

第6回公開シンポジウム

# 人文科学とデータベース

「データ」を読む・観る・解く

2000年12月20日 静岡大学情報学部

第6回 公開シンポジウム実行委員会 主催  
人文系データベース協議会 後援

## 目次

□ 特別講演	
「ええじゃないか」研究を振り返って 田村貞雄（静岡大学）	1
□ 一般講演	
文化資源情報における Topic Map の適用と評価 －「源氏物語」画像DBを使ったモデル作成の試み－ 長瀬真理（静岡大学）	3
文字冗長度による日本文学分析 福田宏（静岡県立大学）、山下泰弘（科学技術振興事業団）、勝矢光昭（静岡県立大学）	11
広領域分野資料の横断的アーカイブ論に関する基礎研究 八重樫純樹（静岡大学）	17
ジャイナ教聖典のデータベース－特殊フォントで表現されたデータのPDF－ 逢坂雄美（仙台電波工業高等専門学校）	27
パスカルデータベースシステム（1） 白石修二（福岡大学）	35
吾妻鏡データベースの構築 安道百合子（国文学研究資料館）	43
全国遺跡データベースの構築 森本晋（奈良国立文化財研究所）	53
地理情報システムを用いた城下町の復原的研究 －彦根城下善利組足軽屋敷地区を中心として－ 生方美菜子、濱崎一志（滋賀県立大学）	63
考古学データベースにおける検索エンジンの研究 三浦宙明、小沢一雅（大阪電気通信大学）	73
□ 公開シンポジウム「人文科学とデータベース」既刊一覧	81

# 「ええじゃないか」研究を振り返って

Looking back upon my study on "*Eejanaika*"  
the nation-wide fanatic boom among the people  
during the last few months of the Edo era

田村貞雄

TAMURA, SADAŌ

静岡大学情報学部

静岡県浜松市城北3-5-1

Faculty of Information of Shizuoka University

3-5-1Jouhoku, Hamamatsu city, Shizuoka prefecture

明治維新直前の1867年(慶応3) 東海道を中心に江戸近郊から四国・山陽道にかけて起こった民衆の狂乱状態を「ええじゃないか」という。「ええじゃないか」では、伊勢神宮をはじめ秋葉三尺坊大権現・津島牛頭天王・豊川稲荷など諸神諸仏の御札が降臨したが、これをきっかけに民衆は街頭に出て狂喜乱舞した。

わたくしは1980年代中葉から「ええじゃないか」研究をし、最初のお札降りと思われる豊橋市の史料を用いて、『ええじゃないか始まる』(1987年)を書いた。この年には、「ええじゃないか」120年を記念して、「ええじゃないか東海シンポジウム」を豊橋市で開いた。

1995年度からわたくしは旧教養部から情報学部に移ったが、工学部から来られた方々と科学研究費の重点領域「人文科学とコンピュータ」に応募して共同研究をした、これはわたしにとっては刺激的な、楽しい研究環境であった。

In 1967 (Keio Year 3), the previous year of the Meiji Restoration, the nation-wide fanatic boom among the people called "*Eejanaika*" came out through Tokaido area including Edo (old name of Tokyo) and its suburb towards Sanyodo as well as Shikoku area.

During this "*Eejanaika*" movement, a large number of charms of famous shrines, including Ise Shrine, Akiba Sanjakubo Daigongen, Tsushima Gozu-tenno as well as Toyokawa Inari, dropped out of the skies, which triggered to make people madly dance

on the streets for weeks. I have been studying "Eejanaika" since the middle of '80s and wrote "Eejanaika Bigins"(1987), with the historical materials from Toyohashi where it was considered the first place of the charm drops. At the same year, celebrating 120th years anniversary of the Eejanaika movement, I coordinated "The Eejanaika Symposium in Tokai" at Toyohashi city.

Since 1995 until 1998, I co-worked in the joint project, the Computers and Humanities, with Dr. NAKATANI, Dr. ITOH, Dr. KONISHI, Dr. AKAISHI and Dr. ABE, all of whom came from the Faculty of Engineering, after I moved to the Faculty of Information. It gave me a great deal of exciting and inspiring experience with very fruitful results.

【キー・ワード】 ええじゃないか、御鋳百年祭、御蔭参り、秋葉信仰

【Key-word】 Eejanaika, Centennial anniversary of god of hoe, Mass visiting to Ise Shrine for thanks, Akiba worship

#### 「ええじゃないか」発端についての諸説

わたくしは、1980年代半ばから明治維新直前に起こった民衆行動である「ええじゃないか」という現象を研究してきた。

「ええじゃないか」のきっかけをなした最初のお札降りについては、戦前には井野辺茂雄「神符の降下に就いて」(1916年)の「七八月の交」名古屋発生説、『維新史料綱要』(1940年)の8月下旬名古屋地方説などがあった。『岡崎市史』第8巻(1930年)のみは典拠をあげないまま、三河発生説をとっていた。なお田村栄太郎「慶応三年”ええじゃないか”」(1960年)のみは8月横浜説である。

藤谷俊雄氏の『「おかげまいり」と「ええじゃないか」』(岩波新書 1967年)では8月中旬より尾州、三州、遠州の三国で発生したとされた。西垣晴次氏の『ええじゃないか』(1973年)は、『磐田市誌』下巻(1956年)の記す8月15日の遠江国見付宿の事例を初例とされた。その後『豊川市史』(1973年)が三河国御油宿の8月4日の史料を載せ、編纂にあたった大久保友治氏の「「ええじゃないか」の発生について」(1974年)で紹介された。7月22日以前とする豊橋市羽田八幡宮の史料は、岸野俊彦氏・伊藤忠士氏からの史料提供により佐々木潤之介氏の「幕末の社会情勢と世直し」(1977年)で紹介された。高木俊輔氏の『ええじゃないか』(1979年)もこの説である。

加藤善夫氏「県東部のええじゃないか」(1985年)は7月18日夜の吉田宿(現豊橋市中心街)の事例を紹介された。

#### 豊橋市牟呂八幡宮「留記」

7月14日発端を示す牟呂八幡宮の史料「留記」は、愛知大学の歌川進氏によって発見され、橘敏夫氏「御札降り発生地域の関連史料」(1984年)ではじめて指摘され、渡辺和敏氏により『新居町史』第8巻(1986年)で全文紹介された。これは牟呂八幡宮近傍で御札を見つかったが、その出所を疑った2人の人物の家族が急死するという異常な事件を記録している。人々は御鋳百年祭を催促する神の仕業と恐れおののき、御札降臨の臨時祭礼をはじめた。その渦中で吉田宿(豊橋市中心部)で、鳥がお札をくわえて来たという噂が発生し、「ええじゃないか」に発展したのである。

わたくしは御鋳百年祭の先行に注目していたので、牟呂八幡宮の史料こそ「ええじゃないか」の発端を示すものと考え、渡辺氏の了解を得て『ええじゃないか始まる』(1987年)を書いた。

#### 人文科学とコンピュータ

1995年静岡大学の情報学部創立にわたしは参加したが、その折り工学部から参加された中谷広正、伊東幸宏、小西達裕、赤石美奈、阿部圭一の諸氏らに誘われて、1995年度に始まる文部省科学研究費重点領域「人文科学とコンピュータ」に参加した。幸い1998年度まで4年間にわたり共同研究をつづけることができたが、わたしにとっては工科系の方々との共同研究は初めてであり、カルチャー・ショックも経験しつつ、多くのことを学び、実に楽しい思い出となった。

## 文化資源情報における Topic Map の応用と評価

---「源氏物語」画像 DB を使ったモデル作成の試み---

### Application and Evaluation of the Topic Map for the Cultural Resource Data ---Experiment for the Graphic Database of the Tale of Genji---

長瀬真理  
Mari NAGASE

静岡大学情報学部、浜松市  
Shizuoka University, Hamamatsu 432-8011

あらまし：本研究の目的は、現在、文書館・博物館・図書館・研究機関・大学・企業などで構築されている様々な文化資源情報を、情報の構築サイドではなく、個別の利用者の視点を反映して効果的に利用・評価するための新しい機構を調査・研究することにある。具体的には、ISO の標準として採用された Topic Map (ISO-IEC13250:2000) を小規模なデータベースに対し実験的に適用を開始し、逐次実験規模を拡大することによって、Topic Map の効果が期待されている XML 型大規模データベースへの適用可能性を検証する。今回は小規模データとして、既に筆者が蓄積した「源氏物語」日英仏ハイパーテキストデータの中から、227 枚の画像データを使用する。画像データは、いくつかの主題に大別することが出来るため、トピックの地図作りのインデックス付加が容易であり、今後ユーザーがこの画像を使うためのモデルを提供することが可能であろう。ただし、本研究は始まったばかりであり、具体的な成果にまでには至っていない。本論文では研究方法の見直し、及びサンプルの DTD 作成までを解説し、読者の批判や助言を仰ぎたい。将来的には、Topic Map を本来の目的である大規模なデータにも適用し、その有効性や問題点も検討するつもりである。更に、それらの成果をデータベース構築にフィードバックし、検討された理論が、これまで蓄積されてきた多様な文化資源情報のマネジメントや評価作業に対して有効であることが証明され、これまで以上にユーザーに使いやすい新たなデータ構築をサポートすることが出来ればと願っている。

**Summary:** The purpose of this research is to investigate effective methods to use large scale data compiled in various archives, museums and libraries, not from developer's sides but from the standpoint of users. For one example of methods, I would like to deal with the Topic Map (ISO-IEC13250:2000) registered recently as a standard of the ISO and do some researches on its possibility and its validity for large data in the XML format. For the case study, I would apply this method to my multilingual and multi-media data of the Tale of Genji developed in the last ten years and would like to create a model for various usages. Further, if it is possible, I will feed back the results to production sides to promote effective usage of large scale data.

**キーワード:** トピック・マップ、ユーザー主体、XML、「源氏物語」

**Keyword:** Topic Map, user-oriented, XML, Tale of Genji

#### 1 序論

伝統的な紙の本には、必ずインデックスがついていた。しかし、電子化されたドキュメントには、利用者側の情報利用に便宜を与えるこの種の配慮がなされていないし、またインデックスを記述するルールも規約も存在していない。確かに現在の高速な検索ソフトウェア（検索エンジン）を使えば、大量の電子文書から自動的に、いわゆる指定された文字列を含むコンテンツを、“正確”にアルファベット順に取り出すことは可能

である。あるいは検索を目的として、yahoo! などを使って、ある特定の項目について膨大な数に及ぶ検索結果集合を導くこともできるであろう。しかし、取り出されたインデックスも検索結果も実際に使うとなると、あまりにも大量であり、無意味な文字列の連鎖にすぎず、かえって我々を途方にくれさせるだけである。

今日、我々は「知や知識の蓄積と共有」を目指して様々なドキュメントをコンピュータに入力・蓄積しているが、それらはいずれも、制作者側の視点から編成されるデータベースであり、個別の利用者からみれば、宝の山ではなく、ゴミの山である。いみじくも Topic Map の提案者の一人である Steve Pepper が「インデックスのない本は、地図 (Map) のない国のような」と言っているように、「インデックスのない電子ドキュメント」は、何ら我々の知識構築には役立たない。こうした状況をユーザーサイドから解決すべく提起されたのが ISO (ISO-IEC13250) の標準として採用された Topic Map なのである。

## 2 Topic Map の基本概念

「トピック地図: Topic Map」とは、膨大な情報資源と情報操作の間に「橋」をかけようとする試みである。

それは、抽象的なデータ同士の意味をつなぎ、新たな知識を構築し、更にそれを新しい情報資源の地図として描き出す手法なのである。

Topic Map のモデルは①トピック、②トピックのデータ、及び③トピックの連合(association)の三要素から構成される。

具体的な解説の為に、例として、B.LeGrand と M.Soto の実験を紹介する。

まず Topic を定義しておこう。Topic とは、電子ドキュメントにおける検索項目、あるいはデータ項目で(本や印刷物のインデックスに相当する)意味連関の役割を担った情報である。

たとえば、会議を開催するための資料情報を考え、そのインデックスを作ってみよう。そして「この会議がヨーロッパのパリで開かれる XML 会議である」と想定してみよう。

すると、個々の”役割(role)”を担った言葉、即ち、論文、ビデオ、表、会議案内、等々で構成されたインデックスが出来あがる。この場合、個々の言語(インデックス)を、更にメタレベルで、異なった”タイプ(type)”のトピック、例えば、会議、都市、国名等など分けるならば、我々は、これらの個々のインデックス

を結合して新しい情報を作ることができる。例えば「XML ヨーロッパ会議がパリで開催される(take place)。パリはフランスに存在する。(is in)」等々である。即ち、異なったタイプのインデックス同士をメタレベルの意味機能で結びつけることによって、二つの結合、「開催される」と「に存在する」を作り出すことができるのである。

このように、Topic Map は資料情報のメタレベルに作成される地図で、トピックと情報ソースの二つの領域を分離させ、更にそれらのトピックとトピックの結合を、基盤となった情報の資料(データが生じるレベル)を超えたところに、新しい意味連関のネットワークを構築する知識創造の新しい方法なのである。

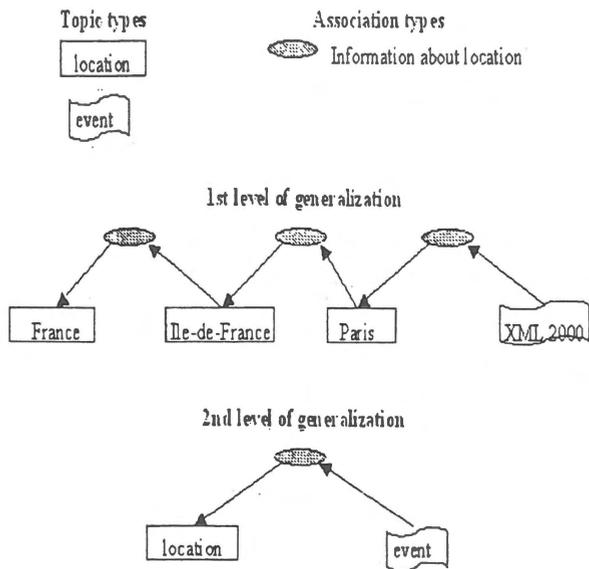
同時に、“地図”という言葉が連想するように、“移動可能な意味ネットワーク”の可視化(visualization)機能も兼ね備える。そこで、次に可視化機能の側面を紹介してみよう。

## 3. 可視化技術としての Topic Map

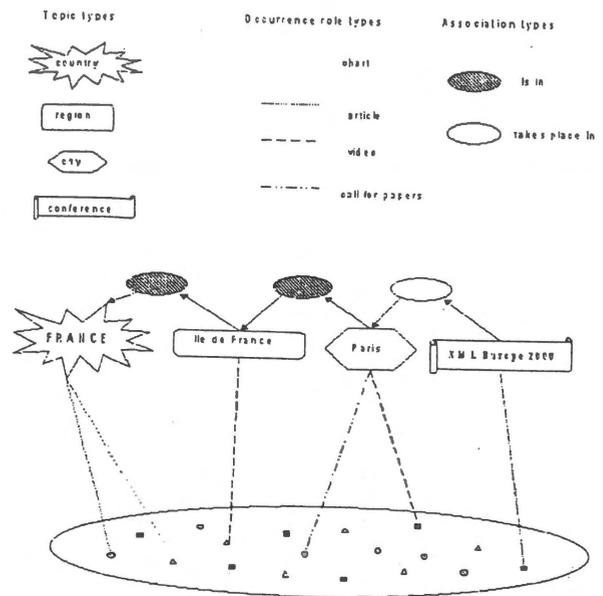
大規模な情報を可視化するために、LeGrand と Soto は開発された MAGE ソフトを利用した。このツールは、もともとは情報に色づけや、三次元表示、動画などの手法を付加して、情報を可視化するために開発された。

これを Topic Map の可視化に利用する。Topic Map は大量のトピックの結合を可能にするが、そのすべてを表示することは困難なため、まずは、関連する情報を選択しなければならぬ。すなわち、関連情報の選択には選別技術が不可欠となる。この選別技術は、XML の属性である、「名前」と「タイプ」によって情報を選別することによって実現される。現在、ネットワーク上に広く用いられている HTML が、情報の表現体裁の制御に限定しているのに比較し、XML は、構造などのタグ情報も提供することから、ネットワーク上での情報共有の新しい機能として注目されつつある。結果出力は次ページを参照願いたい。上段が意味レベルと基本概念、下段が、選択されたインデックスの連合例の可視化である。

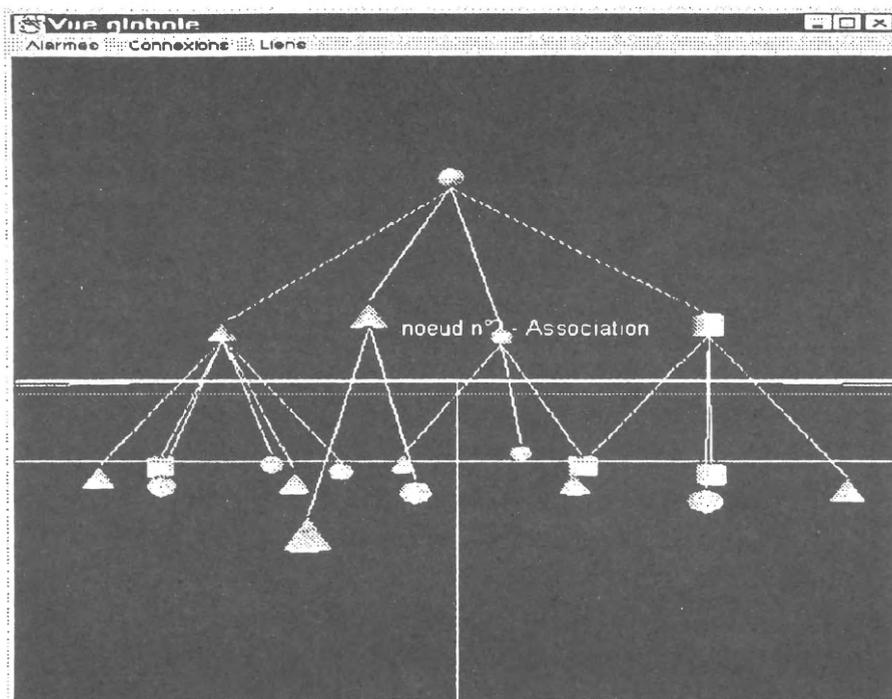
将来は、ユーザーの目的に応じた存在論的な秩序に適合するものへと選別技術が改良されることが期待されている。存在論的知識とは、各分野の基本語からなる語彙で表記される知識を指す。一般に、そこでの概念表記や、当然のことながら、自然言語の記述を基礎付けるものであり、これは、適切な KIF(知識交換形式 knowledge Interchange Format)を指示したものでなけ



Topic Map の意味レベル



Topic Map の基本概念



Topic Map の連合の可視化

ればならない。語彙は新しい概念の挿入がオープン(追記的に可能)である辞書構造と概念構造を持った形で蓄積されるようになる。

こうして、これらの可視化技術や XML 等の基盤技術をもとに、ユーザーは、自己のの関心に従った地図(Topic Map)を作成する。今後、トピックと結合は政治地図、経済地図、あるいは文化地図という風に拡大されると推測される。

そこで、もう一つ、Steve Pepper が「オペラの本」電子ドキュメントを題材に行った研究を紹介しよう。

彼は、Topic Map の推進の代表者であり、ヨーロッパを根拠地に活躍している。ヨーロッパは、現在 Topic Map の普及に非常に熱心である。その理由は、もちろん、当地でも大量データの効率的利用が大きな関心事であるからに他ならなし。しかし Topic Map の拡大の裏には、アメリカに遅れをとったヨーロッパの威信のようなものがある。現在、ヨーロッパは、人権問題やデジタル・デバイド(情報格差)に非常に過敏になっている。多くの情報を発信するばかりでなく、受け手が、より適切に、より公平に情報にアクセスすることへの関心が高いのである。また、「知識は力なり」という伝統もあり、情報を、有効な知識の構築に利用しなければ意味がない、という基本的態度を持つ。そこで、一人一人の個人の立場から、大量データをどのように利用し、その知的活動を支援すべきかという問題に関心が高まり、その答えの一つが Topic Map なのである。

本題の Pepper の研究に戻ると、一般に、オペラの本であれば、どんなものであれ、原則的に、作曲家、作品、劇場など、オペラに関連する”トピック”を含んでいるはずである。

図1 トピック

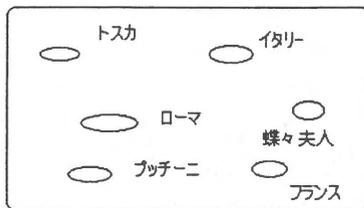


図1は、様々なトピックが、電子ドキュメントの中で並列的に存在する状況を図示したものである。本の裏

のアルファベット順のインデックス同様、このレベルではこれらトピックの役割(role)に差はなく、個々の単語の羅列状態と同じである。

このままでは、個々の語が、同等にテキストに埋没する、いわばプレーン・テキストと同じ状況であり、効果的な意味構築はできない。

そこで、実際の印刷物のインデックス作りでも、言葉の差異化の必要から様々な工夫がなされる。例えば、個々の言葉が特別な意味を持つことを示すために、タイプの違う文字列を作り、それを手がかりに、読者が目的の個所を見つけやすくするようにしている。例えば、人名だけはイタリック表記の文字にするとか、キーワードはボールド・タイプにするといった具合である。あるいはジャンル分けをして、曲名だけをリストアップすることもあるだろう。

図2 トピック・タイプ

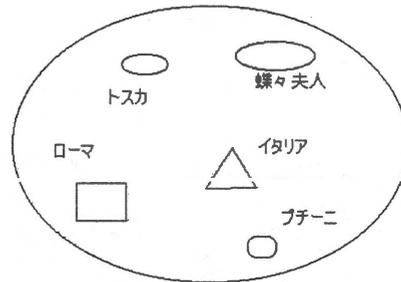


図2は、個々のトピックが、意味レベルで、異なったタイプのトピックであることを図示したものである。

こうすることによって、ユーザーは、個々のトピックの属性を区別することができるようになる。そうすることによって、作曲家と作品のようにトピック同士の結合を行い、有意義な知識の構築を行うこと容易になる。

また、マルチ・インデックス、例えば、名前ごとの、場所ごとの、主題別のインデックスも作成される。

更に、トピックのタイプは、作品名としての「トスカ」と、人物名の「トスカ」も区別できるものにする。あるいは、文中に出現する数字で表記されたデータについては、その数字が、ページ数なのか、脚注の番号を意味しているかなどの区別も可能になる。

トピック的なインデックスとして、ここでも重要なのは、①トピック(名前など一度以上出現する)、②ト

ピック同士の連合、③個々のデータである。

図3トピックの連合

トピックタイプ	出現箇所	連結
△ 国	図	存在
□ 市	論文	開催
○ 人名	ビデオ	
□ 曲名	論文集	

具体的に、先のデータの Topic と連合を行い、データとトピックの間に効果的な橋を架けてみよう。

- \* トスカはプッチーニにより作曲された。
- \* トスカはローマで演じられた。
- \* プッチーニはルッカで生まれた。
- \* ルッカはイタリアにある
- \* プッチーニはヴェルディの影響を受けた。

といった連合が作り出される。

このようなトピックの結びつきによって表現される Topic Map は、情報資源に結びついた利用者の視点にたった情報資産の再表現を可能にする。すなわち、ここに挙げた情報はそれぞれ、ユーザーが作り出した新しい情報なのである。

更に、情報資源と Topic Map が互いに独立していることが重要とされる。これらが独立していることによって、この地図は他のユーザーにも適用されるし、また別の情報資源にも援用可能となる。

このような考え方は、ユーザーサイドの利用を前提した XML の構想にも共通するものであり、現在、XML と Topic Map の融合実験が精力的に行われている。

次に、ここで筆者の行ったデータベース構築を紹介し、Topic Map への応用の可能性について考察を進めたい。

#### 4.「源氏物語」データベース

1996年から98まで、フランスのシャルル・ド・ゴール＝リール第三大学およびフランス科学技術庁、The Japan Foundationなどの協力で、フランス語版の画像データベースの開発に従事した。

画像(227枚)は、1650年、江戸時代随一の蔭絵師、山本春正の描いた木版画を採用した。実際は、貴重本

刊行会出版の「絵本 源氏物語」をスキャナーで入力した。

画像の原本は鶴見大学図書館と東京大学国文学研究室が所蔵しているが、これらは、大部分紙も傷み、画像もかすれており、公開される機会も少ない。そのため、デジタルカメラのやスキャナーでの入力は不可能であった。

一方、貴重本刊行会で出版された画像は、修正が為されておりスキャナーでの入力がたえるものであった。

更に都合の良いことに、個々の画像には三人の「源氏物語」の専門家の手による5～6行の解説がついており、そこを読むだけでストーリー展開を追うことができるようになっている。

#### 4.1 画像のテーマ

227枚の画像のテーマを調べてみると、大体、以下の6つのカテゴリーに分けることができる。

- 1) 和歌の場面(手紙、手習い、絵画を含む) 38枚
- 2) 音楽の合奏 9枚
- 3) 宴会、遊び(狩り、舟遊び、雛人形、囲碁、双六、舞い) 35枚
- 4) 恋(秘められた恋、覗き) 41枚
- 5) 移動(車、馬、船) 16枚
- 6) その他(自然、語り、お化け、死、祈り、勤行、宗教行事、出産、見舞い、散歩、山歩き、夢、教育)

上記の分類の内 1) 和歌、2) 演奏、3) 宴会、4) 恋の場面みを合計すると、123枚で、全体の54.4%を占める。しかしながら、複数のカテゴリーに属する画像も多く、必ずしもこの通りの数にはならない。

カテゴリーのハイライト・シーンを、いくつか紹介する。

#### A) 音楽(数字は帖番号と画像通し番号)

音楽がカテゴリーの一つとして選ばれたのは、数の多さによるが、それ以上に、平安文化に占める音楽の役割が非常に大きいからに他ならない。



上記の場面は、有名な「雨夜の品定め」で知られている第2帖の一場で、左馬頭が愛人の家を訪ねた折り、庭先で笛を吹きその到着を知らせると、中から相手がそれに琴の伴奏で答える場面である。(2-8)

楽器が使われるシーンも多様で、宴会やコンサートの場面はもちろん、逢い引き、心の慰め、また踊りの伴奏にも主要な役割を担う。

和琴、唐琴、笛、太鼓、琵琶、笙、太鼓など様々な楽器が登場する。

### B)恋

当然のことながら恋の場面は、覗きの場面も入れると41枚もあり、どれを選ぶか選択に困る。これは4帖から取られた画像で(4-18)、光源氏と年上の愛人である六条御息所と霧深い朝の別れのシーンで、女主人は、寝所から腫れぼったい目で源氏を見やっている。

源氏は見送りに出た女房の中将があまりに美しいので、御息所に隠れて、廊下の端の高覧に中将を座らせ、寸暇を惜しんで戯れる。同時に、彼は庭に少女をおろし、愛人へのプレゼント用に美しく咲き誇る萩や朝顔を摘ませる。

浮気っぽい源氏の有様を見事に伝える1枚であろう。



### C)宴会

これは右大臣家での藤の花の宴(32-124)の場面である。紅葉や、桜の花見、梅の花見以外に、藤の花見の場面が2回にわたって描かれている。藤の花の色の濃いものは紫色で、蒔絵師は紫式部にふさわしい色と思ったのであろうか。



#### D) 遊



これは、空蟬と軒端の萩が囲碁に興じる場面(3-14)である。全体で囲碁の場面は3回登場する。2回は遊んでいるところ、もう一つは紫の上が碁盤の上に立って、源氏に髪を切りそろえてもらう場面である。物語全体では、囲碁を打つ場面は沢山あり、姫君達が、自慢の木を賭けて勝負をするエピソードなど様々である。

囲碁以外に、狩り、舟遊び、雛人形、双六、舞いなど遊びの場面は35枚もありカテゴリー第2位を占める。

このように、画像は、テキストだけでは分かりにくい平安時代の生活について、貴重な場面を描き出し、様々な情報を提供する。面白いのは、上記の囲碁の場面も含め、音楽の演奏の場面や、恋の場面で、他人から覗かれている画像が多いことである。こうした版画から、プライバシーが無い世界の有り様が容易に読者に明らかになる。ついでに、妻戸、格子、屏風、几帳、御簾などの仕切りで囲われた世界は、当時の建築や生活様式の理解にも大きな助けとなる。

ところで、こうしたカテゴリー分けは、翻って見れば、実は個人的なものに過ぎない。多数の人々が同意する解釈もあるだろうが、山本春正の画像データを目にしたユーザーはもっと多様な関心を抱くであろうし、新たな発見をすることであろう。だとすれば、データ

の製作者や提供者は、そうしたユーザーサイドにたつて、ユーザーがデータを個人化(パーソナライズ)出来るように、データを再構築することが必要となる。

製作者が出来あがったデータをただ提供するだけでなく、より多くの人々に使いやすいデータベースとして電子ドキュメントを開放することが求められる。

その方法の一つが、今回提起したTopic Mapへの応用なのである。

#### 4.2 画像のテーマからトピックへ

画像データを Topic Map の手法で解析できるようにするための以下のような作業が必要となる。

- 1) 画像データを分類し、トピックと関連項目を洗い出す。
- 2) Topic Map を XML 形式で作成する。
- 3) 個々のデータへのリンクを張る。
- 4) Topic Map Navigator 等のソフトを開発する。

1) の作業については、筆者がテーマ別にカテゴリー分けした以上に、ニュートラルに行わなければならない。即ち存在論的な意味空間を構築できるようなトピックの設計を行うことが必要となる。今回は場所(家、部屋、庭等々)、人物(光源氏、紫の上、六条御息女等々)行為(ここでカテゴリーを生かし、遊ぶ、恋をする、笛や琴などの楽器を演奏する、花見をする、等々を設定)を選択肢とする。

2) の作業で、トピックの定義や結合方式を XML 形式でコーディングする。

3) の作業で、トピックの連合先、いわゆる画像へのリンクをはる。ここで有意義な知識の結合が具体化される。

例えば、

- \* 左馬頭が笛を吹く。2-8 図の表示
- \* 光源氏が六条御息女と恋をする。4-18 図の表示
- \* 右大臣が花見をする。32-124 図の表示
- \* 空蟬が囲碁をする。3-14 図の表示

4) のソフト開発に関しては困難が予想されていた。しかし、2000年11月17日に Ontopia Topic Map Navigator が公開され、フリーソフトとしてダウンロードできるという最新アナウンスが出たことから状況は好転している。

このナビゲータを使用するのならば、プログラミングの必要はない。しかしながら、このソフトを利用するためには、Ontopia の定義に添った XML DTD(Document Type Definition)の利用が条件となる。作

成された Topic Map が有効なものであれば、ナビゲータのブラウジングで求める結果を表示できるのである。ただ、残念ながら本論文には間に合わないため、最終的なチェックはできなかつた。いずれにしても、Topic Map のソフトがフリーで入手できることから、今後、この手法の他分野への適用実験が一層進むと予想される。

#### 4.3 DTD サンプル

ここで、実験的なものであるが、XML Topic Map の DTD のサンプルの大枠を作成してみた。

##### XML Topic Map DTD (試案)

```
<?Xml version="1.0" encoding="Shift_JIS"?>
<!ELEMENT topicmap
  (topic | assoc | mergemap)>
<!ELEMENT topic
  (topname | occurs)>
<!ATTLIST topic
  place CDATA #REQUIRED
  person CDATA #REQUIRED>
<!ELEMENT occurs EMPTY>
<!ATTLIST occurs
  instanceOf #PCDATA #IMPLIED>
<!ELEMENT asoc EMPTY>
<!ATTLIST asoc
  id ID #IMPLIED
  xLink #PCDATA #IMPLIED>
```

現在のところ、トピックの指定をどのように行うかは検討中である。また画像データを ENTITY 指定にして外部ファイルとする予定であるが、その場合の ID をどのように記述するかも未定である。

前出の Ontopia Navigator 用の DTD への修正も含め今後の課題としたい。

#### 5. 将来の展望

ネットワークの進展がもたらした大きな変化の潮流の一つに、先にも触れたが、パーソナライゼーション(個人化)が挙げられる。パーソナライゼーションへの傾斜は、例えば One to One マーケティングであり、On Demand Publishing の出現など随所に見いだすことができる。

一方、情報技術の進展とともに、ネットワークを

した情報共有の検討が盛んであり、従来の図書館の機構を発展させた「電子図書館」や「電子博物館」なども実現されつつある。

しかし、上記のパーソナライゼーションの枠組みは、むしろ情報資源そのものを管理対象とするこれらの機構にこそ先んじて適用されるべきと言えよう。すなわち、個々のユーザーの視点に対応した情報提供であり、情報利用者の視点に立ってパーソナライズされた情報の構築が情報の有効利用の鍵を握る。

今後は単にデータの管理者視点からの統制に従った情報蓄積に加えて、個々の利用者の視点に立った電子化ドキュメントの”地図”作りが重要となろう。言い換えれば、Topic Map に代表される情報資源の組織化の研究を通して、情報資源のメタレベルでの再構築・再編成が目指されねばならない。

そのためには、個別分野ごとに、トピックの存在領域に根ざした「トピック地図テンプレート Topic Map Template」を構築し、あわせてその有効性を評価する必要がある。テンプレートの構築には、情報知識、情報工学、知能情報学の研究者の協力を仰ぐ必要がある。また、個別科学を構造的・総合的に構築できる分野ごとの専門化の協力も必要となろうし、ユーザーの利用形態についての実証的研究も忘れてはならない。

今後、Topic Map の研究を通して、情報資源のメタレベルでの再構築の方途を探り、将来的には、各分野のトピックを標準化させ、意味ネットワークとして利用者に流通させることが期待される。本研究は、そのための小さな第一歩である。

#### 7. 謝辞

本論文の執筆にあたっては、(株)シナジーイン・インキュベートの菊田昌弘氏と内藤求氏に様々な助言を頂戴した。紙面を借り感謝の念を表したい。

#### 参考文献:

- 1 Benedicte Le Grand, Michel Soto, "Information management-Topic Map Visualization, Conference Paper at ACM CIKM'99
- 2 Steve Peppe, "The TAO of Topic Maps, Conference Paper at ACM CIKM'99
- 3 ISO/IEC 13250, Information technology-SGML Applications- Topic Maps(ISO, Geneva 2000)
- 4 C. ゴールドバーク「SGMLハンドブック」、アスキー 1993

## 文字冗長度による日本文学分析

Analysis of Japanese literature by redundancy

福田 宏、 山下泰弘\*、勝矢光昭

Hiroshi Fukuda, Yasuhiro Yamashita, Mitsuaki Katsuya

静岡県立大学経営情報学部、〒422-8526 静岡市谷田 52-1

The university of Shizuoka, Yada 52-1, Shizuoka 422-8526

\* 科学技術振興事業団、〒332-0012 川口市本町 4-1-8 川口センタービル

Japan Science and Technology Corporation, Honmachi4-1-8, Kawaguchi 332-0012

あらまし：コンピュータの発達により、統計処理に基づく計量分析の道が文学作品に対して切り開かれた。そこでは、品詞の出現確率や、異なり語の情報を用いて、わが国の代表的な古典文学、「源氏物語」と「手枕」、「雲隠れ六帖」、「山路の霧」等の文体比較が検討されている。更に「源氏物語」における助詞や助動詞の種類とそれらの出現確率が統計数理的に解析され、文学作品に対して計量分析研究分野が確立しようとしている。

本論文では文学作品に対し、情報理論の基本的な概念、エントロピーあるいは冗長度による分析を試みる。これらエントロピーあるいは冗長度等の経年変化や作家による比較等を議論し、これらが文学作品分析の上で興味深い指標となる可能性を指摘する。

Summary: The study of the quantitative analysis of Japanese literature by using the appearance probabilities of the verbs and the auxiliary verbs has been established. In this paper, we point out that the entropy and the redundancy are the interesting indices of the literature in

order to study the secular changes of the Japanese used in the literature and compare the literature works written by the different writers.

