミナー修了証書を手にした若手研究者が全員集まって、東アジア人文情報学の研究成果を発表し、討議し合う集会在企画できるかもしれない。サマーセミナーの時期は、まだまだ残暑が厳しく、旧式の白髪頭になりつつある聴講生には一週間の長丁場はしんどいが、やがてそんな楽しい「同窓会」のある夏休みが来ることを心待ちにしている。

サマーセミナー実施報告（2）

## 人文情報學いまむかし

坂内榮夫

筆者がコンピューターを利用し始めたのは、八十年代後半の頃である。普通のワープロとして論文製作に利用していたほかに、麥谷邦夫助教授（当時）がMS-DOS上で動く一字検索システムをTurboPascalで構築されたので、それを利用して頂くようになった。そこで、検索用元データを作成するため、大型計算機センター（当時）の二階にあったOCR専用機を使ってプレーンテキストを幾つか作成した。データの種類は、花園大學内の禪文化研究所で禪語録研究会の末席に連なっていた関係上、馬祖・百丈・黄檗・臨済の所謂四家のデータを作成する事にした。当時、入矢義高先生の手によって馬祖・黄檗・臨済は譯注が公開されていたので、この三人に関してはその譯注によってデータを作成すれば、底本についてとりあえずの問題はなかった。ところが、百丈については譯注も研究成果も公開されていなかったため、一般に利用されていた宇井伯壽氏校訂の『百丈廣語』（『第二禪宗史研究』所収）を底本にした。これは、『古尊宿語録』や『四家語録』から百丈の説法・問答を集め、そこに『宗鏡録』や『傳燈録』等の引用文との校勘を加えたものである。

しかし、その当時には既に柳田聖山先生の研究により、『百丈語録』の最古のテキストとして、福州東禪寺版や開元寺版『天聖廣燈録』（卷九）の存在が知られていた。更に、『百丈廣語』の最後に付されている「大乘入道頓悟法門」に関しても、引用元である『祖堂集』と『傳燈録』の間に夥しい文字の異同が存在する。この様な、『古尊宿語録』『四家語録』と『天聖廣燈録』との間の文字・文章の異同、また「大乘入道頓悟法門」間相互の異同等の情報は、プレーンテキストの形では保存の方法がなく、異同をカッコに括って書

き加えておくしかなかった。更に、個人的に検出していた『百丈語録』の引用文・典拠などについての情報に至っては、有効な電子化の方法を考えつかなかった。「電子データは、プレーンテキストの形で一字検索に利用する以外に有効な活用法はないのか」と考え始めていた当時の筆者はこの段階で早くも挫折し、以後十年近くこの問題については放置したままであった。

ところが、近年家人が漢字情報研究センターの武田時昌教授の元に内地留學に行き、C. Wittern 助教授や守岡知彦助手の指導を受け「『説文解字』データベース作成の研究」を始めた事から、XMLの古典文獻用タグセットであるTEI (Text Encoding Initiative) や、漢字を文字コードを離れて扱うことのできるCHISEプロジェクトの存在を知った。とりわけ、TEIによるマークアップを行えば、プレーンテキストでは保存できなかった文字の異同情報や引用文・典拠という付加情報も有効に保存する事が可能である事。更に、XMLからHTMLに変換してブラウザ上で表示したりデータ処理する事も可能であり、またXEmacs (Emacs) 上に限定されるのではあるが、XMLテキストをElispにより様々にデータ処理することのできるプログラムも開発中である事も知った。

その結果、長い間お蔵入りしていた『百丈語録』の電子データも、コンピューター上で有効に利用する事ができそうだと考えるに至った。そこで、個人的にWittern助教授や守岡助手の指導を受けに行き、PC-UNIX上で動作する「XEmacs-TEI」や「XEmacs-CHISE」の使用方法について習熟に努めるようにした。そして、今回自らの人文情報學研修の一環として「東アジア人

