第18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

２０１２年１２月２２日（土）

会場：大阪電気通信大学 寝屋川キャンパス
主催：第１８回公開シンポジウム実行委員会
共催：大阪電気通信大学情報学研究施設
後援：人文系データベース協議会

目　　次

● **特別講演**

「情報」という語の成り立ち
小野　厚夫（神戸大学名誉教授）　1

● **一般講演Ⅰ**

ユビキタス個人書斎により和歌の共有・鑑賞・創作を統合的に支援するシステムの提案と試作
陳　泓，鄭　迎花，岸本　知也，小山　敦史，金　群（早稲田大学）　3

アパブランシャ語コーパスを用いた韻律による年代推定
山畑　倫志（北海道大学）　11

刑事判決書の特徴表現パターン抽出に関する複数手法の検討
千本　達也，山本　大介，竹内　和広（大阪電気通信大学），　17
三島　聡（大阪市立大学）

文章の特徴量を用いた質問回答文の印象の因子得点の推定精度の向上
横山　友也，宝珍　輝尚，野宮　浩揮，（京都工芸繊維大学）　25

● **一般講演Ⅱ**

古代を中心とした歴史地理データベースの試み
宮崎　良美（奈良女子大学）　33

考古学データ検索における異種データベースの統一的な利用について
王　鑫，宝珍　輝尚，野宮　浩揮（京都工芸繊維大学），　43
佐藤　哲司（筑波大学）

貝塚データベース　―　作成から利用まで　―
及川　昭文（総合研究大学院大学）　51

● **一般講演Ⅲ**

石造遺物銘文取得のためのアーカイビング手法の開発
上椙　英之（神戸学院大学），上椙　真之（宇宙航空研究開発機構），　59
多仁　照廣（敦賀短期大学）

文化遺産情報資源共有化のためのリレーショナルデータベース構築
－小豆島岩谷地区石切丁場における実践例－
高田　祐一（同志社大学）　67

歴史と数理
小沢　一雅（大阪電気通信大学）　75

第 18 回公開シンポジウム
「人文科学とデータベース」
2012. 12. 22

「情報」という語の成り立ち

When and how did the Japanese word "Joho" appear and widespread?

小野　厚夫
Atsuo Ono
神戸大学名誉教授，神戸市灘区六甲台町 1-1
Kobe University, Professor Emeritus, 1-1 Rokkodai, Nada-ku, Kobe

あらまし：「情報」は日本で造られた漢語で、明治 9 (1876) 年に酒井忠恕が訳した「仏国歩兵陣中要務実地演習軌典」に最初に現れる。その原語はフランス語の renseignement で、敵の「情状の報知」を意味する。この訳本は野営演習の教習本として用いられ、さらに「情報」は 15 年の「野外演習軌典」で公式に採用されたことによって、陸軍内で普及した。当初「状報」も併用されていたが、ほどなく「情報」に一本化されてしまう。新聞では 27 年の日清戦争の記事に最初に現れ、33 年の北清事変で早くも日本語として一般化し、日露戦争の直前から国語辞典に採録されるようになった。

Summary: The oldest book used the word "Joho" was published in 1876 as a Japanese translation version of French army book. The original word is "renseignement" in French. It was used in the official documents in 1882, and spread in the army. In the newspapers, it appeared first during the Sino-Japanese War in 1894-5, and obtained general acceptance among the people from the articles of the North China affair in 1900. After 1904, the word "Joho" was found in the dictionaries of Japanese language.

キーワード：情報，翻訳語，フランス語源，一般化
Keywords: "Joho", "renseignement", derivation of the word, generalization of the word

1. はじめに

「情報」は中国でも使われているが、中国人が日本語来源の中国語として認めている。また最近、「情報」を分析の手段として取り上げるようになった日本近代史の研究者たちが、江戸時代には「情報」という言葉はなかったと書いているので、明治に入ってから日本で造られた語とみなすことができる。

2. 「情報」の初出

「情報」が用いられた書物で最古のものは、明治 9 (1876) 年に酒井忠恕が訳した『佛國歩兵陣中要務實地演習軌典』である。原書は 1875 年にフランス陸軍が刊行した『Instruction pratique sur le service de l'infantierieen campagne』で、「情報」の原語はフランス語の「renseignement」である。

この私本は西南の役後、陸軍の各鎮台で実施された野外演習の教習本に採用され、広く陸軍内に普及したが、版木が摩耗したため、明治 14 年に再版された。陸軍省は明治 15 年に『野外演習軌典』を編集、発刊したが、酒井の訳本を基礎にしており、これが「情報」を用いた最初の公文書となる。

「情報」は敵の「情状の報知（ないしは報告）」の意味で、それを二字熟語に縮めたものと解釈することができる。「情」と「状」には「ありさま、ようす」という共通の意味があり、『野外演習軌典』の発刊後「情報」と「状報」が併用されてい

たが、「状報」は兵語統一の動きの中で明治20年代中頃から出現頻度が激減し、次第に淘汰されていく。

3.「情報」という語の一般化

新聞に「情報」という語が現れるのは日清戦争の時である。明治27年11月29日付けの新聞『日本』の従軍記事が最初の用例で、旅団の命令書の写しに現れる。『日本』は翌30日の付録（号外）で大本営掲示を転載していて、その文中に「情報」が用いられている。しかし、他の新聞社では広島の特派員が送った電文の「ジヤウハウ」ないしは「ゼウハウ」が何を意味するのか判断できなかったとみえ、「ゼウ報、諸報、戦報、詳報、報告」などと、ばらばらの書き換えがなされた。しかし、すぐに「情報」は新聞用語として通用するようになり、見出しにも使われるようになった。

日清戦争から日露戦争までの約10年は、その間に北清事変もあり、日本にとっては国際社会に確固たる地位を占める重要な時期となったが、報道に重点を移した新聞はいずれもこれら三つの戦役時に発行部数を飛躍的に増大させた。

当時の新聞における「情報」の出現頻度を調べてみると、平時は演習記事にたまに現れる程度であるが、戦時に急増する。日清戦争後は、ブール戦争（南ア戦争）と北清事変の勃発によって明治33（1900）年の新聞の見出しと記事に「情報」が頻発している。この2年後あたりから国語辞典に「情報」が採択されるようになった経緯と考え合わせると、「情報」は明治27年の日清戦争で新聞用語として成立し、明治33年の北清事変で一般化したとみなすことができる。

その後の国語辞書の解釈をみると、移入した側の中国では「情報」を軍用語として捉えているのに対し、日本では「情報」を「ようすのしらせ」とか「事情の知らせ」のように、軍事色のない、ごく一般的な定義にしていることが注目される。

［参考文献］

小野厚夫『明治9年、「情報」は産声』「日本経済新聞」1990年9月15日朝刊文化欄

小野厚夫『「情報」という語の由来と変遷』「富士通ジャーナル」17巻1号75頁、1991年1月

小野厚夫『明治期における「情報」と「状報」』神戸大学教養部紀要「論集」47巻81頁、1991年3月

小野厚夫『情報小論』神戸大学国際文化学部紀要「国際文化学研究」創刊号1頁、1994年3月

小野厚夫『情報という言葉を尋ねて』（1）～（3）情報処理学会学会誌「情報処理」46巻4～6号（2005年4月～6月）

小野厚夫『情報という言葉の初出とその一般化―明治期の情報』「情報通信学会誌」89号88頁、2009年3月

［年 表］

明治9年	「仏国陸軍紳士要務実地演習軌典」訳	1876
15	「野外演習軌典」制定	1882
20	「仏和辞林」に「状報」	1887
27-28	日清戦争	1894-5
33	北清事変	1900
37-38	日露戦争	1904-5
大正3年	第一次世界大戦	1914-8
10	外務省に情報部設置	1921
昭和11年	内閣に情報委員会設置	1936
12	内閣情報部設置	1937
14	第二次世界大戦	1939-5
15	情報局設置	1940
16	太平洋戦争	1941-5
23	Shannon「情報理論」の論文	1948
35	情報処理学会発足	1960
38	梅棹忠夫 情報産業論 情報（化）社会	1963
58	長山泰介 欧米造語説	1983
平成2年	小野厚夫 「情報」明治九年に産声	1990

ユビキタス個人書斎により和歌の共有・鑑賞・創作を統合的に支援するシステムの提案と試作

An Integrated Support System Based on Ubiquitous Personal Study for Sharing, Reading and Writing of *Waka* Poetry

陳 泓
Hong Chen
早稲田大学メディア研究所 [1]
Media Research Institute, Waseda University [2]
鄭 迎花, 岸本知也, 小山敦史, 金 群
Yinghua Zheng, Kishimoto Tomoya, Koyama Atsushi, Qun Jin
早稲田大学人間科学学術院 [1]
Faculty of Human Sciences, Waseda University [2]

あらまし:短歌、俳句、川柳は和歌の主な体裁である。季語や自然の描写が和歌の特徴であり、和歌の共有、鑑賞、創作、意味づけには欠かせない情報である。本研究では、それらの情報をキーワードとした検索や個人嗜好を考慮した情報推薦、さらに、誰でも参加できる和歌愛好会のオンラインコミュニティ形成を統合的に支援するシステムを提案し、これまで提案・構築したユビキタス個人書斎(UPS)をベースに、ソーシャルメディアとの連携を図ったうえ、試作システムを構築する。

Summary: *Tanka*, *Haiku*, *Senryu* are the main forms of *Waka* (traditional Japanese poetry). *Kigo* and depiction of nature are representative elements of *Waka*. They are extremely important for sharing, reading, writing and understanding *Waka*. In this study, we propose an integrated support system that incorporates such functions as keyword search of those representative elements, information recommendation that takes personal preferences into account, and online community formation of *Waka* Club as well. We further construct a prototype system based on Ubiquitous Personal Study (UPS), which has been developed in our previous study, being linked with a social media tool.

キーワード:ユビキタス個人書斎、和歌の共有・鑑賞支援、情報推薦、オンラインコミュニティ、ソーシャルメディア
Keywords: Ubiquitous Personal Study, sharing and enjoying *Waka* (Japanese poetry), information recommendation, online community, social media

1. はじめに

近年、スマートフォンを代表とする高性能携帯情報端末が世界で急速に広まっており、また、クラウドコンピューティングが安全かつ膨大な情報保存空間を提供している。その一方、Twitterを代表とするソーシャルメディアの活用により、身近な情報を容易にキャッチすることができるようになった。そこで我々はこれらの技術を有効に利用する、ユビキタス個人書斎(UPS)というフレームワークを提案している[1-3]。本研究ではユビキタス個人書斎をベースに、季語や自然の描写など和歌の共有、鑑賞、創作、意味づけに欠かせない情報をキーワードとした検索、個人嗜好を考慮した情報推薦、誰でも気軽に参加できる和歌愛好会のオンラインコミュニティ形成を統合的に支援するシステムを提

1 埼玉県所沢市三ヶ島 2-579-15
2 2-579-15 Mikajima, Tokorozawa-shi, Saitama

案し、ソーシャルメディアとの連携を図ったうえ、試作システムを構築する。

2. 関連研究

現代の和歌は短歌であるとの解説もあるが、本論文では短歌、俳句、川柳(今日流行っている和歌の形態)を指す。

(1) 和歌の広がり

和歌の鑑賞、作成、投稿は日本国内外で広がっている。和歌の中の連歌はコミュニティの中に生まれ、複数の作者が共同で作成・鑑賞するものである。日本国内では、和歌愛好会の数は2005年の時点で868ある[5]。東洋大学は1987年から「現代学生百人一首」[3]を続けている。「サラリーマン川柳」[4]は第一生命が主催するものとしてよく知られている。上田秋成の国語研究にあたる2500首の古歌を、古代・中世の古歌の詠歌の素材・詠方・表現・内容等について通観すると、約3割を占める歌が広義の名所歌とも称されると言われている[6]。

日本の和歌文化は海外まで広がっている。中国は勿論、英語圏の国でも英語俳句が流行っている[7]。

(2) 和歌関連データベースの構築

和歌情報の取り扱いは、これまで専門家の参加が前提になっている。具体的に、学校教育の一環とする鑑賞や意味づけは専門家達に行われ、情報を収集するための和歌創作活動は教育機関や専門家の個人・コミュニティが持つデータベース等の形でしか存在していない。デジタル化された和歌データ、例えば、国際日本文化研究センターの「和歌データベース」や他の個人で作ったフリーデータベース(「百人一首」、「万葉集」)などが多数存在する。このようなデータベースは、専門研究家を対象にし、必要な情報の検索は非常に手間がかかる。

コンピュータを用いて分析・応用する試みとして、「季語データベースの構築と俳句の季語の自動判定」[8]や「季語データベースと俳句投句鑑賞システム」[9]がある。更に、「俳句に適合する合成画像を生成するシステム」[10]のような、和歌要素の細分化・表面化(俳画)を特徴とする和歌鑑賞システムの構築も研究開発されている。また、スマートフォンを用いた俳句創作支援システム[11]、俳画による体験共有型のコミュニケーション・メディア[12]といった研究開発も報告されている。

これらの関連研究から、情報技術により和歌の鑑賞を支援するシステムはいくつかの問題と課題がある。

- データベースが小規模で分散されている。データベースの拡張・継承に個人が持つパーソナルデータベースが活用されていない。
- パブリックデータベースは存在しているが、アクセス・活用に便利なインターフェースが少ない。
- 専門性が除かれた庶民レベルのコミュニティ環境における共同鑑賞・学習・活用が少ない。
- 現状では愛好会コミュニティや団体間の連携が足りない。情報交換と共有が不十分である。
- ソーシャルメディアとの連携がほとんどみられない。

3. 統合支援システムの提案

スマートフォン、クラウドサービス、ソーシャルメディアが著しく発展するユビキタス社会における情報共有技術が進むなか、本研究では新たなICT技術を活用した和歌鑑賞・意味づけ・自作投稿・共有システムを提案したい。本研究の先行研究として、個人情報アクセスを集約する個人化情報ポータルとして、知識作業に関わる情報の一括管理・組織化・共有化を支援するユビキタス個人書斎(Ubiquitous Personal Study、略してUPS)というフレームワークを提案している。ソーシャルメディアにおける情報の収集、集計、分析、それらの情報に隠された価値のある知識のマイニング、さらに、実世界での情報行動・知識作業とデジタル空間とのシームレスな融合を図ることが特徴である。

本研究はこれまで提案してきたユビキタス個人書斎をベースに、和歌データベースの拡張・継承に繋ぐ和歌の共有・鑑賞・創作を統合的に支援するWAKA-UPSを提案し構築する。和歌の中のキーワード(季語・地名)と、地理情報や個人嗜好による推薦等のITサービスと統合し、和歌鑑賞のスマートフォンアプリケーションを作成する。これを元に、和歌の鑑賞・意味づけ・自作投稿・共有ができるシェアリングサークルを構築する。このようなサークルが和歌の参加型学習、更にデータベースの拡張にも繋がると考えられる。また、地理情報による観光地の情報推薦も考えられる。そのため、和歌に出る地名は歴史上何らかの根拠で有名であると推測される。GPSの地理情報との連携を利用し、それをその土地を訪ねる観光者に適合した和歌を推薦することも可能となる。

---

[3] http://www.toyo.ac.jp/issyu/

[4] http://event.dai-ichi-life.co.jp/company/senryu/

WAKA-UPS システムには、利用者に各自特色のあるユビキタス個人書斎(i-Portal)を持たせ、そしてソーシャルメディアとの連携により、ユビキタス個人書斎の間を繋ぐと同時に和歌創作のログ情報を利用して情報を伝達することを含めた V-Portal として関連情報を公開する。このようなシステムにより伝統文化の伝承・拡張、人文科学知識の普及に貢献できると考えている。

現代のテクノロジを用いて、昔ながらの共同体(コミュニティ)的な環境を提供することができるユビキタス個人書斎システムを用いることによって、和歌コミュニティの組織者に、コミュニティで和歌を創作する環境を提供することが可能となる。

(1) UPS による個人作品の一括管理

ユビキタス個人書斎は、従来の部屋としての物理的書斎をユビキタス環境において仮想的書斎として提案したものである[3][4]。ユビキタス個人書斎は、図 1 に示したように、仮想コンテンツの V-Book と、V-Bookshelf、 V-Desktop、V-Note といったメタファと User Profile で構成される。

図1 ユビキタス個人書斎の構成

ユビキタス個人書斎研究のねらいは、利用者をさまざまな Web サイトの間を渡り歩きながらの情報活動から、再び「個人書斎」というメタファに個人が必要とする個人に適合した知識情報を集中させることである(図2)。個人に適した個人化環境で自分らしい知識作業をするという従来のスタイルを持ちながら、情報の共有、推薦と活用が容易に行える情報環境の構築を目指すものである。さらにソーシャルメディアにおける情報の収集、集計、分析、そして価値ある知識を掘り出し、さまざまなデータと情報を再利用可能なコンテンツとして組織化する。

個人の作品を V-Book として各自のユビキタス個人書斎に集中させれば、個人の独特なデザインにより作品をより一層表現できるだけではなく、様々のメリットをもたらすことが可能である。

図2 UPS で見たい見せたい情報を一箇所にまとめる

(2) Linked Data による和歌作品情報表現

和歌作品情報として、明示的な歌の種類、タイトル、本文と、歌の構造、つまり和歌の固定書式(五・七・五、...)、そして歌の本体にはないが、歌の特徴(ここでは歌の属性という)、そしてコメント、引用などの参照情報が含まれる。

ユビキタス個人書斎に登録する和歌を共有するため、これら情報を Linked Data [13][14]の形でデータを加工する。Linked Data とは、バーナーズ・リーが提唱した、人間だけではなく、機械でも読みこめるデータを作ることである。Linked Data の利点は、RDF などセマンティック Web 技術を利用して、相互運用性の問題を解決し、Web 上のデータをより簡単に再利用できるようになることである。Linked Data では、Web 上の情報が独自の ID を使用する代わりに、ドメイン・ネーミング・システムを使うことで容易に作成することができる。例えば、松尾芭蕉に関する情報をユビキタス個人書斎で提供している場合、

http://haiku.upscloud.net/people/MATSUO-Basho

という ID を使用すれば、対象となる松尾芭蕉を簡単に識別できるようになる。このような形式の ID は、HTTP URI と呼ばれる。またユーザが

http:// haiku.upscloud.net/people/MATSUO-Basho

にアクセスすると、松尾芭蕉の名前、オンライン・アカウント、拠点とする場所、作品一覧などの情報が表示される。

和歌作品情報を Linked Data として表現する利点として、複数の機関や個人が作成したデータを統合できることが挙げられる。和歌を Linked Data で表現するこ

とにより、友人のユビキタス個人書斎から和歌データの取得、属性付け、コメント追加などが可能となり、和歌の分散型データベースを簡単に作成することができる。

作品をユビキタス個人書斎に登録し、Linked Data として公開するもう一つのメリットは、外部からデータのリンクが多いほど、作者ユーザの地位が確立し、作品のクローンができなくなり、著作権の保護になる。

(3) 創作ログの収集

和歌に教養ある人間の高齢化が進んでおり、彼らが持っている和歌修養を後世に残すことが重要である。これらの修養は大部分暗黙知として、本人の脳に存在し、専門家でない限り、これらの知識を形式知にまとめるには難がある。

ユビキタス個人書斎は、スマートフォン、ソーシャルメディア、クラウド技術を総合的に活用することによって、個人の創作活動に関連するログを収集し、蓄積することができる。創作ログには、例えば個人の位置情報を活用し、感慨を発する場所を記録し、感慨する音声を録音し、またソーシャルメディアからその人と周りの友人とのやり取りも同時に収集することが可能である。それにより、作品を理解するため、創作ログをたどり、和歌の愛好家の足跡を追跡し、より深く理解することが可能になる。これらの創作ログのデータが多いほど、コンピュータによる作品解析の手掛りが多くなり、解析の精度向上につながる。

(4) ユーザフィードバックによる意味付け

ユビキタス個人書斎をベースにした Waka-UPS システムは、ユーザのフィードバックによって和歌作品の属性を豊かにすることが期待できる。和歌作品の属性の一部は、次のようなもの組み合わせで構成すると考えられる。

季節(春・夏・秋・冬)
自然(雨・風・雲・雪、・・・)
状況(野遊び・国見・行幸・宴席・うわさ・みやげ・夢・七夕、・・・)
感情(恋心・悲しみ・喜び・怒り・嘲笑・恐れ・望郷・寂しさ、・・・)
比喩(月、・・・)

和歌作品に対して、特有の季節、自然、状況、感情、比喩について、利用者が意見を述べたり、投票したりすることができるようなユーザインターフェースを作り、複数の利用者のフィードバックから作品の属性を作り出すことが可能となる。和歌作品について、コミュニティのメンバが感想などのコメントを書き込み、作者または他の読者はそのコメントに対してコメントまたは単純に「いいね」で意見を述べたりすることによって作品に対する評価を増幅させることが可能である。作者または他のユーザから、作品に似合う画像ファイルの投稿もでき、「いいね」の数により、和歌作品に人気の画像を選ぶこともできる。

(5) ユーザ間の情報共有とノウハウ伝達

ユビキタス個人書斎は、さらに情報推薦、知識共有、共同作業、そして意見や評価の傾向分析、協調学習におけるグループの形成、暗黙知の形式知への変換などの知識作業を支援することができる。それによって、短歌、俳句、川柳など和歌の共有・鑑賞、和歌の創作・投稿、和歌の自作や鑑賞における意味づけなどに関連する情報(季語や地名など)の検索、個人の嗜好を考慮した情報推薦が可能である。

図 3 はSECIモデル理論[15]を用いて、ユーザ間の情報共有と和歌のノウハウ伝達のフローを示す。それにより、和歌に関するノウハウの伝達や習得を支援する。

1. 共同化(Socialization):システムの利用者がユビキタス個人書斎にある素材(和歌・他の投稿)を V-Desktop で各々鑑賞・意味付けをする。この作業は更に個人の感想、コメントや作品等を考えるきっかけにもなる。それは共同化だと考えてもいい。知識伝達の視点から、このプロセスは利用者が持つ生活や和歌の経験が作品に共感を持たせる状態になる。
2. 表出化(Extermalization):利用者が考え出した内容を個人のポータル(UPS の V-Portal)に意味付け、または繋ぎ創作(連歌)、再創作投稿をすることにより、参加者各々の作品(V-Book)の属性拡張に繋がる。知識伝達の視点から、知識を表出化することが実現できる。
3. 連結化(Combination): Cross SNS [4]を利用した個人のポータル(UPS の V-Portal)上での和歌投稿なので、ユーザ間の情報共有が容易にできる。それに、ユビキタス個人書斎における情報(作品とその属性)検索、情報(作品とその属性)推薦機能を加えれば、知識の結合、連結化もユーザ間の相互コメント・参照により達成できると考えられる。

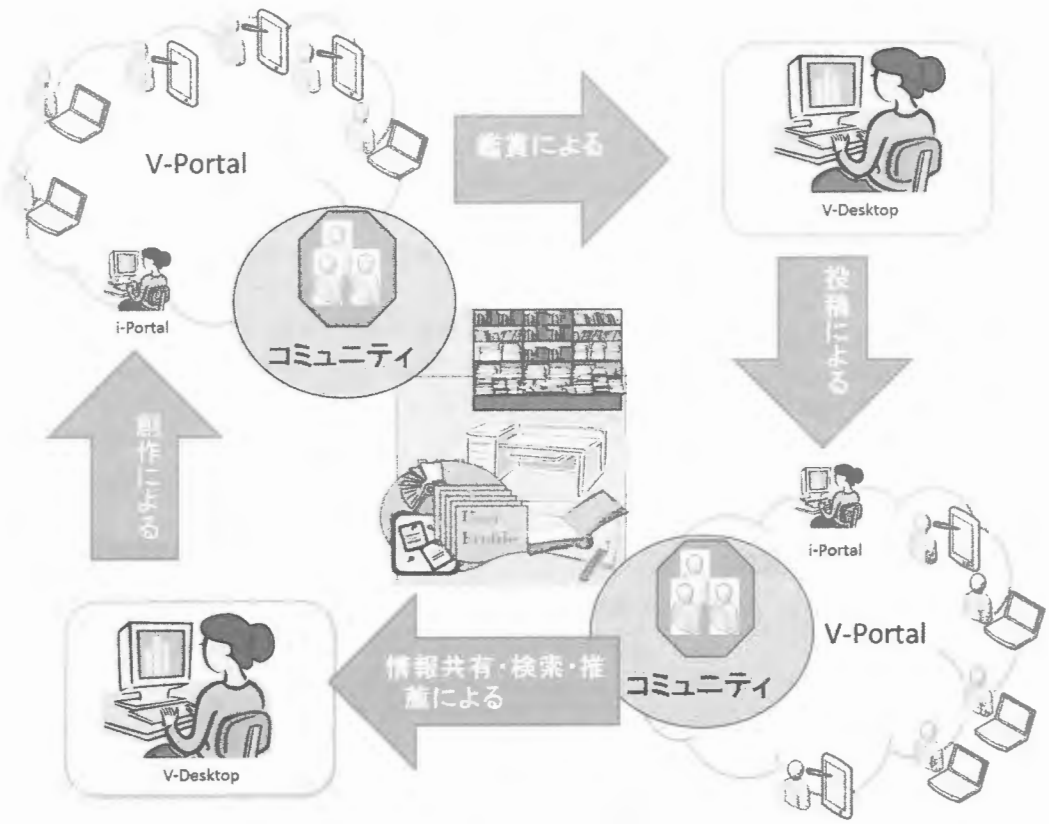


図3 SECIモデルに基づく情報共有・ノウハウ伝達のフロー

4. 内面化(Intermalization): 連結化されたノウハウを次の作品創作と投稿に生せることにより、知識の内面化が行われ、個人の創作経験に変換されることとなる。

上記のプロセスを繰り返すことによって、和歌鑑賞、ノウハウ伝達や作品属性の拡張等多目的の達成が可能である。

(6) オンラインコミュニティ形成の支援

ユビキタス個人書斎の間には、すでに連携する仕組みが備えている。それによって和歌愛好会のオンラインコミュニティ形成を統合的に支援することが可能となる。具体的には、Cross SNS [4]を利用して複数のユビキタス個人書斎をあたかも一つの V-Portal として機能し、さらにソーシャルメディアとの連携により作者と読者の間の交流を促進したり、貢献度によるユーザ・ランキングなどの仕組みによって積極的な参加を促進したりすることができる。また、和歌愛好会のオンラインコミュニティにおいて、和歌の特徴を生かした連歌のようなサービスを提供することも可能となる。

このようにして、本研究で提案しているユビキタス個人書斎 Waka-UPS は、和歌の共有・鑑賞・創作を統合的に支援する個人化情報ポータルとして、和歌に関わる情報アクセスの一括管理を可能にし、情報の推薦・共用活用を支援し、促進することが期待できる。

4. UPS を用いた実装

和歌の共有・鑑賞・創作を統合的に支援することに特化した Waka-UPS のプロトタイプシステムは、これまでの先行研究で構築した UPS 本体に、いくつかの和歌関連モジュールおよびスマートフォンクライアントモジュールを追加して構築している。

和歌愛好者のユーザの特徴を考慮し、堅苦しい検索画面ではなく、ユーザにさまざまな選択肢を与え、ユーザの選択から学習して、より正確にユーザの嗜好を掴むアプローチでシステムを構成している。こうしたデザインにより、ユーザ負担を減らすことが可能なユーザ中心のシステムを構築する。

Waka-UPS のプロトタイプシステムは、ユビキタス個人書斎本体と、形態素解析による和歌属性の自動抽出、ユーザインタラクションによる和歌属性の生成、和歌を Linked Data 形式での公開、和歌 Twitter との連

携など和歌関連モジュールと、和歌スマートフォンクライアントモジュールから構成される。

(1) UPS 本体

ユビキタス個人書斎の基盤システムはユビキタスとクラウドコンピューティング環境を考慮した三層レイヤーの設計になっている。三層レイヤーはクラウドコンピューティングに対応する UPS Cloud Layer、中核の UPS Application Layer (UPS Portal)、そして、ユビキタス環境に対応する UPS Client Layer から構成されている。各レイヤーについては前述の UPS の必要な機能をモジュール化されている(詳細については参考文献[2]を参照されたい)。

(2)和歌関連モジュールの追加

Waka-UPS の試作システムは、下記の四つの和歌関連モジュールが追加されている。

形態素解析による和歌属性の自動抽出
ユーザインタラクションによる和歌属性の生成
和歌を Linked Data 形式での公開
和歌 Twitter との連携

それぞれのモジュールはプラグインとして USP 本体にインストールし、有効化(Activate)して利用する。

(3) Linked Data による和歌知識ベース

Linked Data 形式で和歌を公開するモジュールにおいては、和歌データを Linked Data として活用するためのサービスを提供し、和歌のデータセットに含まれるアイテムへの URI 付与を行うとともに、Linked Data の語彙と URI マネジメントの指針に従って、Linked Data の語彙による和歌のデータ表現を行う。和歌を Linked Data として記述し、表現することにより、和歌に関わるさまざまな知識が蓄積され、あたかもインターネットにおいて巨大な分散型和歌知識ベースを構築される。

(4) 和歌スマートフォンクライアント

和歌スマートフォンクライアントモジュールは、WAKA-UPS の API と、Android の Waka アプリで構成されている。スマートフォンクライアントモジュールにはローカルリポジトリとローカル V-Log が別々に存在し、ネットワークに接続されないオフラインの状態でも手元にある和歌データの読み書きをしたり、和歌作品にコメントを付けたりすることが可能である。

スマートフォンクライアントモジュールが同期された状態では、UPS の V-Desktop、つまり「仮想書卓」上の和歌作品がすべて Android のローカルリポジトリとローカル V-Log に保存される。和歌愛好会オンラインコミュニティのオフ会、または、和歌を歌う集まりへ出かける際、仮にネットワークがつながらなくても、Android 携帯端末に蓄積されている情報を利用することができる。

たとえば、GPS の地理情報を利用して、旅行先へ出かける途中地点にある和歌名所、その時の季節、気候に関連する和歌を提示することができる。V-Desktop から和歌およびその要素に関連するキーワードの検索も可能である。さらに、和歌を鑑賞した感想はテキストまたは音声で記録することもできる。再びインターネットに接続する状態になると、ローカルリポジトリとローカル V-Log に保存されている一時データは、WAKA-UPS システムに自動同期される。

このように旅をしながら記録される鑑賞・創作ログは、本人が後日旅を振り返る時にも利用でき、また友人にバーチャル旅を体験させることも可能である。

(5) 試作システムの構成と試用シナリオ

試作システムは、図 4 に示したように、二つの Waka-UPS を設け、それぞれ「小倉百人一首」と「万葉集」を初期データとして利用される。そして Android の Waka スマートフォンクライアントモジュールでユーザ用クライアント環境を構築する。

和歌愛好者が Waka-UPS 試作システムを使用するシナリオは以下のようなことが想定される。

- スマートフォン利用者が個々のユビキタス個人書斎において、Linked Data を利用して検索し、和歌を鑑賞する。
- 和歌投稿と鑑賞コメントなどスマートフォン利用者からのフィードバックは、ユビキタス個人書斎に蓄積する。
- スマートフォン利用者は、Twitter を利用して和歌とその要素を投稿する。
- ネットワークに接続されないオフライン状態でも利用者が Waka-UPS の V-Desktop を利用する。
- Waka-UPS システムは、利用者のフィードバックから、利用者の嗜好を推測し、季語など和歌要素を推薦し提示する。
- Waka-UPS システムは、利用者のフィードバックと好みから、ユーザグループの形成を提案し、オンラインコミュニティの形成を支援し促進する。

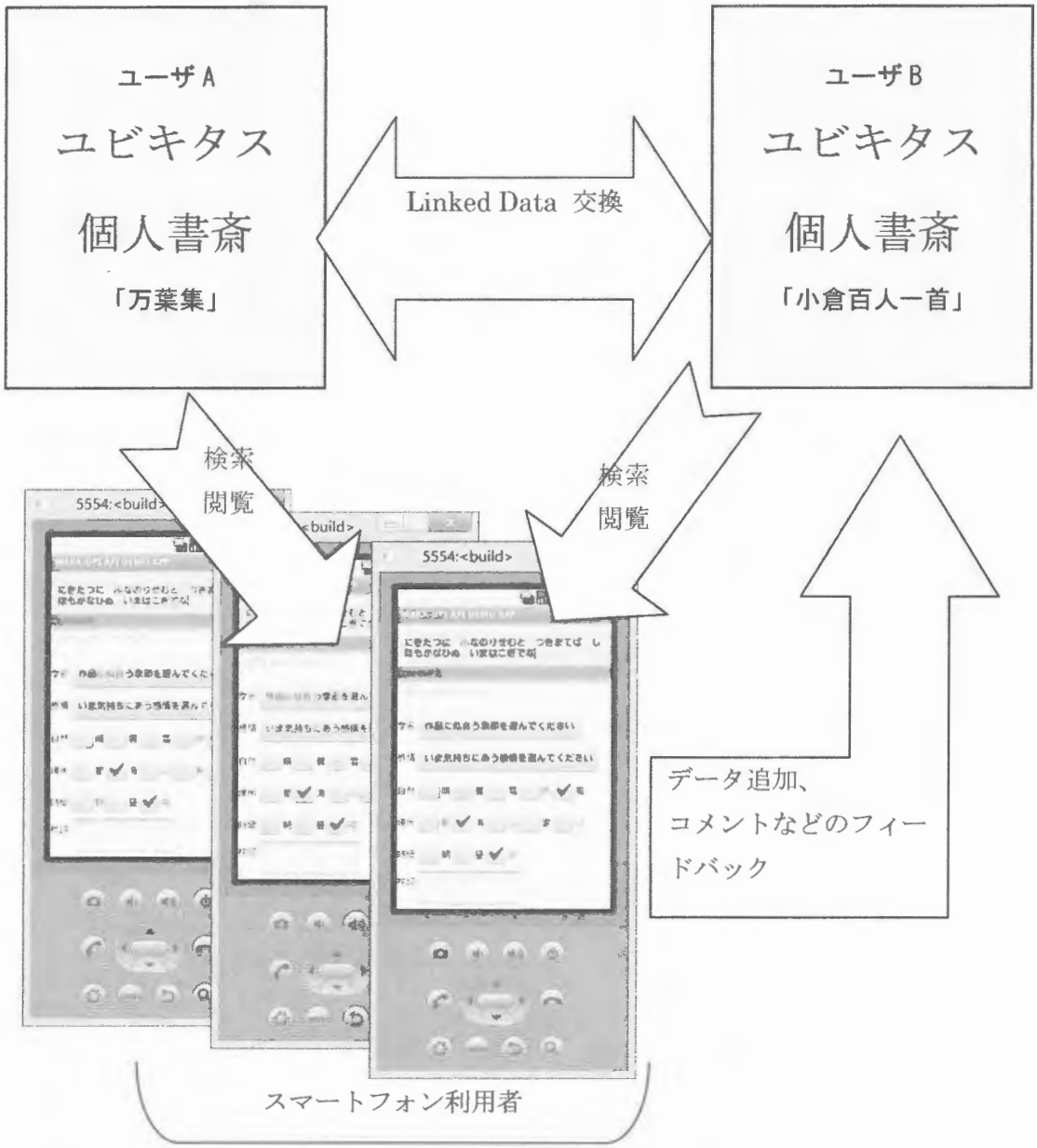


図4 UPS と Android の Waka スマートフォンモジュールで構成される試作システム

5. おわりに

本研究では、季語や自然の描写など和歌の共有、鑑賞、創作、意味づけに欠かせない情報をキーワードとした検索、個人嗜好を考慮した情報推薦、誰でも気軽に参加できる和歌愛好会のオンラインコミュニティ形成を統合的に支援するシステムを提案している。先行研究で構築されているユビキタス個人書斎(UPS)をベースに、和歌の共有、鑑賞、創作に特化したプロトタイプシステム Waka-UPS を構築し、Waka-UPS が必要とする追加モジュールの機能を実装した。さらに、ソーシャルメディアツール Twitter との連携を行ったうえ、スマートフォンで利用する Android アプリを構築した。

WAKA-UPS システムは、和歌愛好者に各自特色のあるユビキタス個人書斎(UPS)を持たせ、そして Cross SNS といった UPS 間の連携を可能とする仕組みにより、和歌の共有だけではなく、鑑賞に関わる感情の共有、和歌創作プロセスの共有も可能となり、それによって、伝統文化の伝承・拡張や人文科学知識の普及を促進することが期待される。

今後の課題として、試作システムの試用と評価実験を行い、ユーザの負担をさらに減らすシステムの使いやすさなどの工夫、システムを利用する動機付けなどさまざまな機能改善をしていきたい。

参考文献

1. H. Chen and Q. Jin, "Ubiquitous Personal Study: A Framework for Supporting Information Access and Sharing," Journal of Personal and Ubiquitous Computing, Vol. 13, No. 7 (Oct. 2009).
2. H. Chen, X. Zhou and Q. Jin, "Socialized Ubiquitous Personal Study: Toward an Individualized Information Portal," Journal of Computer and System Sciences, Vo. 78, No. 6, pp.1775–1792 (Nov. 2012).
3. Q. Jin, H. Chen and R.Y. Shtykh, "User-Initiated Ubiquitous Personal Study under the Ecologically Integrated Framework of Information Environments," *(Poster)*, Intel CSU Transparent Computing and Platform Innovation Summit, Changsha, China (Oct. 2012).
4. 陳、竹井、張、金、"Cross SNS を活用した情報アクセス共有支援環境の提案"、グループウェアとネットワークサービスワークショップ2006論文集 (2006)。
5. 大井,「共同体における認識過程の基礎情報学の分析—俳句を事例として」, 情報処理学会研究報告, 2008-EIP-42 (11) (2008)。
6. 勝倉,「上田秋成の和歌と歌枕(上)」, 福島大学教育学部論集第 40 号 (1986)。
7. 墨岡, 井上, 和田, 田中, Bogdan,「英語俳句サイト Shiki の奇跡—Shiki Team 年代記」, 情報処理学会創立 50 周年記念(第 72 回)全国大会 (2010)。
8. 吉岡,「季語データベースの構築と俳句の季語の自動判定の試み」, 人文科学とコンピュータ, 48-8 (2000)。
9. 吉岡,「季語データベースの構築と俳句投稿鑑賞システムの概要」, 情報処理学会研究報告, 2006-CH-71 (4) (2006)。
10. 湊, 尾内,「俳句の適合する合成画像を生成するシステムの検討」, 映像情報メディア学会技術報告(ITE Technical Report), Vol. 33, No. 44, pp. 43-46 (2009)。
11. 浅見, 高田,「スマートフォンを活用した俳句支援アプリケーションの開発」, 映像情報メディア学会技術報告 (2012)。
12. M. Suzuki, Y. Kobayashi, T. Nakai and K. Yoshida, "Analysis on Empathy-inducing Effect brought by Haiku," IEICE TRANS. INF. & SYST., Vol. E89-D, No. 6 (2006)
13. T. Berners-Lee, Design Issues, Linked Data, http://www.w3.org/DesignIssues/LinkedData.html
14. C. Bizer, T. Heath and T. Berners-Lee, "Linked data - the story so far," Journal on Semantic Web and Information Systems, Special Issue on Linked Data (2009).
15. 野中,「知識創造企業」, 平成11 年 12 月 12 日和敬塾予餞会紀念講演 (1999)。

アパブランシャ語コーパスを用いた韻律による年代推定
Estimating Chronology of Apabhramśa Texts from Meters

山畑 倫志
Tomoyuki Yamahata
北海道大学大学院 文学研究科, 北海道札幌市北区北 7 条西 5 丁目
Hokkaido University Graduate School of Letters, Kita 10, Nishi 7, Kita-ku, Sapporo

あらまし:アパブランシャ語はインド語派の中期インド語に属する言語である。中期インド語から近代インド諸語へ至る過程は未だ不明瞭であり、その間アパブランシャ語はその解明のために重要である。しかし長期間広い地域で使用されたため、言語的に多様な変異を含み、言語事象のみでの分析は困難である。そこで文献がすべて韻文であることを利用し、DB 化した作品群をもとに各詩節と韻律の発展過程を分析し、一定の分類基準を見出す。

Summary: Apabhraṃśa language is a one of Middle Indic Languages. Though Apabhramsa is important for study of Historical Indic Linguistics, there are plenty of different words, morphological forms and meters. This paper applied the Decision Tree classifier and the Naive Bayes classifier to Apabhraṃśa corpus for discovering the conditions of these differences. Therefore, we found a new clue for classification of Apabhramśa language.

