

研究部会報告一抄録

情報知識学会 人文・社会科学系部会主催
第12回 歴史研究と電算機利用ワークショップ

人文・社会科学系部会企画担当 田良島 哲

主催：情報知識学会人文・社会科学系部会・学習院大学史学会

期日：1999年11月27日（土）

会場：学習院大学北1号館308号室

「歴史研究と電算機利用ワークショップ」は、歴史的な方法をとる研究分野におけるコンピュータ活用の経験を情報交換するために、情報知識学会人文社会系部会の企画として、継続して開催している。第12回は、神立孝一氏が企画の主担当となり、長谷川順二（学習院大学大学院）・百瀬康司（東京工業大学大学院）の両氏と田良島哲の3名が報告した。また、今回は初めて学習院史学会と共催し、報告者の選定・会の運営を共同して行った。

村松弘一氏をはじめ、学習院史学会のスタッフの方々には当日の会場の設営をはじめ多大なご協力をいただいた。記して感謝申し上げる。

衛星画像を使用した黄河下流域の河道変更
に関する考察

学習院大学大学院 長谷川順二

黄河は古来より幾度となく氾濫し、その度に流路を変化させてきた。今までに数多くの研究者が様々な方法で、その河道を確定させようと試みてきた。史書には都市や他の自然地形との相対的な比較によって河道の位置を間接的に表現しているため、現在に至るまで黄河故河道の正確な位置は確定されず、研究者によってまちまちとなっている。この不明瞭な状態を打破するために、史書研究とは別の資料、特に科学資料を導入しようと思う。

今回私が導入を試みた研究方法是、「リモートセンシング」と呼ばれる。これは中国語では「遥感」と意識されるように、「離れた場所から目的の場所を観察する」観察方法の総称であるが、現在では特に人工衛星によって撮影された画像を使用した研究を指すことが多い。「人工衛星で画像を撮影」といっても、通

常の航空写真のようにカメラを用いるのではなく、センサーによって撮影された画像である。センサーを用いて収集したデータを数値に応じてグレイデータに変換して、地形写真のように加工したのがいわゆる「衛星写真」であって、通常は白黒で表現されている。センサーによって得られるデータには可視光だけでなく、赤外線を感知することもできるので、適切な処理を施せば、通常の航空写真とは異なる結果を得ることも可能である。「画像処理」とは、端的に言えば「画像内にある同様の特徴を持つ範囲を、目に見えるように色分けする」方法である。処理を行う範囲に含まれる地形の特徴によって、様々な画像処理の方法が存在する。

今回行なった方法は、まず最初に目的の画像に画像処理を行なって、河川やそれに類する特徴を持つ地形を強調する。次に処理画像を観察して、河川と思われるラインを手作業で拾い上げる。最後に拾い上げたラインを現

在の地図と重ね、現在の都市や地形を重ね込む。ここまでが画像関連の作業である。

次に正史などの歴史資料との比較を行う。具体的には、先程処理した河川の近辺に所在する都市の記述と、処理画像との位置の相違を比較する。年代によって河道の走る位置が違うので、資料が十分揃えば年代別の河道を確定することも可能であろう。

「リモートセンシング」を歴史学へ応用した例としては、考古学分野において発掘範囲を事前に確定させたことが挙げられる。処理画像を使用することで、ボーリング調査を用いずに遺跡の所在を確定させることが可能となった。また現在残っている都市遺跡を上空から撮影して、当時の町並みや都市計画を推察することも試みられた。いずれも歴史資料を補完する方法として「リモートセンシング」データが活用されており、本研究も同様の研究方法になる。しかし以前の研究が主に乾燥地帯への適用に留まっていたのに対して、本研究では湿潤地帯の河川を研究対象としている。「リモートセンシング」の歴史学への活用の幅を広げることができれば幸いである。

Windowsにおけるマルチリンガル環境

東京工業大学大学院 百瀬康司

1990年代後半は、まさに「インターネットの時代」であった。「インターネット」や「電子メール」が一般家庭に爆発的に普及したことは、同時に、その端末としてのパーソナル・コンピュータ（IBM PC/AT 互換機）と、そのOS（Operating System）であるMicrosoft Windows98/95が普及したことを意味する。