キーワード：エントロピー、冗長度、日本文学、計量分析、源氏物語

Keywords: entropy, redundancy, quantitative analysis, Japanese literature, Genji Monogatari

### 1 はじめに

シャノン<sup>1</sup>によるエントロピー、相対エントロピー、冗長度等の概念の認識以来、人間の情報伝達手段として用いられる自然言語のエントロピー（自然言語一文字が持つ平均情報量）を評価する試みが多くの研究者によりなされてきている。自然言語を以下の観点から分析することは、自然科学の観点からも、人文科学の観点からも興味あることである。

1 種々の文化圏で独自に発達した自然言語の一文字が持つ情報量や冗長度の比較

2 種々の作家により書かれた文学作品中の一文字が持つ情報量や冗長度の比較

### 3 自然言語の一文字が持つ情報量や冗長さの経年変化

自然言語は典型的な記憶情報源である。この性質を考慮して、上記の観点を議論するには、一定長の文字列の情報量や冗長さから一文字のそれらを求めなければならない。具体的には、N 続きの文字列の平均情報量と (N-1) 続きの文字列の平均情報量を計算し、始めの量から、後の量を引くことにより、一文字当たりの平均情報量 (エントロピー) を求めることになる。そして、この作業を種々の N について実行し、慎重に結論を下さなければならない。

このようにして、自然言語における一文字当たりの平均情報量を評価しようとするれば、膨大な量のデータが必要となる。特に日本語で使用される文字のように、その種類が多い自然言語では、量的に定まっている過去の文学作品や、一人の作家の文学作品から信頼できる一文字当たりの平均情報量 (エントロピー) を求めることは容易ではない。

日本語事象系に対して英語事象系では、もしも大文字と小文字を区別しなければ、文字種は 26 で足りる。この理由のために、英語一文字の持つ情報量や冗長さを求めるには日本語ほど、大量のデータを必要としない。

例えば、Brown 達<sup>2</sup>は 1.75bits を英語一文字のもつ情報量の上限として 1992 年に発表している。Brown 達の計算方法を採用して、Han 達<sup>3</sup>は韓国語について同様の分析をおこない、6.01bits を韓国語一文字のもつ情報量の上限として 1996 年に発表している。これらの値から、その自然言語データを圧縮する際の理論的境界が導かれる。

情報科学の立場では、一文字のもつ情報量は非常に興味深い量であるが、人文科学の立場では、これらの値が単独に扱われるだけでは全く無意味であり、現在に至るまで、情報理論に基づく文学の分析は、注目されることはなかった。

しかしながら、日本の古典文学に対して、助詞や助動詞の種類とそれらの出現確率が統計数理的に

解析され、計量分析研究分野が確立しようとしている現在、情報理論に基づく指標を用いて日本文学の分析を試みることは興味深く思われる。

## 2 文字エントロピーと言語エントロピー

本論文では無記憶情報源の極限として日本語事象系を考察する。日本語に限らず自然言語は典型的な記憶情報源 (マルコフ過程) である。従って無記憶情報源の極限で得られるエントロピーは言語の正しいエントロピーとは言えない。そのためここではそれを文字エントロピーと呼び、言語エントロピーと区別する。そして文字エントロピーから導かれる相対エントロピー、あるいは冗長さの経年変化や作家による比較等を議論する。

文字エントロピーの定義式は以下の通りであり、以下で述べるように言語エントロピーの第一次近似とみなせる。

$$H_0 = -\sum_i p_i \log p_i \quad (1)$$

ここで  $p_i$  は  $i$  番目の事象の発生確率を示す。対数の底は 2 を採用した場合、bit 単位の値になる。情報理論的には「文字エントロピーは自然言語エントロピーの十分安全な上限値である」ことを主張できる。著者達はその上更に文字エントロピーから導かれる相対エントロピー、あるいは冗長さが人文科学における新しい現実的な計量指標となる可能性を期待している。著者達は本論文でこのことを実際に分析し確認したい。

コンピュータの発達により、統計数理の立場から文学作品の計量分析の道が切り開かれた。<sup>5</sup> そこでは、品詞の出現確率や、異なり語の情報をを用いて、わが国の代表的な古典文学、「源氏物語」と「手枕」「雲隠六帖」「山路の露」の文体の比較が検討されている。さらに、「源氏物語」における助詞や助動詞の種類とその出現確率が統計数理的に解析され、「源氏物語」のような文学作品の計量分析研究分野が確立しようとしている。計量分析では統計学に基

づいて注目している品詞の出現頻度等の比較・分析・検定をしている。しかしながら、注目する統計量の取捨選択は、多分に主観的になることは避けられない。

文字エントロピーは情報理論において、明確に定義された平均情報量と密接に関係している統計量である。文字エントロピーは文字の意味は問わないで、単純に文字の出現確率にだけ注目し、文字の情報量の平均値として抽出される客観的な量である。文字エントロピーや、それから導かれる相対エントロピーあるいは冗長度等の量は、現実的な客観的なデータ量であり冒頭に述べた三つの観点の分析を可能にする。このため、これまで進められてきた文学作品に対する計量分析研究の相補的な側面を持つものであると考えられる。

次に言語のエントロピーについて解説する。言語のエントロピーを計算する方法はシャノンにより以下のように定式化された。先ず N-gram エントロピーを以下のように定義する。

$$F_N = -\sum_{i,j} p(b_i, j) \log p_{b_i}(j) \quad (2)$$

ここで  $b_i$  は N-1 続きの文字列を意味する。  $j$  は  $b_i$  に続く任意の文字である。  $p(b_i, j)$  は、N 続きの文字列  $(b_i, j)$  の出現確率を表わす。  $p_{b_i}(j)$  は

$b_i$  文字列の次に文字  $j$  が出現する条件付確率である。条件付確率には以下の (3) 式が成立する。

$$p_{b_i}(j) = \frac{p(b_i, j)}{p(b_i)} \quad (3)$$

(2) 式は (3) 式の性質を使えば、以下のように表される。

$$F_N = -\sum_{i,j} p(b_i, j) \log p_{b_i}(j) + \sum_i p(b_i) \log p(b_i) \quad (4)$$

上記 (4) 式は、一文字の平均情報量は、N 続きの文字列の平均情報量から、(N-1) 続きの文字列の

平均情報量を引いたもので定義されることを示している。

小さい N に対しては、この方法で一文字の持つ平均情報量の計算は可能であるが、大きい N に対しては、特に日本語のように文字種が多い言語ではこれを評価するには膨大なデータを必要とし、現実的ではない。

本論文で著者達が提案している文字エントロピーは一般式 (2) の N=1 に相当するものであり、自然言語のエントロピーとは言えないが、これと密接に関係した客観的な計量分析に利用できる量であり、著者達はこの論文で「文字エントロピー」と呼ぶことにする。

### 3 相対エントロピーと冗長度

事象系の事象数が増加すれば、文字エントロピーは増加し、事象に変化がない同じ事象系を長年にわたり使い続ければ、文字エントロピーは経年により減少することは十分予想ができる。事象系の事象数は、経年により変化する可能性があるし、同じ時代の文学作品であっても、その中で使われる文字種は必ずしも同じではない。従って、文字エントロピーを指標とするよりも、以下で定義される相対エントロピーあるいは冗長度の方が指標として適しているとも考えられる。

$$R = \frac{-\sum_i p_i \log p_i}{\log n} \quad (5)$$

ここで、n は事象の数を意味し、 $\log n$  はこの事象系の最大エントロピーである。この量 R は現実の文字エントロピーの事象系が本来持つことができる最大文字エントロピーに対する割合である。情報科学の観点からは、この量が経年によりどのように変化するか、興味あるところである。

情報理論的に全く無駄のない事象系では、相対エントロピー R は 1 である。無駄が多ければ R は 0 に近づく。そこで、冗長度として以下を定義する。

$$\eta = 1 - R \quad (6)$$

この冗長度は、無駄が多ければ1に、無駄が少なければ0に近づく量である。本論文では文字エントロピー、相対エントロピーあるいは冗長度を指標として約1000年昔の文学作品、「源氏物語」と近・現代文学作品の分析を試みる。

#### 4 データベース

冒頭に述べた3つの観点を議論するための研究対象は、できるだけ古い時代から最近迄の文学作品が望ましい。そこで研究の手始めとして、著者達は古い文学作品の代表として、紫式部が書いたとされる「源氏物語」、そして近・現代文学作品の代表として、芥川龍之介の「羅生門」、川端康成の「雪国」、夏目漱石の「こころ」、松本清長の「点と線」、三島由起夫の「金閣寺」等を選んでデータベースを構築した。本論文に対する評価が得られれば、更に、途中の年代の文学作品をデータベース化し、本研究を進展させたいと考えているが、研究の緒についた現段階では、最古の作品と現代の作品の比較・分析に留めた。

「源氏物語」に出現する文字種や文字数を統計的に分析するため、「源氏物語」54巻すべてをパソコンで統計分析が可能ないようにテキストファイルとしてデータベース化し、これを用いて本論文では分析をしている。データベースの作成は、静岡県立大学経営情報学部・情報科学研究室が独自に行った。データベースの作成に用いた本文は、池田亀鑑編著「源氏物語大成」(中央公論社)である。これの分析に際しては、村上・今西による「源氏物語の助動詞の計量分析」<sup>6</sup>でなされたように54巻を4つのグループに分けて検討しなかった。データベースには空白データがたくさん存在するが、本研究では改行データはカウントし、段落が終わり新しい段落が始まるまでの空白はカウントしないことにした。

近・現代文学作品として、芥川龍之介の「羅生門」、川端康成の「雪国」、夏目漱石の「こころ」、松本清

長の「点と線」、三島由起夫の「金閣寺」等を選んで、統計的な分析が可能ないようにテキストファイルによるデータベースを静岡県立大学経営情報学部・情報科学研究室で構築した。これに用いた本文は、CD-ROM版の「新潮文庫の100冊」である。現在はインターネット上で入手できる文学作品のデータベース化作業を進めている。

#### 5 情報理論に基づく日本文学分析

表1に近・現代文学作品に対する文字エントロピー、冗長度等データを示す。

表1 近・現代作品の文字エントロピーと冗長度

作品名	$H_0$	$\eta$
こころ	7.33	0.335
金閣寺	7.60	0.331
人間失格	7.32	0.321
雪国	7.27	0.322
点と線	7.36	0.313

文字エントロピーでは、例えば三島由起夫の金閣寺のように突出した値も存在するが、冗長度(相対エントロピー)で比較すると、はほぼ他の作品と同じ値になることは興味深い。

表2には、源氏物語における各帖毎の同様の量を示す。図1、2は、これらのデータを図に表したもので、これから分かるように、文字エントロピーの推移は殆ど無秩序であり、所々に出現する極端な文字エントロピーの値は気になる所であるが、源氏物語の各帖の文字数が文字エントロピーの計算には不十分なために生じたものと考えられる。この図から、源氏物語の文字エントロピーは大体5.5から5.8ビットという非常に狭い範囲に収まることが分かる。上記各帖の文字エントロピー、冗長度を独立量とし、箱髭図で示したのが、図3、4である。その右に、近・現代文学作品の代表作(こころ、金閣寺、人間失格、雪国、点と線)によるデータも示した。源氏物語と近・現代文学作品の間には約1000年の経年があり、図から明らかなように、文字エントロピーはこの間、約1.3倍程度増加している。

表1 「源氏物語」における文字エントロピーと冗長度一覧

帖番号	帖名	$H_0$	$\eta$
1	きりつほ	5.78	0.238
2	はは木	5.65	0.244
3	うつせみ	5.55	0.161
4	夕かほ	5.65	0.254
5	わかむらさき	5.67	0.236
6	すえつむ木	5.60	0.195
7	もみちの賀	5.71	0.242
8	花のえん	5.76	0.189
9	あふひ	5.74	0.246
10	さか木	5.75	0.263
11	花ちるさと	5.45	0.125
12	すま	5.72	0.260
13	あかし	5.84	0.265
14	みをつくし	5.72	0.240
15	よもきふ	5.60	0.205
16	せきや	5.59	0.142
17	えあはせ	5.71	0.212
18	松かせ	5.63	0.213
19	うす雲	5.69	0.219
20	あさかほ	5.57	0.181
21	をとめ	5.74	0.265
22	玉かつら	5.69	0.257
23	初音	5.61	0.187
24	こてふ	5.59	0.177
25	ほたる	5.56	0.182
26	とこなつ	5.56	0.178
27	かかり火	5.61	0.139
28	野わき	5.63	0.179
29	みゆき	5.62	0.225
30	藤はかま	5.68	0.194
31	まきはしら	5.68	0.238
32	梅かえ	5.71	0.237
33	藤のうらは	5.81	0.254
34	わかな上	5.77	0.297
35	わかな下	5.71	0.275
36	かしは木	5.60	0.196
37	よこ笛	5.65	0.193
38	すすむし	5.74	0.220
39	夕きり	5.63	0.244
40	みのり	5.69	0.207
41	まほろし	5.77	0.239
42	にほふ兵部卿	5.93	0.226
43	こうはい	5.79	0.213
44	たけ川	5.79	0.246
45	はし姫	5.66	0.239
46	しいかもと	5.73	0.246
47	あけまき	5.64	0.251
48	さわらひ	5.60	0.163
49	やとり木	5.72	0.279
50	あつまや	5.73	0.268
51	うき舟	5.61	0.235
52	かけろふ	5.62	0.243
53	てならひ	5.78	0.284
54	夢のうきはし	5.56	0.168

図1：源氏物語におけるエントロピーの移り変わり

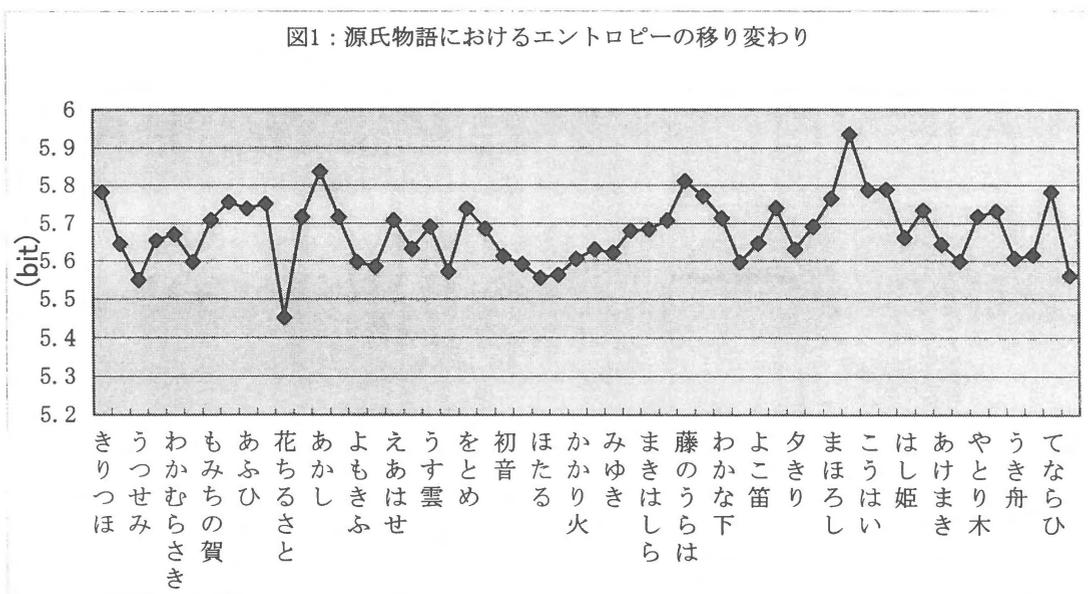
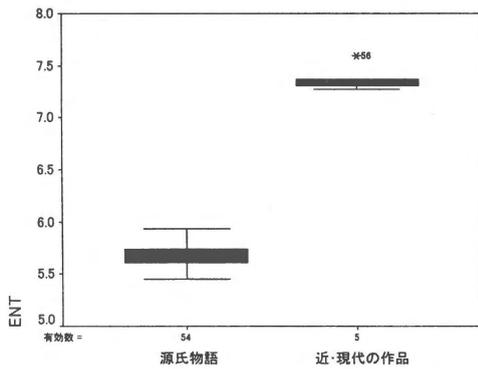
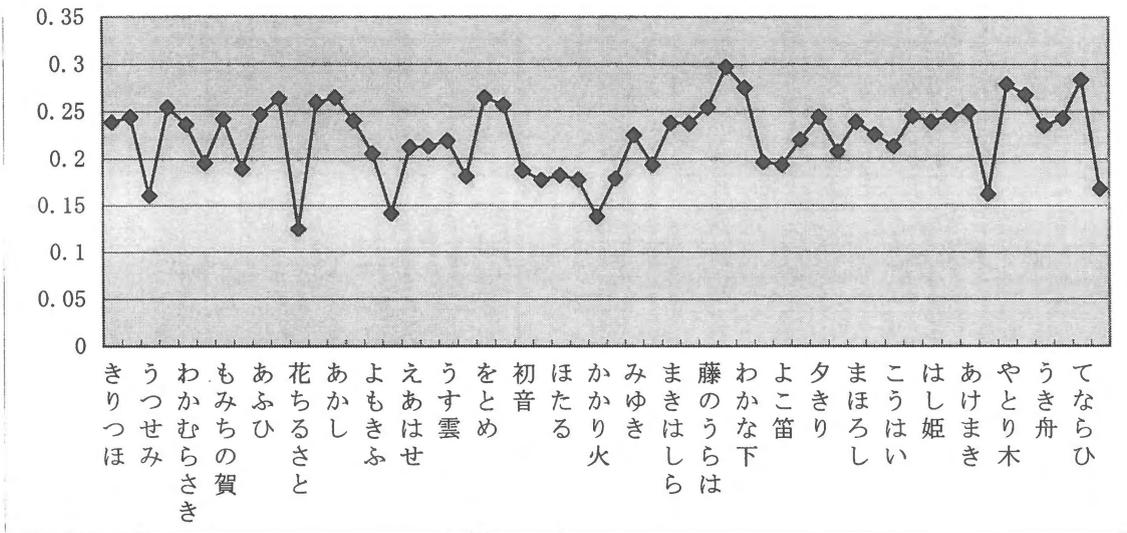
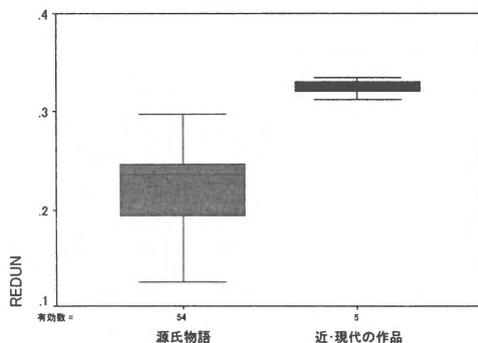


図2：源氏物語における冗長さの移り変わり



作品

図3 文字エントロピー箱髭図



作品

図4 冗長さ箱髭図

参考文献

- 1 C. E. Shannon. The Mathematical Theory of Communication Bell System Technical Journal, Vol. 27(1948), pp. 379-423; ibid pp. 623-656.
- 2 P. F. Brown, S. A. D. Pietrra, V. J. D. Pietra, J. C. Lai, and R. L. Mercer. An. Estimation of an Upper Bound for the Entropy of English. Computational Linguistics, Vol. 18 (1992), pp. 31-40,
- 3 Y. S. Hanm H. R. Park, J. H. Shin and K. S. Choi. An Upper Bound Estimate for the Entropy of Korean. Literary and Lingustic Computing Vol. 1(1996), pp. 141-146.
- 4 村上征勝、今西祐一郎 源氏物語の助動詞の計量分析 情報処理学会論文誌 第40巻 (1999) pp. 774-782.

## 広領域分野資料の横断的アーカイブズ論に関する基礎研究

### A Fundamental Study on the intersectional Archives Theory for the Multidisciplinary Objects

静岡大学情報学部 八重樫純樹

Jyunki.YAEGASHI The Faculty of Information,SHIZUOKA Univ.

〒432-8011 浜松市城北町3丁目5-1 静岡大学情報学部

The Fac. of Information,SHIZUOKA Univ., 3-5-1,Jyohoku-Cho,Hamamatsu,Sizuoka,Japan

#### 【あらまし】

社会の情報技術普及と高度化は急速であり、かつ社会の情報空間は世界規模となってきた。しかしこれら技術進展に比較し、文化財をはじめ社会的共有資料としての情報資源構築が円滑に進んでいない。これは社会共有資料管理・運営において情報記録・管理の基本的な問題解決と普及がなされてこなかったことに起因する。世界のコンテンツは Dublin Core活動に見られるよう標準化の方向に動いている。本論はこれらに早急に対処し、かつ社会共有情報資源構築の基盤であるアーカイブズ論について、分野横断的アーカイブズ論としての可能性とその端緒を探るため、歴史的資料を中心に資料論の整理を行った。

#### 【Abstract】

Despite a rapid growth of the social interest in information and related technologies, Japan falls far behind countries in constructing the social information infrastructure such as databases. Vast majority of objects with educational, academic, and/or cultural importance are left unexplored.

This study aims to establish a basis for the intersectional Archives Theory of information conveyed in these previously unexplored multidisciplinary objects. We demonstrate the feasibility and characteristics of systematic modeling of these objects.

Our approach here is based on the modeling of information activity as a communication activity, and objects concerned as the media for this activity, thereby revealing the characteristics of interdisciplinary information activities.

## 1. はじめに

社会一般の情報化の動向は加速的に急速であり、あらゆる社会分野を巻き込んでいる。これは世界的傾向であり、多少のことはあれ、国内のこの傾向は今後も続くであろう。数年前からすでに社会は情報化の動向にあり、この社会情報化の基盤整備が論議されていた。この課題と目標として、

- 社会情報技術基盤の形成
- 社会情報体制基盤の形成

の二つが必要であるといわれおり、それらを中心に議論が進められた。しかし、社会情報化の基盤をなす社会情報資源構築基盤の形成についてはあまり触れられておらず、上記二つの基盤が整備されれば、社会のなかから自動的に生成されてゆくものと考えられていたのではなかろうか。これは、その後各省庁や自治体で遂行されてきたメディアやコンテンツ構築および情報関係諸研究に関する諸政策や動向の上から推測できる。しかし、現状にいたり上記の技術基盤は米国製のインターネットとウインドウズに完全に席卷されたが、確かに社会に普及した。また、ここ数年、社会論はコミュニケーション論やメディア論が盛んであったが、現実の社会実学としての社会情報体制基盤の整備はこれから問題を探り、どうするかという状況にあるものと認識する。

また、情報技術 (IT:Information Technology) の社会の速効的な効果的としては、欧米との広域性の観点から経済・流通分野や経営管理分野が考え、現在、E-Commerce等の問題が論議されている。しかし、ITが社会に長期的に定着するためには経済活動コンテンツのみならず日常の社会生活や文化、教育、学術、教養等の知的な社会共有情報資源としてのコンテンツの整備が不可欠である。上記の社会情報基盤整備政策決定の上では自動的あるいは小手先で解決する問題であると考えられたのかもしれない。

社会諸資料の社会共有情報資源としてのデジタルコンテンツやデータベース作成の条件としてコンピュータ可読資料となっていなければならない、前提としての資料情報の整理と管理がなされていなければならない (デジタル入力容易

性と資料の整理前作業 (資料内容記録や索引情報、資料整理の管理構造等))。社会共有情報資源の対象となる各種社会分野の資料管理体制と業務がそのように整えられ、すでに多くの蓄積が社会的に累積されているとするなら上記の基盤整備に多少のテコ入れで社会の情報化は円滑に進捗するであろう。

問題はここであり、国内において比較的資料情報整理と管理が業務として定着しているのは貸借関係が主な業務となっている分野、例えば図書館、特権 (特許)、金融・財産管理関係等である。しかし、多くの分野では資料そのものが一過性で棄却対象であったり (主に行政文書関係)、資料保存が重要な意味を有しているので、資料そのものの保存に力点がおかれ専門家に管理が委ねられ、資料情報記録・管理が最小限にしかなされていなかった分野 (主に文化財や博物館資料等の資料) 等、各種存在する。国内においても、特に文化財や博物館資料の世界では、欧米同様に資料一般の情報記録・管理の社会的設定と一般的方法論の確立についての重要性は言われていたが、上記のように社会の学術諸分野単位、あるいは専門領域や専門機関単位の狭い範囲で遂行されてきており、一般化については殆ど触れないできており、現状を迎えている。しかし、世界の情報化動向はきわめて加速的に進展しつつあり、情報技術基盤はインターネットの世界的広域化と社会諸分野への急速な進展・普及により、社会共有情報資源やコンテンツの量が国際的な立場を左右する状況に至りつつあり、この解決にはどうしても資料の情報記録・管理の問題抜きには解決に至らないであろう。これはまさにIT以前の社会共有情報資源構築体制基盤整備の基礎である。

本研究は上記視点から一つの分野や領域にとらわれず、人間・社会・自然等、多様な活動の実態 (情報活動行為) を反映し生産される諸資料群を対象に、そのアーカイブズとしての情報・記録管理の視点として、資料情報のアナロジーと差異の分析を通してモデリングし、社会資料の情報化方法論の基礎を創ることを目的とするものである。現実社会で生成・利用される社会資料群は現実社会において広範に共有知識情報を得ることが可能であり、何らかの収束の方

向性が得られる（これらも時間とともに歴史的資料となるが）。しかし、時間性を含む歴史系資料群は資料そのものが生成・利用された空間認識から、また時間性により物理的、化学的ダメージを受け、多くの社会的情報が壊失されている。この課題を中心に据えて、資料群の分野横断的アーカイブズ論に関する考察を行うものである。

## 2. 資料群の情報概念整理と情報化過程

社会的資料群はもともと学術、教育、産業等、社会や自然の諸活動（情報活動行為）で生成・利用される諸活動の目的であり、過程であり、結果である。これらの情報化における情報モデリングと記述方法論として、資料群そのものと生成される背景空間の構造（Structure）と文脈（Context）の情報認識・記述が要点である。そして技術と人間の諸活動との関わりにおいてはデータが基本的な要素となる。この認識と要点抽出の基盤的として、情報概念について以下整理する。

情報概念については従来から、

- (a) データ概念と情報概念の区分が明白でない理工学系（特に古典的なシステム工学やコンピュータ科学分野の一部にまだ見受けられる）。
- (b) 人間－自然の活動におけるすべての信号－認識事項。社会学系では人間－自然活

動のすべての森羅万象におけるパターン  
の差異を情報として定義している（文献  
(4)(6)(23)(24)等]

の二つのアプローチが存在した。しかし、人間諸活動におけるITとの整合システムを前提とした人間－機会系の組織化とそれらの融合的な創造活動を情報システムシステムとして考え、また情報システム構築をその立場において考える場合、上記二つのカテゴライズでは狭小過ぎる（(a)の場合）かあまりにも取り留めがなさすぎる（(b)の場合）。この立場から、文献(9)を参考にし、要約的に示すと、

(c) 情報はヒトの自己、外界の物事に対する意味の感性、認識であり、データはこれらを形式的、体系的に記述したもの（二次的資料群）。

といえるであろう。情報システムとしての仕様が決定され、あくまでもITを中心に情報システムの一部設計を考える場合、いまだ(a)のカテゴライズでも機械系の効率的設定としての観点から有効かもしれない。しかし、(b)のカテゴライズはこれらが設定され、稼働し、社会と宇宙全体が如何なることかを考える場合であり、森羅万象の宇宙論ではかえって現実の情報システム設定や構築においては、問題や方法論の所在を曖昧にするか見えなくする論理であるものとする。 (c) の概念規定について Fig.1として示す（実世界における資料群はここでは上記(c)との対応では一次資料群であ

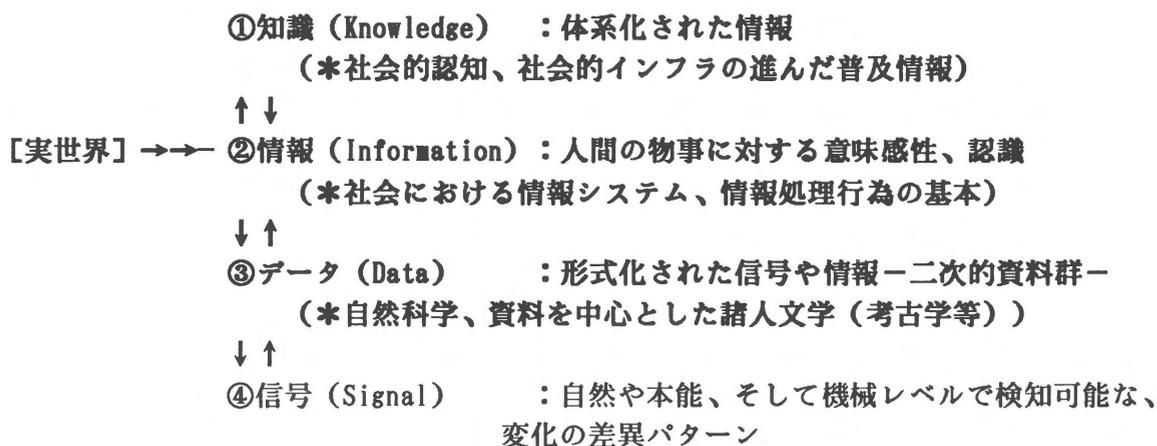
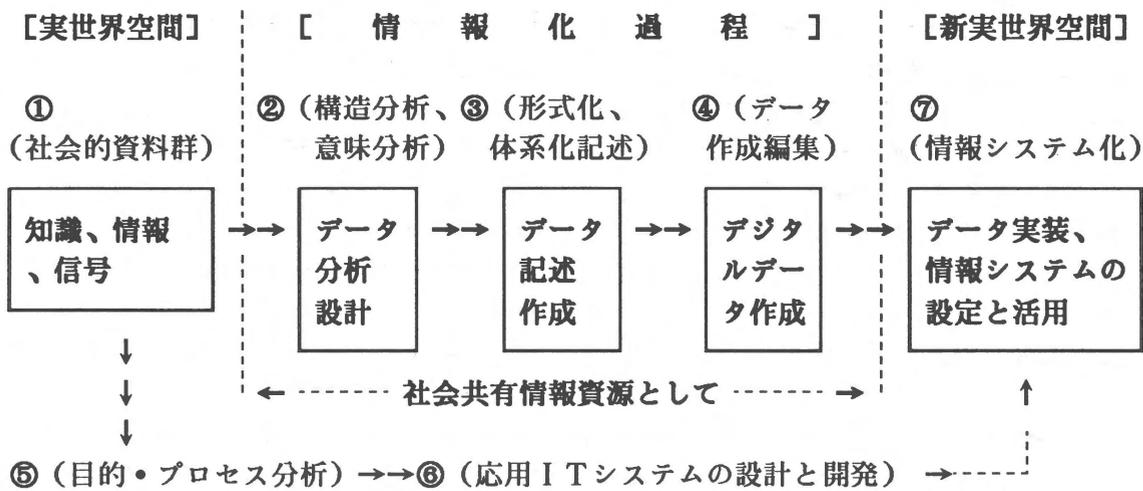


Fig. 1 情報概念の階層性



(注) ・①から⑦への変換系が情報システム化  
 ・②、③、④が基本的データ生成過程  
 ・⑤、⑥、⑦が機械制御系の設定過程

Fig. 2 情報化過程のモデル

る。しかし、以降、本論では一次資料群は資料群、二次資料群はデータとして記述する)。また、実世界の情報化過程におけるデータの位置づけについて、Fig.2に示す。

### 3. メディアとしての資料群

#### 3.1 資料群生成過程とメディアモデル

社会的資料群とは実世界における人間-自然界諸活動の経過か結果として具体的に生成されたものである。したがって、それらが生成される実世界空間の信号あるいは情報が具体的に記述されているか、内包されている存在である。人間-自然界諸活動における一つのまとまった社会的行為を”事象(Event)”とすると、事象はそれが生じる実世界空間の意味において事象として認識され実世界に作用する。また事象そのものは何らかの人為的か自然的な組織体構造(プロセスの集合体)が存在し、それらが偶然か自然あるいは人為的に内部的機能を遂行して生成される一つの結果と考えられる。

したがって、資料群個々は事象の結果あるいは生成過程においてそれぞれ生成されるものであり、これらは事象が生成される組織構造体機

能の一過程あるいは結果の生成体であり、実世界空間と事象のそのものの意味か背景文脈(Context)の意味を内包あるいは表層に記述された存在である。つまり、資料群個々にはそれが生成された実世界空間と事象の文脈が内包あるいは表層に記述されており、そして事象が生成される組織構造体(Frame)の一部に位置づけられ、その機能の意味が内包あるいは表層に記述されている存在であると考えられる。

この事象を生成する組織構造体の単位をプロセスとし、これらプロセス群の連続的に生起され、その活動で事象が生成されるものとし、組織体構造活動の単位が一つあるいは複数プロセスの生起と活動であるとし、がその資料群個々は何らかのプロセスとプロセスが機能するための情報の授々として生成され、機能した結果と捉える事ができよう。これはとりもなおさず、プロセス間通信のメディアとしてモデル化できよう。つまり、資料一つ一つは事象を生成する組織構造体におけるプロセス間通信のメディアである。また、実世界は複数の事象の相互作用によって成立しているとする事ができよう。以上を図式化すると、Fig.3のように示せる。