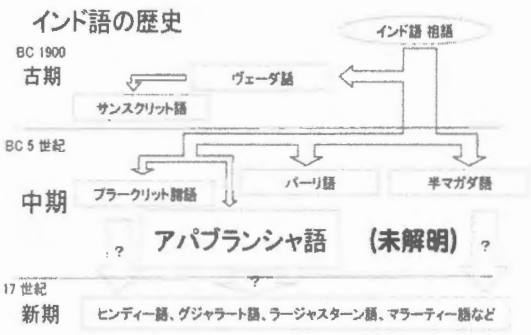
キーワード:データベース, コーパス, アパブランシャ語、文書分類、機械学習
Keywords: Database, Corpus, Apabhraṃśa, Document Classification, Machine Learning

1. アパブランシャ語の概要

本稿はインド・ヨーロッパ語に属するアパブランシャ語についてコーパスと韻律情報を用いて年代推定を目的とする。まずアパブランシャ語の歴史的位置づけ、使用状況および言語特徴について概説する。

インド語(Indic, Indo-Aryan)は紀元前 1900 年ごろからその資料が入手可能な言語であり、またその後も絶え間なく言語資料を作り出し、伝えてきた言語である。そのため言語がどのように変化するのかを探る上で重要な言語である。本発表でとりあげるアパブランシャ語(Apabhraṃśa)はインド語の歴史の中では中期インド語の最新層に分類される言語である。(図表1参照)

中期インド語には多くの言語が含まれ、多様な文献が多く残されている。しかし、それらの歴史的および地域的な整理についてはそれ以前の古期インド語や、ヒンディー語をはじめとした新期インド語と比べ、あまりすすんでいないのが現状である。特に現在インド各地で使用されている新期インド諸語とどのような関係があるのかがいまだ不明なままである。その原因として各言語における言語特徴が錯綜しており、それぞれの地域

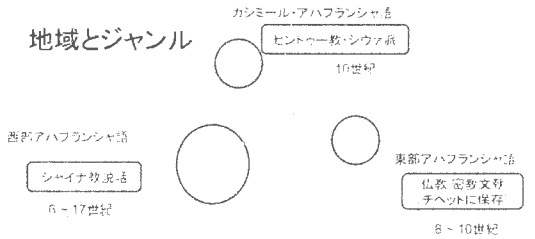


図表 1 インド語の歴史

における新期インド語と簡単に関連づけられないことがある。そのような状況の中で中期インド語の中でももっとも新しいとされるアパブランシャ語は新期インド語への展開を考える上で重要な言語である。

アパブランシャ語は紀元六世紀ごろから十七世紀頃までインド西部を中心に北インドの各地で用いられた言語である。言語名である「アパブランシャ」はサンスクリット語で「堕落した」を意味することから、元来は一地方語であり、サンスクリット語およびマハーラーシュトラ語やシューラセーナ語といった早くから文章語と

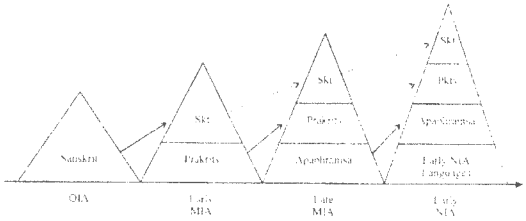
して高い地位にあった言語とは異なった扱いを受けていたことが推測できる。それが具体的にどのような経緯によって文章語となったかは未だ定かではないが、徐々宗教詩や宗教説話の分野での使用が徐々に広がり、十二世紀に至るとグジャラート地方のジャイナ教徒ヘーマチャンドラ(Hemacandra)により文法や詩作法に関する作品が著され、一定程度の規範化がなされるまでになった。その後インド西部においてジャイナ教徒が中心となって類似した形式で多くの説話作品が作られた。それに対しヘーマチャンドラ以前から相当数の作品がアパブランシャ語で作られている。そのため、言語特徴上多くの差異を含むテキスト群が残っている。それらの言語特徴はその作成地域にもとづき図表2のように分類可能である。



図表2アパブランシ ・語の地域とジャンル

この分類は新期インド語との関連を示唆するため、同様の分類から各地のインド語との関連づける試みがなされてきた。[1]しかし、これまでの研究により、そのような地域的な特徴と結びつけることは難しいことがわかってきた。むしろある特定地域の言語が一時的に北インドの広い地域で使われた時代が存在したと考えるのが妥当である。実際に使用されてきた時期の長さ、制作された作品の多さ、また文学スタイルの伝統が類似を考えるとインド西部地域の言語であった可能性が高い。そうした場合、アパブランシャ語に含まれる差異を地域性とは別の形で説明する必要が出てくる。本稿ではそれを文章語としての定型化の過程の中で生じたものと捉え、ヘーマチャンドラによる規範化以前に各地での流行が終わってしまったため、インド西部以外の地域では規範化の途上の特徴を有していると解釈する。Bubeník(1998)はアパブランシャ語が他の諸言語とともに文章語として採用されていく過程を図表3のように示している。

次にアパブランシャ語の言語特徴であるが、インド語の歴史は大まかに言ってサンスクリット語のように複雑な語尾変化で単語間の関係を示す方法から、語尾



図表3インド文章語の展開(Bubeník 1998)

(OIA: 古期インド語、MIA: 中期インド語、NIA新規インド語)

変化を単純化し語順と後置詞によって文法機能を表すという方法に大きく変化した。アパブランシャ語はその過程のちょうど中間に位置するため、どちらの特徴も有している。Bubeník(1998)はインド語の文法構造における歴史的変化という枠組みの中でアパブランシャ語を分析している。そこで示される特徴は次の四点である。

I 形態的に示される格語尾が減少し、また単純化されているため、格標示のための形式がほとんどの語幹に共通している。そのため後置詞とみなしうる語が登場してきている。

II 未完了、アオリスト、完了を一語の活用で表す定動詞の形式が完全に消滅する。それにより過去を表現する形式が過去受動分詞のみになる。

III 本来は「立つ」や「ある」を意味する動詞が助動詞的な役割を持ち、それが現在分詞や過去分詞と共に使われて進行や完了などのアスペクトを表している。

IV 過去の表現が過去受動分詞のみになった結果、完了表現における能格構造の成立が見られる。つまり、自動詞主語と他動詞目的語が同じ形式を取り、他動詞主語が独自の形を取っている。

上記をまとめると一つの文法的機能を複数の語で示すいわゆる孤立語的な傾向を示している。それに加えてIVでは能格構造の出現が見られる。屈折型から孤立型へ、そして主格構造から能格構造という大きな変化の途上にあるのである。このような中間的な性格を持つ言語を分析し、他の典型的な構造を持つインド語以外の諸言語との比較により、言語構造一般の理解にとって得るところの多いものとなる。上に挙げた諸特徴は全て近代インド諸語と共通する。しかし、動詞の活用の体系が存続しているなど、古い屈折的な性格も持つ。インド語史の観点からみれば、まさに中間的特徴を有する言語である。

[1] Tagare (1948)

2. アパブランシャ語コーパス作成の試み

本稿で用いたコーパスは以下のテキストから構成されている。便宜的に図表2で示した地域区分ごとに示す。また制作年代を付しているが、インドの古典文献一般の特徴として作品の年代がはっきりとわからないことが多い。この中で年代推定の確度が高いのは当時の碑文から確認が取れるヘーマチャンドラのみであり、ほかは相対的な年代推定に基づくものがほとんどである。

西部アパブランシャ語

『ヴィクラム王とウルヴァシー』（部分）
カーリダーサ作、6世紀?
古典サンスクリット文学の代表的な詩人であるカーリダーサの手になる戯曲作品。主人公のヴィクラム王が悲しみで我を忘れた際の台詞がアパブランシャ語である。アパブランシャ語の最古の用例となるが、その部分が欠ける写本の系統も存在するため、真偽について意見が割れている。

『パドマの行跡（*Paumacariu*）』（PC）
スヴァヤンブー作、9〜10世紀
インドの代表的な叙事詩である『ラーマーヤナ』はバラモンあるいはヒンドゥーの価値観で書かれた物語であるが、それをジャイナ教の教義に合わせて再構成した作品。De Clercq（2003）により電子テキストが提供されているため、それを用いてコーパスを作成。

『ハリシェーナの行跡（*Hariseṇacariu*）』10世紀以降?
ジャイナ教説話では六十三偉人という伝説上の人物の伝記を大枠としてその中に種々雑多な題材を含ませる形式の作品が多い。この作品もその偉人の一人の伝記の形式をとった作品である。

『サナトクマーラの行跡（*Sanatkumāracarita*）』（部分）
ヘーマチャンドラ、12世紀
アパブランシャ語の規範化を進めたヘーマチャンドラによる作品。全体としてはサンスクリット語の作品だが、アパブランシャ語部分も存在する。

東部アパブランシャ語

『ドーハーの宝庫（*Dohākoṣa*）』
カーンハ作、7〜12世紀?
『ドーハーの宝庫』
サラハ作、11〜12世紀?
これら東部アパブランシャ語作品はすべて仏教徒によるものであり、密教の教義を伝える韻文となっている。

カシミール・アパブランシャ語

『タントラ綱要（*Tantrasāra*）』（部分）
アビナヴァグプタ作、10〜11世紀
同著者の宗教理論書である『タントラ・アーローカ』の綱要であり、各章の末尾がアパブランシャ語の韻文となっている。

『聖典騒動（*Āgamadambara*）』（部分）
ジャヤンタバッタ作、9〜10世紀?
戯曲作品。正統バラモンを自認する主人公が他の宗教宗派を攻撃する作品。登場人物のうちニーラーンバラという宗派の行者がアパブランシャ語の韻文を歌いながら登場する。

以上8種のテキストを入力し、単語単位に分割し、コーパスとして利用している。インド古典語の言語処理においてしばしば問題となる連声（sandhi）[2]の問題であるが、アパブランシャ語では原則連声の現象は消失しているため、その問題は生じない。また複数の名詞を複合語として圧縮し句や文に近い機能をもたせる複合語生成もインド古典語に特有の現象だが、作成したコーパスでは標を付与した上で一語として数えている。

地域	トークン	タイプ	タイトル	トークン	タイプ
東部	3032	1718	DKK	444	319
			DKS	2588	1399
西部	111815	39806	HC	5439	2942
			PC	98651	32208
			SC	7548	4528
			VU	177	128
カシ	508	433	AD	106	90
			TS	402	343
合計	115355	41957			

図表4 各テキストの語数

[2] 語末の音と次の語の語頭の音がそれぞれの音の種類により変化して結合する現象。表記上一体化してしまい区分できない場合もある。

4. アパブランシャ語の地域差と格語尾

アパブランシャ語の中でも特に名詞の格機能を示す曲用語尾は多様な形式を有している。規範化を試みたヘーマチャンドラもその文法書『言語概説(*Śabdānuśāsana*)』では様々な形式を並列している。たとえば名詞の中でもっとも数の多いa 語幹男性名詞は8.4.331—339, 342, 346, 347で規定され、図表5のようになる。

	単数	複数
主格	deva, devā, devu, devo	deva, devā
対格	deva, devā, devu	deva, devā
具格	deveṇa, deveṇaṃ, deveṃ	devahiṃ, devāhiṃ, devehiṃ
属与格	deva, devā, devasu, devāsu, devaho, devāho, devassu	deva, devā, devahaṃ, devāhaṃ
奪格	devahe, devāhe, devahu, devāhu	devahuṃ, devāhuṃ
所格	devi, deve	devahiṃ, devāhiṃ

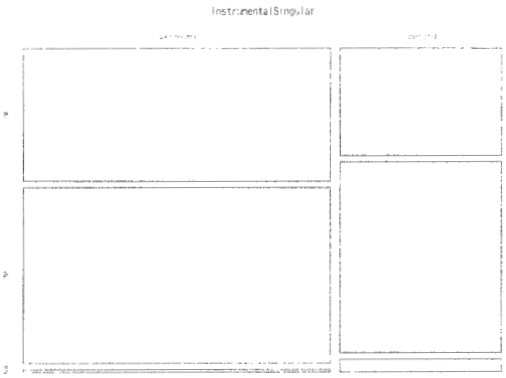
図表5 a語幹男性名詞の変化表

このように同じ機能に複数の語形が含まれることが多々見られ、同一作品内であっても複数出現することがある。そのため、言語特徴を分析する上で、基準となる形式が何であり、どういった条件で使い分けられているのかを突き止める必要がでてくる。

山畑(2011)では韻律の制限のために同一作品においても異なる形式が出現するのではないのかと考え、調査した。コーパス中『パドマの行跡』に対象を絞り、分析を行った。韻律は特に各句の末尾に制限をかけることが多いため、格を示す語尾形式の中でも特徴的なものを用いて韻律との関係を探った。その数は図表5のようになる。属与格と奪格を一つにまとめているため、ヘーマチャンドラの解釈とは異なるが、校訂者のDe Clercqの判断を優先した。図表5に見られる語尾形式のうち*-āṇa*, *-mmi*はごく少数のため無視することにしても、具格単数、与属奪格単数、与属奪格複数においては韻律の制限によって違いが出ているように思える。これらを図示すると、図表7から図表9のようになる。

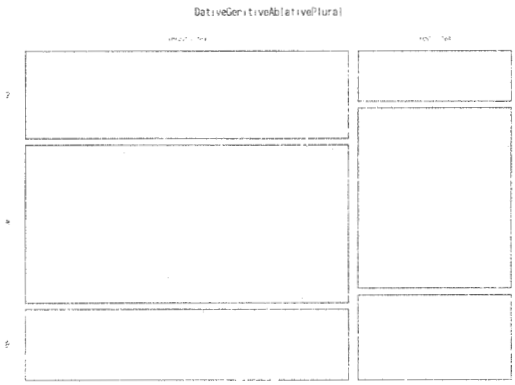
		制限無し	制限有り
具格単数	eṃ	1539	653
	eṇa	2011	1150
	eṇaṃ	30	80
属与奪格単数	ho	2188	911
	su	254	170
	ssa	55	2
具格複数	ehiṃ	1512	845
属与奪格複数	hu	343	94
	huṃ	619	335
	haṃ	278	158
	āṇa	9	20
所格単数	mmi	8	0
	ahiṃ	332	54

図表6 韻律の制限と格語尾



図表7 韻律制限の有無にもとづく具格単数の頻度

図表8 韻律制限の有無にもとづく与属奪格単数の頻度



図表9 韻律制限の有無にもとづ奪属与格複数の頻度

これらの図から韻律の制限がかかることによって次のような影響が出ていることが観察される。

具格単数 *-eṃ* に対して*-eṇa*, *-eṇaṃ* が優勢。
与属奪格単数 *-ho*, *-su* に対して*-ssa* が優勢。
与属奪格複数 *-hu* に対して*-huṃ*, *-haṃ* が優勢。

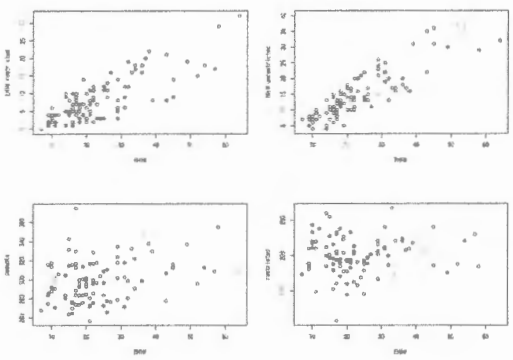
ただ、上記のような表現では全体数から判断されるのみで分布の様態までは表現できない。そこで山畑(2011)に加えてテキスト全体をいくつかのかたまりに分割して、それらのかたまりの性質がどのように分布するかを検討した。

『パドマの行跡』は複合語も一語と考えると総語数98875である。これを便宜上100個の語連続に分け、おおよそ1000語ごとのかたまりを作り、それぞれの中で語尾形式や韻律などの出現をカウントし、その分布との関係を調べた。

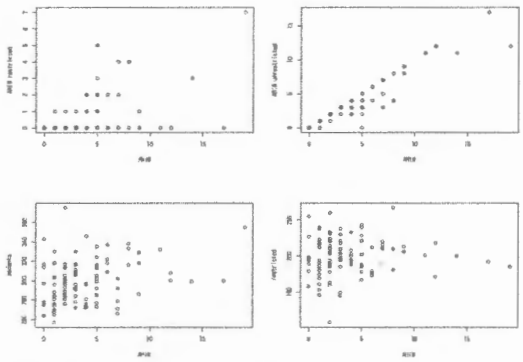
上記で併用される複数の格語尾が検討されなかった具格複数*-ehiṃ*、および所格*-ahiṃ* は比較対象がないため、図表5では出現数の多寡で検討しうるだけだが、テキスト全体を分割することにより、その分布状況がわかる。図表10と11はそれを図示したものである。横軸には一つのかたまりにおけるそれぞれの語尾形式の出現数、縦軸にはそれと関係があると思われる要素をとっている。4枚の図のうち下の二枚はそれぞれ脚末の語数と韻律制限を受ける語数との関係を示している。*-ehiṃ* も*-ahiṃ* もその二要素とは関係が見られない。上二枚はそれぞれの形式のうち、韻律による制限の有無によって分け、それぞれの出現数をみたものである。すでに下の二枚から脚末という要素や制限という要素のみとは相関がないことがわかっているため、それぞれの語尾形式における制限の有無という観点から検討することができる。*-ehiṃ* については制限の有無にかかわらず、正の相関をしめしている。それに対して*-ahiṃ* は制限なしの場合には正の相関を示すが、制限ありの場合は相関を示しているとはいいがたい。

図表10と11から言えることは*-ehiṃ* はどちらの場合でも同様に分布するが、*-ahiṃ* は制限なしの場合のみ、全体の数に呼応しており、制限ありの場合にかなりランダムな出現をするということである。

このように韻律と語尾形式の間には一定の関係があることがわかる。そのため韻律情報を言語分析に利用することによってより詳細な情報を得られる可能性がある。



図表10 具格複数*-ehiṃ* と他要素との関係
(左上：*-ehiṃ* と制限付き*-ehim*、右上：*-ehiṃ* と制限なし*-ehim*、左下：*ehiṃ* と脚末語数、右下：*ehiṃ* と制限付き語数)



図表11 所格*-ahiṃ* と他要素との関係(それぞれの図の意味は図4.*-ehiṃ* と同様)

5. 学習による分類

現在コーパスに格納されているテキストについて現在存在するカテゴリー分類は未だ確定したものではない。そのため、何らかの形での検証が必要である。本稿では単純にテキストにおける単語の出現数に基づき、分類を試みる。手法として決定木およびナイーブベイズを利用した。本章では複合語を分割して利用する。

まず検証として『パドマの行跡』について構成の最小単位にあたるカダヴァカごとに文書を分割し、1414の文書を作成した。本作は全体で『ヴィドヤーダラ・カーンダ』、『アヨーディヤー・カーンダ』、『スンダラ・カーンダ』、『ユッダ・カーンダ』、『ウッタラ・カーンダ』の5部から構成されている。これについてまず決定木の手法により部別の学習を行った。200以上のカダヴァカに出現する語彙34語の出現数を因子として決定木を作成し、それによる自動分類を行った。

	予測				
実際	ayodhya	sundara	uttara	vidyadhara	yuddha
ayodhya	342	0	0	0	0
sundara	0	210	0	0	0
uttara	0	7	48	16	154
vidyadhara	0	6	35	63	198
yuddha	0	2	31	17	283

図表 12 決定木による『パドマの行跡』部別の学習と予測結果

これを見ると『パドマの行跡』においては『アヨーディヤー・カーンダ』と『スンダラ・カーンダ』は語彙から判別可能であるが冒頭の『ヴィドヤーダラ・カーンダ』、および後半部の『ユッダ・カーンダ』と『ウッタラ・カーンダ』では語彙からは判別しかねることがわかる。

さらに別の視点からの分類として『パドマの行跡』以外の7つのテキストをナイーブベイズの手法により学習させ、『パドマの行跡』の各カダヴァカがどの作品に分類されるかを見てみた。図表12がその結果である。結果からは時代や地域とは異なった分類が示唆される。『ハリセーナの行跡』と『ドーハーの宝庫（サラハ）』は地域的にはインドの西部と東部、宗教もジャイナ教と密教化した仏教というように異なる。また『ハリセーナの行跡』と『パドマの行跡』は同地域、同ジャンルで近いのは納得できるが、『サナトクマーラの行跡』も同種のものであるにもかかわらず、分類されたカダヴァカ数はかなり少ない。

タイトル	地域	時代	カダヴァカ分類数
HC	西部	10世紀以降?	507
DKS	東部	11〜12世紀?	495
AD	カシミール	9〜10世紀?	168
DKK	東部	7〜12世紀?	130
VU	西部	6世紀?	106
SC	西部	12世紀	5
TS	西部	10〜11世紀	4

図表 13 ナイーブベイズによって作成した分類によって、『パドマの行跡』（9〜10世紀）の各カダヴァカを配分した結果

この結果からは地域やジャンルよりもむしろ時代的な傾向が読み取れる。これはアパブランシャ語文献群内に存在する多様な差異が地域的な影響によるものではなく、文章語として整理されていった段階の違いによるものであるという考えに合致する。

このようにデータベースを利用することによって新たな視点からアパブランシャ語の分類を見直すことが可能となった。手法の適用方法をより洗練させることにより、さらに有益な結果がでてくることが予想される。

参考文献

De Clercq, Eva. 2003. "Een Kritische Studie Van Svayambhūdeva's Paümacariu."

A Historical Syntax of Late Middle Indo-Aryan (Apabhraṃśa). 1998. John Benjamins Publishing Company.

Tomoyuki Yamahata, 2012, The Classification of Apabhraṃśa —A Corpus-based Approach of the Study of Middle Indo-Aryan—', Corpus-based Analysis and Diachronic Linguistics, pp. 223-249.

山畑倫志、2012、「Suttanipāta の時代区分と韻律との関係」、『印度学仏教学研究』、日本印度学仏教学会、第60巻第2号、pp. 906-911

山畑倫志、2011、「韻律分析によるアパブランシャ語の格形式の確定」、平成23年3月、『印度学仏教学研究』（日本印度学仏教学会）　第59巻第2号、pp.832-839.

刑事判決書の特徴表現パターン抽出に関する複数手法の検討

Experimental Investigation of the Utility of Semi-syntactic Analysis for Phrase Pattern Extraction from Judicial Decisions

千本 達也[†1] 山本 大介[†2] 竹内 和広[†1] 三島 聡[†3]
Tatsuya Senbon Daisuke Yamamoto Kazuhiro Takeuchi Satoshi Mishima
†1 大阪電気通信大学 情報通信工学部 情報工学科, 寝屋川市初町 18-8
Osaka Electro-Communication University, 18-8 Hatsumachi, Neyagawa, Osaka
†2 大阪電気通信大学 工学研究科 情報工学専攻
Osaka Electro-Communication University, Graduate School of Engineering
†3 大阪市立大学大学院 法学研究科
Osaka City University, Graduate School of Law

あらまし:刑事判決書を収集し、その文書集合から特徴的な表現パターンを抽出することを検討する。具体的には、文の句構造を中心とした要素を解析し、文書群すべてから、その機能表現をノードとして、木構造に組織化する。その上で、木に対して複数の枝刈りといった編集をすることにより、当該パターンが特徴的か否かを検討する。また、木を表現し編集するデータ構造について、データ量、編集における計算効率性の観点から工夫を行った。

Summary: In this paper, we propose a data structure to store the various superficial syntactic sequences, which reflects the combination of content expressions and functional expressions in a certain set of specialized documents. We investigate some algorithms that extract characteristic patterns from the previous judicial decisions with the data structure. As a consequence of the experiment on the limited data, we confirm the efficiency of the extracted patterns from the viewpoint of information extraction and retrieval.

キーワード:情報検索, 表層的統語解析, 機能表現
Keywords: information retrieval, superficial syntactic analysis, functional phrases

1. はじめに

本稿では、裁判員裁判において適正・妥当な量刑審理・量刑判断を確保するための基礎資料を提供することを目的とし、既存の刑事裁判決書の蓄積から、複雑な条件下で検索をする、あるいは、情報抽出した結果を計量的に分析するための基礎技術として、自然言語で記述された刑事判決書を収集し、その文書集合から特徴的な表現パターンを抽出することを検討する。

具体的には、表層的な統語構造におけるパターンの組み合わせをできるだけ広く、現実的なデータ構造を用いて保存しておき、その保存データに基づいて特定の文書集合に特徴的な表現パターンを抽出する複数手法を提案する。本稿では、上記データ構造の実装を行い、実際の刑事裁判の判決書をデータとして、それぞれの手法により抽出された表現パターンが類似情報抽出や情報検索といった応用上、どのような利点があるかを検討したい。

2. 文構造の表層的解析

2.1 文型パターン

専門文書に対して高度な検索やテキストマイニングといった処理を行うためには、形態素解析の辞書や係り受け解析のモデルの調整といった言語処理基盤の再整備・調整が課題となってきた。

実際、日本語の統語解析の一つとしてよく用いられる係り受け解析モジュールの cabocha[1)]を法律文書に対して適用した場合、単語に関しては登録されていな

い単語が多く、また、独特の長い文を持つ文体から、係り受けの解析間違いも多い。

他方、統語的な文解析の結果表示には、特定の統語構造のパターンを類型化した文型を利用した表現方法もある。この文型と呼ばれる概念は、特に日本語学習の分野では盛んに用いられている。例えば、日本語能力検定では、学習者があやまりがちな表現の出現環境を同様のパターンによって整理した書籍が販売されている。

文型は一般的には、次のような例文における、「N は、V」、「N1 に N2 を送る」といった形の、特徴的な文構造を表示する。先のカギ括弧内に示したパターンを本稿では文型パターンと呼ぶ。

「太郎は、花子に花を送る」

一般化すると、文型パターンは、N, V といった意味的あるいは統語的性質で抽象化可能な非終端記号と文構造の特徴付け要因となる文書中に実現された実文字列の組み合わせによって記述できる。後者の実文字列は多くの場合、機能表現と呼ばれる表現である。

文型パターンは、文構造の特徴を示すため、パターン内の係り受け構造と整合性を持つ。例えば、「N は、V」、「N1 に N2 を送る」を例にすれば、係り関係はそれぞれ、{(N は、→V)}, {(N1 に→送る), (N2 を→送る)}である。この様に、文型パターンが合致する文の端的に文構造を示している。

文型パターンは、パターン中の非終端記号の抽象度が様々である。例えば、「N1 に N2 を送る」の「送る」という動詞は、「贈る」「あげる」「プレゼントする」といった動詞に代えた場合も、同様の文構造的な特徴をもちうる。このような当該部分に当てはまる複数の対象を抽象化した非終端記号を恣意的に導入するため、非終端記号の性質が統語的なもの、意味的なもの、それらが複合的にからみあったもの、さらには談話的な性質をもつもの、といったように抽象化のレベルや基準が様々である。このことは抽象化に対応する語の集合を規定するという形は可能ではあるが、抽象化の理由は自然言語による記述に頼らざる得ない部分があり、人間が文特徴を把握することを志向した文解析の表示方法といえる。

2.2 機能表現の定義

野田 [2]は、形態素解析用の電子化辞書 Unidic[3] に収録されている、接続詞・連体詞・助動詞・助詞・名詞-助動詞語幹・形状詞-助動詞語幹、及び、松吉ら [4]が編纂した日本語機能表現辞書に収録されている表現を整理した文型パターン分析用の機能表現の集合を定義した。松吉らの日本語機能表現辞書は、「からすると」や「ざるを得ない」のように、国立国語研究所の研究では複数の短単位から構成され機能的な働きをする長単位を、助詞・助動詞型に属する複合辞として扱っている。これらを複合辞の活用や音韻的変化について調査・整理し、機能語に対して認定される表現を、機能型に属する機能表現と定義している。そして、この定義に従い、『日本語表現文型:用例中心・複合辞の意味と用法』[5]、『使い方の分かる類語例解辞典 新装版』[6]に収録されている表現から、341 種の見出し語をもつ 16,771 個の出現形について整理を行い辞書化した。この野田が整理した機能表現(F)を扱う。また、名詞・動詞・形容詞などの、機能表現と対となるものを内容表現(C)とする。そして、ある文における内容表現と機能表現の並びを CF 系列とする。

2.3 CRF による機能表現系列解析

機能表現系列を解析するタスクは、一般に系列ラベリング問題と呼ばれる。ここで、いくつかの要素が連なったものを系列と呼び、系列内のそれぞれの要素にラベルを付けることを系列ラベリングと呼ぶ。系列ラベリングは、ある文では動詞として機能する単語が、直前に形容詞が来た場合には、名詞として機能する場合があるなど、ある要素のラベルが系列内の他のラベルに依存するような場合を扱う。

ここで、機能表現系列を特定するという系列ラベリング問題を処理するために、条件付き確率場(Conditional Random Fields, CRF)[7]を利用する。CRF による識別モデルは形態素解析 [8]や未知語抽出 [9]などの系列ラベリング問題において、隠れマルコフモデル(Hidden Markov Model, HMM)による生成モデルや MEMM(Maximum Entropy Markov Model)よりも高い精度で解析できることが報告されている。他にも、固有表現抽出 [10][11]などでの利用例があり、多くの言語処理タスクにおいて利用される機会が増えている。

本稿では、CRF を用いて各節中の文字系列に対して、各文字に対応するラベルをラベリングする。ラベルの種類は、C と F の 2 種類である。CRF の実装は、CRF++[8]を用いて野田が整備したツールを使用する。

機能表現		begin	を	ていた
内容表現の文字数		7	3	0
内容表現の文字種	漢字	1	1	0
	ひらがな	1	1	0
	カタカナ	0	0	0
	数字	0	0	0
	アルファベット	0	0	0

図 1：CF 系列の抽出

3. 提案手法

3.1 文の解析処理

本節では、刑事判決書の文型的特徴を持った表現の型を機械的に抽出するための前処理として、CF 系列を抽出する。ただし、CF 系列中の内容表現は、実文字列を文字するのではなく、各機能表現にセットとなる内容表現の最大文字列数や文字種などの情報を付与することで、抽象化して扱う。これは、文の持つ機能的な役割の特定と、各機能表現間に入りうる内容表現を制限するためのものである。

入力された節から、文型パターンを抽出するための前処理として、CF 系列を抽出する手続きを以下に示す。

1. 刑事判決書中には記号を伴った箇条書きや特殊な記法によって記された事件番号、裁判の日付などが存在し、これらが以降の解析時にノイズとなるため、テキストから取り除くか、代替記号に置換するクリーニングと呼ばれる作業を行う。また、これ以降の解析において、文を句読点で区切られた節として解析するために、句読点で文を区切り、それぞれ別の系列として扱う。それから、句読点を削除した。さらに、括弧等の記号で囲まれている文中の機能表現は、それらを埋め込まれている文の機能表現の系列中に含んでしまうと以降の解析時に上手く処理することができないため、正規表現を利用して抽出し、別の節として処理を行った。
2. CRF を用いて、節中の機能表現系列を解析し、節を機能表現と内容表現に分解する。
3. 機能表現系列の先頭に begin というダミーの機能表現を追加する。これは以降の解析において、機能表現の先頭情報を保持するためのものである。これにより、機能表現の系列は末尾から見て、内容表現と機能表現の組の繰り返しである CF 系列となる。各機能表現には対となる内容表現の文字数と文字種の情報が付与される。文字種は、内容表現中に特定の文字種が出現するならば 1、そうでないならば 0 というようにフラグの形で情報を保持する。

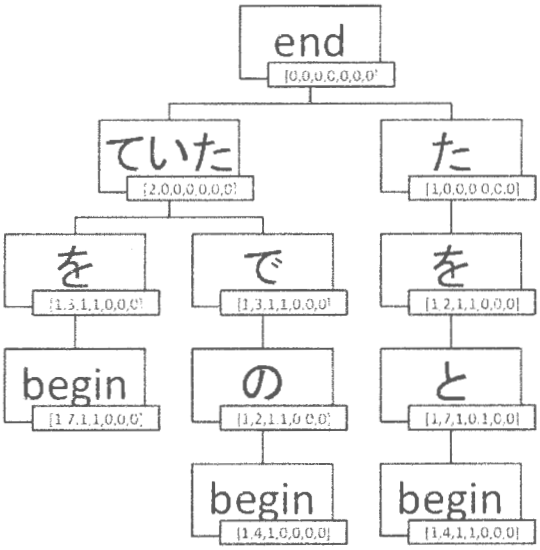
以上、一連の流れで、例として「繰り返し揉め事を起こしていた。」という入力節に対し、前処理を施した様子を図 1 に示す。処理により得られる CF 系列（内容表現に関する情報は省略）は「ていた-を-begin」となる。

3.2 文解析結果の保存法

刑事判決書の各文に対して、3.1 節で述べた前処理を行い、抽出した各 CF 系列から文型パターンを抽出するために、各 CF 系列を一つに統合した CF 系列木として構造化する。各 CF 系列を木として構造化したのはデータサイズの圧縮と、次節において、文型パターンを抽出する際の手続きのためである。木の各ノードは機能表現であり、異なる CF 系列をまとめる際、節末を含む共通の機能表現の部分系列を持つ場合は、それらを一つに統合した。

複数の CF 系列に対して、それらを一つ木構造にするための手順を以下に示す。

1. まずは機能表現の系列に着目した際に、同一の機能表現系列を持つ CF 系列を一つに統合する。そして、統合した CF 系列の数を、その CF 系列の出現数として付与する。ただし、CF 系列を統合する際、各機能表現に付与されている内容表現の情報については、文字列数は大きい方の値を保持し、各文字種については、論理和をとる。
2. 木の根として、end という根ノードを作成する。これは、各 CF 系列の末尾に相当する。
3. 各 CF 系列の末尾の機能表現から順に、end ノードに、再帰的に各機能表現をノードとして追加する。末尾から見た各 CF 系列の並び順は、木における階層と対応する。例えば、end が 0 番目の階層ならば、各 CF 系列の末尾の機能表現は、end の子として追加されるので、1 番目の階層となる。
4. 機能表現をノードとして追加する際、同時に機能表現とセットとなっている内容表現に関する情報と、各 CF 系列に 1.で付与した出現数を、そのノードの属性として持たせる。具体的には、出現数（weight）と木における階層（level）、そして、内容表現の文字数（length）、内容表現の文字種のフラグとして、漢字（kanji）・ひらがな（hira）・カタカナ



※ [weight, length, kanji, hira, kata, num, alpha]

図 2：CF 系列木の抽出

(kata)・数字(num)・アルファベット(alpha)を持たせる。また、ノードを追加する際、すでに同じ機能表現を持つ、既存ノードが存在する場合は、これらの情報に単に上書きせずに、それぞれ更新を行う。weight は追加ノードのものを既存ノードのものに加算、length は値が大きい方を保持、文字種については、それぞれ論理和をとる。

以上の一連の処理の流れで、以下の各節に対し、生成した CF 系列木を図 2 に示す。各節の機能表現は下線で表した。

(1) 繰り返し揉め事を起こしていた。
(2) 被害者方の離れで生活していた。
(3) 強い憤りと精神的ストレスを感じた。

3.3 特徴的文型パターンの抽出

CF 系列木を用いて、刑事判決書の各節に対して、文型パターンを抽出した。文型パターンは、一つの節に対して複数抽出される。

文型パターンの抽出方法について、図 3 に示す CF 系列木と入力節を用いて解説する。これは解説のために擬似的に設定したものである。丸で囲まれた数字は内容表現、アルファベットは機能表現をそれぞれ表す。図 5 のように、E ノードを根とする木を生成する。これは、文型パターンを抽出するための CF 系列木とは異なる木である。この木を文型パターン木とする。文型パターン木から、CF 系列木と整合がある当該文書群に特徴的な文型パターンを抽出する。E ノードを、図 4 に示す 3 つの木の成長規則を用いて成長させていく。ここで、S は任意時点での処理済みの部分木、X は任意の追加ノード、＊は全機能表現に対して整合を持つ追加ノードである。各成長規則を以下に示す

a) S に X を追加する。ただし、CF 系列木中の S の子に X が存在しない場合は、成長を止める。
b) S に＊を追加し、＊に X を追加する。ただし、CF 系列木中の S−＊の子に X が存在しない場合は、成長を止める。
c) S に X を追加せず、level と文中の対象文字を一つ進める。

ただし、機能表現系列の一致だけではなく、CF 系列木中の対応するノードが持つ内容表現に関する情報と整合が取れなければ成長を止めた。具体的には、文型パターン木を成長させる際、解析対象節中のある機能表現と対となる内容表現の文字数が、CF 系列木中の対応する機能表現が保持する文字数より大きいならば、成長を止めた。また、文字種についても同じく、解析対象節中のある機能表現と対となる内容表現の文字種を、CF 系列木中の対応する機能表現が保持していなければ、成長を止めた。これは、各機能表現間に入る内容表現に制限をかけるためである。

最終的に、図 5 における level:5 の B ノードのように、CF 系列木において対象の機能表現の子に begin、つまり、その機能表現の直前の内容表現から、CF 系列が始まるという情報が CF 系列木中に、存在したものを文型パターンとして抽出する。また、各ノードには、3.2 節と同じように、入力節の内容表現の情報をそれぞれ持たせた。

4. 実験

4.1 文型パターン被覆率

刑事判決書 1,171 件の約 640,000 節において、各節に対して CF 系列を抽出し、CF 系列木を生成した。CF 系列木を生成するまでの流れを図 6 に示す。そして、まとめた CF 系列木を用いて、前述の約 640,000 節から、判決書の発行年月が古いものから約 1,000 節を取り出し、各節に対して複数の文型パターンを抽出した。実験に用いた刑事判決書の情報を表 1 に示す。各節に対して抽出された文型パターンの一部を表 2 に示す。各節の下線部は機能表現を表す。そして、各節の文型パターン①は、対象節の全機能表現を並べた状態、つまり CF 系列に相当する。文型パターン②は機能表現数が①に比べて少なく、文型パターン③は、＊を含む文型パターンであり、＊は一定の CF 系列と合致す

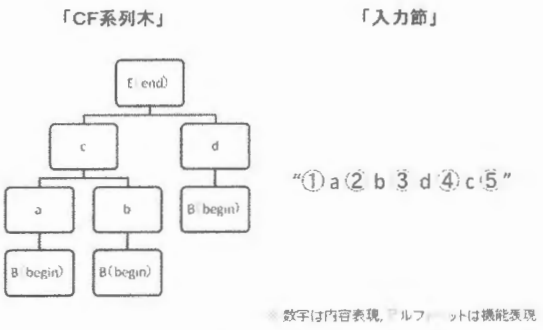


図 3：CF 系列木と入力節

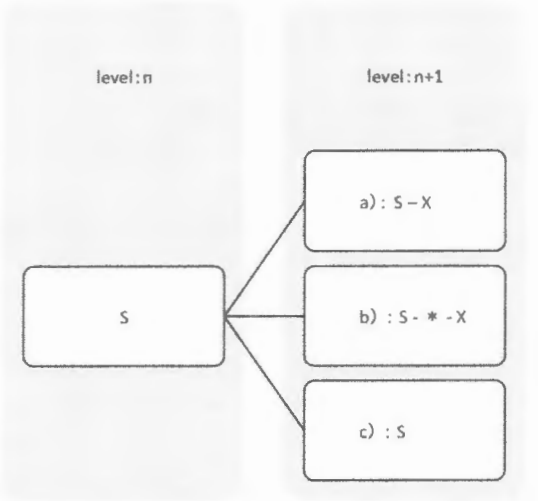


図 4：文型パターン木の成長規則

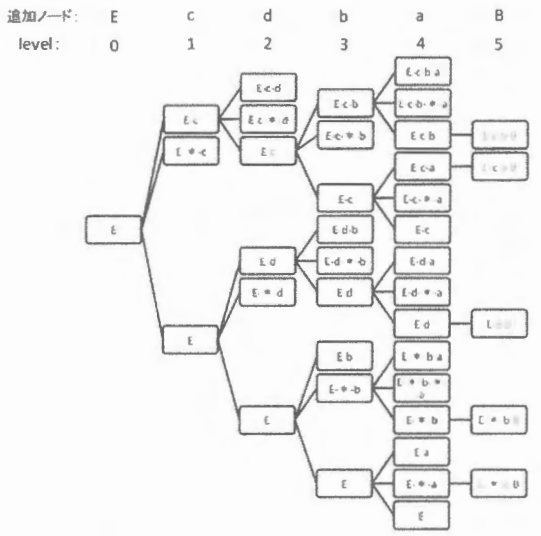


図 5：文型パターン木の成長過程

る。これは、部分木の抽象化に相当する。また、文型パターンは簡易化のために、文型パターン数と各機能表現と対となる内容表現に関する情報などを省略して、CF 系列のみを記載している。

抽出した全文型パターンにおいて、整合を持つ節が多いものについては頻出する表現であると考えられる。そこで、n 個の機能表現からなる文型パターンについて、2 つ以上の節と整合を持つ被覆率 P を算出した。P の算出方法を以下に示す。ここで、ただ 1 つの節と整合を持つ文型パターンの数を PN1、2 つ以上の節と整合を持つ文型パターンの数を PN2 とする。

$$P = \frac{P2}{P1 + P2}$$

機能表現数毎の文型パターンの被覆率を図 7 に示す。n = 1～9 までの結果を出しているが、これは n = 10 以上の文型パターンが抽出されなかったためである。結果としては、n = 3, 4 の文型パターンだけで、全体の約 55%以上を占めており、被覆率 P についても、n = 2 についで高い。よって、刑事裁判決書の特徴な文型パターンの機能表現数は 3 または 4 で調整することが有効であることが推測される。

4.2 CF 系列木を枝刈りしたときの被覆率変化

CF 系列木に対して、複数の枝刈りを行い、そこから抽出された文型パターンの被覆率 P の変化を調査した。ただし、前節の結果を踏まえ、調査を行う文型パターンは n = 3, 4 とする。

CF 系列木に対して、以下の提案手法を用いて枝刈りを行う。各提案手法を閾値 m で適用した場合の CF 系列木から、n = 3, 4 の文型パターンを生成し、m = 1～5 まで変化させたときの n = 3 と n = 4 の文型パターンの P の平均の変化を図 8 に示す。

A. あるノードに着目したときに、そのノードの各子の weight が m 以下の子をカットする。ただし、ノードが根である場合は処理を行わない。
B. あるノードに着目したときに、その親の weight を parentWeight とする。そして、各子の weight の下位 m 個の和が parentWeight よりも小さいとき、それらの子をカットする。ただしノードが根である、または子が一つしかない場合は処理を行わない。
C. あるノードに着目したときに、その親の weight を parentWeight とする。そして、各子の weight の下位 m 個の積が parentWeight よりも小さいとき、それらの子をカットする。ただしノードが根である、または子が一つしかない場合は処理を行わない。

実験の結果より、手法Cの閾値m = 3のときのPの値が他手法よりも高かった。被覆率Pは前述でも述べた通り、実験対象の全節のうち、2つ以上の節に整合する文型パターンの被覆率を表している。これは、多くの節に当てはまる頻出する文型パターン数を表しているので、手法Cは、文型パターン抽出に用いた節集合に対して、頻出する表現を多く保持することがわかった。

4.3 文型パターンによる内容表現の抽出特性

4.2節において、手法Cのm=3で枝刈りしたCF系列木から抽出した文型パターンを用いて、情報抽出と情報検索への応用を考えてみたい。情報抽出も情報検索も文型パターンによる内容表現の抽出が基本となる。

従来の内容表現抽出では、内容表現を抽出するため、ストップリストや形態素解析を用いるが、本稿の手法では、文の表層的な統語的枠組みである文型パターンと合致することにより、内容表現を抽出することができる。例として、表2の「被告人に逆らえなくなったというCの供述が不自然であるとはいえない」というテキストに、合致する文型パターン①および③を適用したときに抽出される内容表現を表3に示す。