今でも、コンピュータは「計算機」と呼ばれる。たしかに、黎明期におけるコンピュータは、ただの「計算機」であった。だが、現在のコンピュータは、それほど単純な機械ではない。ひと昔前のスーパーコンピュータ以上の性能を持ち、多種多様な処理を行うこと

が可能な、いわば「多目的ツール」へと進化したのである。とはいえ、原型が「計算機」であったことの「弊害」は、いまだに残っている。その端的な例が、文字コードの問題である。コンピュータは、「計算機」として使うことを前提に設計されていたがゆえに、そもそも複数の言語（文字）を処理すること自体、まったく想定されていなかった。そのため、様々な国の企業や規格団体が、自分たちの都合でそれを拡張し、多種多様な文字コードを設定してしまった。このようにして規格化された文字コードの背景には、その国の文化的側面が強く反映されており、一朝一夕に解決できる問題ではない。今後、歴史研究に限らず、あらゆる分野の学問がコンピュータとインターネットを利用した研究方法を導入していくことになるだろう。その際、文字コードの問題は非常に重要な要素となるはずである。とくに、歴史研究の場合は、その性質上、マルチリンガル環境の実現が「必要不可欠」である。決して「ひとごと」ではない。したがって、文字コードとマルチリンガル環境について現時点における問題点を整理し、今後いかなる改善が必要なのかを考えることは、歴史研究に関わる人間にとって義務である、といっても過言ではないだろう。

本発表であつかう問題は、以下の5つである。

- (1) 「文字コード」とは何か？
- (2) 「文字化け」の仕組み
- (3) Unicodeはマルチリンガルへの最短距離か？
- (4) Windowsのマルチリンガル環境
- (5) まとめと今後の展望

なお、発表に際しては、専門的な説明は極力省略し、実例をデモンストレーションすることで、「感覚的」に理解できるような形にする予定である。

といった実践的な課題について、若干の提起を行う。

古典籍のテキスト空間と論理構造—電子化の前提としての検討—

文化庁文化財保護部 田良島 哲

現在、コンピュータとソフトウェアの能力の進歩はいちじるしいものがあるから、印刷で用いられている組版のスタイルを、形の上でディスプレイ上に表示することは可能であろう。しかし、古典籍テキストの電子化で、本質的な課題は「印刷されたとおりに表示すること」ではない。テキストの電子化の上で重要なのは、さまざまなスタイルで配置された文字列群の平面空間内における意味あいと、文字列どうしの関係である。いいかえれば、空間的な文字列群の配置を、論理的な構造に変換することが重要なのである。この点は、意外にも、文学研究者や歴史研究者にも十分理解されていないように思う。

本報告では、まず、漢字を主体としたテキストの空間的な配置が、どのように進化してきたかという点を、検討する。中国において、金石文にはじまり紙に定着するテキストの構造は、その歴史の中で、割書・傍書・頭書など付加情報を平面空間内に記述する方法が開発され、11世紀には、これらの方法を集大成した宋版という東アジア世界に大きな影響を与える組版形式が出現し、一つの規範となる。このような規範は7世紀以降、日本に順次移植され、文書や典籍の表現に用いられるようになるが、訓点のように文字の一つ一つをも空間として使う独自の手法も発明されている。このような歴史的な前提を念頭におきながら、種々の記述形式について、その本文に対する論理的な意味づけを確認してゆく。

その上で、古典籍テキストをXMLのようなマークアップ言語の形で構造化するためにどのような点に留意すべきか、また実際に電子的に編集を行うために必要なツールに求められる条件としては、どのような点があるか、

研究部会報告一抄録

情報知識学会 人文・社会科学系部会主催
第13回 歴史研究と電算機利用ワークショップ

人文・社会科学系部会企画担当 田良島 哲

期日: 2000年5月13日(土)

