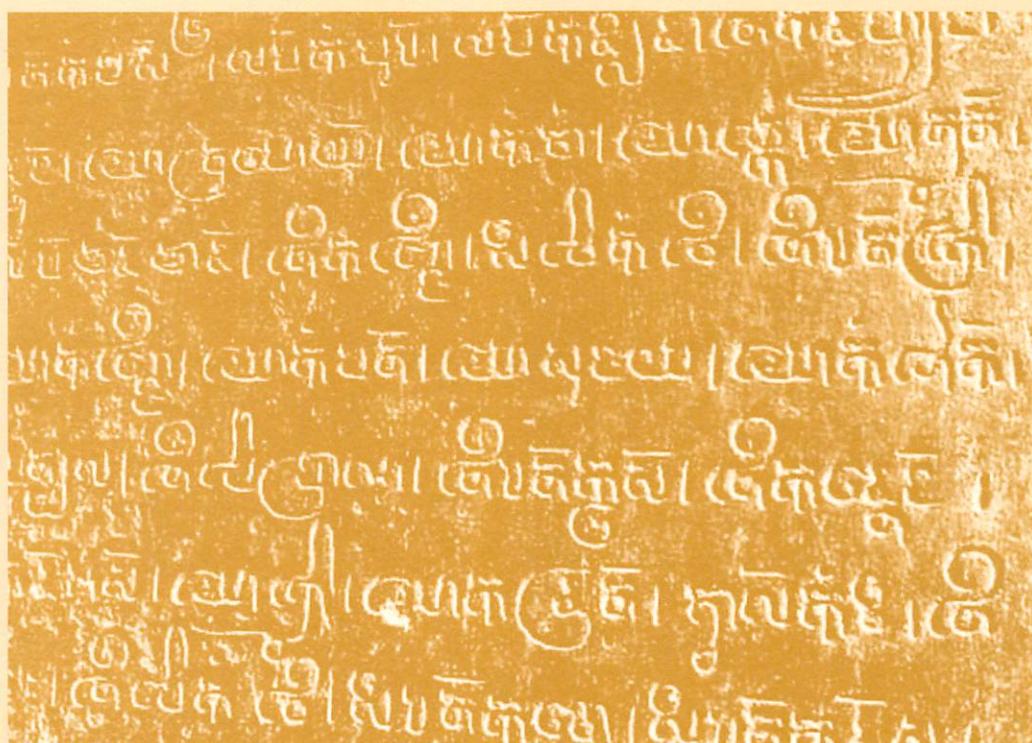


# 論 集

# 情 報 处 理



平成10年度～14年度 文部科学省科学研究費補助金 特定領域研究(A)118

「古典学の再構築」研究成果報告集Ⅳ

A03 「情報処理」班研究報告

IV

平成10～14年度 文部科学省科学研究費補助金 特定領域研究（A）118

「古典学の再構築」研究成果報告集 IV

A03「情報処理」調整班研究報告

論集

「情報処理」

神戸 平成15年3月



表紙：カンボジアのアンコール王朝のインドラ・ヴァルマン王（877-889）が建てた  
ブレア・コー寺院のサンスクリット語碑文。「ブレア・コー」とは「聖なる牛」  
という意味で、この寺院の前面に座す牡牛の像からそう呼ばれているが、「コー」  
というのもサンスクリット語の go (牝牛) に由来している。東南アジア全域  
におけるインド文化の伝播は、古くはサンスクリット語、その後はパーリ語と、  
古典語の力によって担われたものである。

撮影=高島 淳

## はじめに

A03「情報処理」班研究成果報告編集担当 高島 淳

本報告書は、文部科学省科学研究費補助金特定領域研究（A）「古典学の再構築」（平成10年度～14年度）のA03「情報処理」班の研究成果をまとめた論文集である。

「情報処理」班は、古典諸学の連携と高度情報処理技術の応用によって古典学の刷新を図るという「古典学再構築」の大きな目的の中で、特に情報処理を研究の中心に据える研究者を集めた班である。組織としては、前期2年間においては、計画研究が4研究（参加者計20）公募研究が3研究（参加者計5），後期2年間においては、計画研究が4研究（参加者計18）公募研究が2研究（参加者計5），通算31名の研究者によって行われたものである。

情報処理班では、できる限りネットワークを活用するという方針で、研究集会を2回しか開催していないが、その際の報告も2編（一つは外部から参加していただいた永崎研宣氏のもの），本報告書に納められている。

また、公募研究のうちの「抄物の原典参照データベースの構築——『韻府群玉』と『玉塵抄』を例として——」の出雲朝子氏の研究班は、B02「伝承と受容（日本）」班の研究テーマとの関係が深く、これまで両研究班に同時に属しているような形であったため、出雲氏の論文は、読者にとってよりまとまったものとして読めるだろうという配慮から、B02班の報告書に掲載することとした。本報告書には氏の研究分担者である豊島正之氏の報告のみを掲載してある。

本書に集められた諸研究のテーマをおおまかに分類してみると、計量文献学、データベース、画像および文書情報処理といったところが中心とであり、対象分野からすると、日本とインドが中心となっている。これは現在の人文科学における情報処理の利用の現状の反映であるとも言えよう。

昨年のノーベル賞を受賞した二人の日本人は、分野こそ異なれ、いずれも研究のツールを開発した方々であった。共同研究が中心となることが多い自然科学においてはそういう業績を高く評価するという潮流があるが、人文科学においてはいまだにそうした考え方方は稀である。

人文科学においても、データベースの蓄積とその利用によって研究の次元を一段階高めるような研究がようやく始まりつつあるが、こうした研究に対する評価は充分に行われていないために、若手の研究者が積極的に情報処理を活用して新しい研究を切り開いていくこうという姿勢があまり見えてこないようと思われる。

