

目次

プログラム	ii
Session 1 画像・写真	1
小島篤博・青木茂樹・宮本貴朗「中尾佐助資料公開のための画像データベースの構築と運用」	1
守岡知彦「デジタル写真資料のWWWに基づく共有・管理手法について」	9
Session 2 文化遺産	19
近藤康久「2010年代の考古データベースはどう展開するか？」	19
師茂樹・明珍健二「京町家の3次元CG復元をめぐる問題」	29
Session 3 言語情報処理	37
山元啓史「ブーリアン演算による歌ことばモデルの解析」	37
小木曾智信・中村壮範「『現代日本語書き言葉均衡コーパス』のための形態論情報データベースについて」	45
福田智子・竹田正幸・南里一郎「『古今和歌六帖』の本文と表現—デジタル校本の試み—」	53
特別セッション「人文科学とデータベースの教育に関わる現状と課題」ラウンドテーブル	63

プログラム

開催概要

- 開催日 2010年11月27日（土）
- 会場 花園大学 拈花館202（京都市中京区西ノ京壺ノ内町8-1）
- 主催 第16回公開シンポジウム「人文科学とデータベース」実行委員会
- 後援 人文系データベース協議会
- 参加費 一般: 1,500円 学生: 500円

Session1 画像・写真（10:00～11:10）

- 小島篤博・青木茂樹・宮本貴朗（大阪府立大学総合教育研究機構）（10:00～10:35）
「中尾佐助資料公開のための画像データベースの構築と運用」
- 守岡知彦（京都大学人文科学研究所）（10:35～11:10）
「デジタル写真資料のWWWに基づく共有・管理手法について」

Session2 文化遺産（11:20～12:30）

- 近藤康久（東京大学総合研究博物館）（11:20～11:55）
「2010年代の考古データベースはどう展開するか？」
- 師茂樹・明珍健二（花園大学）（11:55～12:30）
「京町家の3次元CG復元をめぐる問題」

Session3 言語情報処理（13:45～15:30）

- 山元啓史（東京工業大学大学院社会理工学研究科）（13:45～14:20）
「プーリアン演算による歌ことばモデルの解析」
- 小木曾智信・中村壮範（人間文化研究機構 国立国語研究所）（14:20～14:55）
「『現代日本語書き言葉均衡コーパス』のための形態論情報データベースについて」
- 福田智子（同志社大学）・竹田正幸（九州大学）・南里一郎（立命館大学〔非常勤〕）（14:55～15:30）
「『古今和歌六帖』の本文と表現ーデジタル校本の試みー」

特別セッション「人文科学とデータベースの教育に関わる現状と課題」ラウンドテーブル

(15:45～17:30)

- 花園大学 情報歴史学
- 同志社大学 文化情報学
- 大谷大学 人文情報学
- 立命館大学 デジタル人文学
- 討論

(17:00～)

中尾佐助資料公開のための画像データベースの構築と運用 Development and Management of Nakao Sasuke Photo Database

小島 篤博 青木 茂樹 宮本 貴朗
Atsuhiko Kojima Shigeki Aoki Takao Miyamoto

大阪府立大学 総合教育研究機構, 大阪府堺市中区学園町 1-1
Faculty of Liberal Arts and Sciences, Osaka Prefecture University,
1-1 Gakuen-cho, Nakaku, Sakai, Osaka

あらまし: 大阪府立大学学術情報センターには、照葉樹林文化論の提唱者である中尾佐助の研究資料（中尾佐助コレクション）が保存されている。本研究では、これらのうち 1955 年から 1984 年までのフィールド調査により撮影された 22,000 枚のスライド画像をデータベース化し、Web による検索が可能なシステムとして構築した。本システムの特徴は、画像を検索・表示するだけでなく、ユーザが独自のキーワードやコメントを付与できるなど、関連分野の研究者に対する支援機能を提供することにより、学術資料としての活用範囲を拡大することを目的としている点である。システムの実装は、Apache、JSP、および PostgreSQL を利用した 3 階層システムとして構成されている。本稿では、画像データベースシステムの構築と運用状況について報告する。

Summary: Library and science information center of Osaka Prefecture University has a collection of slide films taken by the late Dr. Sasuke Nakao who was a broadcaster of *The Theory of Evergreen Forest Culture*. Among them are as many as 22,000 photographs taken in his field survey from 1955 till 1984. To make these valuable materials available to the public, we have developed a Web-based photo database with functions encouraging research activities. In this system, user can add extra keywords and comments on selected photos. The system is based on 3-tier system powered by Apache, JSP and PostgreSQL technologies. In this report, we present a case study of the development and management of the database.

キーワード: 中尾佐助, 画像データベース, 文化資料, 照葉樹林文化論

Keywords: Sasuke Nakao, photo database, cultural materials, theory of evergreen forest culture

1 はじめに

大阪府立大学学術情報センターには、照葉樹林文化論の提唱者として名高い故・中尾佐助大阪府立大学名誉教授の旧蔵書、雑誌、スライド、遠征アルバム、フィールドノート、スクラップブックなど膨大な研究資料（中尾佐助コレクション）が保存され、大阪府立大学の特色あるコレクションの一つとなっている。

照葉樹林文化論とは、東アジアの農耕文化の展開と周辺の文化要素の系譜を体系化して提唱された独創的な仮説であり、今日の日本や東アジアにおける基層文化の成り立ちを考察する基本的な理論となっている [1]。また、自然保護や生態系保全等の観点からも中尾資料を利用した研究が始められている。

現在、中尾資料の一部は公開されているものの、傷みや劣化が激しいものは嚴重に保管され常時一般公開することは難しい。しかしながら、これらのうち写真やスライド、フィールドノートなどについては、スキャナにより画像データとして読み取り、デジタルアーカイブとすることで広く公開することが可能である。

本研究は、中尾資料のうち 22,000 枚のスライドを、フィルムスキャナーにより取り込んだ画像をデータベース化し、インターネットを通じて画像を公開するとともに、関連分野の研究者に対する支援機能を提供することにより、学術資料としての活用範囲を拡大することを目的としている。

学術的・文化的な研究資料をデジタルアーカイブ化し、インターネット上に公開する試みは他にもなされ



図1 中尾佐助スライドデータベース フロントページ

ているが、これらの多くは参照のみを目的としたものである [2, 3]。一方、近年の Web 技術の発展により、SNS (Social Networkin Service) に代表されるような、利用者間のコミュニケーション支援ツールが容易に利用できるようになってきている。本研究の特徴は、こういったユーザ参加型の Web 技術を応用し、研究者が個々の画像ごとに独自のコメントを付与したり、相互参照したりする機能を有することである [4, 5, 6, 7]。これにより、単に静的な資料を提供するだけでなく、資料を核とした研究者間のコミュニケーションを支援するしくみを実現することができると考えている。

本稿では、この中尾資料スライド画像の検索システムの構築事例を報告する。以下、2 章では中尾佐助資料の概要について紹介し、続いて 3 章ではシステムの概要について説明する。4 章では、データベースの構成とシステムの具体的な機能について説明し、5 章で評価および検討を行う。最後に 6 章でまとめる。

2 中尾佐助資料について

中尾佐助コレクションは、旧蔵書 3,500 冊、雑誌 60 タイトル、著書 18 点、学術論文 88 点、新聞・雑誌等の記事 552 点、スライド約 28,000 点、遠征アルバム 77 冊及び写真ネガ、スクラップブック 94 冊、記録ノート 116 冊、フィールドノート 19 冊等からなり、中尾佐助コーナーにおいて閲覧に供している [8]。

このうち、スライドフィルムは 1955 年から 1984 年の海外調査において撮影されたものである。遠征アルバムには、1935 年頃から 1950 年代前半のカラースライドにするまでものが白黒写真で残されている。写真帳の写真の番号とネガの番号が併記しており、個々のスライドを特定できるようになっている。

大阪府立大学学術情報センターでは、これら中尾資料のデジタルアーカイブ化を進めている。まず、1999

年より第1期開発として、中尾が1958年にブータンのフィールド調査を行った際のスライド画像1,300枚を対象とし、個々のスライドにキーワードや場所、日時などを関連付けることで、キーワードのほか地理的、時間的条件から目的の画像を検索することができる画像データベースを構築し、公開してきた[4, 5]。

これに続く第2期開発として、2006年度より、残りのカラースライド画像についてもキーワード付与などの作業を行い、第1期公開分と合わせて約22,000枚の画像データベースとして2009年より公開している(図1)[6, 7]。本稿で説明する研究者支援機能は、この第2期において実装したものである。

さて、これらのスライドフィルムの状態としては、個々のスライドのマウントに、1枚ごとに以下のような識別番号が付与されている。この識別番号をもとに、フィルムごとに分類整理され、保管されている。

Nyy-nn-mm	
N:	Nakao
yy:	撮影年
nn:	フィルム番号
mm:	コマ番号

例えば、“N58-54-12”は、1958年に撮影された、54番目のフィルムの12コマ目となる。

この他、スライドの情報を得る手がかりとしては、中尾自身の著作である『秘境ブータン』[9]、『ヒマラヤの花』[10]、『ブータンの花』[11]があり、フィールドノート(3冊)とブータンのアルバム(12冊)が揃っていた。また、フィールドノートには、キャンプ地と標高、キャンプ地からキャンプ地への移動行程や天候等が明記されており、スライドを特定する際に有用な情報を得ることができた。

3 システムの概要

本システムで実装した機能は以下の通りである。

1. ユーザ登録・管理
2. 様々な条件によるスライドの検索
3. コメントやキーワードの付与
4. ブックマークの付与

まず、本システムではユーザごとに作業履歴等を保存するため、研究者の利用にあたってはユーザ登録

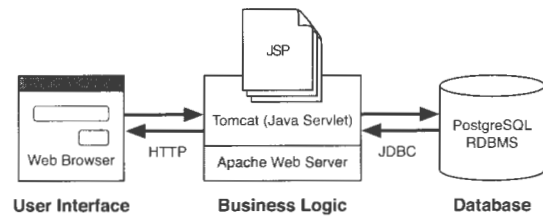


図2 3階層システム

をしていただいている。ユーザは、利用開始時にパスワードによる認証を行い、以後の作業はそのユーザに関連付けて記録される。なお、研究者専用機能を利用しない場合は、ユーザ認証なしで検索や画像表示などの基本的な閲覧機能が提供されている。

スライドの検索は、キーワード、日付、番号、および一覧からの選択のいずれかの方法により行うことができる。スライドを特定した後、必要があれば、そのスライドにコメントやブックマークを付与することができる。コメントは、スライドの内容について自由記述形式で記入することができ、公開の可否も選択可能である。また、ブックマークに登録しておけば、あとで簡単にアクセスすることができる。

以上のような要件を踏まえてシステム構成を検討したところ、Webブラウザをユーザ・インタフェースとして利用し、Webサーバ側でスライド検索やユーザ識別などの処理を行い、バックエンドとしてはRDBMSを利用するのが最適であると結論し、図2に示すような3階層型システム(3-tier system)として実現することとした。

これにより、一般ユーザからのスライド検索や、特定ユーザによるデータ編集作業などはWebインタフェースにより行い、そのためのビジネスロジックはWebサーバに実装し、さらに基本データ構造やそれらの関連性はRDBMSに格納することができる。

本研究では、WebサーバはApacheとTomcatにより実装し、動的なページ生成にはJSP(Java Server Pages)を利用している。JSPでは、具体的なサーバー側の処理をJava言語で記述できるため、Webからの要求に応じて動的にSQLを組み立てて、RDBMSにクエリを投入するなど、きめ細かな処理を行うことが可能である。

バックエンドのRDBMSとしてはPostgreSQLを採用しており、JDBC(Java DataBase Connectivity)を介してTomcat上のJavaと連携している。また、ODBC(Open DataBase Connectivity)を介してMicrosoft Ac-

表1 スキーマ一覧

スキーマ	内容
slide	スライドの基本情報
slide_set	スライドセット単位の情報
itinerary	フィールド調査の旅程情報
keyword	キーワード情報
description	文献のテキスト
slide_keyword	スライド=キーワード対応表
slide_description	スライド=テキスト対応表
person	ユーザの情報
comment	ユーザが付与したコメント
bookmark	ユーザが付与したブックマーク

cessなどのデータベース・フロントエンドと接続することで、容易にデータの編集が可能となっている。

4 システムの構築

4.1 データベースの設計

スライド画像をキーワードから検索するには、適切なキーワードをスライドに対応付ける必要がある。また、キーワードやコメントを付与した作業者を記録するためには、ユーザ情報についてもデータベースに格納する必要がある。本研究では、これらの項目を組み合わせ、1つの検索システムとして機能するようデータベースのスキーマ設計を行った。図3は本システムのスキーマの関連を表したE-R図である。また、各スキーマの一覧を表1に示す。

ここで、slideは個々のスライドの基本情報（撮影日時、場所、画像ファイル名など）、keywordはキーワード情報（キーワード、英表記、同義語へのリンクなど）、またpersonはユーザ情報（ユーザ名、パスワード、氏名、連絡先など）を保持している。

以下に、キーワードによるスライド検索、スライドへのキーワード、コメント、およびブックマークの付与について詳しく説明する。

4.2 キーワードによるスライド検索

キーワードによる検索は、ユーザが入力したキーワードと適合するレコードをkeywordテーブルから検索し、そこからslide_keywordをたどることで目的のslideを選択する。また、複数のキーワードを指定する場合は、それぞれ選択されたスライドの集合に対し、積集合や和集合をとることができる。

実際のキーワード入力画面を図4に示す。キーワー

図4 キーワード検索

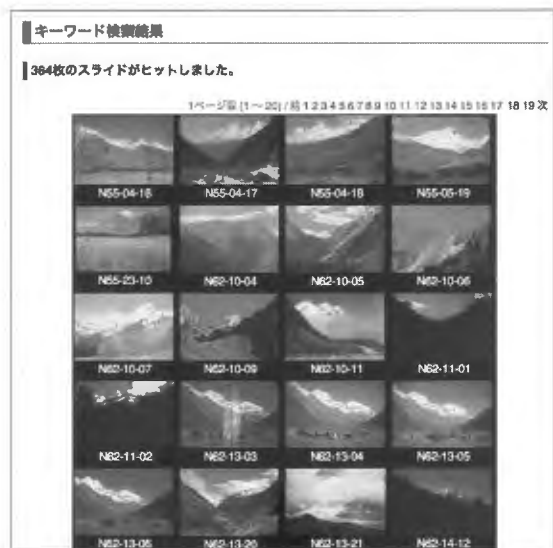


図5 検索結果

ドとして利用できる文字は、漢字、ひらがな、全角カタカナ、英数字である。また、個々のキーワードに対して、前方一致、後方一致、部分一致、完全一致、アルファベットの大文字小文字の区別も指定できるようになっている。

キーワードを指定して「検索」をクリックすると、該当するスライドのサムネイル画像が一覧表示される（図5）。さらにサムネイル画像を選んでクリックすると、スライド画像とともにキーワード、コメントなどの付随情報が表示される（図6）。

4.3 キーワードの付与

スライドにキーワードを付与するためには、まず対象のスライドを選択する。例えば、先述の検索により選択されたスライド画面（図6）の下には編集メニュー

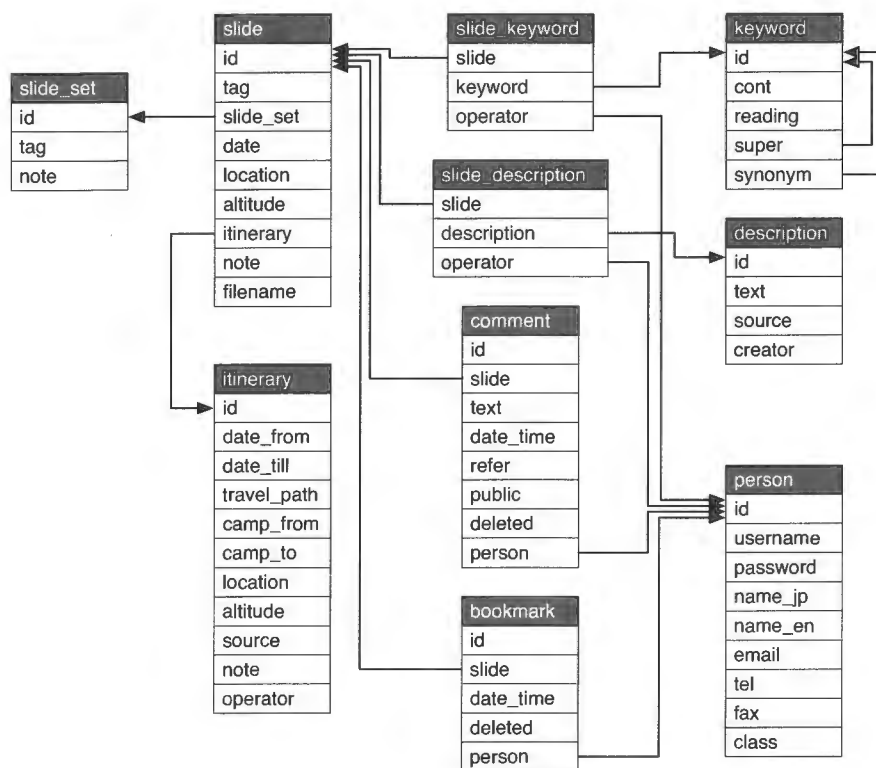


図3 データベースのE-R図(一部)



図6 選択されたスライド画像

が表示され、ブックマーク、キーワード、コメントを付与できるようになっている。ここで「キーワードを付与」というリンクをクリックすると、図7のようなキーワード入力画面が表示される。ここで新たなキーワードの追加と、既存のキーワードの削除ができる。なお、既存のキーワードは、付与者自身または管理者



図7 キーワードの付与

によってしか削除できないように制限されている。

新たに入力したキーワードが、既存の keyword のテーブルに存在しなければ、新たに keyword のレコードを追加する。同時に、slide.keyword のテーブルにレコードを追加し、それぞれ slide と keyword との関連付けを行う。

これにより、スライドに新たなキーワードが付与され、キーワードでの検索が可能となる。

図8 コメントの付与

4.4 コメントの付与

コメントの付与についても、キーワード同様、対象スライドを選択したのち編集メニューから行うことができる。「コメントを付与」をクリックすると、図8のようなコメント入力画面が表示される。コメント入力画面は、コメントを公開するかどうかのチェックボックスと、自由記述形式でコメントを入力できるテキストボックスから構成されている。それぞれ入力し、「送信」ボタンをクリックすれば、対象スライドにコメントを付与することができる。

コメントが付与されると、新たに comment テーブルにレコードが追加され、対象スライドの ID と付与者の ID がそれぞれ対応付けられる。同時に、コメントのテキストは text フィールドに、公開フラグは public フィールドにそれぞれ格納される。

図9はコメントを付与した後のスライドの表示画面であり、入力したコメントが付与者の氏名とともに表示されていることがわかる。

4.5 ブックマークの付与

ブックマークは、大量のスライド画像のうち、特に興味を引かれた画像に対して再びアクセスするのを容易にするために設けられた機能であり、スライド画像ごとに付与／削除が可能である。

編集メニューより「ブックマークを付与」をクリックすると、図10に示すようにスライド画像の上部に

図9 付与されたコメント

図10 ブックマーク

のアイコンが表示され、ブックマークが付与されていることがわかる。一旦ブックマークが付与されると、初期メニュー（図1）の「ブックマーク一覧」より、画像のサムネイルが表示され（図11）、次回以降は容易にスライド画像にアクセスすることができるようになる。

5 評価および検討

5.1 システム運用の現状

まず、現在のシステム運用状況について説明する。本システムは2009年2月28日より正式運用を開始している。登録ユーザ数は、2010年4月現在で94ユーザ、アクセス件数は、2010年2月21日～5月2日ま



図 11 ブックマークされた画像の一覧

での 10 週間で統計を取ったところ、週当たりの平均アクセス数 3420.6 回、ユニーク IP 数 380.6 であった。アクセス元ドメインでは、第 1 位が 'osakafu-u.ac.jp' (大阪府立大学) であり、第 2 位以下は 'msn.com'、'yahoo.net' 等の検索サイトが続いた。

これらは、サイトの情報を収集するロボットによるものと考えられるため、HTTP アクセス時の referer 情報をもとに、検索サイトから参照される際のキーワードについても分析してみた。その結果を図 12 に示す。第 1 位が「ブータン王国」であり、656 件中 148 件、22.6% であった。この他にも、多くはブータンに関するものであることがわかる。これは、第 1 期開発において対象としたのがブータンであったこと、また中尾の業績にもブータンに関連したものが多かったことが理由と考えられる。

続いて、システム運用開始から 2010 年 4 月 30 日までの間に、データベース検索に使用されたキーワードの統計を図 13 に示す。第 1 位は「ブータン」であり、全体で 757 件中 44 件、5.81% であった。また、1 位から 14 位まででキーワード全体の 1/4 を占めていることが分かった。一方、1 度しか使用されないユニークなキーワードは 324 件 (42.8%) であった。

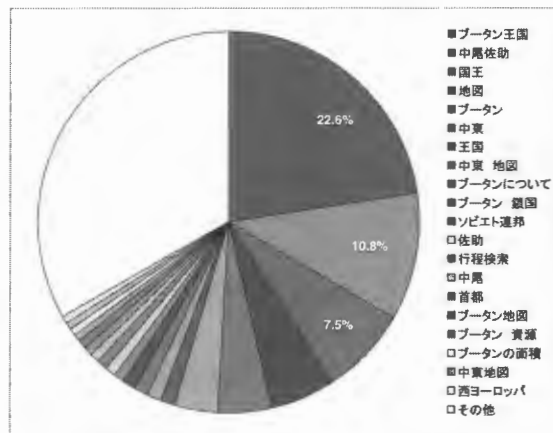


図 12 検索サイトからのキーワードの統計

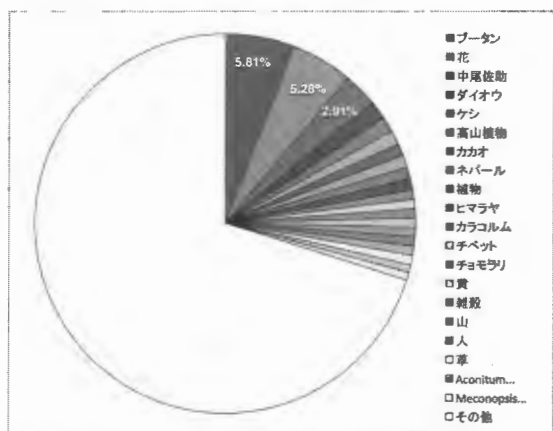


図 13 検索キーワードの統計

5.2 検討と今後の課題

システムの構築に際して、いくつかの点について検討する。3 章で述べたように、本システムで採用した 3 階層型システムは、データベースを利用した Web サイトとしては一般的な構成法であり、本システムにおいてもシステムの目的・用途に合致したものであった。Web のインターフェースは、各種操作のワークフローを検討した上で作成し、対象の画像を選択して操作するという、直感的で簡易な形で実現することができた。

また、データベース設計については、使用する各項目の関連性が明確であり、扱う項目に重複等もないことからシンプルな構成にまとめることができたと考えている。

今後の課題としては、一般的な SNS サイトにおいて標準的になっているメッセージや「友だち」機能など、ユーザ間の連携を高める機能を実装することが挙

げられる。また、ユーザによる使用感の調査などを通して、インタフェース上の問題点の洗い出しと改善も行っていきたい。

6 おわりに

文化資料のデジタルアーカイブ化、ならびに Web による公開が広く行われる中で、単にデジタル化した画像を整理して公開するだけでなく、資料を核とした研究者支援機能を独自に開発し、運用している大阪府立大学学術情報センターの取り組みを紹介した。本システムは、2000 年 4 月より公開しており、インターネットに接続されている端末からであれば Web Browser で閲覧、検索ができる。URL は以下の通りである。

<http://nakao-db.center.osakafu-u.ac.jp/>

今後の展開としては、照葉樹林文化論など様々な研究成果を集約する場として、本システムを実証的に提供していくことを考えている。

謝辞

本研究において、データベースの要件・機能等について有益なご助言いただいた山口裕文 大阪府立大学名誉教授、ならびにデータベース構築にあたって種々ご支援いただいた大阪府立大学学術情報課 赤崎久美氏、木下厚美氏に感謝の意を表します。

参考文献

- [1] 中尾佐助:『中尾佐助著作集 第 VI 巻 照葉樹林文化論』, 札幌, 北海道大学出版会, 2006.
- [2] 中村耕作: “学術調査写真の整理・公開の現状と課題 —大場磐雄博士写真資料整理をふまえて—”, 國學院大學学術フロンティア事業研究報告 人文科学と画像資料研究 第 2 集, Mar. 2005.
- [3] 宮崎紀子, 岡林隆敏, 下田研一: “『幕末明治期口本古写真超高精細画像データベース』の開発”, 第 40 回情報科学技術研究集会予稿集, pp.B52, 2003.
- [4] 富坂敏子, 小池利栄子, 小島篤博, 石井敬三, 宮本貴朗, 山野美賛子: “中尾佐助資料スライドデータベースの構築と利用者間情報共有への展望”, レコード・マネジメント, No.41, pp.8-18, Nov. 2000.
- [5] 小島篤博, 宮本貴朗, 富坂敏子, 小池利栄子, 石井敬三, 山野美賛子: “貴重文化資料デジタルアーカイ

ブの Web による検索・共有システムの構築”, 電子情報通信学会技術報告, ET2001-31, pp.61-68, Jul. 2001.

- [6] 小島篤博, 青木茂樹, 宮本貴朗: “文化資料保存・公開のための画像データベースの開発”, 電気学会情報システム研究会資料, IS-08, 23, pp.37-40, Sep. 2008.
- [7] 小島篤博, 青木茂樹, 宮本貴朗: “研究者支援機能を備えた写真画像データベースの開発”, 電気学会情報システム研究会資料, IS-10, 21, pp.1-6, May 2010.
- [8] 大阪府立大学総合情報センター編: 『中尾佐助 文献・資料総目: 照葉樹林文化論の源流』, 堺, 大阪府立大学総合情報センター, 1997.
- [9] 中尾佐助: 『秘境ブータン』, 東京, 毎日新聞社, 1959.
- [10] 中尾佐助: 『ヒマラヤの花』, 東京, 毎日新聞社, 1964.
- [11] 中尾佐助, 西岡京治: 『ブータンの花』, 東京, 朝日新聞社, 1984.

デジタル写真資料の WWW に基づく共有・管理手法について

How to share and maintain archives of
digital photographs on the WWW

守岡 知彦

MORIOKA, Tomohiko

京都大学人文科学研究所

〒 606-8265 京都市左京区北白川東小倉町 47

Institute for Research in Humanities, Kyoto University

47 Kitashirakawa-Higashioguramachi, Sakyo ward, Kyoto, 606-8265

概要

大量のデジタル写真資料を少ない手間で共有・整理・活用するための方策について検討する。複数人で情報を共有するには WWW ベースで行うのが良いが、このためにはバッチ処理で各種 RAW や JPEG 等から変換処理を行ったりメタデータを抽出する必要がある。また、撮影データではなく、より意味的な情報の付与・編集をどのように（なるべく手間をかけず）行うかも重要である。このために現在取り得る手段や見通しについて議論したい。

Abstract

In this paper, we will discuss how we should share/arrange/use a lot of digital photographs without a lot of time and labor. To share various information among plural members, it is good way to use WWW. However it requires batch processing for various digital photo data, such as conversion from various kind of “RAW format” to JPEG, and reading of various metadata from RAW data or JPEG/Exif. In addition, we don’t need only automatically recorded metadata, but also more semantic information: so it is important how we add/edit semantic annotations of photographs easily. In this paper, we will discuss available methods to treat RAW data and semantic annotations, and possible models and goals to develop archives of digital photographs.