会場: 法政大学ボアソナードタワー9階 サイエンスルーム

「歴史研究と電算機利用ワークショップ」は、歴史的な方法をとる研究分野におけるコンピュータ活用の経験を情報交換するために、情報知識学会人文社会系部会の企画として、継続して開催している。今回は、独特の記述システムを持つ古典籍や古文書を電子化する上での新しい試みについて、谷本玲大（茨城大学）、鈴木卓治（国立歴史民俗博物館）、小口雅史（法政大学）、小畑真帆（株）フラックストーン）の4氏から3本の報告をいただいた。いずれもXML、PDF、LaTeXなどを駆使した実践的な内容であると同時に、今後のテキストの電子化に関する理論的な問題も示唆するもので、報告終了後、熱心な討論が行われた。

なお、会場の準備と当日の運営については、小口雅史氏に多大なご協力をいただいた。記して感謝申し上げる。

XMLによるタグ付けを用いた古典籍の字彙DBについて

一伝本の性質を把握する補助手段として一

茨城大学 谷本玲大

本発表では、古典文学研究を主とする人文系研究の立場から、XML文書+今昔文字鏡TrueTypeフォント+JavaScript+html出力という組み合わせで作成可能なXMLによる変体仮名字母データベースの紹介を行うものである。

今回発表するXMLによる字母データベースは、コンテンツ内容として国宝、尊経閣文庫本『土左日記』を採用したが、コンテンツ素材が何であれ、同様の処理が可能のようにDTDを設計してある。従って、タグ付けの煩雑さを厭わなければ、比較的容易に変体仮名の実際の字形表示と、それを包摂/区別した検索とを行うことができるものである。

文学研究におけるコンピュータ利用について、本文研究の第一歩である文献学的研究に

応用しようとする試みは、既に統計学的な見地から導入されてきた（行動計量学シリーズ6『真贋の科学—計量文献学入門—』村上征勝著 朝倉書店 1994）。

しかし、国文学に於いて、古典本文は一般に現代では用いられることのない様々な字形を複雑に使用して表記するので、いわゆるJIS漢字以外の多種多様な文字を区分して操作できるようにする必要がある。

だが、DB構築に際し、使用文字をJIS漢字に限定し、IDを付加するのみでは視認性に欠け、DBそのものが着眼・発想を支援する役割を担うことはできない場合が多いと思われる。

故に、コンピュータを国文学研究の道具として使えるようにするためには、様々な字形を表現し、かつ、ある程度は原本の改行や字の割り付け組版を含む表示をする機能が用意される事が本来的には望ましい。

更に、様々な変体仮名字形を詳細に弁別するのみでなく、それらの変体仮名を例えば「お」

という読み方をするもののみ一括で抽出したり、あるいはその「お」という読み方をする文字の中で「於」という漢字がその変体仮名の元となる字母であるもの…など、様々なレイヤーを持たせて検索することができるようになっている必要がある。従来は、これらを特定の読み方をする仮名のレベルや、特定の字母のレベルに丸め操作を行って研究されてきたわけだが、SGML や XML を用いれば、視認性を維持しつつ丸め操作／詳細な弁別操作を様々に行えるようになると思われる。

そこで、今回発表者が紹介する方法は、それほど特殊なソフトウェアを必要とせず、パーソナルコンピュータ環境で簡便に行えるという視点から、XML を採用してコンテンツ構築を行った。

本発表で実演するコンテンツは、XML というマークアップ言語本来が持たされているマークアップ言語としての用途や設計思想からは乖離した利用法ではあろうが、XML の利用によって文献学的研究（あるいは、国語学的研究にも転用可能）に着眼・発想を支援するためのDBとして、かなりの要求を満たし得ると思われる。XML によって得られる恩恵と共に、今後の展望、またあわせて実作業に於いて遭遇した問題点などについても提示してみたい。

なお、本発表の元になったのは 1999 年 9 月 16 日から 10 月 8 日にかけて東京外国語大学アジア・アフリカ言語文化研究所短期共同研究員として、同研究所教授芝野耕司氏の指導のもと行った「XML によるタグ付けを用いた『新撰萬葉集』伝本の数量的比較研究」と題する共同研究である。