本報告書が、新しい道を切り開こうとする今後の研究者たちにとって、不充分ではあっても、なんらかの形で役に立ち、刺激となるものであることを切に願うものである。

2003年1月

## 目 次

はじめに .....	i
研究班参加者・研究課題一覧 .....	iv
古典文献の計量的分析 .....	村上征勝, 古瀬順一 1
古典学のための情報処理 .....	安永尚志 他10名 14
古典テキストのデジタル化とデータベース構築・利用支援システムの開発 .....	及川昭文 72
Web 公開のためのデータベース・エンジニアリング .....	及川昭文, 山元啓史 78
ACI-japan -It's Perspectives for the Future- .....	AKIFUMI OIKAWA 85
古近和歌集パラレルデータベースと公開システム .....	山元啓史, 及川昭文 87
古典文献データベースの表記体系確立 .....	徳永宗雄, 山田 篤 95
古典学のための多言語処理システム開発 .....	高島 淳 102
クメール語およびタイ語の翻字について .....	峰岸真琴, 高島 淳 129
チベット文字における結合文字の使用に関する調査 .....	星 泉, 高島 淳 141
平安時代物語文の比較計量的研究 .....	今西裕一郎, 小西貞則, 室城秀之 145
インド古典天文学書の研究と伝統暦プログラムの改良 .....	矢野道雄 175
XML の骨抜き利用法 アジア・アフリカ言語文化研究所データベースの例 .....	豊島正之 178
Web を用いたコラボレーションに関する試み .....	永崎研宣 188
発表成果一覧 .....	195

## 研究班参加者・研究課題一覧

### (1) 計画研究

#### 古典文献の計量的分析

研究代表者 村上 征勝 統計数理研究所領域統計研究系 教授

分担者 古瀬 順一 宮崎大学教育文化学部 教授

#### 古典学のための情報処理

研究代表者 安永 尚志 国文学研究資料館研究情報部 教授

分担者 及川 昭文 総合研究大学院大学教育研究情報資料センター 教授

分担者 武井 協三 国文学研究資料館研究情報部 教授

分担者 松村 雄二 国文学研究資料館研究情報部 教授

分担者 石塚 英弘 国文学研究資料館研究情報部 教授

分担者 柴山 守 大阪市立大学学術情報総合センター 教授

分担者 アンドル・アーマー 慶應大学文学部 教授

分担者 中村 康夫 国文学研究資料館研究情報部 助教授

分担者 原 正一郎 国文学研究資料館研究情報部 助教授

分担者 山田 哲好 国文学研究資料館史料館 助教授

分担者 山田 奨治 國際日本文化研究センター研究部 助教授

#### 古典テキストのデジタル化とデータベース構築・利用支援システムの開発

研究代表者 及川 昭文 総合研究大学院大学教育研究情報資料センター 教授

分担者 吉岡 亮衛 国立教育研究所 研究室長

分担者 山元 啓史 カルフォルニア大学サンディエゴ校

分担者 湯川 哲之 総合研究大学院大学 教授

分担者 出口 正之 総合研究大学院大学 教授

#### 古典文献データベースの表記体系確立

研究代表者 徳永 宗雄 京都大学大学院文学研究科 教授

分担者 山田 篤 財団法人京都高度技術研究所情報メディア研究室 室長

#### 古典学のための多言語文書処理システムの開発

研究代表者 高島 淳 東京外国语大学アジア・アフリカ言語文化研究所 教授

分担者 峰岸 真琴 東京外国语大学アジア・アフリカ言語文化研究所 助教授

分担者 濱田(星)泉 東京外国语大学アジア・アフリカ言語文化研究所 助手

(2) 公募研究

**平安時代物語文の比較計量的研究**

研究代表者 今西裕一郎 九州大学大学院人文科学研究院 教授

分担者 小西 貞則 九州大学大学院数理学研究院 教授

分担者 室城 秀之 白百合女子大学文学部 教授

**インド古典天文学書の研究と伝統暦プログラムの改良**

研究代表者 矢野 道雄 京都産業大学 教授

**抄物の原典参照データベースの構築 —『韻府群玉』と『玉塵抄』を例として—**

研究代表者 出雲 朝子 青山学院女子短期大学国文学科 教授

分担者 豊島 正之 東京外国语大学アジア・アフリカ言語文化研究所 助教授

---

---

# 「情報処理」

---

---

# 古典文献の計量的分析

村上 征勝  
古瀬 順一

## 【要旨】

自然科学の研究に比べ、どちらかというと直感的、主観的、哲学的な色彩の強い古典学研究の分野において、コンピュータに入力した古典テキストの文章の数量的特質（文長・単語長の分布、品詞の出現頻度、単語の出現頻度、語彙の豊富さ等）の計量的分析を中心とした科学的な研究方法を開発することが本研究の目的である。

そのため、以下の3つの具体的な課題

- 1) 『法華経 (Astasahasrika Prajnaparamita)』27章の成立順序に関する研究
- 2) 『源氏物語』54巻の成立順序及び著者に関する研究
- 3) 井原西鶴作品の著者に関する研究

を中心に研究を進めた。

計量分析には語彙、構文、文法、音韻等の情報を含む文献データベースが不可欠である。『源氏物語』に関しては、データベースは作成済であったのでそのデータベースを利用し計量分析を試みた。『法華経』と井原西鶴作品に関しては、計量分析可能なデータベースが存在しなかったため、データベースの作成から始めた。『法華経』に関しては、『法華経』の分析に必要となる『八千頌般若経 (Saddharmapundarika Sutra)』と『十地経 (Dasabhumika Sutra)』のデータベースも作成した。井原西鶴作品に関しては、全作品のデータベースを完成するまでには到らなかったが、散文24作品中19作品のデータベースの作成を終えた。

各課題の研究結果の概略は以下の通りである。

## 1). 『法華経』に関する研究

『法華経 (Astasahasrika Prajnaparamita)』27章

の成立年代を推定するため、『法華経』の他に『法華経』の前に成立したとされる『八千頌般若経 (Saddharmapundarika Sutra)』と、後に成立したとされる『十地経 (Dasabhumika Sutra)』のデータベースをまず作成し、次に『八千頌般若経』と『十地経』の文体の違いを計量分析によって明らかにし、『法華経』の各章の文体がこの2經典のどちらの文体に近いかで各章の成立順を調べることを試みた。