キーワード： デジタル写真、RAW データ、Wiki、アノテーション、データベース

Keywords: digital photography, RAW data, Wiki, annotation, database

1 はじめに

デジタル写真は撮影や処理が容易で、ランニングコストも安く、低コストで大量の写真撮影することが可能である。このため、視覚的な情報を扱うような人文学・人文情報学のさまざまな分野で有用な道具となっている。しかしながら、低コストで大量の写真が撮影できてしまうということ

は、大量の写真データを扱わなければならないということを意味しており、データの管理や整理、処理等に大きな手間がかかり、せっかくのデータが死蔵され、うまく活用されにくいということにもなりかねない危険性もはらんでいるといえる。

こうした問題を解決するひとつの方法は、WWW を用いて複数人でデータを共有し、Wiki 的な手法でもって共同でデータを整理することではないかと思われる。しかしながら、デジタル写真の画像データの場合、テキストデータの場合とはまた別の問題がある。

ひとつは画素数の問題である。今の一般的なデジタル一眼レフカメラの画素数は、エントリーモデルも含めて、通常、1000 万画素以上であるが、XGA で 80 万画素弱、今の一般的なノートパソコン (WXGA) だと通常 100 万画素程度、一般的なデスクトップパソコン (full HD/WUXGA) でも 200 万画素程度であり、撮影されたデータの画素数に比べてはるかに少ない。

また、色の分解能も一般的なデジタル一眼レフカメラでは RGB それぞれについて 10~14bit 程度あるのに対し、一般的なパソコンではそれぞれ 8bit 以上を扱うことができなかったり困難だったりする。表現可能な色域に関しても、一般的なデジタル一眼レフカメラでは AdobeRGB (拡張色空間) が扱えるものが多いが、Windows + IE の場合、sRGB 以外の色空間を扱うことができない。標準的な画像フォーマットである JPEG では、色空間に関しては Exif や ICC プロファイルを用いることで AdobeRGB やその他各種色空間を用いることができるが、色の分解能に関しては 8bit に制限されている。このように、デジタルカメラやディスプレイ、プリンター等の入出力機器の性能向上に比べて、OS や WWW ブラウザー、標準的な画像配布用フォーマットの側の対応が遅れているのが実情であり、オリジナルデータそのままでは WWW 上で扱いづらく、利用者の環境に合わせて適切に変換してやる必要がある。それでも、色を適切に再現するのは必ずしも容易ではなく『色化け』が起こる。『色化け』は文字化けと違って気づきにくく、その点で問題である。

センサーの能力を十分に活かすためには JPEG では不十分であり、デジタル一眼レフカメラや一部のコンパクト・デジタルカメラでは『RAW 形式』と呼ばれるセンサーの出力を基本的にそのまま記録 (圧縮するとしても可逆圧縮を用い、非可逆圧縮は行わない。また、ベイヤー配列のようなセンサーの特性・形状に依存した記録を行い、ピクセルへの展開は行わないなど) する画像フォーマットで撮影した画像を保存することができるようになっている。ただ、この RAW 形式は基本的にカメラ毎に異なった形式になっており、互いに互換性がない。¹

ところで、デジタル写真には撮影時の情報が記録されており、データベースとして利用する上で有用である。しかしながら、こうしたメタデータの幾つかは Exif で標準化されているものの、非標準の形式で記録される情報もあり、Exif で標準化されているような情報であってもベンダー固有の情報として記録されたり、RAW 形式でないと記録されないという場合もある。

このように、カメラで記録された情報を全て利用しようとする RAW 形式で記録する必要があるが、そうすると相互運用性が低くなり、相互運用性を高めるために標準的な JPEG/Exif の範囲内で利用すると取り出せる情報が減ってしまうというジレンマに直面することになる訳である。WWW ベースで写真を管理しようとするなら、GNU/Linux ベースのサーバーなどでバッチ処理可能なプログラムが存在している必要があるが、ここではどの程度の情報が利用可能であるかについて簡単にまとめた。ただ、事例の詳細に関しては著者が保有している Nikon D700 と Canon EOS 5D の 2 機種に限定されてしまい、網羅性を欠くことをあらかじめお断りしておく。

一方、デジタル写真をデータベースとして活用する場合、カメラが自動的に記録したメタデータだけでは不十分であり、ある意図に基づいて整理を行ったり、再配置したり、加工した上で、なんらかのアノテーションを付与する必要があるといえる。カメラが自動的に記録したメタデータは基

¹Adobe は DNG という RAW 形式のための標準形式を提唱しているが、今の所採用しているカメラは少なく、普及しているとはいえないのが現状である。

本的に機材や焦点距離、露出、ホワイトバランス、撮影時刻、経度・緯度（GPS がある場合）等の撮影時（のカメラ）に関する情報であり、被写体や撮影者の意図、写真が意味するものなどについての情報はない訳である。こうした情報の幾つかはパターン認識技術等を用いることで抽出できたとしても、抽出できた情報にも誤りは含み得るし、無論、（ある程度の精度で）機械的に抽出することが難しいものもある。であるからこそ、人手による整理や編集が必要なのであり、そのために Wiki 的な手法を用いようという訳であるが、そのためには、何を対象にどういった種類の情報をどういう風に扱うかということについてモデル化する必要があるといえ、ここではこの問題に関しても議論したい。

2 写真をどうやって管理するか

2.1 日付と時刻の利用

写真を WWW ベースで管理するということは、オリジナルの写真やさまざまに加工された写真、写真にまつわるさまざまなメタデータ等に URI を付けることだという風にとらえることができる。URI はなるべく安定的で変化しないことが望ましいので、Wiki 的な手法で編集可能にする場合、なるべく変化しにくいものをベースに変化しやすいものを属性等として付与するのが良いといえる。こうした観点で考えれば、『撮影された写真』に対して固有の ID を付与し、その ID 自体は固定するのが良いといえる。

撮影者が 1 人、ないしは、カメラが 1 台であれば、この ID として撮影時刻を用いることが考えられる。撮影時刻は機種を問わず標準的に利用可能なメタデータのひとつであり、整理する上でも判りやすいといえる。ただ、通常、時刻の分解能は秒であり、ファイル名に記録されている番号等を併用する必要がある。

カメラが複数台あり、撮影者も複数人いる場合、カメラや撮影者に ID を振ることが考えられる。ただ、カメラの ID を標準的に利用可能な情報から得ることはあまり容易ではない。カメラのベンダーやモデルに関する情報は取れるのであるが、シリアル番号の記録の仕方はベンダー（カメラ）毎に異なっており、同じモデルのカメラが複数台あった場合にはうまく区別できないかも知れない。現実的には、撮影時刻とファイル名を併用すればそうやって作った文字列が衝突することはまれかも知れないが、原理的には衝突し得るので、あまり望ましい方法と言えない。こうしたことを鑑みれば、もし、一人の撮影者が同時に複数台のカメラで撮影することはないと仮定できるならば、撮影者の側に ID を振った方が無難であると思われる。これは写真のアップロードを撮影者が行うという風にワークフローを決めることで自然に実現できると考えられる。

人間にとっての判りやすさを考慮すれば、一人の撮影者が撮影したあるまとまった一連の撮影に対して、なんらかの判りやすいシンボルを付けておくとも良いかも知れない。それが不要であれば日付だけでも良い。シンボルを付与する場合でも、日付はあった方が良いと考えられる。

そうすると、結局、

ユーザー名 + 撮影日 + 付加シンボル + 撮影時刻 + ファイル名

という組で写真に固有の ID を付与することができる。

これは例えば、

`$prefix/<ユーザー名>/<撮影年>/<月>-<日>_<付加シンボル>/<撮影時刻><ファイル名>`

という階層的ファイル名で表現することができ、URL 化することができる。

例えば、2010 年 9 月 28 日 18 時 01 分 38 秒に撮影された写真 _DSC1249.NEF の場合、ユーザー名を “tomo”, 付加シンボルを “CIEAS” とした場合、

```
/opt/pages/tomo/2010/09-28_CIEAS/180138_DSC1249.NEF
```

というファイル名を付けることができ、サーバー名を photo.server とすれば、

```
http://photo.server/~tomo/2010/09-28_CIEAS/180138_DSC1249.NEF
```

というような URL に対応付けることができる。

2.2 『フィルムロール』

デジタル写真は大量の写真を低コストで容易に撮影可能であるために、撮影日単位に整理した場合、ひとつのフォルダー（ディレクトリー）に数百枚以上の写真が並んでしまうことも少なくない。しかしながら、写真の処理は時間がかかるので、ひとつのフォルダーの写真の数は 50~100 枚程度に押えた方が無難であるといえる。これは人間にとっての見やすさという点でも同様である。

幾つかの写真管理ソフト・現像ソフト等では、銀塩フィルム写真のアナロジーに基づいて、『フィルムロール』というような単位を導入しているものがある。すなわち、数十枚程度の一連の写真の集合をかつてのフィルムのようなものに見立てる訳である。

この一連の写真の集合はなるべくなら意味のある単位であって欲しいが、デジタル写真の場合、実際に 24 枚や 36 枚とかからなるフィルムを使っている訳ではないから、その切り方は恣意的にならざるを得ない。これは言い換えれば、安定的な情報となりにくいということを意味している。逆に、例えば、50 枚毎に機械的に分けた場合、ある一連の写真が別のフィルムロールに分かれてしまう恐れがある。人間にとっての判りやすさを考慮すれば、同じような写真やある一連の写真の繋がりはなるべく同じフィルムロールに入って欲しいといえる。とはいえ、人間にとっての意味を重視するあまり、何度もやり直したくなるような単位にすると、ID/URL として不安定になってしまうし、また、整理のための手間も増えてしまうといえる。

こうしたことを考えた場合、ある程度機械的に決定可能で、かつ、人間にとってある程度判りやすい分け方が良いといえる。問題はそのような分け方があるかということであるが、もし、短時間に連続して撮影された写真は似たようなものを撮っているか意味のある一連のシーケンスであると仮定できるならば、撮影時刻の間隔が一定以上（例えば、10 分とか 5 分とか 1 分とか）になっている箇所を『切れ目』と看做して分割するという手法が考えられる。この分割は段階的に行うことも可能である。

例えば、まず 1 時間で分割し、分割されたロール中の写真の数が 100 枚を越えていたら、各ロールを 30 分で分割する。それでも、ロール中の写真の数が 100 枚を越えていたら、次は 15 分で分割する、という風に、分割されたロール中の写真の枚数が一定数以下になるまで切れ目の間隔を短くするという方法が考えられる。あるいは、まず全ての間隔を調べておいて、適切な間隔を推測するという手法も考えられるだろう。ちなみに著者の場合、1 分程度で分割すると良い場合が多いようである。

分割されたフィルムロールはそこに含まれる最初に撮影された写真の撮影時刻（開始時刻）と最後に撮影された写真の撮影時刻（終了時刻）の組で管理するのが安定的でかつ判りやすいといえる。フィルムロールに人間にとって判りやすい名前を付ける場合、フィルムロールの属性として扱うのが良いと考えられる。

ただ、いずれにせよ、写真を同定するという観点でいえば、フィルムロールは冗長な単位であり、あるロール中の写真の数を押えるための中間的なものであるといえる。

2.3 『アルバム』

多数の写真を撮影した場合、通常、撮影した写真の全てを提示・展示するのではなく、ある意図に基づいて少数の写真を選び、それを必要に応じて加工して出力することになる。HDR (High Dynamic Range) 写真やパノラマ写真のように、複数の写真を合成して1枚の写真を作ることも考えられるが、いずれにせよ、オリジナルの写真の全てをそのまま出すのではなく、セレクトされた写真を加工して見せることになる訳である。

この時、セレクトされた写真の集合や加工された写真の集合といった、写真の集合が扱えると便利であり、幾つかの写真管理ソフト・現像ソフト等では、『アルバム』という機構が設けられていたりする。写真の集合の選び方は、人間による指定もありうるが、メタデータを用いて内包的に集合を定義する『スマートアルバム』(例えば、D700 で撮った写真の集合とか、2010 年 10 月 28 日に撮った写真の集合とか。複数の条件を and/or でつないだ複合的な式を書くことも可能である)という機構を備えるものもある。

アルバムは写真の集合であり、写真の ID の集合により表現できる。また、アルバム中の写真の順序関係が必要であれば、集合ではなく、列としても良い。いずれにせよ、写真の ID のリストとして表現可能である。

2.4 ブラケットティング

多くのデジタル一眼レフカメラでは、シャッターボタンを押した時に、3枚とか5枚とかの写真の露出を変えて短時間に連続的に撮影する機能を持っている。これを『(露出) ブラケットティング』と呼ぶことにする。例えば、3枚単位になっている場合、1枚目は標準露出、2枚目は暗めに、3枚目は明るめに撮るという具合である。² これは撮影時にきちんと露出を決めなくても、後で一番望ましい露出の写真を決めることができて便利であるし、複数の写真を合成した HDR 写真を作ることもできる。被写体の移動・変化やぶれを無視すれば、ブラケットティングして撮った3枚とか5枚の一連の写真は基本的に同じものを撮った写真の一部(同じ写真のバリエーション)といえ、関連付けて管理することが望ましい。しかしながら、ブラケットティングに関する情報の記録の仕方はメーカー(カメラ)毎に異なっているし、必ず設定した枚数全部連続して撮影されとは限らないので、撮影時刻等の情報を併用したヒューリスティクスを用いないと違う写真を関連付けてしまうことになる。

2.5 加工された写真の参照

写真を WWW 上で利用する場合、オリジナルの画像データをそのまま表示するだけでなく、画面サイズや用途に応じてリサイズしたり、一部をトリミングしたり、あるいは、露出やホワイトバランス・γカーブを変えるなどの現像処理を行うなど、何らかの加工を行ったものを表示したいことが少なくない。こうした加工の仕方を決め打ちではなく、なるべく柔軟に WWW 上で設定可能にし、かつ、その結果を WWW 上で共有し、他の WWW サービスとの連携を容易にするため

² ホワイトバランスなど他のパラメーターを変えて撮る機能を持っているカメラもある。

には、各種設定や状態を REST アプローチに基づいて扱うのが望ましいといえる。[5] [9] このことを鑑みれば、ベースとなる（オリジナルの画像データ）の URI に対して、加工方法を表すパラメーターを付けた URI を用いて参照するのが良いといえる。

3 RAW 形式を扱うためのツール

3.1 ExifTool

ExifTool [3] は画像のメタデータを読み書きするためのソフトウェアで Perl のライブラリとコマンドライン・アプリケーションを備えている。ExifTool は JPEG/Exif のみならず、さまざまなメーカーのデジタルカメラの RAW 形式を含むさまざまな画像形式・メタデータをサポートしている。

ExifTool はさまざまな画像形式・メタデータをサポートしているため、データ形式や取りたい情報によってプログラムを使い分ける必要がなく、単純化できるというメリットもあるが、より大きなメリットは Exif タグで記録されていない情報を RAW 形式の写真データから得ることができることであろう。

例えば、被写体までの距離の情報は被写体の位置を知る上で有用な情報であるが、Exif タグに記録するカメラはほとんどなく、Exif タグだけでは事実上抽出不可能である。[7] しかしながら、RAW 形式には記録されている場合がある。例えば、Nikon D700 に AF 対応レンズを付けた状態で撮影した RAW 形式のデータの場合、FocusDistance タグにこの情報が記録されており、³ ExifTool を使って読みだすことができる。

3.2 dcraw

dcraw [1] はデジタルカメラの RAW 形式を解釈し標準的な画像形式に変換（このことを『現像』と呼ぶ）するためのソフトウェアである。dcraw は C 言語で書かれており、Linux やその他 UNIX 系 OS で動く他、Windows でも利用可能である。dcraw 自体がサポートする出力形式は PGM/PPM/PAM 形式と TIFF 形式だけ⁴であるが、ImageMagic 等の他の画像処理プログラムと併用することで JPEG をはじめとするさまざまな画像形式に変換することができる。

4 写真のセマンティクスをどう扱うか

4.1 『～の写真』

写真を探すという行為は、ある目的を満たす写真を探すということだといえる。例えば、『土偶の写真』とか『明治時代初期の京都市内の写真』とか『1930 年台のファッションの写真』というのはその一例である。こうした写真の検索にはしばしば写真に付けられたアノテーションやメタデータが用いられるが、これはある意味において、写真を属性の束からなる概念のようなものとして捉える行為だと看做することができるだろう。写真は、本来、『この写真』という風にしか指し指せないものかも知れないが、しばしば、『～の写真』という風に指されるような『概念上の写真』（の

³但し、これは近似値であり、その誤差はレンズ等に依存し、レンズによっては著しく不正確な場合もある。

⁴標準では色の解像度は 8bit であるが、16bit で出力することもできる。

インスタンスのようなもの)として扱われる。これは『写真のセマンティクス』という風に言い換えることができるかも知れない。

属性の束で示されるような『概念上の写真』はそれに属する無数の写真の集合である。前述の『スマートアルバム』は有限の写真の集合に対して、これを実現したものと看做することができる。多数の同様な写真の中から1枚(ないしは少数)の写真を選び出すという行為は、この概念上の写真を示すための代表を選ぶ行為だと看做することができる。これは符号化文字(抽象文字)に対して例示字形を付けることに似ているかも知れない。

『概念上の写真』は指定の詳細さによって、その集合の包含関係を考えることができる。例えば、『土偶の写真』の中には『青森県の土偶の写真』があるし、場所や時代をもっと特定することでより詳細に指定することもできる。こうした属性の束によって指されるものとしての『概念上の写真』の性質は、同様に視覚的に表現される記号である文字(特に、漢字)の場合と共通する部分が多いといえ、素性の集合(属性の束)によって文字を表現する Chaon モデル [6] と同様な手法で表現できるといえる。

しかしながら、『写真のセマンティクス』は文字の場合とは異なる部分もある。文字の場合、通常、作者はいない⁵『概念上の文字』はその文字を解する不特定多数によって共有されており、多少の揺れはあったとしても、『作者の意図』や『受け手の解釈』の介在する部分は少ない。言い換えれば、『文字のセマンティクス』はそれをやりとりする解釈共同体の中に置かれているといえる。しかしながら、写真の場合は写真を撮った人がいるし、写真に撮られた被写体が(少なくともその写真が撮られた時には)存在したはずである。そして、これらは属性の束としての写真、すなわち、『概念上の写真』という枠組に収まり切らないような面がある。

写真資料のデータベース(あるいは、そのための写真のセマンティクス)というものを考えた場合、写真が持つこの2つの側面をうまく扱う必要があると思われる。

4.2 どのようなシステムを構築すべきか

『概念上の写真』は素性の集合(および、それによって表される『写真』間の関係等)によって表されるようなものとして捉えることができ、CHISE-Wiki [10] を一般化したようなもの⁶によって扱うことができると考えられる。⁷

一方、『撮られたものとしての写真』というものを扱う場合、ひとつには、撮影された写真データ自体をどう指し示すかという問題として捉えることができるといえ、2節で議論したような、オリジナルの写真データやそれを加工したものをどう URI 等で参照するかというような問題として扱うことができるかも知れない。他方、『作者の意図』や被写体(被写体が人間であった場合、『モデルの意図』も考えられるし、同様に、『メイキャップアーティストの意図』や『コーディネーターの意図』なども考えられる。被写体が人工物であれば、その作者の意図も考えられる)に関する情報の幾つかは、撮影データや被写体に関するメタデータ等から読み取れるかも知れないし、無論、知り得ない(言語化できないような)ものもあるだろう。いずれにせよ、写真自体やそのメタデータなどで表現し得ない(解析し得ない)ようなものは検索できない訳であり、もし検索の対象としたいのであれば、原理的に暗黙知的なものとならざるを得ないようなものであったとしても、なん

⁵歴史的・伝説上の作者がいる場合はあるが、通常、誰かの著作物とは看做されないし、そのことが文字のセマンティクスを規定したりしない。

⁶現在、CHISE-Wiki を一般化して、文字以外の任意のオブジェクトを対象とした EGT というシステムを開発中である。

⁷この『概念上の写真』はキャラクター(コス)写真を対象にした場合、よりはっきりと意識されるかも知れない。例えば、Cure [2] の場合、ジャンルや作品、キャラクターとそのバリエーションといった階層的なカテゴリが付けられており、指定したカテゴリに属する写真を串刺し検索することができるようになっている。

らかの概念化・記号化を行う必要があるといえる。これは言い換えるならば、『概念上の写真』として扱わなければならないということである。

こうした問題を鑑みれば、結局、2 節で議論したような写真データ自体のストレージとでもいうべきものと、素性の集合によって『概念上の写真』を表現する素性ストレージとでもいうべきものを組み合わせたようなシステムが望ましいのではないかと思われる。これはちょうど CHISE-Wiki と GlyphWiki [8] の関係に似たものという風に考えることができる。ただ、写真に対する視点の多重性は文字よりもはるかに多面的で複雑なものになり得るので、この点に配慮した設計を行うべきだろう。ただ、素性の集合という枠組は非常にプリミティブであるので、どのようなモデル化を行ったとしても、素性ストレージで扱えるようなものとなるのではないかと思われる。現実的には、Wiki 的な方法によって試行錯誤しながらモデルを改良していくということになるのではないかと考えられ、素性ストレージのフロントエンドとなるユーザーインターフェースはこうしたモデル化の試行錯誤を支援するような機能を持つことが望ましいといえる。

5 おわりに

デジタル写真資料を WWW ベースで共有・整理・活用する場合の課題や取り得る手段について議論した。

一般的なデジタル一眼レフカメラに比べて、現在の標準的な配布フォーマットである JPEG や現在の一般的なパソコンの表示能力は、解像度・画素数、色空間や色深度といった点ではるかに限定されており、センサーが捉え得る情報の全てを画面上で表現することができない。しかしながら、こうしたことはハードウェアに関する原理的な問題というよりは、多分にソフトウェア（OS や GUI, WWW ブラウザー等）や画像フォーマットの問題であり、AdobeRGB 等の広色域色空間をサポートするディスプレイは比較的安価に入手できるようになっているし、デジタルテレビなどの家電の世界では 10bit の色深度をサポートする製品も普及しつつある。200dpi を越えるような高解像度のディスプレイが一般的でないのも、GUI や WWW ページが 100dpi 程度の解像度を想定して作られており、解像度独立になっていないことが一因であると考えられる。今話題となっている電子書籍も、iPad や Kindle 等のリーダの解像度がそれほど高くないため、紙に比べてはるかに低解像度の環境を想定して作られてしまう恐れがあるように思われる。しかしながら、人文情報系データベースの寿命や制作のためのコストを考えれば、より長いスパンで考えた方がよいのではないかと考えられる。こうしたことを鑑みれば、現在利用可能な機材を最大限に活かしたデータ化を検討すべきであり、デジタル写真においてはマスターデータとして RAW 形式を基本とするのが望ましい。しかしながら、RAW 形式はペンダー・カメラ毎に形式が異なり、互いに互換性がなく、情報交換の点で難がある。しかしながら、この問題は、現像に関しては dcrw, メタデータに関しては ExifTool を用いることでほぼ解決するといえる。

（Wiki 的な手法を用いて）WWW 上で写真を扱う上で重要なことのひとつは Web API の設計であるが、別の WWW サービスからの簡単に利用するためには REST アプローチを採用するのが望ましいといえる。このことは、言い換えれば、写真に対してどのように URI を対応付けるかということであるが、基本的には撮影者と日付とオリジナルのファイル名の組合せを用いることで写真の ID や URI を作ることができる。そして、撮影（イベント）や『フィルムロール』といった単位で写真のマスターデータをシステムティックに管理することができる。

一方、写真のセマンティクスを考えた場合、こうした『撮影された写真』とは別に、素性（メタデータやアノテーション等）の集合で指定されるような『概念上の写真』というものを考える必要

があるといえる。写真の整理や利用とは『撮影された写真』と『概念上の写真』を対応付ける作業に他ならず、また、写真の検索とは『概念上の写真』に対応する『撮影された写真』を捜し出す行為と看做することができる。『概念上の写真』の持つ性質は (Chaon モデルにおける) 文字 (特に漢字) と同様な部分が少なからずあり、共通の枠組で扱えるのではないかと考えられる。しかしながら、写真にはこうした視覚的記号としての側面の他にも、『撮影した時に存在した被写体を写したものの』という側面や (作者がいて読者がいるような) 『作品』としての側面があり、こうしたことを鑑みれば、『概念上の写真』という視点だけでは明らかに不十分である。⁸しかしながら、コンピューターに載せようとする限り、形式化・機械可読化が必要であるのも事実である。よって、素性の集合やオブジェクト間の関係という枠組を用いて、こうした多面性や暗黙知的・感覚的なものの明示化・モデル化を行う必要があるといえる。このためには Wiki 的な手法の利用は有用ではないかと考えられる。

こうした観点に基づき、現在、『撮影された写真』のストレージ用のソフトウェア、および、CHISE-Wiki を一般化した素性ストレージである EgT というソフトウェアを開発中であり、この両者を統合した環境の実現を目指している。前者は現在の所、リサイズ機能と iPhoto からの変換用ユーティリティを備えるだけの簡単なものであるが、将来的には RAW 形式からの現像機能やトリミング等の簡単な加工機能と一括ダウンロード機能も追加する予定であり、ローカルとの同期機能の実現も検討している。

参考文献

- [1] Dave Coffin. Decoding raw digital photos in linux. <http://www.cybercom.net/~dcoffin/dcraw/>.
- [2] Cure. <http://ja.curecos.com/>.
- [3] Phil Harvey. Exiftool. <http://www.sno.phy.queensu.ca/~phil/exiftool/>.
- [4] ロラン バルト (Roland Barthes). 明るい部屋一写真についての覚書 (原題: La Chambre claire : Note sur la photographie) . みすず書房, 1997 (原書: 1980) 年. 花輪 光 (訳).
- [5] Henry S. Thompson. Web Architecture and Naming for Knowledge Resources. In Takenobu Tokunaga and Antonio Ortega, editors, *Large-Scale Knowledge Resources*, Vol. 4938 of *LNAI*, pp. 334–343. Springer, 2008 年.
- [6] MORIOKA Tomohiko. CHISE: Character Processing Based on Character Ontology. In Takenobu Tokunaga and Antonio Ortega, editors, *Large-Scale Knowledge Resources*, Vol. 4938 of *LNAI*, pp. 148–162. Springer, 2008 年.
- [7] 田島孝治, 高田智和. 「景観文字調査」のための調査結果分類方法に関する一考察. 情処研報, Vol. 2010-CH-87, No. 6, pp. 1–6, 2010 年 7 月.
- [8] 上地宏一. GlyphWiki. <http://glyphwiki.org/wiki/GlyphWiki>.

⁸ロラン・バルトは「明るい部屋」[4]の中で、『ストウディウム』と『プンクトウム』という概念を用いてこうした問題を指摘している。

- [9] 守岡知彦. データを生み出すデータのために. 人文科学とコンピュータシンポジウム論文集『サービス指向のデジタル技術へ～人文科学のポテンシャルへ』, pp. 13-18. 情報処理学会, 2008 年 12 月.
- [10] 守岡知彦. CHISE のセマンティック wiki 化の試み. 情処研報, Vol. 2010-CH-87, No. 8, pp. 1-8, 2010 年 7 月.

2010 年代の考古データベースはどう展開するか？ How Far Archaeological Databases Go in the 2010s?

近藤 康久

Yasuhisa Kondo

東京大学 総合研究博物館, 東京都文京区本郷 7-3-1

The University Museum, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo

あらまし: 昨今、考古学の研究プロジェクトは大型化と学際化が進んでおり、大規模かつ多様なデータを短期間に集約し、解析することが求められている。その一方で、コンピュータの性能向上やインターネットの高速化はめざましく、考古データベースに求められる役割(やらなければならないこと)とデータベースのポテンシャル(できること)は、近年大きく変貌している。そこで本稿では、国内外における 1990 年代以降の考古データベースの研究展開を整理した上で、文献データベースや GIS(地理情報システム)との連携、ネットワーク・コンピューティング、大規模データマイニングを基軸とする 2010 年代の考古データベースのあり方について提言したい。ケーススタディとして、自身が設計・運用に主体的にたずさわっているプロジェクトから、2000 年代型の大規模 GIS リレーショナルデータベースである縄文時代錘具データベース(JOKERBASE)と、2010 年代の考古データベースを特徴づける文献連携・ネットワーク接続型の「旧人・新人交替劇」関連遺跡データベース(Neander DB)の事例を紹介する。

Summary: Recent large, multidisciplinary projects in archaeology have gradually been headed to collect and analyze a large amount of the complicated scientific data in a relatively short term. On the other hand, high performance servers and the high speed Internet have been made available for archaeological computing. Therefore, the role (what to do) and potential (what is able to do) of archaeological databases are dramatically changing. After giving an overview on the development of archaeological database in the past decades, this paper will suggest a perspective to the archaeological databases in the 2010s, that will be characterized by the effective combination of (1) bibliographical information systems and geographic information systems (GIS), (2) network computing, and (3) large-scale data mining. For this purpose, this paper presents an old-and-new couple of the author's case studies: the Jomon Net-sinker Database (JOKERBASE) as a typical GIS-based large-scale relational database in the 2000s and the network-based Neanderthal/Modern Human Transitional Site Database (Neander DB) featured by bibliographical liaison, as a potential standard in the 2010s.