文情報学サマーセミナー」を聴講させて頂いた次第である。

今後は、上記『百丈語録』のマークアップを行ない、テキストの異同や典拠等についてすべて電子化し、コンピューター上で利用できるようにしてみたいと考えている。そのための準備として、現在手始めに讀書會で讀了した南朝宋顔延之の『庭誥』<sup>[1]</sup>について、中華書局版評點本を底本にして、宋版『冊府元龜』と明版『冊府元龜』の文字の異同を中心に個人的にマークアップを行っており、それはほぼ完成している。マークアップ完成後、XSLを用いてどのように文字の異同処理や他の文字データ処理できるのか、現状ではXSLについて殆ど知識がないので、CBETAのソースを参照するなどして色々勉強し、異同の処理だけでも最低限実現してみたいと考えている。サマーセミナーで岩井茂樹教授の研究成果を見せて頂いた感じでは、かなり柔軟に複雑な処理も行なえるようなので、XSLの処理能力については非常に期待している。その次の段階として、譯注についてもどのようにリンクさせて処理できるのか、検討してみる予定である。また、Elispによるデータ処理プログラムについては、その開発状況を睨んで色々利用法を探ってみるつもりである。

なお、この『庭誥』に関しては、諸々のマークアップが完成した暁には宇佐美文理研究代表と相談して、適当なWEB上で公開できればと考えている。そして、『庭誥』で基本的な問題点・疑問点を解決したのち、本来の目標である『百丈語録』のマークアップに取り掛かりたいと考えている。

現在は、先に少し名称を出したように、TEIのマークアップを行なうのにKnoppix<sup>[2]</sup>上で「XEmacs-TEI」を使用している。周知のように、PC-UNIX系OSはそのインストールから環境設定まで非常に煩雑な作業が多く、UNIXシステムに對しての基本的な知識がないと、實際問題として作業を行なうにも困難な場合が多い。更に、

Emacs系のエディターを扱うにも、フォントまわりを始めとして様々な設定が全く複雑怪奇で、誰でもが簡単に使用するにはまだまだ敷居が高いのが現状である。しかし、守岡助手がサマーセミナーで配布された「Knoppix-CHISE」<sup>[3]</sup>は、Knoppixに「XEmacs-CHISE」やフォント等を付け加えて独自に再構成したもので、「XEmacs-CHISE」やフォントの設定及びその他必要な環境設定の類いが豫め施されている。その結果、「Knoppix-CHISE」導入後直ちに「XEmacs-CHISE」が使用可能になっており、「XEmacs-CHISE」や「XEmacs-TEI」<sup>[4]</sup>の普及と利用に多大の福音を齎す可能性を秘めている。今後「Knoppix-CHISE」がよりいっそう進化して、使い勝手のよくなる事を願ってやまない。

[1] 「六朝隋唐精神史研究」(代表: 京都大學文學研究科 宇佐美文理助教授)の科研報告書の一つとして譯注を公刊するため、現在原稿の整理を行なっている。

[2] DebianベースのLinuxディストリビューションの一つ。ドイツのK. Knopper氏が開発している。それに基づいて、獨立行政法人産業技術総合研究所の須崎有康氏が、日本語の使えるように改良した日本語版を配布をしている。URLは以下の通りである。「<http://unit.aist.go.jp/itri/knoppix/>」

[3] 以下のURLで「Knoppix-CHISE」のISOイメージとルートイメージが公開されている。「<http://kanji.zinbun.kyoto-u.ac.jp/projects/chise/dist/KNOPPIX/>」

[4] 残念な事に、C. Wittern 助教授が独自に開発された、TEIの設定は取り込まれていない様である。この点については、是非今後の改良をお願いしたい。



サマーセミナー実施報告（3）

## マークアップテキストがもたらす利便性

秋山陽一郎

2004年9月6日から5日間に渡って行われた「東アジア人文情報学サマーセミナー」は、筆者がTeXについて素人であったこと、さらに唐代ナレッジベース・プロジェクトに従事するためのスキルアップという個人的な需要もあって非常に有意義に受講することができた。受講生は京大大学院の人間・環境学研究科と文学研究科の留学生が過半数を占め、前半がTeX、後半がXML、それを安岡孝一助教授による文字コードの講座で挟むような構成で行われた。人文情報学という学問分野が依然未熟な状態にある現状を考えれば、人文科学の若手院生をターゲットに「マークアップ」という枠組みで集中講義が行われることは、非常に貴重かつ意欲的な試みといえるものの、惜しむらくは5日間という短期間に多くの内容を凝縮してしまったため、受講生にはやや混乱をきたしてしまった感が否めない。半期ないし通年の単位で、このような講座が開講されることがあれば、より大きな効果を期待できるのではないだろうか。