そのため『八千頌般若経』を31章に分割、『法華経』を27章に分割、『十地経』を11章に分割し、分割された各章の長行（散文）部分に出現する13種の不変化辞 (ca, na, api, va, evam, eva, ti, ponar, khalo, yotha, hi, atha, tatha) の出現率を主成分分析で分析した。その結果、『八千頌般若経』、『法華経』、『十地経』の各章が經典ごとにある程度まとまってグループを構成すること、また3つのグループは重なる部分を持ちながら『八千頌般若経』、『法華経』、『十地経』の順に並ぶことが明らかとなった。この3經典は『八千頌般若経』、『法華経』、『十地経』の順で成立したと考えられるが、13種の不変化辞の分析でこれを裏付ける結果が得られた。ただ、『法華経』27章の成立順の推定は、他の情報の分析結果を踏まえ総合的に判断する必要があるため結論を出すまでには至っていない。なお『法華経』、『八千頌般若経』、『十地経』の3經典をまとめ、KWIC索引を作成し、15年3月に出版する予定である。

## 2). 『源氏物語』に関する研究

『源氏物語』54巻の各巻における、名詞、動詞、

助動詞等の品詞の出現率と、「ひと」、「もの」などの出現頻度の多い言葉の出現率の類似性を分析した結果、巻45から巻54までの「宇治十帖」と呼ばれる後半の10巻とそれ以前の44巻との間には出現率の平均値等に違いが見られるという結果を得た。

### 3). 井原西鶴作品に関する研究

井原西鶴作品に関しては、計量分析まで至らなかつたが、西鶴の全散文作品24作品中、19作品のデータベースを完成し、その中の14作品に関しては KWIC 索引 (Keyword in context index) を作成し出版した。また、残り 5 作品に関してもやはり KWIC 索引を平成15年 2 月に出版する予定である。

#### 【位置付け】

人類の貴重な文化遺産として引き継がれてきた文学作品、哲学書、歴史書、宗教書などの著作物の中には、著者や成立時期などが不明であつたり、疑問が持たれているため、当該分野の研究の障害となっているものが数多くある。

これまで、文学、哲学、歴史学、政治学、宗教学などの人文科学の諸領域における文献研究では、記述内容の検討が主要な研究課題であった。しかし多くの研究者の努力にもかかわらず、文献によっては、記述内容の解釈や検討という従来の研究方法だけでは解決されずに残っているものもある。このような文献の未解決の問題の解明には、これまでの研究方法とは異なった視点からの研究が必要となる。

ある文献の文の長さや単語の長さの分布、特定の品詞や単語の出現率などの統計量の中で、記述内容によって左右されることの少ない数量的性質を計量的に分析し、著者の推定、著作の真贋判断、著作年代や著作順序の推定などを試みる計量文献学もそのような従来とは違う視点からの研究法の一つである。

計量文献学は20世紀の初頭から研究が始まり、欧米の文学作品に関してすでにこの研究手法によって多くの成果が得られている。さらに『新約聖書』中の「パウロの書簡」、『旧約聖書』の中の「イザヤ書」、15世紀に書かれた「キリストにならひて」、プラトンの『第七書簡』など、文学作品に限らず宗教、哲学研究の場においてもこの手法が用いられている。

しかしながら、日本語文献やサンスクリット文献などに関しては、研究例は非常に少ない。日本語の文献に関しては、コンピュータで日本語処理が出来るよう

になったのが比較的最近であること、日本文は英文などとは異なり、分ち書きされていないため単語の認定が難しいこと、漢字とカナが混在していること、などが原因で計量分析は諸外国に比べ大きく遅れている。日本語現代文に関しては、形態素分析のソフトの研究が行われるようになり、計量分析が発展する素地は作られつつあるが、古文に関する形態素分析のソフトはいまだ完成していない。したがって古典に関しては、残念ながら本研究の代表者達の手による日蓮遺文の計量分析の研究と、現在、本特定研究で試みた『源氏物語』及び西鶴作品の計量分析など若干の研究しかないというのが現状である。

#### 【連携の成果】

古典学の研究に統計手法を中心とした計量分析を適用する場合、まず必要となるのは、データベースを作成するための高度な言語学的知識である。次に、データベースから研究に必要な有効な情報を引き出すための統計分析法に関する知識であり、最後に必要なのが、分析結果の妥当性を的確に判断するための古典学に関する広く、かつ深い知識である。

したがって、研究を成功させるためには、古典学研究者だけでは不可能であり、統計研究者だけでも不可能である。したがって、必然的に学際研究とならざるを得ない。本計画研究班は統計学と国語学の研究者 2 名より成るが、具体的な分析対象の文献が、仏教書であつたり、国文学でも平安時代と江戸時代というように時代も大きく異なることから、本研究を遂行するに当って、この 2 名の研究者以外に、多くの専門家の協力を得た。

『法華経』に関する研究では、立正大学教授伊藤瑞穂（インド哲学）、『源氏物語』に関する研究では九州大学教授今西祐一郎（中古文学）、白百合女子大学教授室城秀之（中古文学）、西鶴作品に関する研究では共立女子大学名誉教授浅野晃（近世文学）、二松学舎大学教授竹野静雄（近世文学）、早稲田大学教授谷脇理史（近世文学）、武庫川女子大学教授西島政哉（近世文学）の諸氏の協力を得た。

これらの研究者の方々と共同で研究を進めたことによって、インド哲学、中古文学、近世文学分野の研究者に役立つ、精度の高いデータベースや KWIC、さらには計量分析の結果、当該分野の研究に有益な知見が得られたと確信している。

## 【研究成果】

### 1 「法華経」27章の成立順に関する研究

「法華経」27章の成立順の推定には、「法華経」の前後に成立したと考えられる、「八千頌般若経」、「十地経」の2經典のデータベースが必要となったため、3經典のデータベースを作成することとした。

#### (1) データベースの作成と底本

データベース作成に用いた底本は、

##### ・「八千頌般若経」

ABHISAMAYALAMKAR'ALOKA PRAJNA-PARAMITAVYAKHYA THE WORK OF HARIBHADRA  
EDITED BY U.WOGIHARA, TOKYO, 1973

##### ・「法華経」

“SADDHARMAPUNDARIKA-SUTRAM” A Skt.MS.& Tibetan and Chinese translation by Prof. U. Wogihara and C. Tsuchida, Totyo, 1994

である。

「法華経」が荻原本に依拠した理由は以下の理由による。即ち、「法華経」の場合、完全なテキストを追求しても、現在においては入手が不可能であり、研究の便宜上、比較的の信頼性の高いテキストによることが妥当であると思われる。日本においては従来、荻原本が広く参考にされており、その信頼性に関しても問題はないと考える。