キーワード: 考古学, リレーショナルデータベース管理システム(RDBMS), 地理情報システム(GIS), ネットワーク・コンピューティング, 大規模データマイニング

Keywords: archaeology, Relational Database Management System (RDBMS), Geographical Information Systems (GIS), network computing, large-scale data mining

1. はじめに

科学のあらゆる分野の例にもれず、考古学においても昨今は研究プロジェクトの大型化と学際化が進んでいる。それに伴い、審査の透明化と厳格化(中間評価の導入など)、あるいは高い目標設定などを背景として、大規模で多様なデータを短期間に集約し、解析することが求められるようになった。それは、プロジェクトの目的に即した高性能データベースシステムを導入する必要性が高まっていることを意味する。

いっぽう、昨今はマルチコアプロセッサや大容量ディ

スクストレージの普及(低価格化)に伴うパーソナルコンピュータの性能向上がめざましく、通信手段としてのインターネットの高速化(ギガビット級)も進んでいる。そのため、考古学のデータベースに求められる役割(やらなければならないこと)とデータベースのポテンシャル(できること)の両方は、いま大きく変容している。そこで本稿では、国内外におけるこれまでの考古データベースの研究展開を整理した上で、2010 年代の考古データベースのあり方について、自身が設計・運用に主体的にたずさわっている縄文時代錘具データベ

ース(JOKERBASE)と「旧人・新人交替劇」関連遺跡データベース(Neander DB)の例を挙げつつ提言したい。

2. 考古データベース 1990-2010

データベースの高性能化と普及は、コンピュータのそれと軌を一にする。考古学におけるデータベースの利活用については、我が国では日本情報考古学会(1995年発足、前身の「考古学におけるパーソナルコンピュータ利用の現状」研究会と「考古学における計量分析」研究会は1970年代半ばまでさかのぼる。<http://www.archaeo-info.org/>)や情報処理学会人文科学とコンピュータ研究会(じんもんこん、1989年発足、<http://www.jinmoncom.jp/>)、人文系データベース協議会(1999年発足、<http://www.osakac.ac.jp/jinbun-db/>)、欧米では国際学会CAA(Computer Applications and Quantitative Methods in Archaeology, 1973年発足、<http://www.caaconference.org/>)を主たる舞台として研究発表・研究交流が進められてきた。そこで、考古データベースの研究経過を追うために、人文系データベース協議会主催の公開シンポジウム「人文科学とデータベース」とCAAの論文集を中心に、1990年代以降のデータベースに関係する論文を年代順にリストアップした(文献リスト参照)。もちろん、遺漏は多々あるだろうが、考古データベースに継続的に関わっている主な研究者の業績を網羅しているので、研究動向の全体像を見損なうものではないと考える。

考古学にコンピュータ・データベースを利用するというコンセプトは、1980年代にはすでに確立していた(小沢1985)。1991年のCAAで、考古データベースの特集が組まれたことは注目に値する。この時点ですでに、一意識別子を用いて複数の属性テーブルを相互に関連づけるRDBMS(Cheetham 1992; Ryan 1992)や考古学のための文献データベース(Heyworth 1992)の実装例が提示されている。そして、1995年前後を境として、GISと連係する空間情報データベースの実装がはじまる(小林ほか1995; 宝珍ほか1997; 横山・千葉1997)。2000年以降、データベースの収録対象が大規模化し、全国規模の遺跡データベース構築例が見られるようになる(Farcic et al. 1999; Berg 2007; McKeague and Jones 2008; 森本2000, 2001, 2006)。また、個別の考古学研究プロジェクトにおけるデータベースの活用も本格化する(Kadar 1999; Bampton and Mosher 2001; Tsakiraris 2001; 西上・小沢2001)。と同時に、複数の異なるデータベースを統合・集約するためのクリアリングハウス(碓井ほか2000; 河野ほか2002)や自律分散型WebGIS(森2002, 2006)、またデ

ータベース内検索エンジン(三浦・小沢2000)を開発・提供する試みも見られるようになる。

2004年頃を境に、技術・設計フェーズと応用フェーズの両方において、研究対象の多様化が進む。たとえば、技術・設計段階においては、XML・UMLと地理情報標準を用いて考古学が扱う情報全体の構造化を指向するもの(藤本2007)や、考古学の扱う多様な属性情報の集約・階層化をめざすもの(宝珍2004, 2006, 2009)、あるいはGISと連動したRDBMS(松森ほか2009)などが見受けられる。いっぽう、具体的なデータベース対象としては、たとえば国内では旧石器遺跡(日本旧石器学会編2010)、縄文遺跡(横山・千葉1997; 近藤ほか2007)、弥生高地性集落(加藤1998)、古墳(西上・小沢2001; 北條2004; 出田2007; 中村2007; 寺村2008)、地震遺跡(河野2006)などを挙げることができる。世界に視野を広げると、極東旧石器遺跡(Gillam et al. 2009)、イベリア半島旧石器洞穴遺跡(González Sainz and Chacho Toca 1999)、上メソポタミア先史遺跡(Anastasio et al. 2004)をはじめ、枚挙にいとまがない。多くのデータベースは、研究対象の集成と、紙媒体もしくは記憶媒体による集成成果の出版・配布を目的とするが、中には、奈良文化財研究所の全国遺跡データベース(森本2000, 2001, 2006、<http://www.nabunken.go.jp/database/>)に代表されるように、インターネットでの公開・情報提供を目的とするものもある。このように見てくると、ひとくちに考古データベースといっても、その構造や指向性はデータベースを導入する目的によってずいぶん異なり、各々の目的に特化・最適化した設計がなされていることが分かる。

3. 縄文時代 鍾具 データベース JOKERBASE: 2000年代大規模 GIS リレーショナルデータベースの一到達点

東京大学考古学研究室では、1990年代後半より縄文時代遺跡ならびに鍾具(土鍾・石鍾)の全国的な集成作業を推進しており、2005年以降、筆者を主担当者として、GISデータベースへの統合・解析を進めてきた(近藤ほか2007)。その目的は、過去数十年間の発掘調査(報告書)の蓄積に情報処理と分析が追いつかず、全体像の理解が停滞している縄文時代の人間活動の空間的広がり(ここでは土鍾・石鍾を用いる網漁業)を、定量的な全数調査を通じて「見える化」し、再評価することにある。いうなれば、大規模データマイニングによって未知の考古学的パターンを発見しようという試みである。

この研究目的を実現するために、研究チームではGISとの接続を前提とするRDBMSを導入した。この、

縄文時代錘具データベース「JOKERBASE」(Jomon SinKER DataBASE)は、(1)遺跡テーブル、(2)遺物(錘具)テーブル、(3)集計テーブル、(4)遺跡・遺物の記載を正規化するための補助テーブル、(5)遺跡フィーチャという5つの要素から構成される(図 1)。実際のデータベースは Microsoft Excel で作成した個別のテーブルを Microsoft Access 2007 を用いて ESRI ArcGIS のパーソナルジオデータベース(.mdb)に変換・集約する形で管理している。

データベースの基幹部分にあたる遺跡テーブルと遺物テーブル、集計テーブル、遺跡フィーチャはすべて、遺跡ごとに割り当てた一意の識別子(site_id)によって関連づけられる。各々のテーブルを概念的な実体(エンティティ)に見立て、実体間の関係性(リレーションシップ)を観察すると、「錘具」は必ずいずれかの「遺跡」に帰属するという依存関係があり、「遺跡」と「錘具」は「1」対「0以上多」という数値的結合関係(カーディナリティ)にある。これは、ある1か所の遺跡から、錘具の出土が1点報告されることもあれば、2点以上報告されることもあれば、報告されない(0点の場合もあるということ)を意味する。理想的には「遺跡」にはただ1つの遺跡フィーチャが対応する。すなわち、「遺跡」と「遺跡ポリゴン」、「遺跡」と「遺跡ポイント」はそれぞれ「1」対「1」の関係にあるはずだが、現実には消滅等の事情により位置を特定できない遺跡があるので、「1」対「0または1」となる。また、各々の「錘具」は必ずどこかの空間単位すなわち「コンテキスト」から出土し、いずれかの「時期」に帰属し、何らかの「型式」に分類される。時空間分析にあたっては、コンテキスト・時期・型式のほか、寸法や重量といった「錘具」の属性を、さまざまな組み合わせ・方法でクロス集計し、その結果を「遺跡ポイント」と仮想的に結合(テーブル結合)することによって、GIS上に表現する。なお、各テーブルおよび属性の詳細な仕様と解析結果(図 2)については紙幅の都合で割愛するので、近藤 2010 を参照されたい。

4. 2010 年代の課題と解決策

JOKERBASE をはじめ、現行のさまざまな考古データベースは、各々の目的に特化・最適化した設計がなされているとはいえ、ほぼ例外なく一つの共通点をもつ。それは、データベースの構築・編集作業が非公開かつオフラインで進められるという点である。

ところが、冒頭で述べたように、昨今は研究を取り巻く環境が変化し、隣接分野も含めた国内外の研究機関・研究者との協働と研究成果のすみやかな取りまとめ・公表を要するようになってきた。しかるに、従来のオフライン方式でこのような要請に対応するには限界

がある。たとえば、各担当者がオフラインで作成したレコードを集約する方法を取ると、データ項目や分類・表記・入力規則にゆらぎが生じたり、マスタデータベースと各担当者が保存するローカルデータベースの間でバージョンの齟齬が生じたりすることが懸念される。また、上記の理由から、ひとたびデータベースのひな形を配布したあとに、仕様・設計を変更するのが非常に面倒になる。実際、JOKERBASE においても、そのような事態が出来た。

これらの問題を解決する方策としては、ネットワークの活用が考えられる。すなわち、共通のマスタデータベースに複数の担当者が同時にアクセスできる仕組みを調えることによって、協働してデータを入力するのはもとより、作業の中で生じた設計変更や拡張のニーズにも柔軟に対応できるようになるという着想に至る。

しかし同時に、研究プロジェクトには機密性も必要である。ネットワークを利用する際には、情報漏洩やハッキングを防ぐために内部統制・セキュリティ対策を講じなければならない。したがって、Wikipedia のようにすべての編集作業を原則公開で行なうというわけにはいかず、パスワードと暗号化によって適切なセキュリティを設定する必要がある。

5. 「旧人・新人交替劇」関連遺跡データベース Neander DB: 文献情報を統合した次世代ネットワーク・データベース

文部科学省科学研究費補助金(新学術領域研究)「ネアンデルタールとサピエンス交替劇の真相: 学習能力の進化に基づく実証的研究」(領域代表者・赤澤威・高知工科大学教授 <http://www.koutaigeki.org/>)プロジェクトの計画研究 A01「考古資料に基づく旧人・新人の学習行動の実証的研究」(研究代表者・西秋良宏・東京大学総合研究博物館教授)では、ネアンデルタール人(旧人)と現生人類(新人)の交替期(およそ 20 万年前~2 万年前)の人類遺跡を広域的・網羅的に集成することを計画している。集成作業は東京大学・名古屋大学・東北大学・北海道大学・明治大学の研究者が担当地域を分担する形で実施される。

そこで、研究分担者がインターネットを介してマスタデータベースに直接アクセスし、編集できるような仕組みとして、FileMaker Server 11(サーバソフトウェア)と FileMaker Pro 11(クライアントソフトウェア)からなるネットワーク・データベースシステムを導入した。データベースサーバは東京大学総合研究博物館に設置し、ファイアウォールとパスワードによって保護する。

遺跡の登録作業は、発掘報告書から遺跡の名称・経緯度等の基本情報を抽出した上で、文化層ごとの属

性情報を記載し、さらに年代測定値があれば付記するというフローからなる。そこで、このワークフローに最適化する形で RDBMS を設計した(図 3)。基本をなすのは文献書誌情報テーブルと遺跡基本情報テーブルであるが、両者は多対多の関係にあるため、引用関係テーブルを中間にはさんで両者の関係を整理する。また、遺跡基本情報テーブルには遺跡詳細情報としての文化層テーブルを「1」対「1以上多」で関連づけ、年代測定値がある場合はさらに文化層テーブルに年代情報テーブルを関連づける。

この Neander DB はレコードの入力・検索・閲覧・出力機能に特化したものであるが(図 4-5)、将来的に石器製作伝統の Wiki 型データベースと連係させることと、GIS と連係させて人類遺跡の時空間分布を地図化することを計画している。また、作業上の工夫として、経緯度が不明な遺跡については、分担者が遺跡位置図をアップロードし、GIS とリモートセンシングを専門とする担当者が経緯度を算出する体制をとる。これは、ネットワーク・データベースの導入によりはじめて可能になった協働体制である。これにより、データベース構築にかかる時間を大幅に短縮でき、その分プロジェクトの主眼である旧人・新人の時空間分布の把握と学習行動の違いに対する理解を深めることができるものと期待される。

6. 2010 年代の考古データベース展望

Neander DB の最大の特色は、ネットワーク・コンピューティングの活用により、セキュリティを確保しつつ遠隔地に拠点を置く複数の研究者と協働して迅速にデータベースを構築できる点であるが、くわえて文献書誌情報と遺跡情報が関連づけられることにも注目したい。従来、国立情報学研究所の NACSIS-Webcat に代表される図書館系情報システムと、奈良文化財研究所の全国遺跡データベースに代表される文化財系情報システム、さらにいえば列品管理を目的とする博物館系情報システムは、それぞれ別個独立したシステムとして運用されてきた。しかし近年は、自律分散型アーキテクチャ(森 2006)や横断検索システムの実装により、図書館系と博物館・文化財系の情報システムを統合することが可能になりつつある。その意味において Neander DB は、文献情報と考古情報が連係した、次世代のネットワーク型考古データベースの特徴をしめすものといえよう。まとめると、2010 年代の考古データベースは、ネットワーク・コンピューティングを活用して、セキュリティを確保しつつ、文献書誌情報や Wiki、GIS を自在に統合する柔軟な拡張性をもって、大規模なデータマイニングに対応するものとなるであろう。

謝辞

本稿で述べた展望については、Archaeo-GIS Ustream “友引 Night!!”(http://ustre.am/fAyw)における阿児雄之氏(東京工業大学百年記念館)ならびに視聴者諸氏との議論から大いに刺激を受けている。縄文時代錘具データベースは、平成 20 年度日本学術振興会科学研究費補助金(特別研究員奨励費)による成果の一部である。また、旧人・新人交替劇データベースは、平成 22 年度文部科学省科学研究費補助金(新学術領域研究)「ネアンデルタールとサピエンス交替劇の真相」計画研究 A01(前掲)による成果の一部である。

文献

※ 研究経過を明示するため、年次順に配列する。

※ 略称:CAA = Proceedings of the Annual Meeting of Computer Applications and Quantitative Methods in Archaeology (http://www.caaconference.org/). 人文 DB 論文集 x = 第 x 回公開シンポジウム「人文科学とデータベース」論文集

1995 年以前

- [1] 小沢一雅(1985)「考古学研究支援型データベースの構成」情報処理学会論文誌 26:936-945.
- [2] Lewis, P. H. and Goodson, K. J. (1991) “Images, databases and edge detection for archaeological object drawings”. CAA 1990:149-154.
- [3] Arroyo-Bishop, D. (1992) “Further structuring of the ArchéoDATA system”. CAA 1991:89-94.
- [4] Barceló, J. A. (1992) “Programming an intelligent database in hypertext”. CAA 1991:21-28.
- [5] Cheetham, P. N. and Haigh, J. G. B. (1992) “The archaeological database: new relations?” CAA 1991:7-14.
- [6] Heyworth, M. (1992) “The British Archaeological Bibliography: a fully-computerised service for archaeology”. CAA 1991:15-20.
- [7] Ryan, N. (1992) “Beyond the relational database: managing the variety and complexity of archaeological data”. CAA 1991:1-6.
- [8] Murray, D. M. (1995) “The management of archaeological information: a strategy”. CAA 1993:83-88.
- [9] Bullas, S. G. (1995) “ID-MARGARY: an inference database for the mapping, recognition and generation of ancient roads and trackways”. CAA 1994:133-136.
- [10] Rains, M. J. (1995) “Towards a computerised

desktop: the integrated archaeological database system”. *CAA* 1994:207-210.

[11] 小林 努・加藤常員・小沢一雅(1995)「4次元歴史空間システムにおける地理情報処理について」人文 DB 論文集 1:13-18.

1996 年

[12] 宝珍輝尚・中田 充・白井治彦・都司達夫「考古学のためのデータベースシステム」人文 DB 論文集 2:45-54.

1997 年

[13] 及川昭文「考古学データベース:過去を復元するマルチメディア技術」情報処理 38/5:388-391.

[14] 宝珍輝尚・都司達夫・河合秀夫「地理情報を利用した遺物データベースシステムについて」人文 DB 論文集 3:31-38.

[15] 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 3/2:29-40.

1998 年

[16] 加藤常員「高地性集落データベースからみた弥生時代の情報通信」人文 DB 論文集 4:35-42.

1999 年

[17] Bekiari, Ch., Gritzapi, Ch. and Kalomoirakis, D. “POLEMON: a federated database management system for the documentation, management and promotion of cultural heritage”. *CAA* 1998:317-330.

[18] Farcic, V., Benusi, I., Bondzic, N. and Jovanovic, M. “The archaeological database of Serbia”. *CAA* 1998:331-334.

[19] Kadar, M. “Distribution of copper ores in the Carpathians: Data management with relational databases”. *CAA* 1998:345-348.

[20] Schloen, D. “Database design and the electronic publication of archaeological information”. *CAA* 1998:199-202.

[21] González Sainz, C. and Cacho Toca, R. “Resent research on Paleolithic arts in Europe and the multimedia database”. 人文 DB 論文集 5:11-22.

2000 年

[22] 碓井照子・森本 晋・泉 拓良・清水啓治・中 雅明・松村 寛・高瀬啓司・鈴木繭子・水野行子・野田卓・関口靖之「考古学データベースにおける考古学メタデータとクリアリングハウス」地理情報システム学会講演論文集 9:99-104.

[23] 三浦宙明・小沢一雅「考古学データベースにおける検索エンジンの研究」人文 DB 論文集 6:73-80.

[24] 森本 晋「全国遺跡データベースの構築」人文 DB 論文集 6:53-62.

2001 年

[25] Bampton, M. and Mosher, R. “A GIS driven Regional database of archaeological resources for research and CRM in Casco Bay, Maine”. *CAA* 2000:139-142.

[26] Madsen, T. “Transforming diversity into uniformity: Experiments with meta-structures for database recording”. *CAA* 2000:101-106.

[27] Niccolucci, F., D'Andrea, A. and Crescioli, M. “Archaeological applications of fuzzy databases”. *CAA* 2000:107-116.

[28] Tsakirakis, V. G. “Using a Relational Database Management System for the recording of ancient settlements and sites in the Vrachneika Territory in Western Greece”. *CAA* 2000:143-148.

[29] 西上昌治・小沢一雅「前方後円墳データベース検索システムと地理情報処理」人文 DB 論文集 7:73-80.

[30] 森本 晋「全国遺跡データベースの構築 2001 年度の動向」人文 DB 論文集 7:65-68.

2002 年

[31] Couboud, P. “Creating of a database for prehistoric sites: which are the goals, the strategy and what means to put in place?” *CAA* 2001.

[32] Hughes, E. A., Brown, A., Smith, G. R., Marshall, V. A. and Pantos, E. “A web-fronted database of diffraction patterns of archaeomaterials”. In Kiliglou, V. et al. (eds.) *Modern trends in scientific studies on ancient ceramics*. BAR Int. Ser. 1011, pp.85-94.

[33] Osintsov, Y., Chistov, Y. and Gerasimol, D. “The computer catalogue of the Kunstkammer Museum collections and perspectives of an Internet-share anthropological database”. *CAA* 2001.

[34] Schlader, R. Archaeological databases: what are they and what do they mean? *CAA* 2001:517-520.

[35] 河野一隆・塚本敏夫・魚津知克「古環境復原のための考古学情報クリアリングハウスの構築」人文 DB 論文集 8:41-48.

[36] 鋤柄俊夫「中世都市の景観と構造」人文 DB 論文集 8:57-64.

[37] 宮原健吾・内田賢二「平安京における空間情報システムの整備と条坊復原」人文 DB 論文集 8: 49-56.

[38] 森 洋久「分散型 GIS『GLOBALBASE』の実装」人文 DB 論文集 8:33-40.

2003 年

2004 年

[39] Anastasio, S., Lebeau, M. and Sauvage, M. *Atlas*

of Preclassical Upper Mesopotamia. Subartu 13. Turnhout: Brepols.

[40] Asmus, B. and Meister, C. "The Cederberg Rock Art Survey Project: a co-ordinated field record and database structure". *CAA* 2003:230-233.

[41] Joseph, N., Green, D., Cosmas, J. and Itegiaki, T. "Replication in archaeological information systems". *CAA* 2003:216-220.

[42] Zeeb-Lanz, A. "Effective - easy to use - economical: PGIS, the database for archaeological monuments in Speyer (Palatinate), Germany. *CAA* 2003:325-329.

[43] Matschegg, M. "REALonline: IMAREAL's digital image-server". *CAA* 2003:214-215.

[44] Nautiyal, V., Nautiyal, S., Naithani, M., Sanjiv, J., Clark, J. T., Landrum III, J. E., Bergstrom, A., Frovarp, R., Hawley, J., Eichele, D. "3D modelling and database of Indian archaeological pottery assemblages: HNBGU-NDSU collaborative initiative". *CAA* 2003:103-106.

[45] Tringham, R. "Interweaving digital narratives with dynamic archaeological databases for the public presentation of cultural heritage". *CAA* 2003:196-199.

[46] Pardi, G. "The thesaurus of archaeological toponymy". *CAA* 2003:225-229.

[47] 北條芳隆「前方後円墳の築造と方位観念・社会的背景の復元に関するデータベースの活用」人文 DB 論文集 10:43-52.

[48] 宝珍輝尚「遺跡と遺物・遺構表示の階層性に関する一考察」人文 DB 論文集 10:27-32.

2005 年

[49] 津村宏臣「4D-GIS による遺跡空間データベースの構築と先史文化生態解析」人文 DB 論文集 11:29-36.

[50] 宝珍輝尚「階層に基づく遺物データベースシステム」人文 DB 論文集 11:97-103.

2006 年

[51] 宇野隆夫(編)『世界の歴史空間を読む:GIS を用いた文化・文明研究』国際シンポジウム 24, 京都: 国際日本文化研究センター.

[52] 河野一隆「地震考古学データベース(GIS)の開発と災害考古学の展望」宇野隆夫(編)pp.415-424. 宝珍輝尚「様々な集約を可能とする考古学データベースシステム」情報処理学会研究報告 2006-CH-70: 1-8.

[53] 森 洋久「考古学における分散型 GIS の展開」宇野隆夫(編)pp.125-135.

[54] 森本 晋「奈文研版遺跡データベースの構築」宇野隆夫(編)pp.115-123.

2007 年

[55] Berg, E. "Using a GIS-based database as a platform for cultural heritage management of sites and monuments in Norway". *CAA* 2006:321-328.

[56] Meyer, É., Grussenmeyer, P., Perrin, J.-P., Durand, A. and Drap, P. "Integration of heterogeneous cultural heritage data in a web-based information system". *CAA* 2006:329-340.

[57] 石崎研二「GIS・数理モデルによる集落分布の立地分析」人文 DB 論文集 11:73-80.

[58] 出田和久「近畿地方における前方後円墳の分布論的検討:GIS データベース利用の試み」人文 DB 論文集 11:63-72.

[59] 近藤康久・大杉千春・鈴木星「地域スケールで遺物分布を解析するためのデータベース設計とその問題点:南関東の縄文遺跡から出土した網漁具を題材に」日本情報考古学会講演論文集 3:59-66.

[60] 中村一郎「キトラ・高松塚古墳のフォトマップ撮影と画像の保存活用:埋蔵文化財写真でのデジタル画像とデータベース」人文 DB 論文集 13:35-42.

[61] 藤本 悠「地理情報標準応用スキーマ準拠の遺跡空間データベース構築」日本情報考古学会講演論文集 4:109-116.

2008 年

[62] McKeague, P. and Jones, R. "The national database for Scotland: Evolution of the digital resource". *CAA* 2007:186-193.

[63] 寺村裕史「GISを用いた遺跡のデジタル測量と遺跡空間データベースの構築」人文 DB 論文集 12:21-28.

2009 年

[64] Gillam, J. C., Tabarev, A. V., Izuho, M., Nakazawa, Y., Quanjia, C., Tsogtbaatar, B. and Yoo, Y. "The Far East Archaeological Database (FEAD): A maximum 1-minute-resolution dataset for exploring the big picture". *Current Research in the Pleistocene* 25: 197-200.

[65] 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 15:1-12.

[66] 松森智彦・中村 大・木村啓章「GIS と RDB による人文科学研究支援システムの開発事例」地理空間情報 学生フォーラム 2009 in 関西要旨集.

2010 年

[67] 近藤康久「地理情報システムを用いた考古学的時空間分析の方法と実践:西南関東における縄文時

代鍾具分布を題材に『東京大学博士論文』。

※ <http://public.me.com/yaskn> よりダウンロード可能
(著者許諾済、2010年11月27日現在)

[68] 近藤康久「考古学におけるネットワーク・コンピューティング:『旧人・新人交替劇』関連遺跡データベース

スの取り組み」じんもんこん 2010 論文集(印刷中)。

[69] 日本旧石器学会(編)『日本列島の旧石器遺跡:日本旧石器(先土器・岩宿)時代遺跡のデータベース』。

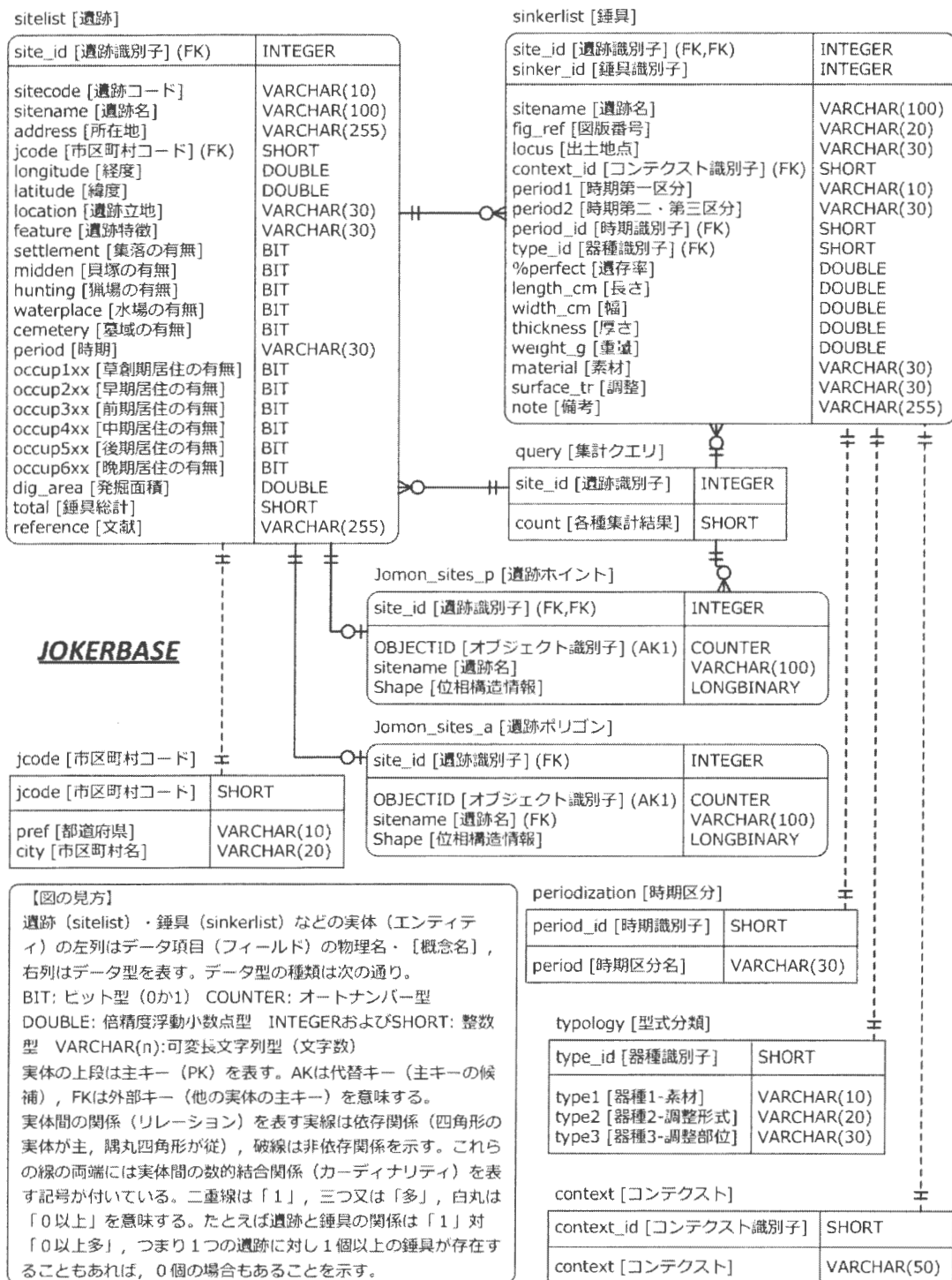


図 1: 縄文時代鍾具データベース「JOKERBASE」の設計図 (ER 図) (近藤 2010: 図 3.13)