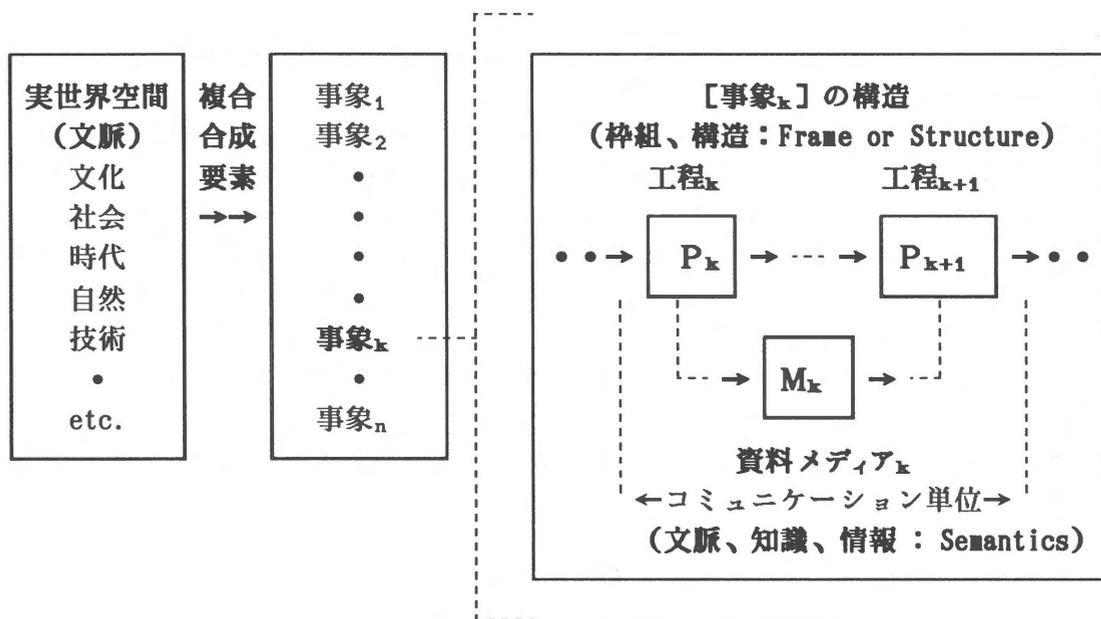


Fig. 3 メディアとしての資料

### 3.2 時間性情報と資料メディア

現時点の実世界空間は、実は時間が経過すると”過去の実世界空間”となり、新たな実世界空間に推移し、新たな事象群と資料メディア群が生成され、時間は止まることがない。そして現時点の実世界空間は常に過去の実世界空間の累積の上に成立している。資料メディア群も同様である。ただし、ヒトは常に現時点で生活せねばならず、そして次に時間的推移する未来実世界空間の生活可能性追求が現時点の生活の活力である。現時点の実世界空間が過去の実世界空間の累積の上に立脚しているとはいえ、ヒトは常に未来空間に生活の視点を置く。したがって過去の実世界空間で生起し、生成された多くの事象群と資料メディア群は、基本的に現時点と未来実世界空間に必要な対象が選別され、忘却あるいは棄却・劣化・消滅してゆくのが自然ともいえる。このため時間は事象群と資料メディア群にとって強烈な雑音空間といえ、忘却や劣化・棄却・消滅の度合いは資料メディア群の時間経過において、置かれた自然環境や素材のよっても異なるが、ほぼ指数関数的である。以上から事象群および資料メディア群の情報性質を考える場合、強烈な雑音空間であ

る時間経過における諸性質の考察が必然である。したがって、これらを、事象が生起された過去の空間を事象空間とし、資料メディアが現在まで経過してきた空間を経緯空間とし、現時点の実世界空間を現在空間とし、資料メディアが経過した異なる三つの空間別に考察するモデルを、ここでは資料メディア群のライフサイクルについて、時間経過との関係をとってFi.4に示す。なお資料メディアを媒介として研究する考古学、歴史学、民俗学、美術史学等が対象としている諸資料メディア群の諸性質とこれら分野研究の特質について、文献[[(16)(19)(26)(28)(29)]等]等で詳しく示した。

現時点における資料メディアと時間性を有した資料メディアとの相違は時間性を有するか、否かである。つまり、経緯空間の存在有無あるいは経緯時間とそれが置かれた環境、時間、素材の問題であり、現時点社会で生成される資料メディアと過去の実世界空間に置いて生成された資料メディアとの差異は経緯空間の存在の問題である。ここをどのように理解、考察するかが根本的な問題解決の基盤である。

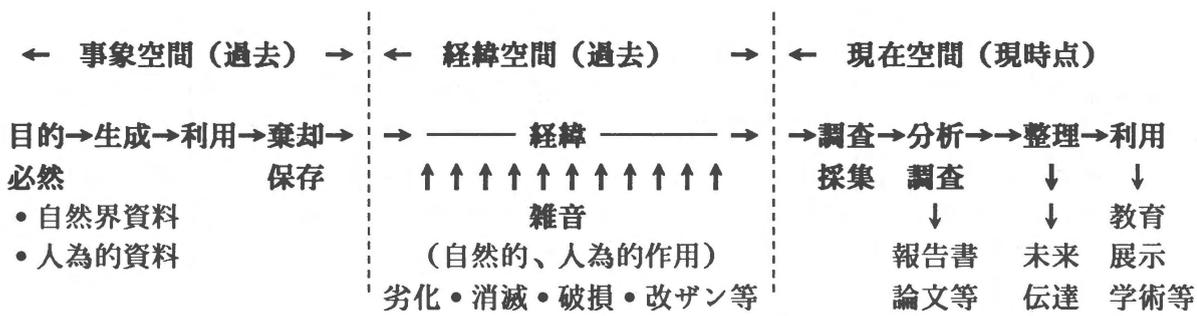


Fig. 4 時間経過と資料メディアのライフサイクル

4. 資料メディアのコミュニケーションとメディアとしてのデータ

4.1 資料メディアのコミュニケーション  
 久保は文献(8)で社会一般のコミュニケーション分類の中で、A型からF型に分類している。この中でアボリジニ研究における諸資料を発信者不明のF型コミュニケーションとして分

類している。現代社会資料メディア群と歴史系資料群の対比を明確にすることと、また、歴史-社会の諸活動の多くはコミュニケーション活動としてとらえる事が可能であり、資料メディア群はこれら諸活動の文脈情報(メッセージ)を記述・内包するメディア群であるのである。

歴史を解釈する歴史学は歴史系資料メディア群をもとに、そのメッセージ解読の行為の歴史

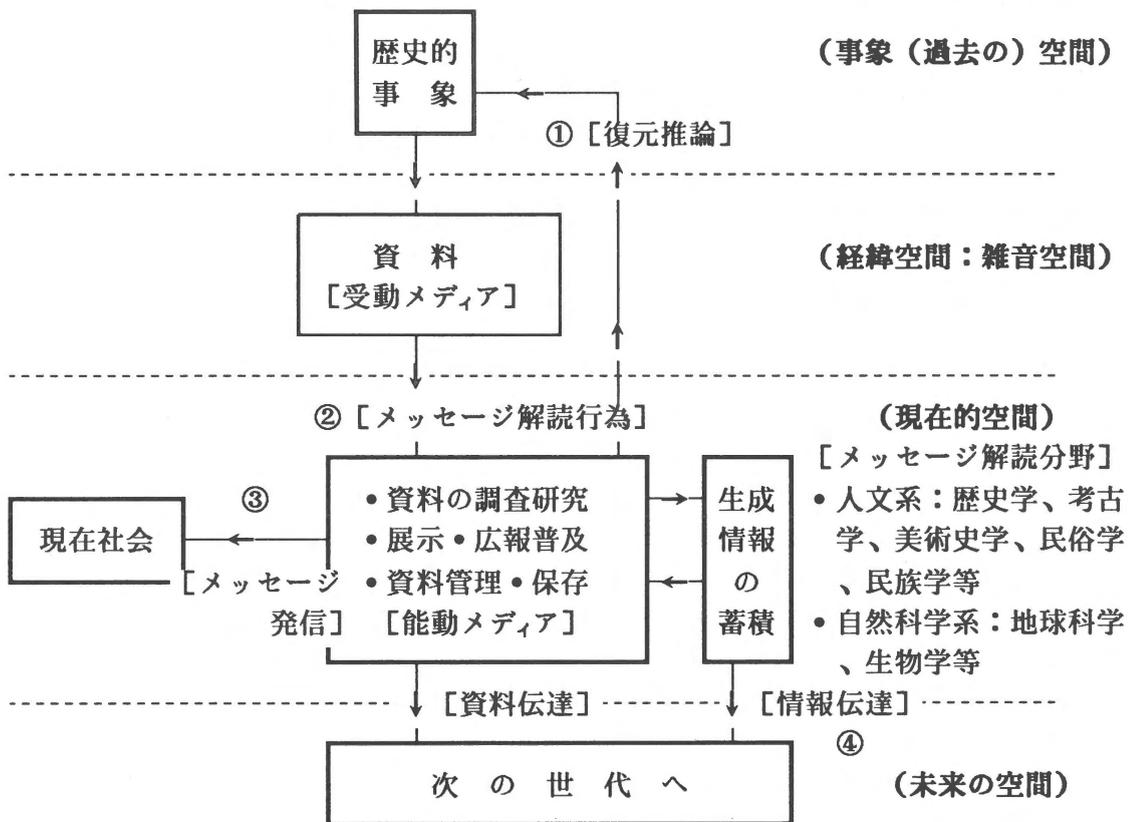


Fig. 5 時間マクロコミュニケーションモデル

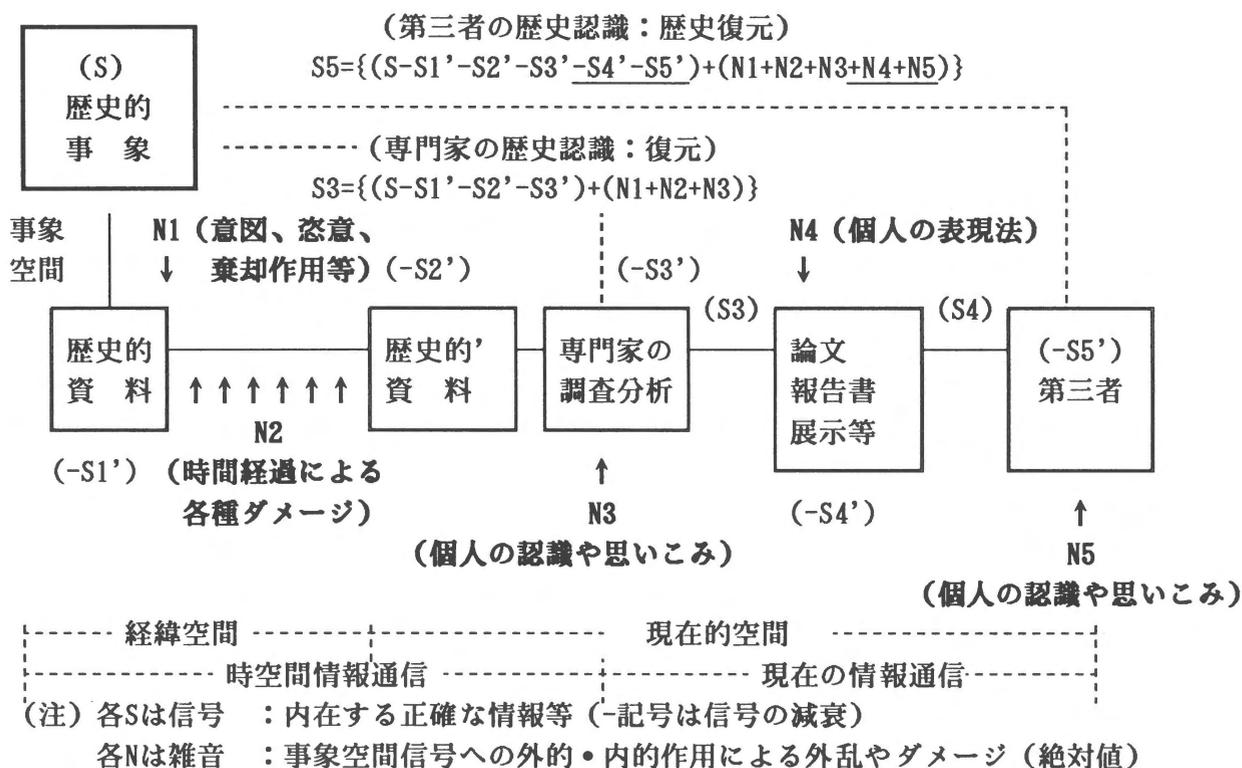


Fig. 6 歴史コミュニケーションのモデル

コミュニケーションであり、またキャッシュカード(メディア)で金を出し入れする行為は何らかのコミュニケーション行為でもある。ここから多分野資料メディア群を共通客体化の基盤で考察する上で、メディア群としてモデリングする。情報を整理するため、三つの時空間別に再配置し、歴史系資料メディア群をもとにしたコミュニケーションモデルをFig.5とFig.6に示す。

Fig.5はミュージアム活動を中心とした社会的能動メディアとしてのコミュニケーションモデルとして、Fig.6は個々の歴史系資料のコミュニケーションモデルとして示す。

#### 4.2 資料情報のカテゴライズと情報記録管理データとしての諸問題

文献 [(2)(5)(8)(14)] 等で歴史系資料メディアの基本的情報を4項組として示した。つまり、全ての資料メディアをEとすると(1)式で示される。

$$\langle E \rangle := \langle T \rangle \langle S \rangle \langle O \rangle \langle C \rangle \dots (1)$$

(T:時間、S:地理的空間、O:対象内容、C:間接的な関連情報)

Fig.3、Fig.4、Fig.5をもとに、以下(1)の各要素について考察する。

##### (1) T:時間

歴史系人文科学研究においては、時間は研究そのもの場合が多い。また、歴史学、考古学、民俗学、美術史学等分野や研究対象において、時間単位の問題(暦(太陽暦と陰暦)、元号、時代等)、時間間隔の問題(年、年代、世紀、時代、時期等)さらに相対時間-絶対時間の問題(歴年代と時代、編年や様式年代等)がある。それぞれの分野、研究対象により個々に使用しているが、現在は西暦年が世界的基準であるが、その文化固有の考古学資料等はそれに相当しにくいものと考えられる。

また、資料情報管理としてのアプローチは、各分野研究における専門資料として考える場合と、社会一般資料として考える場合がある。それぞれの専門分野において資料記述が一般化、標準化されていれば問題は少なく、一部文書関

係には存在しているが、人文学系においてはこれらのための資料論が存在していない。したがって専門資料、社会資料両者の側面からの記述としてはFig. 4の資料ライフサイクルの観点が合理的と考える。つまり、基本的に

- 事象空間 (T<sub>e</sub>) : 発生に関する時間集合
- 経緯空間 (T<sub>p</sub>) : 来歴に関する時間集合
- 現在空間 (T<sub>n</sub>) : 資料化の以降の時間集合

の3つの側面から記述する方法である。これは、(1)における他のS、O、Cにも普遍する事柄である。

### (2) 地理的空間 : S

地理的空間においても時間同様多くの概念が存在する。位置としての点情報、交通など線として情報、また地域や文化圏などの面情報、また地質学等の層としての情報の捉え方が存在する。さらに、緯度・経度のように絶対座標系、海岸線や河川、等高線等の自然地形、道路、建造物、遺跡、市街等の人工物、行政界や地名等の相対的名称など。

これらの多くは時間と共に推移・変化している、現在の地形は百万年、千万年、億年の単位の研究には使えない場合もあり、また歴史性を含む問題は歴史地理学の研究対象そのものの課題で有る場合が多い。つまり確定不能な場合が多い)。かつ分野や研究対象により、その必要な地理的属性は異なる。かつ時間経過と共に推移・変化しており、自然系、人文社会系の対象研究時代によって異なる。時間同様に三つの空間が存在するが、ここは現実的には現在空間から出発し(住所や座標値、海拔等)、事象空間、経緯空間については記述可能な情報から順次記述してゆくしかない。これは対象分野で決定して行くべき事項であろう。

### (3) 対象内容 : O

すべての”もの”資料が対象であり、かつ分野によって資料に対する考え方が異なる。基本的に文字記述された、あるいは何らかの間接的な記述資料の存在する資料群と物体としての”もの”資料群に分かれよう。文字記述資料も生成の時間性が遠いものほど(歴史文書等)情報確認に専門性が必要となってくる。

また物体としての”もの”資料群も専門分野や対象によって異なるとはいえ、基本的に自然

史関係と人文社会関係資料で状況が大きく異なる。自然史系は比較的分類学が進んでおり、かつ標本記録の考えが伝統的に遂行されてきている。しかし、人文社会系”もの”資料群に関しては我国では殆ど情報記録管理のための記述に関する標準化の指向がなされていなかった。これは人類活動の結果生成された多様な資料群であり、その体系化は人類活動そのものが体系化不能であるのと同義であり、困難であることは確かである。

しかし、資料群がその分野の専門家の世界にしか必要なかったことと、専門家あるいは分野の研究視点からの記述方式によりすぎていたことによる。専門分野が一般化されていれば解決の方向性があったかもしれない。が、学説は専門家によって異なり、新資料出現によって大きく変わる専門分野では研究情報を中心とした記述方式の収束の見通しは立たない。また、文化の社会的コミュニケーションメディアであるに関わらず、社会化の視点を伝統的に欠いた結果である。

これは結果であり、現実であるが、社会的コミュニケーションメディアとしての視点から再アプローチが必須である。つまり、各専門分野の文脈理解と知識はどうしても必要であり各専門分野とのコラボレーションが必須ではあるが、社会的メディアの情報管理の側面から管理・索引情報記述の問題を再整理から出発しなければならないものとする。社会はすでにデジタルコンテンツとして画像や音声等、デジタルアーカイブズを容易に受け入れられる状況であるが、索引、管理情報については、特に文化財や歴史系資料に関しては、今後の問題である。

(1)で示した3空間分離した記述の可能性について検討を加える(現在社会資料:事象空間=経緯空間=現在の空間)。

### (4) 間接的関連情報 : C

主に資料に関連した調査報告書、履歴書(カルテ)、論文、所在や管理機関等情報を対象とする。情報化において各資料対象や分野で一番進捗し、かつ基本となるのがここである。これは文献データベース、所在データベース等として構築・運用されている。理系、人文社会系それぞれデータ内容、質、量が異なる。これら分

野それぞれの文献データベース内容調査により分野の研究体制や内容等の文脈が見えてくるものとする。この比較研究も(3)の研究において必要事項となる。これは今後の課題である

## 5. まとめとしての考察

はじめに示したようにIT(情報技術)の社会的普及や技術開発とその応用は急速である。しかし、政府の政策にも関わらず、情報化社会の中味(コンテンツ)としての社会共有情報資源の構築と普及は遅々としている。特に我々社会と文化の基盤をなすべき、全国に散在・管理されている文化財をはじめ各種文書資料や”もの”資料のデジタル化と社会共有情報資源化は未だ軌道に乗ったとはいえない。

この状況は、はじめおよび4章4.2、(3)でも指摘したように社会的アーカイブズ論への無関心さと技術指向に偏した社会の情報化研究活動、及び社会が専門家集団に資料管理を全面的に委託した結果とも言えよう。一方、世界の資料デジタル化の動向は急速であり、サイバー空間が世界規模で展開され、データファイルの標準化はHTMLからXMLの方向へ動き出している。また、資料アーカイブズに関しても

Dublin Core にみられるように、急速に世界的標準化の方向にあり、今後のアーカイブズはこれを無視しては成立しない状況となっている。

この状況において、社会共有情報資源構築の基盤整備は基本的に社会的なアーカイブズ論の整備と普及である。技術が如何に普及し高度化しようと、情報の記録・管理に至る過程は2章で示したように人間が知恵と時間と労力で介在せざるを得ないのである(Fig.2の②、③過程は特に必要である。最近のデジタルアーカイブの動向はともすると、Fig.2における②、③過程の方法基礎論の解明よりも④の技術論と結果としての成果物の生産に力点が注がれている場合を多く見受ける)。また、専門家集団はともすると研究主体で物事の価値を決定する場合が多く、資料や情報の社会化に無関心な場合も多くみられる。したがって資料の社会的客体化視点がアーカイブズにおいても必要となる。こ

のため、時空間コミュニケーションとして歴史的情報活動をモデリングし、資料群を資料メディア群として捉える視点と混在する歴史的専門情報のデータ形成整理方法論として空間情報の分離を4章で示した。

アーカイブズ方法論研究は、情報化社会の情報資源形成のための基本であり、情報化社会形成の上で最も基盤をなすものとするが、国内においては分野個々に、そして各分野の片隅で活動力点の焦点からそれた存在となっている。その結果が現在である。これら活動は分野個々では世界的動向に対処出来るかどうか疑問である。また、特に人間活動痕跡の社会資料群全般を一つ視点から体系的に標準化することは不能である。ただし、人間活動痕跡そして自然活動痕跡に対する社会的視点からの専門分野間における共通的なアナロジーは存在するはずである。そして伝統的経緯の結果として今の各専門分野領域の手法や活動の文脈が存在している。その差異を一本化することは不可能であろう。

本研究は社会資料群全般を視野にいれ、社会的共有情報資源構築基盤としてのアーカイブズ論に関する基礎的方法論の研究を行うため、議論基盤としての資料群と諸研究領域活動について一般化の整理を試みたものである。本論を基盤とし、国際化が相当に進んでいる図書館情報関係専門家、また現在進行中の文書関係アーカイブズ専門家等のノウハウと知恵を注いで頂き、ミュージアム資料関係専門家等や専門分野関係者とコラボレートしつつこれらの問題について研究を進める予定であり、これら専門分野そして研究者、専門家とのコラボレーション活動が必須である。

ヒューマンコラボレーションシステム確立が前提であるが、ここにおいてコラボレーションシステムとしてのITの開発と活用は重要であり、効果してくるものとする。これは社会情報収集や研究、知見交換の協議等において効果するであろう。縄文土偶関連資料研究活動のツールとして現在、プロトタイプを研究開発中である。

## 《参考文献》

- (1) ソシユール (著)、丸山圭三郎 (訳) 1981 『ソシユールの思想』 岩波書店
- (2) 八重樫純樹、小林達雄、野口正一 1984 「縄文時代土偶の情報構造に関する基礎的考察」国立歴史民俗博物館研究報告第3集 pp.251-273
- (3) ハーバース・J (著)、河上倫逸ほか (訳) 1985 『コミュニケーション的行為の理論』 未来社
- (4) マクルーハン、M (著)、栗原裕、河本伸聖 (訳) 1987 『メディア論』 みすず書房
- (5) 八重樫純樹 1989a 「歴史的データの基本構造に関する研究」1989年情報学シンポジウム論文集 pp.115-124
- (6) 吉田民人 1990 『情報と自己組織性の理論』 東京大学出版会
- (7) 八重樫純樹、倉田是 1991 「事例データをもとにした情報検索実験例といくつかの課題」国立歴史民俗博物館研究報告書第30集 pp.207-248
- (8) 八重樫純樹 1991 「歴史系支援情報処理研究の課題」情報知識学会誌Vol.2 No.1 pp.9-22
- (9) 八重樫純樹編著 1992 『国立歴史民俗博物館研究報告—土偶とその情報—第37集』 国立歴史民俗博物館 p.490
- (10) 八重樫純樹 1993a 「歴史系資料・事象情報化に関する研究—その経緯と基礎的課題—」国立歴史民俗博物館研究報告 第50集 pp.461-481
- (11) 八重樫純樹 1993b 「資料の視覚情報化」『科学の目でみる文化財』アボック社 pp.121-149
- (12) 八重樫純樹編著 1993 『国立歴史民俗博物館研究報告—共同研究「歴史学系研究支援情報処理の研究—カタチの情報のデータ形成・索引法—」—第53集』 国立歴史民俗博物館 p.320
- (13) J.YAEGASHI 1994 「An Experimental Study on Organization of Information Resources of Archeological Materials (Remains) Taking as an Example of Clay Figurines of JOMON Period in Japan」『Proc. of 47th FID Conference and Congress』 pp.760-764
- (14) 八重樫純樹 1995a 「歴史情報」『新版情報処理ハンドブック』 情報処理学会編、コロナ社 pp.887-892
- (15) 八重樫純樹 1995b 「”もの” 資料データベースとドキュメンテーション」 第一回アート・ドキュメンテーション研究フォーラム 『美術情報と図書館』 報告書 PP.43-52
- (16) 八重樫純樹 1996b 「資料の情報定性論」『人文学と情報処理誌』 No.11 pp.24-30
- (17) 八重樫純樹、高井健司、宮原健吾、千野裕道 1996 「福岡県における歴史系資料情報化の研究」『MUSEUM KYUSYU誌』 第54号 PP.37-54
- (18) 久保正敏 1997 『マルチメディア時代の起点』 日本放送出版協会
- (19) 八重樫純樹 1998a 「人文系博物館における資料ドキュメンテーションの諸問題」 情報の科学と技術誌 Vol.48, No.2, PP.68-73
- (20) 「土偶とその情報」研究会 (代表: 八重樫純樹) 1998 『土偶研究の地平—「土偶とその情報」研究論集(2)—』 勉誠社、A4.P.419
- (21) 八重樫純樹 1998b 「縄文土偶を例とした資料情報化研究」 Academia 学術新報第174号、PP.41-49
- (22) 八重樫純樹 1998c 「土偶資料を例とした資料情報化研究(3)—広域土偶資料の情報体系調査とコンテンツ形成の基礎課題—」、『土偶研究の地平—「土偶とその情報」論集(2)—』 勉誠社(株)、PP.9-28
- (23) 伊藤守、小林直毅 1998 『情報社会とコミュニケーション』 福村出版
- (24) 林進 1998 『コミュニケーション論』 有斐閣
- (25) 武井恵雄 1999 「情報教育におけるコミュニケーション」『情報教育シンポジウム論文集』 情報処理学会
- (26) 八重樫純樹 1999 「土偶資料を例とした資料情報化研究(4)—情報コンテンツ構築と資料研究の可能性—」、『土偶研究の地平—土偶とその情報』論集(3)—』 勉誠出版(株) PP.13-42
- (27) 坂井知志、高橋信祐、高安礼士、武田国夫、水嶋英治、八重樫純樹 1999 『博物館情報論』、 樹村房、pp.17-27, pp51-60, pp.72-77, pp.145-153
- (28) 「土偶とその情報」研究会 (代表 八重樫純樹) 2000 『土偶研究の地平—「土偶とその情報研究」論集(4)—』、勉誠出版社(株)、A4.P438
- (29) 八重樫純樹 2000b 「多分野資料の情報モデル化に関する基礎的研究」『2000年度研究報告講演論文集』 情報知識学会、pp.29-32

ジャイナ教聖典のデータベース  
— 特殊フォントで表現されたデータのPDF —  
Database to the Canonical Texts in Jain  
- PDF of the Text in Special Font -

逢坂 雄美  
Yumi OUSAKA

仙台電波工業高等専門学校, 仙台市  
Sendai National College of Technology, Sendai 989-3124

**あらまし:**

中期インド・アリアン諸語(Middle Indo-Aryan)で書かれたジャイナ教(アルダ・マガダ語)古文献等の解析のために計算機リソースを開発してきた。これまで、ジャイナ教聖典の5冊の聖典に関するマッキントッシュ版電子ブックを数年前に作成済みである。この電子ブックには、当該研究に必須の基礎資料、電子化テキスト・韻律図式と解析結果・詩脚の正順索引と逆順索引・語彙の正順索引と逆順索引が収録されている。最近のウインドウズマシンの使用状況からみて、電子ブックをこのマシン上でも読むことができるならば、その有用性は大幅に上昇するであろう。特殊フォントを使用しているために、従来のツールではウインドウズ用電子ブックを作成することができなかったが、最近不十分ながらもそのツールが提供されるようになった。ウインドウズ用電子ブック作成とその問題点について議論する。

**Summary :**

We investigated computer analysis of the canonical texts in Prākṛit (called Ardha-Māgadhi), that is, the early Jainist ones in Middle Indo-Aryan (MIA). By using Macintosh we have already published the electronic books for the five important canonical texts on WWW. This book contains the essential data on the study of MIA, that is, the electronic text, metre analysis, pāda index and word index. Recently Win-

dows PC is used for the study of MIA. It would be very useful if we could produce the electronic book on Windows. New computer tool made us prepare this electronic book on Windows. We will discuss how to produce this book in detail.

キーワード: 中期インド・アリアン, アルダ・マガダ語, ジャイナ教, パーリ語, 初期仏教, 電子ブック

Key words: Middle Indo-Aryan, Ardha-Māgadhi, Jain, Pāli, early Buddhism, Electronic Book

**1. はじめに**

中期インド・アリアン諸語(Middle Indo-Aryan)で書かれた古文献は、仏教(パーリ語)とジャイナ教(アルダ・マガダ語)の聖典がその例に挙げられるように、世界の文化の貴重な宝庫であり、現代文明の重要な源流の一つである。初期仏教・ジャイナ教は厳格なカースト制度のバラモン教に対抗する形で同時期に発生したため、相互に関係を持ちながらその教理等を深化させてきた。それゆえ、今日では初期仏教を研究をする上でジャイナ教の研究は不可欠であり、その逆も真なりであると認識されている。しかしながら、これらの諸聖典は難解な言語で書かれていること、さらに質的・量的にも膨大

であることもあって、今後、厳密な研究成果が待たれる分野である。

このような状況下において、我々は中期インド・アリアン諸語に関する正確な知識を有する研究者と情報科学の研究者との密接な共同研究により、図1のような5つの計算機リソースを開発した。

- I. アルダ・マガダ語又はパーリ語のローマ字化システム構築（これをPali96フォントと呼ぶ）。
- II. ローマ字化テキスト作成：ジャイナ教の重要な5つのテキスト（Uttarajjhāyā, Āyāraṅga, Sūyagada, Dasaveyāliya, Isibhāsiyāim）及び仏教の4聖典（Dhammapada, Theragāthā, Therīgāthā, Suttanipāta）等をPali96フォントを使ってマッキントッシュに収録した。
- III. 聖典の韻律解析プログラム：電子化テキストファイルが韻律解析の入力として使用される。
- IV. 語彙の正順索引・逆順索引とパーダの正順索引・逆順索引の作成プログラム。
- V. 聖典の文法解析プログラム。

これらのツール等は、当該分野の研究において極めて重要な役割を果たす基礎資料を作成するに当たって、威力を発揮する。中期インド・アリアン諸語の文献の研究を飛躍的に進展させるためには、韻律解析・文法解析・語彙と構文論等の系統的な研究が不可欠である、と認識されている。テキストの韻律解析は批判的校訂本作成に不可欠であり、語彙の索引は翻訳に当たっての手助けとなる。また、語彙の逆順索引は文法構造解析に必要で、パーダ（詩脚）の正順または逆順索引はテキストの正しい読みの確立と並行詩脚を見いだすのに重要である。

上記のツールを使用して作成されたデータ

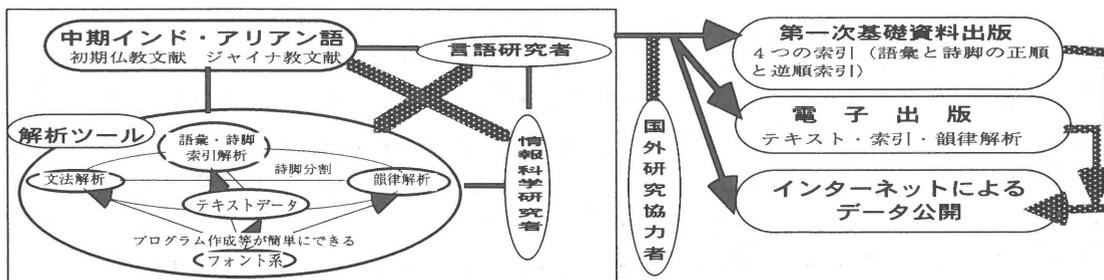


図1. 計算機リソースと研究組織

は、各出版社から単行本等して出版されたもの【1】と、ウェブ上で公開しているものがある（<http://www.sendai-ct.ac.jp/~ousaka>）【2】。ウェブ上で公開している計算機リソース（フォントシステム・電子ブック・プレインテキスト・逆順索引・特殊フォント用エディタ）を表1にまとめてある。

表に示しているように、マッキントッシュ版電子ブック（ジャイナ教聖典の5冊の電子ブック（Eブック））は既に作成済みである。この電子ブックには我々の作成した特殊フォントPali96を使用して作成された電子化テキスト・韻律図式と解析結果・詩脚の正順索引と逆順索引・語彙の正順索引と逆順索引が収録されている。最近のウィンドウズマシンの使用状況からみて、電子ブックをこのマシン上でも読むことができるならば、その有用性は大幅に上昇するであろう。

表1. 我々のウェブサイトアップロードされたデータ。表中の記号○、×は該当する計算機システムで使用可能なデータと使用不可のものを示している。

データ	マッキントッシュ	ウィンドウズ
特殊フォントPali96とキーボード入力法	○	○
ジャイナ教の5冊の聖典のコンピュータブック	○	●
ジャイナ教の5冊の聖典のプレインテキスト	○	○
初期仏教聖典の語彙の逆順索引	○	○
特殊フォントPali96用簡易エディターEDISANとその使用書	×	○

特殊フォントを使用しているために、従来のツールではウインドウズ用電子ブックを作成することができなかったが、最近不十分ながらもそのツール (Acrobat4.0) が提供されるようになった。表1の中で、●の部分、ウインドウズ上のジャイナ教聖典の5冊の電子ブック (Eブック)、つまり Portable Document File(PDF)は、このツールを用いて最近作成されたものである。本論文では、なぜこのウインドウズ用電子ブックを作成するのが遅れたのか、この作成に関連した問題点について議論する。

当該問題の理解には、既に作成しているマッキントッシュ版電子ブックとの比較が不可欠である。第2節では、この比較を主眼点にして、問題の原因となっている特殊フォント及び電子ブックに収納された内容について議論する。第3節ではウインドウズ版電子ブック作成及びその問題点について述べる。第4節では今後の課題等についてふれる。

## 2. マッキントッシュ版電子ブック

ここでは、最初に特殊フォント Pali96 の特性について述べ、次いでマッキントッシュ版電子ブックについて議論する。

### 2.1 フォントシステム, Pali96

計算機解析に当たって、各聖典の刊行本をそのままコンピュータ処理することが出来れば、それが最良の方法である。しかし、デーヴァナーガリー文字 (梵字、卒塔婆等に使われている文字) で書かれたテキストを、自動的に読み込むコンピュータツールが存在しないために、現時点ではこれは不可能である。デーヴァナーガリー文字テキストからローマ字化テキストへの変換には膨大なタイプ入力が必要とされ、多大な時間と経費が要求される。この作業中に、予期できないタイプミス等がローマ字化テキストに忍び込み易い。また、種々の索引を作成した場合、その校正には膨大な時間を要する。以上のことから、認識性の高い高品質の中期インド・アリアン諸語の為のローマ字化フォントを作成できるかどうか、その研究を遂行する上で重要な役割を果たすことになる。さらに、適切に構築されたフォント体系はテキスト解析プ

ログラムの作成を容易にする。これらの条件を全て満たすようなローマ字化フォント (表2参照) を作成した。この表には、通常のローマ字フォントに無い特殊文字 (例えば 'm' 等) のみを掲載している。特殊フォントは全て1バイトコードに割り当てており、拡張アスキーコード上に定義されている。