##### ・「十地経」

##### 散文部分は

Dasabhumikasutra, ed. J. Radher, Louvain 1926  
韻文部分は

“The gathas of the Dasabhumikasutra”, ed. J. Rahder & S. Susa, Eastern Buddhist, vols. 5-6, 1931-2

によった。

ただし、底本の「十地経」散文部分には省略が多くため、Radherの指示にしたがって本文を復元した。

これらの經典のデータベースの作成においては、複合語は原則として、荻原雲来編纂「梵和大辭典」(1986年・東京)にしたがって可能な限り分解(ただし、1000までの数詞と bodhisattva, tathagata など表1に示す若干の複合語を除く)し、それぞれの単語に関し、語形、語幹(語根)、語長や、その単語が複合語の一部であるか否か、長行(散文)部分に出現したか、偈頌(韻

文)部分に出現したか等の計量分析に必要となる情報を付加した。

図1はこのようにして作成した3經典のデータベースを用いて作成したKWIC索引の一部である。

Atmabhava	tathagata
Atmasarira	traidhatuka
Grahapati	namadheya
Janapada	pratyekabuddha
Bodhisattva	rajadhani
Manasikara	lohitamukta
Manaskara	vajrapani
Manojna	senapati
Mahesvara	svabhava
Musaragalva	

表1 分解しなかった複合語

#### (2) 分析結果

ここでは13種の不変化辞 (ca, na, api, va, evam, eva, ti, ponar, khalo, yotha, hi, atha, tatha) の出現率の分析結果を示す。図2は「八千頌般若経」と「十地経」の長行における13種の不変化辞を分析した結果である。(図中の△印は「八千頌般若経」の27章を、○印は「十地経」の11章の長行部分を示している。)

この図を見ると、「八千頌般若経」27章と「十地経」11章はほぼ二分されており13種の不変化辞の用い方にこの2つの經典では差があることがわかる。「十地経」は「八千頌般若経」に比べ、ca, yathaの出現率が高く、一方「八千頌般若経」はhi, na, evam, ili等の出現率が高い。したがって、「法華経」の中でhi, na, evam, iliの出現率が高い章は「八千頌般若経」と不変化辞の用い方が似ていることより比較的早く成立したと推定できる。一方、ea, yathaの出現率の高い章は「十地経」と不変化辞の用い方が似ているため、逆に、遅く成立したと推測できる。

そこで、次にこの2經典に「法華経」も加え、3經典を同時に数量化III類で分析してみたところ図3のような結果を得た。

図3では、「八千頌般若経」のグループと「十地経」のグループが並び、その両グループに「法華経」の各章がまたがるように配置されている。この結果から、「八千頌般若経」と「法華経」の成立時期は、「十地経」の間と推定できる。さらに、この図を詳細に検討してみると「法華経」の2章、13章、15章、19章、21章は「八千頌般若経」のグループの領域に位置することから成立が早く、1章、6章、20章、27章は「十地経」のグループの領域に位置することから成立が遅いとい

うことが推察される。しかし、各章の成立順について  
は他の情報の分析も含め総合的な判断が必要であ

り、13種の不変化辞の分析からはこれ以上の判断は難  
しいと思われる。

### bhakta - Bhagavams

## BH

<sup>o</sup>a gr̥īmat̥ aṇucaṇikramantaḥ paryeṣate  
// paryeṣite bhojanī nīrvṛtaḥ syād  
<sup>o</sup>tā anupūrvaśo tatra gato bhavet /  
<sup>o</sup>īyas tātha pāda-mrakṣapam / salopā-  
tasraḥ parṣadas tam evākam̥ paścād-  
taj-jātyāḥ pudgalā na sevitavyā na  
taj-jātyāḥ pudgalā na sevitavyā na  
āni buddhvā ca vivarjayitavyāni na  
kāmena kalyāpa-mitrāpi sevitavyāni  
ita eva kalyāpa-mitrāpi sevitavyāni  
ita eva kalyāpa-mitrāpi sevitavyāni  
lyāpa-mitrāpi ca tvayā sevitavyāhi  
ah Subhūte pudgalo na sevitavyo na  
liko bhavisyati sacet khalu-paścād-  
-rasa-sparśin upasampharati sevitum  
mahāsattvasya satatasamitaiḥ gandham  
śyām agaru-turuṣka-kunduruka-rasam  
uṇa tatra santi manuṣya-kuṇapānī ca  
nubadhnīyād ānugacched āhāra-hetor  
utpādayitavyam | yadi cen māṇi vyāḍa  
vā amī citūtpādāḥ || anumoditavyā  
antam etad avocat | asaṅkhyeṣṭas te  
ma-paryāyah śruta-pūrvaḥ | imāś ca  
m etad avocat | prajñā-pāramitāyāṁ  
mbodhim abhisambudhyate | paurvakō  
yanti na dhandhāyiṣyanti | kutaḥ sa  
ala-mūlāny anumoditāni || yair api  
vā amī citūtpādāḥ || anumoditavyā  
saṃsthāpayati | niḥspṛhā ca vayam  
arāmo darśayāma upadarśayāmo vayam  
samyaκ-saṃbodhāv anuśiṣṭā ca na ca  
sthavirāḥ Subhūtir āha | gambhirā  
deṣayiṣyanti || Subhūtir āha | yadā  
| gambhirā Bhagavān tathatā | ato  
gatasyānuśikṣemahi | arthino vayam  
tat kasya hetoh | yair nāmāsmābhīr  
sthavirāḥ Subhūtir āha | kiṇ punar  
a-pāramitāyāḥ || tad-yathā 'pi nāma  
dham etad ūcuḥ | imāni khalu punar  
rākramaḥ kṛta iti | kiṇ cāpi vayam  
vantam etad avocat | tān api tāvad  
| tat kasya hetoh | paurvakāpām hi  
ākhyatāḥ || Subhūtir āha | kathaṇ  
vantam etad avocat | katham idāni  
śidhānena sadā 'vinaṣṭena te vayam  
alpa-prahīpā || tad-yathā 'pi nāma  
avantam etad avocat | na tāvad ime  
pi catvāriṇśad-varṣāṇi | tat katham  
sambodhau cittam utpādītaḥ | yeṣām  
ajāt-pāramitāḥ | sarva-sattvānām hi