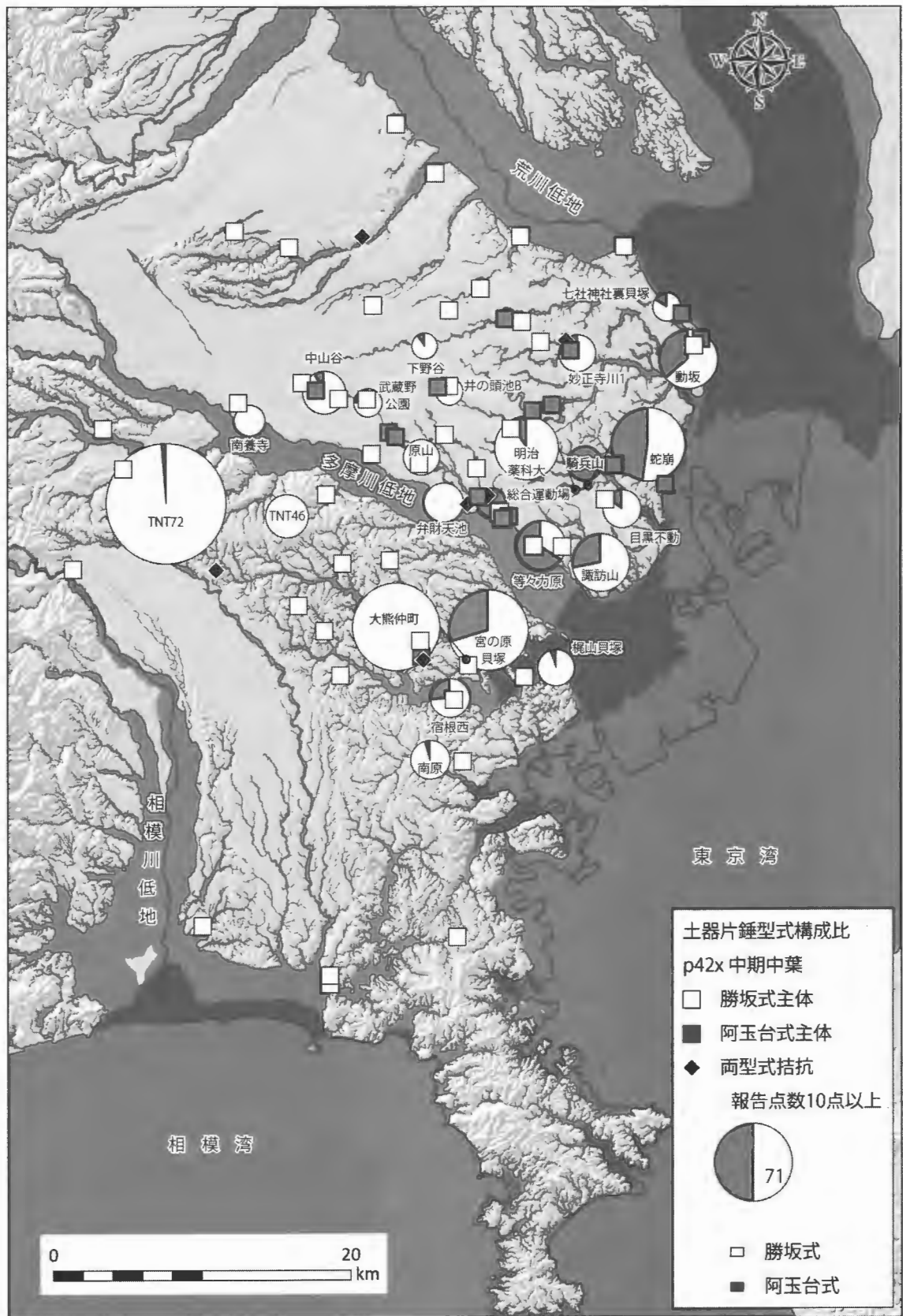


図 2: JOKERBASE と GIS を連係させた解析例。西南関東における縄文時代中期中葉土器片錘の型式構成比。
古東京湾西岸一帯に阿玉台式土器片錘が多く、内陸に入るほど勝坂式の割合が増す(近藤 2010: 図 5.34)。

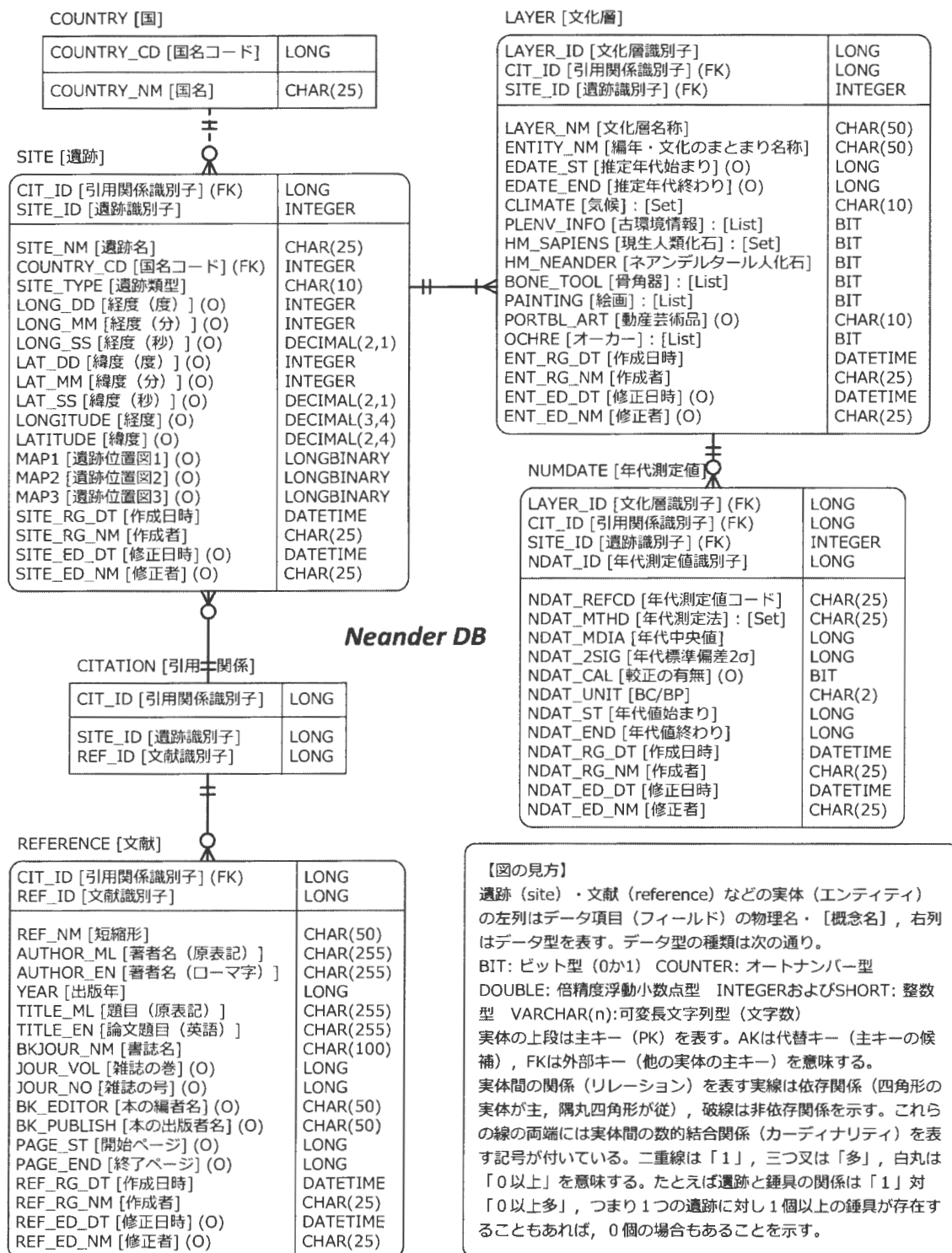


図 3: 「旧人・新人交替劇」関連遺跡データベース「Neander DB」の設計図 (ER 図)

京町家の 3 次元 CG 復元をめぐる問題

Some Problems on the 3DCG Reconstruction of *Kyomachiya*.

師 茂樹

明珍 健二

Shigeki Moro

Kenji Myochin

花園大学, 京都市中京区西ノ京壺ノ内町 8-1

Hanazono University, 8-1 Nishinokyo Tsubonouchicho, Nakagyoku, Kyoto.

あらまし: 花園大学情報歴史学研究室では、2009 年に発見された幕末と推定される間取図（梅忠町家屋敷絵図）をもとに、京町家の 3 次元 CG 復元を行った。本報告では、3 次元 CG による復元の文化遺産研究における意義を述べるとともに、復元の際に浮上したいくつかの問題点について検討する。

Summary: This paper provides a report of a research project on 3DCG reconstruction of a Kyomachiya (traditional town houses of Kyoto) in the end of Edo era based on the newly discovered layout drawings of the houses in Umetada-cho.

1. はじめに

1.1. 問題の所在

近年、京町家¹に対する関心が高まっている。例えば京都市は、新景観政策についての冊子の中で、平安遷都以来、1200 年を超える歴史を積み重ねてきた歴史都市・京都。永い歳月の中で、三方の山々や鴨川、桂川などに代表される山紫水明と称される豊かな自然、世界遺産を含む数多くの歴史的資産や風情ある町並みとが融合して、地域ごとに特色ある多様な景観が創りだされてきました。

と述べ、「風情ある町並み」を代表する京町家をはじめとした歴史的な建造物の保全・再生を新景観政策の五つの柱の一つと位置づけている²。京都市発行の資料をはじめ、多くの町家関連の文献では「今日見られる京町家は近世に成立した」というように概括されているが、後に見るように江戸時代においても前期と後期で町家は大きく変化しており、しかも江戸時代以前の町家は、蛤御門の変（1864 年）に伴う大規模な火災（所謂「どんどん焼け」）などの災害や、近代化に際しての市街地の開発などによってほとんど残存していない。また現存するものであっても、生活財であるという性質上、時代と共に改築・増築等の変化を受けているものが少なくないと思われ、現在の京町家とそ

の町並みに関する認識は、近世における「豊かな自然」と同様³、明治時代以降に作り出されたものがベースになっていると言わざるを得ない。

江戸時代以前の京町家や景観（町並み）に関する研究については、日向進氏が、

近世における京都町家の建築構成（平面形式・構造形式・室内意匠構成など）に関して、遺構による編年的な方法を用いてその類型の異同等を検討することは、遺構がほとんどないことを考慮すると、極めて困難な状況にあるといえる。いきおい文献史料（住戸絵図・普請願書など）による方法に重点を置かざるを得ない。⁴

と述べているように、町触などの文献史料や名所図会などの絵画資料に基づいた研究が中心となっている⁵。しかしながら、それらの史料も数が限られており、江戸時代以前の京町家の実態については、まだ充分には明らかになっていないというのが現状である。

また、前述した京都市の景観政策などを受けて、京町家の実態調査や歴史的な町並みの復元に関する研究においては、GIS やヴァーチャルリアリティなど、コンピュータを用いた研究が行われており、その中には江戸時代の町並みの復元を試みているものもある⁶。古地図や古写真などに基づいた

これら一連の研究は、京都全域を視野に入れた広範囲にわたる景観復元を試みているという点では一定の意義があると思われるが、江戸時代の町家を平屋建てと仮定したり、現代の町家をベースにした3次元モデルを用いたりするなど、江戸時代の京町家に関する先行研究が十分に反映されているとは言えない。もちろん、先に述べた通り、江戸時代以前の京町家の実態についての情報がそもそも少ないという現状に加え、広範囲の復元においてはモデルの単純化等が避けられないことを考えれば、従来の研究の問題点を一概に批判すべきではないだろう。しかしながら、専門知識を持つ研究者によって作成された3次元CGは、その「リアリティ」との相乗効果により、それを見る者（特に専門知識を持たない者）を「必然的に受身にさせる」可能性が強く⁷、限定された情報に基づいた研究者による一解釈（モデル）を提示したものであるということも伝わりにくい⁸。ある特定の見方や考え方を普及し固定化するためではなく、歴史に対する多様な見方や考え方を可視化し、人々に提示するためにも、様々なアプローチによる3次元CGが作成される必要があるのではないだろうか。

本稿では、2009年に発見された幕末期の京町家の間取り図に基づいた、従来とは異なるアプローチによる3次元CG復元（図1、図2）について報告するとともに、その歴史的意義についても検討したい。

1.2. 「梅忠町家屋敷絵図」について⁹

京都下京（現在は中京区）に属する梅忠町は、烏丸通～東洞院通間の三条通りに面した南北の地域（両側町）からなる。現在は北側半分が第二次世界大戦後にNTT京都中央局によって占められ、旧居住区全体を追跡することは困難な状態となっている。

2009年に明珍によって発見された文書群は、幕末期に調査され記録された町内町家の間取り図（以下「梅忠町家屋敷絵図」）や、天保5（1834）年の婚儀の記録（「上川遠藤婚儀諸式控」）などが含まれているもので、江戸後期から幕末期にかけての京都の町衆の生活文化を知る上で貴重な史料であると考えられる。

当時の梅忠町には30軒（南18軒、北12軒）の家屋があり、「梅忠町家屋敷絵図」にはそのうち27軒分の間取りの図面が含まれていることがわかっている。町内全体の町家の間取りをここまで網

羅的に記録した図面はこれまで知られておらず、また2階の間取り図が含まれるものも珍しい。梅忠町には明治2（1869）年に下京区第四組小学校が開校しているが¹⁰、それに関連する史料などから、「梅忠町家屋敷絵図」の成立は幕末期の安政3（1856）年から元治2（1865）年の間であると推定されている。

2. 京町家の3次元CG復元

2.1. 江戸時代の京町家

本稿に關係する範囲で、文献研究が明らかにした江戸時代の京町家の形式について概観しておきたい。もっとも、京町家に関する文献史料は江戸や大阪に比べると数が少なく、丸山俊明氏は「幕府直轄都市の江戸と京都の建築行政に共通する幕府の意向が窺える」ものの、「江戸の触書を京都に当てるには慎重であるべきで、少なくとも一定の論証を要する」と述べている¹¹。

さて、まず17世紀までの京町家の構造に関する規制について、丸山俊明氏は、

…厨子二階で、二階を塗窓や格子窓、壁にする町並景観は、町人が、寛永期の触書の影響を受けて主体的に選択した建築表現であった可能性が大きい。それが整った町並景観の基準となり、町式目など町人間の相互規制で維持されたと考えられる。…

また、所司代や町式目の関心が町並景観つまり表構えを整えることにある点を注視すると、敷地奥は関係がないという事になり、敷地奥に本二階建てを建て込む「おもて造り」や、棟位置を敷地奥にずらして表を厨子二階、奥を本二階建てとすることも可能になる。¹²と述べ、この形式が近世を通して維持されていくが、屋根については、17世紀に瓦葺規制があったものの、

…享保期の瓦葺規制解除後の京都中心部では、棧瓦葺が18世紀後半の宝暦～天明期に普及し、天明大火後の仮屋で一時的停滞を見た後に、寛政期（1789～1800）以降に再び普及して文政期（1818～29）までに完了した、つまり18世紀後半～19世紀初頭に棧瓦葺の町並景観を形成したと思われる。¹³

と述べているように、前半と後半では異なるとする。今回復元する京町家は言うまでもなく文政期以降の形式で復元する必要がある。

2.2. 復元の実際¹⁴

復元作業は、「梅忠町家屋敷絵図」を翻刻し方眼



図1 正面から



図2 裏庭

紙にトレースする段階と、それを元に CG を作成する段階の二つに分けられる。前者については「梅忠町家屋敷絵図」全体に対して行ったが、後者についてはその中の一軒（北側東より 12 軒目、三条烏丸交差点の北東角の家屋敷）を選び、復元を行った。

また、「梅忠町家屋敷絵図」に記載された情報が家を建てるための情報としては不足しているので、どちらの段階でも他の史料を参照しながらの復元的作業を伴う。なお、復元に際して参照した主な史料は以下のとおりである。

- ・「遠藤家文書」（明治 2（1869）年、京都府立総合資料館蔵）
- ・京都の建築規格に関する文献史料（『守貞謄稿』、『京都町触集成』など）およびその先行研究
- ・『三条油小路町並絵巻』（文政三（1820）年、京都府総合資料館蔵）などの絵画史料
- ・日向進『近世京都の町・町家・町家大工』（思文閣出版、1998 年）所収の田中家蔵町家古図
- ・京都市内に現存する町家（NPO 法人四条京町家〔下京区郭巨山町〕、無名舎・吉田家住宅〔中京区六角町〕など）
- ・大阪市立住まいのミュージアムで復元された町家¹⁵およびその図面
- ・白幡洋三郎ほか編『写真集成 京都百年パノラマ館』（淡交社、1992 年）などの写真史料

実際の復元作業について一例をあげると、「十一畳半 座敷」「八畳半 次の間」の間にある境界線をどのように解釈するかが問題となった（図 3）。図 4 はこの中の「八畳半 次の間」を翻刻・トレースし、畳の配置を推定したものである。

部屋と部屋の境界は襖で仕切られることが多く、この場合も襖があると考えられるが、図 4 を見れば分かるように部屋の幅が二間半なので襖が五枚並ぶことになる。しかし、襖は通常二枚一組で置かれることが多いので、五枚並ぶのは不自然であることが、検討会において指摘された。五枚並んでいた可能性や、半間よりも幅の広い襖を使用していた可能性など、様々な可能性が検討されたが、最終的には図 5 のように一番西側を壁にするということが一応の結論となった。

壁にした理由は、「十一畳半 座敷」の「床」についての解釈にある。元の図では部屋の隅に一畳分ほどのスペースがとられているだけであるが、

実際にはその（図で言うところの）下の方まで床があったのではないかと類例などから推定したのである。つまり「八畳半 次の間」との境界まで床（に準ずるもの）があるということになり、その境界は壁になる。もちろんこれ以外の可能性も十分にあり得るが、このような推定作業を各部分、部屋ごとに行ないながらの復元作業となった。今後の様々な検討や復元作業を経て、これらの推定が修正される可能性は十分にあるだろう。

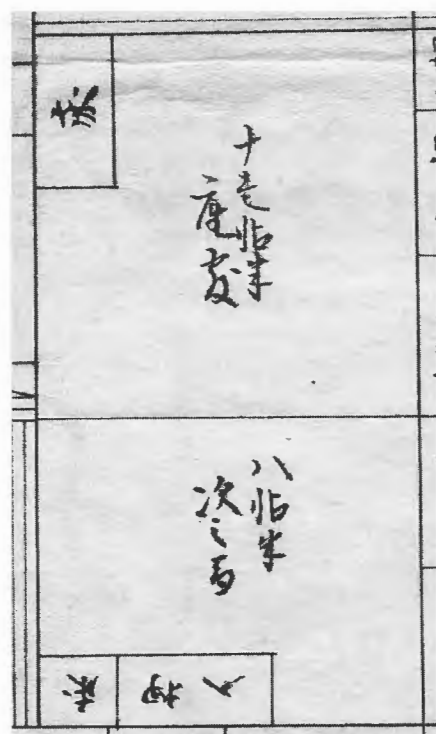


図 3 「十一畳半 座敷」（上）と「八畳半 次の間」（下）

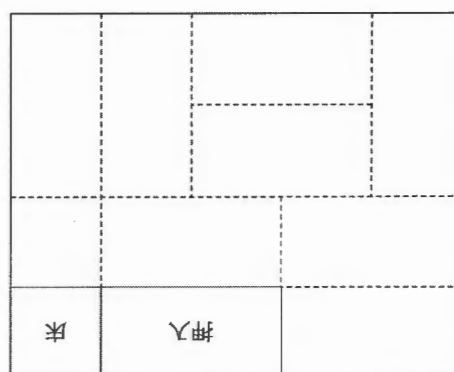


図 4 「八畳半 次の間」の翻刻・トレース図

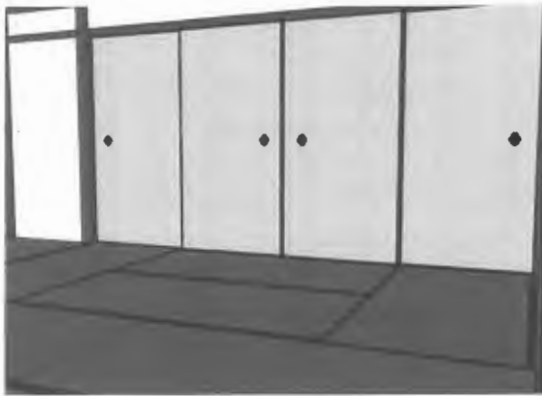


図5 検討結果を経た CG



図6 棧瓦とムクリ

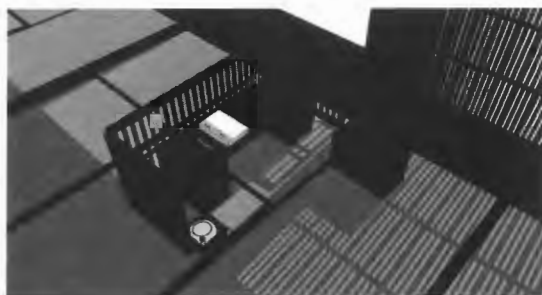


図7 格子からさしこむ陽光の処理

2.3. CG 作成の方針

CG 作成において主に使用したソフトウェアは、Google SketchUp 7（無償版・Pro 版の両方）¹⁶である。このソフトウェアは、建築物のように直線部分が多いものを効率的にモデリングするのに優れており、初心者でも習得が容易であると同時に、同じ Google 社の Google Earth などとも連携ができるので情報公開にも優れた面がある。モデリングにおいては数値を使ったミリ単位の指定も可能であるが、ソフトウェアとしてはメートル法（ミリメートル～メートル）とヤード・ポンド法（フ

ート、インチ）のみに対応しており、日本の尺貫法（間、尺、寸など）には対応していないので、今回のような作業においては変換の手間と若干の変換誤差が生じる問題がある。

先に述べたように、歴史的な文化遺産の 3 次元 CG 復元は、様々なアプローチによる複数の CG が作成されることが望ましいと考えられる。今回の復元においては、従来の方法とは異なり、細かな形状についてもできる限りモデリングをする（テクスチャでごまかさない）ことを試みた。屋根で言えば、瓦を一枚単位でモデリングし、所謂「ムクリ」（屋根の膨らみ）も再現している（図6）。また、日照のシミュレーションを行う場合、格子がただの平面に格子の画像を貼ったものでは室内に差し込む光線などを正しく処理できないので、格子もモデリングをしている（図7）。

ただし、昨今の映画のようなフォトリアスティックな CG の作成を目指したものではなく、研究・教育上、構造を理解しやすくするために輪郭がはっきりした CG にしている（フォトリアスティックな映像を出力することは可能）。

このような CG の作成には当然、コストがかかるうえ、コンピュータへの負荷も大きい¹⁷。しかし、上記のメリットに加えて、後に述べるような教育・啓蒙活動などへの応用においても、高い質感を持った CGの方が、応用範囲が広がるのではないかと考えられる。

3. 3次元 CG 復元の意義と課題

復元 CG を作成する主な目的は以下のとおりである。

1. 史料批判
2. 可視化による問題発見
3. 教育・啓蒙活動などへの応用

以下、これらの点について説明したい。

3.1. 史料批判・問題発見のための 3 次元 CG 化

史料に基づいて実際に家を（仮想的に）建てることで、その史料の信頼性を批判的に検討することが可能になる。今回の復元においても一階と二階の柱がずれる、敷地内に建物が収まらない等々の問題が発生し、以下の可能性が検討された。

1. 図面の数値の（場合によっては意図的な）誤記
2. 度量衡や規格に関する知識が、現在知られているものとは異なる
3. 規格が適用されていない（部分ごとにオーダーメイドされている）

また、復元にあたっては図面には描かれていない多くの部分を再構成しなければならないが、神棚があるべき部屋にそれを飾るスペースがない、図面上で部屋と部屋を仕切る線を襖とすべきか壁とすべきか判断が難しい等々の様々な問題が次々と浮上し、それぞれについて各種史資料との比較検討が行われた。

図8はその一例である。これは同じ位置の一階部分・二階部分を図面から抜き出したものであるが、一階部分が二部屋、二階部分が一部屋として描かれているように見える。この部分をCG復元する際、柱の太さや畳の大きさがすべて統一されており、一階の真ん中の線の部分に柱や襖などが入ると仮定して作業を進めたところ、一階と二階の柱が貫通しないという問題がおきた。すなわち、一階部分の真ん中の線の面積を0とすれば一階と二階の面積は一致するが、当然そのようなことはないので、その分だけ一階の方が広くなり、柱の位置が合わなくなってしまう。

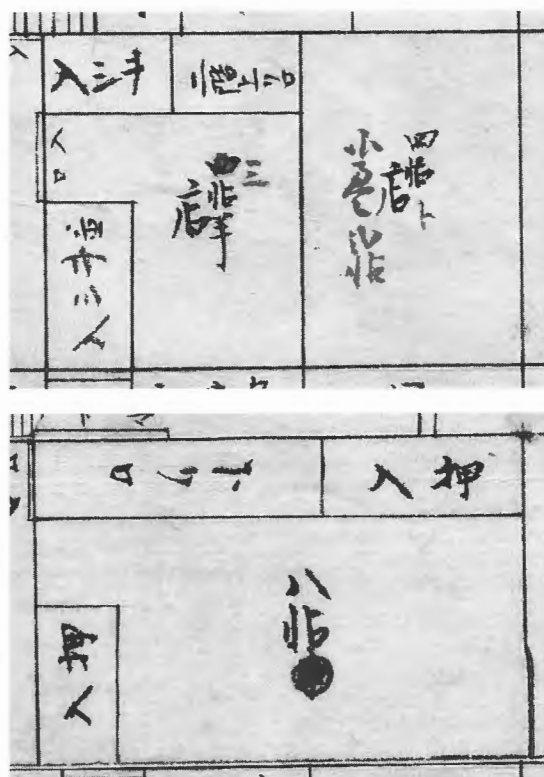


図8 通りに面した部屋(店)の一階部分(上)と二階部分(下)¹⁸

なぜこのような問題が起きるのかについては、いくつかの可能性が考えられる。

1. 一階の仕切り線は便宜的なものであり、実際には襖・柱等がなかった

2. 二階のどこかに誤差を吸収する部分(細い板間など)があった
3. 規格が適用されていない(部分ごとにオーダーメイドされている)

現時点では2で解釈するのが妥当であるとの判断で復元を行なったが、今後、他の町家の復元も含めた検討が必要である。

なお、図面に書き込まれた畳数や寸法情報の単純なミス程度であれば、わざわざCGにしなくても検討は可能であろうが、三次元的な空間としての検討はCGによる復元作業が有効であることが、今回の作業を通じて明らかとなった。

3.2. 教育・啓蒙活動への応用

最初に述べたように、現在普及している町家についての知識は、大正時代以降の様式に基づいたものが多く、近世以前の町家のあり方については必ずしもよく知られていない。3次元CGという説得力のある方法によって、教育・啓蒙活動の場において、あるいは観光の一アトラクションとして¹⁹、伝統的な知識を広く伝えることが期待できる。しかし、先にも述べたように、これがあくまでも研究者の研究活動を通じた一つの見解に過ぎないという点は、留意されなければならないだろう。

4. まとめと今後の課題

以上、幕末期にあったと推定される京町家の復元CG作成について簡単に報告をし、CG復元が文化遺産の研究やその社会還元に資するものであることを述べてきた。

繰り返し述べてきたように、今回の復元CG作成は現時点での限られた情報によって作成されたもので、今後の研究によってはゼロから作り直さなければならない可能性があることは再度指摘しておきたい。特に、一間を六尺五寸あるいは六尺三寸とする当時の規格や、規格通りに建物が建てられたという従来の通説など、再検討の結果によっては今回の復元は根底から覆る可能性も否定できない。これらの問題の解決のためには、残された26軒分の図面についての復元作業を行ない、そこで出てくるであろう共通の問題点について検討をしながら、個々の復元作業にフィードバックをする、という作業を行わなければならない。さらに言えば、梅忠町全体の復元は、単に町家の研究のための方法としてだけでなく、景観復元CGとしても魅力的なものになると期待される。

また、これまで研究の蓄積が少なく不明な点が多かったこともあって十分にできなかった二階の

復元についても、今後の課題としてあげられる。今回はこれまでの研究をふまえて、この建物の二階には（奉公人が寝泊りはするが）人が生活をする場所ではない、という前提で復元なども行ったが、実際には図面上に二階に床の間があり、生活をしていたのではないかという疑問を消すことができなかった。今後の検討如何では、場合によっては奥の方の二階を本二階にするなど、屋根の形状を変化させる必要も出てくるかもしれない。

謝辞

本研究は、2009 年度花園大学特別共同研究費「下京四番組・梅忠町の家屋敷および景観の復元と家督相続の実証的研究」（研究代表者・松田隆行）による成果の一部である。