この拡張アスキーコード上に定義されたキャラクタの属性保持が、このフォントの取り扱いでは最も肝要である。マッキントッシュ上では、全てのワープロソフト等でこの属性を正確に保持しているが、ウインドウズシステムでは基本的にこの属性が保持されていない。このことが、原因でウインドウズ用電子ブックが作成されなかった。

ウインドウズシステムではワープロソフト等でも当然同様なことが発生している。このため、我々はマッキントッシュで作成したテキストを、ウインドウズでも使用できるようにするために、このフォント属性を保持できる簡易エディタ EDISAN を作成している。

表2には、ウインドウズマシンにおける特殊フォント Pali96 のアスキーコードへの割り当てと共に、キー操作法を示している。このフォント系は、基本的にはマッキントッシュの場合と殆ど同一の入力法を採用している。つまり、マッキントッシュの Option key を、ウイ

表2. ウインドウズPCにおけるパーリフォント Pali96 のキー操作法とアスキーコード割り当て。記号 <A>, <S> はそれぞれ Alt key, Shift key を示している。正規の文字は通常通りにタイプできる。特殊文字は Alt key, Shift key と通常のキーを組み合わせることにより入力できる。第3列と6列の数字はアスキーコードを表している。

ã	<A>+n	136	†	<A>+t	160
ā	<A>+a	140	‡	<S>+<A>+y	231
ī	<A>+<S>+`	147	d	<A>+d	182
i	<A>+i, i	148	dh	<A>+i, U	243
ū	<A>+e, u	156	ñ	<A>+n, n	150
ü	<A>+u, u	159	th	<A>+y	180
m	<A>+m	181	dh	<A>+z	189
kh	<A>+k	251	ph	<A>+p	185
gh	<A>+g	187	bh	<A>+b	186
ñ	<A>+g	169	l	<A>+l	194
ch	<A>+c	141	î	<A>+<S>+i	241
jh	<A>+j	198	ś	<A>+s	167
ñ	<A>+<S>+i	246	ş	<A>+x	197

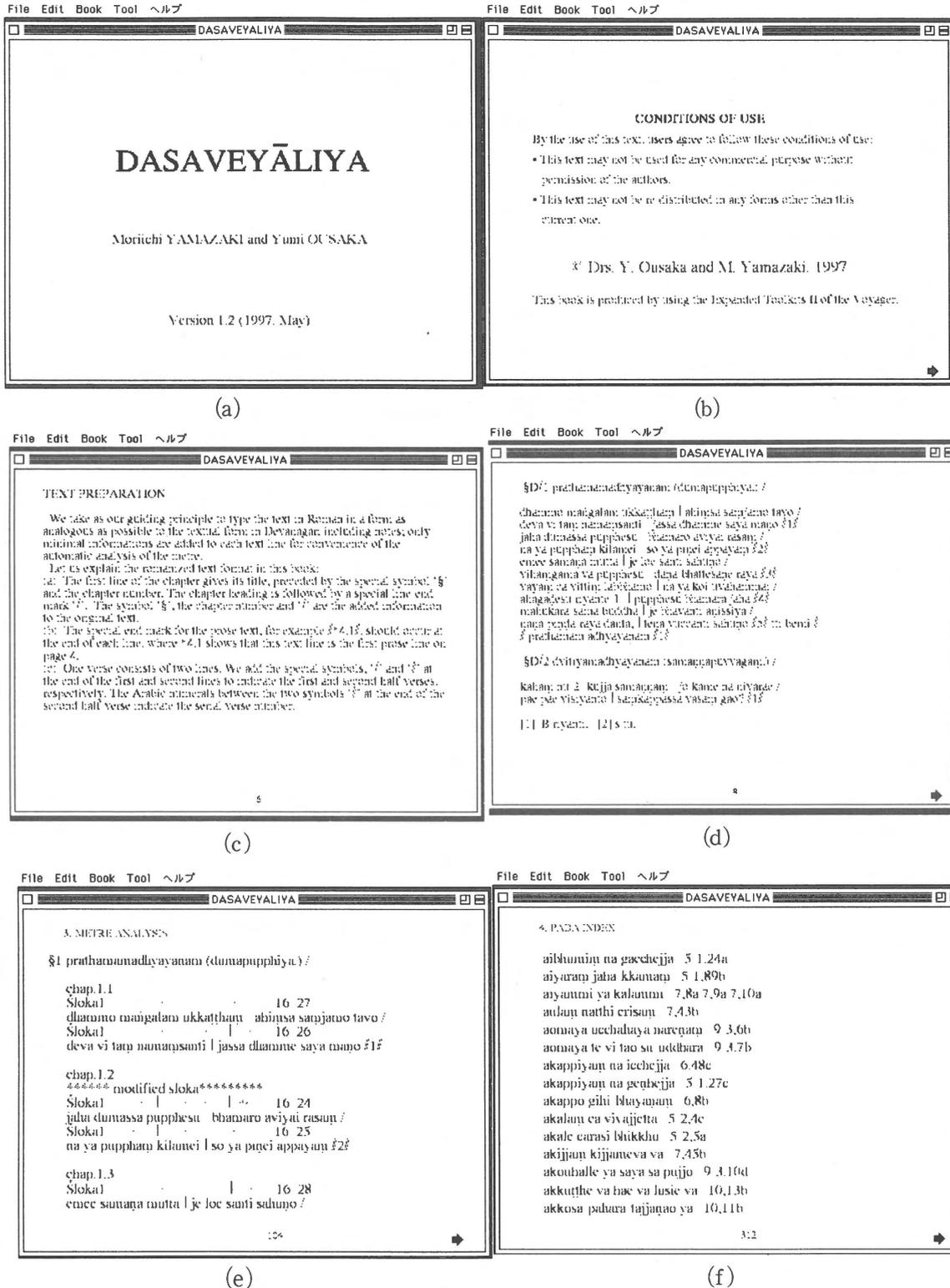


図2. マッキントッシュ版電子ブック

ンドーズマシンではAlt keyに置き換えているだけであり、Shift keyの使い方は全く同じである。また、アスキーコード割り当ては両方の計算機で同じにしている。このことにより、基

本的には、マッキントッシュで作成したテキストは、ウインドーズマシンでも同一コードで表現できることになる。実際、ウインドウズマシン用のジャイナ教プレインテキストは、マッキ

ントッシュ用テキストをウインドウズマシン用に変換して、作成している。(マッキントッシュとウインドウズマシン用のフォントは共にTrue Typeフォントの形式で作成され、我々のウェブサイトアップロードしてある。)

## 2.2 電子ブック

ジャイナ教の重要な5つのテキスト(Uttarajjhāyā, Āyāraṅga, Sūyagaḍa, Dasaveyāliya, Isibhāsiyāim)及び仏教の4聖典(Dhammapada, Theragāthā, Therīgāthā, Suttanipāta)等をPali96フォントを使ってマッキントッシュに収録した。仏教の4聖典の電子化テキストはパーリ文献協会の前会長ノーマン氏より提供されたものであり、ノーマンフォントで作成されている。このテキストを我々のフォントで書き換え、索引作成に活用できるようにした。著作権に抵触するため、電子化テキストは残念ながら公表することができない。

一方、ジャイナ教の聖典には著作権に関する制限がないので、電子化テキスト(図2(d)参照)のみならず、韻律解析結果(図2(e)参照)・詩脚の正順索引(図2(f)参照)と逆順索引・語彙の正順索引と逆順索引を全て電子ブックに収録した。この電子ブックは、Voyagerの日本語版Expanded Bookツールを使用して作成している。このツールでは、文字指定を細かくできる。この指定は煩雑ではあるが、図2(a)より分かるように、特殊フォントも正常に表示されていることが分かる。細かいことを言うと、このツールを使った検索機能のダイアログボックスに特殊フォントを表示できず文字化けが起こる欠点がある。これは、この機能に対して通常フォントのみが使用できるようになっている、ことによる。しかし、基本的には殆ど問題ないと言える。

言語研究上必要な基礎資料がテキスト資料と共に、1冊の電子ブックにまとめられているので、非常に有効な電子化ツールである。この電子ブックもフォントと同様、我々のウェブサイトから自由にダウンロードして使用できる。

## 3. ウインドウズ版電子ブック

ウインドウズ版電子ブックを作成するために

は、これまで議論してきたように特殊フォント属性を保持する処理がまず第一に必要となる。これは、我々の作成した簡易エディタEDISANにより実行される。最初にこのエディタについて述べ、その後Acrobat4.0を使用して作成されるPDFについて議論する。

### 3.1 特殊フォントPali96用簡易エディタ, EDISAN

これまで、我々はプラークリット語・パーリ語・サンスクリット語用のフォント”Pali96”をマッキントッシュ上で作成してきた。このフォントは、拡張ASCIIコード上で定義されたフォントである。これを元に、中期インド・アリアン聖典の解析ツールの構築等をマッキントッシュを使用して行ってきた。

これらの研究ツールをウインドウズマシンでも使用できるように拡張中した。最初に、特殊フォントをウインドウズ用True Type Fontに拡張し、その後このフォントを使って作成されたテキストを移植した。

日本語版ウインドウズOSを搭載しているパーソナルコンピュータ上で、適当なワードプロセッサを使用してこのテキストを読みると文字化けを起し、正確に読むことができないと言う厄介な事態が発生した。これは、2バイトコードフォントを標準にする日本語ワードプロセッサには、特殊フォントを表示するのに不都合な点があることに起因する。

このバグを回避できたら、印刷と画面上の表示が一致する、所謂WYSIWYG(what you see is what you get)表示のテキストを利用できることになり、言語学的な研究の能率も向上し、1バイトの特殊フォントを基に作成された大事な基礎資料を有効に活用することができ、さらに、語彙の索引作成等のプログラム開発も容易にできるという、多くの利点がある。なお、我々のフォントと同種の特殊フォントとしてはNorman Fontがあるが、これに対しても同様な状態になっている。

ウインドウズマシン上では特殊フォント(Pali96 or Norman)の簡便な入力法がなく(入力法がない文字もある)不便であるということから、3年ほど前に、特殊フォント用の簡便なテキストエディタ(EDISAN1.0)を初めて開発し

た。その後改良を重ね、現在はEDISAN3.0である。

EDISAN 開発用のツールは特殊フォントを全て正しく表示できる VisualC++ を使用している。C++ 上での文字コードは普通のアルファベットには ASCII コード、日本語には SJIS コードを使っており、2 バイト文字である日本語を表示させたいときは、文字コードの前後に SJIS コードであることを示すコードを書くことになっている。つまり、VisualC++ では 2 バイトコードが 1 バイトコードよりきちんと分離されており、127～255 番目の拡張 ASCII コードを含めて、0～255 番目のすべてのコード上の 1 バイト文字を正常に表示できるようになっている。この結果、拡張 ASCII コード上に定義されている特殊フォントを VisualC++ では正常に表示できる。

マイクロソフト社より、ある程度のワープロ機能を備えた Word Pad のプログラムが公開されている。我々はこの Word Pad に、特殊フォント入力機能等の必要機能を付加した。この方法により、Word Pad の通常の機能を活かした上で、簡易エディタを迅速に作成できた。

図3にエディタの構成画面を示している。通常のワープロソフトの画面とほぼおなじであり、使用方法も殆ど同じである。このエディタを使用して、リッチテキスト形式のマッキントッシュファイルを読み込むことにより、ウインドウズ用の特殊フォントの属性を固定したファイルを作成できる。このエディタも我々のウェブサイトに掲載してある。

### 3.2 ウインドウズ用 PDF 作成

ウインドウズ用にも Voyager より Expanded Book ツールが発売されている。マッキントッシュ版で作成された電子ブックが、このツールを使って見ることができるならば、ウインドウズ用にわざわざ作成し直す必要がない。しかしながら、このウインドウズ用ツールでは特殊フォントを使ってマッキントッシュで作成された電子ブックを開くことができなかった。これはウインドウズ用ツールでは、拡張アスキーコード上に展開された特殊フォントの取り扱いが正確に定義されていないことによる。

これに代わる PDF 作成ツールも又これまで提供されていない。今年になり、Adobe 社よりウインドウズ版及びマッキントッシュ版 Acrobat 4.0 が提供された。両版のツールを購入し、チェックしてみた

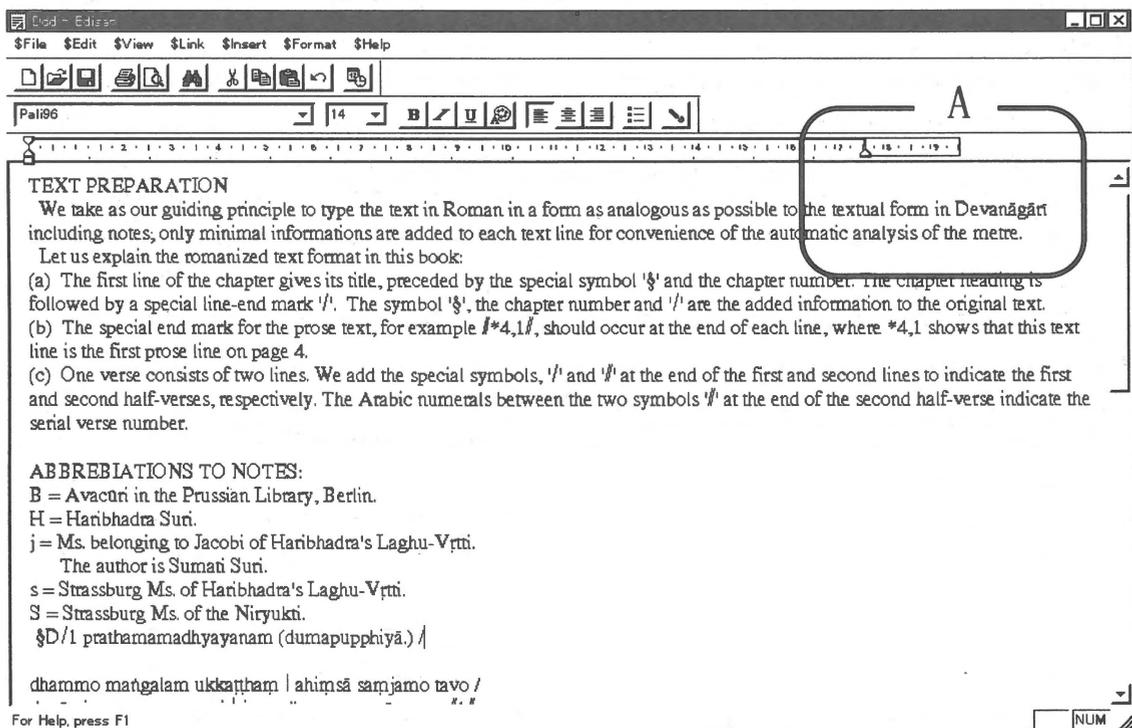


図3. 簡易エディタ EDISAN3.0

ところ、大きな違いがあった。

マッキントッシュ用では、特殊フォントで作成されたファイルのPDFを作成したところ、画像として取り扱われてしまった。その結果、PDFのメモリーは数十倍も大きくなってしまった（メモリーが数十倍になったことから、この現象に気が付いたのであるが）。同じ操作を、ウインドウズ版でも確かめてみた。この場合は、正常にPDFを作成でき、ファイルサイズは数十%程度縮小できた。その結果は、不思議にもバージョン4.0のマッキントッシュ版Acrobatでも正常に読むことができた。

以上のように同一バージョンでも、マッキントッシュとウインドウズ版では極端な違いが現れた。Adobeに問い合わせたところ、どちらのAcrobatとも、特殊フォントを正確に取り扱えるようには設計されておらず、ウインドウズ版ではたまたまよい結果になったのであり、動作保証はできない、とのことであった。なお、マッキントッシュ版Acrobatの以前のバージョンでは、少なくとも画像化されることはなかったのに、新バージョンでは画像化されるようになった。

その後、ウインドウズ版Acrobatで作成されたPDFの動作を何種類かのそれぞれのマシンで調べたところ、何ら異常が現れなかった。それゆえ、このファイルを作成し、ウェブ上にアップロードして公開することにした。

特殊フォントで書かれたウインドウズ版ファイルのPDF作成は以下のように行う。①電子ブック作成用データ(図2に示された全てのファイル)はマッキントッシュ上で作成されて

いるが、全ファイルをMS-Wordのリッチテキスト形式で保存する。この操作は、マッキントッシュ上で行う。②次に、ウインドウズマシンでの操作を行う。①で作成されたファイルを、簡易エディタEdisan3.0で読む。そのとき、フォントをPali96に指定しておく。その後、ファイルをそのままセーブする。この結果、拡張アスキーコード上のフォントの書式が正確に保持される。これは、前小節で議論したように、リッチテキスト形式のファイルを作成したことによる効果である。この後、適当なワードプロセッサを使ってPDF表示形式のファイル内容に整形して、PDFを作成すればよい。なお、この整形されたファイルでは各行末でのハイフネーションに注意する必要がある。このことについては次節で議論する。

図4はジャイナ教聖典 *Dasaveyāliya* のPDFを示している。図の左側のしおりより分かるように、電子化テキスト、韻律解析・詩脚索引・語彙索引等の必要な基礎資料を全て収録している。このしおりは目次の機能を果たしていると同時に、各項目へのジャンプ(リンク)機能をも兼ねている。この結果、このPDFは比較的使いやすいものとなっている。なお、詩脚の逆順索引表示においては、その性質から、詩脚部分の表示には後揃えタブが適し、ライン情報には前揃えタブが適している。図5から分かるように、このPDFではこのようなタブ揃えを実現している。マッキントッシュのExpanded Bookツールではその機能がないので、このようなタブ揃えは不可能である。又、図6は韻律解析結

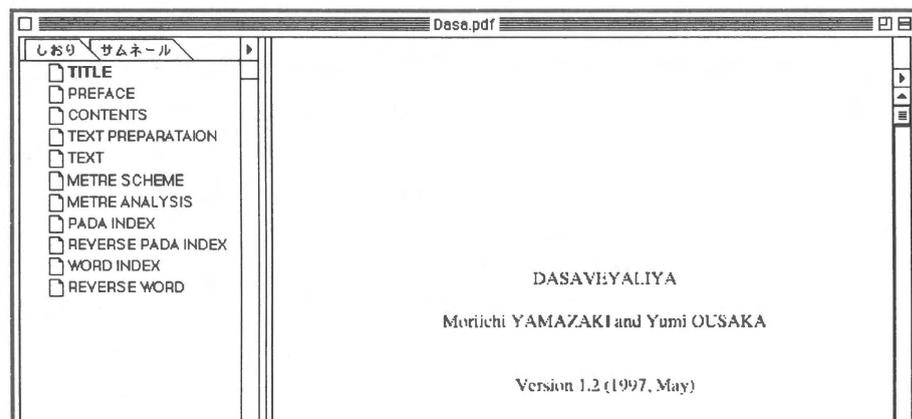


図4. ジャイナ教聖典のPDF例

果である。この例からも、特殊フォントが正常に表示されていることが分かるであろう。現在、ジャイナ教の重要な5つのテキスト(Uttarajjhāyā, Āyāraṅga, Sūyagaḍa, Dasaveyāliya, Isibhāsiyāim)のPDFを作成し、ウェブ上にアップロードしている。

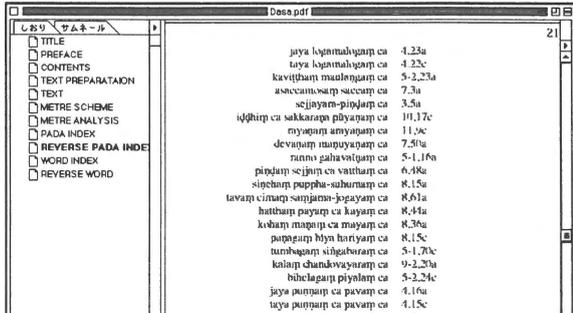


図5. PDFの詩脚索引例

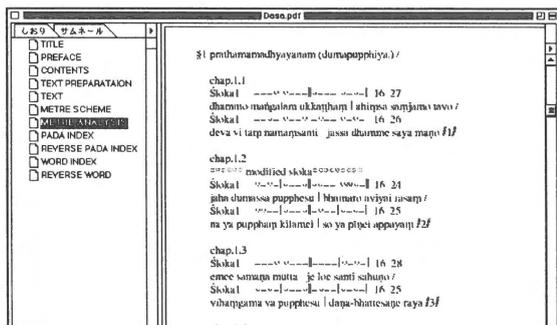


図6. PDFの韻律解析例

#### 4. 討論

前節の最後から2番目の段落で、PDF作成直前のファイル表示において、行末でのハイフネーションに注意を要すると指摘しているが、これは本質的に、簡易エディタEdisan3.0の機能的欠陥に起因すると推測される。この現象は、既に図3に明示されている。図3の、Aで示される部分から分かるように、水平ルーラーの右端よりも更に右側にテキストが表示されている。これはテキストの1行分を、表示可能画面の1行分におさめる機能が正確に動作していないことを示す。この結果、恐らく、単語としての認識も正確に処理されないであろう。今

後、この欠陥の解消について考察する。

#### 謝辞

この研究遂行に当たって適切な助言をしていただいた、山崎守一博士と宮尾正大教授(室蘭工大)に深く感謝いたします。この研究は2000年度文部省・科学研究費補助金、基盤研究(B)の助成の下に遂行したことを付記し、深謝致します。

#### 参考文献

- 【1】 M. Yamazaki, Y. Ousaka and M. Miyao, Indexes to the Dhammapada, Pali Text Society, Oxford 1995.
- Y. Ousaka, M. Yamazaki and K.R. Norman, Index to the Vinaya-piṭaka, Pali Text Society, Oxford, 1996.
- M. Yamazaki, Y. Ousaka, K.R. Norman and M. Cone, Index to the Digha-nikāya, Pali Text Society, Oxford, 1997.
- M. Yamazaki and Y. Ousaka, Theragātha: pāda index and reverse pāda index, Monograph Series 12 of Philologica Asiatica, Chūō Academic Research Institute, Tokyo 1997.
- M. Yamazaki and Y. Ousaka, Therigāthā: pāda index and reverse pāda index, Monograph Series 13 of Philologica Asiatica, Chūō Academic Research Institute, Tokyo 1998.
- M. Yamazaki and Y. Ousaka, Suttanipāta: pāda index and reverse pāda index, Monograph Series 14 of Philologica Asiatica, Chūō Academic Research Institute, Tokyo 1998.
- M. Yamazaki and Y. Ousaka, A Pāda Index and Reverse Pāda Index to Early Pāli Canonical Texts, Kosei Publishing Co., Tokyo 2000.
- 【2】 下記の索引はウェブ上に掲載してある (<http://www.sendai-ct.ac.jp/~ousaka>).
- Reverse word index to the Vinaya-pitaka, M. Yamazaki and Y. Ousaka (1997)
- Reverse word index to the Digha-nikaya, M. Yamazaki and Y. Ousaka (1997)
- Reverse word index to the Mahaniddesa, Y. Ousaka, L.S. Cousins and M. Yamazaki (1999)
- Reverse word index to early Pali canonical texts, Y. Ousaka and M. Yamazaki (2000)

# パスカルデータベースシステム (I)

## Pascal Database System (I)

白石 修二

Shuji SHIRAISHI

福岡大学 理学部 応用数学科

Department of Applied Mathematics, Faculty of Science, Fukuoka University

藤村 丞

Shou FUJIMURA

長崎大学 経済学部 総合経済学科

Faculty of Economics, Nagasaki University

**あらまし:** パスカルデータベースは、メナール編纂パスカル全集既刊本全4巻「ŒUVRES COMPLÈTES DE BLAISE PASCAL」(Desclée de Brouwer)の全テキストをデータベース化したものである。2000年10月からインターネットで試験公開中である。インターネットを介してのいくつかの語彙検索や頻度作成ができるようになっている。検索に関しては、ヒットした文を含む原著でのページ画像も表示できるようになっている。頻度表作成に関しては、ABC順、多寡順の選択ができ、逆引き辞書の作成も可能である。本稿では、パスカルデータベースシステムの概要と使用方法について述べる。

**Summary:** We experimentally release an online database on "ŒUVRES COMPLÈTES DE BLAISE PASCAL"(Desclée de Brouwer publishers) in 4 volumes edited by Jean Mesnard. This database can be accessed through Internet from all over the world, and you can use it for vocabulary searches, making frequency lists, etc. In this paper, we sketch the summary and the usage of Pascal Database System.

**キーワード:** パスカル, データベース, メナール

**Keywords:** Pascal, database, J. Mesnard

## 1 はじめに

1997年、文部省科学研究「人文科学とコンピュータ」の研究資金援助を受けて、当研究グループによるパスカルデータベース構築プロジェクトが本格化した。当初の予想ではテキストデータベース構築はさほどの問題なく進むだろうと考えていたが、そこまで簡単ではなかった。この完成に至るまでにはさまざまな問題が生じ、そのひとつひとつに解答を与えることで着実に構築をめざしてきた。1998年度特定領域研究「人文科学とコンピュータ」報告集 [6] を参照。研究着手から4年を経て、2000年3月にほぼ現在の形にまで完成させることができた。研究成果を多くの研究者に還元することを最終的な目的としているが、その試験的な公開を10月よりインターネットを介して行っている。このインターネット上での試験公開の目的はいくつかある。その一つは、利用者の利便性が最大限に図られているかどうかということの検証である。つまりこのシステムが利用者(研究者)の要求に、でき得る限り答えることができるかどうか、という能力を量ることを意味しているのだが、具体的にはそういう利用者の利便性を追及するために、インターフェースを使いやすくすることを心がけ、また同時に柔軟で強力な検索システムを与えるなど、いろいろな工夫を加えている。インターネット上で公開し、実際に多くの人に利用してもら

うことで、このシステムの利便性に対する配慮が実際上機能するかどうかを確実に把握することができると考える。試験公開の目的のほかの部分、このような実際の利用を通じて、データベースの記述上のミスがあれば、その発見に大きく貢献するだろうし、ひいてはデータベースの精度を高めることにもつながるものである。しかし、この公開に関して注意を払わねばならないことはこれにとどまらない。つまり、利用者の便宜を最大限に図ることを主眼におかねばならないし、おきたいが、利用者が自由自在にデータを操れるようにすることは、オリジナルデータが大量にコピーされる可能性を生み、データベース作成に関する知的財産所有権の侵害を図らずも許してしまう可能性を示唆している。このことに関して、利用者のためには無制限の公開が望ましいし、それこそがデータベース公開の本領だと思われるが、そのような無制限の公開は、デジタルで作成されたものの運命として簡単にコピーし得るという弱点を持ち合わせていることとあわせると大きなジレンマとなるのである。

そういうジレンマを少しでも解消するべく、利用者にはやさしく、しかし製作者の保護も考えたシステム製作を模索しており、現システムにおいては、情報ではなく、情報を引き出す方法において負荷を与えることで知的財産所有権の侵害への抑止策として考えている。

さらにネットワークの速度あるいはコンピュータの能力が充分かどうか、というハードの部分での効率を確かめるという意味がある。

システムは、現在、パソコン上で維持管理及び、保守がなされているが、利用状況及び、改善すべき点に関しては、データの蓄積に伴い追って報告の予定である。偉大なパスカルの多岐に亘る業績の全てを考えたとき、後世に残したい自然や建造物を指定し保存し維持しようという趣旨のもとで設定されている<世界遺産>のような観点をもって、卓抜したその業績と思想に対して、その全ての保存と維持に勤め、後世に残す義務があると考え。そこで、ここまでの完成をみたテキストデータベースを一步先に発展させ、これを中心に据えた<パスカル総合資料館> (パスカルデジタルアーカイブ) の構築を行うことがこのデータベースの次なる使命と考えている。更にこれに引き続いて、この<パスカル総合資料館>もインターネット上での公開を目指したいと考えている。このパスカル総合資料館の早期公開の実現のために、パスカル関係資料のあらゆるものに関し

ての大規模で組織的な収集を予定している。

関係資料収集以外にこのデジタルアーカイブ構築のための準備を平行して行うこととなるが、テキストデータベースと手書きの原稿、写本等の画像データベースを統合する仕組みの研究に着手するとともに、柔軟でかつ強力な検索システムも考案中である。

## 2 システム概要

2000年3月にメナール版パスカルデータベースシステムが完成。2000年10月からデータベースをインターネットで試験公開中である。アドレスは、<http://pascal.sci.fukuoka-u.ac.jp:8080/>である。

システムは、パソコン上にJavaとOracleを用いて構築している。データベースメンテナンスに関しては、アクセス (Microsoft Access) でリモート処理できるようにになっている。

### <システム構成>

コンピュータ本体:  
DELL PowerEdge 1300  
PentiumIII 500MHz  
メインメモリ 256MB

サーバ OS:  
Turbo Linux 4.2 ftp 版

Web サーバ:  
Apache.1.3.6.tar.gz

データベース:  
Oracle8 Workgroup Server for Linux R8.0.5

Javaに関しては、次の2つのパッケージを使用した。

Java 開発キット:  
jdk\_1.1.7-v3-glibc-x86.tar.gz  
(Java Development Kit)  
Java サブレット開発キット:  
jsdk2.0-solaris2-sparc.tar.z  
(Java Servlet Development Kit)

## 3 データベース

データベースは、既刊本全4巻の全文テキストデータベースと見開きの2ページを単位とした画像デー

データベースからなる。テキストデータベースの容量は凡そ 14.6MB で、画像データベースは 176MB である。

### 3.1 テキストデータベース

次は、パスカルデータベースの電子テキストサンプルである。電子テキストには文頭に、文の区切り語 # と同時に原著での位置が分かるようにボリューム番号 (巻数) とページ数、行数が記されている。多段組のページがあるので、列数も入れている。例えば、#2-0232-11-00-P は第 2 巻 232 頁の 11 行目を意味している。P はパスカル本人が書いた文を表す。00 は列数が 1 段組であることを表す。全集には、姉妹やパスカル周辺の人々の書簡集及び編者メナールによる註も含まれているので、ここで区別できるようになっている。したがって、検索時にパスカル本人が書いた文だけを対象にしたり、あるいは姉だけの書簡集に限定することができる。また、パラグラフの区切りとして、記号 @ を入れている。文中での \e\ 等は、フランス語の è を表している。詳しくは、4.5 変換対応表を参照。

#### 電子テキスト例

@#2-0231-38-00-D 4. C'est-\a'\-dire, d'apr\e\'s la d\e\'finition prcmie\'rf, concourantes. En d'autres termes. le point d'intersection des droites NO et PQ se trouve sur la droite MS: les c\o^t\e\'s

@#2-0232-01-00-P LEMME \!R2\

@#2-0232-02-00-P Si par la m\e^me droite passent plusieurs plans, qui soient coup\e\'s par un autre plan, toutes les lignes des sections de ces plans sont de m\e^me ordre avec la droite par laquelle passent lesdits plans.

@#2-0232-06-00-P \L\*\ Ces deux lemmes pos\e\'s, et quelques faciles cons\e\'quences d'iceux, nous d\e\'montrerons que, les m\e^mes choses \e\'tant pos\e\'es qu'au premier lemme, si par les points K, \!R5\, passe une quelconque section de c\o^ne qui coupe les droites MK, MV, SK, SV, es points O, N, Q, les droites

MS, NO, PQ, seront de m\e^me ordre.

#2-0232-11-00-P Cela sera un troisi\e^me lemme.

@#2-0232-12-00-P En suite de ces trois lemmes et de quelques cons\e\'quences d'iceux, nous donnerons des \E'\l\e\'ments coniques complets, \a\' savoir toutes les propri\e\'t\e\'s des diam\e^tres et c\o^t\e\'s droits\L2\, des tangentes, etc., la restitution du c\o^ne presque sur toutes les donn\e\'es, la description des sections de c\o^ne par points, etc.