bhakta tathā 'pi coḍam // paryeṣamāṇo 'pi  
bhaktam udāraṇaḥ avicintayāntaḥ / taḥ cāpi <sup>o</sup>  
bhaktam ca coḍam ca gavreṣamāṇo niveṣanam̥ y  
bhaktam ca dadāmi tubhya ūkām̥ ca ūkām̥ ca  
bhaktam samjānante smēmām̥ ca Sahām̥ loka-dh  
bhaktayā\* na paryupāsitavyāḥ | api tu khalu  
bhaktayā\*\* na paryupāsitavyāḥ | api tu khalu  
bhaktavyāni ārabdha-vīryeṣa smṛtimatā samprajā<sup>o</sup>  
bhaktavyāni paryupāsitavyāni | yāny enām̥ prajñā<sup>o</sup>  
bhaktavyāni paryupāsitavyāni || evam ukte Bhag<sup>o</sup>  
bhaktavyāni paryupāsitavyāni || Subhūtir āha |  
bhaktavyāni paryupāsitavyāni yāni ca ūkām̥ tān<sup>o</sup>  
bhaktavyo na paryupāsitavyāḥ || tat kasya he<sup>o</sup>  
bhaktiko bhavisyati sacet ekāsaniko bhavis<sup>o</sup>  
bhaktum paryupāsitum tām̥ cāsav abhibhūyō<sup>o</sup>  
bhakṣayataḥ campaka-tailaṁ ca pibato dvādaśa-v<sup>o</sup>  
bhakṣayati sma campaka-tailaṁ ca pibati sma |  
bhakṣayantaḥ / teṣām̥ ca nīryāpu pratīkṣamāṇeḥ <sup>o</sup>  
bhakṣayitu-kāmāḥ | atha sa prāṇaka-jāto yena sā mag<sup>o</sup>  
bhakṣayeyus tebhya eva tad dānaṇi dattam bhavatu  
bhagam̥ teṣām̥ bodhisattvānām̥ mahāsattvānām̥  
bhagavānū loka-dhātavo 'gaṇanīyāś citta-bhūm<sup>o</sup>  
bhagavam̥\*\* catasraḥ parṣado vicikitsā-kathaṇk<sup>o</sup>  
Bhagavam̥ caratā bodhisattvena mahāsattvē<sup>o</sup>  
Bhagavam̥ cittūtpādāḥ paścimakena cittūtpāde<sup>o</sup>  
Bhagavam̥ cyuta iḥopapanno veditavyāḥ kiyac<sup>o</sup>  
Bhagavam̥ chandam utpādyā bodhaye bodhisattv<sup>o</sup>  
bhagavam̥\*\* teṣām̥ bodhisattvānām̥ mahāsattvānām̥  
bhagavam̥ tata evān jānlīna etad evāśmākam̥ ba<sup>o</sup>  
bhagavam̥ tato niḥspṛhā samānāḥ | tat kasya  
bhagavam̥ tatrāśmābhīr ekam̥ api spṛhā-cittam̥  
Bhagavam̥ tathatā | ato Bhagavam̥ tathatāto  
Bhagavam̥ tathatā-vinirmukto nāṇyaḥ kaś-cid  
Bhagavam̥ tathatāto buddhānām̥ Bhagavatām̥ bod<sup>o</sup>  
bhagavam̥ tathāgata-jīvāna-darśanena | bhagav<sup>o</sup>  
bhagavam̥ tathāgata-jīvāne 'bhisamboddhavya e<sup>o</sup>  
Bhagavam̥ Tathāgatavatām evācīntyam atulyam a<sup>o</sup>  
Bhagavam̥ Tathāgatasya kaś-cit priyo vā 'pri<sup>o</sup>  
bhagavam̥ tathāgatasya bahūni śrāvaka-koṭi-n<sup>o</sup>  
bhagavam̥ tathāgatasya vacanām̥ śraddhāyā gam<sup>o</sup>  
bhagavam̥ tathāgat'utma-bhāvām̥ tathāgata-ni<sup>o</sup>  
Bhagavam̥ Tathāgatānām̥ arhatām̥ samyaκ-sambud<sup>o</sup>  
Bhagavam̥ Tathāgatānām̥ prajñā-pāramitāyā pañ<sup>o</sup>  
bhagavam̥ tathāgatena kumāra-bhūtena Kapilav<sup>o</sup>  
bhagavam̥ tathāgatena saṃbodhyamānāḥ | mā yū<sup>o</sup>  
Bhagavam̥ Tathāgatenārhatā\* samyaκ-sambuddhe<sup>o</sup>  
Bhagavam̥ Tathāgatenārhatā samyaκ-sambuddhena  
bhagavam̥ tathāgatenēyātā kālāntareqēdām apa<sup>o</sup>  
Bhagavam̥ Tathāgato 'parityakto dharmo 'pari<sup>o</sup>  
Bhagavam̥ Tathāgato 'rhan samyaκ-sambuddhaḥ

Sp. 4 104-24  
Sp. 8 187-12  
Sp. 4 105- 4  
Sp. 4 107- 9  
Sp. 14 256- 1  
Ap. 21 783-26  
Ap. 21 783-26  
Ap. 11 525-14  
Ap. 15 593- 9  
Ap. 22 785-13  
Ap. 22 785- 7  
Ap. 30 930- 5  
Ap. 21 783- 5  
Ap. 21 773-27  
Ap. 30 930-25  
Sp. 22 341-14  
Sp. 22 341-11  
Sp. 3 77-23  
Ap. 3 194-23  
Ap. 19 739- 6  
Ap. 26 833-14  
Sp. 15 269-18  
Sp. 2 32-25  
Ap. 20 749- 5  
Ap. 19 717- 8  
Ap. 7 390-19  
Ap. 26 835-22  
Ap. 26 833-14  
Sp. 4 102-13  
Sp. 4 102-18  
Sp. 4 96- 4  
Ap. 12 559-17  
Ap. 27 858-27  
Ap. 12 559-18  
Sp. 7 161-17  
Sp. 8 185-11  
Ap. 13 571-12  
Ap. 26 841-25  
Sp. 7 161-11  
Sp. 14 265- 4  
Sp. 11 210- 1  
Ap. 2 135-12  
Ap. 12 537-26<sup>o</sup>  
Sp. 14 264- 4  
Sp. 8 186- 8  
Ap. 26 842-13  
Ap. 3 249-23  
Sp. 14 264- 8  
Ap. 26 833-16  
Ap. 4 280- 2