表3が示すように、文型パターンに部分木の抽象部分を含まない時(文型パターン①との合致を利用するとき)、抽出された内容表現は、当該文の内容語となる形態素にほぼ相当する部分となり、形態素解析を使わずとも、すなわち、形態素解析辞書を使わず、機能表現を中心とする辞書のみを使って内容語抽出ができることを示している。

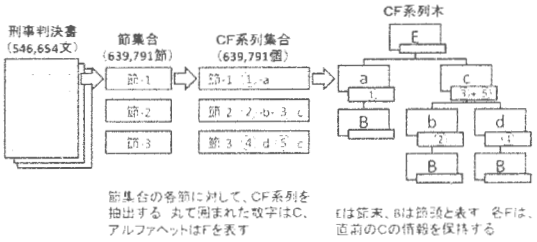


図6：CF系列木までの処理の流れ

表1:刑事判決書の情報

	文数	節数(CF系列数)
全判決書	546,654	639,791
CF系列木生成に用いた判決書	546,654	639,791
文型パターン抽出に用いた判決書	874	1,030

表2:抽出された文型パターンの例

節	文型パターン
被害者の告訴能力の有無の判断に直接影響するものではない。	①end-ではない-に-の-の-の-begin ②end-ではない-の-の-begin ③end-＊-の-の-の-begin
これに基づいて被告人の処罰を求めていると認められる。	①end-られる-ている-と-を-の-て-に-begin ②end-ている-と-を-の-て-に-begin ③end-ている-と-＊-に-begin
被告人に逆らえなくなったというCの供述が不自然であるとはいえない。	①end-ない-とは-が-の-た-という-なく-に-begin ②end-ない-が-の-た-という-なく-begin ③end-ない-とは-が-＊-に-begin

さらに、本手法では、上記の内容語抽出が、文型パターン①との連言条件で成立することを保証するように、内容語抽出の抽出条件を文型パターンの選択により制御することが可能である。表3の文型パターンに部分木の抽象部分を含む場合では、抽出できる内容表現相当部分を形態素に相当する単位だけではなく、波括弧内のような句レベルの文字列を内容表現として抽出可能であると同時に、このような多様な表現を抽象化して検索することが可能となる。

このような文型パターンの抽象化は、制約条件を無制限に緩和しすぎてしまうと内容表現を限定することが困難であるが、本稿の提案手法では、あらかじめ、当該文章集合の可能性あるCF系列の大部分はCF系列木により保存しており、内容表現の抽出制約を当該文書の特質に即して制御可能である。

また、同一文に複数の文型パターンを対応付けることが容易に可能である点を利用すれば、文を限定すれば、当該の文に合致する文型パターンを利用して「同じような書かれ方をする文」を検索することが可能であり、文型パターンと内容表現を複合的に用いた類似検索が検討可能と考えている。

4.4 今後の課題

本稿の提案の主眼は、自然言語文章をデータベース化する際のデータ構造に文型パターンを意識したCF系列によるデータ保存を行う点にある。現在の実装では、このCF系列のCRFに基づく解析とCF系列木の構築に関しては一定の効率的な実装を行えている

が、保存したデータの参照効率には課題が残る。特に、CF 系列からの文型パターンの抽出に関して非常に時間が掛かっており、多くの文章に一般的に出現する CF 系列を符号化しておき、今回の判決書などの専門性の高い、特定分野の文章集合に特徴的な CF 系列をあらかじめ区別して処理する工夫が必要である

具体的には、4.1節と4.2節の実験的検討により、少なくとも刑事裁判判決書の特徴な文型パターンの機能表現数は 3 または 4 で調整することが有効であることが予想され、この知見を生かして、CF 系列木を効率的に保存・参照する工夫を行っていきたい。

今回の検討では、文型パターンの抽象化に関して統計的な検討を行わなかったが、これは、上記の CF 系列の参照高速化の課題に依存性をもつ。また、内容表現を抽象化するための情報も、最大文字数、文字種といった限定的な情報を保持するに留まっており、文型パターンの抽象化について限定的なアルゴリズムを提案している点に課題が残る。

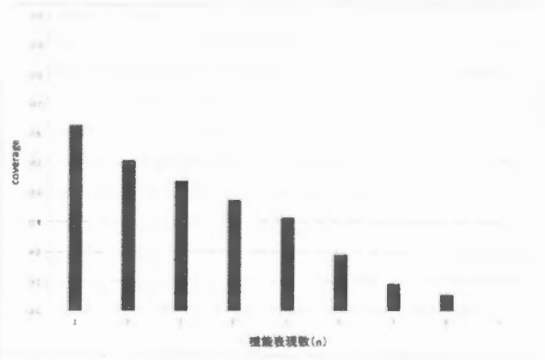


図 7：機能表現数毎の文型パターンの被覆率

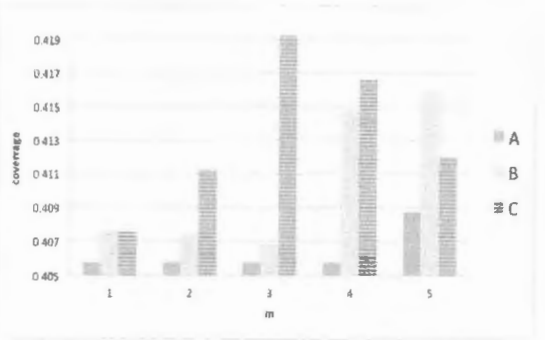


図 8：提案手法毎の被覆率の変化

表 3：文型パターンを用いて抽出した内容表現

文型パターン	内容表現
部分木抽象化なし	• 被告人 • 逆らえ • なっ • C • 供述 • 不自然である
部分木抽象化あり	• 被告人 **{逆らえなくなったというC}** • 供述 • 不自然である

5. まとめ

本稿では、刑事裁判例を計量的に分析するための基礎技術として、自然言語で記述された刑事判決書を収集し、その文書集合から、機能表現を中心とした辞書と CRF を用いることにより、機能表現と内容表現情報の組の繰り返しからなるCF系列を抽出しCF系列木として構造的にデータ保存をする方法を提案した。そして、CF 系列木に基づいて文型パターンを抽出するアルゴリズムを複数検討し、文型パターンといった表層的な統語構造を基準に、当該文章の内容表現の抽出を制御可能なことを実験的に確認した。

現状においては、比較的軽量な処理を実現できた自然言語文章の浅い統語解析に基づいて、構造的な情報を全文書に関して保存しておくことは可能であるが、その利用アルゴリズムを多様化する上での高速な参照を実現する実装が課題である。今後は、より高速な CF 系列木の実装に基づいて、複雑な類似判例の検索・クラスタリングといった応用処理を検討していきたい。

謝辞

本研究は、財団法人日本証券奨学財団の研究調査助成金「裁判員裁判において当事者の主張提示が量刑および判決に与える影響の検討」(研究代表・三島聡)による研究成果の一部である。同財団には謝意を表したい。

参考文献

1) 工藤拓, 松本裕治. Cabocha - Yet Another Japanese Dependency Structure Analyzer - Google Project Hosting (オンライン). 入手先 (http://code.google.com/p/cabocha/).

2) 野田奏. 分野非依存辞書に基づく多様な文章に対する文節分析とその応用. 大阪電気通信大学工学研究科情報工学専攻, 修士論文. 2012.

3) 国立国語研究所:特定領域研究 日本語コーパス 形態素解析辞書 UniDic(オンライン). 入手先(http://www.tokuteicorpus.jp/dist/).
4) 松吉俊, 佐藤理史, 宇津呂武仁. 日本語機能表現辞書の編纂, 自然言語処理, Vol. 14, No. 5, pp. 123-146. 2007.
5) 森山良行, 松木正恵. 日本語表現文型:用例中心・複合辞の意味と用法. 株式会社アルク. 1989.
6) 遠藤織枝, 小林賢次, 三井昭子, 村木新次郎, 吉沢靖. 使い方の分かる類語例解辞典 新装版. 株式会社小学館. 2003.
7) J.Lafferty, A.McCallum, and F.Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the 18th International Conference on Machine Learning. 2001.
8) 工藤拓, 山本薫, 松本裕治. Conditional random fields を用いた日本語形態素解析. 情報処理学 会研究報告. 自然言語処理研究会報告, Vol. 2004, No. 47, pp. 89-96. 2004.
9) 東藍, 浅原正幸, 松本裕治. 条件付確率場による日本語未知語処理, 情報処理学会研究報告, 自然言語処理研究会報告, Vol. 2006, No. 53, pp. 67-74. 2006.
10) 福島健一, 鍛治伸裕, 喜連川優. コーパスからの固有表現辞書の自動構築, 知識ベースシステム研究会, Vol. 79, pp. 19-24. 2007.
11) 齋藤邦子, 今村賢治. タグ信頼度に基づく半自動自己更新型固有表現抽出, 自然言語処理, Vol. 17, No. 4, pp. 3-21. 2010.

文章の特徴量を用いた質問回答文の印象の因子得点の推定精度の向上

Improving Estimation Accuracy of Factor Scores of Impression of Question and Answer Statements by Using Feature Values of Statements

横山 友也*, 宝珍 輝尚*, 野宮 浩揮*, 佐藤 哲司**

Yuya Yokoyama, Teruhisa Hochin, Hiroki Nomiya, Tetsuji Satoh

*京都工芸繊維大学 情報工学専攻, 京都市左京区松ヶ崎御所海道町

Kyoto Institute of Technology, Graduate School of Information Science

Goshokaidocho, Matsugasaki, Sakyo-ku, Kyoto

**筑波大学大学院 図書館情報メディア研究科, 茨城県つくば市春日 1-2

Graduate School of Library, Information and Media Studies, University of Tsukuba

1-2 Kasuga, Tsukuba-shi, Ibaraki

あらまし:質問回答サイトにおける質問者と回答者のミスマッチングの問題を解消するために, Yahoo!知恵袋に投稿された 60 個の質問回答文に対し, 印象評価実験を行ってきた. 実験結果に対して因子分析を施したところ, 文章に関する 9 個の因子が得られた. 構文情報, 単語心像性, 文末表現といった特徴量を採用することで, 全ての質問回答文の因子得点の推定を試みてきている. 本論文では, 単語親密度や表記妥当性を特徴量として採用することにより, 推定精度の向上を図る. これらの特徴量を追加することで, 推定精度が向上できることを示す.

Summary: In order to avoid the problem of mismatch between the questioner and the respondent at Q & A sites, we have experimentally evaluated the impression of 60 question and answer statements posted at Yahoo! Chiebukuro. Nine factors as to statements have been obtained by applying the factor analysis to the scores obtained through the experiment. Factor scores of any other statements have been tried to be estimated by adopting such feature values as syntactic information, word imageability and closing sentence expressions. This paper tries to improve estimation accuracy by adopting word familiarity and notation validity as feature values of statements. It is shown that estimation accuracy can be improved by adding these feature values to the conventional ones.

キーワード:Q&A サイト, 重回帰分析, 因子得点, 単語親密度, 表記妥当性

Keywords: Question & Answer site, multiple regression analysis, factor score, word familiarity, notation validity

1. はじめに

インターネット上における質問回答サイトの利用者が近年急増している. 質問回答サイトとは, インターネット上でユーザ同士が互いに質問と回答を投稿しあうコミュニティの一形態であり, 様々な悩み事・相談事を解決する場であると同時に, 膨大な知識が蓄積されたデータベースとして活用されるようになってきている[1]. あるユーザが質問を投稿すると, 他のユーザがその質問に対して回答を投稿する. 質問者は, 質問文に対して最も適切と判断した回答文を「ベストアンサー」に選定し, その回答を行った回答者に謝礼として手持ちのポイントを贈与する. ここで, 「ベストアンサー」とは, 質問文に対する満足度が最も高いと質問者が主観的に判断した回答文である.

質問回答サイトの参加者が増え, また, 投稿される質問数が膨大になると, 回答者が自身の専門性や興味に合った適切な質問文を探し出すことが困難になるという問題が顕在化してくる. あるユーザが質問文を投稿しても, その質問文が必ずしも適切な回答者の目に留まり, 回答を得られるわけではないという問題である. また, 適切な回答者に巡り会えないミスマッチから, 質問者にも不利益も生じる. つまり, 質問回答サイトの課題は, 日々投稿され続けている幾多の質問と, 様々な興味・関心や専門性を有する回答者とを適切にマッチ

ングすることであるが，質問者や回答者の努力に任せているのが現状である．

これまでの研究において，筆者らは，質問者に適切な回答者を引き合わせるために，質問者と回答者の相性を判断する手段として質問者と回答者の文章の印象評価を行ってきた[2]．Yahoo!知恵袋[1]に投稿された質問文と回答文の計60個の文章に対して，50個の印象語を使用して，印象評価を行った．得られた評価値に対して因子分析を行った結果，文章内容に関する因子が9個得られた．また，得られた因子の因子得点を適宜利用することで，「ベストアンサー」を特定できる可能性を示した[2]．しかし，ここで得られた因子得点は，評価実験を行った質問文と回答文の文章60個に対するもののみであって，他の多数の質問文と回答文に対する因子得点は得られていない．

そこで，どのような質問文と回答文に対しても因子得点の推定を可能にすることを目的として，文章の特徴量からの文章の因子得点の推定を重回帰分析により試みてきた[3]．ここで，(ⅰ)形態素解析を通して得られた「構文情報」，(ⅱ)「単語心像性」，(ⅲ)文末表現という3種の特徴量を用いて因子得点を推定すると，7因子は良好に，2因子はやや良好に因子得点が推定できることを示してきた[3]．しかし，やや良好な推定精度が得られた2因子に関しては，更に推定精度を良くする余地があり，文章の特徴量を更に追加することで，因子得点の推定精度をより向上させる可能性がある．

そこで，本論文では，文章の特徴量として，単語親密度と表記妥当性を追加することにより因子得点の推定精度の向上を試みる．これらの特徴量を用いて推定を行い，推定精度が向上することを示す．

以降，2.では関連研究について述べ，3.ではこれまでの研究について述べる．4.では，新たに追加した特徴量について述べ，5.では因子得点の推定結果について述べる．そして，6.で推定結果に対する考察を示す．最後に7.でまとめる．

2. 関連研究

これまで，「ベストアンサー」を推定する研究が行われてきている[4–11]．Bloomaらは，非テキスト特徴量とテキスト特徴量を用いて，「ベストアンサー」の推定を試みている[4]．Agichteinらは，内容や語法の特徴量を使用することによって，質問文と回答文の質の評価を試みている[5]．また，類推による手法も提案されている[6]．この手法では，過去の知見における質問文と回答文のリンクを使用することによって，「ベストアンサー」を探索する．Kimらは，「ベストアンサー」の選択基準を提案している[7]．情報型の質問は，文章内容の特徴量が重要であり，提案型の質問には有用性が重要であり，選択型の質問には社会的な感情が重要であるとしている．

西原らは，ある1つの質問文に対する回答文群より，「ベストアンサー」になりやすいものを検出する手法を提案している[11]．質問者と回答者の文末表現の相性に着目し，質問文と「ベストアンサー」の組み合わせをクラスタリングすることで，一定の成果をあげている．しかし，この研究では，質問文ならびに回答文の文末表現に着目した手法をとっており，文章内容に着目した手法をとっていない．そこで，本研究では，文末表現も，特徴量として考慮した上で，文章全体の文体や内容から受ける印象評価に着目して検討を行っている．

また，これらの研究は「ベストアンサー」を推定することに重点が置かれた研究である．この点に対して，本研究では，質問文に適切な回答を施すことができると考えられる回答者の選出を図っている点が，大きな特色となっている．更に，既存の研究では，質問回答文の構文情報を直接扱うことにより，「ベストアンサー」の推定を図っている．この点に対して，本研究では，同じ内容の文章でも，表現によって受ける印象が大きく異なることを考慮しているため，質問回答文の印象の度合いにも着目している点にも，大きな特色があるといえる．

一方，熊本は新聞記事を対象として印象評価を行っている[17]．被験者100人が新聞記事10記事を読んで，印象語42語のそれぞれについて5段階(強い—わりと強い—わりと弱い—弱い—なし)で評価するという印象評価実験(アンケート調査)を9度実施して，印象評価データ(印象語42語×9000件)を収集・分析することによって，新聞記事の印象を表現するのに適した印象軸を提案している[17]．本研究では，質問回答文を対象として，印象語を用いた印象評価実験を行う．

3. これまでの研究

3.1. 印象語

Yahoo!知恵袋[1]に実際に投稿された12組60個の4大ジャンル(Yahoo!オークション，パソコン・周辺機器，恋愛相談・人間関係，政治・社会問題)の質問回答文に対して，印象評価実験を行い，実験結果に対して因子分析を施したところ，文章に関する因子が9個得られた[2]．因子，ならびに，因子に対応する印象語を表1に示す．なお，これらの因子の因子得点は，実験で使用した60個の質問回答文から得ており，実験で使用していない質問回答文の因子得点はまだ求まっていない．そこで，文章の特徴量から因子得点を重回帰分析による推定を試みた．

表 1　9 因子と対応する印象語

因子	印象語		
第1因子(的確性)	説得力がある	流暢な	重要な
	素晴らしい	好ましい	巧みな
	真実味がある	清々しい	妥当な
	充実した	美しい	的確な
	丁寧な		
第2因子(不快性)	不快な	憤慨した	非常識な
	残念な	不当な	呆れる
	幻滅した	怖い	
第3因子(独創性)	独創的な	予想外な	特殊な
	斬新な	不思議な	
第4因子(容易性)	易しい	明瞭な	難しい
第5因子(執拗性)	細かい	しつこい	長い
第6因子(曖昧性)	曖昧な	不充分な	
第7因子(感動性)	心温まる	感動的な	
第8因子(努力性)	涙ぐましい		
第9因子(熱烈性)	熱い	力強い	

3.2. 文章の特徴量

文章の特徴量として，形態素解析を通して得られた構文情報，単語心像性，文末表現の 3 組を採用してきた[3].構文情報とは，文の数や文の長さ，名詞や動詞といった品詞の数や割合を，形態素解析により抽出した．また，感嘆符や疑問符の数といった具体的な記号も，文章の特徴量に採用している．単語心像性とは，単語から喚起される様々なイメージが，どの程度思い浮かべやすいかを示す主観的特性を意味している.また，文末表現とは，「全ての質問文と回答文に含まれる文末表現」を表している.ここでは，「ぞ」「だ」「よ」「ね」等の助詞の語数・割合や，文末にくる語数・割合を採用している．

ところで，重回帰分析を実施する際，複数の説明変数同士は無相関であるという前提が必要であり，説明変数は以下の条件を考慮して選択しなければならない[12].

a)目的変数との相関係数が高い説明変数の選択

b)高い相関を示す説明変数の組の一方を説明変数から除外

ここで，b)の事項に反した場合，偏回帰係数が正しく求まらないことがあり，この状態を多重共線性という[12]．多重共線性を回避するために，説明変数同士の相関係数の値を調べ，0.7 以上である組に関しては，一方のみを説明変数として使用し，他方を除外する必要がある．

多重共線性を考慮した結果，説明変数として使用することとした特徴量を表 2 に示す．構文情報は g1-g36，単語心像性は g37-g38，文末表現は g39-g64，にそれぞれ対応している．

表 2　既に採用している特徴量

g	特徴量
g1	助動詞(語彙数)
g2	接頭詞
g3	記号(語彙数)
g4	文数
g5	文の長さ平均(字数)
g6	カタカナ(語数)
g7	全角記号(語数)
g8	全角英数字(語数)
g9	形容詞(語数)
g10	副詞(語数)
g11	連体詞(語数)
g12	接続詞(語数)
g13	感動詞(語数)
g14	ひらがな(%)
g15	漢字(%)
g16	カタカナ(%)
g17	記号(%)
g18	TTR
g19	全角記号(%)
g20	英数字(%)
g21	全角英数字(%)
g22	名詞(%)
g23	形容詞(%)
g24	副詞(%)
g25	連体詞(%)
g26	接続詞(%)
g27	感動詞(%)
g28	「!」の数
g29	「?」の数
g30	句点の数
g31	読点の数
g32	中点の数
g33	3点リーダの数
g34	鍵括弧の数
g35	括弧の数
g36	「/」の数
g37	単語心像性4点台(語数)
g38	単語心像性6.5以上7.0未満(語数)
g39	か(語数)
g40	な(語数)
g41	し(語数)
g42	たい(語数)
g43	ない(語数)
g44	だ(文末語数)
g45	か(文末語数)
g46	な(文末語数)
g47	し(文末語数)
g48	です(文末語数)
g49	ます(文末語数)
g50	たい(文末語数)
g51	ない(文末語数)
g52	ぞ(%)
g53	だ(%)
g54	よ(%)
g55	ね(%)
g56	か(%)
g57	です(%)
g58	ます(%)
g59	ない(%)
g60	か(文末%)
g61	ですか(語数)
g62	ないです(語数)
g63	ますか(語数)
g64	ました(語数)

3.3. 推定結果

9 因子の因子得点を目的変数とし，64 個の文章の特徴量を説明変数として，ステップワイズ選択法による重回帰分析[13]を行った．この時の重相関係数の値を表 3 の「単項」の列に示す．重相関係数は，値が 0.9 以上ならば良好で，0.7 以上ならばやや良好，0.7 未満ならば不良である，という推定精度を表す[14]．第 1 因子から第 6 因子までの 6 因子は推定精度がやや良好であり，残りの第 7 因子から第 9 因子までの 3 因子は推定精度が不良であるという結果が得られた．

また，説明変数の積である二次の項も考慮した上で重回帰分析を行った．二次の項同士の多重共線性を考慮した結果，説明変数の数は 218 個となった．単項の場合と同様に，9 因子の因子得点を目的変数，218 個の特徴量を説明変数として，重回帰分析を行った．この時の重相関係数の値を表 3 の「二次項」の列に示す．第 3 因子と第 6 因子以外の 7 因子に関しては，値が 0.9 以上なので，推定精度が良好であるといえる．第 3 因子と第 6 因子は値が 0.9 に及ばなかったが，それでも推定精度はやや良好といえる．

さらに，二次の項に関しては，推定の良好性を推定誤差により評価する．実験の因子得点とその推定値の平均誤差の絶対値を求め，表 4 に示す．全体の誤差平均は非常に小さく，どの因子も因子得点に近い推定値が得られているといえる．

表 3 重相関係数(特徴量 64 個)

因子	重相関係数	
	単項	二次項
第1因子(的確性)	0.832	1.000
第2因子(不快性)	0.774	0.947
第3因子(独創性)	0.744	0.877
第4因子(容易性)	0.728	0.908
第5因子(執拗性)	0.893	0.966
第6因子(曖昧性)	0.872	0.899
第7因子(感動性)	0.581	0.997
第8因子(努力性)	0.650	0.904
第9因子(感動性)	0.683	0.954

表 4 各因子の残差の絶対値

因子	残差の絶対値
第1因子(的確性)	0.0000420
第2因子(不快性)	0.114
第3因子(独創性)	0.164
第4因子(容易性)	0.124
第5因子(執拗性)	0.0874
第6因子(曖昧性)	0.147
第7因子(感動性)	0.0202
第8因子(努力性)	0.112
第9因子(熱烈性)	0.0858
残差平均	0.0949

4. 新特徴量 -単語親密度と表記妥当性-

NTT データベースシリーズ[15]には，人が主観的に評定を行ったデータと，14 年間にわたる新聞に単語や文字が出現した回数を数えた客観的データが収録されている．これらのデータは，人間の言語処理過程に大きな影響を及ぼすものとして広く知られており，収録されている各特性値や特性値間の関係は，日本語自体の特性を示しているといえる[16]．これらのデータも，文章の特徴量として有用であると考えられる．既に特徴量として採用した単語心像性が，このようなデータに該当する．

ここでは，単語親密度と表記妥当性を新たに追加する．単語親密度とは，ある単語がどの程度なじみがあると感じられるかを 7 段階で表した指標である[17]．また，表記妥当性とは，ある単語の表記のもっともらしさを 5 段階で示した指標である[17]．例えば，乾電池の「たんに」という言葉例とすると，「単に」を意味する場合は単語親密度の値が 5.312，「乾電池」を意味する場合は値が 3.594 となる．これらの値は，被験者の得点を平均して求められている．

また，同じ単語の表記でも，意味または読みが異なる場合がある．例えば，意味が異なる例としては，「アース」という単語は，「電気を逃がすために接地すること」，「地球」，「殺虫剤(メーカー)」の意味がある．読みが異なる例としては，「間」という言葉は，「あいだ」，「ま」の読みがある．このような単語が形態素解析したデータに存在する場合は，文脈から判断しながら手動で意味または読みを決定している．例えば，「娯楽」という単語を例とすると，ひらがな表記の場合は表記妥当性の値が 2.95，カタカナ表記の場合は 1.95，漢字表記の場合は 4.90，である．

このようにして，単語親密度，ならびに，表記妥当性の特徴量を抽出した．これらを表 5 に示す．特徴量としては，単語心像性の特徴量[3]と同様に，単語親密度，または，表記妥当性に該当した単語の数や該当した単語の割合，単語心像性の値が 1 点台(1.0～2.0 未満)，2 点台(2.0～3.0 未満)…のように，1 点間隔で特徴量をとったもの，1.0 以上 1.5 未満，1.5 以上 2.0 未満，…のように，0.5 点間隔で特徴量をとったものを抽出した．表 5 に示した特徴量に対しても，多重共線性を考慮して説明変数を選出した．その結果，表 5 に網掛けを施した 13 個の特徴量を使用することとした．これらの特徴量を g65-g77 と称した上で，表 6 にまとめて示す．

表 5 単語親密度・表記妥当性の特徴量

単語親密度の該当単語(語彙数)
単語親密度の該当単語(語数)
単語親密度の該当単語率(語数)
単語親密度4点台(語彙数)
単語親密度4.5～5.0未満(語彙数)
単語親密度5点台(語彙数)
単語親密度5.0～5.5未満(語彙数)
単語親密度5.5～6.0未満(語彙数)
単語親密度6点台(語彙数)
単語親密度6.0～6.5未満(語彙数)
単語親密度6.5～7.0未満(語彙数)
単語親密度4点台(語数)
単語親密度4.5～5.0未満(語数)
単語親密度5点台(語数)
単語親密度5.0～5.5未満(語数)
単語親密度5.5～6.0未満(語数)
単語親密度6点台(語数)
単語親密度6.0～6.5未満(語数)
単語親密度6.5～7.0未満(語数)
表記妥当性の該当単語(語彙数)
表記妥当性の該当単語(語数)
表記妥当性の該当単語率(語数)
表記妥当性2点台(語彙数)
表記妥当性2.5～3.0未満(語彙数)
表記妥当性3点台(語彙数)
表記妥当性3.0～3.5未満(語彙数)
表記妥当性3.5～4.0未満(語彙数)
表記妥当性4点台(語彙数)
表記妥当性4.0～4.5未満(語彙数)
表記妥当性4.5～5.0未満(語彙数)
表記妥当性5点台(語彙数)
表記妥当性2点台(語数)
表記妥当性2.5～3.0未満(語数)
表記妥当性3点台(語数)
表記妥当性3.0～3.5未満(語数)
表記妥当性3.5～4.0未満(語数)
表記妥当性4点台(語数)
表記妥当性4.0～4.5未満(語数)
表記妥当性4.5～5.0未満(語数)
表記妥当性5点台(語数)

表 6 新たに使用する特徴量

g65	単語親密度の該当単語率(語数)
g66	単語親密度6.5～7.0未満(語彙数)
g67	単語親密度4点台(語数)
g68	単語親密度5点台(語数)
g69	単語親密度5.5～6.0未満(語数)
g70	単語親密度6点台(語数)
g71	単語親密度6.0～6.5未満(語数)
g72	表記妥当性の該当単語率(語数)
g73	表記妥当性3点台(語数)
g74	表記妥当性3.5～4.0未満(語数)
g75	表記妥当性4点台(語数)
g76	表記妥当性4.0～4.5未満(語数)
g77	表記妥当性5点台(語数)

5. 因子得点の推定結果

5.1. 単項のみを考慮した重回帰分析

印象評価実験で使用した 60 個の質問回答文に対し，表 2 と表 6 に示す 77 個の特徴量を説明変数とし，表 1 に示す 9 因子の因子得点を目的変数として，ステップワイズ選択法による重回帰分析[13]を行った．分析の結果，重回帰式(1)が得られた．

$$
\left.
\begin{aligned}
y_1 &= 0.00295g_{37} + 0.150g_{12} + 0.0609g_{59} + 0.0533g_{21} \\
&\quad + 0.0105g_7 - 0.113g_{29} + 0.0896g_9 + 0.0151g_{19} \\
&\quad - 0.750 \\
y_2 &= 1.73g_{44} + 1.79g_{50} - 0.115g_1 + 0.110g_{43} + 0.0341g_8 \\
&\quad - 0.172g_{12} + 0.261 \\
y_3 &= 0.161g_{52} + 0.0682g_8 - 0.0670g_{43} + 0.888g_{50} \\
&\quad + 0.0816g_{60} - 0.0852g_{74} + 0.0197g_{65} - 0.175 \\
y_4 &= -0.0468g_{37} - 0.0843g_{52} + 0.0916g_9 - 0.178g_{45} \\
&\quad + 0.305 \\
y_5 &= 0.0807g_1 + 0.0168g_6 + 0.126g_{29} + 0.0155g_{15} \\
&\quad + 0.00308g_5 + 0.0275g_4 + 0.502g_{49} + 0.731g_{50} \\
&\quad - 0.0792g_{70} - 1.14 \\
y_6 &= -0.0759g_{31} - 0.11163g_{43} - 0.0618g_{56} - 0.0342g_8 \\
&\quad - 0.126g_{55} - 0.772g_{44} - 0.691g_{50} - 0.125g_{25} + 0.698 \\
y_7 &= 0.557g_{13} + 0.0297g_{66} + 0.105g_{23} - 0.243 \\
y_8 &= 0.0871g_1 + 0.0788g_{60} + 0.104g_{23} - 0.00560g_{20} \\
&\quad - 0.00905g_8 - 0.388 \\
y_9 &= 0.193g_{52} - 0.0435g_{75} - 0.0148g_{16} + 0.321g_{50} \\
&\quad - 0.109g_{51} - 0.0260
\end{aligned}
\right\} \quad (1)
$$

この時の重相関係数を表 7 の「単項」の列に示す．第 1 因子から第 6 因子までの 6 因子中，第 5 因子のみは推定精度が良好で，残りの 5 因子は推定精度がやや良好であるといえる結果になった．また，第 7 因子から第 9 因子までの 3 因子に関しては，推定精度が不良であるという結果が得られた．

5.2. 二次の項を考慮した重回帰分析

次に，単項のみの分析で使用した 77 個の説明変数に関して，二次の項を考慮する．二次の項同士の多重共線性を考慮した結果，説明変数の数は 281 個となった．単項の場合と同様に，印象評価実験で使用した

$$
\begin{aligned}
y_1 &= 0.210g_{12}g_{58} + 0.0369g_{50}g_{60} + 0.174g_{30}g_{66} - 0.0909g_{12}g_{6-} - 0.413g_{54}g_{62} - 0.135g_{40}g_{66} + 0.102g_{13}g_{15} + 0.166g_{24}g_{61} - 0.180g_{2<}g_{-1} \\
&+ 0.704g_{51}g_{60} - 0.0893g_{13}g_{<-} - 0.0535g_{52}g_{-3} + 0.149g_{42}g_{<<} + 0.000737g_{5}g_{<-} - 0.0904g_{26}g_{55} + 0.0527g_{32}g_{42} - 0.0504g_{35}g_{18} \\
&- 0.0781g_{21}g_{60} + 0.0142g_{1<}g_{<1} + 0.0174g_{21}g_{42} - 0.00404g_{14}g_{18} - 0.338
\end{aligned}
$$

$$
\begin{aligned}
y_2 &= -0.0486g_{40}g_{5-} + 0.00356g_{-2}g_{-<} - 0.0447g_{32}g_{42} - 0.0111g_{2<}g_{-2} - 0.0141g_{60}g_{-2} - 0.176g_{13}g_{--} - 0.0287g_{12}g_{16} + 1.01g_{28}g_{<3} \\
&+ 0.151g_{20}g_{<<} + 0.165g_{60}g_{-6} - 0.188g_{21}g_{60} - 0.0509g_{49}g_{54} + 0.0415g_{1}g_{18} - 0.325g_{46}g_{69} + 0.0479g_{8}g_{<0} + 0.0345g_{19}g_{51} \\
&+ 0.0288g_{-1}g_{-3} - 0.0568g_{51}g_{60} + 0.00155g_{21}g_{22} - 0.0512g_{43}g_{54} + 0.140g_{32}g_{61} - 0.0196g_{30}g_{66} - 0.00425g_{16}g_{5-} + 0.00648g_{9}g_{16} \\
&- 0.103g_{13}g_{-3} - 0.0921g_{38}g_{18} - 0.504g_{60}g_{64} + 0.0120g_{30}g_{56} + 0.000152g_{<5}g_{16} - 0.00898g_{8}g_{39} - 0.0185g_{21}g_{60} + 0.0412g_{60}g_{-1} \\
&+ 0.00388g_{6}g_{5-} + 0.0204g_{-6}g_{--} - 0.0562g_{55}g_{-6} + 0.00349g_{43}g_{49} + 0.00213g_{4}g_{58} + 0.000243g_{<}g_{24} + 0.00106g_{9}g_{-2} - 0.00116g_{-4}g_{-<} \\
&+ 0.00458g_{6}g_{54} - 0.0000604g_{22}g_{-2} - 0.000833g_{22}g_{-2} - 0.000833g_{-}g_{-5} - 0.00486g_{49}g_{55} + 0.00583g_{25}g_{32} - 0.00219g_{32}g_{54} \\
&+ 0.000484g_{54}g_{5-} - 0.000809g_{41}g_{-6} + 0.00319g_{64}g_{-4} + 0.00120g_{-1}g_{<4} - 0.000160g_{51}g_{-2} + 0.0000497g_{16}g_{48} - 0.00000922g_{3}g_{22} \\
&- 0.0000527g_{25}g_{-5} + 0.00124g_{2-}g_{-<} + 0.00000238g_{29}g_{-3} - 0.00000274g_{<5}g_{-3} + 0.000000177g_{10}g_{<0} - 0.0591
\end{aligned}
$$

$$
\begin{aligned}
y_3 &= 0.0179g_{52}g_{-3} + 0.979g_{22}g_{<3} + 0.113g_{60}g_{-1} - 0.00832g_{33}g_{59} + 0.298g_{2<}g_{60} + 0.0136g_{5<}g_{-2} + 0.0300g_{6<}g_{77} - 0.0212g_{5}g_{5-} \\
&- 0.425g_{<4}g_{63} + 0.0192g_{8}g_{39} - 0.0100g_{16}g_{60} - 0.106g_{<<}g_{-3} - 0.0153g_{19}g_{-4} + 0.176g_{40}g_{<4} + 0.0606g_{24}g_{40} - 0.00110g_{20}g_{58} \\
&- 0.0150g_{49}g_{5-} + 0.450g_{51}g_{60} - 0.0310g_{32}g_{42} + 0.214g_{46}g_{69} - 0.824g_{-}g_{-6} + 0.404g_{35}g_{45} - 0.0331g_{11}g_{50} - 0.128g_{40}g_{<<} \\
&+ 0.0387g_{1}g_{18} + 0.0514g_{2-}g_{36} - 0.111g_{4<}g_{-4} + 0.0386g_{60}g_{--} - 0.227g_{4<}g_{53} + 0.0120g_{0}g_{<-} - 0.101g_{63}g_{-1} - 0.0205g_{24}g_{54} \\
&- 0.00159g_{16}g_{5-} + 0.00128g_{16}g_{-5} - 0.00672g_{12}g_{58} - 0.00299g_{66}g_{-2} - 0.00302g_{56}g_{<0} + 0.0000512g_{5}g_{16} - 0.00122g_{30}g_{65} - 0.0856
\end{aligned}
$$

$$
\begin{aligned}
y_4 &= -0.178g_{48}g_{--} + 0.00717g_{1<}g_{<-} - 0.000238g_{55}g_{6} - 0.486g_{33}g_{4<} - 0.0474g_{46}g_{5-} + 0.961g_{33}g_{-6} - 0.0612g_{<2}g_{-1} - 0.0119g_{-9}g_{-2} \\
&- 0.0734g_{41}g_{<0} + 0.00156g_{3}g_{22} - 0.00339g_{16}g_{3-} + 0.553g_{<5}g_{-1} + 0.00485g_{31}g_{54} + 0.0284g_{10}g_{65} + 0.0116g_{-2}g_{6} + 0.466g_{51}g_{60} \\
&+ 0.0855g_{2<}g_{<0} - 0.106g_{54}g_{64} - 0.0141g_{<6}g_{65} - 0.0232g_{10}g_{2<} - 0.0173g_{51}g_{-2} + 0.00435g_{4}g_{46} - 0.219g_{60}g_{66} - 0.0639g_{2}g_{56} \\
&- 0.0105g_{-}g_{-5} + 0.0334g_{25}g_{32} + 0.0406g_{1-}g_{<9} - 0.0248g_{49}g_{59} + 0.109g_{60}g_{69} + 0.00289g_{5<}g_{-2} + 0.376g_{25}g_{66} + 0.0421g_{18}g_{58} \\
&- 0.0490g_{49}g_{66} + 0.0163g_{53}g_{--} - 0.124g_{46}g_{-0} - 0.000393g_{20}g_{<8} + 0.00546g_{46}g_{56} + 0.00258g_{19}g_{24} + 0.000827g_{68}g_{-2} \\
&- 0.00177g_{16}g_{-0} + 0.00365g_{71}g_{-3} + 0.00925g_{61}g_{-5} - 0.00218g_{19}g_{<<} - 0.000573g_{9}g_{<6} + 0.0145g_{28}g_{53} - 0.0405g_{2<}g_{35} \\
&- 0.00269g_{32}g_{61} + 0.00531g_{2<}g_{-1} + 0.000473g_{6<}g_{-6} + 0.000959g_{<5}g_{46} + 0.000713g_{2-}g_{36} - 0.0000444g_{22}g_{5-} - 0.000255g_{56}g_{68} \\
&- 0.0000329g_{56}g_{<0} - 0.0000544g_{-2}g_{--} - 0.000268g_{63}g_{68} - 0.00000214g_{4<}g_{<0} + 0.0000000382g_{15}g_{18} + 0.183
\end{aligned}
$$

$$
\begin{aligned}
y_5 &= 0.182g_{1}g_{18} + 0.000280g_{5}g_{6} + 0.00467g_{24}g_{60} - 0.0467g_{23}g_{<8} + 0.00985g_{4}g_{<8} + 0.102g_{23}g_{29} + 0.339g_{33}g_{4<} - 0.201g_{66}g_{-1} \\
&- 0.0149g_{51}g_{-2} - 0.266g_{<1}g_{54} - 0.672
\end{aligned}
$$

$$
\begin{aligned}
y_6 &= -0.169g_{1}g_{18} - 0.0548g_{31}g_{44} - 0.0207g_{22}g_{<3} + 0.0826g_{49}g_{<4} - 0.00124g_{5}g_{22} - 0.0144g_{<5}g_{-2} - 0.109g_{60}g_{77} - 0.0204g_{4}g_{18} \\
&- 0.0739g_{68}g_{--} - 0.0749g_{9}g_{54} + 0.0247g_{32}g_{60} + 0.0190g_{68}g_{-2} + 0.00864g_{1<}g_{23} - 0.00686g_{16}g_{-0} + 0.110g_{48}g_{63} - 0.116g_{25}g_{-3} \\
&- 0.0563g_{24}g_{40} + 0.0202g_{69}g_{-<} - 0.0503g_{<2}g_{-3} - 0.0135g_{21}g_{<6} + 0.0504g_{54}g_{--} + 0.0181g_{16}g_{<1} - 0.332g_{51}g_{60} + 0.0901g_{19}g_{60} \\
&+ 0.105g_{3}g_{11} + 0.114g_{<4}g_{69} + 0.0200g_{29}g_{-1} - 0.00800g_{-}g_{2<} + 0.0375g_{25}g_{56} + 0.0396g_{2<}g_{-5} + 0.00247g_{16}g_{<-} - 0.403g_{2<}g_{66} \\
&- 0.00561g_{16}g_{48} + 0.00799g_{<4}g_{5<} + 0.129g_{63}g_{-1} - 0.000959g_{<}g_{<4} - 0.0107g_{-3}g_{-3} - 0.00945g_{45}g_{<9} - 0.00362g_{28}g_{6<} \\
&- 0.00355g_{25}g_{58} - 0.0119g_{2}g_{<6} + 0.000505g_{5}g_{-2} + 0.00146g_{16}g_{66} - 0.00135g_{32}g_{59} - 0.00360g_{29}g_{54} - 0.00124g_{49}g_{56} \\
&+ 0.0000572g_{14}g_{18} + 0.00960g_{4-}g_{69} + 0.00000334g_{<}g_{5-} + 0.000212g_{-6}g_{--} + 0.0000234g_{1-}g_{58} + 0.00000263g_{<}g_{65} \\
&- 0.0000250g_{60}g_{-5} + 0.0000166g_{12}g_{-0} - 0.00000330g_{0}g_{55} + 0.000000122g_{20}g_{-3} - 0.0000000433g_{22}g_{59} \\
&+ 0.000000101g_{-4}g_{--} + 0.795
\end{aligned}
$$

$$
\begin{aligned}
y_7 &= 0.0258g_{10}g_{6<} + 0.643g_{<1}g_{61} + 0.143g_{13}g_{<-} + 0.106g_{30}g_{66} - 0.0235g_{29}g_{-3} + 0.00282g_{<}g_{64} + 0.0296g_{28}g_{6<} + 0.459g_{28}g_{<3} \\
&+ 0.0857g_{52}g_{-3} - 1.28g_{60}g_{61} + 0.708g_{51}g_{60} + 0.0226g_{23}g_{58} + 0.0988g_{45}g_{62} + 0.0351g_{12}g_{50} - 0.00278g_{21}g_{22} + 0.0190g_{50}g_{54} \\
&- 0.00435g_{1-}g_{56} + 0.0394g_{26}g_{55} - 0.326
\end{aligned}
$$

$$
\begin{aligned}
y_8 &= 0.274g_{23}g_{42} + 0.834g_{13}g_{-3} + 0.102g_{11}g_{68} + 0.0571g_{60}g_{--} + 0.0457g_{40}g_{5-} + 0.300g_{60}g_{69} - 0.0471g_{32}g_{58} - 0.223g_{66}g_{-1} \\
&+ 0.00873g_{16}g_{48} - 0.253g_{26}g_{<4} + 0.00512g_{4}g_{56} - 0.0422g_{29}g_{-1} - 0.299g_{30}g_{66} + 0.0529g_{32}g_{40} + 0.143g_{60}g_{66} - 0.0585g_{61}g_{-<} \\
&- 0.00436g_{10}g_{24} - 0.232
\end{aligned}
$$

$$
\begin{aligned}
y_9 &= 0.133g_{42}g_{<0} + 0.103g_{<2}g_{-3} + 0.192g_{16}g_{-6} + 0.0242g_{<6}g_{65} + 0.0238g_{-4}g_{-<} - 0.310g_{4-}g_{<4} - 0.0532g_{56}g_{<0} - 0.000204g_{5}g_{0} \\
&+ 0.0422g_{54}g_{<5} + 0.0313g_{9}g_{<0} - 0.141g_{18}g_{5-} - 0.113g_{32}g_{42} + 0.254g_{42}g_{--} - 0.0501g_{60}g_{-6} + 0.0372g_{23}g_{<-} + 0.514g_{25}g_{66} \\
&+ 0.0442g_{2-}g_{36} + 0.319g_{60}g_{64} - 0.0290g_{8}g_{<0} - 0.00793g_{4}g_{18} - 0.00400g_{16}g_{23} - 0.00750g_{19}g_{25} + 0.0202g_{26}g_{65} \\
&- 0.204g_{55}g_{-6} + 0.0163g_{16}g_{51} + 0.0202g_{54}g_{-3} - 0.0427g_{2<}g_{48} - 0.0987g_{48}g_{63} - 0.00774g_{2<}g_{-2} + 0.00371g_{29}g_{72} \\
&+ 0.0383g_{12}g_{61} - 0.0251g_{49}g_{60} + 0.00784g_{11}g_{<8} + 0.0223g_{2<}g_{10} + 0.0389g_{2-}g_{-<} - 0.0142g_{1<}g_{-3} - 0.0125g_{25}g_{-5} \\
&+ 0.0157g_{43}g_{-1} + 0.00123g_{21}g_{42} - 0.0350g_{51}g_{69} - 0.000426g_{16}g_{-<} - 0.00543g_{11}g_{6<} + 0.0716
\end{aligned}
$$