¹ 丸山俊明氏は「京都の町家の固有性は、実はただ一点。他の大都市より多く残る点だけです。しかし、これは非人道的な空襲の結果です。日本人にとって不本意な状態なのです」と述べ、「京町家」という用語は空襲による他地域の町家の消失の上に成立している「京ブランド」にすぎないと述べる（『京都の町屋と町なみ一何方を見中様に作る事、堅仕間敷事』（昭和堂、2007 年、p. v）。丸山氏の「京ブランド」に関する意見には傾聴すべき点が含まれると思う反面、町家の形状などの面に京都の固有性がないという意見には賛同しかねる部分もあるため、本稿では「京町家」という用語を用いる。

² 『京都の景観』（京都市都市計画局都市景観部景観政策課、2009 年）

³ 京都市の新景観政策における「京都の景観」論は、あたかも「1200 年を超える歴史」の中で「豊かな自然」が維持されてきたかのような印象を与えるが、江戸時代には大規模な森林破壊があったことが指摘されており（田中淳夫『森林からのニッポン再生』平凡社、2007 年など）、実際『都名所図会』などを見ても比叡山や東山は山肌を露出させた禿山として描かれている。日本の森林が現在のような形になったのは近代以降の植林によるものが大きく、その意味で「豊かな自然」を安易に歴史に結びつけるべきではないと思われる。

⁴ 日向進『近世京都の町・町家・町家大工』（思文閣出版、1998 年）、p. 22。

⁵ 川上貢ほか「京都近世町家の宅地割と家屋平面」（『日本建築学会大会学術講演梗概集』、1971 年）、高橋康夫「近世京都における町家の諸類型について」（『日本建築学会近畿支部研究報告集』、1974 年）、高橋康夫「町家普請における工費と標準工数について」（『日本建築学会大会学術講演梗概集』、1974 年）、高橋康夫「近世京都の町の変遷と住居構成について—下京太子山町を事例として—」（『日本建築学会学術講演梗概集』、1975 年）、日向進前掲書、丸山俊明前掲書など。

⁶ 矢野桂司ほか「都市 3 次元 GIS/VR による京都バーチャル時・空間の構築」（『情報処理学会研究報告』2004-CVIM-142、2004 年）、河原大ほか「バーチャル京都：歴史都市の景観変遷」（『日本バーチャルリアリティ学会論文集』、2004 年）、矢野桂司ほか「都市景観のデジタル化：京都バーチャル時・空間—京都の過去の町並みを歩く」（『地図情報』25-2、2005 年）、矢野桂司「4D-GIS

を活用した歴史都市京都のバーチャル時・空間の構築と社会的合意形成」（『測量』55-5、2005 年）、矢野桂司ほか「歴史都市京都のバーチャル時・空間の構築」（『E-journal GEO』1、2006 年）、小阪佳宏ほか「GIS データに基づく京都の町並みの仮想空間自動生成：面・線・点データの立体化」（『情報処理学会研究報告』2006-CG-125、2006 年）、矢野桂司ほか編『バーチャル京都—過去・現在・未来への旅』（ナカニシヤ出版、2007 年）、矢野桂司「バーチャル京都を用いた都市景観の復元とマネージメント」（『環境と防災』1、2008 年）、矢野桂司ほか「バーチャル京都：時空間を取り入れた 4 次元 GIS」（『地学雑誌』117-2、2008 年）など。

⁷ 師茂樹「「デジタルアーカイブ」とはどのような行為なのか」（『情報処理学会研究報告』2005-CH-66、2005 年）参照。

⁸ 新谷尚紀「映像民俗誌論—『芸北神楽民俗誌』とその制作の現場から」（『民俗学の資料論』、吉川弘文館、1999 年）では、記録映画などの制作において、民俗調査における「調査研究者の主観による作為」を指摘した上で、「民俗映像だけが成果というのではなく、その制作自体についての記録資料が併せて必要」であり、「いつ、どこで、だれによって、どのような意図で、制作されたのかを文字記録によって残しておく必要がある」と述べているが、これは 3 次元 CG においても当てはまるのではないだろうか。

⁹ 本史料群についての詳細は、近日刊行予定の松田隆行ほか「下京・梅忠町の家屋敷景観の復元および家督相続の実証的研究」（『花園大学文学部研究紀要』43、2011 年 3 月刊行予定）を参照。

¹⁰ 『日彰百年誌』（日彰百周年記念事業委員会、1971 年）。

¹¹ 丸山俊明「17 世紀の京都の町並景観と規制—江戸時代の京都の町並景観の研究（その 1）—」（『日本建築学会計画系論文集』581、2004 年）。

¹² 丸山俊明前掲論文。

¹³ 丸山俊明「18～19 世紀の京都の町並景観と瓦葺規制—江戸時代の京都の町並景観の研究（その 2）—」（『日本建築学会計画系論文集』587、2005 年）。

¹⁴ 作業過程の詳細については、松田隆行ほか前掲論文を参照。

¹⁵ 大阪市立住まいのミュージアム『住まいのかたち 暮らしのなれ 大阪市立住まいのミュージアム図録』（平凡社、2001 年）参照。

¹⁶ <http://sketchup.google.com>（2010 年 11 月 9 日最終確認）。本稿執筆現在はバージョン 8 が配布されている。

¹⁷ 室内のムービーの作成においては、レンダリング速度をアップさせるために、不要な外装、特に瓦を取り除くことでオブジェクト数を減らしたりしている。

¹⁸ この図では二階部分に畳が敷いてあり、押入れが二つ設置されていることがわかる（他にも類例あり）が、江戸時代には通りに面した二階の居室利用が制限されていたことが先学によって指摘されている（丸山俊明前掲書 p. 76 など）。この他にも、「梅忠町家屋敷絵図」には二階の積極的な住居利用や三階の存在を示唆する図面も存在するので、近世の京町家における二階以上の利用形態について、多くの情報を提供してくれるのではないかと考えられる。

¹⁹ 歴史遺産の 3 次元 CG と iPhone・iPad 等を組み合わせた観光客のための情報提供システムの計画が進められている（「iPhone、歴史遺産を CG 再現 京都高度技研がソフト開発」京都新聞朝刊、2010 年 6 月 22 日）。

ブーリアン演算による歌ことばモデルの解析

山元 啓史

東京工業大学大学院社会理工学研究科

要 旨

八代集 (905 年頃～1205 年) の和歌 (約 9500 首) を対象にグラフによる歌ことばのネットワークモデルを作成し、分析を行っている。語彙はノードとエッジの集合であり、それらで構成されるネットワークである。この集合に対して、和・差・積を求め、八代集における歌ことばの変遷を分析する。

キーワード：和歌、ブーリアン演算、ネットワーク、八代集、語彙、日本語史

An Analysis of the Models of Classical Japanese Poetic Vocabulary using Boolean Operation

Hilofumi Yamamoto

Graduate School of Decision Science and Technology, Tokyo Institute of Technology

Abstract

We have been analyzing the transitions of meanings of Japanese words using the network models of classical Japanese poetic vocabulary in the the *Hachidaishū* (ca. 905–1205). Vocabulary can be expressed as a class of nodes and edges, networks, which allow us to operate them mathematically. This paper addresses the analysis of network structures of classical Japanese poetic words using boolean operation: union, intersection, subtraction.

Keywords: classical Japanese poetry dictionary compilation morphological parser, the *Hachidaishū*, boolean operation

1 はじめに

語彙とは「語の集まり」のことであって、数えられる個々の語のことではない¹。本研究は「語の集まり」を集合とし、ブーリアン演算を用いて分析する方法について述べる。

語彙研究には「単語の離散的な集まり」として、単語の計量分析を主とする研究と、「組織的なまとまり」として単語と単語の類縁関係を分析する研究がある [24, 3]。語彙を「組織的なまとまり」として捉

える研究としては、日本語ソーラスのひとつである分類語彙表 [4] を基準に分類カテゴリ別に語彙の出現頻度を計算する手法が多く報告されている (たとえば, [12, 13, 23, 1, 16] など)。ただし、語相互の結びつきや依存関係に関する研究はあまり報告されていない²

筆者はこれまでに和歌用語を中心にグラフ表現を用いた語彙の分析を行ってきた [17, 18, 19, 21, 20]。これらの研究では、一首に共に出現する 2 語のパ

¹したがって、語彙数ではなく語彙量といい、英語でも vocabulary は uncountable である [11, 序論]。

²語という概念を用いず、n-gram 統計 (任意数の文字列長の統計量) を用いて歌ことばのジェンダー (男ことば、女ことば) を明らかにした研究はある [5, 6]。

ターンを1単位として、その集合をネットワークで表現し、分析を進めている。1語ではさまざまに解釈される語の意味も2語で分析すれば、その2語の示す文脈が想像しやすくなる。また、それらをグラフで示すことにより、鳥瞰図のように「語の集まり」が一瞥できる利点もある³。このようなグラフ図形は数理的表現で、論理和、論理差、論理積などのブーリアン演算を施すことができる。またその数理的性質をそのまま語彙研究に応用することができる。本論は、この点に注目し、ネットワーク中に見られる語彙の構成要素や依存関係の分析を示すことを通して、語彙研究の枠組みを提案するものである。

以下に用語Aと用語Bを中心とする2つの語彙ネットワークに対するブーリアン演算の種類を整理する。

1.  統合／論理和: AとBの2つのネットワークの統合。
2.  交差／論理積: AとBの2つのネットワークに共通して出現したもの。2つのネットワークが共有している語相互のつながり方と語相互の接続の量を視覚的に示す。
3.  差分A／論理差: ABの論理和から用語Bのネットワークを差し引いたもの。用語Bを排除し、用語Aにのみ関わる語彙を抽出する。
4.  差分B／論理差: 上記の逆。
5.  排他／否定論理積: 統合から交差を排除したもの。用語Aと用語Bの相違を強調する。

上記の演算を用い、a. 語の比較（類似する2語の具体的な相違の分析）、b. 時間の比較（語の2時代における比較）、c. 作者の比較（2名の作者の比較や性別による相違）、などが分析できると考えている。本稿では、aとbについて報告する。

³語彙を空間的に分析する方法はグラフではないが、マトリックスやデンドログラムを用いて語と語の相互関係を計算する研究は以前よりあった（主に[8, 9, 10]など）。

2 方法

材料は、国文学研究資料館編集正保本版「八代集」（古今集、後撰集、拾遺集、後拾遺集、金葉集、詞花集、千載集、新古今集）収録のすべての和歌9503首を用いる。和歌テキストは新編国歌大観の番号を付けた上でファイルにセーブした。それぞれの和歌テキストは、古文品詞タグ付けシステム kh [22]で単位分割し、品詞タグを付けた。分割の単位は国立国語研究所β単位にしたがった。単位分割だけでは、異表記同義語の問題があるので、それぞれの語を t2c⁴を使って、シソーラスコードに変換した。

モデルはあらかじめ出現する個々の語について $idf[14, 15]^5$ を計算し、次に共出現パターン（テキストに共に出現する任意2語の組み合わせ）を生成し、先程の idf 値とパターンの頻度を使って、各パターンの重みを計算して作成する。共出現パターンは単なる2語の組み合わせリストではあるが、共出現パターンで描画されたグラフには、もとの文にある文脈が含まれることがわかっている[17]。その点が単語リストによる頻度集計と異なる。

すべてのパターンを描くとグラフは真っ黒な塊になってしまうので、各パターンがそのテキスト群において、どの程度重要なパターンであるのかを評価し、重要なパターンから描き出す手続きが必要となる。そこで、テキスト群(d)において任意の1語(t)が特徴的であるかを評価する式 $tfidf(1)[7]$ を拡張し、任意の2語のパターン(t_1, t_2)がどの程度特徴的であるかを評価する式(2)を用い、パターンの重み(cw)を計算する。

$$w(t, d) = (1 + \log tf(t, d)) \cdot idf(t) \quad (1)$$

$$cw(t_1, t_2, d) = (1 + \log ctf(t_1, t_2, d)) \cdot cidf(t_1, t_2) \quad (2)$$

$$cidf(t_1, t_2) = \sqrt{idf(t_1) \cdot idf(t_2)} \quad (3)$$

ただし、(2)の前半は t_1 と t_2 の2語が共出現した時のテキストの数。(2)の後半 $cidf(t_1, t_2)$ は、(1)の $idf(t)$ を拡張し、2語の idf 値の幾何平均(3)としたものである。以上の方法で得られた cw 値を相互に比較できるよう、一旦標準得点に変換し、正規化を行い、1σ以上の共出現パターンを Graphviz (<http://www.graphviz.org/>) で描いた。

⁴t2c: Token to Code, 自作。単位切りした語を入力すると分類語彙表準拠のシソーラス体系コードを返すプログラム。

⁵ idf はある特定のテキストにしか出現しない語か、どんなテキストにも出現する語なのかを示す値。 $idf(t) = \log N/df(t)$ ただし、 N はすべての資料の数、 $df(t)$ は、語 t の出現する資料の数。

3 結果

3種類の演算の結果を示す。はじめに（これは厳密にはブーリアン演算ではないが）コアノード（分析する用語）を全体の集合より削除し、その余りの集合を分析する方法、つぎに、2語の集合を統合した時の論理積（交差）をグレーで示し、2語の近さを分析する方法、最後に、2語の関係を時代を隔てて分析する方法について述べる。

3.1 コアノードの削除

削除はプルーニング（枝の刈り込み）とも呼ばれ、検索キーに関わるノードとエッジ（以下コアノード）をすべて削除する方法である。一般的にコアノードの共出現ウェイト（*cw*）の値がきわめて大きい時、すべてのノードはコアノードと結ばれ、放射線状に真っ黒な図形となる。これを自転車の車輪に喩えて、「スポークエフェクト」と呼んでいる。特に、地名のような特定の和歌にしか用いられない語の場合、よく見られる。そもそもコアノードにあたるキーワードで検索した歌のデータを用いてネットワークを描いているのであるから、すべての歌はコアノードと関係する。このことを前提に分析するなら、コアノードを刈り込んで見通し良くしてもかまわない。

図1は古今集のデータを用いて「梅」ネットワークを描いたものである。(a)は「梅」ノードの削除前、(b)は削除後である。古今集の場合、コアノードを削除しなくてもある程度、語相互のつながりは観察できるが、削除した方がよりわかりやすい。「梅」「鶯」「梅の香」「鶯が縫う梅の花笠」「梅花を折る」など、古今集特有の語のつながりが見えるようになった。

一方、図2は新古今集のデータを用いた「梅」ネットワークより「梅」ノードを削除する前(a)と削除した後(b)である。新古今集の場合、コアノードを削除しないと、「梅」以外の語のつながりは見えにくい。図2(a)をサッカーボールのような球体と見るならば、ボールの中心に「梅」があり、「梅」から伸びるエッジが球面を支えている（あるいは、つなぎとめている）ように見える⁶。「梅」を取り除くと、ちょうどボールの展開図が開くように、語相互のつながりが広がって見える(図2(b))。

⁶しばしば、図2(a)のような二重輪の構造（外側の輪と中心に集まるモコモコとした雲）になる。樹形図の描かれ方と同じなのであるが、その理由はまだよくわからない。

古今集の図1(b)と新古今集の図2(b)を比較すると、前者では「鶯」と「花」「色香」の関係が見えるのに対し、後者では、前者には見られなかった「鶯」と「雪」の関係が見える⁷。

3.2 2語の共有ノードの違い

つぎに、2語のネットワークの統合と交差による分析を示す。

図3(a)は「鶯」と「桜」の統合と交差、図3(b)は「鶯」と「梅」の統合と交差を示したものである。交差部分はグレーで示されている。

「鶯」は『万葉集』から数多く詠まれ、梅の花に鳴く鶯が最も多く、初春に鳴く鶯が春の最初に咲く梅の花とともに詠まれるのは当然といわれている[2, pp. 71-2]。図3が示すように、(a)と(b)の交差部分を比較すると、共に「鶯」と「桜」の各ノードはグレーで示されておらず、互いに同じ歌では出現しないことがわかる。共有するノードの数も3と少ない。一方、「鶯」と「梅」のノードは共にグレーで示されており、同じ歌に2語が使われていることがわかる。共有するノードの数も15であり、この2語の関係がよく詠まれることがわかる。

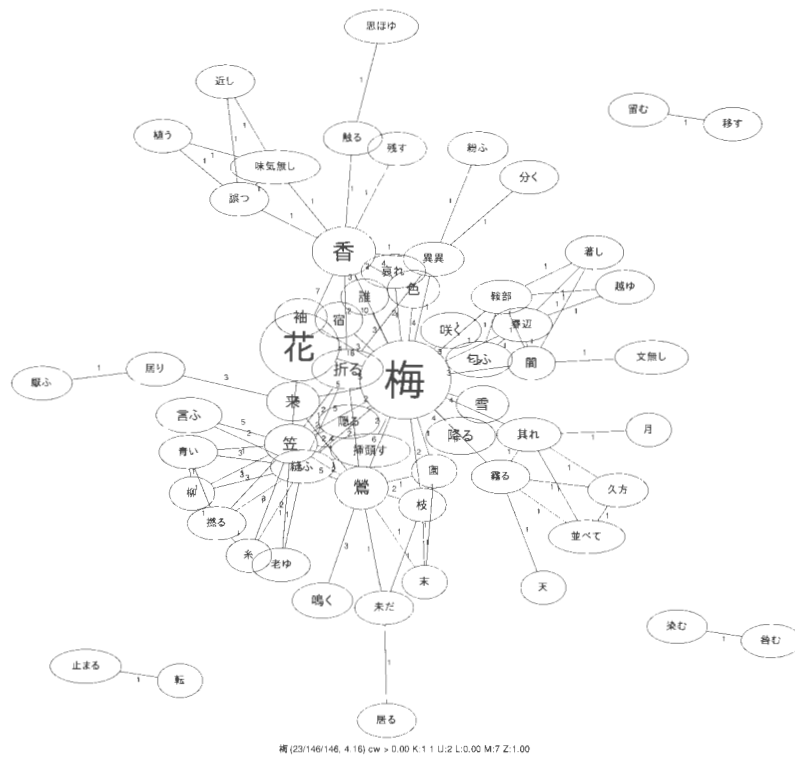
3.3 歌集で変化する共有ノード

最後に「桜」と「吉野」の関係が歌集によって変化するを示す。

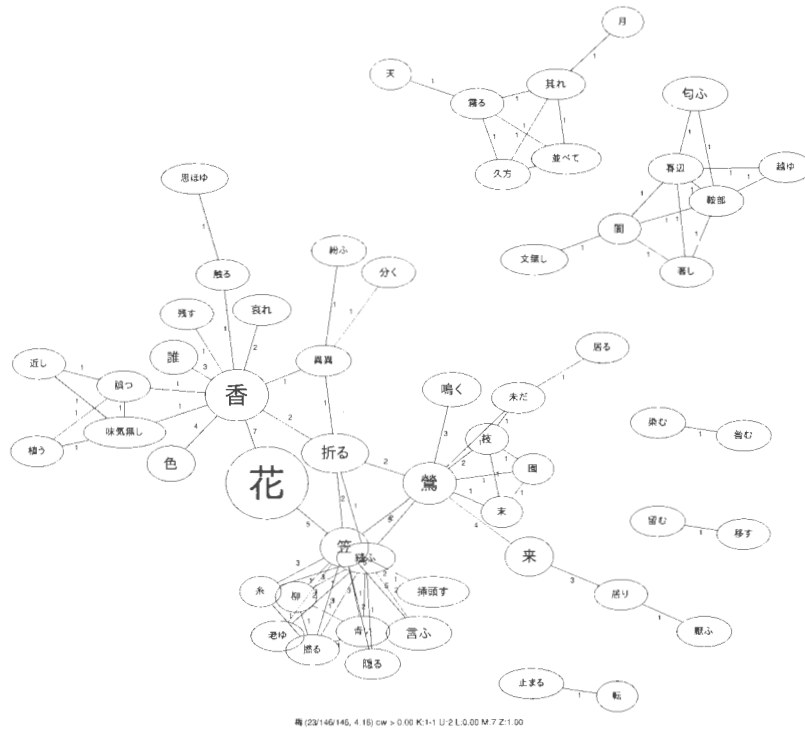
図4は、古今集における「桜」と「吉野」の関係(a)と新古今集における「桜」と「吉野」の関係(b)を示したものである。今でこそ「桜」と「吉野」の関係は有名であるが、「吉野山と桜の関係が決定的なものになるのは、やはり『吉野山去年こぞのしをりの道かへてまだ見ぬ方の花をたづねむ』(新古今集・春上)を代表とする数々の歌をよんだ西行とその時代[2, p. 436]で、古今集の時代では、「桜」と「吉野」の関係より「雪」と「吉野」の関係の方が強いといわれる[2, p. 435]⁸。図4(a)を見ると、確かに古今集の「吉野」は「桜」との関係よりも「雪」

⁷この関係は、新古今集30番（読人不知）「梅か枝に／なきてうつろふ／鶯の／はね白妙に／あは雪そふる」に見られる。

⁸片桐[2, p. 435]はよと「山岳信仰と結びついた吉野の山々のたたずまいがますます神秘的イメージになって行ったのであろう。(略)山岳信仰の地・隠遁の地としての吉野山であったが、そのような神秘的なイメージは雪をいただく山々の姿とマッチして、吉野山といえば雪がよまれるというようになった」という。

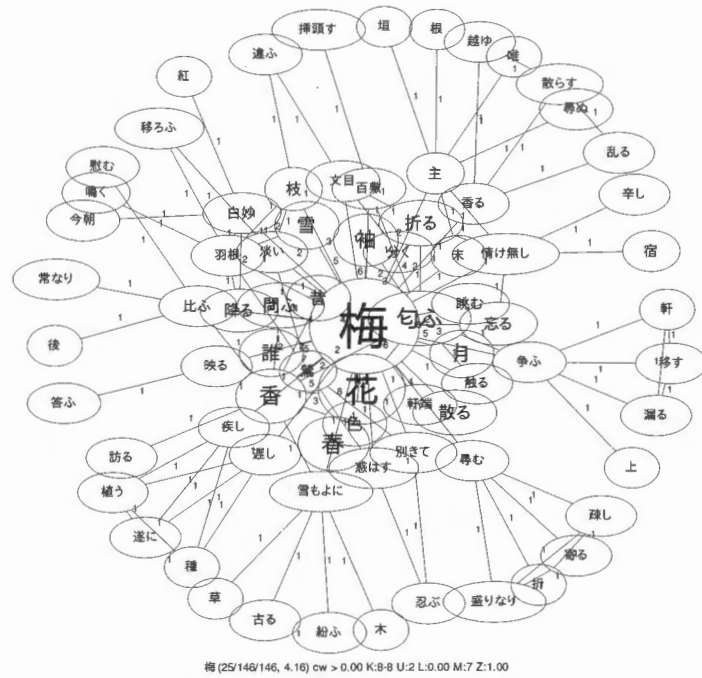


(a)

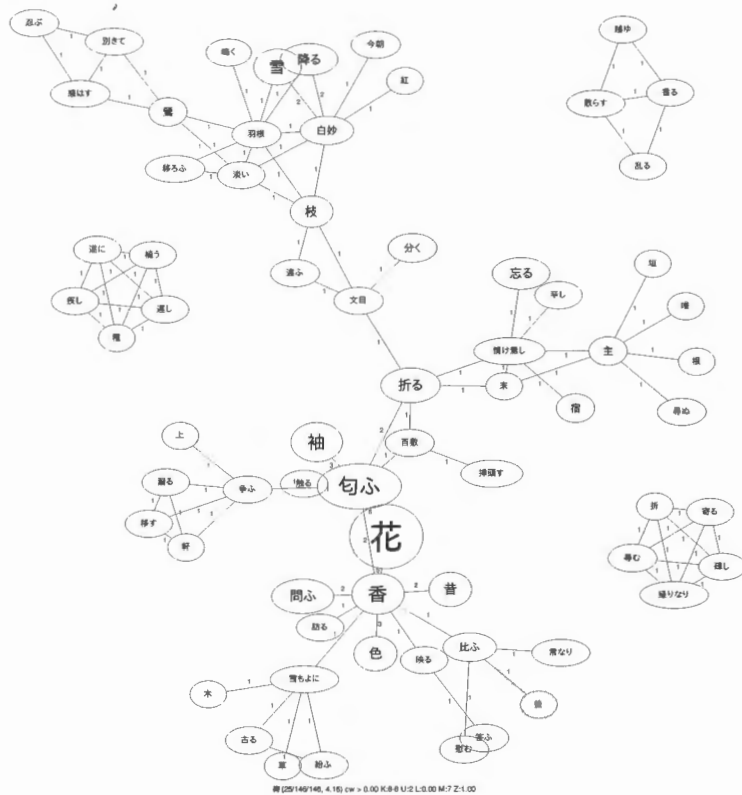


(b)

図 1: 古今集データにおける「梅」ノード削除前 (a) とノード削除後 (b)



(a)



(b)

図 2: 新古今集データにおける「梅」ノード削除前 (a) と削除後 (b)

との関係の方が強く、「吉野」のネットワーク中に「雪」「白雪」「御雪」「寒し」のように「雪」を表す語や、「隠れ家」「(雪道) 踏み／平らす (馴らす)」のように「隠遁」を表す語が見られる。図 4 (a) と (b) の 2 つの歌集 (約 905 年と 1205 年の成立) を比較することによって、「桜」と「吉野」の関係が時代につれて変化していることがわかる。

4 おわりに

本稿は、ブーリアン演算で語彙の集合を分析する方法について述べた。任意の 2 語の共出現パターンの違いを統合・交差を用いて示すことができた。また、同様の方法により、時代にわたって 2 語の関係の変化を示すことができた。どの演算を利用するかは、あらかじめ部分的に出力された図を見た上で、研究目的に応じて、適宜判断しなければならない。どの演算がどのような局面に有効であるかは、今後の課題としたい。

参考文献

- [1] 犬飼隆：平安末期複合動詞の意味構造, 国語語彙史研究会 (編), 国語語彙史の研究, 第 9 巻, pp. 272–258, 和泉書院 (1988).
- [2] 片桐洋一：歌枕歌ことば辞典, 角川小辞典, 第 35 巻, 角川書店, 東京 (1983).
- [3] 計量国語学会 (編): 計量国語学事典, 朝倉書店 (2009).
- [4] 国立国語研究所 (編): 分類語彙表／フロッピー版, 国立国語研究所言語処理データ集, 第 5 巻, 大日本図書, 東京 (1994), 『分類語彙表』は 1964 年に国立国語研究所資料集 6 林大担当として 刊行された。
- [5] 近藤みゆき: n グラム統計処理を用いた文字列分析による日本古典文学の研究—『古今和歌集』の「ことば」の型と性差—, 千葉大学「人文研究」, Vol. 29, pp. 187–238 (2000).
- [6] 近藤みゆき: n-gram 統計による語形の抽出と複合語—平安時代語の分析から—, 日本語学, Vol. 20, pp. 79–89 (2001).
- [7] Manning, C. D. and Schütze, H.: *Foundation of statistical natural language processing*, The MIT press, Cambridge, Massachusetts (1999).
- [8] 水谷静夫: 共出現関係に拠る語彙分類の試み, 計量国語学, Vol. 77, pp. 1–13 (1976).
- [9] 水谷静夫: 語の共出現に拠る語彙構造探究の諸法, 計量国語学, Vol. 79, pp. 1–18 (1976).
- [10] 水谷静夫: 用語による梅・桜の歌の弁別, 計量国語学, Vol. 12, pp. 1–13 (1979).
- [11] 水谷静夫: 語彙, 朝倉日本語新講座, 第 2 巻, 朝倉書店, 第 1 版 (1983).
- [12] 中野洋: 新聞語彙調査の類別語彙表について, 電子計算機による国語研究 II, 国立国語研究所報告, 第 34 巻, pp. 38–54, 秀英出版, 東京 (1969).
- [13] 西端幸雄: 「歌物語」3 作品の使用語彙の比較, 「歌物語」語彙の数量的分析と研究, pp. 3–18, 第 1 版 (1996), 文部省科学研究費: 重点領域研究「人文科学とコンピュータ」研究成果報告書.
- [14] Robertson, S.: Understanding inverse document frequency: on theoretical arguments for IDF, *Journal of Documentation*, Vol. 60, pp. 503–520 (2004).
- [15] Rocchio, J. J.: The SMART Retrieval System: Experiments in Automatic Document Processing, in Salton, T. G. ed., *Relevance feedback in information retrieval*, pp. 313–323, Prentice-Hall, Englewood Cliff, NJ, 1 edition (1971).
- [16] 山田進: 意味分類辞書, 国語学, Vol. 53, No. 1, pp. 30–43 (2002).
- [17] 山元啓史: 古今集データベースによる歌語の視覚化, 人文科学とデータベース, 第 11 回シンポジウム, pp. 81–8, 人文科学とデータベース協議会, 大阪 (2005).
- [18] 山元啓史: コンピュータによる歌枕の分析, イタリア日本語教育協会, 第 3 回シンポジウム論文集, pp. 373–382, イタリア日本語・日本語教育学会 (2006).
- [19] 山元啓史: 歌ことばの可視化とコノテーションの抽出—グラフによる共出現パターンの作り方—, じんもんこん 2006, 人文科学とコンピュータシンポジウム, Vol. 2006, No. 17, pp. 21–28 (2006).
- [20] 山元啓史: ネットワークによる歌ことばのモデリング, 語彙研究, Vol. 2007, No. 5, pp. 21–32 (2007).
- [21] 山元啓史: モデリングによる歌ことばの変遷と分析—八代集・歌ことばシソーラスの開発—, じんもんこん 2007, 人文科学とコンピュータシンポジウム, Vol. 2007, No. 15, pp. 163–170 (2007).
- [22] 山元啓史: 和歌のための品詞タグづけシステム, 日本語の研究, Vol. 3, No. 3, pp. 33–39 (2007).
- [23] 山内洋一郎: 連歌分類語彙表 (体の類) 試案—宗祇関係千句連歌七種による—, 国語語彙史研究会 (編), 国語語彙史の研究, 第 6 巻, pp. 358–348, 和泉書院 (1985).
- [24] 田中章夫: 国語語彙論, 明治書院 (1978).