図1 「法華經」、「八千頌般若經」、「十地經」のKWIC の一部

般若經・十地經13不変化辞

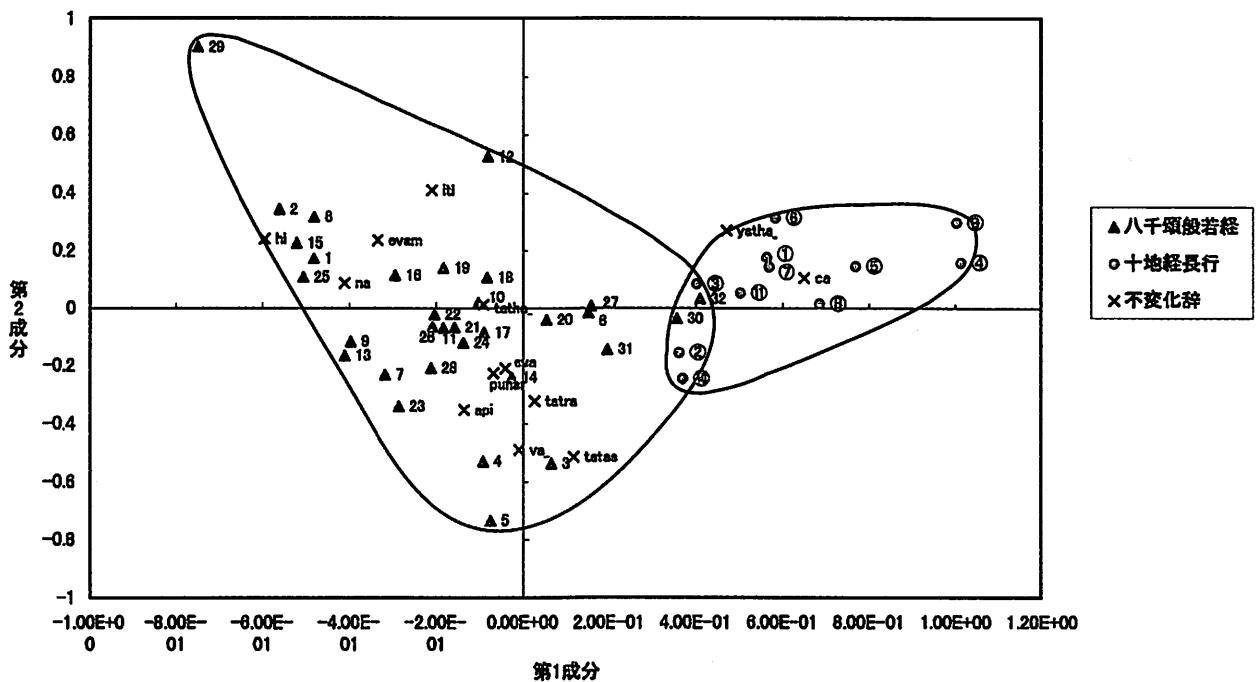


図2 13種の不変化辞の数量化III類分析結果  
（「八千頌般若經」及び「十地經」）

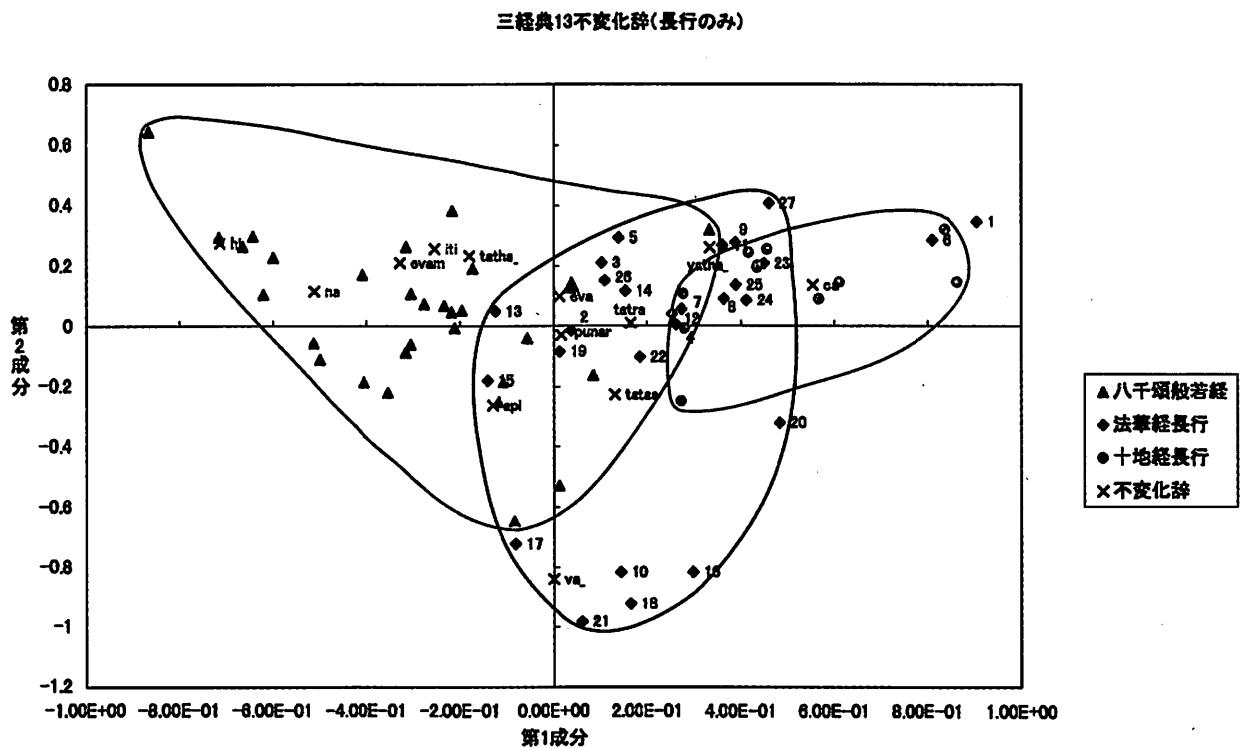


図3 13種の不変化辞の数量化III類分析結果  
（「八千頌般若經」、「法華經」及び「十地經」）

## 2 「源氏物語」54巻の成立順序及び著者に関する研究

平安時代の貴族生活を背景に、光源氏の恋と栄華を描いた大河小説「源氏物語」の成立は、西暦1004年頃といわれている。したがって成立後千年にならんとしている。この全54巻からなる「源氏物語」は、我が国古典文学の最高峰であるばかりでなく、その芸術性の高さ故に諸外国にも広く翻訳され、日本人はもとより多くの外国人にも親しまれると同時に、多くの文学者の研究対象にもなってきた。