### 3.2 画像データベース

既刊本全 4 巻を見開き 2 ページをひとつの単位として、全て画像データベース化した。検索された文章を原著で確認できるようになっている。原著に含まれる数学や物理関係の図等は、テキスト化していないので、当面このような処置をとっている。図 1 は、画像データの例である。閲覧時、コピー防止のため PDF 形式とした。PDF は Adobe Acrobat のファイル形式で、Portable Document Format の略である。

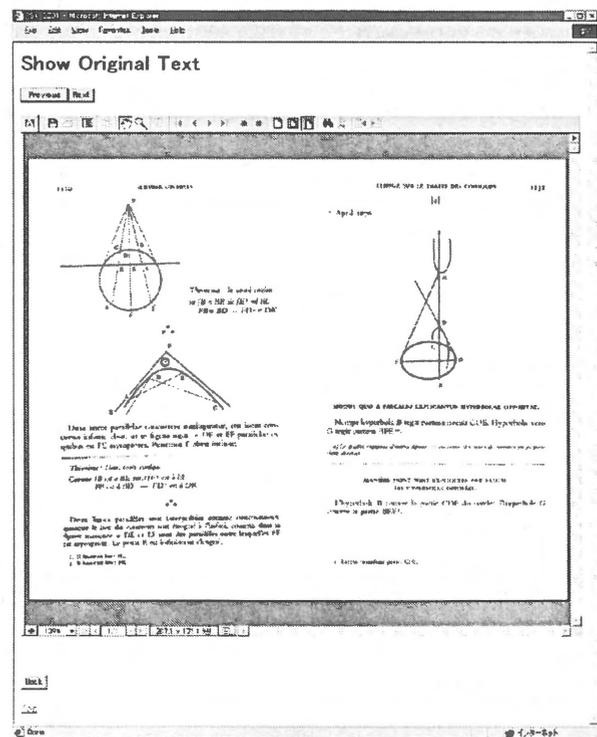


図 1: 画像データベース

## 4 データベース使用方法

### 4.1 試験公開要項とパスワードの取得

パスカルデータベースを使用するには、試験公開要項をよく読み、パスワードを入手する必要がある。試験公開要項は、現在、日本語版、韓国版、英語版、ドイツ語版が用意されている。順次他の言語での要項も準備中である。試験公開要項及びパスワードの入手方法に関しては、

<http://pascal.sci.fukuoka-u.ac.jp:8080/> にアクセスする。例として、英語とドイツ語の試験公開要項を上げている。図 2,3 を参照。

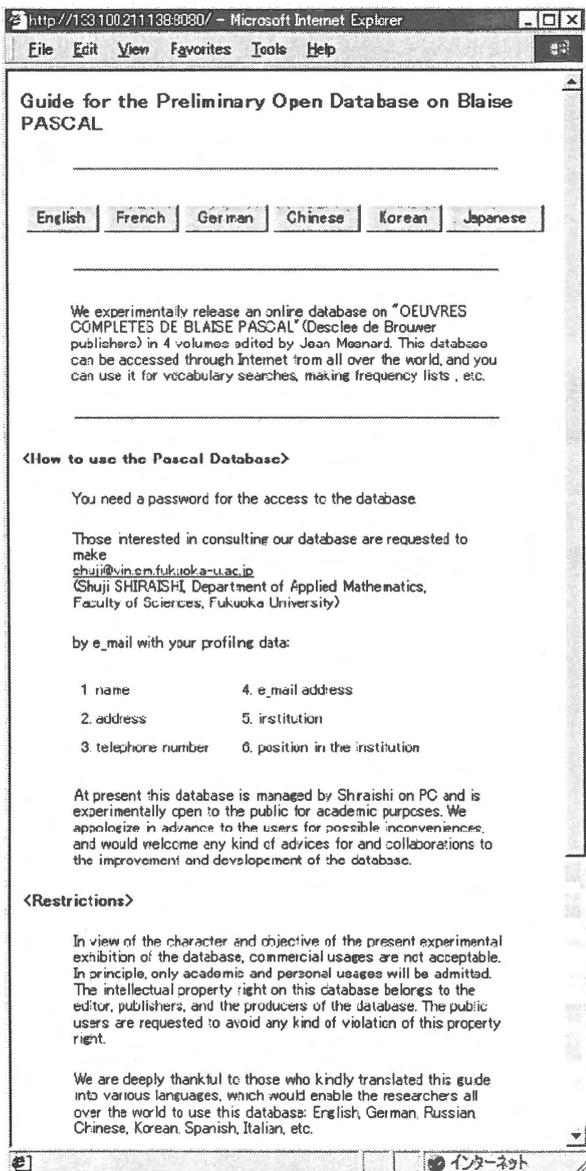


図 2: 試験公開要項 (英語)

試験公開要項には、利用要項、必要連絡事項、制限事項が簡潔に記されている。このデータベースの使用については、実験公開の目的からいっても、商業的な利用の不可、そして原則として個人の使用のみとしている。また、個人の資格としても真に学術的な目的をもつ利用者限定している。



図 3: 試験公開要項 (独語)

### 4.2 データベース接続

パスワードを取得したら、それを用いてデータベースに接続する。

ブラウザを起動して

<http://pascal.sci.fukuoka-u.ac.jp:8080/>

にアクセスする。そのページにある Go ボタンをクリックすると、パスワード入力画面 (図 4) になる。

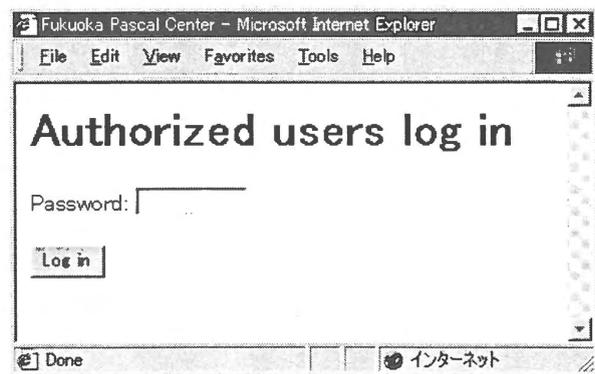


図 4: パスワード入力画面

“Password:” の欄に入手したパスワードを入力して、“Login” をクリックする。認証に成功すると、Search & Frequency 画面になる。(図 5) この画面

には3つのボタンがある。search と frequency, input\_character\_table である。search は語彙検索用ボタン、frequency は頻度作成のボタンである。input\_character\_table は変換対応表である。

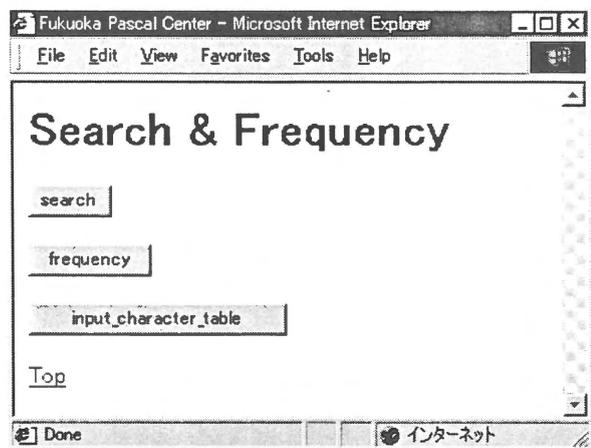


図 5: Search & Frequency 画面

### 4.3 語彙検索

検索には、語彙検索と目次検索の2種類を用意している。目次検索に関しては、作品ごとにアクセスすることができるようになっている。アクセスすると、作品の先頭の頁が表示され、頁の前後はボタンで捲れるようになっている。語彙検索を行うには、Search & Frequency 画面で“search”をクリックする。そうすると、Search 画面になる。

ここで、検索項目を指定する。識別 (P:パスカルが書いた部分等)、言語区分 (F:フランス語部分、L:ラテン語部分)、巻、頁、行、列検索文字列を入力して“Submit Query”をクリックする。

例えば図6では、パスカルがフランス語で書いた部分で、第2巻の最初から1400頁までの中で、4つの単語または文字列 je、ne、pas、et を含む文が何処にあるかを検索する (And 検索)。je の前後には空白を入れているので完全な単語として指定している。et の後には空白を入れているので、単語の接尾語になりうる。ne と pas に関しては、前後に空白を入れないので単語の部分列の可能性が充分にある。文字列を含む文にヒットしたら、その位置情報 (巻・頁・行など) がまず表示される。その位置データに関しても10個ずつ表示するようになっている。図7参照。

Next ボタンをいくつか押した後、位置情報 2-0523-

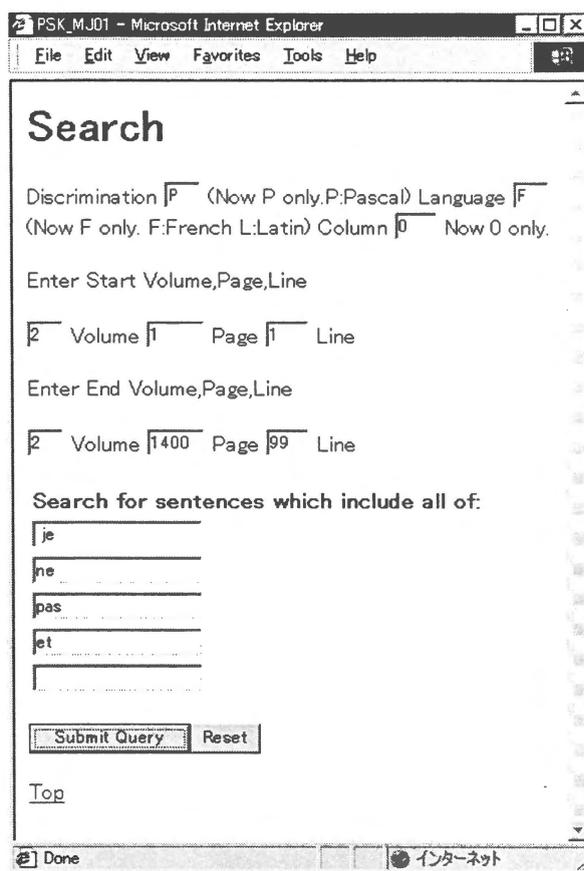


図 6: 語彙検索画面

16 を選択すると実際のフランス語の文が表示される。図8参照。

ここで、さらに Show\_Original\_Text ボタンを押すと、その頁の画像データが表示される。3.2 画像データベース参照。

### 4.4 頻度表

データベース作成における頻度表作成は、誤入力の発見に貢献し、データベースをより完璧なものに近づけるものである。と同時に希少語の発見などにみられるように、1次的な資料としての役割も担うので、大いに意義のあることとして、細心の注意を払って取り組んできたものである。

その成果の一つであるパスカル頻度表第2巻の1[5]は、パスカルが1640年初めから1654年の15年間に書いた論文、手紙など全ての作品を網羅したもので、もっとも華々しい活躍をした時期であると同時に宗教思想に目覚めた時期のものであり、内容は多岐にわたり、作品ごとの頻度には著しい変化がみら

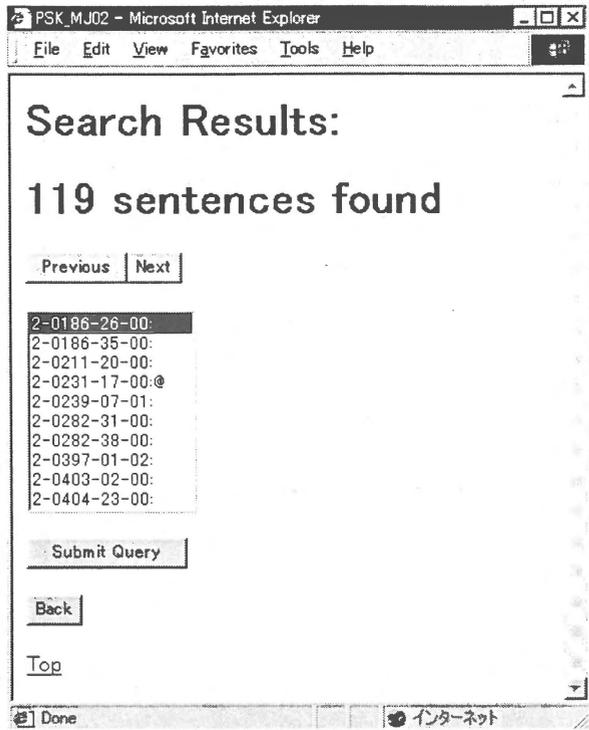


図 7: 検索結果一覧表示画面

れる。頻度表は、まず 2 巻中の収録作ごとに分けてアルファベット順、頻度順、逆引き順と作成し、また同時に、メナール版の特徴として、時系列的に作品を並べていることから、全体を併せて一つの作品と考え、この頻度表も作成した。これについても同じようにアルファベット順、頻度順、逆引き順と作成した。

頻度表を作成するには、Search & Frequency 画面(図 5)で“frequency”をクリックする。頻度表条件入力画面(図 9)が現れる。

ここで、次の頻度作成項目を指定する。

識別 (P:パスカルが書いた部分等)、言語区分 (F:フランス語部分、L:ラテン語部分)、検索範囲 (巻、頁、行、列)、頻度表を作成したい単語の範囲、検索結果の表示方法 (ABC 順、多寡順、逆引き順)

例えば図 9 は、パスカルがフランス語で書いた部分、検索範囲は、第 2 巻の 518 頁の 28 行から 527 頁までとしている。また、単語の範囲は、“p” から “r” までの間にある単語を対象にしている。

ABC 順 をチェックするとアルファベット順に頻度表が作成される。図 10 を参照。

また、図 9 で “Freq.Order” をチェックすると、図 11 のように多寡順に頻度表が、“Reverse dic” をチェックすると図 12 のように逆引き辞書が作られる。



図 8: 検索文表示画面

#### 4.5 変換対応表

パスカルデータベースにはアルファベット以外の記号は、変換対応表に基づいてデータベース化している。その変換対応表を見るには、Search & Frequency 画面(図 5)で“input.character.table”をクリックする。図 13 参照。

例

é → e'    à → a'    è → e'  
 ù → u'    â → a^    ê → e^  
 î → i^    ô → o^    û → u^  
 œ → oe    ç → c/

### 5 今後の課題 – パスカルデジタルアーカイブの構築に向けて

人類共通の財産として位置付けられた建造物・自然環境を保存するために、各国が協力し連携しようというのが「世界遺産」の主旨だが、人類共通の遺産は建造物や環境自然だけではない。同じく人類が誕生して以来、築き積み上げてきた「人類の叡智」も、われわれの大切な遺産である。具体的には各学問の知識、芸術、技術なども人類の歴史文化資産として残さねばならない「世界遺産」である。文字、記号、映像などによって表現されてきたこのような分野は、

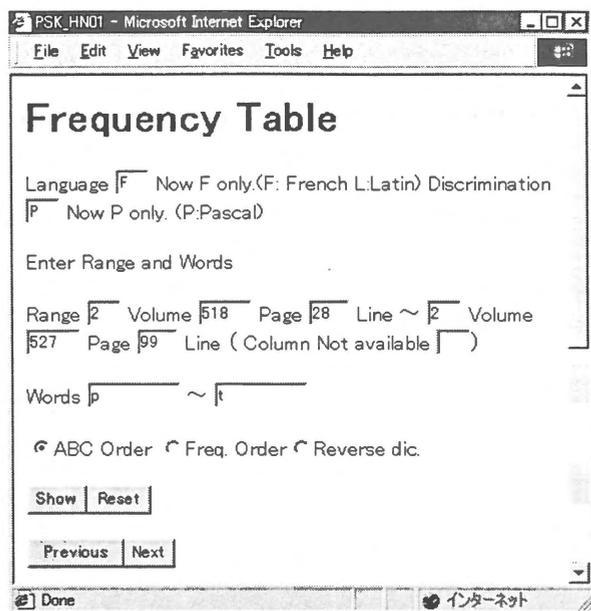


図 9: 頻度表条件入力画面

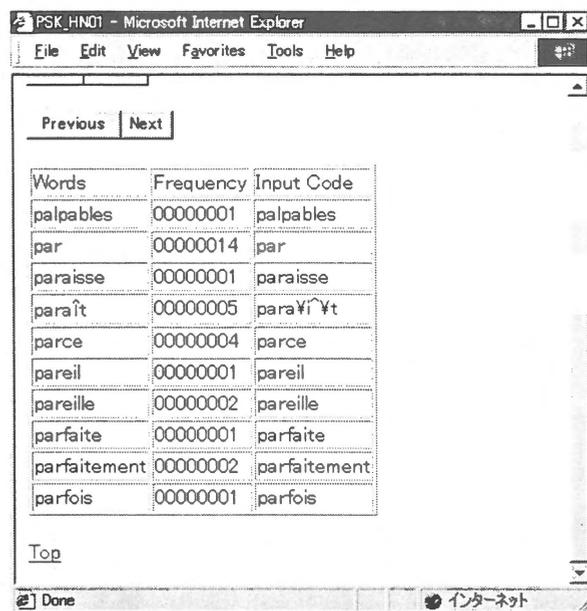


図 10: 頻度表表示画面 (ABC 順)

なまじそのような方法で保存されているために、保存方法の恒久性に付いての検討が後回しにされがちであったが、ここにきて従来の方法とは全く違う形でのデジタル技術を使用した保存方法によってこれら人類の叡智を後世へ継承しようという試みが生まれている。これが「デジタルアーカイブ」である。デジタルアーカイブの必要性についての認識は急速に高まってはいるが、例えばインターネット公開に向けてを前提にしても、一般利用者、資料利用関係者などからの利便性という視点が欠けているのが現状である。一般的には過去のマイクロフィルムによる保存のようにただ単に写し取るだけの作業に多かれ少なかれ終始して、デジタルアーカイブの縦横無尽な可能性についての研究はまったくもって不十分な上、そのために解決しなければならない問題も山積みの状況である。しかし当研究グループは、数年前よりメナール版パスカル全集の編者本人であるメナール氏の協力を得、国内外で唯一というパスカルデータベースのデジタルアーカイブの研究を先行しており、資料を実際に引き出す側の使い勝手にも配慮し構築を試みる研究をし続けている。また、インターネット上での公開に向けて、諸問題も想定し、これに解決を与えてきている。それに加えての第2段階として、このデータベースの充実にヴィジュアルの側面も盛り込む方法を試みている。具体的にはパスカルの直筆の文章、彼に関するあらゆる物品及び映像(多数のものが主にフランスに存在する)

などの画像処理をしなければならないもので、これを収録して初めてこのデータベースが完成に近づいたといえるもので今後はこのデータの収集を目的とするものである。今後の研究では、パスカルに関する資料を収集し、デジタル化し、組織化することにより、多角的、微細に分析、整理、研究することができる。

哲学者、数学者、科学者、宗教家と多彩な顔を持ち、そのいずれの分野においても超一級の仕事を成し遂げた天才パスカルの全生涯、全思想を体系的に網羅したデータベースは世界にも類を見ない規模のもので、これを明快な形で等しく全世界共通の遺産として位置付け、インターネット上の公開を考えている。

## 6 底本

*Œuvres complètes* de Blaise Pascal, tomes.1-4, texte établi, présenté et annoté par Jean Mesnard, DDB.

## 7 謝辞

本研究は、1997年度 重点領域研究「人文科学とコンピュータ」研究課題名 パスカル全データベース作成と言語解析、1998年度 特定領域研究「人文科学とコンピュータ」研究課題名 パスカル全データベース作成と言語解析、1998年度 研究公開促進費 学術

Words	Frequency	Input Code
que	00000130	que
qui	00000044	qui
si	00000030	si
pour	00000028	pour
pas	00000027	pas
peut	00000021	peut
qu'il	00000020	qu'il
plus	00000018	plus
qu'elle	00000017	qu'elle
par	00000014	par

図 11: 頻度表表示画面 (多寡順)

Display Char.	Input Code
Ü	Uo
U	Uv
U	U~
W	W~
Y	Y~
Y	Y'
Y	Y'
Z	Z'
Z	Z.
Z	Zv
a	a
ä	a"
á	a'
ã	a-
â	a^
à	a`

図 13: 変換対応表

Words	Frequency	Input Code
quand	00000010	quand
s'entend	00000002	s'entend
place	00000001	place
préjudice	00000001	pr`Ye`Yjudice
substance	00000003	substance
parce	00000004	parce
succède	00000001	succ`Ye`Yde
précède	00000001	pr`Ye`Yc`Ye`Yde
solide	00000002	solide
rende	00000001	rende

図 12: 頻度表表示画面 (逆引き)

- [2] パスカルデータベース作成と言語解析 (II), 1998 年 3 月. (情報処理学会「人文科学における数量的分析」にて共同発表)
- [3] パスカル全データベース作成と言語解析 (III), 情報処理学会, 特定領域「人文科学とコンピュータ」研究情報誌, 6 (1998) 49-55. (共著: 輪田裕, 藤村丞)
- [4] パスカル全データベース作成と言語解析 (IV) 1999 年 8 月. (第 4 回「言語・認識・表現」研究会 印刷中)
- [5] パスカル頻度表 第 2 巻の 1, 多賀出版, 1999 年 3 月. (共著: 輪田裕, 柴田勝征) 533 頁
- [6] 特定領域研究「人文科学とコンピュータ」1998 年度研究成果報告書 1999 年.

図書図書名 パスカル頻度表第 2 巻の 1、の成果に基づくものであることを付記する。

## 参考文献

- [1] パスカル全データベース作成と言語解析 (I), 情報処理学会, 情報研報, 36, 6 (1997) 31-36. (共著: 輪田裕, 藤村丞)

# 吾妻鏡データベースの構築 Construction of "Azumakagami-Database"

安道 百合子  
Yuriko ANDO

国文学研究資料館  
〒142-8585 東京都品川区豊町 1-16-10  
National Institute of Japanese Literature  
Yutakamachi 1-16-10, Shinagawa-ku, Tokyo 142-8585, Japan

## あらまし：

鎌倉幕府が編纂した漢文体の古記録である「吾妻鏡」のデータベースの構築過程とその結果を報告する。これは国文学研究資料館で行われている「原本テキストデータベース事業」の成果の一部である。このデータベースは、本文の領域（漢文体）・付訓情報を記した領域（カタカナ）・訓読文を示した領域（漢字とカタカナが混在する読み下した文体）・暦日情報などの注釈に関わる情報を記した領域、という四つのデータ領域を持つ。漢文という枠にとらわれずに、どの領域からも自在に検索できるデータベース構築を目指した。

## Summary:

"Azumakagami" is the history in a Kamakura era and be written in classical Chinese .

I will report the method for proceeding the database of "Azumakagami".

There are five fields of data in this database.

- (1)Original texts that was written in classical chinese character
- (2)Ruby (Phonetic characters written by Katakana)
- (3)Classical japanese that transrated classical Chinese ,it is written chinese character and Katakana
- (4)Notes such as calendar day information
- (5)Images of the original texts

This was made for the Japanese literature researcher. This is a part of the result of the project at the National Institute of Japanese literature.

**キーワード：**吾妻鏡, データベース, 古記録,

**Keywords :** azumakagami, database,

## 1. はじめに

国文学研究資料館の原本テキストデータベース事業は一つの古典文学作品を三年間かけてデータベースに仕上げ毎年刊行する計画のもとに行われている。一年目には業者による基礎入力作業、二年目には該当作品を専門とする複数の研究者による監修作業、三年目には資料館内でデータを統合、データベースとして仕上げるという総監修の作業をする。これは、毎年新たな作品を選んで作業にかかるという仕組みで、現在では一年間に三つの古典文学作品がいずれかの作業工程に置かれている状況である。したがって、資料館では、基本的に毎年新たな作品を総監修しており、昨年度は「吾妻鏡」であった。発表者はその総監修作業に関わった一人として、以下に吾妻鏡データベースの構築方法および結果を報告するものである。

古典文学研究は現在に伝わる諸本を読み解くことから研究がはじめられる。現在活字として流通している作品本文は、原本そのものの情報をそのままに伝えていくものではなく、読みやすさを考慮して表記を訂したり、他本と校合のうえより良いと思われる本文を採るといった校訂作業を経たものが多い。古典文学作品のデータとしては、底本そのものの固有の情報がまず求められるが、さらに、他の作品や同一作品で他の伝来系統を持つ本と比べようとするならば、表記をある程度統一させた情報が必要になってくる。つまり、古典文学作品のデータベースには、レベルを異にするいくつかの層の情報が求められるわけである。資料館のデータベース作成はそうした古典文学作品特有の問

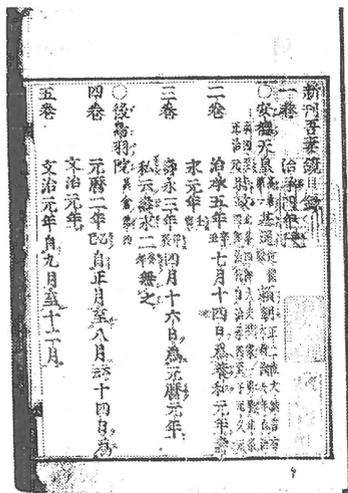
題点をできるだけ解決しようとしたものである。

なお資料館の原本テキストデータベース事業の成果は、古典作品のデータと、そのデータを検索利用するためのシステムを同梱して刊行している。システムは韻文作品用「和歌検索システム」と散文作品用「散文検索システム」の二種類で、「吾妻鏡」に関しては散文検索システムを用いて利用する。以下の本文においては、散文検索システムのことを「システム」、そのシステムに登録して利用する作品データのことを「マスターデータ」と称する。

## 2. 底本について

データベースの底本としたのは国文学研究資料館蔵寛永三年（1626）版の「吾妻鏡」25冊本である。「吾妻鏡」は鎌倉幕府が編んだ歴史書で、治承四年（1180）の頼朝挙兵に始まり、文永三年（1266）の六代將軍宗尊親王の辞任・帰京に終わる。中世以降、江戸時代まで広く武士の愛好書となっていたようで、寛永版本に大量の付訓情報があるのもその一端と言えよう。付訓情報は中世以降の読みをさぐる意味でも興味深い。

巻名はないが、年次を追って記述されているので、年月日が意味上の指標となる。物理的な位置情報は版本のV（冊）P（頁）L（行）で示す。完成したマスターデータは、テキストデータで約10メガの大きさである。



〈版本「吾妻鏡」第一冊 一丁表の画像〉

## 3. システムについて

まず、システムの構造について簡単に説明をしておきたい。散文作品は基本的に底本のVPLという位置情報で管理し、五層のデータ領域を自由に検索できるように設計している。五層とは①本文領域、②注記領域、③標準化領域、④メモ領域、⑤画像、である。資料館で仕上げるマスターデータはこの五層の領域を持ち、①②③④のどの領域からも一致文言・不一致文言を指定して検索することが可能で、検索結果を見る際にもその結果の位置に該当するすべての領域の情報を確認することができる。システムのブラウザ画面からは、作品全体を概観でき、利用者が必要な箇所の必要な領域のデータを見る（読む）ことも無論、自在である。②③④の各領域については、ブラウザ画面に編集ボタンがあり、利用者が検索をかけながら見つけた情報に対して、自分のルールでマーキングしていたり、メモ領域に情報をためたりして、記述されたデータのレベルを超えて活用することができる。①本文領域についても、マスター出力で、マスター全体をテキストデータとしてファイル化し、編集可能である。

マスターデータのフォーマットは公開しているので、どの領域に関しても利用者が自由に使用・作成することが可能になっているわけである。もちろん利用者が自分だけのマスターデータを作成してシステムに登録することもできる。コンピューター操作を不得手とする国文学者が、比較的簡単に、データベースの作成・利用・改変に関わることができることを目指して設計されたものである。

## 4. マスターデータの作成方法

### (1) マスターデータの構造

マスターデータの構造は次のようになっている。

¥A 作品名（データ冒頭に一つのみ）

¥T 和暦

¥M 月日情報

¥B 本文領域

V

P

L

¥C 標準化領域

V  
P  
L  
¥D 注記領域  
V  
P  
L  
¥E メモ領域  
V  
P  
L

¥M 次の月日情報

(以下次の¥Tが立つまで同様に繰り返す)

これは吾妻鏡に限らず、散文検索システムに登録できるマスタの基本構造である。作品に応じて¥T・¥Mの内容は自由に設定でき、¥Tは必須タグだが¥Mは任意である。

## (2) データ入力

初期入力には版本にある文字情報すべてと連辞符・返り点などの符合をすべて入力した。\*図1参照。

## (3) 漢文訓読文の作成

初期入力のデータは版本に忠実なデータであるが、版本に付されている情報すべてが、漢文を訓読するために必要にして十分な情報とは言えない。実際には、返り点の片方が落ちている場合があるため、返り点情報を抽出し、データの不備を補ったうえで、プログラム処理によって、訓読文を作成した。具体的には、返り点にそって漢字を並び替え、その漢字の送り仮名にあたる部分を補うという作業をしたわけである。\*図2参照。あわせて、他の二層の領域作成を行った。漢字だけを集めて本文領域を作成し、連辞符を熟語認定の手がかりとしつつ振り仮名にあたる部分を集めて注記領域を作成。版本のどの位置の情報に関してもすべて五層の構造を持つため、作業としては画像を除いた四層を並べた「四行ファイル」という形式を作成した。一行に与えられる四層の情報を並べて作業を進めることができるため、データの訂正をしやすい形である。

## (4) 四層のデータ領域の特徴

ここで各領域の特徴を簡単に示しておく。

①**本文領域** 底本を忠実に再現する領域。漢文の部分は、返り点・連辞符を抜いた白文の形で翻字する。漢字・仮名交じり文の部分(消息文など版本のごく一部に見える)は、そのまま翻字する。誤刻もそのままとする。漢字は原則として底本のままの字体。

②**注記領域** 吾妻鏡本文にある傍記を書き込む領域。漢字の振り仮名にあたる部分のかな(付訓)を原本のままおこす。音読符や漢字に濁点が振られている場合は歴史的仮名遣いで該当漢字の読みを記す。

③**標準化領域** 本文領域を訓読する領域。国史大系本「吾妻鏡」と大きく異なる本文は、その異同が同時に検索できる。

④**メモ領域** 校異や注釈等を書き込む領域。

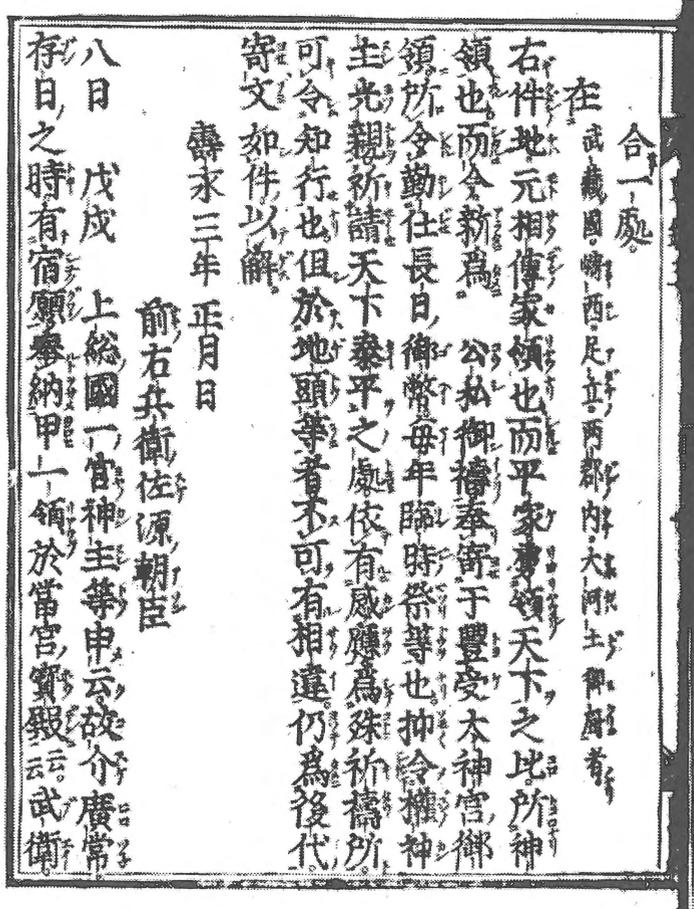
この段階では、④メモ領域には何も書き込まれていない。②注記領域と③標準化領域についても、形のうえでは出来ているが、その内容はまだ不十分なものである。なお「標準化」という名称は、古典文学作品全般を検索するために至便なように表記のゆれを最小限におさえるべく手を加えた領域を意味してここでは用いているが、なおその実際を言い得ているとは言い難く、用語そのものには問題を残していることを申し添えておく。

## (5) 監修員による読みおよびデータ訂正

吾妻鏡やその周辺を専門とする史学及び国文学の分野の先生方十一人を監修員として資料館が選び、それぞれの先生に四行ファイルの状態データの確認訂正をしていただいた。まずは本文・注記の二領域が底本を忠実に再現したものになっているかの確認を行う。特に漢字については、単純な入力ミスがあれば改め、入力不能な漢字を入力可能な漢字に置き換えることが可能かどうかといった判断を行う。プログラム処理によって作成した訓読文は版本の連辞符などが正確でない場合もあり、必ずしもこなれた訓読文とはなっていない。また版本の誤刻やあやまった付訓と思われる場合もある。そうした部分についてコメントを付すことや適切な読みを提示していただくこととした。また現在流通している書籍のなかではもっとも一般的に用いられている国史大系本との校異をとり、メモ領域にその異同を書き込み、標準化領域にそれを反映させるという作業も行う。

この監修員が目と手を使って確認する工程は、デー

\* 図1 初期入力の状態



\* 左図：版本画像

\* 下図：初期入力の状態

※初期入力は底本のとおりに翻字

漢字に附属する情報は

(振り仮名)

{送り仮名/返り点}

の書式でおこす。

漢字と漢字の間の連辞符は

真中にある場合：-

左端にある場合：=

の記号を入れる。

¥T 新刊吾妻鏡卷第三

V 0 3

P 0 0 1

<中略>

P 0 0 2

L 0 1 △△合 {セテ/} -- 處、

L 0 2 △在 {ノ/} <武=藏 {ノ/} 國、崎=西 (サキニシ)、足=立 (アダチ) {ノ/}、兩=郡 (ガン) {ノ/} 内 (ウチ)、大=河=土 (オホカハド) {ノ/} 御=厨 (ミクリ) {ニ/} 者 (ノ) {ナリ/}。>

L 0 3 右 (ギ) 件 (クダン) {ノ/} 地 (チ) {ハ/}、元 (モト) 相=傳 (ソウデン) {ノ/} 家=領 (カリヤウ) 也 (ナリ)。而 (シカルニ) 平=家 (ケ)、虜 (リヨ) {ノ/} 領 (リヤウ) {スルノ/} 天=下 {ヲノ/} 之比 (コロ)、所 (トコロ) {ナリ/} 神=

L 0 4 領 {トスルノ/} 也。而= (シカルニ) 今、新 (アラタ) {ニ/} 爲 (タメ) {ニ/} △公=私 (コウシ) 御= (ン) 禱 (イノリ) {ノ/}、奉 {リノレ} 寄 (ヨセ) {ノ/} 于豐=受 (トヨケ) 太=神=宮 {ノ/} 御=

L 0 5 領 {ニ/}。所 (ロ) {ノ/} 令 (シム) {ルノレ} 勤= (キン) {ノ/} 仕 (ジセ) 長=日 {ノ/} 御=幣 (ゴヘイ)。每 (マイ) 年臨=時 (リンジ) {ノ/} 祭=等 (マツリトウ) {ヲノ/} 也 (ナリ)。抑 (ソモト) 令 (シム) {シテ・ルノ下ノ} 權= (ゴン) {ノ/} 神=

L 0 6 主 (カンヌシ) 光=親 (ミツチカ) {ヲノ/}、祈= (キ) {ノ/} 請 (セイ) {セノ} 天=下泰= (タイ) 平 {ヲノ上ノ} 之 (ノ) 處 (トコロ) {ニ/}、依 {テノレ} 有 {ルニノ} 感=應 (カンヤウ) {ノ/}、爲 (シテ) {ノ} 殊 (コト) {ニ/} 祈=禱= (キタウ) 所 {トノ/}、

## 図2 漢文訓読文作成の過程

〈初期入力の状態〉

L05 領 |ニノ|。所 (ロ) |ノレ| 令 (シム) |ルノレ| 勤- (キン) |ノ二| 仕 (ジセ) 長-日 |ノノ| 御  
一幣 (ゴヘイ)。毎- (マイ) 年臨-時 (リンジ) |ノノ| 祭=等 (マツリトウ) |ヲノ| 也 (ナリ)。抑 (ソ  
モ++) 令 (・シム) |シテ・ルノ下二| 權= (ゴン) |ノノ| 神=  
L06 主 (カンヌシ) 光=親 (ミツチカ) |ヲノ|、祈- (キ) |ノ中| 請 (セイ) |セノ| 天-下泰- (タイ)  
平 |ヲノ上| 之- (ノ) 處 (トコロ) |ニノ|、依 |テノレ| 有 |ルニノ二| 感-應 (カンヲウ) |ノノ|、爲 (シ  
テ) |ノ二| 殊 (コト) |ニノ| 祈-禱- (キタウ) 所 |トノ|、

〈返り点の不備を●で示した状態〉

L05 領 |ニノ|。所 (ロ) |ノレ| 令 (シム) |ルノレ| 勤- (キン) |ノ二| 仕 (ジセ) 長-日 |ノノ| 御  
一幣 (ゴヘイ)。毎- (マイ) 年臨-時 (リンジ) |ノノ| 祭=等 (マツリトウ) |ヲノ| 也 (ナリ)。抑 (ソ  
モ++) 令 (シテ・シ) |・ルノ下●| 權= (ゴン) |ノノ| 神=  
L06 主 (カンヌシ) 光=親 (ミツチカ) |ヲノ●|、祈- (キ) |ノ中| 請 (セイ) |セノ| 天-下泰- (タイ)  
平 |ヲノ上| 之- (ノ) 處 (トコロ) |ニノ|、依 |テノレ| 有 |ルニノ二| 感-應 (カンヲウ) |ノノ|、爲 (シ  
テ) |ノ二| 殊 (コト) |ニノ| 祈-禱- (キタウ) 所 |トノ|、