(2)

60 個の質問回答文に対し，281 個の特徴量を説明変数とし，9 因子の因子得点を目的変数として，ステップワイズ選択法による重回帰分析[13]を行った．この結果，重回帰式(2)が得られた．この時の重相関係数を表 7 の「二次項」の列に示す．どの因子も重相関係数の値が 0.9 以上となっており，どの因子も推定精度が良好であるといえる．

6. 考察

3.で示した文章の特徴量が64個の場合と，5.で示した文章の特徴量が77個の場合とで，重相関係数の値が変動しているかどうかを比較検討する．表3と表7の結果を表8にまとめて表す．単項のみの場合，第3因子と第5因子に関しては値が向上しているが，第7因子と第9因子に関しては，値がやや低下している．残りの5因子に関しては，値に変動が見られなかった．

二次項に関しては，第2因子，第3因子，第4因子，第6因子，第8因子，第9因子の6因子に関しては値が向上している．一方で，第 1 因子，第 5 因子，第 7 因子の 3 因子に関しては値が低下している．しかしながら，特徴量が 64 個の場合では重相関係数が 0.9 に及ばなかった第3因子と第6因子に関しては，値が大幅に向上している．結果的には，全ての因子において重相関係数の値が 0.9 を上回っているため，どの因子も推定精度が良好であるといえる結果となっている．従って，文章の特徴量を追加したことにより，推定精度の更なる向上に成功したといえる．

表 7 重相関係数(特徴量 77 個)

因子	重相関係数	
	単項	二次項
第1因子(的確性)	0.832	0.989
第2因子(不快性)	0.774	1.000
第3因子(独創性)	0.788	0.999
第4因子(容易性)	0.728	1.000
第5因子(執拗性)	0.908	0.925
第6因子(曖昧性)	0.872	1.000
第7因子(感動性)	0.552	0.963
第8因子(努力性)	0.650	0.950
第9因子(感動性)	0.674	1.000

表 8 重相関係数の比較

因子	特徴量64個の場合		特徴量77個の場合	
	単項	二次項	単項	二次項
第1因子(的確性)	0.832	1.000	0.832	0.989
第2因子(不快性)	0.774	0.947	0.774	1.000
第3因子(独創性)	0.744	0.877	0.788	0.999
第4因子(容易性)	0.728	0.908	0.728	1.000
第5因子(執拗性)	0.893	0.966	0.908	0.925
第6因子(曖昧性)	0.872	0.899	0.872	1.000
第7因子(感動性)	0.581	0.997	0.552	0.963
第8因子(努力性)	0.650	0.904	0.650	0.950
第9因子(感動性)	0.683	0.954	0.674	1.000

7. まとめ

本論文では，質問者と回答者の相性の判定をめざして，文章の因子得点の推定精度の向上を行った．ここでは，構文情報，単語心像性と文末表現に加え，単語親密度と表記妥当性を文章の特徴量として使用し，質問回答文の因子得点の推定を試みた．その結果，全因子において精度良く推定できることを示した．

今後は，質問文に対して適切な回答が見込める利用者を選出する手法の確立を行い，そのようなプロトタイプシステムの構築・評価を行う予定である．また，今回得られた重回帰式を用いて，まだ求まっていない質問回答文の因子得点を求める．これをもとに「ベストアンサー」の推定を行い，その推定手法を確立させていく．更に，客観的な視野が加味された「グッドアンサー」[18]の推定手法も確立した上で，「ベストアンサー」との比較検討を行っていく予定である．

謝辞

本研究は一部，科研費(21500091)の助成を受けて行われたものである．また，実装・評価に際し，大学共同利用機関法人国立情報学研究所から提供を受けた，Yahoo!知恵袋のデータを利用している．ここに記して謝意を示す．

参考文献

[1] Yahoo!知恵袋, http://chiebukuro.yahoo.co.jp/

[2] 横山友也，宝珍輝尚，野宮浩揮，佐藤哲司：質問回答サイトの質問文と回答文の印象評価とベストアンサーの推定，日本感性工学会論文誌，Vol.10, No.2, pp.221-230, 2011.

[3] 横山友也，宝珍輝尚，野宮浩揮，佐藤哲司：文章の特徴量を用いた質問回答文の印象の因子得点の推定，第 14 回日本感性工学会大会，B1-01, 2012.

[4] Blooma, M.J. and Chua, A.Y.K. and Goh, D.H.L.: A Predictive Framework for Retrieving the Best Answer, Proc. of 2008 ACM Symposium on Applied Computing (SAC08), pp.1107-1111, 2008.

[5] Agichtein, E., Castillo, C., Donato, D., Gionis, A. and Mishne, G.:Finding High-Quality Content in Social Media, Proc. of the Int'l Conf. on Web Search and Web Data Mining (WSDM08), pp.183-194, 2008.

[6] Wang, X. J., Tu, X., Feng, D. and Zhang, L.: Ranking Community Answers by Modeling Question-Answer Relationships via Analogical

Reasoning, Proc. of 32nd Int'l ACM SIGIR Conf., pp.179-186, 2009.

[7] Kim, S., Oh, J. S. and Oh, S.:Best-Answer Selection Criteria in a Social Q&A site from the User-Oriented Relevance Perspective, Proc. of American Society for Information Science and Technology (ASIS&T) 2007 Annual Meeting, 2007.

[8] Adamic, L. A., Zhang, J., Bakshy, E. and Ackerman, M. S.:Knowledge Sharing and Yahoo Answers: Everyone Knows Something, Proc. of 17th Int'l Conf. on World Wide Web (WWW2008), 2008.

[9] Jurczyk, P. and Agichtein, E.: Discovering Authorities in Question Answer Communities by Using Link Analysis, Proc. of 16th ACM Conf. on Inf. and Know. Management (CIKM2007), pp.919-922, 2007.

[10] Hovy, E., Gerber, L., Hermjakob, U., Junk, M. and Lin, C.-Y.: Question Answering in Webclopedia, Proc. of 9th Text Retrieval Conf., pp. 655-664, 2000.

[11] 西原陽子, 松村真宏, 谷内田正彦:Q&A コミュニティでの質疑応答パターンの理解, 第 22 回人工知能学会全国大会, 1H2-7, 2008.

[12] 菅民郎:初心者がらくらく読める多変量解析の実践上, pp.42-45, (社)現代数学社, 1993.

[13] 重回帰分析(ステップワイズ変数選択), http://aoki2.si.gunma-u.ac.jp/R/sreg.html

[14] 菅民郎:初心者がらくらく読める多変量解析の実践上, (社), p.39, (社)現代数学社, 1993.

[15] 佐久間尚子, 伊集院睦雄, 伏見貴夫, 辰巳格, 田中正之, 天野成昭, 近藤公久:単語心像性①, NTT データベースシリーズ日本語の語彙特性 第 3 期(第 8 巻), (社)三省堂, 2005.

[16] NTT データベースシリーズ, http://www.kecl.ntt.co.jp/mtg/goitokusei/

[17] 天野成昭, 近藤公久:単語心像性①, NTT データベースシリーズ日本語の語彙特性 第 1 期, (社)三省堂, 2003.

[18] 情報学研究データリポジトリ :「Yahoo!知恵袋データ(第 2 版)」の提供について, http://www.nii.ac.jp/cscenter/idr/yahoo/chiebkr2/Y_chiebukuro.html

第18回公開シンポジウム
「人文科学とデータベース」
2012. 12. 22

古代を中心とした歴史地理データベースの試み
The trial of the with historical GIS database about ancient Japanese

宮崎 良美
Yoshimi Miyazaki
奈良女子大学 古代学学術研究センター，奈良市北魚屋東町
Nara Women's University, Kitauoyahigashi-machi, Nara, Nara

あらまし：古代から中世の荘園や村落景観を研究する上で重要な条里（耕地を1辺約 109m の格子状に区画した地割と、これを「三条」「一坪」などの数詞によって示す条里坪付呼称法）は、仮想的にグリッドデータとして表せるため GIS にもなじみやすい。本報告では、古代から中世にかけての奈良盆地歴史地理データベースのひとつである「条里・条坊関連史料データベース」の構築と活用の試みについて報告する。

Summary: In this paper the author has examined the matters which were caused in constructing a historical GIS database of ancient Nara Basin, Japan, where agricultural land readjustment had been practiced in ancient times and has made a brief memoir on the result of a trial use of the database. The readjustment changed paddy fields into the grid pattern of approximately 109meters in length on each side and introduced a uniform system that indicated locations by a numbering method which was based on the areal unit of *cho*, the rectangular land division. The land system, the *Jori* system, is very important when studying the landscape of ancient times to medieval times. Since each *cho* has individual number according to the *jori* system, it can be virtually expressed as grid data with geographical location and easily adopted to GIS analysis.

キーワード：歴史 GIS，条里，奈良盆地
Keywords: Historical GIS, *jori* system, Nara Basin

1. はじめに

奈良盆地は、古代において平城京・藤原京をはじめとして数々の宮都が営まれ、長岡京に遷都した後も長く重要な地域であり続けた。興福寺をはじめとする大寺院の膝元でもあり、多くの荘園が経営され、土地売買などの証文や検地帳はじめとする史料や、土帳などの地図資料などが数多く残されている。

このような土地関係史料は、最も基礎的な歴史地理研究の資料であり、奈良盆地に関わる豊富な史料群は、古代・中世の荘園や村落の構造や寺院経済史などの視点から詳細な分析・考察が加えられ、研究成果が蓄積されている[1]。

史料にめぐまれた場所では景観変遷を追究することが可能であり、例えば皿池といわれる溜池の築造と耕地の関係[2]など開発史的な視点からも興味深い成果が期待できる。そこで、奈良時代以前の奈良盆地における開発状況が明らかにできれば、古代の宮都が営まれた前後の時期を通じた景観変遷をとらえることができると期待される。そこで、これらの史料群や既往の研究成果をもとに、土地利用に関わる情報を収集して「条里・条坊関連史料データベース」の構築を進めている[3]。

このデータベースの根幹は、奈良盆地を覆う条里と条坊の地割である。条里地割は1辺約 109m の坪と呼ばれる方格地割を基本とする。また、奈良盆地北部のかつての平城京域には、条里地割よりも一回り大きい方格の条坊地割が残存している。

そして、平城京域では「左京二条六坊十三坪」の

ように条坊の呼称で、その他の地域は条里の坪付に基づく「添上郡京南五条五里一坪」のような郡名や数詞による坪付(以下、条里呼称とする)によって所在地を示している。

さらに、奈良盆地については112葉の1/5000地図上に条里坪の区画と呼称を復原した、奈良県立橿原考古学研究所編『大和国条里復原図』[4](以下、『条里復原図』とする)が刊行されている。平城京域は『条里復原図』中に条坊が復原されているほか、奈良文化財研究所編『平城京条坊復元図』、『平城京条坊総合地図』[5]などが作成されている。

これらの成果により、GIS上に条里・条坊の土地区画を条里呼称によって識別されるグリッドデータとして復原でき、条里呼称が記載された史料群は位置情報を伴う地理情報として扱うことが可能となる。両者を関連づけることで、古代・中世の史料群の内容を約109m四方の解像度で地図上に示すことができ、様々な分析が可能となる。

そこで、条里の概要と、Historical GISによる「条里・条坊関連史料データベース」の構築の経過について簡単にふれた後、GISを利用することにより条里についてどのようなことが検討可能となるのか若干の試みを行なったので報告することにしたい。

2. 奈良盆地の条里

先述のとおり、条里は主に古代から中世に行われた耕地区画と、条里呼称の制度である。概ね奈良時代に畿内を中心に施工が始まり、平安時代にかけて全国的に拡大したとされる。その遺構は奈良盆地をはじめとして大阪平野や滋賀県の湖東平野などの沖積平野を中心に、東北地方の秋田平野や横手盆地から南九州の鹿児島県国分平野や川内川下流平野、さらには佐渡島、隠岐島などの離島や、阿蘇火口原などの山間盆地にまでみられた(図1)。

近代に入り、1899年(明治32)耕地整理法の制定により、用排水改良事業や区画整理が行われ[7]、さらに1961年に農業基本法が施行されて圃場整備事業が行われるようになると、全国的に条里地割が消滅することになった。しかし、幸い奈良盆地では圃場整備事業が進展せず、多くの条里地割が残存している。条里関連地名も豊富に残されており、『条里復原図』に見られるように、条里呼称や地割の復原を可能にしている。

この奈良盆地の条里の特徴についてふれておきたい(図2、図3)。1辺が1町(60 歩≒109m)の坪区画を基本とし、6 町間隔に縦横に平行する道路・水路・畦畔などで方格に区画し、奈良盆地の場合、東西の列を条と呼び、南北の列を里とするため、6町×6町=36 町で 1 里をなす[8]。そして、奈良盆地の主要な条里は、旧平城京の南端を1条として、南へ2条、3条と進む。条は路西条里の一部に、郡により数え方が異なる箇所があるものの、現・大和郡山市付近の「添下郡京南一条」から御所市・五條市の市境付近の「葛上郡四十二条」まで連続する。里は盆地のほぼ中央を南北に通る下ツ道を起点とし、盆地縁に向かって1里、2里・・・と進む。下ツ道に接しない場合もこれに近い方を起点とする。

このように平城京の南に広がる奈良盆地の主要な

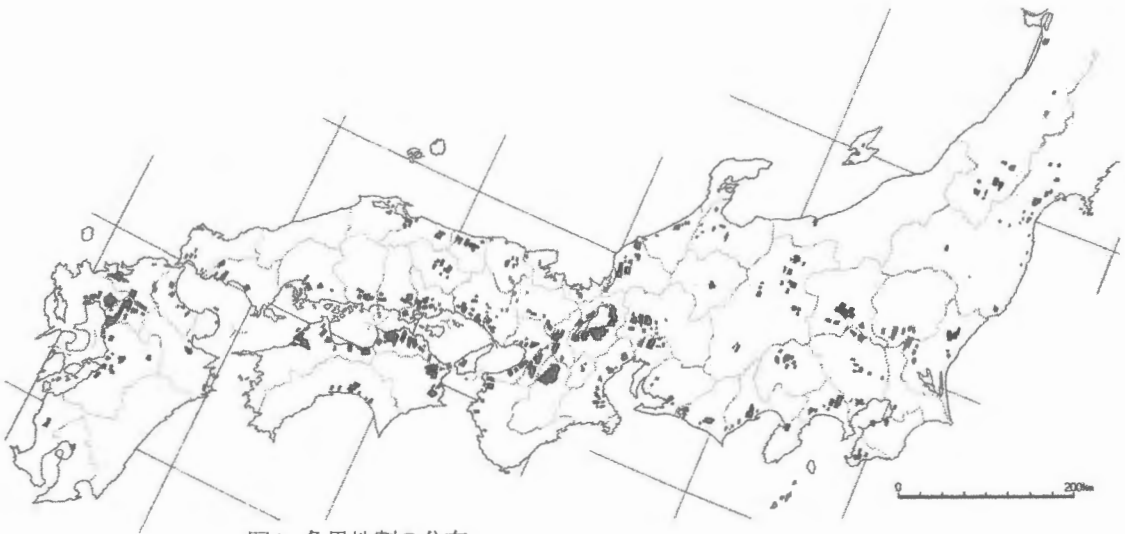


図1 条里地割の分布

出典:石井素介ほか編『図説日本の地域構造』(古今書院、1986年)p28の図を一部変更

条里は条里呼称が連続することから「統一条里」あるいは「基本条里」と呼ばれる。そのうち下ツ道の東を路東条里、西を路西条里と呼び、平城京があった添上郡・添下郡では特に京南条里という。

一方、奈良盆地周辺には、小盆地の内平群条里や、馬見丘陵の地割は確認できないものの史料に表れる葛門条、真野条里があった。平城京周辺では左京の南に統一条里とは坪並が異なる京南辺条条里、京の東には京東条里、右京の北には西大寺所蔵「大和国京北班田図」に描かれた京北条里があった。これら京周辺の条里は、条里の施工時期と平城京造営の問題に関わって、造営規格などの問題を残している。

3. 条里・条坊関連史料データベースの構築

次に、条里・条坊関連史料データベースの概要を、作業工程に沿って述べておきたい(図4)。

まず、竹内理三編『平安遺文』などの史料集から奈良盆地の条里の坪付記載を含む古文書を集め、本文テキストデータを作成した。そして、土地利用に関わる条里呼称、荘園名、地目、面積、土地所有者などの項目について情報を抽出し、Microsoft Excel に入力した[9]。また、史料の本文は html ファイルにし、データベースとして利用する ArcGIS 上でも閲覧できるようにリンクさせている。

現在、約 900 点の文書が入力されている[10]。日付が最も古いものは霊亀二年(716)で、最も時代が下るものは応永十三年(1406)であり、約 700 年間にわたる。日付が明記された文書は約 340 ヵ年についてあり、記載の多寡にかかわらず、約 2 年に 1 点の割合で何らかの土地史料が得られたことになる。

条里に関する坪付史料についてみると、天平二十年(748)「弘福寺三綱牒」(『大日本古文書』3-41)が最も古く、天平勝宝二年(750)「出挙銭解」(『大日本古文書』3-405)、神護景雲元年(767)「太政官符」(『新訂増補国史大系』25 巻 446 頁)と続く。「弘福寺三綱牒」は偽作文書の可能性が指摘されており[11]、「出挙銭解」は里名を欠くため、8 世紀後半

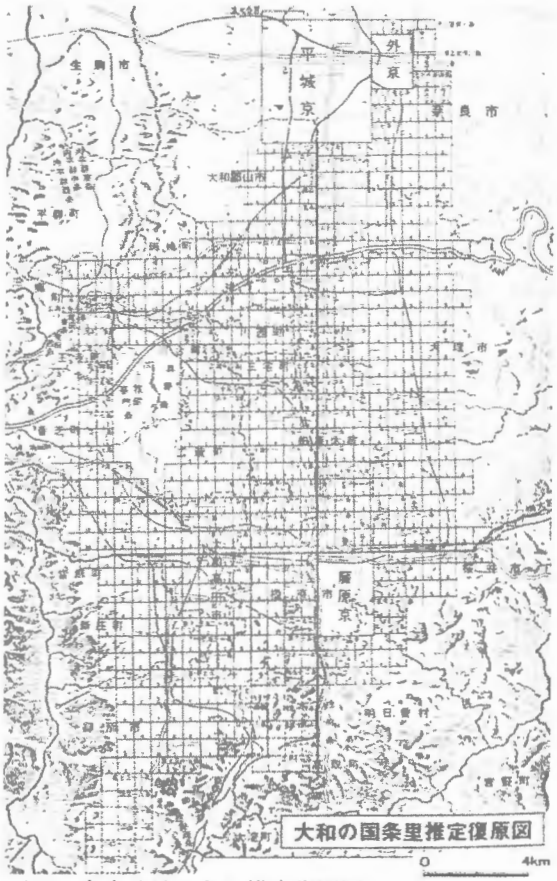


図2 奈良盆地の条里推定復原図

出典:池田末則・横田健一監修『日本歴史地名大系 30 奈良の地名』、平凡社、1981 年、巻頭図

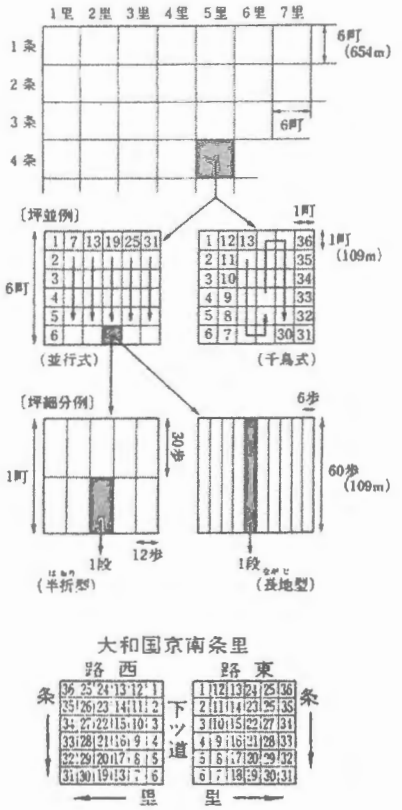


図3 条里区画の方法

出典:奈良県史編集委員会編『奈良県史4 条里制』名著出版、1987 年、p4 の図を一部変更

以降の史料について本データベース上で扱えるということになる。

一方、GIS 上の条里・条坊坪グリッドは、次章で詳述する条里モデルの作成と、『平城京条坊復元図』のトレースにより、盆地全域で約 2.7 万坪ある。そのうち史料に表れる坪は約 5000 である。1坪あたりの文書数では、広瀬郡十三条三里十坪、十七坪(現在の奈良県広陵町)が最多で 31 点あり、東大寺の荘園に関係して検地帳などが数多く残されていることによる。一般的には約 1.3 点程度である。最も多く坪付記載を持つ史料は、延久二年(1070)「興福寺雑役免坪付帳」(『平安遺文』4639、4640)で、奈良盆地に限れば約 3900 坪分の情報が得られる。

本節では条里・条坊関連史料データベース構築について概要について述べた。次に条里モデルの作成について述べていくことにしたい。

4. 条里モデルの作成

条里区画の GIS データ化に際し、『条里復原図』により遺存する、あるいは復原された条里地割を、GIS 上でトレースしてフィーチャを作成することも考えられた。だが、河川や旧河道の周辺や山麓部のように条里遺構が遺存しない、あるいは地割が乱れるなどトレースが困難な箇所も多い。一方で、奈良盆地は統一条里にほぼ覆われることから、条里の土地区画をグリッドデータとして作成しても、現実の区画と位置や大きさ等に大きな齟齬が生じないことも予想された。そこで、試みとして条里モデルを作成し、条里モデルと実際の条里遺構について比較・検討を加えることにした。

その工程は次の通りである。まず、『条里復原図』の地図112葉を、ESRI 社 ArcGIS 上で位置情報を与えて幾何補正した後、図中の条里界線の交点について国土座標を計測した(図5)。下ツ道は道路の両端と条里界線の交点が図中に示されているため、両端の点を計測の対象とした。横大路は道路端が路東 24 条の北から 3 坪目にあたるため、条里界線の交点としての計測の対象となっていない。地割が乱れて交点が確認できない箇所もあるため、結果として計 394 点の計測値を得た(図6)。

次に、奈良盆地の坪長の平均、南北 109.5m、東西 109.2m[12]をそれぞれ 6 倍して、南北 657m、東西 655.2m で、国土座標の北に対し傾き 0° の格子状データを、統一条里にほぼ重なるように作成した(図7)。下ツ道部分では『条里復原図』上で道路端の座標が取得できていることから、路東1里と路西1里にあたる格子間に道幅分の余剰部を設けることにした。発掘調査により下ツ道の道幅が確認されている箇所

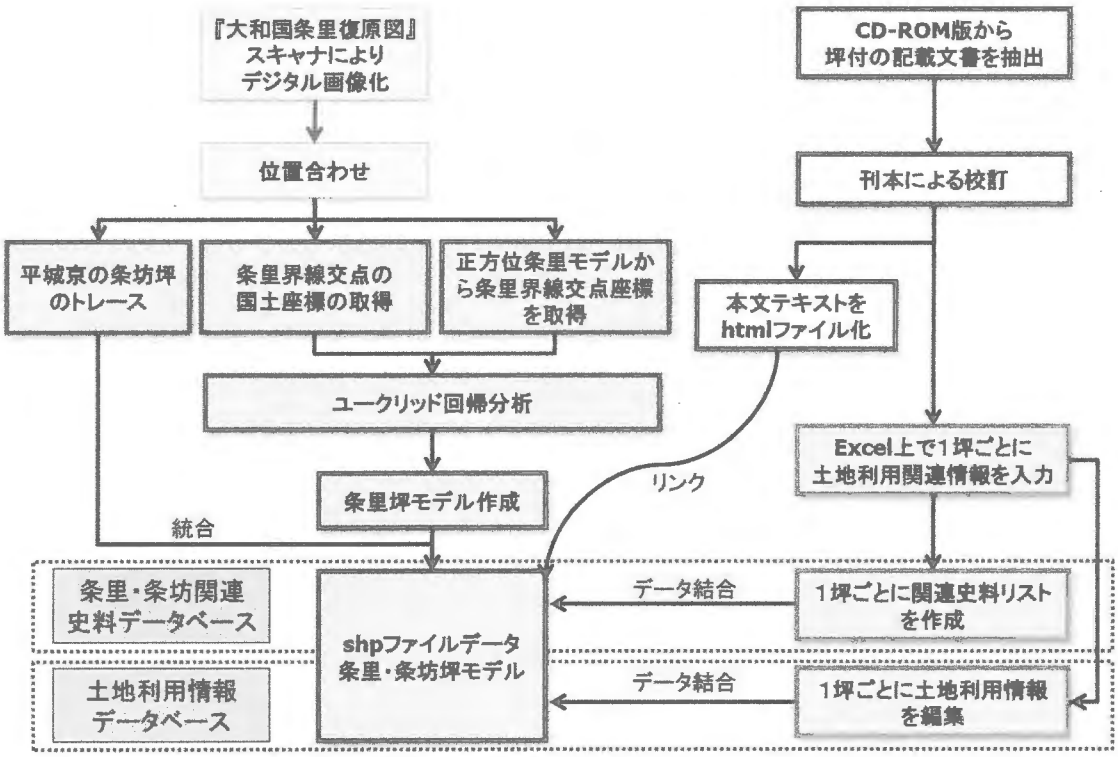


図4 条里・条坊関連史料データベースの工程

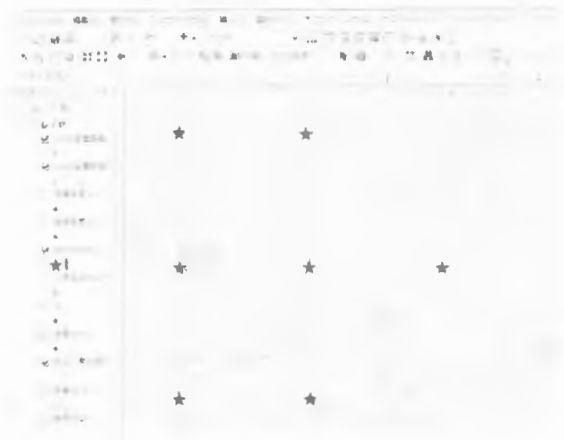


図5　条里界線交点の座標計測
地図出典::奈良県立橿原考古学研究所編(1981)
『大和国条里復原図』

もあるが場所によって異なる。統一条里の基点とされる下ツ道と横大路の交差点は図中に示されていないため、横大路の道路痕跡上で、下ツ道から1坪東と西にある坪界線の交点間の距離から2町分の長さ109.2×2mを引いた約46.3mとして設定した。このようにして作成した格子状データの各頂点の座標を計測し、仮想の条里の統一規格の座標値とした。

そして、現実と仮想の規格との間で対応する交点について、仮想の規格の交点座標を現実のものへ近づけるためユークリッド回帰分析を行った[13]。結果として得られたユークリッド回帰式のパラメータの推定値を表1に示した。

表1　奈良盆地の条里モデルに関するユークリッド回帰分析の結果

パラメータ	値
伸縮(c)	1.0009
回転角(θ deg)	0.247
切片(b_1)	41.832
(b_2)	−59.063
2次元相関係数	0.999884

回転角は方向統計量である。

スケールの伸縮率を示すパラメータcは1.0009となり、現実の条里長がわずかに大きかったことを意味している。回転角を表すパラメータθは0.247°で、仮想の条里を反時計回りに0.247°回転すると現実の条里に最も合うことを表す。平行移動を表す切片はb_1が41.832、b_2が−59.063で、仮想の条里を東に約41.8m、南に59.1m平行移動すると、両者のずれが最も小さくなることを意味する。

以上のパラメータをもとにユークリッド変換された仮想の条里の座標値に基づいて再度格子状のデータを作り、このなかを36等分して、条里モデルとした。条里モデルでは坪長は南北109.6m、東西109.3mとなり、座標北から西へ約14.82′振れていることになる[14]。

「条里・条坊関連史料データベース」のフィーチャとして利用するために、平城京域の条坊データとの統合などを行った。また作業の便宜上作られた現実

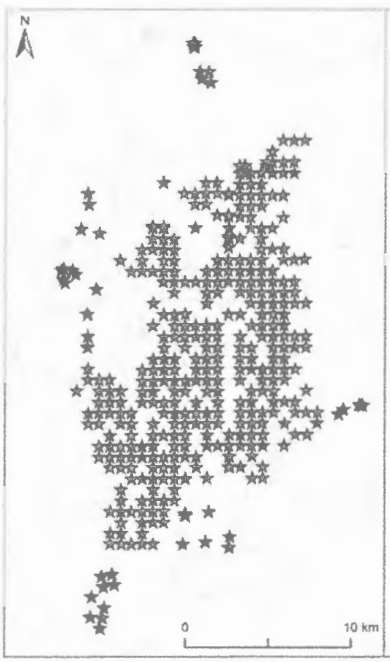


図6　計測した条里界線の交点

図7　格子状データの頂点

図8　条里・条坊モデル

に存在しない坪は、『条里復原図』中に条里界線を示していなくても、坪付史料がある場合もあるので、現在のところ削除せず残している(図8)。

4. 条里地割とモデルのずれの検討

奈良盆地の条里地割の施工時期をめぐっては見解の相違があるものの、古代のある時期に短期間に盆地全体に施工されたものではない、という認識でほぼ共通している[15]。とすれば、条里施工の基本的な設計プラン(統一的なプラン)が存在したのか、それとも、異なる設計プランで施工された条里地割が並存しているのか検討の余地がある。

これまでの地割研究、例えば条里地割や古代のアガタやミヤケ、近世の城下町プラン研究等により、地割線の方位のずれや差は施工時期や施工単位の相違に起因するとの見解がなされることが多い。これを参考にすれば、作成された条里モデルと現実に地表面に遺存する条里地割とのずれが地域的なまとまりを示すならば、それは施工時期や施工区の違いを反映するものである可能性があることになる。

そこで、現実の条里と条里モデルとの間のずれについて分布の点から分析し、地域的な区分が可能なのか検討することにしたい。

(1)現実の条里とモデルのずれの分布

条里モデルの条里界線の交点を基準として、対応する現実の交点までの距離、つまりずれの大きさとその方向を地図に示した(図9、10)。

ずれの大きさをみると盆地周縁部の小条里区、葛城川や初瀬川、高取町の吉備川の谷や、桜井市の寺川が盆地に出る谷口部では、100m 以上つまり 1 町以上のずれが生じており、最大約 600m に達する。平城京周辺でも、京北条里や京南辺条条里のずれは大きい。京北条里については、平城京を隔てているためにモデルの見直しなど検証が必要であろう。以下ではこれら小条里区は除いて分析を進める。

統一条里では、ずれの量は最も小さいもので 1.7m、最大 80m であり、1 町以上のずれる交点はなかった。また、盆地中央部では相対的にずれは小さく 40m を超えるものはなかった。盆地南部をみると、平地部から盆地東山麓に向かってずれが累積するように大きくなる傾向がある。葛下～葛上郡や馬見丘陵南側でも同様である。奈良盆地北部では、一見山麓・丘陵部でずれが大きくなる傾向があるようだが、北東部の京東条里地区では平地部の京南条里よりもずれが小さく、西部の平群郡の、現・安堵町付近では、盆地中央に近い東側でずれが大きく表れる。

また、ずれの大きさが切り替わる境界となっている箇所がある。例えば、路東条里の山辺郡から式下郡にかけての 11～18 条 4 里にあたる部分では、その東側で、西側よりずれが大きくなっていることがわかる。

(2)ずれの方向

次に、条里モデルから見て、現実の条里地割がどちらの方向にあるのか、ずれの方向について図10からみていくと、まず、いくつかの矢印が一定の方向を向いている地域的まとまりがあることに気づく。この矢印の方位が変わる箇所を、ずれの大きさと比較してみると、当然のことながら先に挙げた山辺郡から式下郡にかけての11～18条4里の交点列のように、ずれの大きさが切り替わる境界に一部が合致する。

さらに、路東 23 条・24 条のラインでは、矢印の向きが北東と南東に盆地を横断するように切り替わっている。ここに横大路が通っているため、古代道路の道幅がずれに反映されている例である。また、馬見丘陵南東麓の広瀬郡条里では、低地部側から取り囲むように東向き～北東向き～北向きの矢印の列がある。これは広瀬郡条里の南側では広瀬郡界、東側では郡界あるいはこれに沿って流れる曽我川に規定されたものであろう。ここでは、曽我川の西に葛城川が北流するが、その影響のようなものは顕著でないため、郡界がずれを規定する要因となっていると考えられる。

以上から、現実と条里モデルのずれは、地割の施工に際しての基準線・境界とされてきた古代道路や河川、郡界などの存在が反映されると考えてよい。この地域的な傾向は、地割の形状や地形・傾斜などの要素もふまえた検討を経る必要はあるが、統一条里における施工区のような単位の範囲や基準線、規格の変換点をマクロスケールで検出する手がかりとなるであろう。

5. ずれの地域性と条里

条里地割とモデルの比較検討により得られたずれ大和川・布留川から一定の距離をおくと、山辺郡・式下郡それぞれに小グループが存在するのがうかがえる。山辺郡ではこれは(3)群の一部と見ることがで

図9　条里地割とモデルのずれの量　　図 10　条里地割とモデルのずれの方向

DEM データ国土地理院『数値地図 50m メッシュ(標高)』より作成。後掲図も同データによ

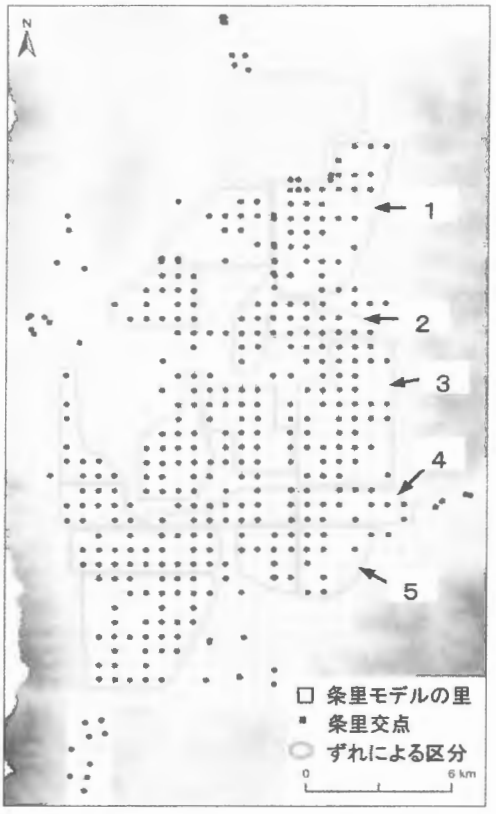


図11　ずれによる条里の区分

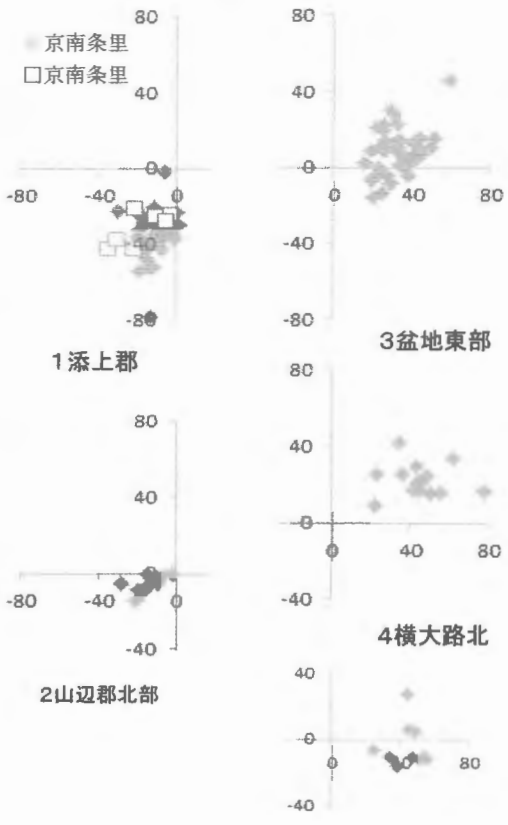


図 12　条里区分のずれ(単位:m)

きる。式下郡では南東へ6～9mずれる2点が東西に並ぶが、すぐ西のドツ道や南の式下郡路東15条の東ずれのグループも異なる。

この分布について整理することで、統一条里はおおよそ図11のようなまとまりに区分される。なお、この区分は数個以上の条里界線交点を含むものとし、1～3点程度のものは保留としている。この地域区分ごとに分析を行いやすいように、ずれの東西成分をX軸に、南北成分をY軸にとって散布図を作成した(図12)。紙数の都合もあるので、路東条里を取り上げてみてみよう。

(1)添上郡

南西～南へのずれが卓越する。ずれの大きさは平均37.3m、南北方向のずれの大きさは、南へ2.0～77.0m、東1m～西34.9mの範囲にある。このうち、京東条里は、東大寺の占地や地割の起源についてなど課題が残る地域であり[16]、区分は慎重を期さねばならないが、ずれの傾向に限っていえば京南条里と同じまとまりとみられる。しかし、先にふれたとおり盆地中央から離れても、ずれの量が比較的小さいことなどから周囲のずれの要因などの検討も必要だが、別グループに分けられる可能性も残る。

山辺郡界を越え、山辺郡7条を含むことについては、郡界の移動が考えられたが、「条里・条坊関連史料データベース」によって11～13世紀の郡界について推定してみたところ[17]、当時の郡界が近世末とは異なるとは推定できたが、この7条を含む程南にあったとは考えられず、課題としておきたい。

(2)山辺郡北部

西へ平均約14.6mずれる交点群で、山辺郡8条～10条にあたる。だが、10条2、4里の交点は、ずれの方向が異なるため、本グループから外している。南北方向のずれが小さいことも特徴的である。

この10条のラインが(1)(2)群など主に南西方向へずれるグループと(3)群以南の東～東北方向へずれるグループの境界になっているようである。

(3)盆地東部

山辺郡11条4里付近から、式下郡15条2里付近、十市郡23条2里付近以東を1つの範囲とした東方へのずれが特徴的で約16.0～60.2mのずれが見られる。ずれの平均は約35.9mである。式下郡域ではずれの境界が山辺郡より1里分西側に表れている。散布図で、北東方向に飛び出しているのは、式上郡20条5里の西南角にあたる点である。三輪山麓の傾斜地であることや、巻向川の影響も考えられる。

(4)(5)盆地東南部－横大路北・南

横大路を挟んで北と南に分けたが、東へ大きくずれる傾向は(3)群と同様である。横大路の北については散布図の点にあまりまとまりがない。より東にあるのは式上郡、相対的に西にあるのは十市郡内の交点のようである。

横大路の南では、ずれの南北成分から、南へ6～16mずれるものと、北へ4～26mのものの2グループに分かれる。後者は飛鳥川右岸側にある。

このほか、山辺郡西部では、ずれの向きに一定した傾向がないため地域区分を設定していない。ここは山辺郡・式下郡の郡界にあたり、初瀬川(大和川)と布留川の合流点にも近い。ふたつの河川や旧河道の痕跡に伴う地割の乱れがあり、これがずれに影響しているものと考えられる。

以上、現実の条里と条里モデルの交点座標のずれと分布について検討し、奈良盆地の統一条里内の区分を試みた。路東条里では山辺郡10条付近を境にして、北部と南部に分かれる。また、北にあるグループが南西へ、南にあるものは北東にずれる傾向があることから、路東条里全体では、条里モデルよりやや西に振れた基準線を持ち、坪区画はやや小さい規格で施工されていると推測される。さらにこの中が5グループのようなブロックに分かれているものと捉えられる。