「源氏物語」には紫式部（970？－1019？）の自筆原稿が存在せず、写本でのみ伝えられてきている。そのため、様々な写本の不備を訂し、るべき本文を復原する本文の校訂に始まり、語句の注釈、作者とされる紫式部についての伝記、時代的背景など、諸分野からの研究が古注釈以来現代に至るまで数百年にわたって着実に積み重ねられてきた。しかしながら、現時点においてもなお研究課題は数多く存在する。なかでも54巻がすべて紫式部によって書かれたかどうかという基本的な問題に関しては、古くから疑問が出され、数多くの作者複数説が出されてきた。

たとえば、

- ・年記、官位などの叙述面での矛盾から（巻42）「匂宮」、（巻43）「紅梅」、（巻44）「竹河」の3巻を別人の作とする説
- ・（巻1）「桐壺」から（巻41）「幻」までを紫式部の作とし、（巻42）「匂宮」、以後の巻は紫式部の娘大式の三位の作とする説
- ・与謝野晶子が主張した（巻33）「藤裏葉」までを紫式部の作とし、物語の中で用いられる歌が非常に優れている（巻34）「若菜上」以後の作は大式の三位（藤原賢子）の作とする説
- ・宇治を舞台とする（巻45）の「橋姫」以後の10巻（まとめて「宇治十帖」と呼ばれている）を大式の三位の作とする「宇治十帖」後記説

このような「源氏物語」の作者複数説の中でもっとも広く知られているのは、「宇治十帖」を大式の三位の作とする説である。

この他、54巻の成立順序や、書写者による部分的な補筆の可能性などに関しても疑問が出されており、これらの問題の多くは、数多くの研究者の努力にもかかわらず、依然として未解決のまま持ち越されている。

そこで、「源氏物語」の文章をあらたに計量的観点から分析するため、「源氏物語」54巻の文章を単語に分割し、品詞情報、活用形の情報などを付加したデータベースを構築した。

分析では、表2のように、54巻を4つのグループに分けて検討することとした。これは次のような理由による。今日、「源氏物語」54巻は全体の構成の観点から、次の3部、すなわち

第1部：（巻1）「桐壺」～（巻33）「藤裏葉」

第2部：（巻34）「若菜上」～（巻41）「幻」

第3部：（巻42）「匂宮」～（巻54）「夢浮橋」

に分けるのが通説となっている。しかしこの分析では、54巻の成立順序や、後半の10巻（「宇治十帖」）の他作者説などの可能性を検討するため、通説とは多少異なり、表2のような、A、B、C、Dの4つのグループに分割した。

グループ名	巻数	グループに含まれる巻の番号	備考
A	17	1, 5, 7, 8, 9, 10, 11, 12, 13, 14, 17, 18, 19, 20, 21, 32, 33	紫の上系物語（第1部）
B	16	2, 3, 4, 6, 15, 16, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31	玉壁系物語（第1部）
C	11	34, 35, 36, 37, 38, 39, 40, 41 42, 43, 44	第2部 「匂宮三帖」（第3部）
D	10	45, 46, 47, 48, 49, 50, 51, 52, 53, 54	「宇治十帖」（第3部）

表2 「源氏物語」54巻のグループ化

第1部の33巻を、紫の上系物語の17巻（Aグループ）と玉壁系物語の16巻（Bグループ）に分けたのは、玉壁系の16巻が紫の上系の17巻の後に成立したとする説があるからである。また第3部に含まれる匂宮三帖（（巻42）「匂宮」、（巻43）「紅梅」、（巻44）「竹河」）を第2部の8巻と合わせてCグループとしたのは、筋立て、文体等において緊密な関連を有する「宇治十帖」の10巻を1つのグループ（Dグループ）とするためである。

### [作者複数説の検討]

表3は「源氏物語」に出現する品詞のうち、主要な7つの品詞の、巻1～44の一巻当たりの平均出現率と、「宇治十帖」の10巻の一巻当たりの平均出現率を比較

したものである。

(A, B, C グループ) (D グループ)			
品詞	巻1～44	巻45～54	t 値
動詞	0.16190	0.16505	-1.3591
形容詞	0.05907	0.05566	1.8597
形容動詞	0.02388	0.02426	-0.2963
名詞	0.18189	0.16423	5.1624
副詞	0.04000	0.04106	-0.8770
助動詞	0.11309	0.12391	-5.0477
助詞	0.31517	0.31872	-1.6130

表3 「源氏物語」の巻1～44と「宇治十帖」(巻45～54)の7つの品詞の一巻当たりの平均出現率

ところで、どのような書き手でも、品詞の使用率を常に同じにすることはできない。したがって、紫式部真作の巻でも当然巻によって品詞の出現率に違いが出る。

この表に示されたt値は、そのような巻ごとの出現率のバラツキを考慮した上で、巻1から巻44までと、「宇治十帖」の10巻の一巻当たりの平均出現率の差が、紫式部が54巻すべてを書いた時によく生じるような差であるのか、生じる可能性のほとんどない差であるのかを判定するために用いる量であり、次の式で計算される。

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

ここで、 $n_1$ ,  $\bar{x}_1$ ,  $s_1^2$ は、それぞれ巻1から44までの巻数、44巻の当該品詞の平均出現率、及び出現率の分散であり $n_2$ ,  $\bar{x}_2$ ,  $s_2^2$ は「宇治十帖」の巻数、「宇治十帖」での当該品詞の平均出現率及び出現率の分散である。統計学的にはt値の絶対値が2以上ならば(つまり+2以上あるいは-2以下ならば)、紫式部が54巻すべてを書いた時に、よく生じるような差ではないと考えられている。したがって、表3のt値を見ると、名詞、助動詞の平均出現率については、巻1～44と「宇治十帖」との間には紫式部がすべてを書いた場合に生じる可能性のほとんどない大きな差があると考えられる。言い換えると、紫式部が54巻すべてを書いた場合に、巻1～44と「宇治十帖」の一巻当たりの名詞や助動詞の平均出現率について、表3で示したような差が生じる可能性は非常に低いということになる。また形