さて、この小論では当セミナーを通じて感じたことを、特に「マークアップテキストの意義」という焦点に絞って綴ってみたい。

TeX や XML を使えらるとどう幸せになれるのか？

当セミナーの受講者は、国籍や関心、スキルなどがまちまちであったこともあって、自然、受講者は講義内容を必要に応じて銘々の関心や研究分野にアジャストして理解を深めていくことになるが、この点で具体的な事例の紹介を交えるといった受講者各自が明確なイメージや目的を持てるようにするための配慮が若干不十分であった気がする。特にXMLについては、具体例が出たのが最終日（第9講）の岩井茂樹教授の時点になってし

まったこともあって、かえってXMLが面倒なフォーマットなのではないかと戸惑った受講者も少なくなかったことと思う。この点、PDF変換方法についての説明が第7講においてChristian Wittern 助教授よりあったが、OpenOffice.orgのような既成のエディタに実装されているPDF生成機能や、2日目までに紹介されたTeXのPDF変換と比べると、well formedなマークアップを施しておく必要がある分、受講生の間では逆に「なぜわざわざXMLにするのか？」という疑問を生じる結果となってしまった。XMLというフォーマットを利用する目的や場面に対する説明が足りなかったことも、そうした疑問を増幅させてしまう結果に繋がってしまったのだろう。TeXについても、第2講に真下裕之氏より非常に懇切な説明をいただいたが、TeXのメリットとして挙げられていたものの大半（たとえば文書スタイルの管理など）が実はMS Wordなど既成のWYSIWYG型のエディタに実装されているため、環境移行の面倒さをおしてまでTeXに乗り換えようという意欲が受講生の間では起きにくかった印象がある。

このように、TeXやXMLを使えらるとどのよ



うに便利なのか、というイメージが具体的に湧かなかったことが受講者の泣き所になってしまった。

### データ交換形式としての利便性

TeX や XML がもたらしてくれる最大の利点は、何といてもクロスプラットフォームの下(異なる OS や端末間)でのデータ交換にある。TeX も XML もテキストベースのマークアップ言語であるため、Windows や Mac OS, Unix といった様々なプラットフォーム上で全く同一のデータを共有することができ、しかも比較的認知度の高い規格であるため、対応しているツールが多数公開されているのも強みといえる。特に組版ツールとしての TeX などは、この部分での利便性が指摘のレベルにとどまることなく、もっと強調されても良かったのではないだろうか。

昨今、オフセット入稿によって自費出版が安価で容易にできるようになってきたが、執筆者と編集者との原稿のやりとりを MS Word や一太郎といったワープロ原稿で行っている例をしばしば見る。ところが、少なくともこれまでの筆者の経験では、このようなケースの多くはレイアウト面で必ずと言って良いほどトラブルが発生している。執筆者と編集者との間で使用している OS が異なる場合はもちろんのこと、OS やアプリケーションが全く同じ場合においてさえも、微妙な環境差や設定の違いだけでレイアウトは少なからず変わってしまう(特に段落がえの改行絡みでページ数が変わってしまったり、図表の位置やマージンでトラブルが発生することが多い)。このような作業環境差に由来するトラブルの大半は TeX を利用することで回避できるはずで、TeX のスキルを身に着けるという煩わしさは依然としてあるものの、クロスプラットフォーム下でほぼ同一レイアウトを共有でき、しかもツールの多くがフリーで入手できるため、少なくとも編集側にとっては頭痛の種を減らす有力な選択肢になるはずである。

### データの再利用性という要素

テキストベースのマークアップ言語の今ひとつの利点として、データの再利用性という要素も欠

かすことができない。XML はもちろんのこと、TeX もまた HTML への変換が容易に行えるため、人文系研究者として執筆原稿をたちどころに PDF や Web ページに変換することができる点ができるという点は大きな魅力といえる。この方面において最も重要な意味を持っていたのは、何といても守岡知彦氏の第 5 講「マークアップ概論」における論理構造的マークアップの話であろう。論理構造のしっかりしたマークアップが施されているデータは、XML であろうと、また TeX であろうと、データの再利用性が非常に高いというマークアップのいわば基本を強調された内容であったが、非常に重要かつ実用的な内容だっただけに時間が足りなくて消化不良に終わってしまったことが悔やまれる。

マークアップテキストの再利用性については、SGML のサブセットの一種である HTML の作成を通じて筆者も日頃から痛感しているところで、一つだけ簡単な事例を挙げるとすれば、

#### (1) 見た目の書式でマークアップした場合

```
<font size="6"><b>第一章</b></font>
<font size="4"><b>第一節</b></font>
<p>本文<font size="2">注釈</font></p>
```