なお、南部で全体的に東にずれていることについては、北部はドツ道から盆地縁まで含むのに対し、南部は2里以東のみを含むためで、ドツ道を中心とする地域を含めると、この中間に収まってくる。

また、それぞれのグループ内で特異なずれ方をするものについては、河川などによる地割の乱れなど地形の影響によるものが多く見られた。

7. むすびにかえて

本稿では、古代を中心とした歴史地理データベースとして、奈良盆地の条里地域を中心としたデータ

ベースの構築について報告した。このなかで、条里地割のフィーチャとするために、条里モデルを作成した。これを現実の条里と比較し、そのずれを分析することで、統一条里内に地域的なまとまりを見出し、条里施工区の範囲やその基準線などについて検討する手がかりが得られるなど、地図表示に用いるばかりでなく、積極的な活用ができることを見出した。

本報告で行った条里の地域区分は試みとしてのものであり、本データベースによる史料、坪長や条里の内部地割などの関連、古地形の復原的研究や、近年の発掘調査成果などをふまえた多角的な検討を重ねる必要がある。また、今の条里モデルは周辺の小条里区を含めたユークリッド回帰分析によるものであるため、小条里区と統一条里の大きなずれに引きずられている可能性も考えられる。今のままでもずれの地域性分析は可能であったと考えるが、小条里区を除いた分析を行うことで、統一条里により精度よく重ねられるモデルが得られ、より的確な知見を得ることができるであろう。これらは今後の課題としたい。

また、歴史資料を扱う上で、偽作文書のような史料そのものの信頼性の確認と、偽作文書も作成された時期の何らかの事情を反映すると考えられることから、その経緯にまつわる情報などもデータベースに反映することも課題といえよう。

条里研究については多くの研究が積み重ねられている。本報告も基礎資料として『条里復原図』を利用することにより実現できたものである。しかし、条里のように広範囲にわたる地割を大縮尺図で検討することは、大変な困難を伴う。それに対して GIS には、例えば大縮尺での坪長の計測やその分析結果を一覧できる形で地図上に示せたり、位置情報を与えておけば複数の資料を一度に重ね合わせて閲覧できたりするメリットがある。『条里復原図』刊行以後、奈良盆地内の発掘調査成果の蓄積もなされており、条里・条坊関連史料の検討をあわせて総合的に分析することで新たな知見が得られる可能性は高い。

最後に、幸い奈良盆地では条里遺構がよく残り、大規模な破壊を免れ、条里プランの復原もなされている。しかし今、全国的に圃場整備完了から年数が経ち、かつての土地利用の記録・記憶が失われつつあり、平成の大合併や公図のデジタル化等に伴う旧地籍図類等の保存等の問題もある。Historical GIS データベースは、そのメリットを活かすことで、景観復原のための公図をはじめとする土地改良関連等の諸資料の保存のための有効な対策の 1 つとなる。これを視野に入れた研究の進展も課題としてあげておきたい。

謝辞　本稿の作成にあたり奈良女大学出田和久先生、石崎研二先生に多くのご助言をいただきました。ここに感謝の意を表します。

文献

1)朝倉弘(1984):『奈良県史 10　荘園』、名著出版。泉谷康夫(1972):『律令制度崩壊過程の研究』、鳴鳳社。片平博文(1978):「大和国乙木荘の歴史地理学的研究」。人文地理 30-2、pp136-153。室伏朝子(1986):「延久 2 年「興福寺大和国雑役免坪付帳」の地理資料としての検討」、人文地理 38-2、pp73-88。渡辺澄夫(1956):『畿内庄園の基礎構造』、吉川弘文館、など多数ある。

2)金田章裕(1993):『微地形と中世村落』、吉川弘文館。

3)「条里・条坊関連史料データベース」は、奈良女子大学 21 世紀 COE プログラム「古代日本形成の特質解明の研究教育拠点」において構築が始まり、出田和久教授の構想と石崎研二准教授の GIS に関する技術的なアドバイスに基づいて、実務を著者が担当している。奈良盆地歴史地理データベース群の概要は次の文献に紹介されている。拙著(2010):「奈良盆地歴史地理 GIS データベースの構築と課題」古代学 1、pp55-68。出田和久(2011):「条里・条坊関連史料データベースについて」(山田奨治編『近代日本の歴史的時空間データマイニングのための基盤整備』日本学術振興会科学研究費補助金平成 19～22 年度基盤研究(A)課題番号 19200019 研究成果報告書(代表:山田奨治))、pp108-114。出田和久(2012):「奈良盆地歴史地理データベースの構築とその利用」(HGIS 研究協議会編『歴史 GIS の地平』、勉誠出版、pp197-207。

4)奈良県立橿原考古学研究所編(1981):『大和国条里復原図』財団法人由良大和古代文化研究基金。

5) 奈良文化財研究所編『平城京条坊復元図』(1995年作製1万分の1奈良市全図に1996年編集焼付)。奈良文化財研究所編(2003):『平城京条坊総合地図』奈良文化財研究所。

6) 前掲注3)など

7) 今村奈良臣・佐藤俊朗ほか(1977)『土地改良百年史』、平凡社。

8) 岩本次郎(1987)「条里制－大和における寸描」(木村芳一ほか編『奈良県史4　条里制』、名著出版、pp3-8)

9) 奈良女子大学舘野和己教授の監修により竹内亮助教(当時)、大戸香美氏、中川由莉氏(同大学院生:当時)が史料からのデータ作成を担当した。

10) 竹内理三編『平安遺文』、同『鎌倉遺文』(東京堂出版)、筒井俊英校訂(1971):『東大寺要録』(国書刊行会)、藤田経世『校刊美術史料　寺院編』上～下、中央公論美術出版、東京大学史料編纂所編『大日本古文書』(編年)、同『大日本古文書家分け18　東大寺文書』について入力作業を進めている。

11) 福山敏男(1970):「かわらでら」(『日本歴史大事典』河出書房)、井上寛司(1972):「弘福寺領大和国広瀬荘について」(『赤松俊秀教授退官記念国史論集』)。石上英一(1987):「弘福寺文書の基礎的考察－日本古代寺院文書の一事例－」東洋文化研究所紀要第103冊、pp115-161など。

12) 木全敬蔵(1987):「条里地割の計測と解析」(木村芳一ほか編『奈良県史4　条里制』、名著出版、pp.101-121)

13) 認知地理学における認知地図のゆがみの分析手法などを参考にした。若林芳樹(1999):『認知地図の空間分析』、地人書房。中谷友樹(1998):「空間のゆがみの統計分析－2次元回帰分析の実際－」、立命館文学553、pp287-310。

14) 1974年の奈良市柏木町の発掘調査で検出された下ツ道から、下ツ道の方向は座標北より17分46秒振れることが報告されている。(奈良国立文化財研究所編(1974):『平城京朱雀大路発掘調査報告書』、奈良市)

15) 落合重信(1967):『条里制』、吉川弘文館。木村芳一ほか編(1987):『奈良県史4　条里制』、名著出版。金田章裕(1985):『条里と村落の歴史地理学研究』大明堂。中井一夫(1982):「地域研究－奈良県における発掘調査から」『条里制の諸問題』Iなど。

16) 岩本次郎(1987)「大和国条里制の地域的諸様相」(木村芳一ほか編『奈良県史4　条里制』、名著出版、pp69-97)

17) 古代の郡界推定については、出田(2011):前掲書で報告されている。

付記

本研究は奈良女子大学21世紀COEプログラム(革新的な学術分野)「古代日本形成の特質解明の研究教育拠点」の成果をもととし、科学研究費補助金(研究活動スタート支援)「条里地域のHistorical GISによる景観復原の試み」(代表:宮崎良美)の成果の一部を利用した。

第18回公開シンポジウム
「人文科学とデータベース」
2012.12.22

考古学データ検索における異種データベースの統一的な利用について

A unified use of heterogeneous databases in archeological data retrieval

王　鑫，宝珍　輝尚，野宮　浩揮
Xin Wang, Teruhisa Hochin, Hiroki Nomiya
京都工芸繊維大学　情報工学専攻，京都市左京区松ヶ崎御所海道町
Kyoto Institute of Technology, Goshokaido-cho, Matsugasaki, Kyoto

あらまし：考古学データの検索の際に，異なるデータベースシステムを使用している利用者が，使用しているデータベースを意識せずに同じサービスを受けられるようなデータベースシステムの実現について述べる．ここでは，クライアント側で JDBC を使い，いくつかのデータベース（MySQL，SQLite，PostgreSQL，Excel 等）に接続する手法について述べる．

Summary：This paper describes the realization of a database system enabling users, who are using different database systems, to receive the same service without considering the type of database during the search of archaeological data. Here we will describe a method to connect to several database systems including MySQL, SQLite, PostgreSQL, and Excel by using JDBC on the client side.

キーワード：考古学，データベース，JDBC，接続
Keywords：archeology, databases, JDBC, connect

1. はじめに

近年，コンピュータ技術の発展に伴い，考古学に関するデータを電子媒体としてコンピュータ上で取り扱っている考古学者も少なくない．遺跡データや遺物データといった考古学データは，その特性から，よく地理情報と関連付けて保存されている[1-6]．そういったデータをコンピュータ上で管理するためには，ユーザ自身が地図データを用意し，利用する必要がある．そのため，Web 上で公開されている地図サービスを利用することで，ユーザ自身が地図データを用意する必要がないシステムも提案されている[1]．しかし，こういった従来の考古学データ検索システムを利用するためには，ユーザの所持するデータベースを，自分の手が完全に行き届かないサーバーなどに保存する必要があることが多い．

考古学者にとって，自分の所持する研究データは大変貴重なものであり，安易に手の届かないところに置きたくないのが現状である．そのため，これまで提案されてきているシステムは魅力的なものではあるが，自分の所持するデータベースをサーバー側へコピーすることは，考古学者にとってためらわれるものである．

そこで，本研究では，ユーザの所持するデータベースを，サーバー側へコピーすることなく利用可能な考古学データ検索システムの実現を目的と

して，検討を行う．このために，クライアント側で検索を行い，その検索結果のみをサーバー側に送信することで，地図上に分布の表示が可能なシステムについて検討する．本研究と従来の研究との違いは，クライアント側で，ユーザの所持する考古学データベースへの接続，検索を行っているため，サーバー側にデータベースをコピーする必要がないことである．したがって，クライアント側で，ユーザの所持する考古学データベースに対しての処理を行うような，デスクトップアプリケーションが必要となる．本研究では，それらは JDBC を用いて実現している．

本論文では，関連研究や対象とする考古学データ，本システムにおけるクライアント側プログラムとサーバー側プログラムの設計と実装，その動作例を報告する．

以降，まず 2. で関連研究について述べる．次に，3. で具体的な設計について述べ，4. でクライアント側とサーバー側の実装について述べる．そして，5. で動作例を示し，最後に 6. でまとめる．

2. 関連研究

2.1 情報共有

2.1.1 遺跡における情報共有

このデータベース（DB）ではサーベイヤーが集まった情報を基礎にし，ベトナムと国外に住んでいる信徒から適当なデータや情報を収集するため専門的なウェブが使われている[7]．

情報の正確性と信頼性のために DB は定期的なアップデートを行わなければならない．しかし，セキュリティと信徒が便利に使えるのを考慮されており，この DB は公共使用は不可能である．

この DB の情報は三種類に分けてある．一つは基本的な情報で，教会堂の名前や位置などである．二つ目は専門的な情報を含んでいる（サーベイヤーの研究等）．建築スキーマと詳細なイメージもこの種類に入っている．最後はインタビューサーベイである．

2.1.2 ラッパーとメディエータ

TSIMMIS（The Stanford-IBM Manager of Multiple Information Sources）は複数の情報を統合するためのシステムである．それは，データモデルと多数の異なるソースからの情報の結合をサポートするように設計されている一般的なクエリ言語を提供している．また，自動的に情報を統合するためのシステムを構築する際に必要な要素を生成するためのツールが用意されている．

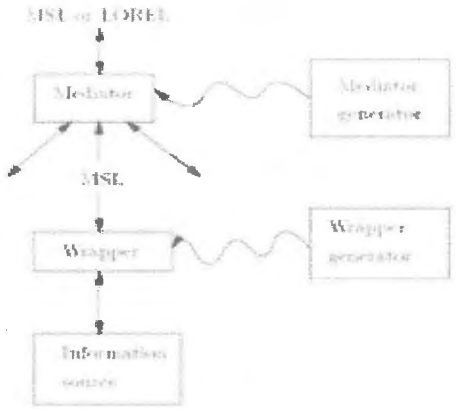


図 1　TSIMMIS の主成分

TSIMMIS の中心的な目標は，それが簡単に調停するシステムを生成することである．この目標のための一つのサポートは，共通のデータモデル言語で，もう一つは自動的にラッパーと仲介の要素を生成するためのツールである．

（1）ラッパーは，特定のクエリまたはコマンドへのアプリケーションのクエリを変換することにより，異種情報源へのアクセスを提供している．生成されたラッパーは，クエリを調べ，これとラッパーの仕様ファイルに指定されている他のパターンと比較する．

（2）メディエータは，複数の異種情報源の統合に使用されている．クエリ言語として MSL（mediator specification language）を使用することができる．MSL は 2 つの重要な機能を提供している．一つ目はこの言語はオブジェクトの再帰的な定義が可能なことである．二つ目は調停にインポートされるオブジェクトへのセマンティックオ

ブジェクト ID の割り当てが可能なことである.

2.1.3 XML 異種情報源統合における XML 構造統一化

XML は柔軟な表現能力を持ち,構造データおよび半構造データを表現することができる.この特性のために,XML を共通データモデルとする異種情報源統合に対するニーズが高まっている.

```
<xmql-results>
  <publisher-list>
   <publisher>pub-01</publisher>
   <author-list>
    <author>
     <name>Konishi</name>
     <book-list>
       <book>XML</book>
     </book-list>
    </author>
   </author-list>
  </publisher-list>
    <publisher-list>
<publisher>pub-02</publisher>
<author-list>
    <author>
<name>Hayashi</name>
     <book-list>
       <book>Web Service</book>
     </book-list>
    </author>
   </author-list>
 </publisher-list>
</xmql-results>
```

図 2　併合処理例

提案のシステム MediPresto/XM は,XML を共通データモデルとする不特定多数の異種情報源統合における構造統一化機能要件を満たす XML 構造変換を提案している.

MediPresto/XM は,不特定多数の情報源のスキーマやデータに関する動的な異種性解消,およびビューの動的な生成,取得という特徴を持つ.動的な異種性解消は,問い合わせに応じた統合対象情報源の動的選定,および各情報源のスキーマやデータの異種性の動的解消を行う.利用者からの問い合わせに応じて静的ビューから動的ビューを生成し,演算を行う.MediPresto/XM ではあらゆる情報源を XML 形式で扱い,統合する.利用者は各情報源の XML を参照せずに問い合わせを行う.MediPresto/XM は問い合わせに応じた対象情報源の探索,各情報源問い合わせの生成,そして各情報源からの実行結果の統合および構造変換処理を行う.

2.2 地理情報システムと考古学 DB

2.2.1 地理情報システムを用いた遺跡 DB

横山らは遺跡立地と地形及び，自然環境との関連を解明するためのデータベースの概要を説明し，その構築における問題点を述べている[3].

青森県，岩手県，秋田県及び宮城県を対象範囲とし，台帳データ，遺跡の輪郭図，標高図，斜度図，地形データ，水系図，集水地図，衛星データを格納している.

遺跡データベースの機能には，データベース管理システムの機能とデータ解析の機能がある．データベース管理システムの機能には，データの入力，追加，修正，データファイルの生成，削除，複製，名称変更，注記の表示，フォーマット変換など,検索と表示がある.データ解析の機能には,データの空間的な変換解析処理を行う地理データの操作や統計処理がある.

遺跡データベースの拡張として，(1)データの種類の充実，(2)対象範囲の拡張，ならびに，(3)遺跡データの更新体制の整備を考えている.

2.2.2 遺跡データと数値地図

森本は，千葉県の遺跡データベースと国土数値情報を組み合わせた試験的なプログラムを作成し

ている[4]．プログラムの目的は，(1) 現在利用されている「千葉県埋蔵文化財分布図」に準じて，遺跡を分類しながら表示させる，(2) 検索によって選択された遺跡の文献について，「収蔵図書データベース」に登録された千葉県文化財センター図書室から，一覧表示させる，(3) 野外調査用に，現地周辺の遺跡を抽出して，その位置を地図画面上に表示させるというものである．

遺跡データベースの加工と検索では，まず，調査された遺跡データベースである「既調査遺跡データベース」のテーブルと，全体的なデータベースである「遺跡台帳データベース」のテーブルを合体させることにより，既調査の遺跡とそうでない遺跡を，両者合わせて一度に検索できるようにしている．また，合体したテーブルから，遺跡の管理番号と文献番号の項目データだけからなるテーブルを新たに作成し，遺跡と所蔵図書をリンクさせるための中間的なテーブルを作成している．また，元来の情報システムには，地形図に関するデータが含まれていない．そこで，地図画像の検索と表示に利用する地形図データを新たに追加している．遺跡の検索は様々な要素に対応してできるようにしている．検索で選択できる項目は，遺跡名，所在地，市町村，地図，遺跡種類，立地，水系である．

数値地図画面の表示では，国土数値情報として，2万5千分の1の地形図の地図画像を使用している．これは TIFF 画像ファイルとして CD-ROM に収録されている．地図画像データをそのままコンピュータ画面上で表示すると，かなり限られた範囲でした表示されない．そこで，本来の地図画像データを半分に縮めた縮小画像データを別途作成し，それにも遺跡の位置を表示できるようにしている．

最後に，GPS の利用である．現地踏調査の場合，野外で周辺を検索して，地図上で確認するという利用方法も考えられる．それには，野外で現在地を測定し，その位置を中心にして周辺の遺跡を検索し，それから地図上に遺跡の位置を表示させれば良い．現在地の測定には，GPS(Global Positioning System)を利用する．

2.2.3 島根県遺跡データベースの構築と運用

島根大学地域貢献推進協議会では「島根県遺跡データベース」を構築・運用している[5]．「島根県遺跡データベース」は，考古学調査研究，埋蔵文化財行政の効率化を第一の目的とし，また，利用者は考古学研究者や考古学専攻学生，自治体文化財行政担当者といった専門家から，小中学生や一般市民まで，幅広く対象としている．

データテーブルは主に，遺跡テーブル，遺構テーブル，遺物テーブル，調査テーブル，文献テーブル，画像テーブルという基本テーブルに分割管理されている．そして，利用者に合わせて 3 種類の検索が用意されている．

初心者向けの「遺跡メニュー検索」では，市町村・時代・遺跡種別・『出雲国風土記』の記述から，項目を選び，該当する遺跡を検索する．検索条件に合致した遺跡は一覧表示される．

ある程度遺跡に興味や知識を持つ一般利用者向けの「遺跡簡易検索」では，遺跡名称・遺跡名称よみ・市町村・時代・遺跡種別・遺構種別・遺物種別などから検索する．検索条件に合致した遺跡は一覧表示される．

研究者や自治体文化財行政担当者などの専門家向けの「複合詳細検索」では，遺跡・遺構・遺物・調査・文献の諸条件を組み合わせて設定し，検索できる．検索条件に合致したデータはそれぞれ選択すれば一覧表示できる．また，一覧表示されたデータの中から一データを選択すると，それぞれのデータの詳細表示画面に移動する．詳細表示画面では，関連付けられた画像データがあればサムネイル表示されリンクが張られている．画面下部では関連付けられた他テーブルのデータにリンクが張られており，各画面に移動できる．

遺跡データは位置情報をもっており，GIS システムと連携することによって，地図表示や分布図作成ができる．遺跡は代表する場所を点として表示するに留めている．拡大，縮小，移動，表示レイヤの選択操作などが出来るほか，表示された遺跡位置のドットをクリックすると，遺跡データベースシステムに返信し，該当遺跡の詳細情報画面が表示される．レイヤは，遺跡位置，市町村境界図，大字境界図，等高線図，旧版地図，現在の地図などを座標付けし，調整したラスター画像からなる．

2.2.4 様々な集約を可能とする考古学 DB システム

宝珍は，遺跡からの遺物・遺構の分布表示を行う多様なシステムを統一的に扱うことを目的としたデータベースシステムの実現を目指している[6]．また，システムでは，分布表示における，「全体」，「検索対象」，ならびに「集約単位」を表すそれぞれのデータ実体と表示要素のデータ実体を柔軟に対応付け可能とすることにより，遺物等の分布表示システムを汎用的にするとともに，実行時に動的に「全体」，「検索対象」ならびに「集約単位」を変更可能としている．

2.2.5 goo 地図を利用した遺跡 DB システム

山内らは，遺跡から多数出土する遺物を地図情報と連携して管理する遺跡データベースシステムの設計と実装について報告している[1]．管理者にも一般の利用者にも使い易い考古学データベースシステムの実現を目的として，一乗谷朝倉氏遺跡の遺物データをもとに，遺跡から多数出土する遺物を地図情報と連携して管理する遺跡データベースシステムの設計と実装を行っている．

従来では，地図情報は管理者自らが用意していたが，Web 上で提供されている地図サービス「goo 地図」を利用することで，地図情報の管理の簡便化を図っている．

2.3 マイ・データベース・システム

及川は，研究者個人が，分析の対象として，もしくは研究の成果をまとめたものとして作られるデータベースの概念を考えている[10]．

一般に，人文系のデータベースは標準化が困難で，データベースの内容は，研究者それぞれの研究内容や成果と深く関連している．したがって，データベースの対象となる研究資料が同じであっても，データベースの内容は研究者ごとに異なったものになりがちで，それぞれの研究者のニーズに応えたデータベースが作られることが必要となる．本来，データベースは，周到な準備を経て，組織的，計画的に作成され，複数人で共有されるものであるが，マイ・データベース・システムでは，研究者個人（場合によっては，グループもあり得る）のデータベースとして，データベース内の項目の定義を変更したり，再編集したりといった作業を容易に行えるよう設計されている．また，人文系の研究者でも簡単に作れ，それを研究に活用できるようなシステムになっている．

3. 設計

3.1 システム構成

まず，図 3 に本システムの目指す動作モデルを示す．図 3 に示すように，本システムは，大きくクライアント側プログラムとサーバー側プログラムの 2 つに分けられる．

3.2 クライアント側プログラム

クライアント側プログラムでは，ユーザの所持するデータベースへ完全にローカルな環境で接続する．したがって，ユーザの所持するデータベースの接続に必要な情報を外部へ漏らすことなく，システムの利用を可能にする．接続したデータベース内のテーブルを用いて，ユーザが所持するデータを地図上にマーカーの形で表す．デフォルトのマーカー（地図の中心を確定するため）も同時に生成する．

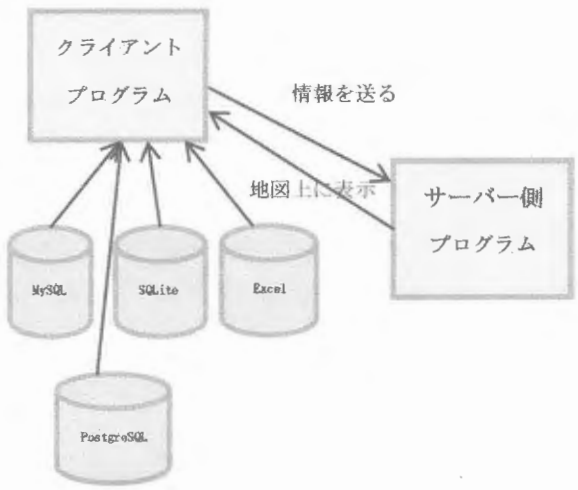


図3　本システムの動作モデル

3.3 サーバー側プログラム

サーバー側プログラムは，クライアントから送信されてきたデータをもとに，遺物の分布をGoogle Map 上に表示することが目的である．遺物の出土分布の表示に必要な情報は緯度と経度である．また，Google Map に遺物の出土分布を表示するには，Google Maps API が必要だが，Google Mapを表示するのはサーバー側なので，ユーザがそれを導入する必要はない．サーバー側でブラウザを起動し，必要な URL を渡す．現在，Google Map による遺物分布の表示は，Web ブラウザ（ユーザのコンピュータ上で，標準ブラウザとして設定されているもの）で表示されるようになっている．

3.4 情報の管理

遺物の出土分布を地図上に表示する上で必要になってくるのは位置情報である．これらの位置情報は，サーバー側にコピーする必要がなく，検索する時だけサーバー側と連結すれば良い．これも本手法の一つの利点である．

4. 実装

4. 1 クライアント側の実装

本データベースシステムではクライアント側プログラムのインターフェースは，HTML＋javascript で実現している．

4. 1. 1 データベースの操作

データベース接続や検索などのデータベースの操作は，JDBC を用いて行っている．現在は MySQL，PostgreSQL，Sqlite の三種類のデータベース管理システムに対応している．Excel の対応は検討中である．

4. 1. 2 検索結果のサーバーへの送信

検索結果のサーバーへの送信は，POST メソッドで行っている．送信される変数は，経度と緯度の検索結果数が格納された配列$_POST である．$_POST でデータを読み込み，サーバー側プログラムに送信する．

4.2 サーバー側の実装

本データベースシステムにおけるサーバー側プログラムは，データベースを扱うため，JAVA＋PHP で実現している．

クライアント側プログラムから送信されてきた変数を受信し，地図上に分布を表示する．

まず，データベースの型の識別と接続を行う．JDBC で各データベースに対応するドライバをロードし，それぞれに接続する．次に，SQL 文を利用し，データベースにデータ検索を行う．最後に，検索結果を用いて地図の初期位置の中心点の緯度と経度を設定する．そして，検索結果数とデフォルトの結果（lat1=35.09&lng1=135.07）をまとめ，Web ブラウザを起動し，必要な URL を渡し地図上にマーカーを立てていく．

5. 動作例

ここでは，現在実装済みの部分の動作例を示す．

5. 1 システムの起動

システムを起動すると，図 4 に示すようなウインドウが表示される．ウインドウ内には，タブパネルが表示されており，「データベースリスト」

のタブが用意されている．DB 接続設定のタブにて，ユーザの所持するデータベースへ接続を行う．このパネルでは，接続するデータベースの指定を行う．図 5 で示されるようなセレクトボックス形式で，データベースの管理方法を選択する．選択できる項目として，「MySQL」，「SQLite」，「PostgreSQL」，「Excel」があるが，現在 Excel には対応していない．

次に，図 6 に示すように，データベース接続に必要な情報をフォームに入力する．例えば，MySQL の場合，データベース名，データベースに含まれたテーブル名，経緯度に対応するカラム名である．データベース型は図 5 で選択するとき自動的に入力されるので，ここでは入力不要である．フォームに入力された情報を用いて，データベースへ接続する．

接続に成功すると，図 7 に示すようなブラウザが開き，ユーザ所持のデータが Google map 上に表示される．

図 4 システム初期ウインドウ

図 5 データベースの選択

図 6 接続に必要の情報の入力

図 7　結果の表示

6．まとめ

クライアント側でユーザの所持するデータベースに接続することで，それをサーバー側にコピーすることなく，考古学データを地図上に分布表示可能にすることを目的として検討を行った．

システムの基本部分の実装は行ったが，実際にユーザに利用してもらえるほど十分な機能は実装されていない．経緯度しか検索できない点である．今後の課題として，まだ実装していない機能の実装，既に実装されている機能の改良などが挙げられる．また，実装したシステムを実際に考古学研究者に使用してもらい，性能や使い勝手の評価を行うことも今後の課題である．

参考文献

[1] 山内洋裕，宝珍輝尚：地図 API を用いた遺跡データベースシステムの設計と実装，情報処理学会研究報告 2007-CH-74(11)，vol.2007，no.49，pp.81-88（2007）．

[2] Teruhisa Hochin, Fumiaki Kobayashi, Hiroki Nomiya : Seamless Usage of User's Databases in Archaeological Database System, Proceedings of CIPA2009, PS1-13 (2009).

[3] 横山隆三，千葉史：地理情報システムを用いた遺跡データベース構築，情報考古学，Vol.3，No.2，pp.29-39 (1997)

[4] 森本和男：遺跡データと数値地図，日本情報考古学会第9回大会予稿集，pp.13-22（2000）

[5] 会下和宏：「島根県遺跡データベース」の構築と運用，情報考古学，Vol.12，No.2，pp.29-39 (2006)

[6] 宝珍輝尚：考古学データベースシステムにおける様々な集約の一実現法，情報考古学，Vol. 15，No.1・2，pp. 1-12 (2009)

[7] Tomohara KATANO, Yukimasa YAMADA: The prospects and possibilities of an interactive database for information sharing and rebuilding for a historical and cultural community, Proc. of CIPA, pp. 141-146 (2009)

[8] Hector Garcia-Molina, Yannis Papakonstantinou, Dallan Quass, Anand Rajaraman, Yehoshua Sagiv, Jeffrey Ullman, Vasilis Vassalos and Jennifer Widom: The TSIMMIS approach to mediation: data models and languages, Journal of Intelligent Information Systems, 8, pp. 117-132 (1997)

[9] 小西一也，鈴木源吾，堀口恭太郎，林孝志，芳西崇：異種情報源統合における XML 構造統一化手法，情報処理学会研究報告（データベースシステム），vol.2002，no.67，pp.139-144 (2002)

[10] 及川昭文：研究者のためのマイ・データベース・システムの開発，第13回公開シンポジウム「人文科学とデータベース」，pp.25-34，(2007)．

[11] SQLite：http://www.sqlite.org/

[12] Ext JS：http://www.extjs.com/

[13] MySQL：http://www.mysql.com/

[14] 壬川純：はじめて学ぶMySQL

[15]http://mountainbigroad.jp/fc5/pgsql_java.

[16] 稲葉一浩:Google Maps API 徹底活用ガイド

[17] Rasmus Lerdorf Kevin Tatroe Peter MacIntyre プログラミング PHP

貝塚データベース　－作成から利用まで－
Creation and Use of Kaizuka Database

及川 昭文
Akifumi Oikawa
総合研究大学院大学，神奈川県三浦郡葉山町
The Graduate University for Advanced Studies, Hayama, Miura, Kanagawa

あらまし：貝塚遺跡は，それが存在していた時代の社会，文化，自然環境等を復元するための有用な手がかりを与えてくれる。すなわち，貝塚遺跡の立地環境，分布状況，時代，出土遺物（とくに貝，魚骨，獣骨等）を調べることによって，さまざまな考古学情報を引き出すことができる。そのためには，まず貝塚遺跡に関するデータの一覧，すなわちデータベースの構築が不可欠となってくる。筆者らは 1970 年代に貝塚データベースの作成を開始し，現在総合研究大学院大学の附属図書館のホームページから公開しており，5,000 カ所以上の貝塚遺跡が収録されている。この貝塚データベースを作成するにあたっての課題やデータベースの利用例について報告する。

Summary: Kaizuka(Shell mound) sites give us useful information for figuring society, culture and natural environmet in the times when they existed. That is, by examining natural environment of each site, sites distribution, and remains(especially shell, animal bones and so on), we can get the various archaeological information. In order to do so, it is essential to create the database about Kaizuka. We started to work in late 1970s and Kaizuka database is now opened through the homepage of Sokendai's(the Graduate University for Advanced Studies) library. More than 5,000 sites are recorded in this database. In this report some aspcets about the creation of database and some results of researches which refer Kaizuka database.

キーワード：貝塚遺跡，データベース，統計解析
Kewords: Kaizuka(Shell mound) Site, Database, Statistical Analysis

はじめに

貝塚データベース作成のそもそもの発端は，酒詰仲男が表した「日本縄文石器時代食料総説」に含まれている貝塚遺跡を，1970 年代の後半に電子化したことに始まる。この時作成したデータベースについては，「貝塚データベース－その作成と応用－」（国立民族学博物館研究報告第 5 巻 2 号，1980 年）に詳しく報告した。このデータベースは「酒詰ファイル」と呼んでいたが，収録されていた遺跡数は約 900 にすぎなかった。その後，本格的な貝塚データベース作成を開始したが，酒詰ファイルとは異なり，収録する対象を貝塚遺跡だけに限定せず，動物遺存体を出土しているすべての遺跡に広げた。したがって，貝塚データベースという名称は実態にそぐわないところもあるが，これまでの「貝塚データベース」の拡充ということで，名称はそのまま使用している。

1. 貝塚データベースの構築

ここでは，貝塚データベースの具体的な内容，構築上の諸問題について述べる。

1.1 データベースの種類と内容

考古学は遺跡を始め「モノ」をその研究対象としており，それらの「モノ」に関するあらゆる情報が研究を進めていく上で必要になってくる。発掘調査や分布調査等から生み出される分布図，写真，実測図，拓本，あるいは報告書等の文献といった多種多様な情報のデータベース化が検討されなければならない。貝塚データベースにおいても実測図や写真等についてのデータベース化が望ましいが，今回の研究においては次の 3 種類のデータベースを作成した。

遺跡・遺物　貝塚データベースといった場合，このデータベースを指し，表 1 のような項目が収録されている。現在までに蓄積されているレコード数は約 5,000

であるが，これは 1990 年頃までに発見されている遺跡数で，現在はかなりの数の遺跡増が想定される。

文　献　上記の貝塚データベースを構築するために利用した 1 次資料である調査報告書を主体とした文献の書誌情報データベースである。約 1,200 文献が収録されている。

貝属性　約 800 種類の貝について，生息域（水深，潮間帯等），生活圏（内海，河川，汽水域，地上等），生活状態（付着，着生等），分布（本州以南，瀬戸内海等）等の項目を収録した辞書ファイルである。

表 1　貝塚データベース項目一覧

遺跡番号	遺物－節足類
遺跡名	遺物－棘皮類
遺跡名（よみ）	遺物－魚類
所在地	遺物－両生類
県・市区町村コード[1)]	遺物－は虫類
時代コード	遺物－鳥類
遺構コード	遺物－哺乳類
土器編年	遺物－人骨
絶対年代	遺物－植物
文献番号 [2)]	遺物－その他 [3)]
遺物－貝類	経緯度データ [4)]

1）JIS X0401, X0402 に準拠
2）文献データベース含まれている文献の ID
3）貝製装身具等の特記すべき遺物
4）国土地理院が作成した「平成 12 年度版　日本の市区町村位置情報要覧」に基づいて電子化したデータベースと市区町村コードをキーとしてマッチングを行い，プログラムで自動的に収得したものである。要覧には各市区町村行政界の「東端，西端，南端，北端，重心」及び役所（役場）の位置情報が記載されているが，電子化したのは重心のみである。市区町村の重心位置であることから，貝塚遺跡に取り込んだ経緯度データは，正確なデータではなく，近似的な位置情報となっている。

1.2 構築上の諸問題

一般的にデータベースの構築には多大な労力と時間を必要とするが，その過程で生じる問題には，実際に携わった者でなければ理解できないものが少なくない。それらの多くは本質的な問題ではないが，データベースを利用する段階では大きな問題となることが多く，できるだけ早い時期に解決しておく必要がある。とくに，データの品質を一定に維持することは，データベース構築において，最も重要なことの一つであり，そのためにいろいろな工夫が必要になってくる。

(1) データの均質性の維持

考古学データベースを作成するための基本的な一次資料は，「発掘調査報告書」である。これらは一つの調査単位で作成されるが，その様式は統一されておらず，報告書のそれぞれで遺跡や遺物の記述は異なっているといってもよい。一冊の報告書に複数の遺跡の報告がなされていることも少なくないが，同じ報告書の中でも報告者が違えば，その表現や記述様式が異なっていることは珍しくない。このため，一次資料からデータシートを作成する際に，データ記述の統一を図ることに多大な労力と時間を費やすことになる。今回のデータベース作成でとくに問題となったことをいくつか挙げると，以下のようになる。

遺物名の表記の不統一　現時点で貝塚データベースに含まれている遺物の種類は，約 3000 種類になる。このうち種類が多いのは貝類（約 1700）魚類（約 500）鳥類（約 200）哺乳類（約 300）で，当然といえば当然であるが，貝類が他を圧倒している。これだけの数になると，人手で表記の統一を図るのは不可能であるため，チェックのためのプログラムを作成し，かなりの部分を機械的に処理できるようにした。とくに，貝類については，「原色日本貝類図鑑」「続原色日本貝類図鑑」（吉良哲明著，保育社刊）に記載されている貝の名称（約 2800）をコンピュータに入力し，この辞書ファイルとのマッチングによるチェックを行った。誤りが一次資料にあるのか，データシートへの転記，あるいは機械可読化の時点で起こったのか判断することは難しいが，一次資料のレベルですでに間違っている例は少なくない。不統一の具体的な例をいくつか挙げると，次のようなものがある。（●印に統一）

●クジラ くじら 鯨	文字種（カタカナ，仮名，漢字）
アマオフネ ●アマオブネ アマオブネガイ	・濁点のあるなし ・～カイ（ガイ）の有無
●ウラウズガイ ウラウヅガイ	「ズ」と「ヅ」
ニッポンジカ ●ニホンジカ	「ニッポン」と「ニホン」

種の同定　貝塚遺跡から発見される貝殻にしろ動物の骨にしろ，必ずしもその種が同定できるとは限らない。非常に少量であったり，腐敗や損傷がひどかったり，あるいはもともと同定しにくいものであったりして，何であるかを特定できない場合には，

ウシ OR ウマ	シジミ属
イノシシ？	タラ科種不明
ニシン・ウグイ類	ニシン類
ニシン科	ネズミ等中小動物の微細骨
ニシン科の一種	ブリ・タイ類似の魚

等と報告書に記載されている。正確を期すためには，遺物そのものをもう一度調べ直すということが必要になるが，現実には不可能である。現在は例に挙げたようなまま入力されているが，できるだけもとの情報を生かし，かつ統一的に表記できるような方法を検討する必要がある。

(2) 存在しない所在地

遺跡の所在地は基本的に一次資料である報告書等に記載されているとおりに入力されているが，市町村合併や町名変更等の理由で，現在は存在しない所在地となっている例が少なくない。また，現在は正しくても，将来同様の理由で不正な所在情報となることもある。この問題を解決する方法としては，次の 2 つが考えられる。

①市町村合併や町名変更の情報を確実に収集し，それをデータベースに反映させる。

②物理的な位置情報を緯度・経度の数値データとして，あるいは遺跡の位置を示した地図をイメージデータとしてデータベースに取り込む。

①は費用，時間，人手という観点からみれば，あまりに非現実的である。②の物理的な位置情報は，むしろ積極的にデータベースに取り込むべき情報で，それが入力されることにより，遺跡分布図の自動作成や，あるいは空間分析といった研究に大いに資することは明らかである。

(3) 読めない遺跡名

一般的に遺跡の名称は，その遺跡のあるところの地名，あるいはそれに類した名称（例えば，○○塚とか△△洞と昔から呼ばれていたというような場合）等からつけられることが多い。したがって，難読の地名があるように遺跡名にも難読なものが少なくない。また，一般的な読み方でなく，特殊な読み方をするような場合もあり，データベースとしては遺跡名の読みは必須項目ということができる。

ところが，実際には一次資料である報告書のどこを探しても，遺跡名の読みが見つからないものが少なくない。最近は遺跡名にふりがなを振った報告書が多くなりつつあるが，データベース化を前提にした報告書の様式の標準化ということを検討する時機が到来しているのではないだろうか。

貝塚データベースにも「**鉞切遺跡，海戸貝塚，砲原貝塚，土穴瀬貝塚，大鼠蔵貝塚，定留貝塚**」のように遺跡名の読めないものが数十あり，これらは現地の人に問い合わせるしか方法がなく，これもデータベース構築に時間がかかる理由の一つである。

2. 研究支援ツールとしての検索システム

研究者自身でデータベースを作成しようと思ったら，DBMS を初歩から学習したり，講習を受講する等，それなりの時間が必要になってくる。半日程度の学習でデータベースが作れるようになり，かつ研究支援ツールを有する DB システム SOARS を開発した。[1]

2.1 SOARS が目指したもの

SOARS とは Sokendai Academic Resouces System の略で㈱富士通の Shunsaku をもとに開発したデータベース・システムである。開発においてもっとも重視した目標は，「人文系の研究者でも SE やプログラマ等の情報技術者の支援なしで，データベースの作成から Web での公開までできる」，「単なる検索ツールとしてではなく，データベースを活用できる研究支援ツールとして機能する」の 2 つである。

今日 Excel 等のソフトを利用して，資料を整理したり，分析したりしている人文系の研究者は少なくない。Web 上で簡単にデータベースが定義でき，Excel のデータをそのままアップロードし，それがデータベースとして構築されるようになれば，データベース・システムを知らない研究者でも，自分の資料をデータベースとして広く公開することが可能になる。詳細な説明は避けるが，SOARS はこれを可能にしたシステムである。

図 1 SOARS の構成

2.2 研究支援ツール

本来データベースは蓄積されたデータを分析したり，人事システム，給与システムあるいは大学における学務システム等の業務を効率よく運用するために開

発，発展してきたものである。ところが情報化時代の到来とともに大量の情報の中から必要なものを探し出す，いわゆる検索機能が重要視されるようになった。その結果検索やそれに関連した機能や性能は，ハードウェアや OS の発展とともに格段に充実してきているが，データベースに格納されているデータを分析したり，活用するための機能は不十分なままの状況にある。SOARS は，検索機能のみでなくデータベースの分析やそれらを活用するための機能を充実させることを大きな目標として開発を行った。

文献目録等の類のデータベースはいわば参照情報を提供するためのデータベースであるが，現在 SOARS で公開されている貝塚データベース等は，データベースそのものが分析対象となるものである。もちろん貝塚データベースを参照目的で利用することもあるが，検索結果を対象としてさまざまに活用することがデータベース構築の本来の目的である。SOARS では，データベースそのものを分析する機能や活用を容易にするためのダウンロード機能の充実を図った。

(1) 統計処理機能

統計処理機能としては，「基本統計」と「頻度集計」の 2 つの処理ができるようになっている。

基本統計：項目定義でデータ種別が「数値」と定義された項目を対象に，「最低値，最高値，平均値，標準偏差値」を計算する。

頻度集計：項目定義で「頻度集計」対象とした項目について，その項目に入力されている「テキスト」ごとの頻度集計を行う機能である。この機能はデリミッタ項目を対象とした場合，とくに効果的である。2 つの項目を組み合わせたクロス集計も可能となっている。