「現代日本語書き言葉均衡コーパス」のための
形態論情報データベースについて
On the Morphological Information Database
for Balanced Corpus of Contemporary Written Japanese

小木曾 智信

中村 壮範

Toshinobu Ogiso

Takenori Nakamura

国立国語研究所 言語資源研究系, 立川市緑町 10-2

National Institute for Japanese Language and Linguistics, Department of Corpus Studies,
10-2 Midori-cho, Tachikawa city, Tokyo

あらまし: 国立国語研究所を中心に構築が進められている「現代日本語書き言葉均衡コーパス」のために開発した形態論情報データベースについて報告する。このデータベースは辞書データベースとコーパスデータベースとからなり、語彙表を介してお互いのデータの同期がとられている。辞書データベースの内容は形態素解析辞書 UniDic のソースデータとして利用される。コーパスデータベースは形態論情報を埋め込んだ XML 文書を出力する。

Summary: This paper reports the morphological information database for the Balanced Corpus of Contemporary Written Japanese, which is under way at the National Institute for Japanese Language and Linguistics. This database consists of dictionary database and corpus database, and two databases are synchronized via the lexicon table. Contents of dictionary database are used for the source data of UniDic, an electrical dictionary for morphological analysis. Corpus database can output XML documents annotated with morphological information.

キーワード: コーパス, 形態素解析, 辞書, データベース

Keywords: corpus, morphological analysis, dictionary, database

1. はじめに

本発表では、『現代日本語書き言葉均衡コーパス』(以下 BCCWJ)のために開発した「形態論情報データベース」の設計と実装について報告する。BCCWJ は総語数 1 億を目標に、人間文化研究機構国立国語研究所コーパス開発センターを中心に構築が進められている大規模なコーパスであり、2011 年の公開を目指している。このコーパスにはすべてのテキストに長短二種類の形態論情報が付与され、語を基本とした検索や集計が可能となる。

形態論情報データベースは、このコーパスの形態論情報を付与し整備するとともに、コーパスを利用した研究を可能にするためのデータベースである。国立国語研究所で開発・運用を行っている。

このデータベースは辞書データベースとコーパスデータベースからなり、語彙表を介してお互いのデータ

の同期がとられている。辞書データベースは、形態素解析辞書 UniDic の元となる見出し語のデータを格納しており、コーパス中に現れる新語を随時追加している。コーパスデータベースは UniDic によって BCCWJ のテキストを解析した結果を取り込んだものであり、形態論情報の人手による修正を行うとともに、様々な手段によって検索することが可能となっている。

形態論情報を格納したコーパスのデータベースは、全文テキスト検索が可能であると同時に、当該のテキストの構成要素であるすべての語について詳細な属性(形態論情報)を保持し、これらを組み合わせた検索にも対応する必要がある。また、データの整合性を保つために、辞書データベースと関連づけて管理を行う必要がある。

こうした点で、一般的なデータベースとは異なる特徴を持っている。

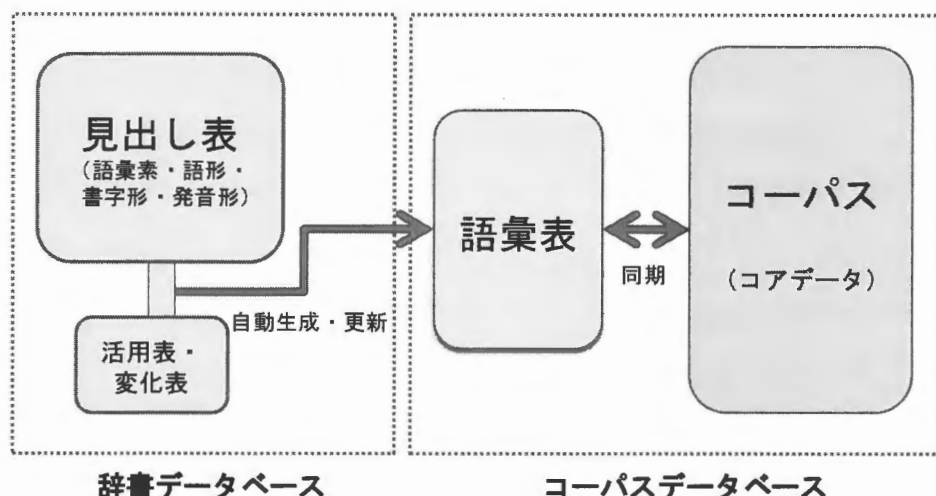


図1 形態論情報データベース全体図

2. 形態論情報データベースの概要

形態論情報データベースの主な利用目的は、次の3点である。

- (1) 形態素解析辞書 UniDic の元となる見出し表・活用表を格納し、見出し語の追加・修正作業を行う
- (2) BCCWJ の短単位で解析されたテキストを格納し、人手による修正を行って精度を高めたデータ(コアデータ)を作成する
- (3) 短単位で解析されたテキストを格納し、コーパスを利用した研究に利用する

(1)は辞書の見出し語、(2)、(3)はコーパスのデータを扱うことになる。これに対応して、形態論情報データベースは、(1)の辞書見出しを格納する「辞書データベース」と(2)、(3)のコーパスを格納する「コーパスデータベース」に分かれている。コーパスの形態論情報と辞書の情報を同一に保つ必要があるため、それぞれのデータベースは中間に辞書見出し表から生成される「語彙表」を挟んで連係している。コーパスに出現したすべての語は、原則として語彙表のいずれかのレコードと関連付けられる。(図1)

形態素解析辞書の作成という観点から見たときには、(1)、(2)は形態素解析辞書 UniDic の元となるデータを用意するための作業に相当する。(1)の見出し表を組み合わせることで解析辞書の見出し表(辞書)が生成され、(2)のコアデータから学習用コーパスが作られる。この二つのデータ元に、機械学習により形態素解析辞書が作成される。

(3)はこの形態素解析辞書によって解析されたテキストデータを学習コーパスと同様の形式で格納したものである。このデータは言語研究に利用するだけでな

く、辞書の整備(未登録の語を見つけ出し追加する等)のためにも利用される。

「形態論情報データベース」は、データベースソフト(DBMS)に Microsoft SQL Server を、クライアントに Microsoft Access で作成した専用アプリケーションを用いるクライアント・サーバ型のシステムとして構築されている。サーバの OS には Windows 2003 Server R2 Standard、データベース管理システム(DBMS)として Microsoft SQL Server 2005 Standard Edition (SP2) を利用している。十分なメモリを利用するためどちらも 64 ビット版 (x64 Edition) を利用している。

3. 辞書データベース

辞書データベースは、形態素解析辞書 UniDic の元となる見出し語のデータベースである。見出し語のテーブルのほか、活用表などの辞書作成に必要な情報からなる。

辞書データベースの基本となる見出し表は、UniDic の見出し設計にあわせて作成された「短単位語彙素」、「短単位語形」「短単位書字形」「短単位発音形」の4つである。

UniDic では図2のような階層化された見出し語が設定されている(伝ほか 2007)。

「語彙素」は国語辞典の見出し語に相当するレベルで、語の意味や語の出自などの情報はここに記述される。「語形」は異語形を区別するレベルで、たとえば「アマリ(余り)」に対する「アンマリ」「アンマシ」「アンマ」といった異語形、上一段活用と文語上二段活用といった活用の違いのほか、可能動詞形もここで区別される。「書字形」は異表記を区別するレベルで、漢字を使うか仮名書きするかといった違いのほか、送り仮名の揺れ

もここに記述される。「発音形」には発音やアクセントなどの情報が記述される。

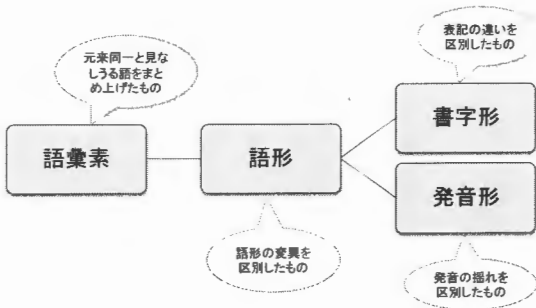


図2 UniDic の見出し設計

辞書データベースの見出し表はこの階層をそのまま反映し、「短単位語彙素テーブル」「短単位語形テーブル」「短単位書字形テーブル」「短単位発音形テーブル」の4つからなっている(図4)。

各見出し語は、具体的には図3のように階層化された形で格納されることになる(発音形は語形から直接結合する)。

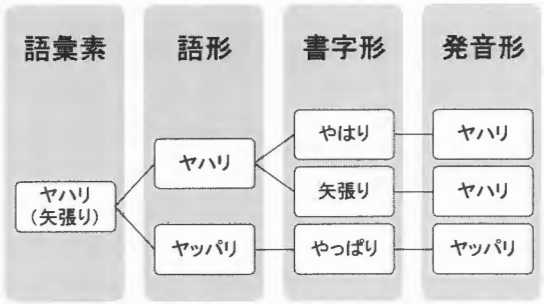


図3 UniDic の見出し構造の例

見出し表

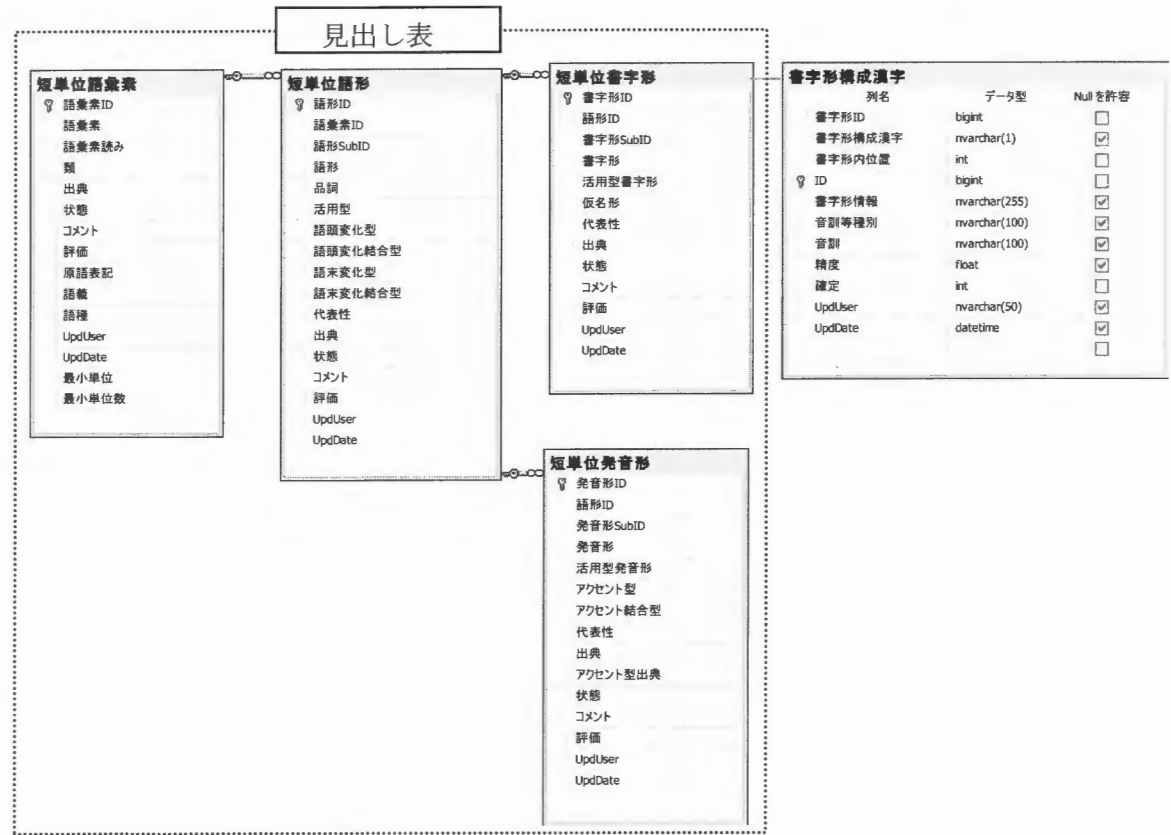


図4 辞書データベースの見出し表

語彙表の展開

辞書データベースには、見出し表のほかに、活用語を展開するための「活用表」と「活用型表」、語頭変化形を展開するための「語頭変化表」、語末変化形を展開するための「語末変化表」が存在する。

語形は、語頭変化(連濁など)・語末変化(促音化など)・活用の3種の変化により語形を派生する。個々の語形は、語頭変化・語末変化・活用の順に反映され、出現形に展開されることになる(図5, 図6)。



図5 語彙表の展開 (例: 辛い)

結果	メッセージ	id	lemma	reading	type	meaning	form	pos	cType	synCType	cForm	origin	pronBase	orthBase	kanBase	pronToken	orthToken	kanTok...	shModType
1		10653267969514795	読む	ヨム	用		ヨム	動詞一般	五段マ行一般	五段マ行	仮定形一般	lcp	ヨム	読む	ヨム	ヨム	読む	ヨム	
2		10653267969514796	読む	ヨム	用		ヨム	動詞一般	五段マ行一般	五段マ行	仮定形融合	lcp	ヨム	読む	ヨム	ヨム	読む	ヨム	
3		10653267969514849	読む	ヨム	用		ヨム	動詞一般	五段マ行一般	五段マ行	命令形	lcp	ヨム	読む	ヨム	ヨム	読む	ヨム	
4		10653267969514659	読む	ヨム	用		ヨム	動詞一般	五段マ行一般	五段マ行	意志推量形	lcp	ヨム	読む	ヨム	ヨム	読む	ヨム	M1@0
5		10653267969514658	読む	ヨム	用		ヨム	動詞一般	五段マ行一般	五段マ行	意志推量形	lcp	ヨム	読む	ヨム	ヨム	読む	ヨム	M1@1
6		10653267969514657	読む	ヨム	用		ヨム	動詞一般	五段マ行一般	五段マ行	意志推量形	lcp	ヨム	読む	ヨム	ヨム	読む	ヨム	M1@1
7		10653267969514625	読む	ヨム	用		ヨム	動詞一般	五段マ行一般	五段マ行	未然形一般	lcp	ヨム	読む	ヨム	ヨム	読む	ヨム	
8		10653267969514731	読む	ヨム	用		ヨム	動詞一般	五段マ行一般	五段マ行	終止形一般	lcp	ヨム	読む	ヨム	ヨム	読む	ヨム	NULL
9		10653267969514753	読む	ヨム	用		ヨム	動詞一般	五段マ行一般	五段マ行	連体形一般	lcp	ヨム	読む	ヨム	ヨム	読む	ヨム	NULL
10		10653267969514689	読む	ヨム	用		ヨム	動詞一般	五段マ行一般	五段マ行	連用形一般	lcp	ヨム	読む	ヨム	ヨム	読む	ヨム	
11		10653267969514693	読む	ヨム	用		ヨム	動詞一般	五段マ行一般	五段マ行	連用形撥音便	lcp	ヨム	読む	ヨム	ヨム	読む	ヨム	
12		10653305148384033	読む	ヨム	用		ヨム	動詞一般	文語四段マ行	文語四段マ行	命令形	活	ヨム	読む	ヨム	ヨム	読む	ヨム	
13		10653305148384001	読む	ヨム	用		ヨム	動詞一般	文語四段マ行	文語四段マ行	已然形	活	ヨム	読む	ヨム	ヨム	読む	ヨム	
14		10653305148383841	読む	ヨム	用		ヨム	動詞一般	文語四段マ行	文語四段マ行	意志推量形	活	ヨム	読む	ヨム	ヨム	読む	ヨム	M1@1
15		10653305148384060	読む	ヨム	用		ヨム	動詞一般	文語四段マ行	文語四段マ行	未然形<省略>	活	ヨム	読む	ヨム	ヨム	読む	ヨム	NULL
16		10653305148383809	読む	ヨム	用		ヨム	動詞一般	文語四段マ行	文語四段マ行	未然形一般	活	ヨム	読む	ヨム	ヨム	読む	ヨム	
17		10653305148383915	読む	ヨム	用		ヨム	動詞一般	文語四段マ行	文語四段マ行	終止形一般	活	ヨム	読む	ヨム	ヨム	読む	ヨム	NULL
18		10653305148383937	読む	ヨム	用		ヨム	動詞一般	文語四段マ行	文語四段マ行	連体形一般	活	ヨム	読む	ヨム	ヨム	読む	ヨム	NULL
19		10653305148384039	読む	ヨム	用		ヨム	動詞一般	文語四段マ行	文語四段マ行	連用形<省略>	活	ヨム	読む	ヨム	ヨム	読む	ヨム	NULL
20		10653305148383875	読む	ヨム	用		ヨム	動詞一般	文語四段マ行	文語四段マ行	連用形ウ音便	活	ヨム	読む	ヨム	ヨム	読む	ヨム	
21		10653305148383873	読む	ヨム	用		ヨム	動詞一般	文語四段マ行	文語四段マ行	連用形一般	活	ヨム	読む	ヨム	ヨム	読む	ヨム	
22		10653305148384043	読む	ヨム	用		ヨム	動詞一般	文語四段マ行	文語四段マ行	連用形撥音便	活	ヨム	読む	ヨム	ヨム	読む	ヨム	NULL
23		10653305148383877	読む	ヨム	用		ヨム	動詞一般	文語四段マ行	文語四段マ行	連用形撥音便	活	ヨム	読む	ヨム	ヨム	読む	ヨム	

図6 展開された語彙表の例

4. コーパスデータベース

BCCWJ のデータは XML で記述されている。コーパスデータベースでは、この情報を関係データベースの一般的な表で表現するために、Stand-off Annotation の方法により、「文字表」「短単位表」「文字修正表」「数字タグ表」「ルビ表」「タグ表」の各表に分けて取り込んでいる。形態論情報の処理に直接関連するタグのみ専用テーブルに書き込み、その他のタグは一括してタグ表で保管する。いずれのテーブルもサンプル ID と原文における文字位置をキーとして関連付けられている。

コーパスデータベースには各種のコーパスが格納されている。そのうち、人手修正を施したデータをコアデータと呼ぶ。コアデータは形態素解析辞書 UniDic の学習用コーパスとして利用される。コアデータ以外のデータは、見出し表に登録するための未登録語の採集や、コーパスを利用する研究のために用いるデータである。

コーパスデータベース内のテーブルは文字テーブルを軸として、サンプル ID と文字開始位置・文字終了

位置をキーにして関連付けされている。また、辞書データベースとは語彙表テーブルを介して関連付けされている。これによりコーパスデータベース用のアプリケーションからはコーパスデータベース内の全てのデータにアクセスできるようになっている。

短単位テーブル

短単位テーブルは形態素解析結果を取り込んだもので、コーパスデータベース内でも最も重要な役割をもつテーブルである。UniDic のもつ豊富な形態論情報のほか、コーパス中の位置情報をもつ。形態論情報は、「語彙素」「語彙素読み」「品詞」「活用型」「活用形」「書字形」「発音形」でユニークとなり語彙表中の 1 エントリと対応し、辞書データベースと関連づけられる。位置情報は、サンプル ID とサンプル中の連番で一意に特定される。

長単位テーブルと文節テーブル

長単位は、BCCWJ の形態論情報として付与される言語単位の一つで、文節をもとに、そこから付属語等

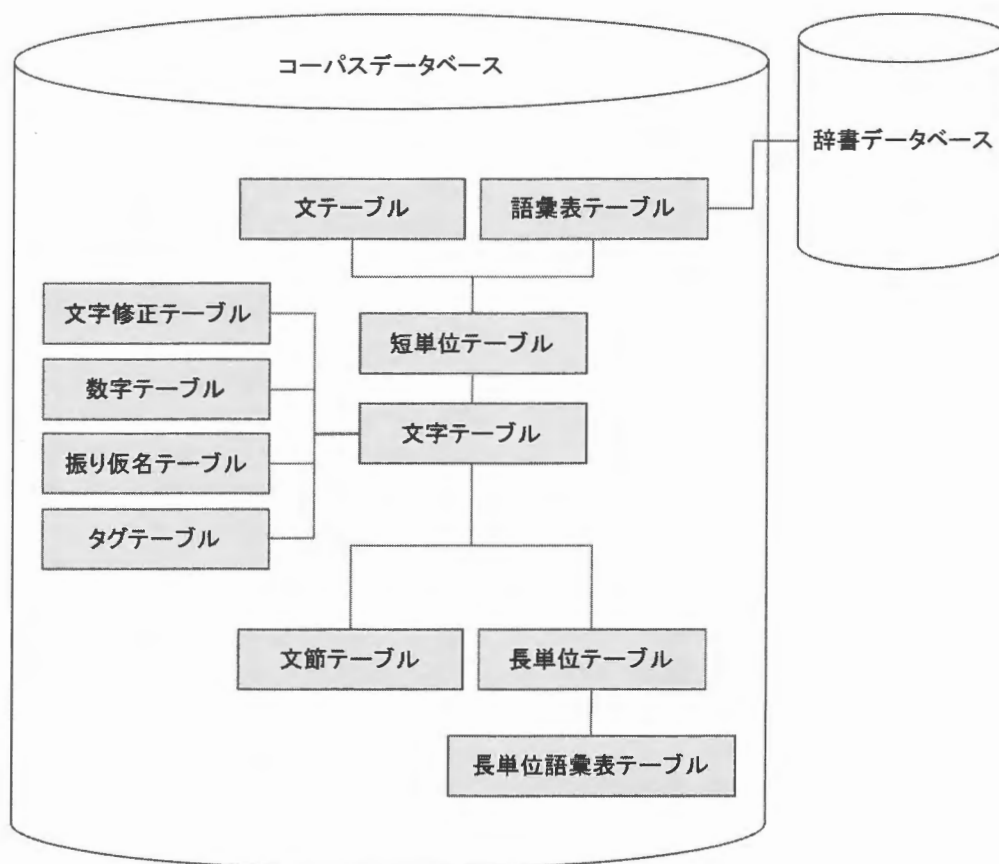


図7 コーパスデータベースのテーブル

を切り出したものに相当する。一つの長単位は、一つの短単位または複数の短単位の連続となる。

短単位と長単位・文節は、表 1 のような関係にある(表中の B は境界位置であることを示す)。文節境界は常に長単位境界であり、文節・長単位境界は常に短単位境界となる。また、文節や長単位は短単位の連続からなる。ただし、注釈的な括弧などにより、長単位が短単位の連続とならない場合がある。短単位と長単位は、語彙素・品詞・活用型等の情報をもつが、文節は境界のみを記録している。

表1 短単位・文節境界・長単位の例

短単位境界	短単位	文節境界	長単位境界	長単位
B	文化	B	B	文化庁文化交流使事業
B	庁			
B	文化			
B	交流			
B	使			
B	事業			
B	は		B	は
B	,		B	,
B	芸術	B	B	芸術家
B	家			
B	,		B	,
B	文化	B	B	文化人等
B	人			
B	等			
B	,		B	,
B	文化	B	B	文化
B	に			に
B	携わる	B	B	携わる
B	人々	B	B	人々
B	に			に
B	,		B	,
B	一定	B	B	一定期間
B	期間			

長単位はコーパスに出現したものを単位として認めるという形を取っており、コーパスから切り離した見出し表としては管理しない。そのため形態論情報データベースではコーパスデータベースの中でのみ取り扱われ、辞書データベースとは直接関係しない。

長単位に関係するテーブルとしては、長単位テーブルと文節テーブル、長単位語彙表テーブルがある。長単位テーブルは、出現した長単位の情報を格納するテーブルであり、語彙素・品詞・活用型などの情報が、短単位の情報を利用して付与される。文節テーブルは文節境界を記録するためのテーブルである。長単

位テーブルと文節テーブルは、短単位テーブルと同様に、文字表テーブルのサンプル ID と文字開始位置・文字終了位置によって関連付けされている。

5. クライアントアプリケーション

辞書データベースとコーパスデータベースは、それぞれ Microsoft Access による専用のクライアントアプリケーションによって修正・管理される。辞書管理ツール「UniDic Explorer」とコーパス修正ツール「大納言」である。

辞書管理ツール・UniDic Explorer

「UniDic Explorer」は辞書データベースに見出し語を追加するための中心となるツールである。見出し語の追加・修正作業には、見出し語表の階層をそのまま表示し、修正することが可能となっている。また、階層構造を保ったまま、見出し語をコピー・移動することができる。各見出し語はコーパス中の用例と関連づけられており、コーパス中の頻度が表示されるほか、用例の参照も可能である(図 8)。

コーパス修正ツール・大納言

「大納言」はコーパスの検索・修正を行うためのツールである。

コーパスの検索は、「短単位」を用いて、その語彙素読み・語彙素・書字形などの形態論情報を用いた検索ができるほか、全文テキスト検索も可能となっている。全文検索は、短単位テーブルから生成した文テーブルにインデックスを付与することにより実現している。さらに、前後に出現する短単位の形態論情報を複数組み合わせた「高度な検索」機能をもつ。

コーパスの修正は、短単位表については、語彙表を参照しながら、辞書データベースの見出し語と関連づけながら修正を行うシステムとなっている。これにより、常にコーパスと辞書との整合性を保ちつつ、随時修正を行うことができるようになっている。

短単位のほかに、長単位や文節境界の修正、誤った本文テキストの修正など、コーパスの各種テーブルについて、関連する単位との整合性を保ちつつ修正を行うことが可能となっている。