〈返り点の不備を正した状態〉

L05 領 |ニノ|。所 (ロ) |ノレ| 令 (シム) |ルノレ| 勤- (キン) |ノ二| 仕 (ジセ) 長-日 |ノノ| 御  
一幣 (ゴヘイ)。毎- (マイ) 年臨-時 (リンジ) |ノノ| 祭=等 (マツリトウ) |ヲノ| 也 (ナリ)。抑 (ソ  
モ++) 令 (シテ・シ) |・ルノ下| 權= (ゴン) |ノノ| 神=  
L06 主 (カンヌシ) 光=親 (ミツチカ) |ヲノ|、祈- (キ) |ノ中| 請 (セイ) |セノ| 天-下泰- (タイ)  
平 |ヲノ上| 之- (ノ) 處 (トコロ) |ニノ|、依 |テノレ| 有 |ルニノ二| 感-應 (カンヲウ) |ノノ|、爲 (シ  
テ) |ノ二| 殊 (コト) |ニノ| 祈-禱- (キタウ) 所 |トノ|、

〈プログラム処理を行い訓読文にした状態〉

L05 領ニ寄奉り。長日ノ御幣。毎年臨時ノ祭等ヲ勤令ル所仕也。抑權ノ神  
L06 主光親ヲ、天下泰平ヲ祈請セ令・ル之處ニ、感應有ルニ依テ、殊ニ祈禱所ト爲、

〈監修員によって適切な訓読文に訂した状態〉

標準 L05 領ニ寄奉り、長日ノ御幣、毎年臨時ノ祭等ヲ勤仕セ令ムル所也。抑權神  
標準 L06 主光親ヲシテ、天下泰平ヲ祈請セシムルノ處ニ、感應有ルニ依テ、殊ニ祈禱所ト爲テ、

タの信頼性を高めるために必須の工程である。

## (6) 総監修作業

監修員の確認を終えたファイルを統合・訂正・総仕上げの作業である。

監修員の確認をおえたファイルはさまざまな見解が入って、データの内容そのものはより高度なものになっているが、同時に、見解の相違や不統一な箇所が多く混じるといふ難点も生じる。総監修の工程では、そのデータを内容的に統一のとれたものに仕上げ、さまざまな検索に耐えるように訂正していく。

標準化領域の統一化作業の実際は、項目の選択(何を統一させる必要があるか)・検索利用をしながらの検討(どのように統一させるのか適切か)・版本付訓のばらつきを検証などを繰り返して進める。また、異体字テーブルを作成し、この領域については、通行漢字に改めるという処理を行っている。これらの処理の一例は凡例に示している。

この段階で新たに追加したこととしては、次の二点である。

第一に、歴史書であるという作品の性質を最大限に生かすタグの構造が求められると判断し、暦日情報をタグ付けすることとした。初期入力段階では¥Tに巻数を入れてあるのみで、任意タグの¥Mは空状態であったが、ここで、¥Tに和暦情報を、¥Mに月日情報を入れた。それにより、システムのブラウザ画面の左に年月日が一覧できるようになったわけである。同時に暦日情報からの検索に耐えるべく、メモ領域の各日最初の行に暦日情報を一定の書式で書き入れた。システムは検索のみをするために開くのではなく、ブラウザ画面であたかも書籍をひもとくように、本文を縦覧できる設計となっているため、その最初の画面に暦日情報が見出しとして見えるかたちにすることで、縦覧のしやすさは格段にあがるものと考えられる。

第二に、底本にない情報の追加である。底本には、脱落部分が存するため、他本で補う必要がある。これも、歴史書であるという作品の性質上、底本にはなくても、現在伝わる「吾妻鏡」諸本の情報を集めた全体を検索する利用者の必要性に応じるためである。底本にない情報については、伝本のなかではもっとも情報量の多い「吉川本」を用いて、補うこととした。本文領域には「吉川本」の本文を引用し、標準化領域には独自に訓読文を作成した。注記領域は空となるが、メモ領域の情報は底本にある情報の部分とまったく同じ

体裁である。

## 5. 完成したマスタと検索の実例

完成したマスタデータの内容は次のようになっている。その一部をここにひく。

- ¥A 吾妻鏡
- ¥T 新刊吾妻鏡目録
- ¥M 小見出名なし
- ¥B 本文領域
- V 0 1
- P 0 0 1
- L 0 1 新刊吾妻鏡目録
- L 0 2 一卷△△治承四年〈庚子〉
- L 0 3 安徳天皇、〈高倉第一〉基通〈近衛殿〉頼朝〈正  
〈中略〉
- ¥C 標準領域
- V 0 1
- P 0 0 1
- L 0 1 新刊吾妻鏡ノ目録
- L 0 2 一卷△△治承四年〈庚子〉
- L 0 3 安徳天皇、〈高倉ノ第一。〉基通〈近衛殿。〉頼朝  
〈中略〉
- ¥D 注記領域
- V 0 1
- P 0 0 1
- L 0 1
- L 0 2 (6) セウ / (10) カノエ
- L 0 3 (1) アントク / (7) タカクラ / (12) モ  
〈中略〉
- ¥E メモ領域
- V 0 1
- P 0 0 1
- L 0 1
- L 0 2  
《中略》
- ¥T 序
- ¥M 小見出名なし
- ¥B 本文領域
- V 0 1
- P 0 1 4

—以下略—

付録1. 異体字テーブル

標準化領域については異体字処理をおこなった。以下の表の左側の字を右側の字に変換することによって、通行の字体に改め、検索の便を高めるためである。

亞 亜	惡 惡	蘆 芦	鯨 鯨
壓 圧	菴 庵	圍 囲	爲 為
醫 医	毓 育	弋 一	壹 壹
稻 稻	飲 飲	媵 淫	隲 隲
韵 韻	卵 卵	鬱 鬱	感 感
殿 殿	閨 閨	審 審	營 營
曳 曳	榮 榮	颯 頰	衛 衛

凡例の一部： 異体字テーブル（602字のうちの一部）

◆ 送り仮名の補い方について。

a) 送り仮名の補い方は次のようにする。

(連体詞・副詞等は多くカナ一字のみを送るのを基準とするが、  
版本の付訓の傾向も考慮して決めた)

兼テ(かねて)	尤モ(もつとも)	甚ダ(はなはだ)
太ダ(はなはだ)	悉ク(ことごとく)	忽チ(たちまち)
聊カ(いささか)	凡ソ(およそ)	但シ(ただし)
散テ(あへて)	剩ヘ(あまつさへ)	争カ(しかでか)
盡モ(いへども)	頗ル(すこぶる)	殆ド(ほとんど)
先ヅ(まつ)	若シ(もし)	就テ(ついて)
依テ(よつて)	因テ(よつて)	以テ(もつて)
周章テ(あはてて)	合セテ(あはせて)	并ニ(ならびに)
纒ニ(しまりに)	次ニ(ついでに)	
相ヒ互ニ(あひたがひに)	相ヒ~	
然シテ(然而・しかして)	然レドモ(然而・しかれども)	
而シテ	而ルニ	
云ク(いはく)	日ク(いはく)	同ク(おなじく)
其ノ(その)	此ノ(この)	是レ(これ)
向フ(むかふ)	違ハス(つかはず)	訪フ(とぶらふ)
行フ(おこなふ)	賜フ(たまふ)	賜ハル(たまはる)
御フ(たまふ)	奉ル(たてまつる)	奉ハル(うけたまはる)
訪ヌル(たづぬる)	生痛ル(いけどる)	勿カレ(なかれ)
甚シ(はなはだし)	知シ召ス(しるしめす)	聞シ食ス(きこしめす)

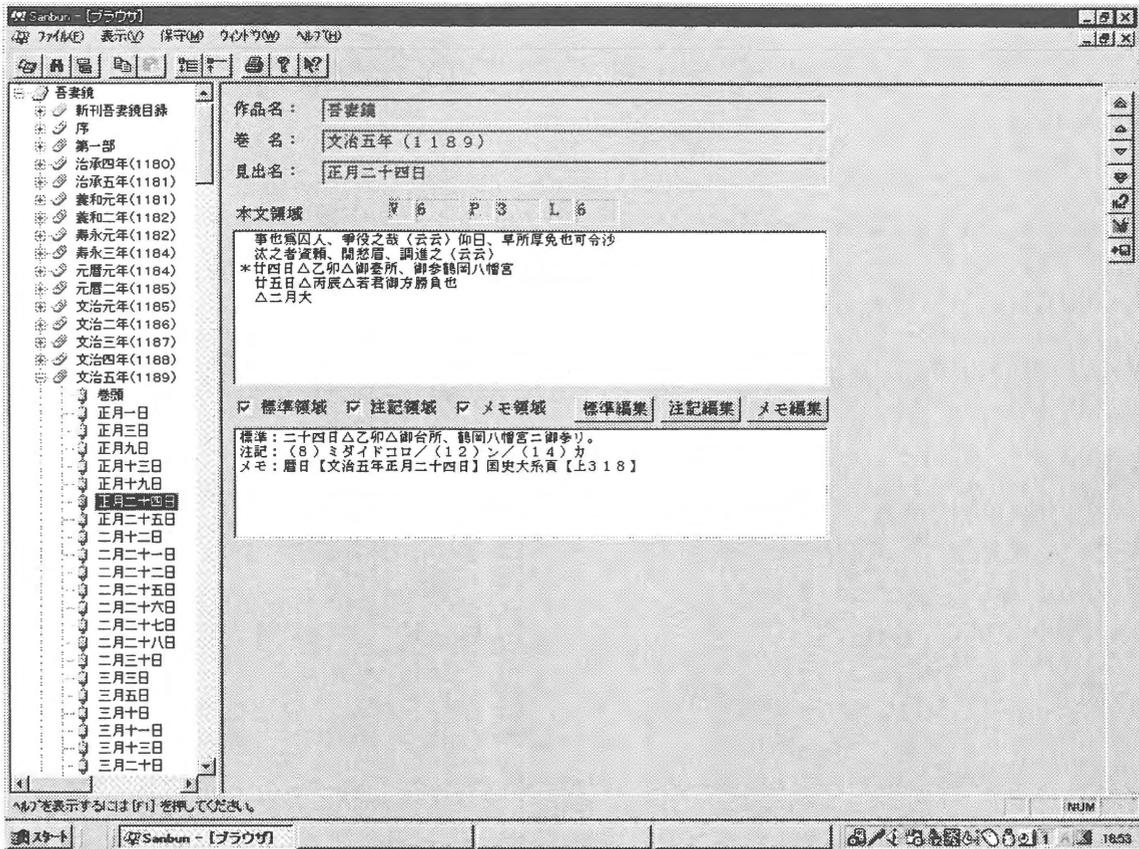
b) 次の語(接続詞の一部・名詞)にはカナを送らない。

又・亦・之(これ)	矣(かたがた)	使(つかひ)	次(つぎ)
據(よんどころ)	且(かつ)	故(ことさら)	繼(なほ)
兼(ともがら)	尚(なほ)	由(よし)	旨(むね)

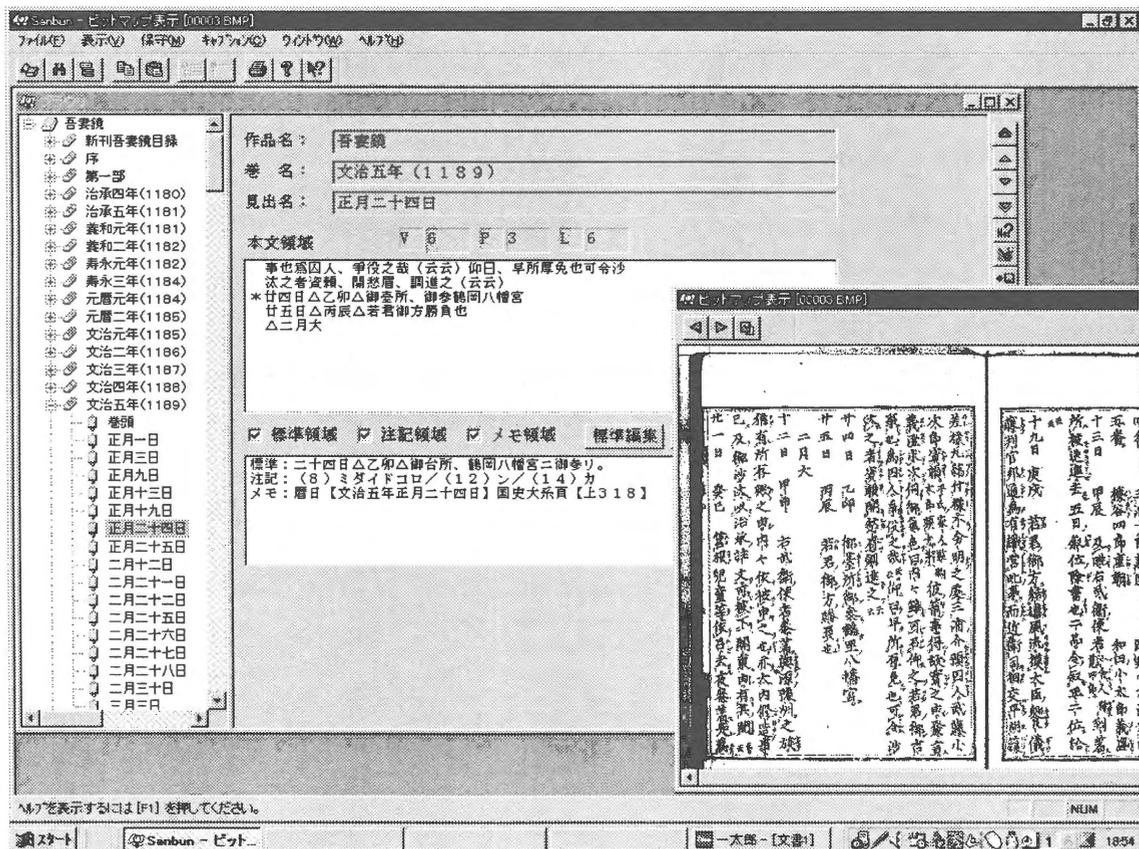
c) 一つの漢字が名詞としての読みと動詞としての読みを持つ場合は、  
名詞には送り仮名を添わず、動詞にはカナを添ふ

凡例の一部  
： 標準化  
領域におけ  
る送り仮名  
の補い方

左：検索条件入力画面 ・ 上：検索文言入力画面



〈図：ブラウザを開いた画面〉



〈図：本文に対応する画像を開いた画面〉



## 実際に検索を行う手順

検索の一例である。

- 1) 結果画面を開く
  - 2) 検索作品を選ぶ
  - 3) 検索領域を選ぶ
  - 4) 検索文言（一致文言・不一致文言）を入力
  - 5) 複合検索の方法を設定（本文 and 標準 など）
- \* 検索
- 6) 検索結果画面
  - 7) ヒット文言一覧から用例を選ぶ
  - 8) 本文表示（ブラウザ）

## 6. 本データベースの可能性と課題

このデータベースの独自の点は標準領域に読みやすい訓読文を示したことにある。単純に読みやすさからの利用者増は期待できる。また全文の訓読文がデータベースとして提示されたことは前例がないため、その完成度はともかく、初の試みとしてたき台にはなると思う。システムを使いこなして利用しようとする研究者にとっては、平家物語などの同時代の作品をマスターデータの形にして登録しておけば、本文の形態の異なる文献を同時に検索できるという点も魅力であろう。

課題としては、「標準化」の不徹底である。資料館内では、事業の発足当初から、原本に特化することなく、ひろく古典文学作品を検索するために必要な仮名遣い等の統一化の作業を経た情報の領域を「標準化領域」と呼んできた。既に刊行された「絵入源氏物語」のこの領域では、本文をすべて統一された歴史的仮名遣いによって平仮名にひらくこととしている。その際、濁音を清音に、音便形を音便のない形へそろえる処理も行っている。

「吾妻鏡」の場合は、本文が基本的に漢文体であるために、すべてを仮名にひらくことが必ずしも検索の便を高めることにはならないと判断した。とはいえ、現代の利用者の読みやすさ・成立時代の訓読法・版本の訓読法（付訓）のいずれを優先させるべきかという点が最後まで問題となり、いずれと確定もされなかった。基本的に版本の付訓は注記領域に持たせていることと、全頁の画像があることとで、補えると考えている。強いていえば、検索もれを最小にすることに注意をはらったため、版本の付訓をできるだけ生かしつつも、現在の研究者の読みになじむまいと判断されたも

のは訂正し、ある程度、用字・送り仮名・仮名遣いなどを統一させることにした。結果にいたるいちいちの統一化の検討については、当然反対意見もあろうが、その方針を凡例に掲げることで利用者には説明することとした。

最終的に現在出版されている「絵入源氏物語」の標準化とはずれを抱えているわけである。漢文体の作品の場合、五層の構造が不十分なのか、あるいはもっと別の形態があるのか、という点については、今後利用者の意見を聞きながら模索していきたい。

## 7. おわりに

本データベースは利用者参加型のデータベースである。書籍として出版されている古典本文には、校訂者の解釈が提示されるのが普通であるが、このデータベースはできるだけ、底本の情報を忠実に再現することにつとめている。画像はその最たるものであるが、全画像を公開することにより、利用する研究者は、国文学研究のどの段階からでも研究対象に向かうことが可能となっているわけである。日本語、とりわけ古典文学の検索にはできるだけ、解釈の入らない、けれども検索もれを最小限におさえるべく配慮されたデータが必要である。それをここでは標準化領域と呼んだわけだが、底本の情報を尊重しながら、一方で検索に備えるための措置である。利用者は利用しながら、自分の解釈を入れていったり、自分なりの注釈をつけていったり、自分だけのデータを作り上げることができる。版本（写本や古筆切の場合もある）を読むことから始めて、一冊の注釈書を作り上げるまでの、古典文学研究の流れを、どの段階からでもはじめられ、どのレベルまでも高めることができるというのが、国文学者にとってのこのデータベースの最大の魅力であると考えている。

## 謝辞

吾妻鏡データベースの構築は国文学研究資料館の事業の一環であり、本稿は研究情報部データベース室長中村康夫助教授の指導のもとに作成した。データベース監修員として関わってくださった方は、佐伯眞一・櫻井陽子・早川厚一・永村眞・大友一雄・田渕句美子・大津雄一・小川剛生・高山有紀・磯水絵・市川浩史・近藤成一諸氏である（敬称略）。記して御礼申し上げる。

# 全国遺跡データベースの構築

## Database of Archaeological Sites in Japan

森本 晋  
Susumu MORIMOTO

奈良国立文化財研究所情報資料室、奈良市  
Nara National Cultural Properties Research Institute  
Nara 630-8577

あらまし： 奈良国立文化財研究所では1987年度より全国遺跡データベースの構築を行っている。1996年度よりインターネット対応データベースを導入してデータの更新を続けており、1998年度からはインターネットを経由しその成果を公開している。

**Summary:** Nara National Cultural Properties Research Institute (ab. Nabunken) began to study about the database of archaeological sites in Japan in 1987. More than 400 thousand sites are existed in Japan, and Nabunken has registered about 230 thousand of sites in the database. Now one can get the data via the Internet home page of the Institute (<http://www.nabunken.go.jp>).

**キーワード:** 遺跡、データベース、考古学、位置情報、標準化

**Keyword:** site, database, archaeology, location, standard

### 1 遺跡データベースの必要性

遺跡に関する情報は考古学の研究の上でも埋蔵文化財行政の上でも必要不可欠なものである。考古学研究においては、遺構も遺物も遺跡とのつながりをもった情報としてとらえなくてはならず、各種の遺物に関するデータベースも遺跡情報を必要としている。

従来、遺跡の情報はある地域の遺跡一覧という地理的な範囲を限ってか、ある遺物についての出土地名表といった限定的な形でしか提示されていないのが普通である。全国規模での網羅的な遺跡データベースの作成は行われていなかった。

### 2 開発の経緯

文化庁では1987年度より全国文化財データベース構築計画の調査費が計上され調査検討が始まった。奈良国立文化財研究所（以下、奈文研）では全国不動産文化財データベースの一環として全国遺跡データベースの設計・構築を行っている。

不動産文化財とは土地と結びついていてそこから切り離すことのできない、遺跡（史跡を含む）、名勝、天然記念物、建造物（指定建造物を含む）、伝統的建造物群保存地区をさす用語で、美術工芸品、無形文化財、民俗文化財といった動産文化財と対応する。

全国遺跡データベースについては文化財情報システム全体の設計と合わせてその実現に向けた設計調査を行っていた。その一環として1990年2月27日には奈文研で全国遺跡データベース検討会を開催し全国の担当者から意見を求めた。1992年9月には奈文研の『埋蔵文化財ニュース』75号で「全国文化財データベースについて」という特集を組み、全国遺跡データベースの詳細を公表している。

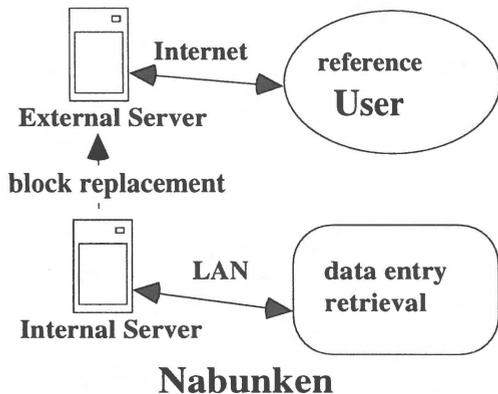
その後、こういった設計を実現するハードウェア、ソフトウェアの検討が続けられていたがシステムを実現するには相応の予算処置を待たなければならなかった。幸いにも1995年度末に補正予算によって奈文研の情報システムの一新が可能となったので、遺跡のデータとしてまず都道府県別の遺跡地図から文字情報の入力を1996年度より外部発注により開始した。

ところが当初運用を予定していたワークステーションのOSのバージョンに対応したデータベースソフトのバージョンアップの遅れからそのままでは公開の見通しがたたなくなった。

このため、別のワークステーションを用意し、全

文テキスト検索エンジンを利用したデータベースの一般公開を開始した。しかし、このシステムもアクセス時にエラーが発生することがあるなど安定せず、暫定的に全国のデータを手作業で市町村別に分割してリストの公表を合わせて行っていた。

1996年度の終りになってPCサーバを導入し、M言語を用いたデータベースを構築した。ハードウェア、ソフトウェアの更新はその後も行っているが、これが現行のシステムである。奈文研内部でデータベースの更新を行うためのシステムと外部にデータを公開するためのシステムの2本建てとなっている。内部用、外部用ともにフィールドは同一で、更新の日時により若干の新旧はあるものの内容もほぼ同一である。いずれもインターネット・イントラネット対応となっており、検索・更新はクライアントパソコンのブラウザから行う。ただし、外部用データベースに対しては通常、検索のみが可能である。現行システムでの一般公開は1999年9月28日より行っており、利用者は登録や課金なく自由に検索が可能となっている。



構築初期の遺跡地図データに加えて、各地の新規の遺跡地図や文化財地名表、古墳の集成といった各種遺跡地名表から情報収集を行い、レコードの追加や内容の更新を行っている。また、遺跡の発掘調査報告書に添えられている報告書抄録の内容によるデータベース更新を合わせて行っている。

### 3 内容

全国遺跡データベースの設計については、『埋蔵文化財ニュース』75号に詳しい。現在の知見からすると改良すべき点が多々あるが、それらについては後述することとする。

全国遺跡データベースの基本的な考え方は、奈文研が提供するのデータの核となる部分であって、

都道府県や市町村などの遺跡を調査する機関において必要なフィールドがあれば、それは各地での事情や需要に応じて独自項目としてローカルシステムを構築してリンクし利用するというものである。このためセンターシステムである全国遺跡データベースのフィールドは簡単なもの、地域的な偏在の少ないものとなっている。

研究者が必要とするような独自のフィールドについても個々の研究者においてローカルシステムとして構築してもらうのが最も効率的なやり方であると考えている。

以下、現行の全国遺跡データベースのフィールドについて解説する。このフィールド名などは『埋蔵文化財ニュース』75号段階ではなく、現在公開中のデータベースに従っている。

全国遺跡データベースは次の27フィールドで構成する。

- ID
- 市町村ID
- 種別
- 名称（漢字）
- 名称（かな）
- 所在地コード
- 所在地
- 境界文化財ID
- 所有者種別
- 主な時代
- 指定区分
- 調査・発掘の有無
- 文献の有無
- 地形図番号
- 緯度
- 経度
- 時代・遺跡種別
- 立地
- 現況
- 保存状況
- 面積
- 遺跡地図番号
- 群集遺跡ID
- 遺構概要

#### 遺物概要

#### 発掘概要

#### その他概要

これらのフィールドのうち、「ID」から「経度」までを共通項目群と呼ぶ。これらは全国遺跡データベースだけでなく、ほかの不動産文化財のデータベースにも共通するフィールドである。「時代・遺跡種別」以下のフィールドは全国遺跡データベースに固有のフィールドであり、遺跡項目群と呼んでいる。

#### ID

「所在地コード」と「市町村ID」を組み合わせたものの末尾に「種別」コードを加えて、遺跡のIDとする。12桁の数字として計画。

#### 市町村ID

遺跡の所在する市町村ごとに遺跡に付与されたユニークな番号。6桁の数字として計画。

#### 種別

不動産文化財の中での区分のためのフィールドで遺跡にはすべて「1」が与えられる。ちなみに名勝には「2」、天然記念物には「3」、建造物には「4」、伝統的建造物群には「5」、その他には「9」を与える計画である。

#### 名称（漢字）

遺跡の名称、ひとつの遺跡に対していろいろな呼び方がある場合は / で区切って並記する。

#### 名称（かな）

遺跡の名称の読み方をひらがなで記載する。ひとつの遺跡に対していろいろな呼び方がある場合は「名称（漢字）」と対応させて / で区切って並記する。この「名称（かな）」は実際に発音した場合の読みを記載するので、例えば「No.1遺跡」は「な ンぱーいちいせき」となる。

#### 所在地コード

自治省が定める全国地方公共団体コードの市町村コード（5桁のアラビア数字）を使用する。複数の市町村にまたがる遺跡の場合 / で区切って並記

する。

#### 所在地

都道府県名から記載する。町名変更などがあった場合、資料があれば最新のものに更新する。複数の市町村にまたがる遺跡の場合所在地コードの記載と対応させて / で区切って並記する。

#### 境界文化財ID

複数の市町村にまたがる遺跡の場合に、それぞれの市町村IDを記載する。

#### 所有者種別

遺跡地の主な土地所有者についてコードを記入する。所有者が国は「1」、都道府県は「2」、市町村は「3」、寺院は「4」、神社は「5」、法人は「6」、個人は「7」、その他は「9」を与える。

主な時代 コード

項目名	コード	
旧石器	10	
縄文	20	
弥生	30	
古墳	40	
古代	飛鳥白鳳	50
	奈良	51
	平安	52
	細分不明	69
中世	鎌倉	60
	南北朝	61
	室町	62
	戦国	63
	細分不明	69
近世	安土桃山	70
	江戸	71
	細分不明	79
	明治	80
不明	90	

#### 主な時代

遺跡の主な時代についてコードを記入する（前ページの表参照）。

#### 指定区分

遺跡に関しては、国指定特別史跡は「11」、国指定史跡は「12」、都道府県指定史跡は「13」、市町村指定史跡は「14」を入力。

#### 調査・発掘の有無

「0」をなし、「1」をありとして入力。詳細については「発掘概要」を参照することとする。

#### 文献の有無

「0」をなし、「1」をありとして入力。

#### 地形図番号

遺跡の範囲を含む国土地理院発行の25,000分の1地形図の地図番号。複数にまたがる場合は / で区切って列記する。

#### 緯度

遺跡位置代表点の緯度。度分秒で表した数値を単位記号を入れずに表記する。

#### 経度

遺跡位置代表点の経度。度分秒で表した数値を単位記号を入れずに表記する。

#### 時代・遺跡種別

時代と種別を別のフィールドとして設けただけでは混乱が生じるため、その組み合わせについてコード化したもの。次ページに表を掲げる。

組み合わせを用いるのは、例えば、複合した遺跡で縄文時代の集落と弥生時代の墓が発見されていた場合、時代が「縄文」「弥生」、種別が「集落」「墓」とだけデータ登録されていたのでは、縄文時代の墓も発見されているような誤解が生じるためである。

#### 立地

コードを入力。山地は「1」、丘陵は「2」、台地

「3」、扇状地「4」、低地平地「5」、水底「6」、それらに該当しない島については島嶼「7」を与える。

#### 現況

コードを入力。市街地・集落は「1」、道路は「2」、墓地は「3」、耕作地は「4」、山林・原野は「5」、河川は「6」、その他は「9」を与える。

#### 保存状況

コードを入力。遺跡全体が良く保存されている場合は保存として「1」、遺跡の一部が破壊されている場合は不良として「2」、遺跡全体が消滅している場合は消滅として「3」、不明は「4」、その他は「9」を入力。

#### 面積

遺跡の面積の実際の値を1平方メートル単位で記入。概数の場合も末尾の0を略さずに記述する。

#### 遺跡地図番号

文化庁編『全国遺跡地図』での遺跡地図番号と遺跡一連番号。この遺跡地図にないものについては空欄となる。

#### 群集遺跡ID

古墳群の場合などで、個々の遺跡のほかに群全体にも市町村IDが振られている場合それを記入する。

#### 遺構概要

遺構に関する概要を情報源とともに記入する。

#### 遺物概要

遺物に関する概要を情報源とともに記入する。

#### 発掘概要

発掘調査に関する概要を情報源とともに記入する。

#### その他概要

当該遺跡に関する記載のある文献名や、変更前の遺跡番号などを記入する。

時代・遺跡種別 (2010) コード

遺跡種別 時代		居 住 集 落							生 産 関 連					墓 ・ 祭 祀						そ の 他		
		集落	洞穴	貝塚	宮都	官衙	城館	交通	窯	田畑	製塩	製鉄	その他	墓	古墳	横穴	祭祀	経塚	社寺	集石	散布地	その他
旧石器		1001	1002	1003				1007					1012	1013			1016			1019	1020	1021
縄文		2001	2002	2003				2007			2010		2012	2013			2016			2019	2020	2021
弥生		3001	3002	3003				3007		3009	3010		3012	3013	3014		3016			3019	3020	3021
古墳		4001	4002	4003				4007	4008	4009	4010	4011	4012	4013	4014	4015	4016			4019	4020	4021
古 代	飛鳥白鳳	5001	5002	5003	5004	5005	5006	5007	5008	5009	5010	5011	5012	5013	5014	5015	5016	5017	5018	5019	5020	5021
	奈良	5101	5102	5103	5104	5105	5106	5107	5108	5109	5110	5111	5112	5113	5114	5115	5116	5117	5118	5119	5120	5121
	平安	5201	5202	5203	5204	5205	5206	5207	5208	5209	5210	5211	5212	5213			5216	5217	5218	5219	5220	5221
	細分不明	5901	5902	5903	5904	5905	5906	5907	5908	5905	5910	5911	5912	5913			5916	5917	5918	5919	5920	5921
中 世	鎌倉	6001	6002	6003	6004	6005	6006	6007	6008	6009	6010	6011	6012	6013			6016	6017	6018	6019	6020	6021
	南北朝	6101	6102	6103	6104	6105	6106	6107	6108	6109	6110	6111	6112	6113			6116	6117	6118	6119	6120	6121
	室町	6201	6202	6203	6204	6205	6206	6207	6208	6209	6210	6211	6212	6213			6216	6217	6218	6219	6220	6221
	戦国	6301	6302	6303	6304	6305	6306	6307	6308	6309	6310	6311	6312	6313			6316	6317	6318	6319	6320	6321
近 世	細分不明	6901	6902	6903	6904	6905	6906	6907	6908	6909	6910	6911	6912	6913			6916	6917	6918	6919	6920	6921
	安土桃山	7001	7002	7003	7004	7005	7006	7007	7008	7009	7010	7011	7012	7013			7016	7017	7018	7019	7020	7021
	江戸	7101	7102	7103	7104	7105	7106	7107	7108	7109	7110	7111	7112	7113			7116	7117	7118	7119	7120	7121
	細分不明	7901	7902	7903	7904	7905	7906	7907	7908	7909	7910	7911	7912	7913			7916	7917	7918	7919	7920	7921
明治		8001	8002	8003	8004	8005	8006	8007	8008	8009	8010	8011	8012	8013			8016	8017	8018	8019	8020	8021
不明		9001	9002	9003	9004	9005	9006	9007	9008	9009	9010	9011	9012	9013	9014	9015	9016	9017	9018	9019	9020	9021

## 4 検索

実際の利用はWWWのブラウザから行う。従って比較的新しいバージョンのブラウザが動作するインターネットにアクセスできるパソコンからであれば利用可能である。

データベースは奈文研のホームページ、<http://www.nabunken.go.jp>、からアクセスする。簡単な説明のページがあり、検索を選択すると、単純な構成の検索要求画面となる。検索は基本的に全文検索でフィールドを特定せずに行う。オプションとしてフィールドを指定した検索も可能である。

例えば「茶白山」を検索する場合はただ単に検索の言葉の欄に「茶白山」と入力して探せばよい。現在、全体で224260件の遺跡が登録されていて、「茶白山」でヒットするのは379件であり、1ページに16件ずつが表形式で表示される。これを一覧表示と呼んでいる。次ページ上段を参照。フィールドとして「名称(漢字)」に「茶白山」を含むものとして指定すると、360件がヒットする。

検索を繰り返すことで絞り込みを、1回の検索に際して検索語句の欄に複数の検索語句をスペースで区切って表記することによってOR検索を行うことが可能である。

それぞれのデータについて詳細な情報を得たい場合は、一覧表示の各列の先頭にある番号をクリックすることで、詳細表示を行う。次ページ下段参照。後に述べるようにフィールドによってはデータが入力されていないものがある。

## 5 問題点

現行のシステムはフィールド構成の設計から時間がたっていることもあり、種々の問題点がある。もともとシステム自体の問題ではなく、遺跡あるいは遺跡地図そのものが持つ性質による問題点も多い。まず、前者から検討する。

### 5.1 現行のフィールドに関する問題点

#### ID

下記のように適切な「市町村ID」を得ることができないために「ID」をキーとすることができないレコードがたくさん存在する。

#### 市町村ID

データベースの設計時には整数を想定していたが、実際の資料では、英字やハイフンあるいは枝番などが含まれていることもある。都道府県単位の番号だけで市町村としての遺跡番号がないところがある。