(2) ダウンロード機能の充実

SOARS では以下のようにさまざまな画面でダウンロード機能を利用できるようになっている。

①利用者管理
②データベース定義
③項目定義
④詳細表示
⑤検索結果一覧
⑥ KWIC リスト
⑦統計処理

一般の利用者が利用できるのは⑤～⑦である。検索されたレコードのダウンロード機能はほとんどの DBMS に標準的に装備されているが，レコード・データ以外のデータをダウンロードするには，複雑な操作や利用者側でマクロを組む等の作業が必要であり，人文系の研究者にそれを望むのは，多くの場合困難である。

検索結果，たとえば図 2 のようなリスト（CH 関連文献データベースを「及川昭文」で検索）を論文等に一覧表というような形で引用する場合，これまでは「カットアンドペースト」といった面倒な作業を必要とすることが多かったが，SOARS ではこの作業を大幅に省力化することができる。

まず，得られた検索結果一覧を希望する順序に並べ替える。図 2 の場合，ヘッダーである「ID，タイトル，著者，発行年」のいずれかをクリックすれば，その項目をキーとしてソートされた一覧が表示される。次にそれを CSV ファイルとしてダウンロードし，ワープロに取り込み，必要であれば編集を行い望むようなリストを作成する。

図 9 のような基本統計の結果や図 10 のような頻度集計の結果も，同じように CSV ファイルとしてダウンロードできるようになっており，Excel 等の表計算ソフトを使って，より高度な統計的分析等を行うことも可能になる。

得られたデータをどのように活用するかは利用者側の課題であるが，ダウンロードできる場面を数多く設定することで，データベースの活用の機会が拡がったことは間違いないであろう。

ID	タイトル	著者	発行年
1 10006	国際会議にみる人文科学分野へのコンピュータ応...	小沢一雅（大阪電通大），及川昭...	1989
2 10007	人文科学におけるコンピュータ利用の現状と課題	及川昭文（国教研）	1989
3 10008	遺跡データベースと映像化	及川昭文（国教研）	1990
4 10087	日本語教育支援システム	大深悦子（日本IBM）坂谷内勝（...	1991
5 10080	講演会における情報伝達度についての一考察	及川昭文（茨城大）	1992
6 10096	日本語教育・学習支援システムの機能構成とその...	高木清（ノス），吉岡亮衛（国教...	1992

図 2　検索結果一覧

3. 数量的分析

数量的分析を行うためには「数値」が必要である。したがって，多くの考古学資料はそのままでは数量的分析の対象とはならない。大きさ等が計測された土器や石器は，その計測値をもとにさまざまな分析が可能であるが，土器の文様や形，前方後円墳の形態等，もともと数値で表現できないものを分析するためには，何らかの数量化が必要になってくる。[2]

3.1 数量化の方法

表 2 は貝類，魚類，哺乳類のうちで，100 ヵ所以上の遺跡から出土している遺存体の一覧である。この一覧に，それぞれの項目で「有」というのがあるが，これは種の同定ができなかったことを意味している。この種の記載は報告書によく見られることであるが，同定できないほど少量あるいは細片だったのか，専門家がいなくて同定できなかったのか，いずれか判断できないのは問題である。つまり，後者の場合であれば，専門家に依頼して同定することは可能であることが多いからである。

貝類についてもう少し詳しく考察してみる。貝塚データベースに収録されている遺跡から発見される貝類は，全部で約 2,200 種類になるが，その実態は表 3 のようになる。この表から分かることは，1 つの遺跡からしか発見されない貝類が 987 種（全体の約 44.8%）あり，全体の 80%の貝は 10 ヵ所以下の遺跡からしか出土していないことが分かる。2,200 種の貝の大部分は，日本全国の数ヵ所からしか発見されてい

表 2　100 以上の遺跡から出土している貝類，魚類，哺乳類

貝類

1.	ハマグリ	1931	23.	イボウミニナ	336	44.	コシダカガンガラ	179
2.	アサリ	1436	24.	カワニナ	312	46.	タマキビ	176
3.	アカニシ	1194	25.	イボキサゴ	305	46.	イタヤガイ	176
4.	サルボウ	1040	26.	キサゴ	302	48.	オキアサリ	175
5.	オキシジミ	1020	27.	イガイ	296	49.	カワアイ	174
6.	シオフキ	1015	28.	アワビ	293	50.	ホソウミニナ	173
7.	カキ	1008	29.	シジミ	283	51.	ウチムラサキ	171
8.	マガキ	978	30.	ヘナタリ	251	52.	オオヘビガイ	170
9.	ヤマトシジミ	920	31.	イシダタミ	244	53.	オオタニシ	169
10.	ハイガイ	900	32.	アラムシロ	233	54.	カリガネエガイ	160
11.	ツメタガイ	808	33.	ウバガイ	227	55.	フトヘナタリ	157
12.	ウミニナ	793	34.	チョウセンハマグリ	223	56.	イシガイ	142
13.	オオノガイ	730	35.	ナミマガシワ	220	57.	テングニシ	138
14.	カガミガイ	697	36.	クボガイ	216	57.	ダンベイキサゴ	138
15.	イボニシ	503	37.	バカガイ	213	59.	コタマガイ	132
16.	サザエ	423	38.	ベンケイガイ	211	60.	ヒダリマキマイマイ	121
17.	バイ	408	39.	有	207	61.	キセルガイ	117
18.	スガイ	398	40.	ウネナシトマヤガイ	195	62.	ムラサキガイ	111
19.	イタボガキ	384	41.	ホタテガイ	192	62.	アマオブネ	111
20.	レイシ	372	42.	ミルクイ	189	64.	ヒメエゾボラ	108
21.	アカガイ	369	43.	マシジミ	186	65.	ナガニシ	103
22.	マテガイ	360	44.	ニホンシジミ	179			

魚類

1.	スズキ	546	9.	フグ	202	17.	コイ	144
2.	クロダイ	450	10.	有	182	18.	サバ	135
3.	マダイ	419	11.	ヒラメ	172	19.	カツオ	134
4.	サメ	307	12.	コチ	161	20.	ウグイ	133
5.	ボラ	257	13.	サケ	158	21.	ウナギ	126
6.	エイ	252	14.	ブリ	156	22.	ニシン	123
7.	マグロ	245	15.	カレイ	151			
7.	タイ	245	16.	カサゴ	145			

哺乳類

1.	イノシシ	1217	8.	ウシ	265	15.	ニホンザル	121
2.	シカ	1119	9.	イルカ	229	16.	ウサギ	114
3.	イヌ	627	10.	ノウサギ	200	17.	エゾシカ	113
4.	ウマ	532	11.	有	197	18.	テン	107
5.	ニホンジカ	453	12.	アナグマ	187	19.	アシカ	100
6.	タヌキ	398	13.	ネズミ	157			
7.	クジラ	360	14.	キツネ	134			

ないということは，少数出土例の貝をそのまま分析の対象とするのは適切でないことを意味している。しかし，これらの貝を除外することは，ほとんどの貝のデータを捨てることになるので，何らかの方法でこれらの情報を利用できるようにする必要がある。

表 3　出土頻度別貝の種類数

出土遺跡数	貝の種類	%	累積%
1	987	44.8	*44.8*
2	259	11.7	*56.5*
3	148	6.7	*63.2*
4–10	358	16.2	*79.4*
11–20	170	7.7	*87.2*
21–50	145	6.5	*93.8*
51–100	72	3.2	*97.9*
101–	65	2.9	*100.0*

林は，その「数量化の方法」の中で「**現象を仕分けしてうまく段階を分けて括ってかかることである，括られたものはそれを一つのものとして大きな枠の内で処理し，括られた内部のものはその内部においてあらためて解析を考えてゆくという仕方である。**」[3]と述べているが，少数の遺跡からしか発見されない貝について，何らかの方法でより大きなグループにまとめていくことができれば，分析の対象とすることが可能となってくる。

大きくまとめていく方法として試みたのが，貝の属性でまとめるということである。具体的には，貝の生息域，たとえば暖かい海，冷たい海，浅いところ，深いところに生息しているといった基準で，貝をグルーピングすることを目論んだ。そのために，まず「貝属性データベース」を作成し，その内容を分析した。

最初に行ったのは，生息域の表記例をすべてコンピュータに入力し，それを 50 音順にソートして生息域リストを作成した。データベースのための 1 次資料として利用した「日本近海産貝類図鑑」の編者が「本書の使い方と用語」で「**分布の表記は情報の不均一や種の特性のため全巻必ずしも統一されていないが了とされたい。**」と述べているとおりに，実に統一がとれていない。たとえば，「伊豆七島鳥島沖」「伊豆諸島鳥島」「伊豆諸島鳥島沖」「伊豆鳥島近海」は，ほぼ同じ地域を指すと思われるが，表記が異なっている。「伊豆七島」と「伊豆諸島」，あるいは「伊豆諸島」「伊豆諸島沖」「伊豆諸島海域」は同じなのか，違うのか，判断することはほとんど不可能である。

当初は日本の周りの海をいくつかのゾーンに分け，それぞれの貝の生息域はどのゾーン（ひとつあるいは複数）に対応するか設定することを試みたが，あまりに生息域の表記が不統一であるため，そのようなゾーンの設定そのものができないことが分かった。そこで次善の策として，それぞれの貝の生息域の南限，北限のみを設定することにした。

たとえば，「北海道以南」「北海道～九州」「北海道東部以南」は，すべて「**北海道以南**」とし，「北海道南部～サハリン」「北海道以北」「北海道西部～ベーリング海」などは，すべて「**北海道以北**」とした。「貝属性データベース」に生息域を設定した後，貝塚データベースとのマッチングを行って，生息域毎の貝種などをまとめたものが表 4 である。

表 4　生息域毎の貝種・出土貝種・遺跡数
（各地域より以南，すなわち地域が北限となる）

地域	辞書貝種	出土貝種	遺跡数
北海道	289	114	2043
三　陸	320	57	1441
房　総	974	228	1268
東京湾	29	8	144
相模湾	506	19	100
伊　豆	332	57	67
駿河湾	91	5	18
遠州灘	96	6	825
熊野灘	9	1	2
紀伊半島	958	126	145
四　国	253	11	47
瀬戸内海	110	5	16
九　州	289	16	696
奄　美	666	87	141
沖　縄	344	22	48
小笠原	109	2	33
男鹿半島	79	5	534
佐渡島	79	0	0
能登半島	190	2	6
隠　岐	8	0	0
山口県	182	4	7

表 2 には，100 以上の遺跡から出土している貝 65 種を挙げたが，これらの大部分は生息域が「北海道以南」となり，全国に生息していることから，その出土例が多くなることはうなずける。一方，生息域が限定されているかかわらず，多くの遺跡から出土している貝もある。具体的には，サルボウ，フトヘナタリ，シオフキ，ハイガイ，オキシジミ等の暖かいところに生息している貝，ウバガイ，ホタテガイ，ヒメエゾボラ等の寒いところ生息している貝である。とくに「サルボ

ウ」の生息域は「東京湾～有明海」の範囲にもかかわらず，仙台湾から三陸沖，また青森の方でも出土している。一方「ホタテガイ」は，その生息域は「東北～オホーツク」となっているが，茨城県，東京湾，渥美半島，北九州まで分布している。これらの貝について，出土遺跡数と，それぞれの貝が生息域以外で見つかった遺跡数を時期別に集計したものが表 5 である。

たとえば，「オキシジミ」の場合，生息域は房総半島以南が設定されているので，房総半島（緯度は 35 度 42 分 52 秒とした）より北に出土した遺跡を生息域以外の遺跡数として数えた。この表では草創期に房総半島以北で見つかった遺跡は 89 で，全体で 135 になっている。

この表から分かることは，「アマオブネ～ヘナタリ」のグループ A（暖かいところに生息している貝）は北限を越えて多く発見されているの比べて，「ウバガイ～ホタテガイ」のグループ B（寒いところに生息している貝）は，その大部分が生息域で発見されているということである。とくに越境組の最たるものは「ハイガイ」で，全体の 73%が生息域外で見つかっている。何故，このようなことが起きたのかについては，いろいろな説が展開できるが，もっとも有力なのはいわゆる縄文海進といわれる地球温暖化との関連であろう。

図 3 は，各グループの合計の生息域外の出土比率を時期別に表したものであるが，グループ A とグループ B ではその傾向が逆になっている。これは次のように説明できる。

表 5　生息域外で見つかった遺跡数

貝名(以北)	緯度[注]	不明	草創期	早　期	前　期	中　期	後　期	晩　期	合　計
アマオブネ	35.42.52	0/ 5	1/ 6	2/ 10	3/ 10	8/ 19	5/ 42	1/ 17	8/ 58
イタボガキ	35.42.52	2/ 6	39/ 59	63/131	63/ 93	87/148	87/160	14/ 45	155/295
オキアサリ	35.42.52	0/ 0	15/ 28	37/ 74	26/ 43	33/ 85	41/105	13/ 30	69/164
オキシジミ	35.42.52	11/ 18	89/135	181/323	167/248	204/368	254/461	31/104	480/849
カワアイ	35.42.52	3/ 4	22/ 29	30/ 51	19/ 35	37/ 71	46/ 81	9/ 21	76/137
サルボウ	35.40.12	11/ 15	106/146	223/346	178/247	247/362	295/450	60/109	551/832
シオフキ	39.00.00	0/ 16	5/133	14/317	19/212	17/341	5/422	2/102	35/786
ダンベイキサゴ	38.00.00	0/ 2	1/ 23	1/ 35	1/ 19	2/ 42	2/ 47	0/ 16	2/ 93
チョウセンハマグリ	35.42.52	0/ 0	23/ 32	37/ 62	33/ 51	59/ 96	50/101	13/ 34	93/163
テングニシ	35.42.52	1/ 3	10/ 18	10/ 43	8/ 25	14/ 50	12/ 58	4/ 18	22/ 98
ハイガイ	35.08.19	17/ 28	77/ 96	221/290	181/251	203/279	276/371	44/ 87	528/722
フトヘナタリ	35.40.12	0/ 2	18/ 28	12/ 42	10/ 35	19/ 54	21/ 79	3/ 22	37/119
ヘナタリ	35.42.52	2/ 5	24/ 44	25/ 75	23/ 59	29/ 89	33/104	5/ 34	64/193
グループ A		47/104	430/777	856/1799	731/1328	959/2004	1127/2481	199/639	2120/4509
貝名(以南)		不明	草創期	早　期	前　期	中　期	後　期	晩　期	合　計
ウバガイ	35.42.52	0/ 0	3/ 31	8/ 62	5/ 64	6/ 71	10/ 70	6/ 28	11/155
ヒメエゾボラ	38.00.00	0/ 0	1/ 7	1/ 18	0/ 30	0/ 35	1/ 31	1/ 10	2/ 62
ホタテガイ	35.42.52	0/ 0	5/ 17	8/ 40	9/ 59	12/ 54	20/ 53	7/ 13	26/110
グループ B		0/ 0	9/ 55	17/120	14/153	18/160	31/154	14/ 51	39/327

注）範囲の境界を示す緯度。35.40.12 は 35 度 40 分 12 秒のこと。

グループAは，大部分の生息域の北限が房総沖，あるいはそれより南となっているが，温暖化が進む間その境界線は，だいたい仙台湾から三陸沖まで北上していたと考えられる。逆に，グループBは温暖化から寒冷化が進むにつれ，その境界線は南下してきたと判断できる。その結果，それぞれの北限，南限を超えた分布が増減してきた状況が観察されることになる。

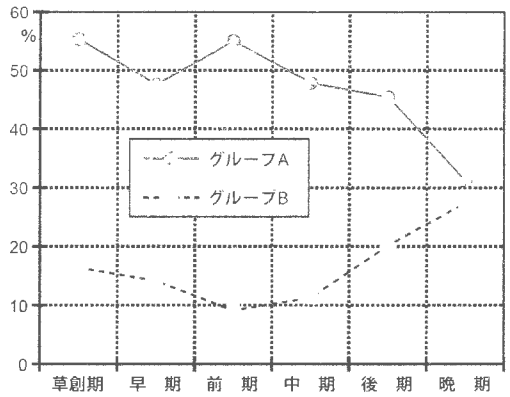


図3　各グループの生息域外出土の時期別割合

おわりに

これまでの考古学研究においては，一つひとつの事実や資料を丹念に収集し，それらを詳細に観察，解釈し，仮説を組み立てていく，いわゆるボトムアップの手法が主流であった。しかしながら，この手法では何千，何万といった量のデータを対象とした分析は困難で，個人のレベルで集められる，あるいは分析できる範囲にとどまらざるを得ないということがある。また，発掘調査件数が増え，自然科学的分析等の資料の増加も相まって，集めるべきデータが飛躍的に増加するとともに，一つひとつの資料を精査することが困難にもなってきている。そのような状況の下では，個人研究は土器の文様のほんの少しの違いを論じるといった，よりミクロな方向に向かわざるを得ないところがある。そのような一種閉塞的な状況にある考古学研究を変えるものとして，考古学的資料の数量化を提案しているわけであるが，その利点としては以下のようなことがある。

データの共有化：データベース化の目的のひとつにデータの共有化ということがあるが，それを実現するためには，共有のための基準をつくる，いいかえればこれまで個々の研究者がそれぞれに解釈していた考古学的資料に対して，共通の認識を持つということが必要になってくる。その共通認識を深めていくための有効な手段として数量化がある。数量化を行うためには，何らかの基準を設けなければならず，結果的に数量化された数値というフィルターを通じて共通の認識が確立されていくことになる。そして，そのことによって考古学的資料の共有化が促進されることになる。

主観から客観へ　ー定性的データの分析ー：土器の文様や形状を数量化するということは，必然的に数量化のための基準を策定し，それを示すことが不可欠になる。そのことは，それぞれの研究者が提示した「数値」に対する研究者同士の相互批判が可能となることを意味する。この相互批判が同じ土俵の上でできるということは，「科学的」であることの重要な要件であり，どちらかといえばこれまでの考古学に欠けていた部分である。そして，これまで主観的なデータとしてしか取り扱えなかった定性的データも，より客観的なデータとして分析できるようになる。

大量データの分析：一つひとつの資料としてしか取り扱えなかったものが，これまでとは比較にならない大量データの処理も可能となってくる。また，貝塚データベースのように，出土例が少数であっても，大きなまとまりとして，さまざまな方法で数量化することによって，これまで分析をあきらめていたデータも分析可能となる。

このように多くの利点があるが，数量化を行うことの最大の意義は，考古学的資料の数学的モデルに基づいた分析，すなわちトップダウン的な研究手法が可能となることである。これまでの考古学は，2000年に起こった石器捏造事件に代表されるように，一つひとつの事象を重大視するあまり，それらの事象に振り回される傾向が少なからずある。このことが仮説検定型の研究の進展を阻害している一要因であるが，数量化，そしてそれに基づいた数理的手法による研究は，このような状況を打破する大きなきっかけとなる可能性を秘めている。

[1]及川昭文，藤沢桜子，洪政国，山元啓史「研究支援機能を強化したデータベース・システムの開発」人文科学とコンピュータシンポジウム 2007 論文集，pp.229-236，2007

[2]及川昭文，山元啓史「大規模考古学データの分析ー貝塚データベースを例にしてー」『日本情報考古学会第15回大会発表要旨，pp.77-84，2003

[3]林知己夫「数量化の方法」東洋経済新報社，1974

石造遺物銘文取得のためのアーカイビング手法の開発
Development of Archiving Technique for images of text on Stone Monuments

上椙 英之† 上椙 真之†† 多仁 照廣†††
Hideyuki Uesugi† Masayuki Uesugi†† Teruhiro Tani†††
†神戸学院大学 人文学部 神戸市西区伊川谷町有瀬 518
††宇宙航空研究開発機構 相模原市中央区由野台 3-1-1
†††敦賀短期大学 地域総合科学科，敦賀市木崎 78-2-1
†Kobegakuin University, 518 Arise Ikawadani, Nishiku, Kobe
††Japan Aerospace Exploration Agency, 3-1-1 Yoshinodai, Sagamihara
†††Tsuruga Junior College, 78-2-1 Kisaki, Tsuruga

あらまし：本論文では，石造遺物の表面に刻まれた銘文の画像解析による取得を目的とし，簡便で普遍性を持った撮影方法と，石造遺物の状況に応じた解析手法を柔軟に登録・検索・適用する解析型のデジタルアーカイブについて述べる．

Summary: In this paper, we developed a new data archive system for stone monuments, which includes image processing tools for pictures of the characters on the monuments. We can store and search the algorisms of the image analysis as well as the images and data of the samples to the system, and can semi-automatically apply the algorisms of the image processing which applied to the previous samples those have similar characteristics of the new sample.

キーワード：石造遺物，データベース，画像処理，金石文
Keywords: stone monuments, database, image processing, epigraphs

1. はじめに

近年，デジタルアーカイブ構築の隆盛の中，石塔や石碑，水鉢といった石造遺物のデジタルアーカイブは，歴史的価値が高いものなど一部のものを除いては殆ど構築事例が無い．その一因として，石造遺物からの文字情報取得の困難さが挙げられる．従来，拓本によって文字を取得し，その拓本をデジタル化することで，石造遺物のデジタル化は進められてきた(「金石文拓本資料データベース」東京大学史料編纂所[1]「日本金石拓本コレクションデータベース」早稲田大学[2]，など)．しかし拓本は一件のデータ取得に時間がかかる上，ある程度の経験を重ねる必要がある．また，石造遺物は歴史的・宗教的な観点から，汚損が許されず非接触・非汚損での観察しか許されないために適用できないことも多い．近年では，デジタルカメラや 3D スキャナーによる，3次元形状の復元が可能となり，文化財のデジタル化に大きな役割を果たしている．しかし，これらのデバイスを用いた石造遺物のアーカイビングには未だ課題は多い．先ず 3D スキャナーはフィールドワーク中に使用可能なハンドタイプのものが開発されている．しかし，高額なため，複数導入することは難しく，フィールドに無数に存在する石造遺物のデータ取得には不向きと言える．次に比較的安価に手に入るデジタルカメラを使用する場合も課題は多い．石造遺物のほとんどは屋外に在り，長期にわたって風化作用を受けているため，

1.表面がコケで覆われている
2.風化のため通常撮影では文字判読が困難
3.試料前面に十分な撮影距離がとれず，統一されたデータ取得方法が採れない

などの理由により，通常撮影だけでは判読出来ない場合が多い(図1)．

加えて，拓本・デジタルカメラの 3D スキャナーいずれの場合も，文字情報取得のためにはデータ取得後，さらに煩雑な画像処理が必要となる．その画像処理にしても，石造遺物は，円柱・角柱・自然石などの形状の差や，風化の度合い・材質など多くの要因によって，採るべき手法は変わってくる．しかし，現在稼動しているデジタルアーカイブは博物学的な展示システムの延長であり，画像解析に使用できるアーカイブシステムについては未整備である．

図1 石造遺物の残存状況例.
決まった手法によって，このような多彩な画像を分析するためには，データフォーマット，およびイメージ獲得のための撮影方法を標準化する必要がある.

従来の博物的なデジタルアーカイブでは，基本的に同一の視点からは 1 枚の画像しか取得・提示されず，またその必要もなかった．しかし，画像処理という機能をアーカイブシステムに導入する場合，1枚の画像からの情報量が変化しない以上，どれほど優れた画像処理法でも限界がある．そこで，連続的なデータを規格に従って取得することで，決まった手法で効率的に画像処理をすることが可能となり，一枚の画像からは得ることの出来なかった情報を容易に引き出すことができるようになる．例として月探査衛星かぐやでは，同じ地点の「光の波長」の違う画像データを用いて，これまでの月形成モデルでは形成されないと考えられていた Purest anorthosite と呼ばれる岩石を月表面上に発見するなど[3]，既に広く用いられている技法である.

本研究ではこの技術を応用し，図 2 のように上・下・左・右・正面の五つの光源を用い，試料に陰影コントラストを作り，同一の視点からそれぞれの光源の画像を撮影した．この複数の斜光画像を新しい軸とした試料の画像群を取得する．この手法は，複数の照明条件で画像データを取得し対象となる物体の 3 次元形状を推定する Photometric Stereo[4]に代表される手法であり，Reflectance Transformation Imaging[5]で既に文化財へ適用実績がある．また，shape from shading[6][7]を活用すれば，陰影から三次元形状を復元することも可能となる．これらの解析手法は研究の蓄積も多く，大いに参考にすべき技術である．しかし，どのような優れた解析手法であっても，解析が行われた場合，解析の手順やパラメータは解析担当者の経験の中、アプリケーションの中，マニュアルに転記される事もあり得る．再度，過去の解析で使用した手順やパラメータを利用する場合，ある程度の労力を必要とする．しかし，これらの手法は意図する光源以外を排除しなければならないが，石造遺物は，その環境が多様で，安定した撮影環境を用意することは不可能である．また，試料の移動も不可の場合が多く，どのような撮影環境であっても，望む画像が取得できる撮影方法を開発しなければならない．さらに，試料表面は色情報や乱反射も多く，それらが画像の解析を妨げる要因となっていることも併せ，既存の画像処理の手法では各地に点在する石造遺物の文字情報を効率的に抽出することは難しい.

本研究では，このような石造遺物のアーカイビングに山積する問題を解決するため，(1)一枚で石造表面の文字の判読可能な画像を提示することを目的とする，(2)画像処理の為，複数光源を軸とした画像群を取得する，(3)システムに Sips の他にも Image Magick や OpenCV といったライブラリを設置し，過去の画像処理の例を新しいデータにも適用することで，画像の閲覧/抽出だけでなく，画像処理のルーチンを検索/再帰的に実行するアーカイブシステムの構築を含めたトータルなソリューションの開発を目指す.

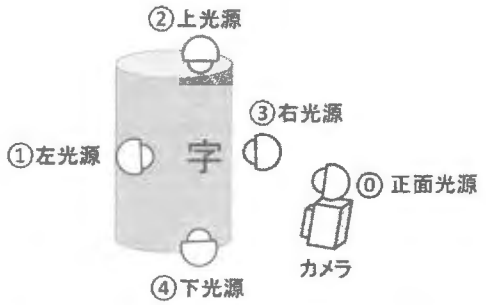


図2 撮影状況略図.
上下左右正面の五方向の光源下で撮影する．撮影時にカメラが動くと画像処理時に位置合わせが必要となるため，リモコンを使用して撮影中はカメラを動かさない.

2. ワークフロー

文字情報取得までのワークフローは以下の通りである.

(1) 4光源+正面光源の5つの光源毎に撮影
(2) 解析サーバに画像を登録
(3) メタデータ登録
(4) 画像の切り抜き
(5) 光源毎の画像平均を取得
(6) グレースケール化し5枚の画像を取得
(7) 各画像の輝度幅を揃える
(8) 正面光源から斜光の差分を取得
(9) 最も明るいピクセルで合成

(1)が撮影，(2)～(3)がデータ登録，(4)～(8)が全試料に共通した画像処理となり，画像データが登録されると，自動で(8)まで実行する．(9)以降がモジュールであり，試料毎に内容が変化する．最も明るいピクセルで合成する手法は，極めて単純な手法であり，斜光によって文字の影がはっきりと認められる場合に効果がある.

3. 撮影機材・手法

撮影機材は，デジタルカメラは NIKON 社の D5000（約 1,230 万画素）を使用した．画像解析の際に位置合わせの問題が起こらないよう，フォーカス等は全て手動に設定し，ノイズ低減と光による白とびを防ぐために ISO 感度は低めの 200 に設定した．光源には，高輝度 LED を306個配置した TEEDA LED L306P を使用した．この光源は単三電池でも十分な光量を確保できるため，屋外の電源確保のための大型バッテリーを持ち歩く必要がなく，フィールドワークでの使用に適している．

撮影時は，以下の点に注意する．

- 撮影枚数は光源間で一致させる
- 撮影時，カメラが動くと位置合わせが必要となるため，リモコンを使用して撮影中はカメラには手を触れない
- 地面が柔らかいと周囲を踏みしめただけで位置がずれることがあるため，極力カメラにも近づかない
- 光源を写り込ませない
- 一つの光源を撮影し始めたら，決めた枚数を撮影し終えるまで，光源を他の角度に動かさない
- 試料に直接日光が当たっている場合，陰影の正確な取得が困難になるため，日除けとなる傘を用いて日除けとする

以上の点を踏まえ，それぞれの光源位置からの写真を4枚の計20枚の画像データを取得した．本撮影手法では，光源を手持ちとし，カメラ操作はリモコンで行うため，図3の様に一人で実行可能である．

図3 学部生に撮影を依頼．

4. システム構成

本研究ではデータ取得後の画像解析をより効果的に進めるためのデジタルアーカイブシステムを開発した．システムの構成は以下の通りである．

- プログラム言語：bash + perl
- データベースサーバ：Apache + CGI
- データベースクライアント：Web ブラウ
- OS：MacOSX

MacOSX であれば UNIX 環境，Apache，perl が標準装備されており，サーバ設定が比較的容易である．また，OS に高性能な画像変換プログラム“sips”が標準搭載されており，このプログラムを利用することで，高価な画像処理ソフトを購入する必要が無い．また，Shell Script を使用し，Sips の他にも Image Magick や OpenCV といったライブラリが使用可能である．

本研究のデータベースシステムは，本来地球惑星科学分野において，個人運用を目的とし，データベースの設置から運用までのすべての過程を計算機システムに習熟していない研究者にも設置・運用が可能となることを目的として開発されたものである[8]．設置方法は指定されたディレクトリにファイルを設置し，サーバを公開するだけである．そのため，Mac で Web 共有設定が可能であれば，設置は数回のクリックだけで完了する．

5. データ登録

撮影した画像群は FTP でサーバにアップロードする．サーバに画像を登録する際，ファイル構成を記述した text ファイルを併せて登録する．text ファイルは全て一行一文字で，0から9までの数字を記述する（図 4）．

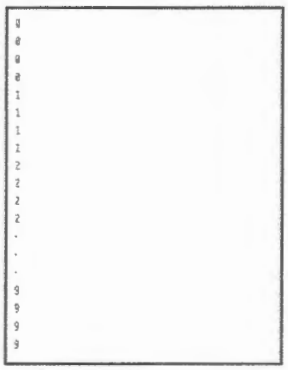


図 4 ファイル構成テキスト．数字一つが一つのファイルに対応し，各数字は光源方向を表している．5〜9 は右上・右下・左下・左上の光源表す．基本的には 4 方向で十分な効果が得られる．サーバでは画像ファイルを昇順にソートするので，登録前にソートした上で，画像の光源方向を確認し数字を記述する．同じ数字のファイルは全て平均化される．

次いでメタデータを登録する．本システムはこの登録されたメタデータを利用して，近似した条件の過去の試料の画像解析手順を検索し，その過去の試料に適用された手法を，新規の登録画像にも適用する．過去の解析実績が無い場合，若しくはシステムが適用したモジュールによる結果が思わしくない場合は，新規に「使用画像処理用ライブラリ」「画像処理パラメータ」「手順」を検討・開発し，その一連の手順を登録する．そのため，本発表のメタデータは画像処理に影響を与えると想定されるものを優先して定めている．[1]

[1] 現時点では，書誌的な情報をメタデータとして記述し，公開することを目的としていないため，ダブリンコアに沿うメタデータとはなっていない．データが増え，情報提供の質を考える必要が出て来た時，改めてダブリンコアに沿うメタデータを検討していく．

表1 メタデータ

メタデータ	説明
ID	試料 ID
address	現所在地住所
history	移設前旧所在地住所
place	所在地概況(寺社名・道路・河川など)
build	造立年月日
transport	移設年月日
longitude	緯度
latitude	経度
altitude	高度(海抜)
direction	文字面の方向
height	高さ
width	幅
depth	奥行き
shape	石像物形状(角柱, 円柱, 自然石など)
material	素材
owner	試料の造立主. 施主
surface	表面形状(凹曲面・凸曲面など)
surfaceID	複数文字面がある場合の識別番号
condition	文字の視認度を以下の 5 段階で評価 1. 通常撮影画像でも十分判読可能 2. 現地判読は容易だが撮影では困難 3. 現地判読が不可能ではないが困難 4. 字形は認識できるが判読不可 5. 字形が残っていない
text	銘文内容
names	銘文中の人名リスト
object	石像物俗称(石灯籠, 鳥居, 水鉢など)
images	総画像枚数
direct	正面光源の有無
dark	斜光源の数
set	画像処理のセット番号
comment	コメント

6. 画像処理

6.1. 共通処理

画像処理には全ての試料に適用できるものがある. それを「共通処理」として, 画像が登録された際に自動で全ての画像に適用される.

(4)trim…石造遺物以外が映り込まない様, トリミングを行う. (6)で輝度を揃える際, 背景迄映り込んでいると, 輝度が上手く揃えられない可能性があるため.

(5)average…光源毎の複数枚画像の average 平均をとることで, 虫や塵などの映り込みや自然光の揺らぎの影響を低減する(図 5).

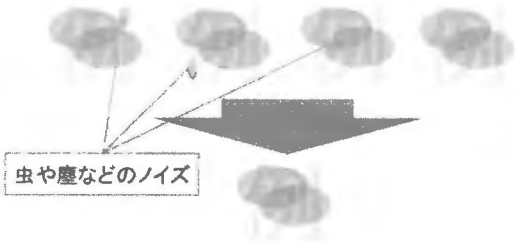


図5 average(平均)による効果.

(6)gray…本研究において重要なのは濃淡であり, 色情報は必要無いため, グレースケール化し5枚の画像を取得する.

(7)brightness…各画像輝度を揃える(図 6)

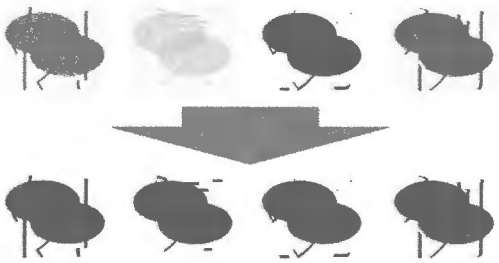


図6 輝度の調整.

(8)subtract…正面光源画像から斜光画像の差分を取得(図 7)

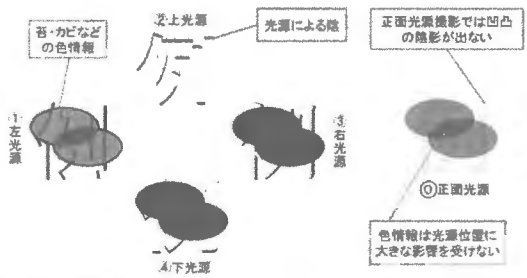


図 7 差分画像の取得.
正面光源では凹凸の陰影が写らない. 一方, 斜光画像には, 凹凸の陰影が写る. また, 苔やカビなどの色情報は斜光による影響を大きく受けないため, 差分を取得すると, 色情報が除かれ, 陰影が強調された画像が取得できる.

6.2.個別処理

個別処理は，試料毎の特徴（メタデータ）によって手順やパラメータが異なる．今回は差分をとった画像のMax(最も明るい部分)で画像合成する，単純な合成方法で画像処理を行う．この処理を行うと図 8 に見られる様に，上下左右の光源による陰影が合成され，文字が浮かび上がる．

図 8 差分画像の合成．差分を取る事で取得された陰影が合成される．

7. データベース画面

図9はデータベース画面の例である．試料(1)は，神戸市垂水区高丸一丁目の瑞丘八幡神社の石塔で，処理前のオリジナルの画像が保存されている“org”ディレクトリが表示されている．ディレクトリ内の画像は，「ディレクトリ内画像」にサムネイル表示され，そこから選択された画像が「選択画像」に表示される．この画像からオリジナル画像では，文字の判読が出来ないことが明らかである．「ディレクトリリスト」には“org”ディレクトリの他に，共通処理後の画像が保存される“src”ディレクトリ，個別処理後の画像が表示される“result”ディレクトリが表示され，表示ディレクトリの切り替えが可能である．

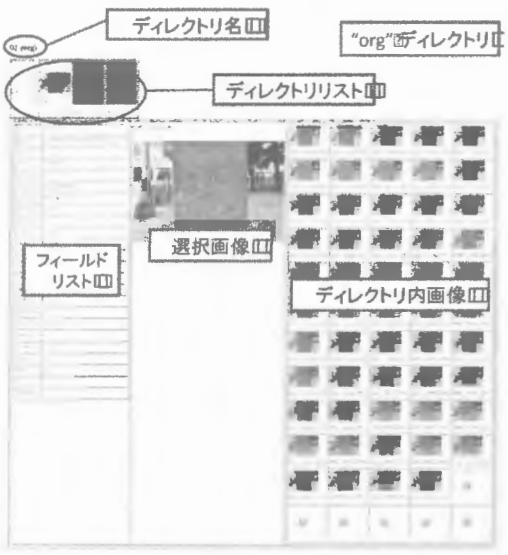


図9 データベース画面例・試料(1)．
“org”ディレクトリは撮影後の画像がアップされるディレクトリに当たる．フィールドリストには登録したメタデータが表示され，このメタデータでソート・検索も可能である．

画像処理の過程で共通処理後の画像を“src”ディレクトリ（図 10）に保存する．画像はトリミングされ，平均化によってノイズが低減され，差分取得によって表面の不要な色情報が除かれている．この例の様に，共通処理後の時点で文字が判読出来そうな試料もある．

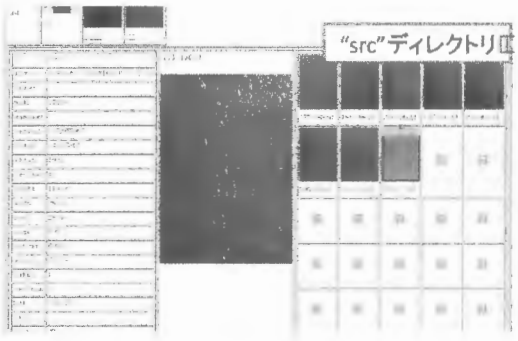


図 10 “src”ディレクトリ．
共通処理を終え，石造遺物表面の不要な色情報が除かれ，斜光による陰影が強調された画像が光源毎に保存されている．

“result”ディレクトリ（図 11）には，個別処理後の画像が保存される．ディレクトリ内画像には，個別処理後の画像が表示されている．この中から，文字が判読できる画像を選択・登録する事で，今後は類似したメタデータを持つ画像に同じ処理方法・パラメータが優先して適用される．

図12は，左からaverage合成・max合成・min合成（最暗部の合成）の結果である．本試料の場合，中央のmax 合成が有効な画像処理方法であると判断される．

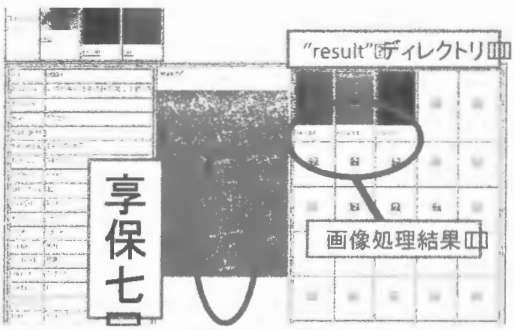


図 11 “result”ディレクトリ．試料(1)の画像処理結果の画像．

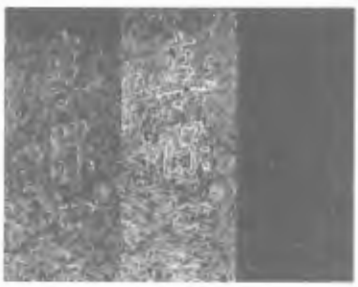


図 12 「享保七」が判読可能である．

8. 事例

成功事例

試料(2) 神戸市垂水区 瑞丘八幡 狛犬台座

図13はオリジナル画像であり，図14は図12と同じくaverageとmaxとminの結果である．以下の通り，オリジナル画像での文字の判読は難しいが，図12同様max合成ならば，「嘉永七甲寅十一月吉日」と，最もよく判読できる．一方でmin合成は全く判読できない．

図13 瑞丘八幡 狛犬台座 オリジナル画像

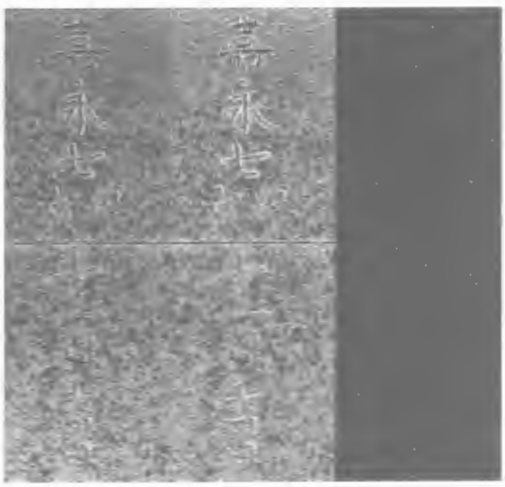


図14 瑞丘八幡 狛犬台座 画像処理後

試料(1)と(2)は両者とも，メタデータ“condition”で言えば，2の「現地判読は容易だが撮影では困難」に相当する，風化状態である．

試料(3) 伊勢市 北山墓地 宝篋印塔

図15は，左がオリジナル，右がmax合成による結果である．この試料は，表面の色情報(カビ)が文字の判読を妨げている．また砂岩製のため風化が早く，表面が摩耗するように文字情報が失われつつある．文字の凹みはかろうじて残っているが，現地での判読には慣れが必要なため“condition”は3に相当する．

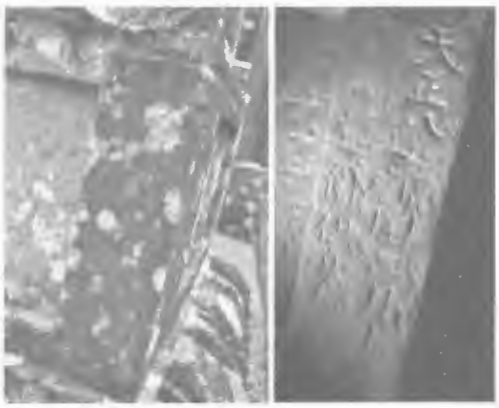


図15 伊勢市 宝篋印塔

失敗事例

試料(4) 神戸市西区 厳島神社 狛犬台座

図16は狛犬台座部分である．max合成処理後の画像をトリミング前の位置に重ねているが，文字面の大きな湾曲が，光のムラを発生させ，処理結果にその影響が見られる．

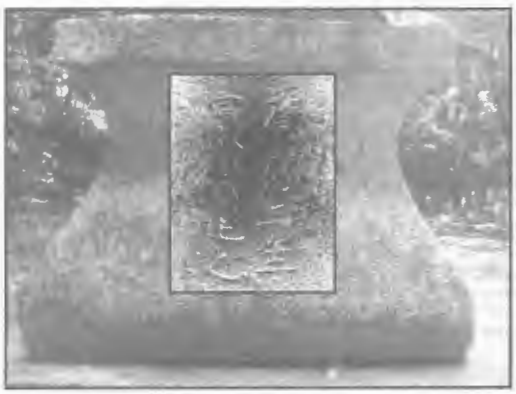


図16 厳島神社 狛犬台座

試料(5) 神戸市垂水区 瑞丘八幡 石鳥居

図17は円柱形の石鳥居である．試料(1)と年代・風化共に同程度だが，形状が異なるため，上手く処理出来ていない．

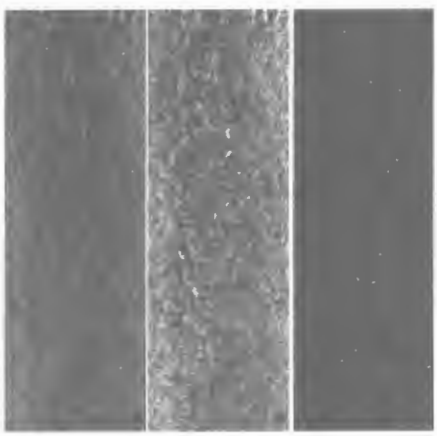


図17 瑞丘八幡 石鳥居

試料(6) 敦賀市　三嶋八幡　石灯籠

図18は円柱の石灯籠である. この事例の場合, 既に字形が失われており, 判読は不可能である.