#### (2) 論理構造でマークアップした場合。(見た目のスタイルは CSS で一括指定)

```
<h1>第一章</h1>
<h2>第一節</h2>
<p>本文<span class="note">注釈</span> </p>
```

(1) のような見た目のデザイン本位のマークアップをしたのでは、融通の利かないマークアップデータにしかならない。文章の論理構造は、文章そのものを改変しない限りは原則として不変だが書式情報は可変である。たとえば、ウェブサイト全体のデザインを変更する時など、全ての HTML ファイルの全てのマークアップを一つ一つ書き換えていく必要があるが(無論、事前に一定のマークアップルールさえ定めてあれば、機械的に一括置換をかけることも可能だが)、(2) のように“見出し 2”・“見出し 3”として論理構造本位のマークアップを施し、デザイン面を外部スタイルシートに任せておけば、そのスタイルシー

トを差し替えるだけで、個別の HTML ファイルを何ら編集することなく全てのページのデザインを一括変更することができ、また既成のエディタに取り込んで印刷用文書として再利用するのも容易にできるようになる。またこれがもし第三者にとっても有用なコンテンツを含んでいるとすれば、第三者がこれを取り込んで再利用する上でも非常に扱いやすくなるはずである。(たとえば The Web Kanzaki で公開されている自動 RDF 生成スクリプト — <http://www.kanzaki.com/docs/sw/dc-a-matic> などを利用して RSS という形態で再配信することもできる。)

#### 人文系研究者の視点でマークアップを最適化

このような論理構造本位のマークアップを施していく利点はほかにもある。花園大学の師茂樹氏の講座にあった DOM による XML 操作を少しばかり活用しただけでも、ただの検索が飛躍的に便利になる。たとえば、十三經注疏のテキストデータを以下のようにマークアップすれば、経・伝・注・疏・陸徳明音義を区別して検索することができるし、各ソースに成書時期を属性として付与すれば、各ソースが編まれた時代を条件として指定することもできるようになる。

```
<div>
<篇名>關雎</篇名>
<div>
<經>關關雎鳩，在河之洲。…</經>
<傳>興也。關關，和聲也。…</傳>
<注>箋云，摯之言至也。…</注>
<音義>雎，七胥反。鳩，九尤反。…</音義>
<疏>正義曰，毛以爲關關然，聲音和美者…</疏>
```

```
</div>
(中略)
</div>
```

ただ、こうした論理構造のマークアップは、書物全体の構造が正しく理解されていないと行えない。ここに人文系研究者の需要がある。実は上記のような、XML を利用した十三經注疏の検索システムは、すでに台湾の中央研究院歴史語言研究所が「新版漢籍電子文獻 3.0 測試版」(<http://serv2.sinica.edu.tw:8081/HandyWeb>)として試験的に公開しているのだが、上記の<音義>に相当する陸徳明の《經典釋文》や<注>の部分が阮元の校勘記と一緒にたにされているなど、十三經注疏の論理構造が正しくマークアップされていないがために、折角の魅力的な新検索システムもあまり功を奏していない。このように最適なマークアップを模索することが、次なる段階として求められているのは自明であり、それは人文系研究者の参与なくしては成しえないことであることもまた明白である。

趣味と研究の境界は、それが自己完結に終わるか、他者との共有を指向しているかで分かれる。後者を志向する者は常に研究データの交換性や再利用性に配慮することが求められるだろう。その意味で XML や TeX はひとつの有力な選択肢であるといえる。後学の便宜のために我々の先人が多くの工具書を残してくれたことは我々のよく知るところであり、それらの工具書の恩恵にあずかっている今日の我々もまた、後学のために新たな選択肢を残していくことが今日の我々に課せられた責務ではないだろうか。

## TOPICS

### 2004年度の事業

- ・2004年9月6日～10日 2004年度東アジア人文情報学サマーセミナー「インターネット時代の人文学の技術（スキル）」を開催（於京都大学人間・環境学研究科棟）
- ・2004年11月18日～21日 イタリア国立東方学研究所・フランス国立極東学院との共催により「中国宗教文献研究国際シンポジウム」を開催（於京都大学百周年時計台記念館）
- ・2005年1月22日 中国国家図書館との共催により日中共同シンポジウム「漢字文献資料庫の新技术」を開催（於北京市湖北大厦）
- ・2005年2月13日 京都新聞社との共催により「オープンフォーラム漢字文化の今（第二回） 東アジアの人名・地名と漢字」を開催（於京都新聞文化ホール）

Chinese Characters  
and Culture



発行日 2005年2月28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/>