LaTeX による古文書組版の一経験

国立歴史民俗博物館情報資料研究部

鈴木 卓治

筆者は、博物館における情報システムに関する研究を業務とする情報系研究者である。

古文書はまったくの素人であるが、1998 年 10 月 20 日から 12 月 6 日にかけて開催された本館企画展示「収集家 100 年の軌跡— 水木コレクションのすべて—」の出し物のひとつとして、水木家資料中世文書 154 点のマイクロ写真とその釈文を並べて表示するコンテンツ (http://www.rekihaku.ac.jp/gallery/mizuki/index_mizuki.html にて公開中) を開発した際、LaTeX を利用した古文書釈文の組版を経験した。本発表は、このときの経験をベースに、pLaTeX2e+ 金水マクロ + 今昔文字鏡 TrueType フォント + bookworm マクロ + Acrobat という組み合わせで、海外の環境でも TrueType 品質で表示印刷できる釈文データの作成を実演するものである。

歴史学におけるコンピュータ利用を考えると、その第一の基礎をなす対象は、かたや古文書であり、かたや屏風・絵巻・浮世絵などの図像情報であろう。古文書は一般に複雑な表記規則をもち、いわゆる JIS 漢字以外の多種多様な文字を含み、組版の位置関係や加筆・訂正・削除その他の編集痕といった重要な付帯情報を含んでいることなど、情報工学の観点からみても、その扱いには研究が必要であり興味深い対象である。

コンピュータを古文書研究の道具として使えるようにするためには、組版を含む表示・印刷の機能、入力・訂正・切り貼りなどの編集（しばしば組版指示作業と同時進行）の機能、および検索の機能が用意されなければならない。現時点での最適解は、Microsoft Word などのプログラミング可能なワードプロセッサシステムをベースに、各機能を満たす道具を開発することであろう。

今回筆者が紹介する方法は、対象を組版に特化していることと、組版システム自体が人文系一般になじみのうすいものであることから、広くおすすめてできる方法とはいえない。しかしながら、現在望める最高品質に近い組版データが比較的安価に作成できることや、

指示したい組版情報が確定している場合、組版システムを利用したほうが、WYSIWYGの原則に基づくワードプロセッサよりもかえって少ない手間で組版の指示が行なえることなど、押さえておくべきいくつかの重要な原則を含んでいるように思われる。

日本古代文書の電子化をめぐる諸問題 ーデジタル古文書集作成の実体験をふまえてー

法政大学第一教養部 小口 雅史
(株)フラックス・トーン 小畑 真帆

本報告は『デジタル古文書集日本古代土地経営関係史料集成』（同成社刊）の製作経験をふまえて、現状の問題点と今後への展望を考えるものである。

日本古代の代表的古文書群である正倉院文書は、二次利用されたために残ったという特殊な経緯を有する。また江戸時代後期の整理の過程で、二次利用面が、不用意に截断され一次利用面への回復が試みられたため、現状は、一次利用面も二次利用面も、ともに当時の生の状態ではなくなっている場合が多い。こうした接続の問題と、さらに数次わたって書き込みがなされているという重層性によって、この一大古文書群を詳細に分析するためには、紙媒体の翻刻よりは電子媒体による翻刻こそが、研究上、大きな効果を持つ。

多数のJ I S外漢字を有する古文書群であるが、その問題は、今昔文字鏡によってほぼ解決のめどがついている。縦書き表示はPDFで処理すれば問題ない。残された課題は、ダイナミックな検索プログラムの開発である。

今回は以下のようなステップを踏んだ。

○古文書のデータ処理

- ・原典図版を座標系にマップする
- ・単語区分の識別
- ・検索アルゴリズムの重層化

この過程で、日本古代文書の電子化をめぐ

る諸問題がいくつかうかびあがった。必要なのは(既存のシステムの拡張も含め)、古文書ページ記述システムなのだろうか。主要な開発環境すべてに標準プログラムソースをライブラリとして用意すべきなのか。標準辞書として、完全な辞書には割注や文字サイズに関する情報も必要で、さらに、その影響力の調整機能は実現してみたい。ではオブジェクトデータベースとして理想的な結論はどこにあるのか。今回は実践例をふまえての諸問題の提示に留まらざるを得ないが、当日の議論の展開に期待するところも大きい。