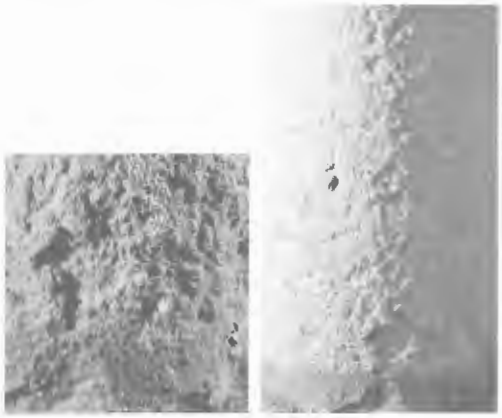


図18　三嶋八幡　石灯籠

要検討事例

試料(7) 伊勢市　北山墓地　宝篋印塔

試料(7)は, 試料(3)と同じ宝篋印塔であり, "condition"は 3 である. また, 試料(3)と同じくカビ由来の色情報が文字の判読を大きく妨げている(図19). (3)との違いは, 銘文の記述面の広さと, 銘文の刻まれた面が凹形になっているため, 斜光を当て難いという点が挙げられる.

結果, 図20に見られるように大体の文字は判読できるが, 文字面全体を囲む壁の為, 斜光が端の文字に当たらないことが原因で画像右下や左下に黒く文字の判読できない箇所が存在している.

また, 本誌料は文字の記述面が広い為, トリミングを殆どせず, 広い範囲を, 光源を平行移動させながら取得した. その為, 図21のように光のムラが発生し, 図20の上部と下部の明るさの違いなど, 均一な処理結果が得られ難くなっている.

図19　伊勢市　宝篋印塔　オリジナル画像

図20　伊勢市　宝篋印塔　max 合成処理画像

図21　伊勢市　宝篋印塔　差分画像

試料(8) 尼崎市　生島神社　石灯籠

図22は, 御影石製の石灯籠であり, "condition"は3である. この試料は風化による凹みの深い部分と, 刻まれた文字の深さが同じ 2mm であるため, 斜光によって陰影を写そうとすると, 風化の陰影も同じ程度に強調される. そのため, 処理後画像の文字が判読し難くなっている.

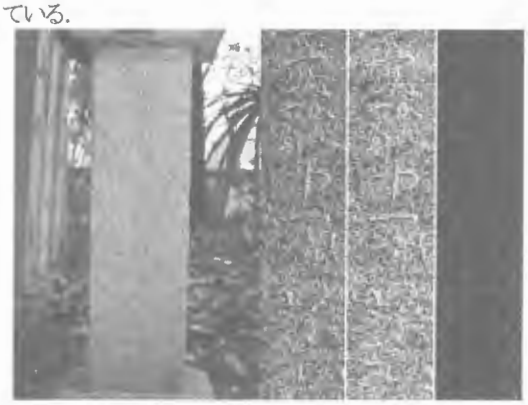


図22　生島神社石灯籠

9. まとめと課題

本論文では, 石造遺物の文字情報取得法として, 光源の位置を変えて撮影し, 斜光源画像を新しい軸とした画像群の撮影方法を提案し, その画像群を利用した解析型デジタルアーカイブの発表を行った.

その結果，通常の撮影では文字の判読が不可能なほど風化した石造遺物表面の銘文を，ある程度，判読可能な画像として処理することが出来た．

しかし，円柱など，形状が複雑になれば max 合成だけでは，うまく処理できない事例も多く，また，文字情報とは関係の無い凹凸まで陰影を残すので，それがノイズとして画像に残る事で，判読が妨げられる等，課題は多い．

先ず，形状の問題は，Reflectance Transformation Imaging や，shape from shading など画像処理分野の手法を，個別処理の手法に取り入れる事で，形状に応じた解析手法を提示することが期待出来る．次に，解析結果画像をヒストグラム解析し，閾値を決め 2 値化する．さらに 2 値化した画像からノイズを除くために，拡張(dilation)・収縮(erosion)の順に処理を行う，クロージング(closing)と，逆に収縮(erosion)・拡張(dilation)の順に処理を行うオープニング(opening)を行った．結果，図 23 の様に判読できた試料(1)は，より鮮明になり，判読が難しかった試料(8)は，「文政」という文字が浮かび上がる程には復元できる．

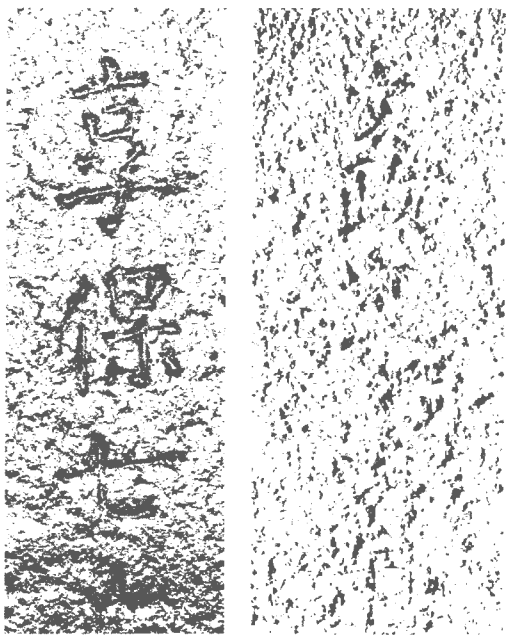


図 23 試料(1)と試料(8)の 2 値化画像

今後は，これらの手法を個別処理に取り入れ，文字の認識精度向上を図る．また，解析手順の検索/適用機能を検証・活用する為には多くの試料画像が必要となる．今後は多くの研究者に登録・解析の検証を行って貰える様，システム公開を前提に，セキュリティやシステムの安定度を高めていきたい．

10. 謝辞

本研究は科研費 23650122 の助成を受けたものである．

11. 文献

[1] 金石文拓本史料データベース
http://wwwap.hi.u-tokyo.ac.jp/ships/_shipscontroller

[2] 日本金石拓本コレクションデータベース
http://www.enpaku.waseda.ac.jp/db/kinseki_takuhon/

[3]Ohtake et al.: The global distribution of pure anorthosite on the Moon, Nature, 461, pp.236-240 (2009).

[4]R. J. Woodham.: Photometric Method for Determing SurfaceOrientation fromMultiple Image, Optical Engineering, Vol. 19, pp.139-144(1980).

[5]Reflectance Transformation Imaging
http://culturalheritageimaging.org/_Technologies/RTI/

[6]Ramachandran, V. S. Perceiving shape from shading, Scientific American, August, pp.76-83(1988).

[7]Ramachandran, V. S. Perception of shape from shading. Nature. 331, pp.163-166(1988).

[8] 上椙真之，上杉健太朗:X 線 CT の 3 次元データのためのデータベース開発，日本惑星科学連合大会予稿集.Disk2, MGI015-06 (2010).

[9] 上椙英之，上椙真之，多仁照廣：石造遺物デジタルアーカイブ構築のための撮影手法の開発，情報処理学会研究報告，Vol.2012-EIP-55 No.11, pp.99-100 (2012).

[10] 上椙英之，上椙真之，多仁照廣：石造遺物銘文取得のためのデータベース開発，人文科学とコンピュータ シンポジウム論文集，No.7, pp.179-184(2012).

文化遺産情報資源共有化のためのリレーショナルデータベース構築
-小豆島岩谷地区石切丁場における実践例-
Relational database for sharing information on cultural heritagede
-Practices in quarries Iwagatani Shodoshima-

高田 祐一
Yuichi TAKATA
同志社大学 文化遺産情報科学研究センター, 京田辺市多々羅都谷 1-3
Research Center for Knowledge Science in Cultural Heritage ,Doshisha University, 1-3
Tataramiyakodani, Kyotanabe, Kyoto

あらまし:近世城郭普請に関連する調査研究は、近年活発化しつつあるが、研究者や一般の関心ある人々が情報を利活用するためのデータベースは現在存在していない。そして城郭普請の研究では、石材の生産地である石切丁場から消費地である城郭石垣を一連の工程として捉える必要があり、調査研究情報をデータベース化する際には、有機的に結合する機能が求められる。徳川大坂城の普請では、西日本各地の大名が各地で採石しており、本発表のモデルを展開すれば、各地の情報を大坂城をキーに結合することができ、有益な情報基盤として成立する。

Summary:Recent, research related to the construction castle of early modern times, has been activated.But, There is no database that can be used by researchers and the general public.The study of castle construction, it is necessary to consider a series of movements, a stone wall is a place of consumption from a quarry is a place of production.Lords of western Japan, were collected stones in various areas.You can expand the model of this presentation, you will be able to combine information from various places. The database will be established as a foundation of useful information.

キーワード: リレーショナルデータベース, 石垣,石切丁場
Keywords:Relational database, Stone wall, quarries

1. はじめに

近世初期、江戸幕府は公儀御普請として大坂城や江戸城の普請をおこなった。城郭普請で特に労力が多い工程が石垣普請である。石垣の構築は、石材を調達・運搬し、石積みをおこなう。これらの作業には遠方の大名を動員し、遠方の石切丁場から石材を調達する場合があった。当然、研究には広域な視野が必要となる。しかし、石垣普請に関する調査研究情報を蓄積し、共有・分析するためのデータベースは現在存在していない。

そこで本稿では、石垣普請に関する調査研究情報のデータベースを構築し、その有用性を検討する。実践例として、元和寛永期におこなわれた大坂城(現在は大阪城だが、本稿では歴史用語としての大坂城を用いる)を対象とし、石材調達地である小豆島岩谷地区の石切丁場を採り上げたい。

2. 近世の公儀御普請

江戸幕府は、慶長・元和・寛永年間に日本中の大名を動員し、各地で城郭普請をおこなった。幕府直轄城では公儀御普請として動員した。幕府が主導した城郭として江戸城、二条城、大坂城、名古屋城、篠山城などがある(1)。江戸城では慶長8年(1603)から工事が始まり、寛永まで断続的に工事が行われた(2)。主に外堀掘削などの掘方は東国大名、石垣・枡形構築の石垣方は西国大名が担当した。寛永13年(1636)の普請では、掘方として東国52大名、石垣方として西国61大名動員している。

また各大名は、所領を近世都市化する過程で、居城を整備もしくは築城をおこなった。例えば慶長6年

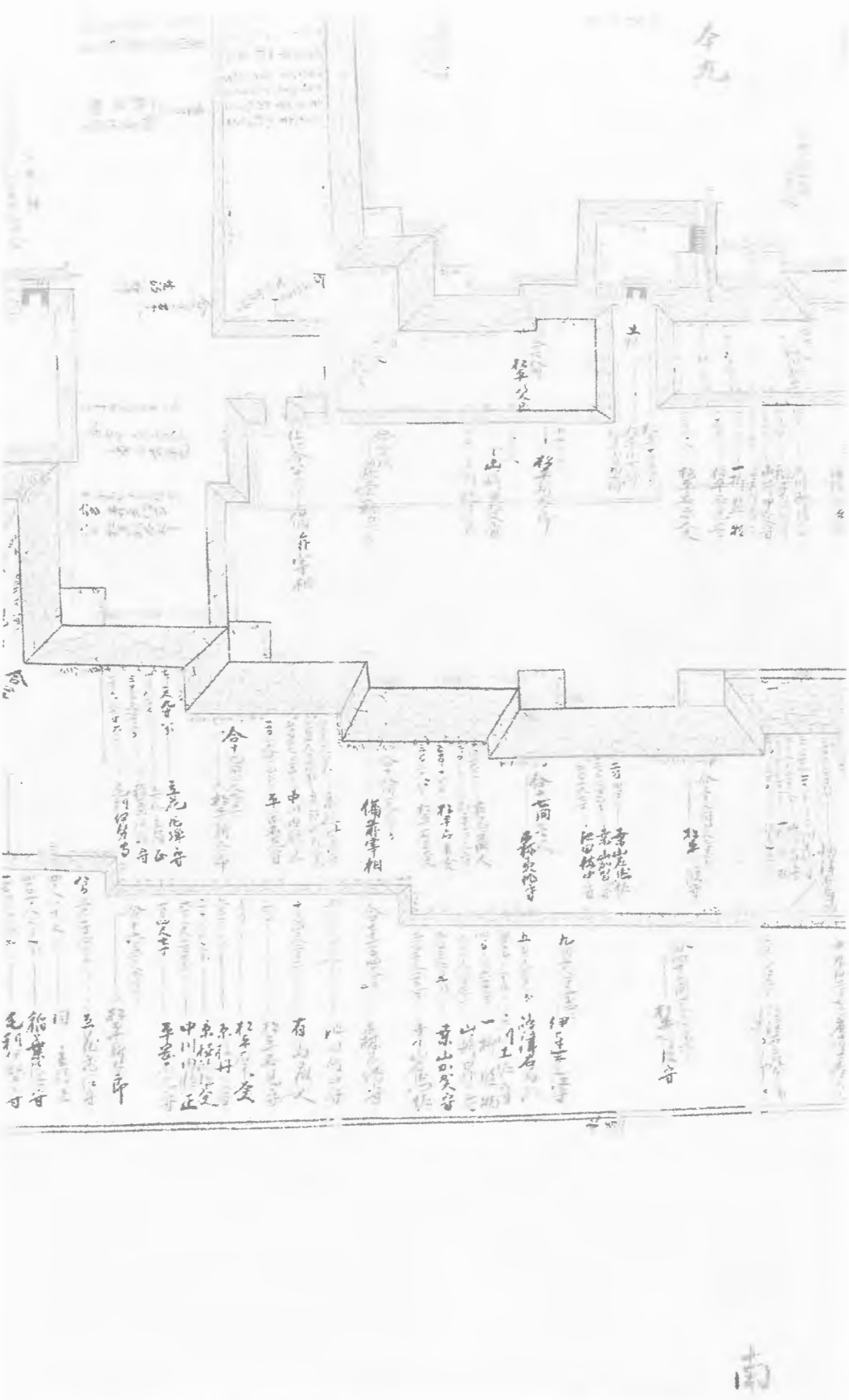


図1「大坂城普請丁場割之図」(大阪府立中之島図書館蔵)

(1601)の福岡藩黒田家による福岡城築城、慶長7年(1602)の佐賀藩鍋島家による佐賀城築城などあり、まさに日本中が築城ラッシュの状態であった。

3. 大坂城普請と石切丁場

慶長20年(1615)大坂夏の陣によって大坂城は落城し豊臣家は滅亡した。その後、江戸幕府は大坂城再築を図り、石垣・堀の構築に西国・北国の大名60数家を動員した。工事は主に元和6年(1620)の第一期工事、寛永2年(1625)の第二期工事、寛永5年(1628)の第三期工事に分けて進行した。大名は、割普請によって担当する石垣を割り当てられた。大名の担当石垣は「大坂城普請丁場割之図」によって把握することができる(図1)。この図によって大坂城で担当した石垣や担当量が判明する。

そして石垣の構築部材である石材は自らの責任で確保した。現在確認されている石切丁場としては、京都府木津川市加茂、大阪府生駒山系、兵庫県東六甲山系、岡山県前島、香川県小豆島(写真1)、広島県尾道、山口県大津島、福岡県行橋市沓尾、佐賀県唐津市谷口などがあり、瀬戸内海沿岸で広範囲に確認されている(図2)。北垣聰一郎氏はこれらを「花崗岩ベルト地帯」と仮称している(3)。石材調達地がこれだけ広域に求められたことは、良質な石材を膨大な数を必要としたためだろう。また日本の築城技術のピークとされる大坂城で高石垣が成立しえた要因として、「石材の規格化」(4)とそれを実現可能にする「採石技術の平準化」(5)は必要不可欠な要素であった。

4. 石垣石の刻印

昭和34年(1959)、大坂城総合学術調査が実施された。考古学・歴史学・地質学などの学識者によって「大

写真1 小豆島岩谷 豆腐石石切丁場

写真2 大坂城刻印調査の様子

注(6)報告書

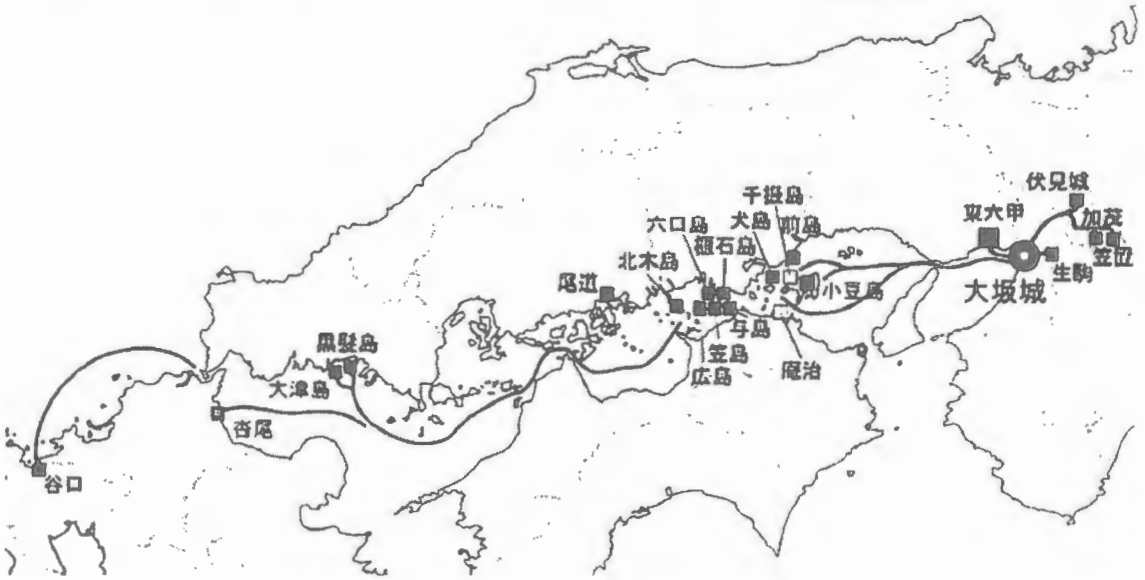


図2 大坂城と石切丁場

注(3)文献

坂城総合学術調査団」が結成された。最大の成果は、豊臣大坂城は現在の徳川大坂城の地中にあり石垣はすべて江戸幕府による再築であることが判明したことである。調査の一環で石垣の刻印が入念に調べられた(写真 2)。現存の大坂城石垣に壁番号が割り当てられ壁ごとに刻印種が集計されている(図 3)。

刻印とは石に様々な目的で印を彫ったものであり、大名の家紋、石工の作業用の印、石切丁場での傍示など多様な機能があると言われている。大坂城総合学術調査では刻印の種類が基本 200 種内外、派生形を含めると 1247 種類確認されている(7)。刻印は石材調達のあり方や担当壁の他藩との境を判断するために有効であり、北野博司氏は丁寧に石垣構造と刻印を読み解くことによって大名家中組の存在を実証してい

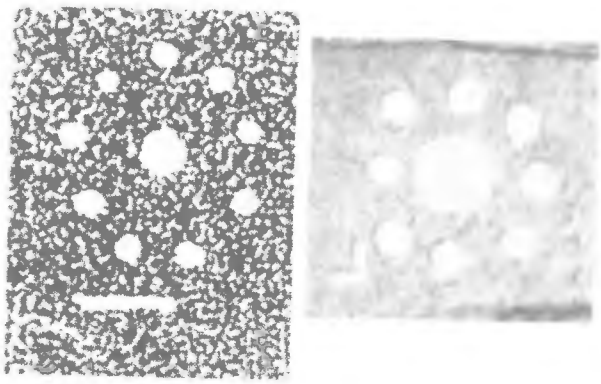


図4 兵庫県芦屋市で見つかった十曜紋(左)と大坂城南外堀の九曜紋(右)

注(11)文献

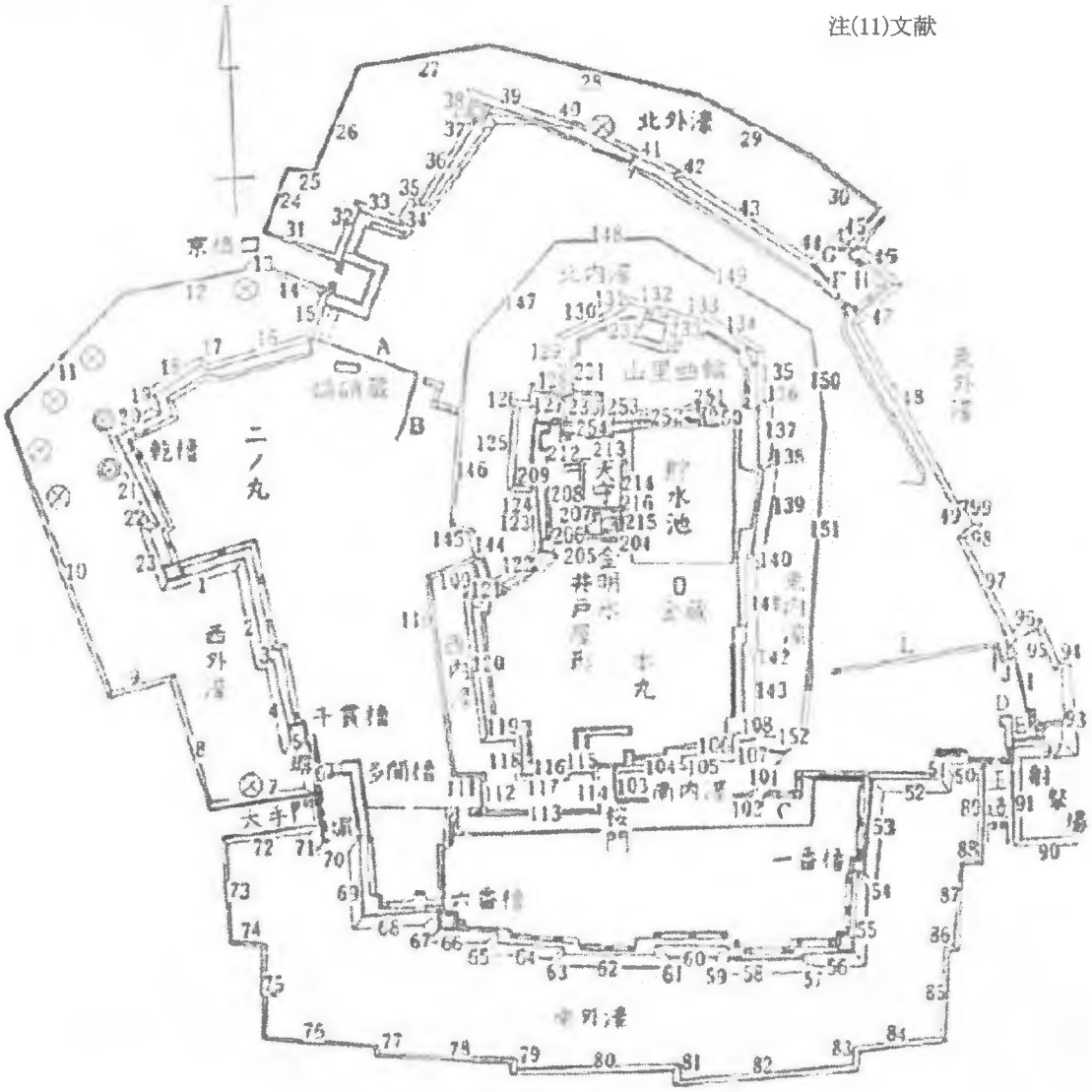


図3 大坂城石垣の壁番号

注(7)文献

る(8)。また石切丁場内の領域のグルーピングに刻印を用いることは有効であり、石切丁場と石垣をつなぐ非常に重要なキーでもある。兵庫県東六甲山系の甲山刻印群の検証では文献、石切丁場の状態、大坂城石垣の状態を総合的に検討することで佐賀藩鍋島家の刻印であると実証した(9)。

刻印の扱いには注意を要する。例えば、従来では九曜紋(図4)の使用大名は豊前小倉藩細川家であると理解されてきた(10)が、多賀左門氏は慎重論を打ち出しており、そもそも兵庫県芦屋市で見つかった紋は十曜紋であるとしている。刻印の解釈によって石切丁場の理解は大きく変わる。このように、石切丁場で確認された刻印と石垣を安易に結びつけることは危険で、文献等を用い複合的に検証する必要があり、刻印とその使用大名を再度検証する時期にきていると考えられる。また刻印は全国の近世城郭・石切丁場で確認されており、それらの刻印は同一もしくは類似している場合がある。当然、採石した集団の関連性、集団内の労働編成、時期、技術の系譜、石材の流通など分析しうる情報資源として有用だと考えられるが、現時点では情報を統合した例はない。統合できれば全国の城郭石垣や石切丁場について時空を超えて有機的に結合でき有用な情報基盤となる可能性がある。定量的分析など情報科学を応用することも可能となろう。

5. 石垣普請データベースの要件・設計

そこで刻印をキーに情報を結合し、分析できる情報基盤を整備するために石垣普請データベースを構築することにした。まずは代表的なモデルケースとして大坂城と石切丁場として唯一の国指定史跡である小豆島を取り上げデータベース化したい。データベースとしてはリレーショナルデータベースを採用した。理由は主キーによって様々なデータと結合することができ、今後の拡張性に優れているためである。また枯れた技術であるため容易に構築でき扱える技術者も多い。そしてSQLによって柔軟に分析ができるため採用した。

今回のデータベースの要件としては以下の通りである。各地の石切丁場と城郭石垣の刻印種類と個数に関する情報を蓄積し、各種の分析を可能とすること。石切丁場の分布調査では、石材1点ずつ把握しており、石材毎の刻印種と位置情報のデータインプットが可能である。本稿では位置情報は扱わず、段階的開発として刻印種に限定する。城郭石垣の調査では個別石材毎の把握はなされていない。大坂城総合学術調査では壁ごとに刻印種と個数を集計しているため、

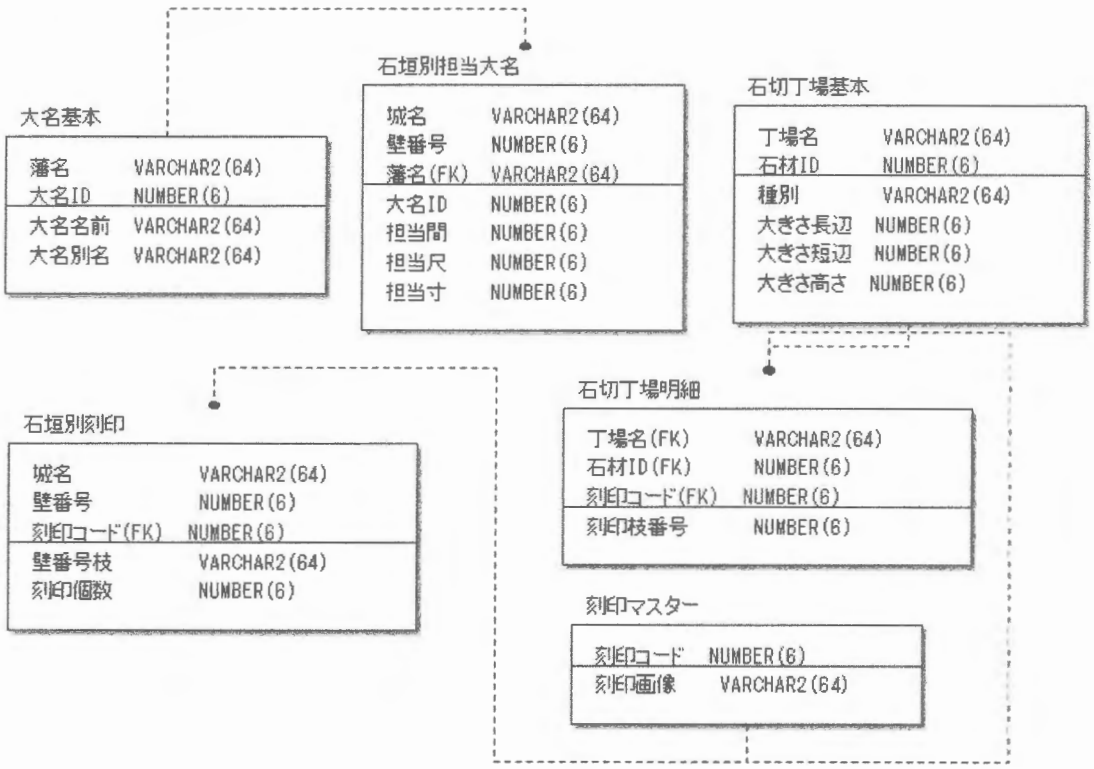


図4 石垣普請データベースのデータモデル(案)

※サイズは仮指定

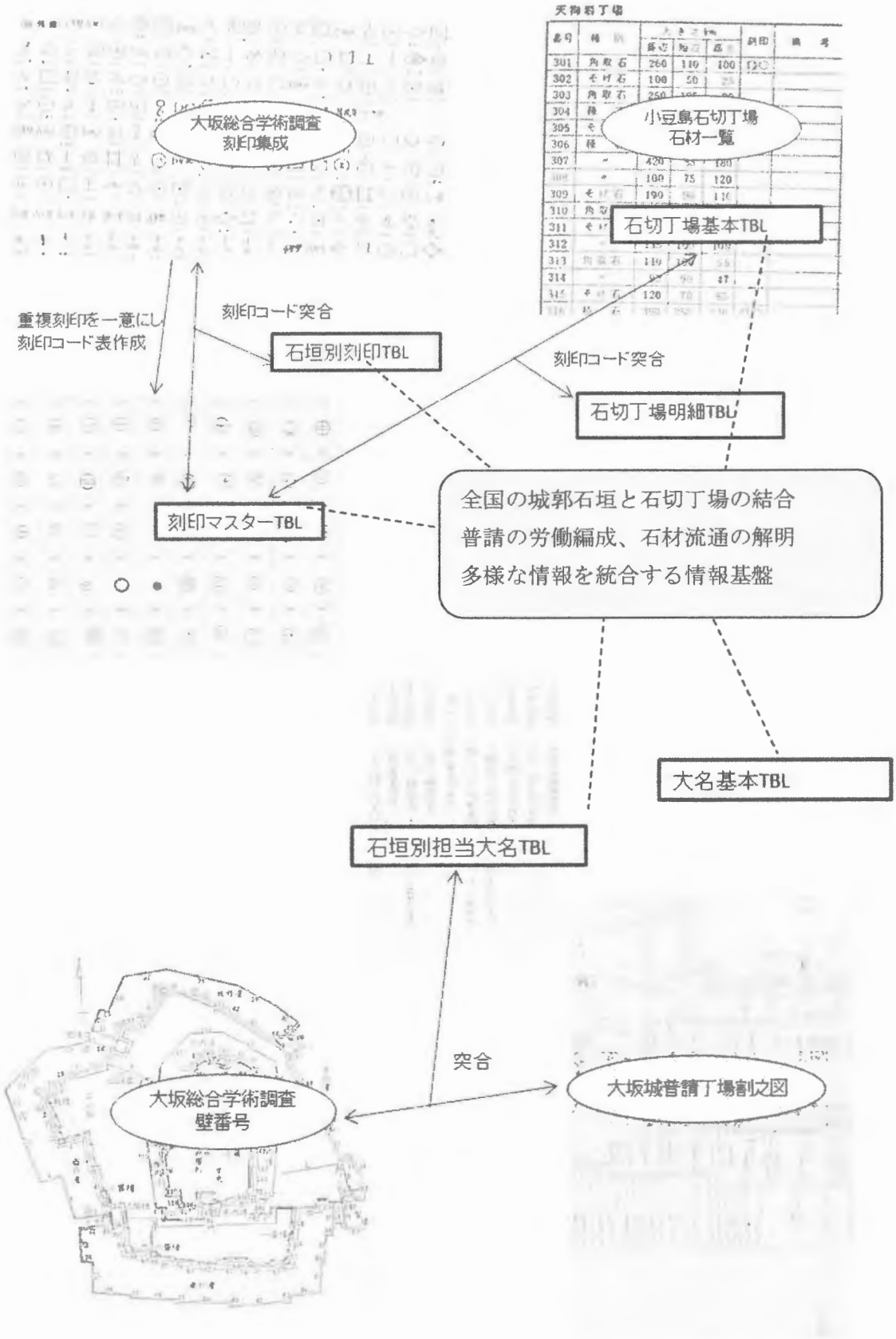


図5　調査情報のデータ化プロセス

壁ごとの情報をインプットデータとする。大坂城の場合、前述の「大坂城普請丁場割之図」を確認すれば壁ごとに担当大名が判明するため、この情報もインプットデータとする。

データベースのデータモデルを試験的に作成した。実運用を考えれば、レコードの論理削除を管理するための削除フラグのカラムを設定し、修正履歴などの項目を設定したり、参照性制約等のビジネスルールを踏まえた設計を行う必要がある。しかし、今回はGUIの入出力インターフェースは検討対象外とし、データベースに直接SQLコマンドにて分析することを目的とする。実運用のための検討やアーキテクチャの設計は別途行う。そのため、今回は石垣・石切丁場・担当大名・刻印に関するカラムに絞り、最低限の構成とした。

6. データ化作業とのデータ内容

上記のデジタルデータは存在していないため、まずは調査データのデジタル化を図った。その作業方法や作成データについて報告する。今回使用したデータは、小豆島の『史跡 大坂城石垣石切丁場跡保存管理計画報告書』(12)と『大坂城の謎』(13)を使用した。作業プロセスを(図5)に示した。補足を要するテーブルのみ解説する。

大坂城総合学術調査の成果である石垣ごとの刻印集成を基に刻印マスターテーブルを作成した。集成にて掲載されている刻印についてすべて重複をなくし、1種類の刻印に1コードを付与することで一意にした。村川氏は刻印の種類を1247種類と報告しているが(14)、筆者が数えると1333種類となった。微妙に形が違う刻印を同一とみなすかどうかで差異が発生したものとみられる。総合学術調査時の原資料は公開されておらず、詳細な検証はできない。また石垣は1号壁から254号壁まであるが、公開資料では1号壁から152号壁の記載にとどまる(15)。理想は網羅して情報化すべきであるが、実際上解決できないため、システム化で回避できない一種の現実的な「割り切り」と考えている。

石切丁場の石材一覧をテーブルにした。ひとつの石材に複数の刻印があるケースがあるため、基本テーブルと明細テーブルに分離した。基本テーブルの1レコードがひとつの石材となり、複数刻印がある場合、明細テーブルに刻印の数だけレコードが挿入される。大坂城石垣になく小豆島石切丁場にある刻印が確認された。すべての刻印種類を網羅し一意なコードを付与する必要があるため、刻印マスターテーブルにレコードを追加した。このような例は他にもあると想定される。兵庫県東六甲山系甲山刻印群で確認されている刻印は加工石材の控え(16)に打たれており、石切丁場では多数確認されているその刻印も大坂城では崩落石垣で見つかった数例しか検出されていない。石切丁場で見つかる加工石材に対し面(17)以外に打たれている刻印は、石積み完了までにその目的・機能を喪失していると考えられ、そのような刻印は多数あるだろう。そのため刻印種類はまだまだ増加する見通しである。

7. 今後の課題と展望

近世初期では多数の藩が日本各地の城郭石垣構築のために集結し、各地で石材を調達した。本稿ではこのような近世の公儀御普請の大きな流れ、大坂城普請と関係する石切丁場を概観したうえで、これらを有機的に結合しうるキーが刻印であるとした。刻印をキーにリレーショナルデータベースの技術で石垣普請データベースを構築すれば多様な分析が可能となる。

データベース化する際の課題にメタデータの統一がある。例えば2005年から2008年に兵庫県教育委員会が実施した徳川大坂城東六甲採石場詳細分布調査では矢穴の有無や矢穴形式なども調査成果をリスト化している(18)。一方、小豆島岩谷地区の調査では矢穴形式などのデータ項目はない(19)。また用語や石材分類の統一も課題となる。兵庫県の分類では割石・矢穴石・調整石であり、小豆島の分類では種石・角取石・そげ石となっている。データマッピングが可能か検討するとともに、メタデータの検討も必要となろう。

将来的にデータベースのWEB化を図り広く利活用される必要がある。利活用され新たに得られた知見をデータベースに循環させることで、データベースの維持管理が可能となりデータの質・量を成長させていくことにつながると考えている。課題は多いが、ひとまず諸賢の叱正を請いつつ、小稿を閉じたい。

<引用・参考文献>

(1) 善積美恵子「手伝普請一覧表」(『学習院大学文学部研究年報』1968年)。
(2) 横田冬彦『日本の歴史16 天下泰平』(2009年、講談社)。
(3) 北垣聰一郎「近世石切丁場研究の現状とその課題」(『ヒストリア別冊 大坂城再築と東六甲の石切丁場』、2009年、大阪歴史学会)。
(4) 北垣聰一郎「石垣構築技術の発達と石材の規格化」(『ヒストリア別冊 大坂城再築と東六甲の石切丁場』、2009年、大阪歴史学会)。
(5) 森岡秀人「築城石・石切場と切石規格化をめぐる一試考」(『橿原考古学研究所論集』第15、2008年)。

(6) 森岡秀人・竹村忠洋編『徳川大坂城東六甲採石場VI　岩ケ平刻印群発掘調査報告書　第32・33・45・67・70・79・81・91地点－平成9・11・14・15・16年度国庫補助事業－』(芦屋市文化財調査報告>第64集、2006年)。

(7) 村川行弘『大坂城の謎』(改訂版)(学生社、2002年)。

(8) 北野博司「大坂城再築における石垣普請の組織と技術」(『城郭石垣の技術と組織』、2012年、石川県金沢城調査研究所)。

(9) 高田祐一・望月悠佑A「甲山刻印群E地区と肥前鍋島家の関係について」(『関西学院考古』10、2007年、関西学院大学考古学研究会)、同B「徳川大坂城にみる大名の石垣普請～肥前佐賀藩鍋島家を例として」(『徳川大坂城東六甲採石場―国庫補助事業による詳細分布調査報告書―』、2008年、兵庫県教育委員会文化財室)、同C「東六甲採石場甲山刻印群」(『別冊ヒストリア　大坂城再築と東六甲の石切丁場』、2009年、大阪歴史学会)。

(10) (7)と同じ。

(11) 多賀左門「東六甲採石場城山刻印群と「十曜紋と一」の刻印」(『歴史と神戸』277、2009年、神戸史学会)。

(12)『史跡　大坂城石垣石切丁場跡保存管理計画報告書』(1979年、内海町教育委員会)。

(13) (7)と同じ。

(14) (7)と同じ。

(15) 大坂城総合学術調査時に254号壁まで網羅して調査されているか不明である。もしくは未整理なのか不明。総合学術調査団としての公的な報告書が刊行されなかった以上、現時点ではどういう状況かわからない。

(16) 石材の奥行のこと。当然、石積みすると表面からは見えなくなる。

(17) 石垣で見えている石の表面。

(18)『徳川大坂城東六甲採石場―国庫補助事業による詳細分布調査報告書―』(2008年、兵庫県教育委員会文化財室)。

(19) (12)と同じ。

歴史と数理
Mathematical Understanding of History

小沢 一雅
Kazumasa Ozawa
大阪電気通信大学，大阪府寝屋川市
Osaka Electro-Communication University, Neyagawa, Osaka.

あらまし：歴史は，さまざまな数値情報を包含している。本稿の研究目標は，歴史における数値情報から歴史の理解に役立つ法則性を抽出することができるかどうか，できるとすればどのような方法があるか，などについての実証的探究にある。今回とくに注目したのは巾乗則であって，日本古代における数値データの分布を巾乗則の観点から分析することによって古代社会の進展に関する新たな知見が導かれている。この分析では，数理モデルにもとづくシミュレーションも導入され，重要な役割をはたしている。こうして得られた知見を総合して，「歴史はいかに進行するか」という課題についても簡潔に論ずる。

Summary: This paper presents a consideration on how to perform mathematical understanding of history. Power law has especially been taken into account for analyzing numerical data appeared in history: It has been shown that either distribution of volumes of 23 ancient Japanese tomb mounds or that of families belonging to each of 30 ancient counties can reasonably be explained in terms of power law. A mathematical model-based simulation to grow a distribution of variables with power law has been carried out. Its results have been playing an important role in our work. The concluding remark is presented, which includes a short discussion on "How does history progress?".