また、市町村合併により番号などは当然重複が出てくるがそれをどのように処置するかが決まっていない上、同一市町村内においても番号の変更がかなり頻繁なことがある。

以上のように現在入力に用いている資料からは同一市町村内においてすべての遺跡に付与されていかつユニークな数字としての「市町村ID」を得るには無理がある。

#### 名称(漢字)

現状ではこのフィールドが遺跡を弁別するためのキーとなっている。キーとするためには同一市町村には同一名称の遺跡はひとつしか存在してはいけないことになるが、完全にそうになっているわけではない。また市町村合併により同一名称となってしまう場合もあり得る。

名称を持たない遺跡も数多く遺跡地図には記載されておりそれらを適切に処理する必要がある。所在地名から○○所在遺跡といった仮称を与えるのも一案であるが、その際は仮称が本データベースによる独自ものであることを「その他概要」に示す必要がある。

遺跡名の変更も頻繁に起こるのでどれとどれを同一遺跡と判断するかむづかしい。これは遺跡範囲の変更とも連動している問題である。

名称を複数持つ遺跡も存在する。この場合は主たる名称がどれであるかを判断しなくてはならない。

#### 所在地

所在地の記載をどこまで詳しくするかは資料によって異なる。広範囲に渡る遺跡の場合、所在地をいくつか / で区切って並記することになる。地名表記の変更もかなり頻繁におこる。旧地名を「その他概要」に記載しなくてはならないこともあろう。地名だけが変更された時、その地名に基づいて命名された「名称」だけが残って「名称」の由来が不明になるためである。

#### 境界文化財ID

奈良国立文化財研究所 全国遺跡データベース

検索 0. [遺跡データベース] 224260件  
 絞り込み 1. [茶白山] 379件



RecNo.	ID	名称(漢字)	所在地	時代・遺跡種別
223707		下開発茶白山遺跡	石川県能美郡辰口町下開発/チャウスヤマ	2001/6901
223708		茶白山裏敷跡群	石川県能美郡辰口町徳久	9011
223833		茶白山城跡	山口県山口市大字黒川字河内	6906
224046		茶白山城跡	山口県防府市大字富海(茶白山)	6906
224105		茶白山城跡	山口県徳山市大字通字山田	6906
224512		茶白山城跡	山口県柳井市大字日禰字小園	6906
224547		前田茶白山遺跡	山口県下関市前田町1丁目	5905/5918
224576		鉢伏茶白山遺跡	石川県河北部宇ノ気町鉢伏	3001
225232		茶白山古墳群1号墳	三重県四日市市大字泊村字盆ノ井738-3	4014
228118		太田北遺跡(太田茶白山古墳・伝説(天皇陵外堤部外周))	大阪府茨木市太田3丁目202-1	4014/6920
228550		別所茶白山古墳/円筒寺茶白山古墳/宝泉寺茶白山古墳/宝泉村5号墳	群馬県太田市大字別所字茶白山	4014
228684		赤堀茶白山古墳/赤堀村200号墳	群馬県佐波郡赤堀町大字今井字薄島	4014
229306		毛生茶白山古墳/羽生田茶白山古墳	栃木県下都賀郡壬生町羽生田古敷	4014
229417		畑米茶白山古墳	栃木県佐野市畑米町茶白山	4014
233293		四天王寺旧境内・茶白山古墳	大阪府大阪市天王寺区四天王寺2-55-3、逢坂2-17-3・16・17	6918/7101
233512		北山茶白山西古墳/北山西古墳	群馬県富岡市南後園	4014

全体の内、16件が表示されています。↑↓ボタンでスクロールします。一覧表ボタンで一覧表を選択します。メニューボタンでデータベース選択に戻ります。一覧表のID番号をクリックすると、そのカードを表示します。

語句をスペース区切りで並べるとOR検索になります。  
 フィールド指定で、指定したフィールドのみを検索対象にできます。

言葉

奈良国立文化財研究所 全国遺跡データベース

検索 0. [遺跡データベース] 224260件  
 絞り込み 1. [茶白山] 379件

メニューボタンでデータベース選択に戻ります。↑↓ボタンでカードをめくりします。一覧表ボタンで一覧表に戻ります。



RecNo=#228232	
ID	
市町村ID	138
種別	1
名称(漢字)	茶白山古墳群4号墳
名称(かな)	ちゃうすやまこぶんぐんよんごうぶん
所在地コード	24202
所在地	三重県四日市市大字泊村字盆ノ井738-3
境界文化財ID	
所有者種別	
主な時代	H0
指定区分	
調査・発掘の有無	
文献の有無	
地形図番号	
緯度	345632
経度	1363546
時代・遺跡種別	4014
立地	
現況	
保存状況	
面積	
遺跡地図番号	
群集遺跡ID	
遺構概要	市報16(なし)
遺物概要	市報16(5世紀後半-土師器(甗・高坏・甗)・須恵器(ハソウ・二重ハソウ)・円筒埴輪(淡輪系)・形染埴輪(朝顔形・馬形・太刀形))
発掘概要	地中送電線新設工事
その他概要	市報16-1996.3

「市町村ID」を設定できない遺跡についてはこのフィールドも十分に機能できない。

#### 所有者種別

ハードウェアの高速化・大容量化によりコード化する必要性は薄れている。ある程度の範囲のある遺跡に対してその土地所有者を決定するのは判断基準の詳細な提示がないと不可能である。現在使用している素材からは所有者について十分な情報が得られないためにこのフィールドは空欄であることが多い。

#### 主な時代

資料の時代の提示の仕方がまちまちであり、記載が難しい。特に長い時代にわたる遺跡の場合どれが主たる時代であるかの判断材料がないことが多く、羅列した記載になる傾向がある。現状では複数の「主な時代」を時代順に表記しており、主な「主な時代」順になっているわけではない。

#### 指定区分

ハードウェアの高速化・大容量化によりコード化する必要性は薄れている。指定区分の記載については種類がごく限られているので、自然言語で記述しても問題は少ないと考えられる。

#### 調査・発掘の有無

設計ではこのフィールドの値としては有無を表す「0」か「1」がはいるだけである。このフィールドを設ける意義は少ない。

#### 文献の有無

設計ではこのフィールドの値としては有無を表す「0」か「1」がはいるだけである。このフィールドを設ける意義は少ない。設計では全国遺跡データベースと平行して不動産文化財文献情報データベースを構築することとなっていたが、実現していない。全国遺跡データベースと不動産文化財文献情報データベースとは「ID」でリンクを貼ることが計画されていたが、前記の理由で「ID」を有効に利用することができない状況にある。

#### 緯度

秒の部分については小数点以下の表記も認める。

#### 経度

秒の部分については小数点以下の表記も認める。

#### 時代・遺跡種別

遺跡について検索する場合このフィールドが一番キーとなる。従って、時代および種別についての区分や表記は常に最善のものでなくてはならない。

北海道や沖縄県方面での時代区分との対照も明示する必要がある。

#### 立地

ハードウェアの高速化・大容量化によりコード化する必要性は薄れている。広域遺跡において主な「立地」をどうやって認識するか、その区分とともに検討する必要がある。

#### 現況

ハードウェアの高速化・大容量化によりコード化する必要性は薄れている。広域遺跡において主な「現況」をどうやって認識するか、その区分とともに検討する必要がある。

#### 保存状況

ハードウェアの高速化・大容量化によりコード化する必要性は薄れている。「保存状況」についても前ふたつのフィールドと同様の問題が存在する。ほんの一部でも破壊されていたら「完存」ではなくなると「完存」の遺跡はなくなってしまうことにもなるので、区分については基準の提示が必要となる。

#### 面積

「面積」についての記載がある情報源が少ない。

#### 群集遺跡ID

「ID」がうまく機能していないことからこのフィールドも十分に機能していない。

#### 遺構概要

どこまで詳しく記述するかについて検討が必要。ただ、表記について標準化することは困難である。

# 地理情報システムを用いた城下町の復原的研究

—彦根城下善利組足輕屋敷地区を中心として—

## Restoration-Study on the old castle town Hikone with the GIS

生方 美菜子 濱崎 一志  
Minako UBUKATA Kazushi HAMAZAKI

滋賀県立大学大学院 人間文化科学研究科 地域文化学専攻  
〒522-8533 滋賀県彦根市八坂町 2500

Graduate Student of human cultures, The University of Shiga Prefecture  
2500 Hassaka-cho, Hikone, Shiga 522-8533

あらまし：都市には多くの情報が隠されている。例えば街区、宅地の形態・規模や、その場所に関わる情報などが現況の中に潜在している。これらの性格の異なる情報は文字情報としてとらえることもできるが、情報量が膨大なものとなるため地域の歴史研究をすすめていく上で、大きな困難と混乱が生じている。

そこで地理情報システムを用いてこれらの情報の有効活用ができないかと考えた。本研究は彦根城下町を対象とし、城下町景観の特徴を顕著に残す足輕屋敷の変遷やその現況を分析する。方法としては位置情報を表すデジタルマップに属性情報を表すデータベースをリンクさせる。さらに、測量精度の高い古絵図をラスターデータとしてはりつけることにより、歴史的都市の骨格の変遷をたどり、文字情報とリンクして新たな視点から都市の復原ができないかと考えた。今回は、測量精度、描画技術の高く城下町景観を詳細にあらわす「御城下惣絵図」をその背景にはり、現状及び異なる時代との比較の時空間分析を試みた。さらに景観要素を抽出し、デジタル化を検討する。本研究は、それらの要素のより適切な関連づけを、空間情報と時系列情報の両方面から目指し、地域文化財の保存を念頭におき、地理情報システムを用いた歴史的景観復原を試みた。

Outline: Many information is hidden in the city. For example, the form and the scale of a town-block and a housing site, the information in connection with the place, etc. are latent in present condition. When recommending history research of an area since the amount of information will become huge although the information from which such character differs can also be caught as character information, big difficulty and big confusion have arisen.

Then, it was thought whether effective practical use of these information could be performed using a Geographic

information system. This research analyzes changes and its present condition of the common foot soldier mansions which leave the feature of landscape notably for the Hikone castle town. The database showing attribute information is made to link to the digital map which expresses spatial information as a method. Furthermore, it was thought whether changes of the frame of a historical city would be followed, it would link with character information, and city restoration could be performed from a new viewpoint by sticking old map with high survey accuracy as raster data. "Gojoukasouezu" which survey-accuracy and drawing technology are high, and expresses a castle town landscape in detail was stuck on the background and the analysis between spatial and temporal was tried the comparison with the present condition and a different time. Furthermore, landscape elements is extracted and digitization is examined. This study tried the historical landscape restoration of series information using aim at both from a field and the Geographic information system bearing preservation of regional cultural assets in mind related attachment more suitable than those elements' at spatial information and temporal information.

キーワード：地理情報システム/地域文化財/城下町/古絵図  
Key words : GIS/Geographic Information System/Regional Cultural asset/Castle town/old map

## 1、はじめに

近年、地域文化財に焦点をおき、その地域の景観復原を試みるにあたり地理情報システム (GIS: Geographic Information System) が利用されはじめている。地理情報という視点から景観復原を考えていくと、それらの代表的な資料として、古絵図、紙地図、位置情報がテキストとして含まれた文献史料、航空写真等がある。これらの位置情報と歴史的情報を用いることによって、様々な時空間分析が可能となる。

しかし、それらの地理情報システムを用いて時空間分析を行っている現状は、手法を試行錯誤する段階である。古絵図のデータベース化による時系列分析、歪み補正や正図化による空間分析が行われているが、指標となるものは未だできていない。歴史的環境及び景観、地域文化財の情報を含むものをデジタルデータとして変換し、地理情報システム上で用いていくのがいかに困難であるかをあらわしている。

地理情報システム上で用いられている代表的な要素としては、ベースマップとしてのベクターデータとその参照のためのラスターデータ (画像データ)、そしてデータベースとしてテキストデータのリンクが考えられる。本研究は、それらの要素のより適切な関連づけを、空間情報と時系列情報の両方面から目指し、具体的に単一の地域を対象にし、地域文化財を念頭におき、地理情報システムを用いた景観復原への試みしたい。

## 2、研究の対象・目的

彦根は、城郭と城下町が揃って残っている。その城下の足軽屋敷地区は町人地の立花町・魚屋町とともにその面影を色濃く残す地区のひとつである。彦根市は将来的に足軽組屋敷を残していきたいと考え、また、現在そこに住んでおられる方々からも残していきたいという強い希望が出ている。特に、善利組足軽屋敷では保存会も設立されており、地域の人たちの足軽屋敷への強い愛着のある地区である。これまでには彦根藩足軽組屋敷の平面型についての研究 (1998 室谷) がなされている。

それらの保存という視点からそのさきがけとして、足軽屋敷の景観復原を地理情報システムを用いて試みるのが本研究の目的である。古絵図、文献史料に含まれるテキストデータのデータベース等の位置情報及び属性情報を地理情報システムのソフトウェア<AutoCAD Map 2000>上にとりこみ、現状と当時の状態を彦根城下・足軽屋敷中心に分析する。

## 3、方法

本研究においては、GIS ソフトウェア<Auto CAD Map 2000>を用い、ベースマップとしてのデジタルマップ、背景としての画像データ、テキストデータであるデータベースをとりこむ。基本となる彦根デジタルマップのオブジェクトにテキストデータであるデータベースをリンクさせ、建物と照合させる。同時に、背景として絵図・航空写真の画像データをレイヤー構造のなかに配置し、現状と当時との比較を行う。

### 3-1 ベースマップ

#### 3-1-1 ベクターデータ

##### ○彦根 デジタルマップ

地理情報システム利用という観点からベースマップとして不可欠なのがベクターデータである。その基礎資料として代表的なものが国土地理院発行の数値地図 2500 (空間データ基盤) である。しかし彦根市近郊はまだ発行されていない。そこで当研究室において平面直角座標系第VI系にのっとり紙地図からデジタルマップを作成した。入力に用いた紙地図は、彦根市現状平面図 (縮尺 1/500) 及び都市計画図 (縮尺 1/2500) である。従って、オブジェクトひとつひとつに座標情報が盛り込まれている。

主に閉じられたポリラインで構成している建物オブジェクトをトポロジー化しポリゴンとした。建物ごとに、地籍地番の情報を付し、それを<キー>とし、建物とデータベースをリンクさせる。地籍地番を用いる理由は、建物の位置情報が住居表示の番号と地籍地番がそれぞれ別個に併在しているところにある。また、地籍地番は規則性をもって配置され、今後変化する可能性が少ないと推測し、位置情報の基盤として採用した。ここでは「BLUE MAP 住居表示地番対象住宅地図」を用いて入力した。

#### 3-1-2 ラスターデータ (画像データ)

##### ○「御城下惣絵図」(天保七年 1836)

—彦根城博物館蔵—

6枚分割の絵図として伝わり、伝存する城下町絵図としては最も詳細なものとされる。方位や距離などは若干の誤差はあるものの、江戸時代後期の城下の形態をほぼ正確に伝えていると考えられる。善利組足軽屋敷地は約七百戸を数え、城下の足軽組屋敷のうちで最も戸数の多い地区であることを示している。

構成要素としては、屋敷地領域の位置情報とそれに付されている氏名・身分・間口サイズの文字情報を主として、道・河川・溝等が描かれている。

・部分画像データとして解像度 600dpi で街区ごとにスキヤニングしたものを貼り付ける。歪みの著しいものは伸縮させる機能（ラバーシーティング＝幾何補正）で調整する。これは表記されている氏名が読み取れる程度である。

・全体画像データとして、解像度 200dpi で足軽屋敷全体をスキヤニングしたものを貼り付ける。これは屋敷地割の輪郭を視覚で捉えられる程度である。

○航空写真（昭和五十七年 1982）

・部分画像データとして解像度 600dpi で足軽屋敷の部分をスキヤニングしたものを、背景として貼り付ける。これは建物の輪郭を視覚で捉えられる程度である。

・全体画像データとして解像度 200dpi で区画ごとにスキヤニングしたものを、背景として貼り付ける。

○航空写真（昭和三十六年 1961）

・解像度 300dpi でスキヤニングしたものを、背景として貼り付ける。

○航空写真・米軍撮影（昭和二十二年 1947）

・解像度 300dpi でスキヤニングしたものを、背景として貼り付ける。

### 3-2 テキストデータ

○彦根民家調査カルテデータベース

当大学人間文化学部生活文化学科土屋研究室が調査しデータベース化したものを用いた。主に昭和前半以前の民家を対象にしている。

・地籍地番を<キー>とし、彦根デジタルマップ及び県税務課作成の昭和 20 年以前建築の民家データベースとリンクさせる。

○家屋台帳・昭和 20 年以前建築の民家データベース

県税務課が、昭和 20 年以前に建築された建物をデータベース化したものである。所在地は地籍地番で表記されている。

・地籍地番を<キー>とし、彦根デジタルマップ及び彦根民家調査カルテデータベースとリンクさせる。

○「御城下惣絵図」（天保七年 1836）データベース

屋敷地領域に付されている氏名・身分・間口サイズの文字情報をデータベース化したものである。そのひとつひとつのレコードは屋敷地領域ポリゴンに付していく。

### 3-3 分析手法の検討

ここで二つの視点でもって、分析のより有効な手法を検討する。

#### 3-3-1 マクロ的

彦根城下町全体から足軽屋敷地区に向けて視点を運び、いかにも上空から彦根城下町を見ているかのように、城下町全体の景観を概観する。地形・街区・河川・堀割等の変化を分析する。

○彦根デジタルマップ

○「御城下惣絵図」（天保七年 1836）

○航空写真（昭和五十七年 1982）

○航空写真（昭和三十六年 1961）

○航空写真・米軍撮影（昭和二十二年 1947）

○家屋台帳・昭和 20 年以前建築の民家データベース

#### 3-3-2 ミクロ的

足軽屋敷地区から屋敷地及び民家等に向けて視点を運び、いかにもその屋敷群を歩いているかのように、足軽屋敷の町なみを概観する。データベースとリンクさせることによって、町なみの塀や生垣の連続面、道幅、溝、民家一軒一軒そしてそのディテールに及ぶまで分析する。

○彦根デジタルマップ

○「御城下惣絵図」（天保七年 1836）

○航空写真（昭和五十七年 1982）

○航空写真（昭和三十六年 1961）

○航空写真・米軍撮影（昭和二十二年 1947）

○彦根民家調査カルテデータベース

○家屋台帳・昭和 20 年以前建築の民家データベース

○「御城下惣絵図」（天保七年 1836）データベース

#### 3-3-3 分析対象の善利組足軽屋敷について

足軽居住区というのは 50 石以下の武士の居住していた地区で、江戸時代には簡単な塀、棟門のついた平屋の妻入り、ないし平入りの足軽屋敷が並んでいた。その足軽居住区において長屋の建築が多いのが特色である。今回はケーススタディとして足軽屋敷における長屋に焦点をおく。このデータは、当大学人間文化学部生活文化学科土屋研究室が調査したものである。

長屋は都市における借家人住宅で、外見上は町家とほぼ同じで道に面しており、異なるのは境壁を隣同士で共有して建っている点である。また、間口が町屋より狭いなど、建築物としての質は高くない。長屋の分

類として a 平屋、b 低長屋、c 高長屋がある。b は低二階長屋をしめし、c は高二階長屋を示し、隣と 2 軒にまたがって、入母屋破風のついたものもある。データ上では a 平屋、b 低町家、c 高町家と記している。長屋の経年変化をみると、明治元年では圧倒的に低町家が多く建設されているものの、それ以降はほとんど建てられていない。逆に、高町家は明治後期頃から徐々に建て始められ、大正後期には大半は高町家になっている。この傾向は町家と同様で使い勝手のよい高町家タイプの長屋が受け入れられていたことが考えられる。足軽居住区は武士居住区と同じような変遷がみられ、江戸時代からの足軽屋敷は多く残ったものの、明治に入ってから町家や塀付 2 階建が主流となっている。

このほかに間口サイズにも顕著な傾向が見られる。

このように、彦根民家調査カルテデータベースの各項目から、身分的居住区、建築種別、建築年代、表構え等の分類によって足軽屋敷の町なみの変遷、現状における残存状態をつかむことができる。

### 3-4 「御城下惣絵図」デジタル化の試み

ここで用いた「御城下惣絵図」は絵図をスキャンした画像データ、いわゆるラスターデータである。彦根デジタルマップは建物の位置情報の盛り込まれたデジタルデータで構成されている。今回ミクロ的にみた場合、ラスターデータの状態では「御城下惣絵図」はデータベースとリンクできない。さらに、絵図は屋敷地領域を示しており、現況との比較は建物の重なりを視覚的でしか判断できない。しかし、ラスターデータには重要な景観要素が含まれているので、それらをいかに抽出・認識するか、言い換えてみれば、ラスターデータの景観要素をいかにベクター化するかが重要な問題である。方法としては、認識・抽出の判断基準を定め、ラスターデータをトレースし、ベクターデータを作成することである。今回は屋敷地領域と道、河川、水路、溝等をトレースしベクターデータとする。そして「御城下惣絵図」の主構成要素である氏名・身分・間口サイズ等をデータベース化したものを位置情報でリンクさせる。

## 4 考察

彦根城下町形成後において「御城下惣絵図」という精度の比較的高い絵図を背景にして、彦根デジタルマップ、航空写真を貼り時系列で比較した。絵図に座標値は表記されていないため、城下町当時の地点と

変化がないと予想のつく地点を対照点として貼りつける方法がある。この方法はその地点の一致が大前提となるが、その同定ができれば、絵図の貼り込みは非常に有効になる。あまりにも歪みが著しい場合上記したとおり伸縮させる機能（ラバーシーティング=幾何補正）で調整するが、今回マクロ的にみる場合にその必要性はなく、2点を対照させ縮小拡大編集のみで済ませることができた。

上記の二つの分析手法から以下のことが言えるであろう。街区及び道幅はほとんど変化せず現在にいたると考えられる。また、「御城下惣絵図」のひとつひとつの屋敷地領域をみていくと、当時の屋敷地配列を推測することができる。建物の変遷はわからないが、現況と比較して屋敷地の配置にそれほどの変化がないことが推測できる。このようにラスターデータという画像としての情報のままに、時系列で照合できる絵図の貼りこみは有効であると考えられる。

また今回は「御城下惣絵図」のベクターデータ化を試みた。上記のラスターデータの貼り込みから得られる情報以上のものが期待できる。それは絵図の屋敷地領域ポリゴンと建物ポリゴンのオーバーレイにより、絵図から作成したテキストデータベースの構成要素の氏名・間口サイズの比較ができることである。

しかし、より有効かつ適当に利用するには課題は多い。文字情報及び図や絵をどのように分析させていくかが、それは要素ごとにデータベース化することが適当であろう。どのような項目を設定したらいいのか等いかにデータベース化するか、が課題である。

今回は主に「御城下惣絵図」を用いて、景観要素の抽出とベクターデータ化、文字情報のテキストデータ化、そしてそれらのより適切なリンクを目指し、現況の比較を行い、手法の一つの試みとした。しかし、そのより有効かつ省力的な方法を検討して必要がある。例えば、今回は、「御城下惣絵図」で試みたが、彦根城下町に関する絵図の一つである「彦根藩侍屋敷之図」（文政十二年 1829）の写しからも同様の試みをする中で、更にその変遷と分布分析を多元化できるだろう。

## 5、今後の課題

地域文化財という観点から景観復原を考える時、文献史料・絵図が重要な史料である。今後はその史料から、いかに景観要素の抽出をしていくか、が最重要課題であると考えられる。地域文化財の要素としては、寺社

仏閣・樹木植生・河川・水路・溝等が考えられるが、それらは位置情報の確立が困難である。例えば絵図の場合、測量精度や描画技術、紙ベース資料ゆえの劣化等の問題も潜在している。それらの問題も加味しながら、その要素を史料から認識・抽出し当時と現状とを比較・検証し、地域文化財を分析するということが重要であろう。もう一段階としてその要素をいかにベクター化するかという手法の検討が多分に必要である。今後は、系統立てて時間軸で追ういわゆる時系列分析の可能性を探り、その手法を検討していきたい。

\*景観…本研究において、地理学的視点から、景観とは周辺空間とはっきり識別できるような一定の特徴を有する空間単元を形成しているもの、として捉えている。

\*地籍図・地籍地番…明治6年(1873)7月、地租改正条例が公布され、土地の測量作業が本格化する過程で、村ごとに縮尺600分の1の地籍図が作成された。地券取調総絵図、地引絵図などと呼ばれるこの地籍図は土地一筆ごとに境界、地番、面積が記入され、地目が屋敷、田、畠などによって彩色されており、小字名が記載されている。

#### 参考文献

- 出田和久・木村圭司・宮崎良美 「古地図に描かれた内容のデータベース化のためのシステム構築」『古地図に描かれた内容のデータベース化のためのシステム構築』1999
- 黒川隆夫 「江戸図データベースとその応用」『講座 人文科学のための情報処理 データベース編』尚学社1998
- (社)日本観光協会 『城下町彦根の町なみ-歴史的景観の調査と保存修景-』2000
- 中村和郎・手塚章・石井英也 『地域と景観』古今書院1991
- 彦根市教育委員会 「彦根藩データノート 善利組屋敷(リーフレット)」
- 彦根市史編纂委員会 『彦根市史』
- 彦根城博物館 『彦根の歴史-ガイドブッカー-』1991
- 星野秀和・久保紀重・飯村威・田宮彰弘・飯田剛輔・平井政二・大伴真吾 「空間情報と時系列情報の統合化に関する研究 -道路・交通情報管理のためのプロトタイプング-」『GIS-理論と応用,1999,Vol.7 No.2』GIS学会誌1999
- 室谷 誠一 「彦根藩足輕善利組屋敷の平面型について」日本建築学会学術講演梗概集1998
- 油浅耕三 「古絵図による地域の歴史的環境に関するGISの構築-柏崎市における試み-」『地理情報システム学会講演会論文集 Vol.9』地理情報システム学会2000

○ラスターデータ (画像データ)

	貼り付け範囲	縮尺	解像度
「御城下惣絵図」 (天保七年 1836)	足軽屋敷 1 街区	—	600dpi
	足軽屋敷全体		200dpi
航空写真 (昭和五十七年 1982)	足軽屋敷全体	1/10000	600dpi
	彦根城下ほぼ全域		200dpi
航空写真 (昭和三十六年 1961)	彦根城下ほぼ全域	1/10000	300dpi
航空写真・米軍撮影 (昭和二十二年 1947)	彦根城下ほぼ全域	1/40000	300dpi

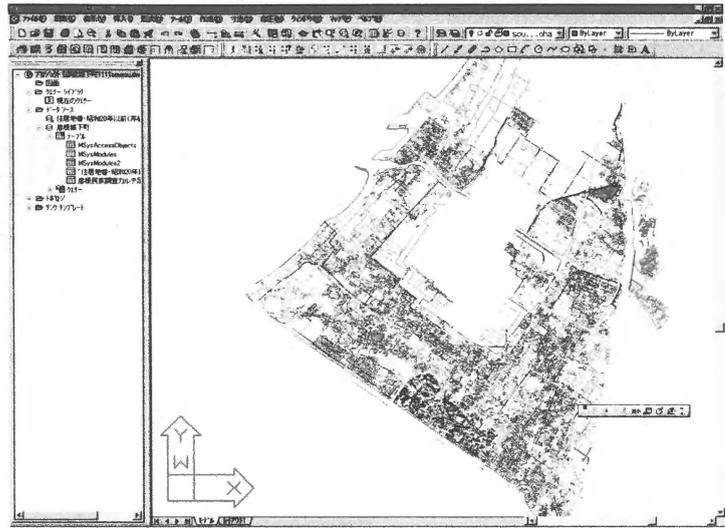
○テキストデータベース (フィールド一覧とキー設定)

	彦根民家調査カルテ DB	家屋台帳・昭和 20 年以前建築の民家 DB	「御城下惣絵図」 (天保七年 1836) DB
主な項目 <キー>	旧住所 (地籍地番) 現住所 建築年 氏名 土地身分  間口 1 階軒高 2 階軒高 建築種別武士住宅 建築種別町家 建築種別長屋 改造 (度数評価) 町家 屋根 卯建 (有無) 地棟 (本数) 袖壁 (有無) 軒裏 2 階壁面 2 階後室 2 階開口部 1 階開口部 門 塀 塗籠  他	所在地 (地籍地番)  建築年  用途 延床面積 1 階面積	氏名 身分  間口

### 3-2 ベースマップ

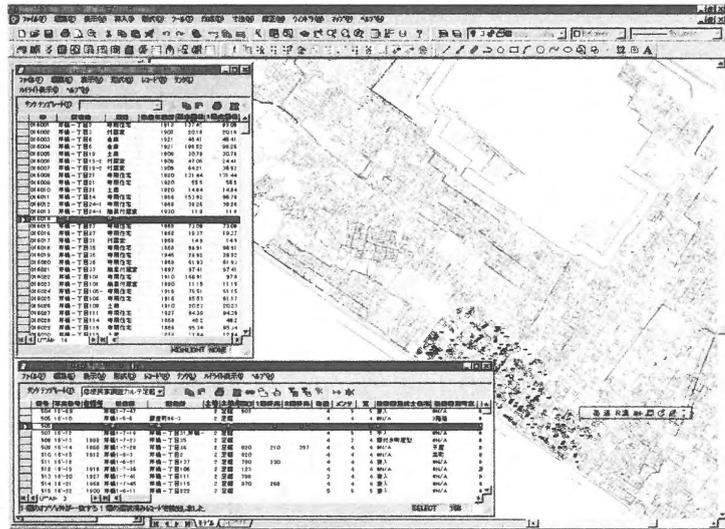
#### 3-2-1 ベクターデータ

- 彦根デジタルマップ  
彦根市現況平面図 (S:1/500)  
彦根市都市計画図 (S:1/2500)による

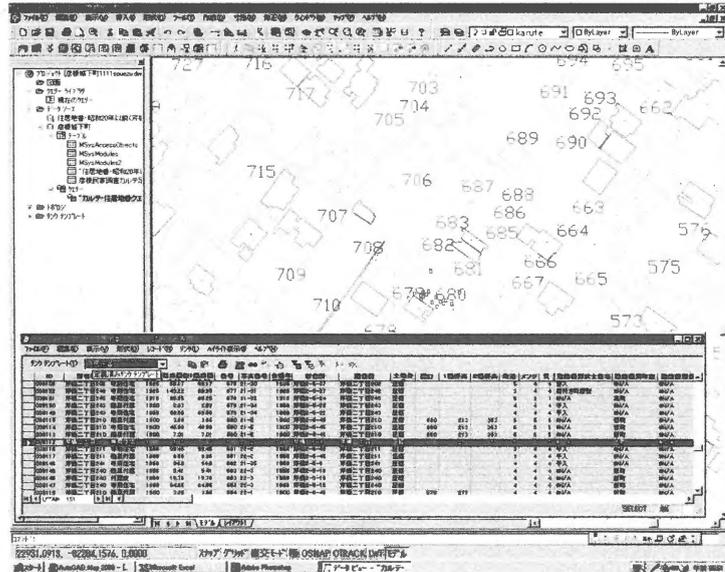


#### 3-2 テキストデータ

- 彦根民家調査カルテデータベース
- 家屋台帳昭和20年以前建築の民家データベース

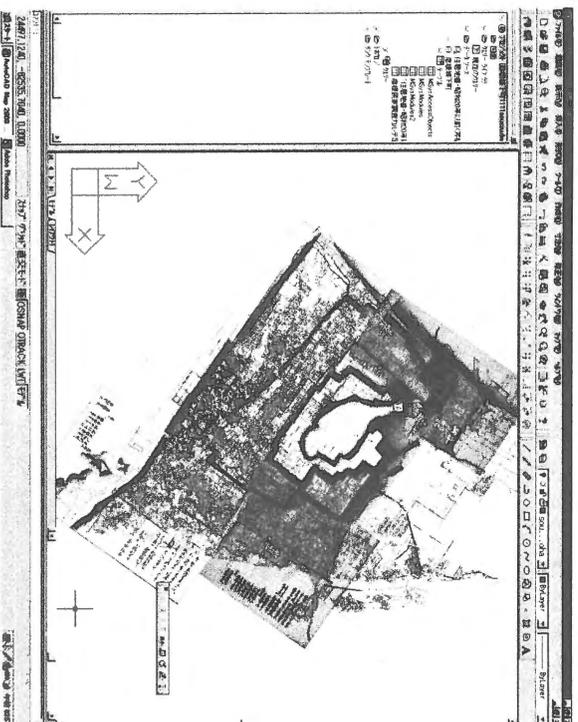


- 彦根民家調査カルテデータベースと家屋台帳データベースのクエリにリンクした建物オブジェクト



### 3-3-1 マクロ分析の検討

- ・御城下惣絵図(1836)  
6枚分割の絵図を貼る際、  
編集は尺度変更のみである
- ・彦根デジタルマップ



- ・航空写真(1947) 米軍撮影
- ・彦根デジタルマップ



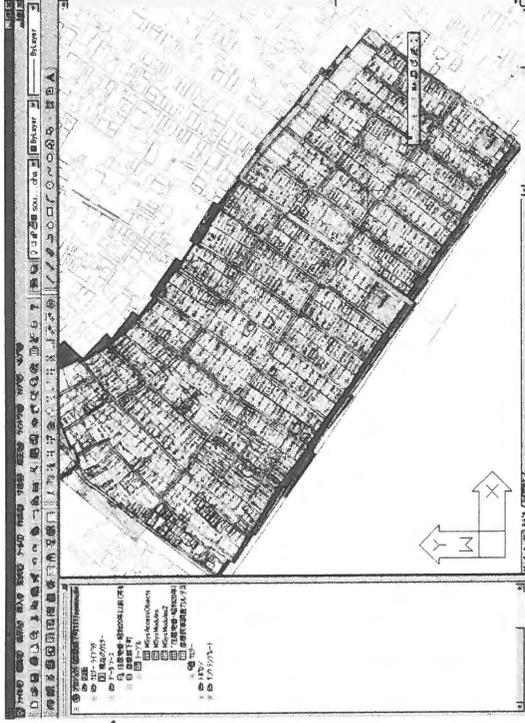
- ・航空写真(1982)
- ・彦根デジタルマップ



### 3-3-2 ミクロ分析の検討

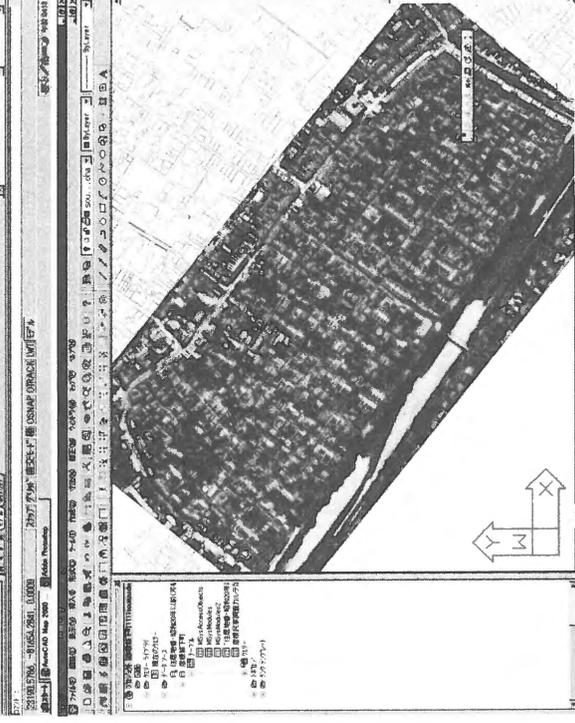
- ・御城下惣絵図(1836)  
歪みを小さくするため街区毎に貼るが、  
編集は尺度変更のみである