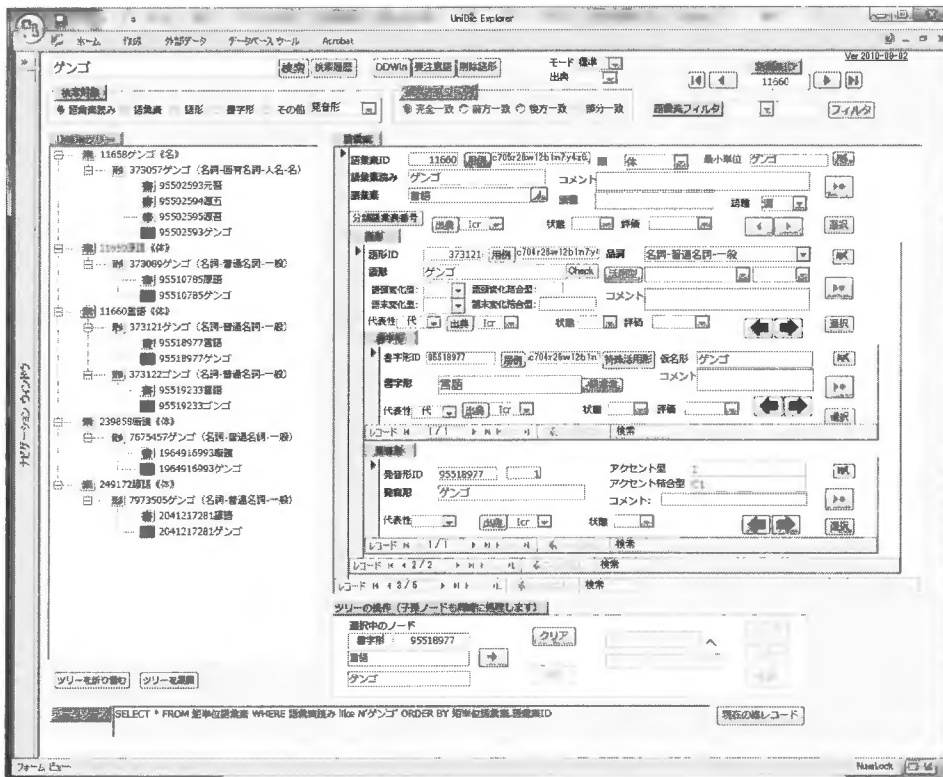


図8 辞書管理ツール・UniDic Explorer

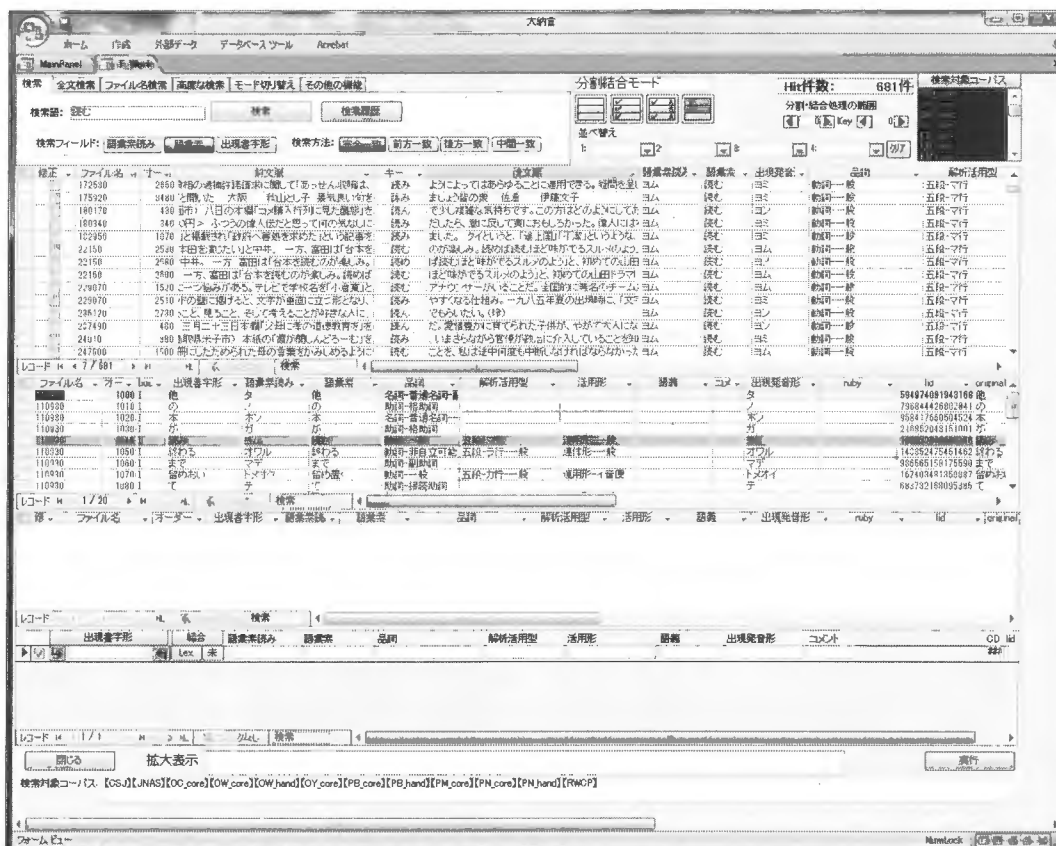


図9 コーパス修正ツール・大納言

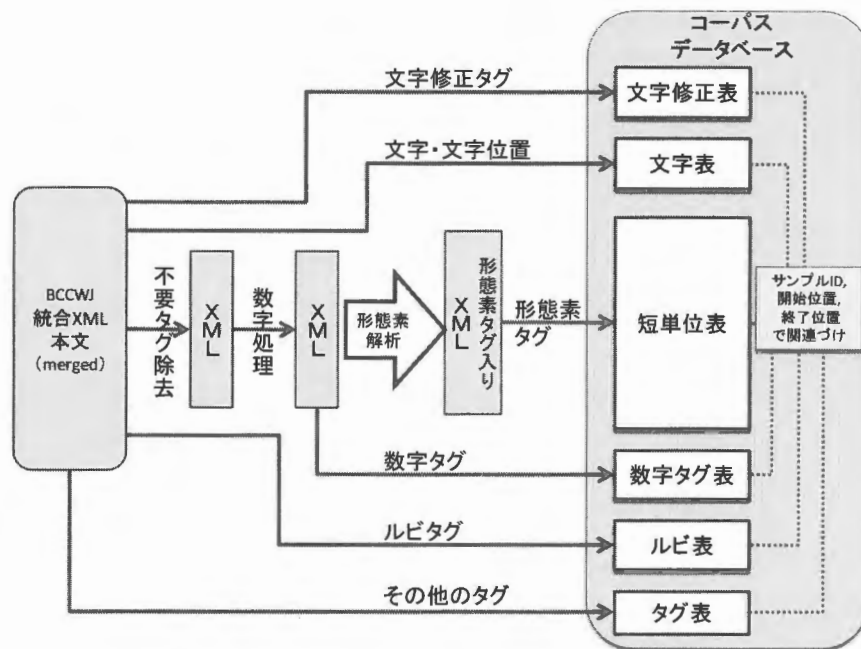


図10 BCCWJ サンプルの形態素解析とインポート

6. データのインポートとエクスポート

BCCWJ のサンプルに形態素解析を施し、コーパステーブルにインポートする際には、図 10に示すような流れになる。形態素解析や数字処理の邪魔になるタグの除去や、数字変換などの処理が加わるため、それぞれの表の情報を取り出す段階が異なる。タグ・文字テーブルは、元の XML 文書から直接取り出す(したがって、文字テーブルとタグテーブルからXML 文書が完全に再現できる)。数字タグは数字処理後のデータから取り出すことになる。

取り込んだデータは、人手で修正した後、元の XML 文書に形態素タグを埋め込んだ XML 形式でエクスポートすることができる。DBMS の管理ツール (SQL Server Management Studio) 上で、SQL 文を実行することによって出力される。エクスポート用の SQL 文では、各テーブルを結合し、データベース内部で XML 型のデータとして生成した後、ファイル出力している。データベース内でXML型のデータを生成するため、この時点で整形形式の XML であることが保証される。

テーブルの結合時には、タグテーブルを参照するが、このとき、ルビや数字などの別テーブルで管理されているタグはタグテーブルから出力せず、各テーブルの情報を元にタグを再構成して出力することになる。

7. おわりに

BCCWJ のために開発した「形態論情報データベース」の概要を紹介した。大規模コーパスの構築には、

注意深く設計されたデータベースが必須である。また、豊富な情報を持ち、日々人手修正が加わる1億語規模のコーパスデータベースを運用していくことは容易ではなく、多くの経験を積むこととなった。こうしたノウハウを、今後のより大規模なコーパスや、歴史コーパスなどの構築に活かしていきたいと考えている。

参考文献

- Masaaki Asahara, Ryuichi Yoneda, Akiko Yamashita, Yasuharu Den and Yuji Matsumoto (2002) Use of XML and relational databases for consistent development and maintenance of lexicons and annotated corpora , In *Proceedings of the Third International Conference on Language Resource and Evaluation (LREC 2002)*, Las Palmas, Canary Islands, Spain, pp.1372-1378.
- 伝康晴・小木曾智信・小椋秀樹・山田篤・峯松信明・内元清貴・小磯花絵 (2007) 「コーパス日本語学のための言語資源:形態素解析用電子化辞書の開発とその応用」(『日本語科学』22 号, pp.101-122)
- 小木曾智信・中村壮範 (2009) 『現代日本語書き言葉均衡コーパス』形態論情報データベースの設計と実装』国立国語研究所内部報告書 (LR-CCG-08-04)

参考 URL

形態素解析辞書 UniDic <http://download.unidic.org>

『古今和歌六帖』の本文と表現—デジタル校本の試み—

Text and Expression of Kokin Waka Rokujo: An Attempt to Computer-assisted Construction of Annotated Textbook

福田 智子* 竹田 正幸** 南里 一郎†

Tomoko Fukuda Masayuki Takeda Ichiro Nanri

*同志社大学 文化情報学部, 京田辺市多々羅都谷 1-3

Doshisha University, 1-3 Tataratsudani, Kyotanabe, Kyoto

**九州大学大学院 システム情報科学研究院, 福岡市西区元岡 744

Kyushu University, 744 Motooka, Fukuoka

†立命館大学情報理工学部非常勤講師

概要: 約 4,500 首の和歌を収める『古今和歌六帖』は、平安時代の歌語や表現の研究に欠かせない資料である。その主要伝本の電子テキストを作成し、新たに文字列データ解析ツールを開発して研究に活用する。伝本間の本文異同によって生じた複雑な書き入れを記述するための表記規則を案出し、改良を重ねた。現在、6 種の伝本について入力作業を完了している。開発したツールは、書き入れから生成される本文をも対象とした検索・解析機能を備えており、これを用いて複雑な本文の整理と分析を試みる。

Summary: "Kokin Waka Rokujo", an anthology including about 4,500 classical Japanese poems, is indispensable material for studies on words and expressions in Japanese poems at the Heian era. Of its ancient manuscript copies which have survived to the present day, we focused on the most significant six, built up their electronic text files. There are a number of handwritten notes in each of the copies, and most of them describe text differences. We devised and improved a set of description rules for representing such complicated notes. We developed a software tool for analyzing them. The tool is able to find occurrences of user-specified keywords even in (virtual) text strings implied by the notes. By using this tool we try to analyze complicated texts of the anthology.

キーワード: 校本, 古今和歌六帖, 平安和歌, ハイパーテキスト上のパターン照合, ビット並列化手法

Keywords: annotated textbook, Kokin Waka Rokujo, Japanese poem at the Heian era, pattern matching in hypertexts, bit-parallel technique

1. 『古今和歌六帖』の伝本

十世紀後半に成立したと思われる『古今和歌六帖』は、我が国における最も初期の類題和歌集といわれる。『万葉集』や『古今和歌集』『後撰和歌集』、さらには私家集や歌合などから採歌、分類した歌数は 4,500 首にのぼり、平安中期の和歌研究のみならず、次世代の『源氏物語』をはじめとする王朝女流文学にとりこまれる和歌表現の集合体としても、重要視されてきた(注 1)。

ところが、その一方で、『古今和歌六帖』に信頼できる古写本が存在しないというのも、周知の

事実である。すなわち、文禄四年(1595)の書写奥書を有する北岡文庫蔵永青文庫本が、確認できる書写年代の最も古いものであり、それ以前の古写本の現存はいまだ確認されていない。

また、『六帖』諸本の中でも、版本以前の「写本として伝存されていた時期の古今和歌六帖の古態を残す」(注 2)といわれる宮内庁書陵部蔵桂宮本(以下、「桂宮本」と略す)においてさえ、『古今和歌六帖』諸本の本文の乱れについて、次のように述べる。

すへてこの六帖いかにやらんいつれもくみな
かくのみしとけなきものにて侍れは本のまゝ
にしるしをく。のちに見ん人心えさせ給へし。

第一帖奥書(注3)

もとより約4,500首という大部な歌集である。
このような諸本の現存状況の上、複雑な書き入れ
をもつ伝本も少なくない。整理すべき本文情報は、
膨大な量に上ることになる。

『古今和歌六帖』諸本全体を視野に入れた基礎
研究は、1960年代に活発に行われた。平井卓郎氏
『古今和歌六帖の研究』(明治書院、1964年2月)、
中西進氏『古今六帖の万葉歌』(武蔵野書院、1964
年6月)、そして『図書寮叢刊 古今和歌六帖』上・
下巻(宮内庁書陵部編、養徳社、1967年3月・1969
年3月)である。最後の『図書寮叢刊』は、桂宮本
を底本として、主要伝本の基礎研究を精緻かつ広
汎におこなっており、現在においても、『古今和
歌六帖』研究の基本文献である。

これらの『古今和歌六帖』全体論は、その刊行
時期からみて、いわゆる手作業で行われたことは
いうまでもない。そのような時代に、これだけ広
汎な伝本を視野に入れ、大量のデータ整理がなさ
れたことは、敬服に値する。

本研究では、近年発展が著しい情報科学の技術
を生かして、文学研究者と情報科学研究者との共
同研究によって、それらの先行文献の驍尾に付し、
『古今和歌六帖』全体論の新たなかたちの実現を
試みる。

2. デジタル校本という発想

『古今和歌六帖』に限らず、伝本研究において
は、一つの伝本を底本として定め、他本との本文
比較により異文を表示する、校本の作成が必要不
可欠である。これによって、諸本の本文異同を、
一目で具体的に把握することができる。

そしてさらに、研究者側の視点に立てば、どの
伝本に研究の主眼を置くかによって、底本を自由
に替え、他の諸本との本文異同を一覧したいとい
う要望が少なくないであろうことは、容易に想像
できる。だが、校本を紙媒体で作成するかぎり、
諸伝本中の一本を底本に定めれば、その後の作業
において、容易に底本を変更することはできない。

そこで、たとえば、久曾神昇氏編『古今和歌集
成立論』資料編(注4)のように、あえて底本を定め
ず、諸本本文を列挙するのみにとどめた本文集成
も編纂された。近年においても、頼政集輪読会(代

表: 中村文氏『頼政集本文集成』(2009年1月)(注
5)は、PDF ファイルをCD-ROMに収録したものであ
るが、底本を定めず、諸本本文を列挙するという
体裁をとる。いわゆる校本のあり方からすれば、
本文異同を明示しないという点において退行して
いるともいえなくはないが、研究者の問題意識に
よって、中心とする本を替えて本文をチェックで
きるという点から見れば、必要に応じて底本を自
由に選びたいという要望から生まれた、一つのか
たちであるともいえよう。

そこで、紙媒体にテキストを固定するのではな
く、PCの画面上にテキストを表示することにより、
底本を任意に選択でき、本文異同も視認しやすく、
また、伝本を逐一増補できる、全く新しい発想のデ
ジタル校本システムを開発する。このシステムは、校本
画面の表示の他、和歌本文は、原態のままの表記
でも、清音仮名・歴史的仮名遣いに改めた本文で
も検索を可能にする。傍書によって生じる和歌本
文は、プログラムによって自動生成し、見せ消し
や補入記号の箇所も検索対象とする。つまり、あ
らゆる墨付きの検索に対応するものである。このデ
ジタル校本は、大量で複雑な伝本本文を効率的に処
理するという点で、実証的な文献学的文学研究をお
こなう際の利用価値は高い。諸本の本文異同を把
握し、作者(編者)自筆本により近い写本(原典)を
探求するとともに、享受史の観点から、後世の流
布本の性格を把握した上で、異同が生じた理由を
探り、書き入れをも含めた諸本本文の性格を総合
的に整理するための足掛かりをつくるのに大いに
役立つものと思われる。

3. 諸本原態の機械可読化

3-1. 作成したデータとその構造

現段階において、データの作成が完了しているの
は、次の6種の伝本である(注6)。主要伝本から、
本文系統の偏りがないように選択した。E)以外は
すべて写本である。A)は『新編国歌大観』(1984年
刊)の底本、また、E)は近世以降の流布本であり、
『続国歌大観』(1925~1926年刊)の本文は、その
系統に属す。

- A) 宮内庁書陵部蔵桂宮本
- B) 北岡文庫蔵永青文庫本
- C) 内閣文庫蔵「和学講談所」印本
- D) 内閣文庫蔵「江雲渭樹」印本
- E) 寛文九年版本
- F) ノートルダム清心女子大学附属図書館蔵
黒川本

各伝本は、“katural.txt”“kanbun6.txt”というように、帖ごとに六つのテキストファイルからなる。

次にデータの構造について述べる。一首分のレコードは、下記のように7行で構成する。具体例は、本ページ下部に示した。

- [1 行目] 歌番号(『新編国歌大観』に依拠)
- [2 行目] 題, 題についての書き入れ
- [3 行目] 作者名, 作者名についての書き入れ
- [4 行目] 歌本文 A(原態のままの表記)
- [5 行目] 歌本文 B
(清音仮名・歴史的仮名遣いに改めた本文)
- [6 行目] 左注, 左注についての書き入れ
- [7 行目] 歌本文 A についての書き入れ

歌番号は、『新編国歌大観』番号に対応したもので、伝本ごとの通し番号ではない。したがって、諸本間の歌順の相違により、歌番号が前後することがある。

また、題・作者名・左注についての書き入れは、それぞれ本行の本文の末尾にカンマを付した後に記す。歌本文 A の書き入れのみ、本行の本文とは別の行(7 行目)に改めて記す。

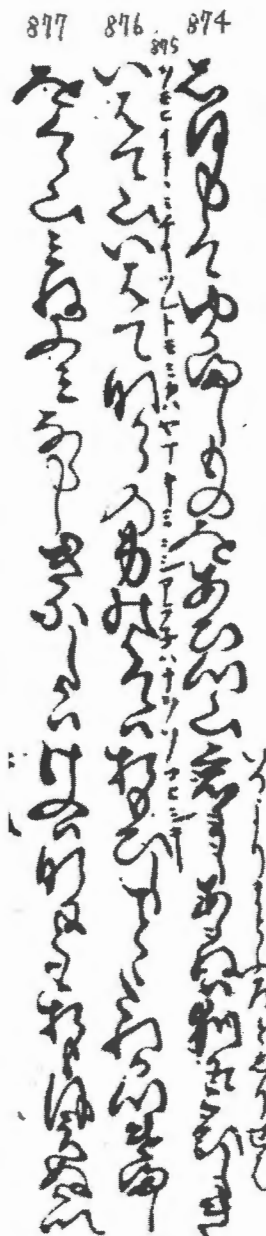
歌本文は、A と B を設ける。A では、原態の表記を可能なかぎり示す一方、B は、清音仮名・歴史的仮名遣いに校訂した本文である。

3-2. 歌番号の扱い

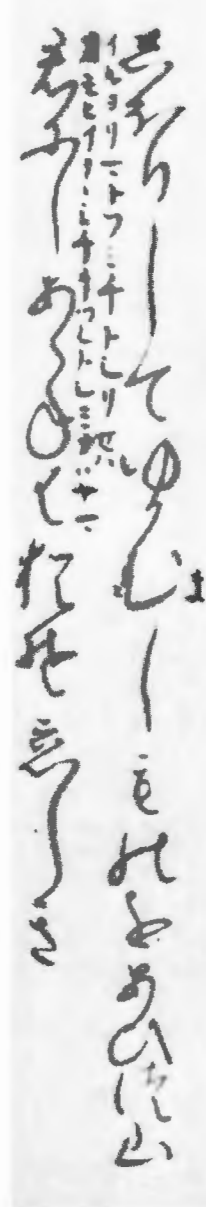
『新編国歌大観』所収の『古今和歌六帖』は、桂宮本を底本とし、その歌順に基づいた歌番号が付されている。現代の研究者にとっては基準となる番号であり、それゆえ、本研究で作成するデータに含まれる一首一首のレコードにも、その番号を振ることにした。

しかし、『新編国歌大観』は、底本の桂宮本を独自に校訂した上で歌番号を付している。伝本の原態から見ると、すべての歌を均質なものとして一律に扱うことはできない。

たとえば、右の桂宮本 875 番歌は、片仮名小書きによる書き入れである。本行は、平仮名に若干の漢字を混ぜた表記で、文字も大きく、書写の段



桂宮本 875 番



永青文庫本 874-875 番

階が明らかに異なっていることがわかる。そこで、書き入れ和歌の歌番号には、前に“\$”を付け、“\$875”とした。

とはいえ、このように一首全体が書き入れである例は、さほど問題にはならない。むしろ、さらに右に示した永青文庫本のように、本行の上行と

(例)「桂宮本」5 番歌

5

*

*

やま風に/とくる氷の/ひまことに/打出る浪や/春のはつ花

やまかせに/とくるこほりの/ひまことに/うちいつるなみや/はるのはつはな

*

s{古一春上},1[1-2]r{谷イ|たに}

下句の間に、下句と上句の書き入れがある場合は、データ作成の際に特別な処理が必要である。

この例では、874 番歌の上句と 875 番歌の下句で本行の一首が構成されている。だが、その行間に 874 番歌の下句と 875 番歌の上句が書き入れられているため、この片仮名小書きによって、本行の一首の歌が、二首に分かれてしまう。

そこで、永青文庫本のこの歌については、本行の本文をそのまま一首の歌本文として採用し、歌番号は、874・875 の両方を振ることにした。同様の事例は、永青文庫本 1094 番・1095 番にもある。

3-3. 表記規則の案出

伝本に見える墨付きすべてをデータに反映させるためには、以下の手順を踏まなければならない。

- A) 文字を読む。
- B) 音数律を読み取る。
- C) 本行と書き入力を区別する。
- D) 書き入れの位置を把握する。
- E) 書き入れの意味を解釈する。

その上で、得られた情報を、論理的に記述することが求められる。それは、人の目で見て解ればよいという次元ではなく、計算機処理が可能な表記規則である必要がある。

とはいえ、人間がデータを入力する際に、必要以上の負担を感じないような工夫はすべきであろう。表記規則は、伝本から読み取れるデータの「解釈」であって、具体的な事例の抽象化である。したがって、人間にも理解しやすく、かつ、入力ミスを極力抑制できるような表記規則であることが望ましい。

さらには、いったん策定した規則でも、常に修正と更新とが必要である。入力作業が進むにつれ、また、対象とする伝本を追加するたびに、想定外の例を見いだすことになる。その都度、表記規則を見直し、タグの問題点を修正することは避けられない。

そこで、現時点では、次のような表記規則によってデータを作成している。これにより、6 種の伝本については、ある程度の重要な情報の記述を可能にした。以下、その概要を示す。

基本的には、英数字と、“[]” “{}” “|” “*” を中心に用い、その他の記号を補助的に使用する。

A) 英字

英字には、次の三種類の機能を持たせる。

[位置] 書き入れの位置情報を示すもの

[よみ] 書き入れによって本行のよみを置換して異文を生成するもの

[位置][よみ] 上記の両方を行うもの

以下、その一覧を示す。

- t 行頭注 [位置]
本行の上部に付いた書き入れ
- T 行頭注 [位置][よみ]
表示は“t”と同じ、よみの生成もする
- d 行末注 [位置]
本行の末尾に付いた書き入れ
- D 行末注 [位置][よみ]
表示は“d”と同じ、よみの生成もする
- g 行間書き入れ [位置]
行間の書き入れ
- s 集付 [位置]
本行の上部に付いた歌集名・部立などの書き入れ
- S 集付 [位置]
本行の末尾に付いた歌集名・部立などの書き入れ
- r 右傍書 [位置][よみ]
本行の右側の書き入れ
- l 左傍書 [位置][よみ]
本行の左側の書き入れ
- f 補入 [位置][よみ]
脱落した文字の本行右書き入れ
- F 補入 [位置][よみ]
脱落した文字の本行左書き入れ
- m ミセケチ [位置][よみ]
見セ消チ記号の位置表示、よみの生成もする
- k 重ね書き [位置][よみ]
重ね書きされた下の文字
- c 擦り消し [位置][よみ]
擦り消された文字
- x 墨滅 [位置][よみ]
墨で消された文字
- h 圈点 [位置]
挿入・書き換え位置を示す点
- b 傍線 [位置]
文字に付いた傍線
- e 文字の除棄 [よみ]
書き入れの指示で文字を除棄する
- z 左注由来の異文 [よみ]
左注の内容からよみを生成する

B) 数字と括弧、および“|”(縦線)

数字は該当する句と、句中の何文字目かを示す。二種類の括弧“[]”“{}”と組み合わせて記述する。括弧の中を“|”(縦線)で区切り、縦線の前は表示位置、後はよみの置き換え箇所を示す。これを、先に示した英字と組み合わせて記述する。

以下、レコード7行目に記述する、歌に関する書き入れの例を示す。

(例) 2[3|4-6]r{桜|さくら}

第2句の3字目の右に「桜」の書き入れあり。

歌本文Bの第2句の4字目から6字目を「さくら」に置き換え。

ただし、“h” “f” “F” が関係する場合は、「当該字の後に」の意とする。

(例) 3[2]f{あ}

第3句の2字目の右下に「あ」の書き入れあり。歌本文Bの当該箇所に「あ」を挿入。

なお、このように、表示箇所とよみの反映箇所が同じ場合は、縦線とその後の内容を省略することができる。

その他、句をまたがる書き入れなど複雑な例があるが、ここでは割愛する。

3-4. 漢字と異体字の扱い

各伝本の表記は、仮名にしても漢字にしても、現代の通行字体とは異なる事例が多い。もちろん、これは周知の事実であるが、こうした資料のデジタルデータ化には、常に、どの文字で表記するかという問題がつかまとう。

たとえば、仮名の場合、現代の通行字体とは異なる字体が多く用いられるが、字母が「安」でも「阿」でも区別せず、「あ」とするのが通常処理であろう。仮名字母の研究をするのでなければ、それらを区別して記述する必要はないという判断である。

漢字に関しても、異体字関係にある字をすべてまとめて、同一の字体で記述しておく、という立場もありえよう。検索の便を考えると、さまざまな字体を想定する必要があるという利点がある。しかし、漢字に関しては、字体の顕著な違いは、区別して記述しておくほうがよい場合がある。ある漢字の字体の違いが、表記から見る伝本間の近似性や、書写者の属する文化圏を示すということもありうるからである。

具体例を示そう。「声」という漢字は、「聲」とは異体字関係にあり、コード上でも区別して入力しておくのは容易である。ところがその一方で、

右に示すような字体、つまり最終画を左に払うのではなく、右に曲げてはねた形の「声」がある。これを通行字体の「声」と同一視して、一律に「声」と記述する立場もあろうが、伝本での出現箇所を見ると、その出現箇所に明らかな傾向が看取される。

最終画を右に曲げてはねた「声」は、内閣文庫蔵「江雲渭樹」印本の第五帖2738番歌以降、同本の第六帖に集中して分布する。しかも、この本とごく近い関係にあると思われる島原図書館蔵松平文庫本でも、同じ字体の「声」が、同様の箇所に使用されている。このような現象は、今後の伝本研究に何らかの示唆を与えてくれる可能性がある。そこで、「声」と区別するために「@声」と記述して差異を明確にした。

このように、漢字の異体字については、慎重に取り扱い、可能なかぎり字体を区別するという方針で、作業を進めている。

なお、使用する漢字の字体は、互換性を重視して、JIS X 2080 の範囲、すなわち第2水準までに今のところ限定している。そのため、第3水準・第4水準、またそれ以外の字体の漢字については、“@”との組み合わせ表記とした。たとえば、「秋」の異体字「𠂔」は頻繁に使用されているが、これも「@秋」と表記する。



内閣文庫蔵「江雲渭樹」印本
2738 番

4. 解析ソフトウェアツールの設計

今回作成したデータは、1レコード7行からなるテキストデータであるが、通常と異なり、以下の特徴をもつ。

- (1) 同じ作品についての複数の伝本である。
- (2) 書き入れデータを含む。

(1)に示されるとおり、伝本による異同をみるのが本データ作成の主たる動機の一つである。このために、各レコードには『新編国歌大観』による歌番号が付してある。検索結果などを表示する際には、基本的に、同じ歌番号をもつレコードを複数の伝本にわたって同時に表示することになる。しかし、① 同一の歌番号をもつレコードが同じ伝本内に複数存在すること、② 一つのレコードが複数の歌番号をもつこと、により、インタフェース

がやや複雑になる。いずれにせよ、これ自体に特に技術的な難しさはない。

(2)によって、以下の二つの機能が必要となる。

- A) 書き入れを忠実に再現したテキスト表示機能。
- B) 書き入れによって指示される異文を反映したキーワード検索機能。

A)は、影印をイメージデータとして画面で確認できることが理想であるが、権利面の問題から現実的ではない。したがって、どの程度影印に近い表示で折り合いをつけるか、という問題といえる。本研究ではこの機能の充実には深入りしない。

一方、B)は、非常に興味深い。書き入れによるテキストの異文の指示は、①文字列の挿入、②文字列の削除、③文字列の置換の3種類に分けられる。このうち②③は対象となるテキストの範囲の指示を伴うが、この範囲はネストすることがあるばかりでなく、ネストせずにオーバーラップすることもある。したがって、このような書き入れ指示をXMLなどのような括弧構造としてタグ付けで表現することはできない。

このようなテキストは、文字列処理の分野では、ハイパーテキストとしてモデル化されている[1]。ここでいうハイパーテキストとは、以下のように形式的に定義される。

【定義】ハイパーテキストとは、頂点ラベル付き有向グラフ $G=(V, E)$ で、頂点ラベルが任意の長さの任意の文字列であるものをいう。

さて、ハイパーテキストに対する文字列照合問題を解くことを考えよう。ハイパーテキストが DAG であるとき、すなわち、サイクルを含まないとき、ハイパーテキストの表現するテキストの集合は有限集合である。本研究で作成するデータに対応するハイパーテキストは DAG となるため、これを展開して通常のテキスト集合にして扱う方法が考えられる。しかしながら、この方法では、テキスト集合のサイズ(テキストの個数)が指数的に増大しうる。実際に展開してみたところ、データサイズは 100MB を超してしまった。したがって、このようなナイーブな手法は現実的ではない。

ハイパーテキスト上の文字列照合問題に関しては、次の結果が知られている。

【定理】 ([1]) ハイパーテキストに対する文字列照合問題は、 $O(N+m|E|)$ 時間で解くことができる。ここに、 N はテキスト長の総和、 m はパターン長とする。