キーワード：歴史，数理，べき乗則，シミュレーション，日本古代史，前方後円墳
Keywords: History, Mathematics, Power law, Simulation, Japanese ancient history, Ancient tomb mounds

1． はじめに

歴史とは何かという根源的な問いについて厳密な回答を与えるのは，（筆者には）難しいが，辞書的な表現でいえば，「歴史とは，人間社会が経てきた変遷の軌跡」ということになろう。本稿では，とりあえずこうした一般的な理解にしたがうことにする。一方，歴史との関連でいうところの数理とは，そこにある数値情報をどのように理解するか，つまりそれがどういう意味をもつのか，あるいは何らかの法則性があるのかどうかといった，数学的な姿勢を交えて考えていく思考法をいう。起点はあくまで数値情報である。

歴史にはさまざまな数値情報があるが，重要であっても計量できないものがある。歴史の進行とともに一貫して単調増加してきた知識の総量（情報量）がそれである。いまや人類のもつ知識の総量は，（何ビットなどと）計量することもできないほど膨大な量に達していると想像される。人類という知的生物の比類なき能力は，知識の獲得とその蓄積能力にあるといっても過言ではない。近代における産業の進歩や生産力の急激な増加といった現象も，根源的にみれば，累々と蓄積された知識の活用の成果にほかならない。たとえば，人類史における数値情報として人口を例にとってみると，増加期，停滞期，あるいは減少期が不規則に反復し，必ずしも増加一辺倒ではなかった。減少期が続くと，祖先たちは滅亡への危機感からそれを克服する手段を必死に模索したであろう。首尾よくそれが獲得できれば，知識として蓄積できたと考えられる。成功も知識，失敗も知識として蓄積されたはずである。

いわゆる文化・文明は，このように一貫して単調増加してきた知識の反映として生み出されてきたものと考えるべきであろう。一方，歴史を形成する人間社会の変遷には，ここでいう知識の総量と同様な，一貫して単調増加するような決定論的（非確率的）な方向性は存在しない。筆者は，むしろ歴史は偶発

的な事象によって生じる変動の軌跡ととらえるべきであると考えるに至った。こうした歴史認識は，マーク・ブキャナン（Mark Buchanan）のいう，人間社会を非平衡状態（あるいは臨界状態）にある個体の集団とみるとらえ方と軌を一にする[1]。

本稿では，歴史を数理という視点から理解する方法論を探究するにあたって，これまで筆者が試行したいくつかの事例研究をとりあげ，そこで現れた数値情報（データ）について新たな角度から再検討する。すなわち，巾乗則（べき乗則）が成立するかどうか，もし成立する場合には，どういった知見がみちびかれるか，などについて検討を行う。くわしくは後述するが，巾乗則とは，データの分布を表現する関数タイプの1つにかかわる事項であって，近年フラクタルや複雑系の科学などの話題の中でもよくとり上げられている[2]。

2．歴史における数値情報

歴史を考える主な学問分野として，文献史学や考古学がある。いまわれわれが知っている歴史，たとえば日本史とは，これらの伝統的な諸分野における長年の研究が総合された成果にほかならない。とくに，考古学は「もの」（遺物・遺構）を対象とした学問分野であるため，多くの数値情報が現れる。筆者は，前方後円墳の形態研究を行ってきたが，人工造形物である墳丘の規模や形状はまさに数値として計量してはじめて詳細な分析が可能になる。ほかの遺物等についても同様であって，考古学はその成り立ちからしてとりわけ数理との関係が深い分野といえるであろう。

一方の文献史学では，一見すると数値情報がそう大きな役割をはたしていないように見えるが，じつはそうではない。たとえば，年代は明らかに数値情報である。歴史上重要な事蹟がいつ起こったか，その年代は歴史を読み解く上で鍵となる基本情報である。文献史学では，こうした年代は通常文献から抽出できるし，それが問題になることは比較的少ない。ただし，文献に記載されている年代が信頼性を欠く場合が出てくると，事蹟の年代が正しく特定できないことになり，重大な問題が発生する。つまり，歴史を読み解くための基本情報が欠落するといった事態に陥るのである。文献が豊富ではない古代史では，このような状況がしばしばみられる。邪馬台国問題もその一例である。

たとえば，3つの事蹟A，B，Cがあり，事実としての年代順序がA→B→Cであったとき，もしBの年代が特定できない事態が起これば，B→A→C，あるいはA→C→Bのように誤ったストーリーが歴史の解釈として現れる危険性があるわけである。

年代についていえば，考古学において行われる遺物の年代決定（土器の編年など）にも問題が含まれている。すなわち，おなじ遺物についても年代観が研究者によって異なる事例が現実にあり，もしその年代が歴史解釈上の鍵になる場合には，まったく異なった解釈が並立することになるのである。

年代以外にも数値情報は存在する。人口もその例であるし，戦争における死者数なども重要な数値情報である。こうした歴然とわかる直截的な数値情報ではなく，複数の異なった量を複合して導出される数値情報もありえる。現代では，たとえばＧＤＰ（国内総生産）といった数値情報が社会の動向を読み解く上で重要な役割をはたしている。このようないわば抽象的な数値情報の導入も，歴史と数理という観点からは積極的に検討すべき課題であろうと思われる。

3．歴史に現れる巾乗則

3．1　巾乗則の数学

近年，フラクタルや複雑系の科学という話題の中で，しばしば巾乗則という数学用語が現れる[2]。ここでは，具体的な事例をとりあげて解説する：

いま，日本の 2012 年の都市人口を降順にならべてみる（表1参照）。トップはなんといっても東京であって，８９５万人（特別区内）。第2位は横浜市で３６９万人。第3位は大阪市で２６７万人。以下，名古屋市，札幌市，神戸市，京都市・・・と続いていく。最後尾は，第７８９位の歌志内市（北海道）で，人口約4千人である。７８９都市が人口という数値で順位づけられている。固有名詞である市名で議論するのは煩雑なので，つぎのように数式的に簡潔に表現しておく。

$$P_m = \text{第 } m \text{ 位の都市人口} \qquad (m = 1, 2, \cdots, 789)$$

さて，表1にある順位づけられた都市人口はつぎのような数式（巾乗関数）で高精度に近似できることが判明している。

$$P_m = A\, m^{-D} \qquad (1)$$

ただし，A と D はデータ（ここでは都市人口データ，表１参照）によって決まる定数である。D は（フラクタル）次元とよばれている。一般的にいえば，次元とはデータの分布全体との対比において，局所への量の集中度を表す尺度になっている。ちなみに，表１から算出される２つの定数の値は，それぞれ，A = 1000 万 および D = 0.91 である。図１に，この定数の場合の巾乗関数を曲線としてグラフ化している（60 位までを表示，以降は省略）。

表１　都市人口（抜粋）

順位	府県名	市名	法定人口
1	東京都	特別区部	8,949,447
2	神奈川県	横浜市	3,689,603
3	大阪府	大阪市	2,666,371
4	愛知県	名古屋市	2,263,907
5	北海道	札幌市	1,914,434
6	兵庫県	神戸市	1,544,873
7	京都府	京都市	1,474,473
8	福岡県	福岡市	1,463,826
9	神奈川県	川崎市	1,425,678
・	・	・	・
・	・	・	・
・	・	・	・
786	北海道	赤平市	12,637
787	北海道	夕張市	10,925
788	北海道	三笠市	10,225
789	北海道	歌志内市	4,390

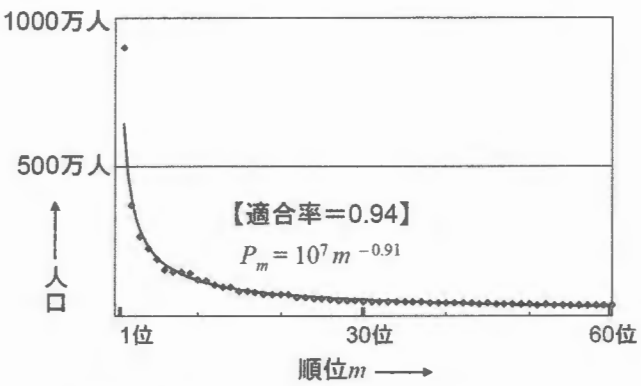


図１　都市人口の巾乗則分析（60 位まで表示）

いま，「高精度」で近似できると述べたが，図１の曲線が実際の都市人口の分布にどれだけフィットしているかを適合率（R^2 値）として定量的に評価することができる。図１の曲線の適合率は 0.94（94%）であって，非常に高い。まさに高精度である。

一般に，データ分布が式（１）の形で高精度に近似できるとき，巾乗則にしたがうという。適合率がどの程度以上なら高精度といえるかについては，経験則として，筆者は 0.90 以上という基準を想定している。

３．２　考古学における巾乗則（前方後円墳）

考古学において巾乗則が成立する事例があるかどうか，あるとすればいかなる適合率で次元はどの程度か等々について，何らかの検討を行った先行研究はこれまでにない。ここでは，前方後円墳を具体例として，巾乗則との関連性を検討する。

表２　最古級前方後円墳（墳丘長 50m 以上）

古墳名	墳丘長	墳丘体積	順位
箸墓	281m	350000 m^3	1
椿井大塚山	170	79000	2
浦間茶臼山	138	42000	3
中山大塚	132	37000	4
ヒエ塚	125	31000	5
弁天山A1号	120	28000	6
中山茶臼山	119	27000	7
石名塚	111	22000	8
石塚山	110	21000	9
森1号	106	19000	10
丁瓢塚	104	18000	11
元稲荷	94	13000	12
纏向石塚	93	12800	13
網浜茶臼山	92	12600	14
植月寺山	91	12000	15
原口	81	8500	16
美和山1号	80	8200	17
那珂八幡	75	6800	18
操山109号	74	6500	19
聖陵山	70	5500	20
御道具山	65	4400	21
弘法山	63	4000	22
川東車塚	61	3600	23

前方後円墳は，４世紀初頭から６世紀末にいたる古墳時代，全国各地で盛んに築造された日本古代の個性的なモニュメントである．慣習上，古墳時代は前期・中期・後期に区分されることが多い．前期から中期へ，さらに後期へと，前方後円墳の形態は時間的に変化していくが，その形態変化には規則性が認められる[3]．また，石室の構造，副葬品，あるいは墳丘上の埴輪なども時期の推移にともなって変化することが知られている．

ここで，数値情報として，前期の中でもっとも古い古墳たち（「近藤編年」[4]でいう“１期”に相当）であって，古墳時代の幕開けに登場する最古級前方後円墳の墳丘体積をとりあげる．表２に，最古級前方後円墳２３基を墳丘体積の降順に示している．なお，同表にある体積値はすべて筆者の算定（概算値）によるものである．巨大古墳である箸墓古墳（奈良県桜井市）は同表のトップに現れている．

さて，表２に示される体積の分布は巾乗則にしたがうことが判明した．すなわち，適合率＝0.95 および次元＝1.23 の巾乗関数で高精度に近似できることがわかった（図２参照）．そこで，導出された次元値（＝1.23）がどういう意味をもつかについて少し考えてみる．

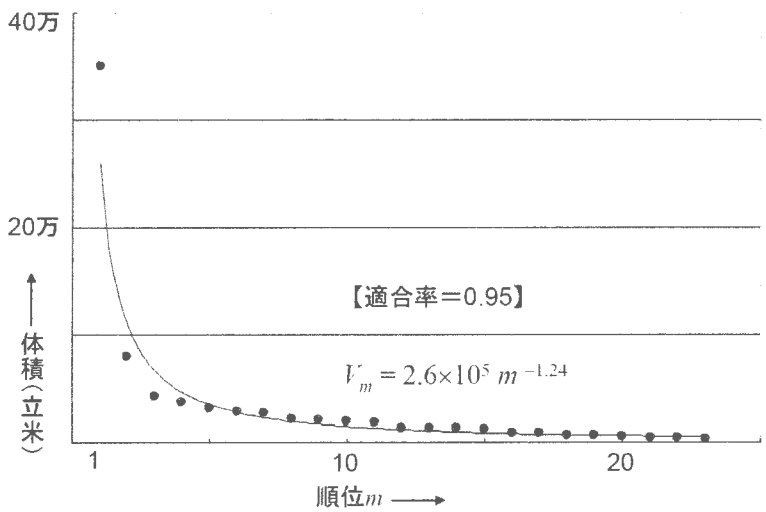


図２　最古級前方後円墳の巾乗則分析

そもそも前方後円墳の築造は土木工事であって，たとえば，仁徳天皇陵（堺市）の墳丘築造に要する総費用の試算（大林組）では，計算の根拠は墳丘体積におかれている[5]．つまり，掘削や盛土などの作業が中心になる前方後円墳の築造では，費用の大半はこうした土工に必要な労働（労役）に投入されることになり，結果として土量，すなわち墳丘体積を基準としてほぼ比例的に諸費用が計上されていくのである．

単純化すれば，前方後円墳の築造に要する総費用は，墳丘体積に正比例すると考えてよい．この前提で２３基の古墳が体積の降順にならんでいる表２の体積分布を考えてみる．慣習としてよく用いられる上位１０％という線引きで上位を抜き出すと，この場合は箸墓古墳と椿井大塚山古墳の２基になる．２３基全体の総体積に占める上位１０％（２基）の割合はほぼ５０％である．この２基の所在地は，奈良県（箸墓）および奈良に近接する京都府（椿井大塚山）であって，ほぼ畿内の中心地区（奈良盆地周辺）に位置するといってよい．したがって，最古級前方後円墳を体積でみれば，畿内の中心地区におよそ５０％の割合で集中化が進んでいるとみなすことができる．上述の次元値（＝1.23）は，このレベルでの集中化の度合（集中度）を意味すると考えられる．なお，次元が１前後の値をとるとき，上位１０％が全体に占める割合はつねに５０％程度であって，前述の都市の人口分布でもほぼ同様の占有率を示す．

さて，畿内中心地区における最古級前方後円墳の体積の占有率が５０％に達していることと，前方後円墳築造の費用が墳丘体積に比例するという，２つの知見を総合すれば，前方後円墳時代の幕を開いた大和政権の「財」は，その支配地域全体との対比の中でかなり集中化が進んでいたという推理も成り立つであろう．

３．３　魏志倭人伝の巾乗則

魏志倭人伝には，よく知られているように帯方郡から邪馬台国をはじめとして古代の国々に至る道程と国々の戸数が記されている．ここでは，国々の戸数に着目して巾乗則が成立するかどうかについて検討することにする．

魏志倭人伝には，戸数が明記された８つの主要な国が記載されている．加えて，戸数は不明であるが女王国と敵対する狗奴国がある．表３に，これら９ヶ国とそれぞれの戸（家）数を示している．さらに，戸数その他の詳細は不明と断っているが，女王国に属する国々として２１ヶ国（斯馬国，己百支国，伊邪国，都支国，弥奴国，好古都国，不呼国，姐奴国，

對蘇国，蘇奴国，呼邑国，華奴蘇奴国，鬼国，爲吾国，鬼奴国，邪馬国，躬臣国，巴利国，支惟国，烏奴国，奴国）が列記されている。国名が明記されているのは，以上の３０ヶ国である。

また，別記として，「女王国の東，海を渡ること千余里，また国あり，皆倭種なり」という記述があるが，文脈上これら「倭種」等は重視されてはいない。

国名	戸数
対馬国	千余戸
一大国（壱岐国）	三千家
末盧国	四千余戸
伊都国	千余戸
奴国	二万余戸
不弥国	千余家
投馬国	五万余戸
邪馬台国	七万余戸
狗奴国	—

表３　魏志倭人伝の主要国と戸数

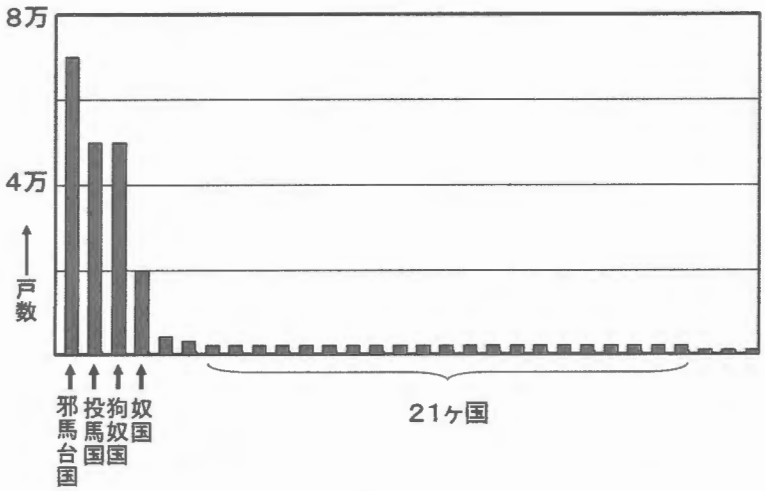


図３　魏志倭人伝３０ヶ国の戸数分布

やはり，魏志倭人伝は，国名を記した上述の３０ヶ国を当時における主要な国々とみなしていると考えてよいだろう。そこで，これら３０ヶ国の戸数が巾乗則にしたがうかどうかの検証を行いたいが，戸数がわかっているのは表３の中の８ヶ国だけであって，その他２２ヶ国の戸数が不明な点が問題である。これらについては筆者の推定値を用いることにする。筆者の推定値とは，狗奴国を５万戸，女王国に属する２１ヶ国をすべて均等に２千戸とするものである。詳細は拙著を参照されたい[6]。

こうした推定値を導入するなどして３０ヶ国にすべて戸数を付与した（「三千余戸」→ 3000 戸のように，戸数データはすべて完数化している）。図３に，３０ヶ国の戸数を降順に棒グラフで示している。横軸に主要な国名などを記入している。つぎに，戸数を順位づけるわけであるが，ここでいう順位というものを厳密に規定しておく必要がある。

【順位の決定】戸数の順位とは，その戸数以上の戸数をもつ国の数である。図３で具体的にいうと，７万戸以上の戸数をもつ国は１つしかないから，７万戸は第１位である。つぎの５万戸の順位は，それ以上の戸数をもつ国が３つあるから，３位である。「２位」はない。以下同様に戸数の順位を算定していくと，戸数の順位は表４のようになる。２１ヶ国の戸数を推定値としてすべて２千戸としたので，戸数２千戸の順位は２７位になっている。１千戸の順位は３０位である。もし，国々の戸数がすべてちがっていれば，１位から３０位まですべてが埋まり，表４のような順位の「飛び」は起こらない。

表４　戸数の順位

順位	戸数
1	70000
3	50000
4	20000
5	4000
6	3000
27	2000
30	1000

さて，表４の戸数データに巾乗関数のあてはめを行うと，適合率は 0.81 であって，近似としてみたときの精度は低い（図４参照）。３０ヶ国の戸数を順位でみた分布は，巾乗則にしたがわないと判定できる。

ただし，順位の上位１０％にあたる国々（３位以上として３ヶ国）の戸数が全体に占める割合は７

0%に達している。つまり，上位10%による占有率だけは高いが，分布そのものは巾乗則ではないということである。これは，じつは巾乗則の意味を理解する上できわめて参考になる事例であって，(後述の) シミュレーションによって巾乗則を考察する場面で再度とりあげることになる

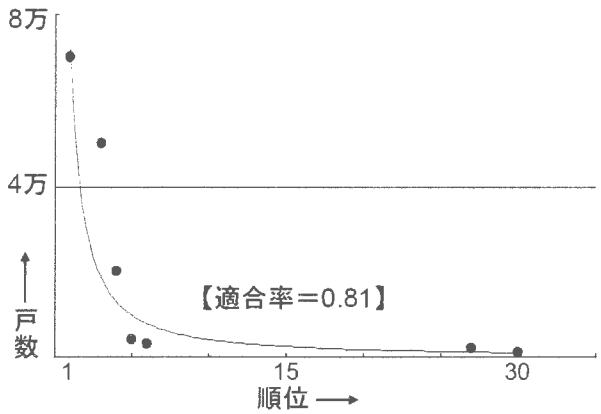


図４　３０ヶ国の戸数の巾乗則分析

４．シミュレーション研究

４．１「人口」・「勢力」・「財」の集中化モデル

日本の都市人口（表１参照）は，順位でみたとき，完全な巾乗則にしたがう分布であることが判明した。前述のように，最古級前方後円墳の墳丘体積もおなじく巾乗則の分布をなすことが明らかになった。ブキャナン（M. Buchanan）は，巾乗則の成立に関して，映画や音楽のヒット作品の売上収入，アメリカの都市人口，個人の年間収入（資産）等々，多くの事例を紹介している[1]。さらに，こうした具体例の外見のちがいにかかわらず，巾乗則が出現してくる根底には，ある共通な原理が働いていると述べている。この原理なるものについて，ブキャナンは数式を使わず文章で説明しているため，細かいところであいまいさが残っている。筆者は，この原理なるものを「集中化モデル」と名づけて自己流に定式化した。以下のように数学的にはきわめて簡単な定式化である。

いまK個の同質の変量体があり，それぞれの値を$P_1, P_2, \cdots, P_K$とする。K個の変量体とは，K個の都市人口でもよいし，K人の資産としてもよい。変量体の値は時間的に変化すると想定し，用語的には「世代」の進行にともなう変化と表現することにする。すなわち，各変量体の値は，世代n $(=1, 2, 3, \cdots)$の関数としてつぎのように表すことにする。

$$P_k(n) \quad (k = 0, 1, \cdots, K)$$

さて，集中化モデルとは，はじめは「ドングリの背比べ」のように値に関して大きな格差のなかったK個の変量体の間に，世代の進行にともなって次第に格差が現れ，巾乗則的に集中化が進んでいく数学モデルである。集中化モデルの定義を，n世代の変量体の値からつぎの$n+1$世代の値が導かれる漸化式としてつぎのように与える。

$$P_k(n+1) = P_k(n) + \delta r P_k(n) \qquad (k = 1, 2, \cdots, K) \qquad (2)$$

ただし，δは，確率0.5で正負（つまり+1か−1か）が決まる変数，およびrは1世代進むときの変量体の値の変動分を決める変化率(<1)である。つまり，つぎの世代の値は，いまの世代の値に変動分（正負）を加えたものであり，さらにその変動分の大きさは，いまの値に比例するという形式である。

４．２　シミュレーションの実行

集中化モデルによるシミュレーションは，初期値（$n=1$におけるK個の変量体の値）と変化率rを設定して実行する。図５にシミュレーションの実行例を示している。この例では，前述の魏志倭人伝の国々の戸数をイメージして変量体を３０個設定し，初期値をすべて均等に3000とした。すなわち，

$$P_k(1) = 3000 \qquad (k = 0, 1, \cdots, 30)$$

と設定し，変化率については$r=0.1$とした。

世代の進行にともなって各変量体の値が増減していくわけであるが，式（２）は，ある変量体の値がいったん０（ゼロ）に落ち込めば，それ以降は永久にその状態から脱出できない形になっている。図５のシミュレーションの実行例では，こうした消散的現象を止める目的で下限（=1000）を設定している。すなわち，ある変量体の値が1000以下になると，強制的に1000にもどすという例外的処理である。

さて，このシミュレーション例では，最初すべておなじ初期値（$n=1$において3000）から出発した３０個の変量体の値の分布は，図６に示すように１００世代（$n=100$）ぐらいに達すると適合率0.95で完全な巾乗則にしたがう集中化状態に到達する。さきの魏志倭人伝の国々の戸数分布は適合率0.81であっ

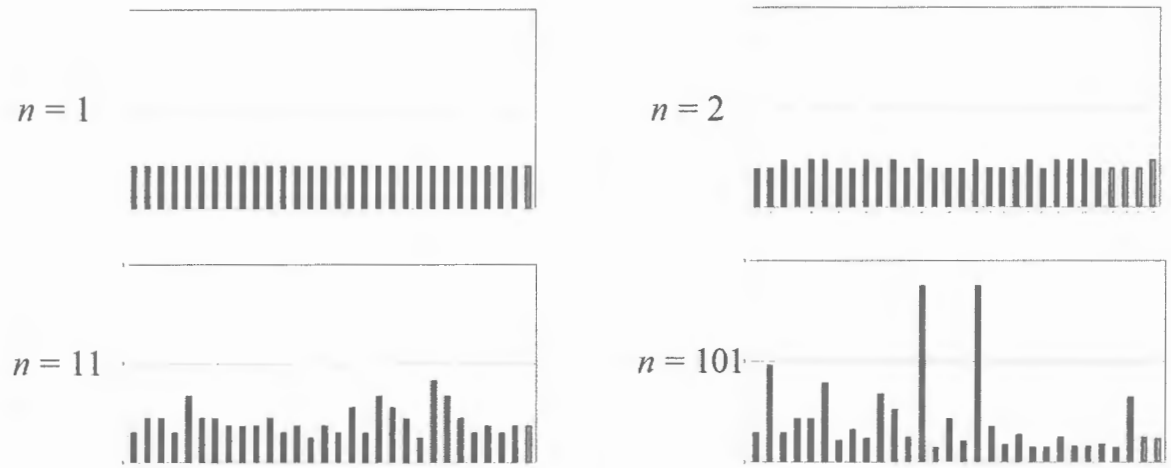


図５　均等な初期値からスタートしたシミュレーション実行例（サンプル系列）

（*n* は世代。３０個の変量体を横にならべている。縦軸は変量体の値）

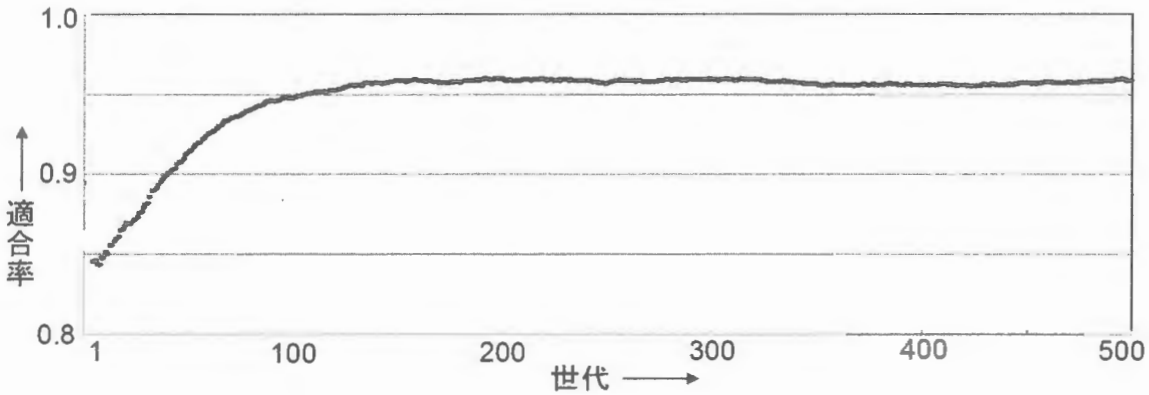
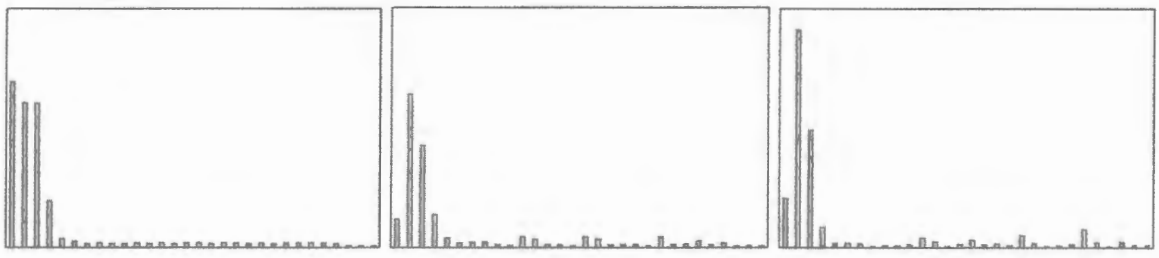


図６　世代の進行と適合率の変化

（1000 回のシミュレーションで発生した 1000 系列の平均値）



n＝1〈初期値(30ヶ国)〉
適合率0.81

n＝ 50
適合率0.95

n＝ 100
適合率0.94

図７　30 ヶ国の戸数（図 3）を初期値とする仮想シミュレーション

て，巾乗則にしたがわないと判定された。

図６において，適合率が 0.90 未満となるのは，５０世代までのごく初期の過渡的な区間であることがわかる。古代の国々の戸数分布は，世代の進行とともに巾乗則的な集中化に向かっていく過渡的な段階に位置づけられると考えたい。図７に，古代の３０ヶ国の戸数（図３）を初期値として，その後１００世代に至る戸数分布変遷の仮想シミュレーション（サンプル）を示している。この例では，グラフ左端に位置し最初トップの戸数を誇っていた邪馬台国

が５０世代で没落する一方，もとは２位につけていた投馬国が戸数をやや増加させながら１位につけるなどの大きな変化がみられる．１００世代になると，トップの投馬国がさらに躍進し，その地位を確立していく．このような流れを「政権」の交代と解釈することができれば，集中化モデルの意味がさらにふくらむことになろう．

このシミュレーションでは，最初は適合率が0.81で巾乗則にしたがわないとされた３０ヶ国の戸数分布が，５０世代になると適合率が0.95に達し，完全な巾乗則成立の意味で集中化状態に到達している．

あくまで確率過程であるため，若干の変動はありえるものの，５０世代以降についてもしばらくの期間，適合率はほぼ高水準で推移し，巾乗則成立レベルを保つものと考えてよい．

４．３　巾乗則的な集中化の意味

集中化モデルは，世代の進行とともに変量体の値の分布を巾乗則的な集中化に向かわせるという機能をもっている．はなはだ単純なモデルであるが，多くの社会現象にみられる巾乗則の生成原理の１つを提示していると考えてよいだろう

ここで，巾乗則的な集中化の意味を考えるために，そもそも巾乗則とは何か，あるいは巾乗則という“分布則”の背後にあるものは何かを理解する必要がある．じつは，巾乗則と表裏一体の関係にあるものはフラクタルである．本稿の話題と関係する例として樹形フラクタル[7]をとりあげよう．

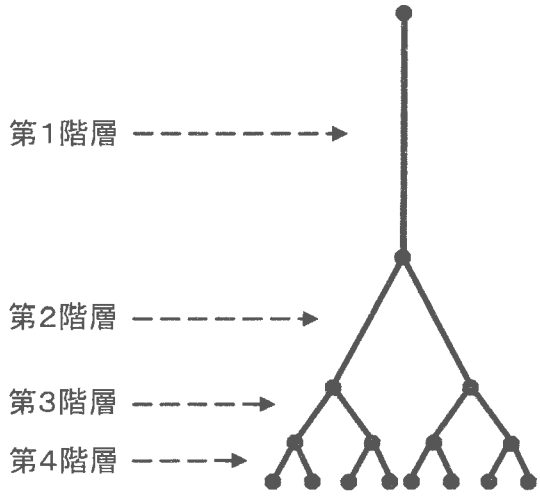


図８　樹形フラクタル

図８は，２分岐の樹形フラクタルの説明図（実際の樹木とは天地を逆転して表示）であって，最上位の１本の枝（幹）が，つぎの第２階層で２つの枝に分岐する．つぎの第３階層では，それぞれまた２つの枝に分岐する．このように，階層が下降するにつれ分岐がくり返されていくのが樹形フラクタルである．もし，階層が１つ下降するたびに，枝の長さが半分になるとする．この場合，最上位の枝の長さを１としたとき，第６階層以上で生成されたすべての枝（６３本）の長さを順位づけると，表５のようになる．

表５　樹形フラクタルの枝長の順位

階層	枝の数	枝長	順位(枝長)
1	1	1	1位
2	2	1/2	3位
3	4	1/4	7位
4	8	1/8	15位
5	16	1/16	31位
6	32	1/32	63位

ここで，表５の順位でみたときの枝の長さ（枝長）の分布に巾乗関数のあてはめを行うと，適合率は0.99（99%）になり，ほぼ完全な巾乗則を示すことがわかる（図９参照）．

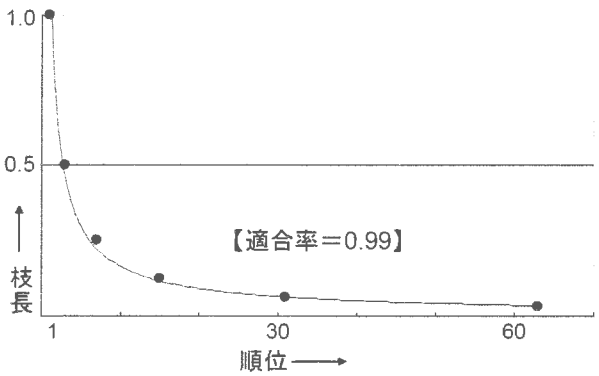


図９　樹形フラクタルの枝長の巾乗則分析

さて，図８の樹形フラクタルは，見方を変えれば，最上位の枝をトップとする階層的な支配関係（ヒエラルキー）を表す組織図とみることができる．適合率99%というほぼ完全な巾乗則の背後にはこうしたフラクタル構造が潜んでいるわけである．したがって，これまでみてきた変量体の値の分布と巾乗関数との適合率とは，その分布の背後にひそむ変量体間の関係がフラクタル構造（階層的な支配関係）にど

れだけ近いかを示すものと考えてよいだろう。

そこで，図３に示した魏志倭人伝の国々の戸数の話にもどってみる。前述の結論は，上位１０％にいる３ヶ国の戸数占有率は高いが，巾乗則にはしたがわない，ということであった。巾乗則にしたがわないということは，国々の関係としてフラクタル構造（階層的な支配関係）が未成熟であることを示唆している。実際，魏志倭人伝が伝えるように，邪馬台国を本拠とする卑弥呼は，国々の協議を経て倭国の女王に「共立」されたのであって，強権によって国々を支配下においた女王ではなかった。図３における上位３ヶ国の中には卑弥呼と交戦中の狗奴国も入っている。３ヶ国の戸数占有率が高いというさきの結果は，単なる統計上の数合わせにすぎないのであって，本質は，むしろ巾乗則的な集中化という意味において当時の倭国はいまだ未成熟なレベルにあったと考えるべきであろう[6]。

４．４　歴史はいかに進行するか

人類は，この地球上に姿を現して以来，一貫して社会的動物として群れをなしつつ生活を営んできた。群れとは，烏合の衆ではなく，トップに支配者がいて，群れ全体の指揮をとる形態をいう。支配者の存在は，群れが生存するための必須条件である。大きな群れになると，内部に何層もの階層的な支配構造が必然的に発生する。さらに，群れと群れの競合関係が，支配と被支配の関係に変化すれば，そこにも階層的な構造が生まれていく。これも最大多数の生存にとって必要な社会的変化であった。

こうした動きの全体が，これまで述べてきた巾乗則的な集中化という方向性を具現化していくと考えてよいだろう。このような方向性が生まれる根源には，そもそも社会を構成する人々の関係が，絶対安定な平衡状態にあることはなく，一定の拘束の中でつねに流動性をもつ臨界状態にあるという“遺伝的性向”を想定せざるをえない。

このような臨界状態においては，個々の人たちの偶発的な動きが起点となって多数を巻き込む動向が生まれていく。社会の大きな変動に進展することもありえる。歴史上にみられる社会変動とは，つきつめるとこうした偶発的な事象によって惹き起こされるものであって，個々の人々の動きを超越した何らかの必然的法則に導かれて社会が変動するというようなものでは決してない。当然ながら，地球上で地域を超えて，類似の社会的変動が起こるというような“普遍的法則”もありえない。地域ごとにある偶発的な事象や個人の動きによって，それぞれ質的に異なった歴史が編まれていくのである。もし，あえて法則というべきものがあるとすれば，人間の遺伝的性向から集中化に向かおうとする方向性をあげる以外にはない。

はじめに述べたように，歴史において一貫して単調増加してきたものは知識の総量である。知識の増大は，巾乗則的な集中化のありようにも影響を及ぼしてきたであろう。人々の個々の動きの中で，知識としての「経験」が影響を及ぼす一面はつねにありえたであろうし，知識によって生み出された技術が影響を及ぼすこともありえる。文化・文明も，知識の反映である。地球上では，多くの地域を包括するいくつかの文明圏が成立してきた。１つの文明圏内では，おなじ共通の文化的文明的拘束の中で人々が流動的に動きながら個性的な臨界状態をつくってきたとみられる。しかし，集中化に向かうという方向性だけは，古今東西を通じてつねに不変であったと考えてよいだろう。

一方，巾乗則的な集中化が進んで一定のフラクタル的構造ができあがったとしても，それがそのまま永遠に持続する保証はない。臨界状態における人々の偶発的な動きが引き金となって，その崩壊化がはじまる事態も確率的にはありえる。が，仮に，いったんできあがったフラクタル的構造が崩壊したとしても，遺伝的性向である集中化の動きがすぐにはじまり，究極としてつぎの新たなフラクタル的構造が姿を現すことになる。大地震で社会基盤が一挙に崩壊しても，ただちに復興の動きがはじまることに似ている。まさに遺伝的性向である。

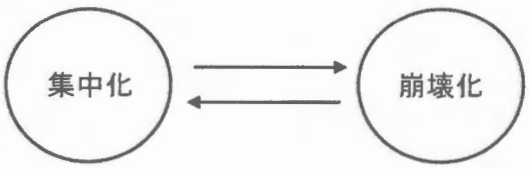


図１０　集中化と崩壊化

歴史の進行とは，巾乗則的な集中化と，その逆の崩壊化の反復過程であるといえよう（図１０）。反復のきっかけとなるものはすべて偶発的な事象によるものであって，そこでは人々の個々の動きが起点となる。反復過程において，一貫して増大を続ける知

識が良きにつけ悪しきにつけ影響を及ぼし続けてきたことは，これまでの歴史が物語っている。

歴史とは偶発的な事象をきっかけに惹きおこされた社会の変遷の軌跡である。「歴史に“ｉｆ”はない」とよくいわれるが，本稿の文脈からしてまさに真理をついた言葉である。

５．むすび

本稿は，歴史を数理という視点から理解する方法論を探究するにあたって，巾乗則に注目し，過去に筆者が試行したいくつかの事例研究で現れた数値情報にこれを適用して考察を試みた。

本文中でも述べたように，最古級前方後円墳の墳丘体積に関する分析では，高精度の近似で巾乗則が成立することが判明した。フラクタル的構造の存在が暗示されるのである。前方後円墳の築造に必要な「財」の動員力が墳丘体積に比例するという知見からすれば，大和の政権がこの時代にフラクタル的構造のトップに位置していたという推論が成立することになる。少なくとも筆者の関心についていえば，この点は新しい重要な知見であって，さらにくわしく検討してみたいと考えている。

本文中で言及した巾乗則的な集中化とは逆の，崩壊化についてはくわしく論ずることができなかった。当然，偶発的な事象をきっかけとして雪崩のように起こると考えられる崩壊についても適正なモデルがありえると思っている。今後の課題として考えてみたい。

本稿で紹介した集中化モデルは，都市人口，ウェブサイトの訪問者数，個人資産，音楽作品の販売数等々に関する変動のシミュレーションにも有効な一定の普遍性をもつと考えている。歴史とは直接関係のないこうした多くの数値情報についての分析が進めば，巾乗則のもつ意味がさらに明瞭に浮かび上がってくるものと期待している。

最後に，本稿で述べたシミュレーションの実施，およびデータ入力に協力してくれた奥島健一氏に感謝する。

【参考文献】

[1] Mark Buchanan, *Ubiquity*, Crown, New York, 2001.
本書の翻訳版（水谷淳訳）は，2003 年に早川書房より単行本『歴史の方程式―科学は大事件を予知できるか』として刊行された。さらに，2009 年，同社より改題した文庫本『歴史はべき乗則で動く― 種の絶滅から戦争までを読み解く複雑系科学』が刊行された。

[2]高安秀樹・高安美佐子『経済・情報・生命の臨界ゆらぎ』，ダイヤモンド社，東京，2000.

[3]小澤一雅 『前方後円墳の数理』，雄山閣，東京，1988.

[4]近藤義郎『前方後円墳集成』（全５巻），山川出版社，東京，1992.

[5]中井正弘 『仁徳陵―この巨大な謎』，創元社，大阪，1992.

[6]小澤一雅『卑弥呼は前方後円墳に葬られたか―邪馬台国の数理』，雄山閣，2009.

[7]小澤一雅『パターン情報数学』，森北出版，東京，1999.

主催：第１８回公開シンポジウム
　　実行委員会

共催：大阪電気通信大学
　　情報学研究施設

後援：人文系データベース協議会

第１８回公開シンポジウム実行委員会

委員長：加藤常員（大阪電気通信大学）
委　員：川口　洋（帝塚山大学）
　　宝珍輝尚（京都工芸繊維大学）
　　小森政嗣（大阪電気通信大学）
　　竹内和宏（大阪電気通信大学）

人文系データベース協議会

議　長：田田和久　（奈良女子大学）
副議長：加藤常員　（大阪電気通信大学）
副議長：川口　洋　（帝塚山大学）
　　赤間　亮　（立命館大学）
　　江澤義典　（関西大学）
　　及川昭文　（総合研究大学院大学）
　　桶谷猪久夫（大阪国際大学）
　　小沢一雅　（大阪電気通信大学）
　　金田明大　（奈良文化財研究所）
　　小島篤博　（大阪府立大学）
　　五島敏芳　（京都大学）
　　後藤　真　（花園大学）
　　小森政嗣　（大阪電気通信大学）
　　柴山　守　（京都大学）
　　関野　樹　（総合地球環境学研究所）
　　高橋晴子　（大阪樟蔭女子大学）
　　中谷広正　（静岡大学）
　　八村広三郎（立命館大学）
　　原正一郎　（京都大学）
　　深海　悟　（大阪工業大学）
　　寶珍輝尚　（京都工芸繊維大学）
　　三宅真紀　（大阪大学）
　　村上征勝　（同志社大学）
　　森下淳也　（神戸大学）
　　師　茂樹　（花園大学）
　　米澤　剛　（大阪市立大学）

人文系データベース協議会 第１８回公開シンポジウム「人文科学とデータベース」

発行日　2012年12月22日
発行所　第１８回公開シンポジウム実行委員会
〒572-8530　大阪府寝屋川市初町１８－８
大阪電気通信大学 情報通信工学部 情報工学科
加藤常員（シンポジウム事務局）　Email:kato@ktlab.osakac.ac.jp