- ・彦根デジタルマップ



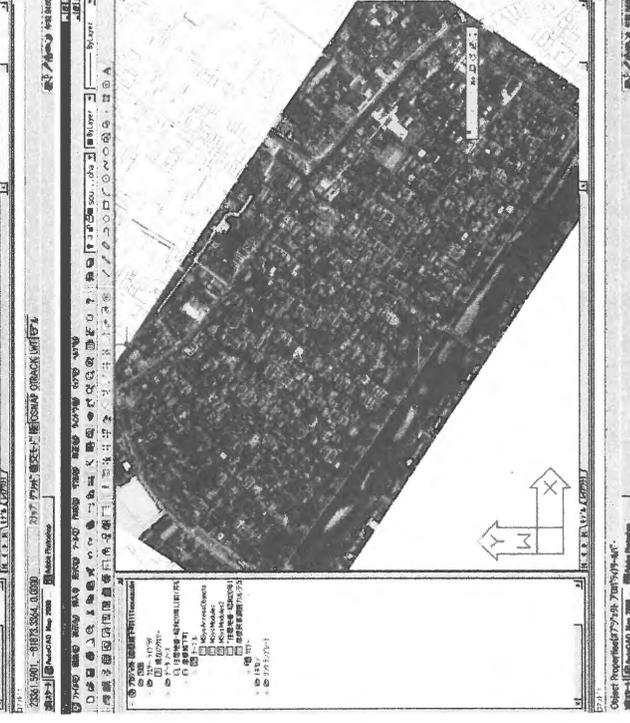
- ・航空写真(1961)

- ・彦根デジタルマップ



- ・航空写真(1982)

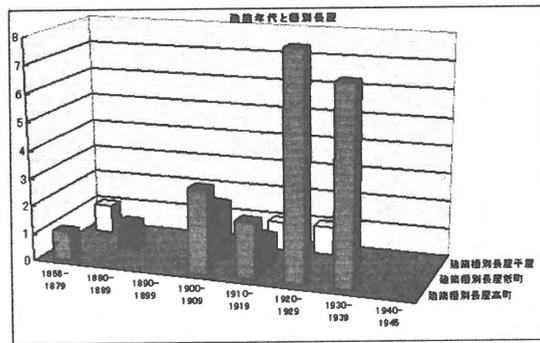
- ・彦根デジタルマップ



番号	建築年代	新住所	旧住所	地番	開口	開口区分	1階軒高	2階軒高	建築種別(大住宅)	建築種別(定)	建築種別(長屋)	建築種別(近)	建築種別(その他)	屋根	基礎	地蔵木数	地蔵有無
512	18-19	1316	茅葺一丁目36	茅葺一丁目106	足跡	1237	12m以上		BN/A	BN/A	BN/A	高町	BN/A	瓦	BN/A		なし
530	17-2	1926	茅葺一丁目4-14	茅葺一丁目331	足跡	670	8m以上9m未満	234	540	BN/A	BN/A	高町	BN/A	瓦	BN/A		なし
535	17-9	1926	茅葺一丁目4-32	茅葺一丁目402	足跡	880	8m以上9m未満	223	454	BN/A	BN/A	高町	BN/A	瓦	BN/A		なし
540	17-17	1920	茅葺一丁目3-41	茅葺一丁目510	足跡	1745	12m以上	272	473	BN/A	BN/A	高町	BN/A	瓦	BN/A		なし
541	17-18	1910	茅葺一丁目3-41	茅葺一丁目511	足跡	1005	10m以上11m未満	205	445	BN/A	BN/A	高町	BN/A	瓦	BN/A		なし
545	17-24	1900	茅葺一丁目13.14.15	茅葺一丁目632	足跡	2050	12m以上	248	499	BN/A	BN/A	高町	BN/A	瓦	BN/A		なし
547	17-27	1921	茅葺一丁目2-40	茅葺一丁目610	足跡	920	9m以上10m未満	228	472	BN/A	BN/A	高町	BN/A	瓦	BN/A		なし
552	17-24	1900	茅葺一丁目37.38.39	茅葺一丁目704	足跡					BN/A	BN/A	高町	BN/A	瓦	BN/A		なし
555	17-35	1926	茅葺二丁目9-25	茅葺二丁目638	足跡					BN/A	BN/A	高町	BN/A	瓦	BN/A		なし
555	17-E	1900	茅葺一丁目1-43	茅葺一丁目707	足跡					BN/A	BN/A	高町	BN/A	瓦	BN/A		なし
568	18-17	1934	茅葺二丁目8-41	茅葺二丁目407-1	足跡					BN/A	BN/A	高町	BN/A	瓦	BN/A		なし
516	18-27	1926	茅葺二丁目7-11	茅葺二丁目434	足跡					BN/A	BN/A	高町	BN/A	瓦	BN/A		なし
690	22-11	1933	茅葺二丁目6-54	茅葺二丁目217-1	足跡	1980	12m以上	291	504	BN/A	BN/A	高町	BN/A	瓦	BN/A		なし
701	22-20	1921	茅葺二丁目5-57.58	茅葺二丁目121	足跡					BN/A	BN/A	高町	BN/A	瓦	BN/A		なし
702	22-21	1927	茅葺二丁目5-58	茅葺二丁目120	足跡					BN/A	BN/A	高町	BN/A	瓦	BN/A		なし
711	22-31	1930	茅葺二丁目5-39	茅葺二丁目104	足跡					BN/A	BN/A	高町	BN/A	瓦	BN/A		なし
715	22-33	1930	茅葺二丁目4-29	茅葺二丁目149-1	足跡					BN/A	BN/A	高町	BN/A	瓦	BN/A		なし
722	23-8	1930	茅葺二丁目4-58	茅葺二丁目32-1	足跡					BN/A	BN/A	高町	BN/A	瓦	BN/A		なし
724	23-10	1888	茅葺二丁目3-11.12	茅葺二丁目42-1	足跡					BN/A	BN/A	高町	BN/A	瓦	BN/A		なし
728	23-16	1930	茅葺二丁目3-53.54	茅葺二丁目708	足跡					BN/A	BN/A	高町	BN/A	瓦	BN/A		なし
734	23-24	1925	茅葺二丁目2-43	茅葺二丁目827	足跡					BN/A	BN/A	高町	BN/A	瓦	BN/A		1両方

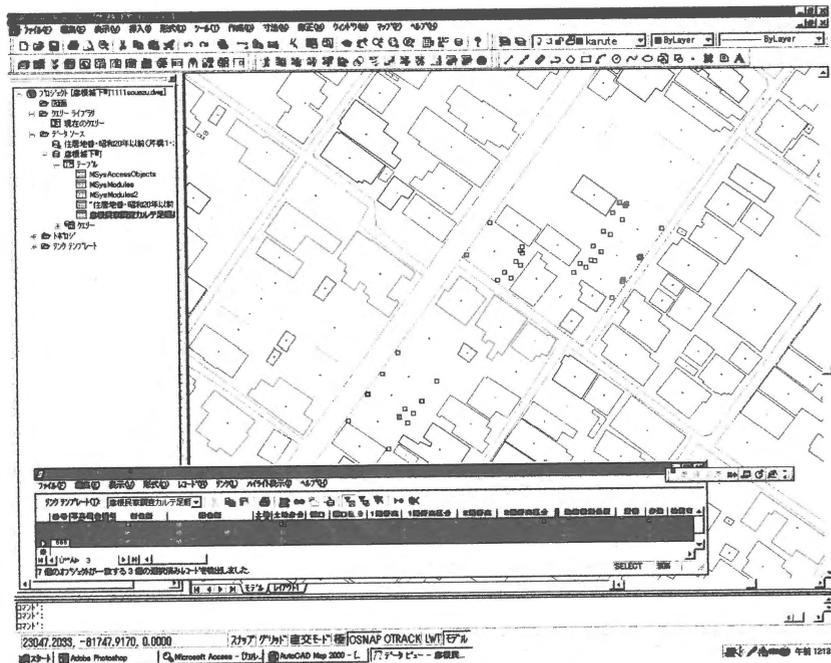
①彦根民家調査カルテデータベース  
“建築種別長屋高町”を抽出

	建築種別 長屋高町	建築種別 長屋低町	建築種別 長屋平屋	合計
1868-1879	1		1	2
1880-1889		1		1
1890-1899				0
1900-1909	3	2		5
1910-1919	2	1	1	4
1920-1929	8		1	9
1930-1939	7			7
1940-1945				0
総計	21	4	3	28



②建築年代別による建築種別長屋の分類

③ AutoCAD Map 2000上で  
“建築年代 1900-1909”  
“建築種別長屋高町”を抽出



# 考古学データベースにおける検索エンジンの研究

## Implementation of a Retrieving Engine for an Archaeological Database

三浦宙明 小澤一雅

Hiroaki MIURA and Kazumasa OZAWA

大阪電気通信大学大学院

〒572-8530 大阪府寝屋川市初町18-8

Graduate school, Osaka Electro-Communication University,

Neyagawa-shi, Osaka 572-8530, Japan.

### あらまし

本論文では前方後円墳データベースにかかわる検索エンジンについて述べる。最初に、前方後円墳データベースを構成する主な項目について概説する。これらのデータ項目は、前方後円墳に関する考古学的研究において重要な役割をはたすと考えられるものに示ぼられている。次に、前方後円墳データベースがインターネット上で公開され、考古学的視点から多くのユーザーが関心を持って利用することを前提に、考古学データベースのための検索エンジンがいかにあるべきかについての考察を述べる。本論文では、試案として具体的に検索エンジンを構築し、これについて検討を行う。

### *Summary*

This paper presents a retrieving engine for an archaeological database of Japanese ancient tombs. First, the basic items composing the database are illustrated, which have been playing key-roles in archaeological study on Japanese ancient tombs. Second, a technical discussion is made on how to build a retrieving engine, provided that many people use the database through the internet. It would be very important that our retrieving engine should be friendly to users in archaeology who are necessarily not familiar to any machine retrieving procedure. This paper presents tentative implementation of a retrieving engine and some examples of information retrievals from the database using the engine.

キーワード 考古学/データベース/古墳/検索エンジン

*Keywords* archaeology / database / ancient tomb / retrieving engine

## 1.はじめに

前方後円墳は約1500年前（4世紀から6世紀にかけて）につくられた巨大な墳墓のことである。前方後円墳は北海道・青森・秋田・沖縄をのぞく43県に存在し、およそ5800基ほど遺存していると考えられる。

本研究ではその古墳1基ごとに、築造されたと考えられる“推定時期”、古墳の大きさや形態を表す“7部位”、現在その古墳がどの程度の状態で依存しているかの“崩壊度”などのデータ項目を設け各データ項目を原簿カードに記述している。近年、データベースをwebで公開するために、特に必要とされる40項目についての電子化作業を進めてきた。

これら大量のデータから特定の前方後円墳にかかわる学術的情報を検索することは容易ではない。そこで、できる限り効率よく所望の情報を検索するためのシステムが必要となる。また、前方後円墳データベースを利用すると思われるユーザは考古学に対して仮りに理解が深くとも、検索エンジンで用いられるAND・OR等の論理演算については、やや抵抗があると考えられる。本研究では、よりユーザーフレンドリな情報検索手順をめざしてCGIを用いた検索エンジンを試作した。

## 2. 前方後円墳データベース

前方後円墳は時代と地域によって型式が変化する。データベース化にあたり40種類の項目を設定している。以下主要項目について要約する。

### 2-1 主要項目

#### 古墳名にかかわる項目群

古墳名を漢字で表記し、またカタカナで読みを示す。さらに将来検索エンジンの英語化を見越してローマ字（ヘボン式）表記も併用する。

#### 緯度・経度にかかわる項目

前方後円墳の緯度・経度は各基ごとに後円部の中心の位置をデジタイザで測定した計測

値である。

#### 古墳形態にかかわる項目群

前方後円墳データベースを構成する古墳は、ほとんどが前方後円墳であるが（図1 前方後円墳）、少数の例外的古墳も収録している。すなわち、帆立貝式古墳・前方後方墳・双方中円墳・上円下方墳である。一方、前方後円墳の基本形状を決定づける重要な要素としての墳丘の7つの計測値も収録している（図2 7部位）。

#### その他にかかわる項目群

他のデータ項目として、巨大古墳に近接する小規模の古墳で、巨大古墳に関連して営まれたとされる陪塚、古墳の表面の装飾にかかわる葺石や埴輪、古墳内部の遺物や石室などの項目が存在する（表1 データ項目）。

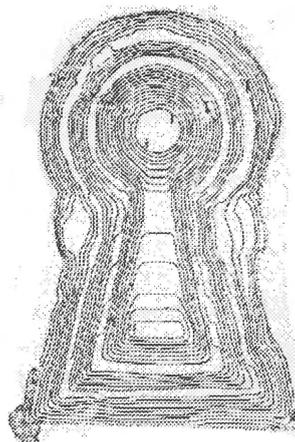


図1 前方後円墳（仲津山古墳）

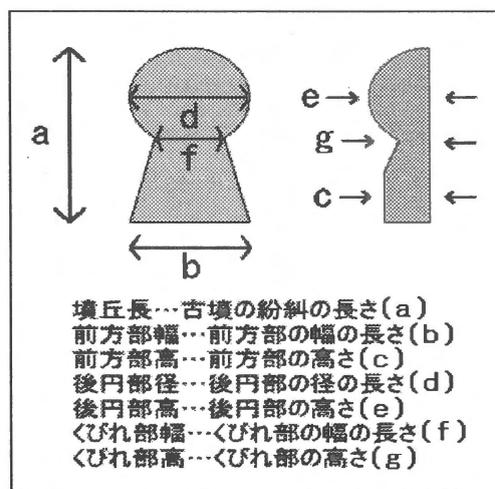


図2 7部位



### 3. 検索エンジン

考古学データベースのための検索エンジンがいかにあるべきかを考慮したうえで、シンプルなキーワード入力形式の検索方式を選択した。また、AND・OR等の論理演算を表面上に出さず、ボタン操作の繰り返しによる絞り込みを行えるよう、ユーザーフレンドリな検索エンジンをめざした。

本研究での特徴として、全国一斉に古墳の検索を行うのではなく、各県または地方ごとの選択を行い、該当地域から古墳の検索を行うことにある。これは全国に対して単純な検索（例えば古墳形態が前方後円墳を検索）を行うと 5000 基を越す古墳が該当データとして返ってきてしまい、取り扱いが複雑になるからである。

検索エンジンの流れとしては、始めに古墳データを先に述べた府県データファイルから 1 行ずつ読み込み、“#” ごとにデータを切り分ける。次に検索側の入力する検索文字（キーワード）が、対応したデータ項目に含まれているかどうかについてパターンマッチを行う。この作業を選択したすべての古墳データに対して行い、必要であればこれら該当古墳に AND 演算を用いた絞り込み検索を行う。

しかし、単に文字列のパターンマッチでは問題が起こることがある。例えば、方位の場合において“東”を選択すると“東”だけでなく“北東”のように“東”を含むすべての方位が該当となってしまう。そこで、方位の場合はデータ項目とキーワードが完全一致の場合のみを該当古墳として検索を行う。また、墳丘長など検索項目が数値の場合、“486m の古墳を調べたい”というような検索要求はまれであって、大方は“300m以上 500m以下”のように特定範囲で検索を行うと考えられる。そこで“300-500”と入力する事で数値の範囲検索が可能となり、目的的古墳を絞り込むことが出来ように設計した。

上記の繰り返しにより絞り込まれた古墳か

ら、詳細データを表示させたい古墳を選択し情報を表示させる。また、データ項目の 7 部位から前方後円墳の復元形を推測し、数値だけでなく画像として古墳形態を表示させている。

本検索エンジンは、Perl 言語でプログラムされた CGI (Common Gateway Interface) によって検索を行うものであり、web に繋がっているパソコンであればブラウザを用いてすべてのマシンで利用できる。また、古墳の復元画像のアプレット部は JAVA によって構築し、ブラウザの画面に組み込んで表示をさせている。

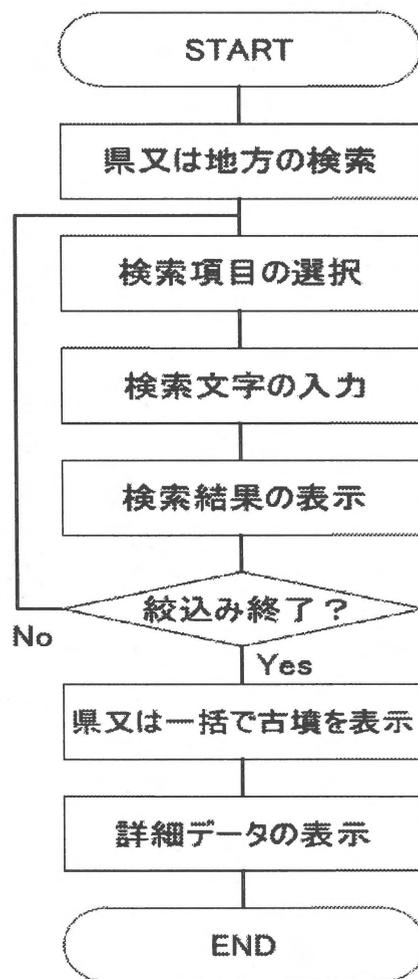


図 6 検索の流れ

#### 4. 実行例

ここでは、つぎの実例をもとに、本検索エンジンによる絞り込み過程を示す。

**【例】検索キーワード**

地域 = “近畿地方”，古墳名 = “天皇” を含む，墳丘長 = “300m以上 500m以下”

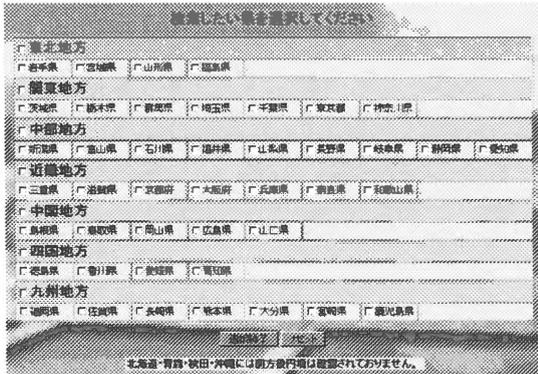


図 7

**Setp 1**

左図が検索エンジンの Top ページである。チェックボックス型の選択画面であって、地方別の近畿地方にチェックを入れることで地域の選択を行うことができる（地方別で近畿地方を選択し、また県別で大阪や京都を選択しても二重検索は行われない）。

[論理演算 OR]

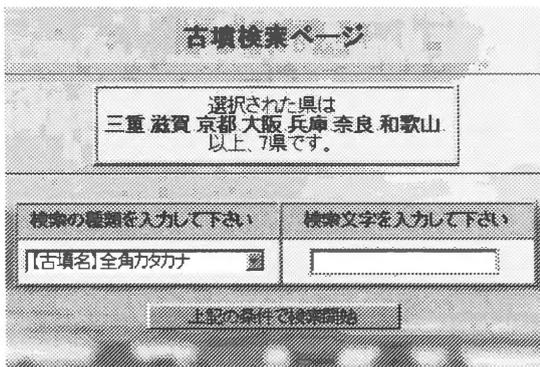


図 8

**Setp 2**

Step 1 より “選択された県は、三重・滋賀・京都・大阪・兵庫・奈良・和歌山、以上 7 県です” と表示される。

次に “古墳名に天皇” を含む古墳を検索したいので、検索項目を “【古墳名】漢字” 検索文字を “天皇” と入力し検索開始ボタンを押す。

[論理演算 AND]

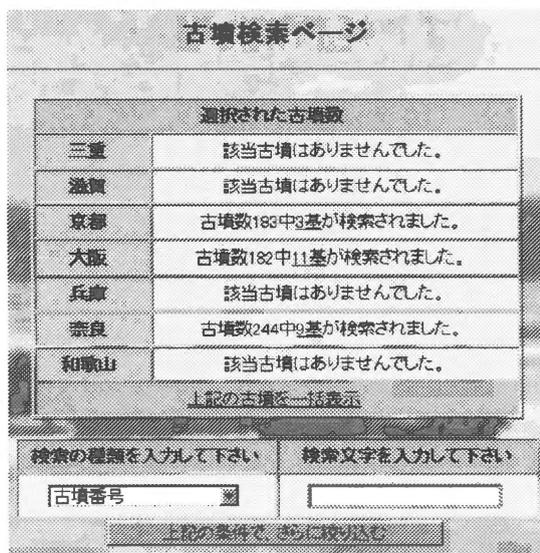


図 9

**Setp 3**

絞り込みにより、京都 3 基・大阪 1 1 基・奈良 9 基の古墳が該当古墳であると結果が表示される。

ここで、さらに古墳を絞り込みたいので検索項目を “【7 部位】墳丘長” 検索文字を “300-500” と入力し再び絞り込みボタンを押す。

[論理演算 AND]



## 5. CD-ROM パッケージ化

現在、本研究の目的である検索エンジンは web 上で利用可能となっている。しかし、すべてのマシンが web に繋がっているわけではない。そこでローカルマシン上でも検索が行えるよう、前方後円墳データベースの CD-ROM パッケージ化を実現し、より多くの人々に古墳データベースを利用してもらえるよう鋭意作成中である。ただし、ローカルマシン上では web で使用される CGI の検索エンジンは利用できないので、JAVA 言語で検索エンジン部を新たに作成し、web ブラウザさえあれば検索が可能になるように配慮する予定である。

また CD-ROM パッケージ化によって、逆に web 上で容量の問題で実現できなかった立体的地理情報の表示が可能となる。これにより、古墳の築造される地形の立地条件が視覚的に明瞭に確認できるようになると予想される。

## 6. おわりに

本稿では、考古学データベースにおける情報検索のありかたを想定しつつ、実際に構築した検索エンジンを示した。今後は、さらにユーザーフレンドリな検索エンジンをめざして、アプレットの強化による古墳復元図の立体化 (図 13) などを追加し、実用性の高い検索エンジンの構築をめざす。

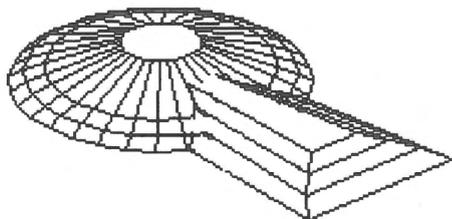


図 13

## 参考 URL

小澤研究室

<http://www.ozlab.osakac.ac.jp/>

古墳検索エンジン

[http://www.ozlab.osakac.ac.jp/cgi-bin/kSearch/search\\_kofun.cgi](http://www.ozlab.osakac.ac.jp/cgi-bin/kSearch/search_kofun.cgi)

## 参考文献

- [1] 小澤一雅『前方後円墳の数理』  
雄山閣 (1988 年)
- [2] 小澤一雅『講座 人文科学のための  
情報処理 [第 2 巻 データベース編]』  
尚学社 (1998 年)
- [3] 近藤芳郎『前方後円墳集成・近畿編』  
山川出版社 (1992 年)
- [4] 文化庁文化財保護部『全国遺跡地図』  
文化庁文化財保護部 (1974 年～)
- [5] 石川昇『前方後円墳築造の研究』  
六興出版 (1989 年)

表1. データ項目

データ項目	データのタイプ	参考：仁徳天皇陵の場合
実測図番号	数値	27113
古墳名（カタカナ）	文字	ニントクテンノウリョウ;ダイセン
古墳名（漢字）	文字	仁徳天皇陵;大仙
古墳名（ローマ字）	文字	Mausoleum of Emperor Nintoku;Daisen tomb
古墳群	文字	百舌鳥古墳群
県名	文字	大阪
所在地	文字	堺市大仙町
所在地コード	数値	27201
データ精度	文字	LBA
緯度	数値	1332927
経度	数値	343342
1:2500地形図	文字	堺
古墳形態	文字	前方後円墳
陪塚（有無）	文字	有
墓石（有無）	文字	有
造出（有無）	文字	有
埴輪（有無）	文字	有
遺物（有無）	文字	有
石材質（有無）	文字	有
石室（有無）	文字	有
棺（有無）	文字	不明
立地	文字	台地
方向	文字	南南西
崩壊度	文字	完存
推定時期	文字	中期
墳丘長	数値	486
前方部長	数値	300
前方部幅	数値	34
後円部径	数値	244
後円部高	数値	36
くびれ部幅	数値	163
くびれ部高	数値	25
体積	数値	不明
周濠（有無）	文字	有
築成段数	数値	不明
築成段数（前方部）	数値	4
築成段数（後円部）	数値	4
埴輪（備考）	文字	円筒;人;馬
石室（備考）	文字	竪穴式石室
棺	文字	不明
参考文献	文字	原島礼二「巨大古墳と倭の五王」（1981年）

## 公開シンポジウム「人文科学とデータベース」既刊一覧

### ■ 公開シンポジウム「人文科学とデータベース」

1995年12月25日(月), 26日(火) 大阪電気通信大学(寝屋川キャンパス)

#### 特別講演

古地震データと活断層 (pp. 1-4)

寒川旭(通産省地質調査所)

#### 一般講演

IntelligentPad システムを用いた歴史学研究支援データベースの構築 (pp. 5-12)

赤石美奈, 中谷広正, 伊東幸宏, 阿部圭一, 田村貞雄(静岡大学)

4次元歴史空間システムにおける地理情報処理について (pp.13-18)

小林努, 加藤常員, 小沢一雅(大阪電気通信大学)

視点に依存する属性付け機構を持つ木簡研究支援システム

—構造化型データベースの概念— (pp. 19-28)

森下淳也(姫路獨協大学), 上島紳一(関西大学), 大月一弘(神戸大学)

古典籍と JIS 漢字 —テキストの本文校正との関係において— (pp. 29-36)

當山日出夫(花園大学)

手書き文字時系列筆跡パタンの一解析と今後の計画 (pp. 37-42)

東山孝生, 山中由紀子, 澤田伸一, 中川正樹(東京農工大学)

絵画 DB とイメージ検索 —浮世絵の線画表現とデータ圧縮効果— (pp. 43-48)

濱裕光, 志賀直人(大阪市立大学)

画像データベースの自然言語インタフェースについて (pp. 49-54)

伊東幸宏, 中谷広正(静岡大学)

多視点距離データを用いた3次元形状モデリング (pp. 55-60)

横矢直和(奈良先端科学技術大学院大学), 増田健(電子技術総合研究所)

ハイパーメディア・コーパスの構築と言語教育への応用について (pp. 61-66)

上村隆一(福岡工業大学)

「歌物語」語彙の数量的分析と研究 (pp. 67-74)

西端幸雄(大阪樟蔭女子大学)

高次辞書データベースのための語彙知識自動獲得システム (pp.75-82)

亀田弘之(東京工科大学), 藤崎博也(東京理科大学)

社会調査結果の視覚化データベース (pp. 83-88)

吉田光雄(大阪大学)

「間」に関するデータベースの構築 (pp. 89-98)

中村敏枝(大阪大学)

方言音声データベースの作成と利用に関する研究 (pp. 99-104)

田原広史, 江川清, 杉藤美代子, 板橋秀一(大阪樟蔭女子大学)

■ 第2回公開シンポジウム「人文科学とデータベース」  
1996年12月21日(土) 大阪電気通信大学(寝屋川キャンパス)

招待講演

三浦梅園の主著『玄語』のデータベース化と解読の試み  
—江戸時代のハイパーテキストを読み解く— (pp. 1-10)  
赤星哲也(日本文理大学), 北林達也(三浦梅園研究所)

江戸図データベースの作成と今後の課題 (pp. 11-22)  
黒川隆夫(京都工芸繊維大学)

一般講演

Shape from motion を応用した什器類の立体データ作成 (pp. 23-34)  
中島重義, 岡本次郎, 濱裕光, 細川省一(大阪市立大学)

古地図に描かれた内容のデータベース化への試み (pp. 35-44)  
出田和久, 正木久仁, 小方登, 山近博義(奈良女子大学)

考古学のためのデータベースシステム (pp. 45-54)  
宝珍輝尚, 中田充, 白井治彦, 都司達夫(福井大学)

インターネット・イントラネットにまたがる分散型図書館目録データベース  
の構築と運用 (p. 55)  
芝勝徳(神戸市外国語大学)

音楽における印象語検索システムの開発とその有用性 (pp. 57-66)  
原田章, 吉田光雄(大阪大学)

「方言認知地図」プログラムと統計処理地図 (pp. 67-78)  
ダニエル・ロング(大阪樟蔭女子大学)

■ 第3回公開シンポジウム「人文科学とデータベース」

1997年12月20日(土) 大阪電気通信大学(寝屋川キャンパス)

招待講演 「邪馬台国大和説を科学する」

倭人社会と卑弥呼の王権 (pp. 1-10)

吉田晶 (岡山大学名誉教授)

前方後円墳成立の歴史的意義

—「邪馬台国時代」における首長と農民層の共同幻想— (pp.11-17)

広瀬和雄 (奈良女子大学)

一般講演

歴史学研究支援システムの構築 (pp. 19-30)

三浦崇, 伊東幸宏, 小西達裕, 田村貞雄(静岡大学), 赤石美奈(北海道大学)

中谷広正, 阿部圭一(静岡大学)

地理情報を利用した遺物データベースシステムについて (pp. 31-38)

宝珍輝尚, 都司達夫(福井大学), 河合秀夫(大阪電気通信大学)

木簡研究支援データベースシステム —知見と仮説に基づく再構造化 (pp. 39-46)

森下淳也, 大月一弘(神戸大学), 上島紳一(関西大学)

大庭脩(皇学館大学), 杉山武司(姫路獨協大学)

短編推理小説の論理構造の分析 (pp. 47-54)

西島恵介, 神山文子, 藤田米春 (大分大学)

形状分析ツールの開発とその応用 —浮世絵に描かれた役者の同定と分類— (pp. 55-64)

モハメド・アラミン・ブイヤン, 阿古弥寿章, 濱裕光(大阪市立大学),

松平進 (甲南女子大学)

技術紹介

3次元形状入力へのおさそい —人文科学の道具として— (pp. 65-72)

濱裕光(大阪市立大学)

■ 第4回公開シンポジウム「人文科学とデータベース」

1998年12月19日(土) 大阪電気通信大学(寝屋川キャンパス)

招待講演

邪馬台国は北九州甘木市付近にあった (pp. 1-22)

安本美典 (産業能率大学)

特別講演

選挙研究とデータベース (pp. 23-24)

三宅一郎 (関西大学)

一般講演

中期インド・アリアン聖典のデータベース (pp. 25-34)

逢坂雄美, 山崎守一 (仙台電波工業高等専門学校), 宮尾正大 (室蘭工業大学)

高地性集落遺跡データベースからみた弥生時代の情報通信 (pp. 35-42)

加藤常員 (大阪電気通信大学)

データベース倫理について (pp. 43-52)

江澤義典 (関西大学)

階層構造グラフによるデータモデルの適用例: 木簡データベース (pp. 53-59)

杉山武司 (姫路獨協大学) 森下淳也, 大月一弘 (神戸大学) 上島紳一 (関西大学)

遺物破片の計測—照度差ステレオ計測装置の製作と計測— (pp. 61-70)

結城宏和, 宝珍輝尚, 都司達夫 (福井大学), 河合秀夫 (大阪電気通信大学)

技術紹介

地場工芸品立体展示システム —人文科学の道具として— (pp. 71-78)

橋本隆之, 青木功介, 釣裕美 (インテックシステム研究所)

■ 第5回公開シンポジウム「人文科学とデータベース」

1999年12月18日(土) 関西大学総合情報学部

招待講演

顔の認知研究における顔データベースの利用 (pp. 1-10)

加藤隆 (関西大学)

Recent Research on Paleolithic Arts in Europe and the Multimedia Database (pp. 11-22)

César González Saintz, Roberto Cacho Toca (University of Cantabria, Spain)

一般講演

PhotoVR 考古資料データベース『北スペインの旧石器洞窟美術』

日本・スペイン産学共同 プロジェクトの実現と諸問題 (pp. 23-34)

深沢武雄(テクネ)

バーチャルリアリティによる遺跡探訪 3D Archaeo-Copter (pp. 35-38)

中村健, 小沢一雅(大阪電気通信大学)

照度差ステレオ法を用いた遺物の表裏形状の計測 (pp. 39-46)

結城宏和, 宝珍輝尚, 都司達夫(福井大学)

文学データベースのための文書の構造化と意味管理 (pp. 47-58)

横田一正, 三宅忠明, 國島丈生(岡山県立大学)

劉渤海(岡山理科大学), 田槇明子(リョービスシステムサービス)

相対インデックス法を使った文構造分析 (pp. 59-66)

雄山真弓, 岡田孝, 黒崎茂樹(関西学院大学)

仮想電子辞書システムの設計と構築 (pp. 67-78)

芳野学, 都司達夫, 宝珍輝尚(福井大学)

主催 第6回公開シンポジウム実行委員会

後援 人文系データベース協議会

代表 中谷広正 (静岡大学情報学部)  
江澤義典 (関西大学総合情報学部)  
阿部圭一 (静岡大学情報学部)  
伊藤敏彦 (静岡大学情報学部)  
伊東幸宏 (静岡大学情報学部)  
小西達裕 (静岡大学情報学部)  
白井靖人 (静岡大学情報学部)  
田村貞雄 (静岡大学情報学部)  
長瀬真理 (静岡大学情報学部)  
八重樫純樹 (静岡大学情報学部)

議長 小沢一雅 (大阪電気通信大学)  
出田和久 (奈良女子大学)  
江澤義典 (関西大学)  
及川昭文 (総合研究大学院大学)  
雄山真弓 (関西学院大学)  
加藤常員 (大阪電気通信大学)  
川口 洋 (帝塚山大学)  
黒川隆夫 (京都工芸繊維大学)  
柴山 守 (大阪市立大学)  
高橋晴子 (大阪樟蔭女子大学)  
都司達夫 (福井大学)  
中谷広正 (静岡大学)  
中村敏枝 (大阪大学)  
八村広三郎 (立命館大学)  
濱 裕光 (大阪市立大学)  
深海 悟 (大阪工業大学)  
宝珍輝尚 (福井大学)  
正木久仁 (大阪教育大学)

人文系データベース協議会 第6回公開シンポジウム「人文科学とデータベース」

---

発行日 2000年12月20日

発行所 第6回公開シンポジウム実行委員会

432-8011 浜松市城北3-5-1 静岡大学情報学部情報科学科中谷広正

[nakatani@cs.inf.shizuoka.ac.jp](mailto:nakatani@cs.inf.shizuoka.ac.jp)