[1]の手法は KMP 法に基づいており、DAG に限らず一般の有向グラフに適用可能である。しかし、

今回のデータは DAG であることから別の方法を採った。すなわち、パターンに対する NFA をビット並列化手法により実装[2]し、DAG の頂点をトポロジカルソートした上で順次動作させることにした。文字列照合に要する時間は、 $O((m/w)(N+|E|))$ である。ここに、 w はワード長を表わす。特に、パターン長 m が w 以下であるとき、計算時間は $O(N+|E|)$ となる。

以上述べたように、本ツールは歌番号による検索とキーワードによる検索の機能を持ち、キーワード検索では、検索結果の中からユーザがレコードを選択することにより、同じ歌番号をもつレコードが伝本ごとに表示される。その際、基準において伝本と比較した歌の異同の有無が句単位で色表示される。

さらに研究者向けの機能として、基準と定めた伝本と比較した伝本ごとの異同のパターンごとに統計をとる機能なども順次整備している。

5. デジタル校本を用いた文学研究

5-1. 『古今和歌六帖』の本文研究

『古今和歌六帖』は約 4,500 首という多くの和歌を収めている。その上、類題和歌集であることから、一首一首の和歌の出典考証を行う必要もある。したがって、関連データは膨大な量にのぼり、しかも、相互に複雑な関係性をもっている。

本研究で開発しているデジタル校本は、二つ以上の諸本間において、本文異同の存する歌の数と具体例を表示することができる。

一例を示そう。『古今和歌六帖』諸本は、いずれも藤原定家所持本を源家長が書写・校合した本の系統であると考えられるが、詳細にみると、写本系と版本系に分けられることが、すでに指摘されている(注7)。そこで、写本系本文の桂宮本と、寛文九年版本との本行の本文を比較してみる——この場合は、和歌データにおける歌本文 B を分析対象とし、表記上の異同ではなく、歌の意味に関わる、語句の異なりをもつ歌を抽出した——と、明らかな脱字・衍字も含めて、異同を有する歌は全体の 30 パーセントあまりを占めることがわかる。

この機能は、『古今和歌六帖』諸本本文の全体像を把握するために、きわめて有効であろう。『古今和歌六帖』のような大部な作品を対象とする文学研究は、往々にして作品の特定の部分に着目して行われることがあるが、大まかな全体像を把握した上で個別の事象に向かうことは、研究視点の

客観性を保証するという点においても、重要であろう。

また、異同を有する歌のみを抽出・列挙するデジタル校本の機能を用いることによって、具体的な異文調査を能率的に進めることができる。前述の桂宮本と寛文九年版本との異同においても、1,400首あまりの異文のある歌を抽出した。その中には、次のような歌もある。

よも山を うちこえくれは かさぬひの
しまこき帰る たなゝしを舟 (桂宮本)
しはつ山 打越くれは かさぬひの
嶋こきかくる たなゝしをふね
(寛文九年版本)

これは、『古今和歌六帖』第三帖、1817番歌であるが、出典は、『万葉集』巻第三、272(274)番(注8)と考えられる。

四極山(シハツヤマ) 打越見者(ウチコエミレバ)
笠縫之(カサヌヒノ) 嶋榜隠(シマコギカクル)
棚無小船(タナナシワブネ) (注9)

初句の「四極山(シハツヤマ)」は、『古今和歌六帖』においては、寛文九年版本と一致するが、桂宮本の「よも山」も、『万葉集』の古写本である類聚古集・神田本に見え、細井本・京都大学本は、漢字表記の左に「よもやま」とある(注10)。とすれば、『万葉集』の古写本の本文と一致する桂宮本文も、故なしとしないことになる。

ところで、古典和歌の本文を検索する際、今日一般に用いられるのは、『新編国歌大観』である。そこに収載されている『古今和歌六帖』テキストの底本として採用されたのが桂宮本であったことはすでに触れた。また、それ以前に、『古今和歌六帖』の本文検索をする場合には、『続国歌大観』が用いられたが、本文が、江戸期の流布本である寛文九年版本系統であることも、前述のとおりである。

『続国歌大観』と『新編国歌大観』とで、それぞれ「しはつやま」を検索すると、『古今和歌六帖』においては、各一首の、まったく別の歌が見出される。すなわち、『続国歌大観』では、寛文九年版本の系統の本文であるため、前掲の歌が検出されるが、桂宮本を底本とする『新編国歌大観』では、当然この歌は該当せず、第二帖1425番歌が浮かび上がる。

しはつ山 いはさかはかり われのれる
むまそつまつく 家こゝぬらし
(桂宮本)

しほつ山 打越ゆけは 我のれる
馬そつまつく 家こふらしも
(寛文九年版本)

ちなみに、この歌の出典も『万葉集』であり、以下のように、初句の本文は、寛文九年版本と一致する(注11)。

塩津山(シホツヤマ) 打越去者(ウチコエユケバ) 我乗有(ワガノレル) 馬曾爪突(ウマゾツマヅク) 家恋良霜(イヘコフラシモ)

『万葉集』巻第三、365(368)番

したがって、この歌は、『続国歌大観』では、「しはつやま」と検索しても用例として出てこない。

『校本万葉集』に拠っても、初句について、『万葉集』諸本に異同はない。『古今和歌六帖』の「しはつ山」(桂宮本)、「しほつ山」(寛文九年版本)の本文異同が、『万葉集』本文の揺れによるものではないとすれば、次には、「は」と「ほ」という平仮名字体の類似による可能性を疑ってみねばなるまい。『古今和歌六帖』諸本間には、第二句に大きな本文の対立があり、それも含めたさらなる本文の検討が必要であろう。

なお、『古今和歌六帖』と出典となる歌集との間には、複雑な本文異同が存在することがある。とくに『万葉集』を出典とする場合には、それが顕著になる場合が多い。そこで、文字列解析器 eCSA (efficient character string analyzer) Ver.2.00 (竹田正幸開発)の類似歌抽出機能を用いて、『古今和歌六帖』約4,500首すべての出典考証を行い、一通りの調査を終了している。両者のデータを組み合わせることにより、『古今和歌六帖』諸本の性格を知るための、より詳細なデータが得られる。

5-2. 『古今和歌六帖』の表現研究

(大伴宿祢家持更贈・紀女郎 歌五首)

野干玉能(ヌバタマノ) 昨夜者令還(ヨフベハ
カヘル) 今夜左倍(コヨヒサヘ) 吾乎還莫(ワ
レヲカヘスナ) 路之長手乎(ミチノナガテヲ)
『万葉集』巻第四、781(784)番

大伴家持が紀女郎に贈ったこの万葉歌は、『古今
和歌六帖』第五帖、3023 番にも載っている。

むはたまの よは人^たはかへる こよひさは
われをかへすな うちのたま姫 (桂宮本)
うは玉の よむへは帰る 今宵さは
我をかへすな 道のなかくてを
(寛文九年版本)

『古今和歌六帖』における桂宮本と寛文九年版本
の結句の異同は甚だしい。だがここでも、出典の
万葉歌と本文が一致するのは、寛文九年版本であ
る。

一方、桂宮本の結句「うちのたま姫」の例は、
『新編国歌大観』を検してもきわめて少ない。だ
が、先行例として、『古今和歌集』異伝歌を挙げ
ることができる。

さむしろに 衣かたしき こよひもや
我をまつらむ うちのたまひめ
又は、うちのたまひめ
『古今和歌集』巻第十四恋歌四、689 番(注 12)

ここで想起されるのが、『古今和歌六帖』に収
められた『万葉集』巻十二の歌で、異伝注記のあ
る歌は、すべて異伝歌が採られているという指摘
である(注 13)。当該歌では、桂宮本の結句のみのこ
とではあるが、『古今和歌集』の異伝歌に共通す
る歌語を用いていることは、きわめて興味深い。

ところで、「うちのたまひめ」は、『古今和歌
集』では、男性が通ってくるのを待つ女性のイメ
ージである。ところが、『古今和歌六帖』の場合、
題は「くれとあはす」であり、男性の来訪を拒絶
する女性を描く。そもそも、この歌は紀女郎にあ
てた歌であるが、「うちのたまひめ」と紀女郎とを
特に結びつける要素は、今のところ見出せない。
とすれば、ここは、『古今和歌六帖』桂宮本が、『万
葉集』の詠歌状況から離れ、さらに、『古今和歌集』
異伝歌の歌語を用いながら、そのイメージからも
脱して、新たな「うちのたまひめ」歌の世界を
構築しようとした結果生まれた本文を掲載したと
見られよう。

では、桂宮本の本文は、いついかなる事情で生
まれたのか。この点は、にわかに決しがたいところ
ではある。だが、出典と見られる万葉歌の結句
に比し、これほどまでの異文が生じているという
ことは、やはりそこに、偶然ではない意図的な作
為が働いたのではなかったか。

歌語をめぐるこのような模索の跡は、十世紀後
半から十一世紀において見られるところである(注
14)。『古今和歌六帖』編纂時に、当該歌が、桂宮
本本文のかたちですでに流布していたのか、ある
いは、『古今和歌六帖』編纂時の本文改変であっ
たのかは即断し難いが、当時の和歌表現の特色を
示す事例として、留意しておくべきなのではない
だろうか。

6. 今後の課題

現存する伝本は、その本がいったん書写されて
から、さらに文字が書き入れられ、あるいは摺り
消されて今日に至っているものが多い。また、そ
の親本が、どのような本の写しであったのか、明
確であることのほうが稀である。

そのような状況にあって、諸本の墨付きすべて
を、本行と書き入れという階層に分け、検索する
のみならず、諸本間の異同のパターンによって、
用例を抽出するデータマイニング機能を開発する
ことは、実証的な文学研究に寄与するところきわ
めて大であろう。

ただし、伝本の原態を機械可読化するためには、
伝本によって性格の異なる、様々な事象に対応し
ていかなければならない。それまでに定めた表記
規則にあてはまらない要素が見出された場合は、
新たなタグを付け加えるか、あるいは、既存の表
記規則の枠組みを変更することになる。

たとえば、朱筆をはじめとする書き入れの色情
報についても、今後、電子テキストに盛り込む必
要がある。だが、この点については、新たなタグ
を拡充することにより、比較的容易に対応できる
であろう。

また、和歌本文を一行書きにしているか、それ
とも二行書きかという点は、伝本研究の際に、か
なり重要になってくることがある。

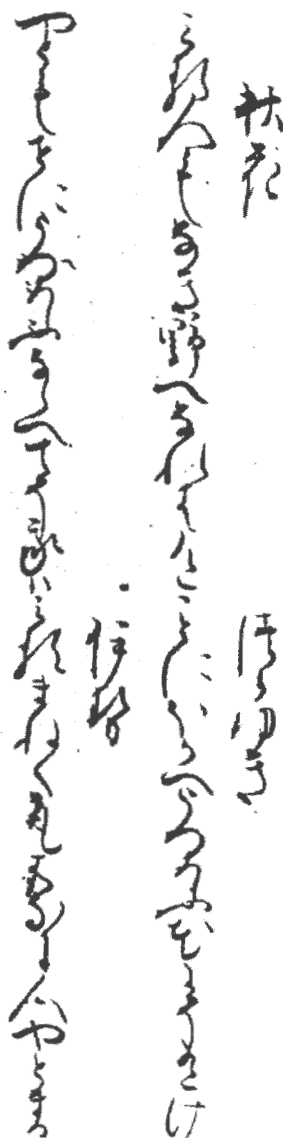
次ページの影印は、桂宮本の第六帖、4057・4058
番の歌である。以下に本文を示すと、次のように
なる。

秋花 つらゆき
みる人も なき野へなれは 色ことに

ほかへうつろふ 花にそ有ける(4057番)

伊勢

やともせに うつろふならへてそ 我はみる
まねく尾はなに 人やとまると(4058番)



桂宮本 4057・4058 番

なお、4058 番歌、第二句の見せ消チ一文字目の右横には、「へ」という傍書がある。見せ消チと傍

書をあわせると、第二句は「うへならへてそ」という本文になる。

4058 番歌が、第二句を誤って「うつろふ」と書写してしまった原因を考える際に、この歌の直前に配されている 4057 番歌の第四句「ほかへうつろふ」に目移りしてしまったのではないかという推定は、それほど的外れではないだろう。そして、この桂宮本の親本が和歌二行書きだったとすれば、4057 番歌の第四句と 4058 番歌の第二句は、まさに隣り合わせの位置に書写されていたことを想定しうる。このようなアプローチを可能にするためには、和歌の書式は、ぜひデータ化したい情報の一つに数えられよう。

さらに、現存『古今和歌六帖』には収載されないが、他の歌集や歌論書に『古今和歌六帖』歌として載っている歌もある。たとえば、山本明清校注『古今和歌六帖標注』(1831 年成、1832 年刊)には、「古今六帖拾遺」として、47 首というまとまった数の歌を列挙する。現存『古今和歌六帖』に、少なからぬ歌の脱落があるとするならば、その復元のために(注15)、それらのデータも、今後、増補していく必要があろう。

デジタル校本は、いったん校本画面を構築した後でも、伝本の入力データの変更・増補をおこなうことができる。校本を用いて研究を進めながら、データの質の向上と充実を図ることが可能である点、紙媒体での従来の校本とは一線を画すといえるであろう。今後も、デジタル校本の枠組みの構築とその整備に努めると同時に、文学研究に活用していきたい。

註

- (1) 近年においては、真野道子氏「定家の源氏注釈における万葉歌」(『中古文学』第78号、2006年12月、pp. 69-82)をはじめとする論考がある。
- (2) 『新編国歌大観』古今和歌六帖解題(橋本不美男氏・相馬万里子氏・小池一行氏)。
- (3) 以下、本書の引用は、宮内庁書陵部蔵桂宮本の影印紙焼資料に拠る。
- (4) 上・中・下・研究編、全4巻。昭和35年3月・9月・12月、昭和36年10月刊、風間書房
- (5) 文部科学省私立大学学術研究高度化推進事業(オープン・リサーチセンター整備事業)「人文科学におけるマルチメディア情報通信技術の教育利用に関する共同研究」、2008年度早稲田

大学戸山リサーチセンター個別研究課題「オンデマンド授業システムの共同研究基盤への応用研究」研究成果(申請者・早稲田大学文学学術院兼築信行氏)。

- (6) 科学研究費補助金基盤研究 (C) 「文字列データ解析システムの構築と平安中期歌語生成に関する研究」(2007 年度～2009 年度、課題番号: 19500217、研究代表者: 福田智子)による研究成果の一部である。テキスト作成作業には、研究分担者(連携研究者)および研究協力者にご尽力いただいた。
- (7) 『図書寮叢刊 古今和歌六帖』下巻「古今和歌六帖」解題、321 ページ参照。
- (8) 『万葉集』の歌番号は、旧番号(新番号)の順に示した。
- (9) 以下、『万葉集』本文の引用は、とくに断らない限り、『新編国歌大観』に拠る。
- (10) 『校本萬葉集』による。なお、京都大学本は、緒のカタカナ表記。
- (11) 寛文九年版本の本文が『万葉集』本文と一致する場合は多い。版本が、出版に際し、『古今和歌六帖』本文を整備したであろうことは、富永洋子氏「古今和歌六帖の研究—細川家永青文庫本及び松平文庫本を中心として—」(『国語と国文学』第 42 巻 1 号、1965 年 1 月)、岸上慎二氏「古今六帖本文覚え書—写本の形による読みⅡ—」(『語文』〈日本大学国文学会〉第 67 号、昭和 1987 年 3 月)などによって、その可能性が指摘されている。
- (12) 『古今和歌集』本文の引用は、『新編国歌大観』に拠る。
- (13) 新沢典子氏「古今和歌六帖と万葉集の異伝」(『日本文学』〈日本文学協会〉第 57 巻第 1 号、2008 年 1 月)。なお、『古今和歌六帖』本文は、桂宮本に拠る。
- (14) 福田智子「平祐舉の和歌——一条朝和歌の側面——」(『和歌文学研究』第 75 号、1997 年 12 月。後に『平安中期私家集論—歌人・伝本・表現—』(2007 年 2 月、勉誠出版)では、「うちのはしひめ」から「吹きではまのはまひめ」という語句が生み出された可能性を指摘した。
- (15) 黒田彰子氏「古今和歌六帖と和歌童蒙抄—古今和歌六帖本文の復元をめざして—」(『愛知文教大学論叢』第 12 巻、2009 年 11 月)他の論がある。

参考文献

- [1] A. Amir, M. Lewenstein, and N. Lewenstein. Pattern matching in hypertexts. *J. Algorithms* 35(1):82-99 (2000).
- [2] G. Navarro and M. Raffinot. *Flexible pattern matching in strings: practical on-line algorithms for texts and biological sequences*. Cambridge Univ. Press, 2002.

謝辞

資料をご提供くださった宮内庁書陵部・永青文庫・国立公文書館・肥前島原松平文庫・ノートルダム清心女子大学付属図書館、および国文学研究資料館に厚く御礼申し上げます。

附記

本稿は、科学研究費補助金基盤研究 (C) 「文字列データ解析システムの構築と平安朝文学の伝本と表現に関する総合的研究」(課題番号 22500236、平成 22～24 年度)における研究の一部である。

特別セッション

「人文科学とデータベースの教育に関わる現状と課題」ラウンドテーブル

このシンポジウムを企画・支援する人文系データベース協議会は、人文科学におけるデータベースの構築と応用を中心とした研究分野における学術的交流を目的とした大学間の連絡協議会である。この協議会を中心とするシンポジウムも回数を重ね、多くの成果を得るにいたった。しかしながら、いままで、この協議会で大学のもうひとつの機能である教育を正面からとりあげてくることは、多くなかったのではないだろうか。

無論、それは、人文系データベース協議会が、この問題を捨象してきたわけではない。人文系データベース協議会が立ち上がってきた当初は、この領域そのものが黎明期であり、まとまった教育などを行える環境が整わなかったためだと理解している。

その後、この「人文科学とデータベース・人文科学とコンピュータ」の領域が成熟してくるにつれ、この領域への理解も進み、多くの研究者を輩出し、すばらしい成果が多く生まれるようになってきた。その成果を踏まえ、日本全国、とりわけ、この人文系データベース協議会の「中心」である関西の多くの大学で、研究組織に加えて教育組織が立ち上がるようになってきた。それは、長く協議会に関わってきた人々の尽力の一つの形である。

教育組織が多く立ち上がり、研究者の「再生産」を行える体制が整いつつあるということは、ひとつの領域がひとつの「学問分野」へと成熟していく過程であるといえるだろう。その教育組織の体制が整い、再生産への道筋が整いつつある一方で、課題が見えてくる時期にもなってきた。

そこで、このラウンドテーブルでは、それぞれの大学の特色・成果の報告と課題をあぶりだすことを目指す。とりわけ、この領域の教育に関わる共通の悩みとそれを解決する／してきた方法などを忌憚なく討論し、相互に協力し合う体制を見つけることで、より新しいステップに進めるのではないだろうか。文理のバランスをどのように考えるか、キャリアパスの課題、人材の確保、最終的にはどのような人材を輩出することを目指すのか、など多くの課題について議論できれば幸いである。

(文責・後藤 真)

1. 花園大学 情報歴史学 【師 茂樹】

授業科目

- ・情報歴史学研究Ⅰ・Ⅱ (2回生以上、それぞれ前後期2単位)
- ・情報歴史学概論Ⅰ・Ⅱ (全学年、それぞれ前後期2単位)
- ・情報歴史学実習Ⅰ・Ⅱ (2回生以上、それぞれ前後期2単位)
- ・文化遺産学研究入門演習 (2回生、2単位。前期と後期で学科内の異なるコースを履修)
- ・文化遺産学演習 (3回生、4単位) など。

2009年3月に刊行した後藤真・田中正流・師茂樹『情報歴史学入門』(金壽堂出版)は、2002年に始まった情報歴史学コースにおける教育の実践をまとめたものであるが、ここでは大きく分けて以下の2つの能力の養成を目指している。

1. デジタル化された歴史情報に対するリテラシー能力
2. 学生が自ら問題を見出し、卒業制作と卒業論文によって解決する能力

1については、既存の歴史学のためのデータベースや文化遺産のデジタルアーカイブなどを批判的に「読む」(評価する)とともに、それがどのような目的・方法で作られたものなのかを読み解く訓練を行う。また、様々なデジタル化の技法を学びつつ、それらの一つを用いて実際に対象(歴史学的な知識や文化遺産)をデジタル化する実習を行っている。

2については、他のディシプリンにおける卒業論文指導などと大きな差はないと思われるが、情報歴史学では特に、コンピュータを使ってどのような貢献ができるのか(自己満足にならないか)という点と、

「対象」と「方法」を区別して先行研究調査などを行う点を指導している。また、成果物の公開などにあたっては、対象となる人々に対してそれがどのような影響を持つのか等、デジタル化や公開という活動がもつ社会的な影響力についても考えるよう指導している。

【参考文献】

- ・師茂樹「情報歴史学の課題 花園大学・情報歴史学コースのための主観的航海図」(『漢字文献情報処理研究』第3号、2002年)
- ・佐藤誠・田中正流「花園大学「情報歴史学コース」の現状と課題」(『漢字文献情報処理研究』第4号、2003年)
- ・師茂樹「情報歴史学のこれから 花園大学・情報歴史学コースの4年間をふり返って」(『漢字文献情報処理研究』第7号、2006年)
- ・師茂樹「情報歴史学の教育に挑む」(『歴博』第140号、2007年)

氏名：師茂樹

所属：花園大学文学部

主要業績：

- ・師茂樹「GraphText 紙テープに呪縛されないテキストデータの試み」(『漢字文献情報処理研究』第10号、2009年10月)
- ・師茂樹「携帯電話の絵文字における semantics の問題」(『東洋学へのコンピュータ利用 第21回研究セミナー』、2010年3月)
- ・師茂樹「^{メモリア}「公共の記憶」としての電子書籍」(『ユリイカ』2010年8月号、青土社、2010年7月)

2. 同志社大学 文化情報学 【阪田真己子】

氏名：阪田真己子

所属：同志社大学文化情報学部

主要業績

- ・「学びと身体空間」,『「学び」の認知科学事典』(佐伯胖監修・渡部信一編,大修館書店,共著(2009))
- ・Motion Feature Quantification of Different Roles in Nihon-Buyo Dance,"Advances in Human-Robot Interaction ", Edited by Vladimir A. Kulyukin,INTECH, (2010)
- ・Comparison of Differentiation of Basic Dance Motions of Nihon-buyo Using Motion Capture,"Proceedings of the International Conference on Humanized Systems", (2010)

3. 大谷大学 人文情報学 【柴田みゆき】

授業科目

- ・情報社会の権利と法律 (現在担当)
- ・情報産業論 (過去担当)
- ・マルチメディア論 (過去担当)
- ・文化財映像化演習
- ・人文学テキスト処理論
- ・プログラミング演習
- ・データベース演習

1年次に対しては過去、Windows OS を中心としたPC活用能力を学ぶ週2回の実習「専門の技法」を担当した。2年次は現在、「人文情報学演習II」を担当している。学生のコミュニケーション能力向上

のため、前期はディベート能力、後期はプレゼンテーション能力の強化に特に力を入れている。これらの経験から、総合的な情報活用能力を客観的に判断可能な試験を2年生に課す提案を行い、2年前からRasti試験が導入されることとなった。この試験の成績をゼミ選択に反映させることで学生の勉強と受験の動機づけを行い、同時に学生に外部評価を自覚させることで向後の勉強の動機づけを行っている。

3・4年次は、学科開設以来継続してゼミを担当している。コンテンツ（動画）製作とその法的問題点がテーマである。大きな問題点は、参加学生の学力により、毎年、技術を教えるか、対象の内容を深めるかのどちらかに偏りがちになることである。卒論は映像、オンラインゲーム、コミュニケーション、法律に関わるテーマが多い。

氏名：柴田 みゆき

所属：大谷大学文学部

主要業績：

- ・柴田みゆき, "資料空間から任意の関心領域を柔軟に抽出・提出する新たな方法論の研究", 大谷学報 第88巻第1号, 大谷学会, 2008-12-10.
- ・Seiji Sugiyama, Atsushi Ikuta, Miyuki Shibata and Tohru Matsuura, "An Event Oriented Data Management Method for Displaying Genealogy: Widespread Hands to InTErconnect BASic Elements(WHITEBasE), The 9th International Conference on Computer Information Systems and Industrial Management Applications(CISIM), 2010-10-10.
- ・柴田みゆき, et. al., "北京版チベット大蔵経の高再現性デジタル画像化：写真撮影過程", 情報処理学会研究報告, 人文科学とコンピュータ研究会報告 98(97), 73-80, 1998-10-23.
- ・京都ソフトウェアリサーチ メディアデザイン, "GNU Emacs 入門", Login:UNIX シリーズ, オーム社, 1994-3.
- ・柴田みゆき, 高橋真共編, "情報リテラシーの基礎【2010年版】(II)", 2010-9.

4. 立命館大学 デジタル人文学 【河角龍典】

授業科目

- ・文学部副専攻イノベーションプログラム「デジタル人文学」コース（2010年新規開講）
- ・「コンピュータグラフィックス基礎演習Ⅰ（2次元CG）」（2回生・前期・2単位）
- ・「コンピュータグラフィックス基礎演習Ⅱ（2次元CG）」（2回生・後期・2単位）
- ・「WEB技術基礎演習」（2回生・後期・2単位）
- ・「テキスト情報処理演習」（3回生・前期・2単位）
- ・「デジタルアーカイブ演習」（3回生・後期・2単位）
- ・「空間情報処理演習（GIS）」（3回生・後期・2単位）
- ・「コンピュータグラフィックス演習Ⅰ（3次元CG）」（3回生・前期・2単位）
- ・「コンピュータグラフィックス演習Ⅱ（映像）」（3回生・後期・2単位）

※2回生配当科目の3科目6・単位と、3回生配当科目の2科目4単位以上を選択し、計5科目10単位以上の単位取得が修了要件となる。

立命館大学文学部では、2010年から副専攻イノベーションプログラムとして、「デジタル人文学」コースを開講した（定員30名）。このコースは、文学部で教養科目として開設される情報系科目30科目とは別に提供される科目である。「デジタル人文学」コースでは、Bに示したとおり合計8科目が開講され、人文科学の分野に必要な情報技術を実践的にかつ系統的に履修できる。各科目の担当教員は、現在のところ人文科学出身の教員で構成されている。

1回生時の12月までに志望理由書を執筆し、受講登録にエントリーする必要があるため、専攻分野とコース提供科目との関係の理解が容易になるように「テキスト・画像情報処理系」、「空間情報処理系」、

「ビジュアル情報処理系」3つの履修モデルを提案している。しかし、2回生時には、自分自身が行う研究に情報技術がどのように活かされるのかということが、明確にならないまま履修している学生が多いのが現状である。

氏名：河角龍典

所属：立命館大学文学部

主要業績：

・河角龍典（単著）「バーチャル長岡京・平安京 3D マップ」，（国立歴史民俗博物館編『桓武と激動の長岡京時代』，山川出版社，2009 年，所収）

・河角龍典（単著）「平安時代のバーチャル京都」，（矢野桂司・中谷友樹・磯田弦編『バーチャル京都』，ナカニシヤ出版，2007 年，所収）

・河角龍典・原澤亮太・吉越昭久（共著）「中世京都の地形環境変化」，（高橋康夫編『中世都市研究 12 中世のなかの「京都」』新人物往来社，2006 年，所収）

第 16 回公開シンポジウム実行委員会

実行委員長 師 茂樹（花園大学）
実行委員 後藤 真（花園大学）
実行委員 出田和久（奈良女子大学）
実行委員 川口 洋（帝塚山大学）

人文系データベース協議会

運営委員

奈良女子大学 出田和久（議長）
帝塚山大学 川口 洋（副議長）
大阪電気通信大学 加藤常員（副議長）

正会員（氏名 50 音順）

関西大学 江澤義典
総合研究大学院大学 及川昭文
大阪国際大学 桶谷猪久夫
花園大学 後藤 真
京都大学 柴山 守
大阪樟蔭女子大学 高橋晴子
静岡大学 中谷広正
立命館大学 八村広三郎
京都大学 原正一郎
大阪工業大学 深海 悟
京都工芸繊維大学 宝珍輝尚
同志社大学 村上征勝
神戸大学 森下淳也
花園大学 師 茂樹

事務局

大阪電気通信大学 小沢一雅
大阪電気通信大学 小森政嗣

本論文集の発行にあたっては、花園大学「学内で開催される学会等への助成」を受けた。

本論文集は、日本学術振興会・科学研究費補助金「東アジア仏教論理学史研究のための逸文データベースの構築」（研究課題番号 20720013）、「写真史料学構築のための情報学的手法の研究—上田貞治郎写真コレクションを題材に一」（課題番号 21700280）による成果の一部である。

第 16 回公開シンポジウム「人文科学とデータベース」論文集

2010 年 11 月 27 日 発行

編集・発行 第 16 回公開シンポジウム「人文科学とデータベース」実行委員会

印刷 株式会社テイスト（京都市中京区西ノ京南円町 83）