容詞の差も名詞や助動詞ほどではないにしろ、紫式部が全てを書いた場合には生じる可能性の低い差であるといえる。

表4には、「源氏物語」によく出現する言葉の中から、やはり巻1～44と「宇治十帖」との間で平均出現率に大きな差があると考えられる言葉をリストアップした。

(A, B, C グループ) (D グループ)			
言葉	巻1～44	巻45～54	t 値
ひと/人(名詞)	0.1437	0.02000	-2.9273
もの/物(名詞)	0.00429	0.00581	-1.9426
おもひ/思ひ(動詞)	0.00251	0.00146	-3.0332

表4 「源氏物語」の巻1～44と「宇治十帖」(巻45～54)の「ひと／人」「もの／物」「おもひ／思ひ」の一巻当たりの平均出現率

「源氏物語」の自立語の中での、名詞の「ひと(人)」の出現頻度は第3位(総出現度数3441回)、「もの(物)」は第17位(総出現度数1041回)、動詞の「おもひ(思ひ)」(動詞「思ふ」の連用形)は第15位(総出現度数1170回)である。このような出現頻度の多い言葉に関しても、巻1～44と「宇治十帖」では平均出現率に同じ人が書いた場合には生じる可能性の少ないと考えられる大きな差が生じている。

次に巻ごとの品詞や単語の出現率を調べてみる。図4～図8は、「源氏物語」の各巻における3種類の品詞(名詞、助動詞、助詞)と2つの単語(「ひと」、「こと」)の出現率を、A, B, C, Dの4グループに分けて示したものである。各巻は各グループ内で巻番号の大小順に並んでいる。また図中の横線はグループの平均出現率である。

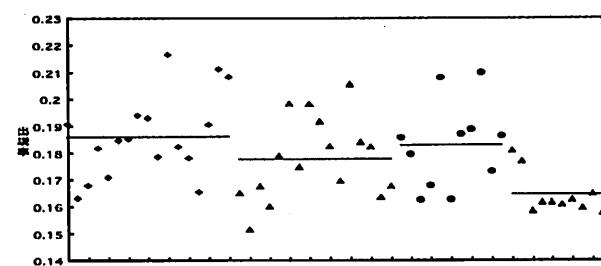


図4 グループ別名詞の出現率

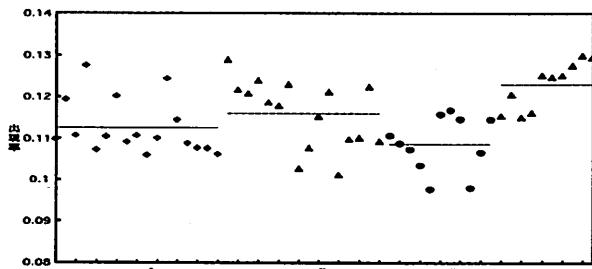


図5 グループ別助動詞の出現率

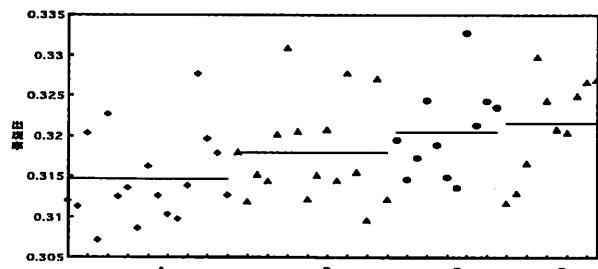


図6 グループ別助動詞の出現率

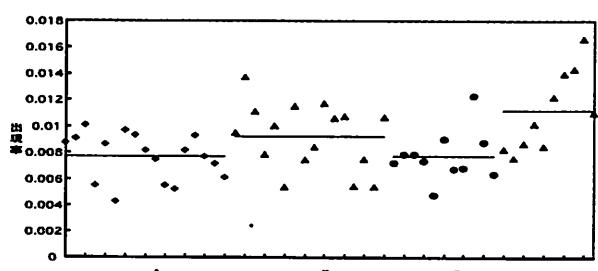


図7 グループ別名詞「ひと」の出現率

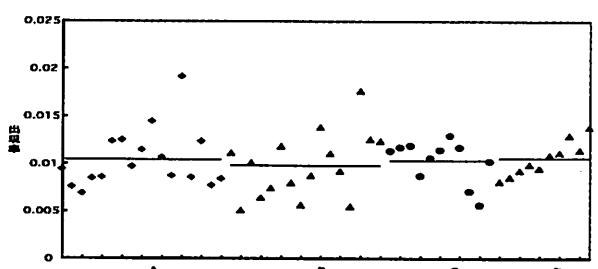


図8 グループ別名詞「こと」の出現率

これを見ると、Dグループの平均出現率が他の3つのグループの平均出現率と大きく異なっていたり（図4、図5、図7）、またDグループの巻の出現率に他のグループには見られない一定の増加傾向が見られたり（図5、図6、図7、図8）と、Dグループの巻の文章が他のグループの巻の文章に比べて明らかに異なることがわかる。

このように、計量分析でDグループ「宇治十帖」の文章はそれ以前の巻1～44までの文章と幾つかの点で異なるという興味深い結果が得られているが、このような分析結果から、「宇治十帖」の作者が紫式部ではないと結論づけるのは現段階では難しい。それは、これまでの分析は紫式部の文体は変化しないという前提に立っているが、54巻という長編を執筆している間に紫式部自身の文体が変化した可能性も考えられるからである。

『源氏物語』の作者複数説を検討するにはさらに、『源氏物語』と成立時期が近い『うつほ物語』や、他の人物が紫式部の文章を模倣して書いた作品である『雲隠六帖』、『山路の露』、『手枕』などの文章と比較するなど、より詳細な分析が必要である。

#### [54巻の成立順序の検討]

現段階では作者複数説について明確な結論を下すことができないため、ここでは54巻がすべて紫式部の手によるものと仮設した上で、助動詞の出現率に注目し54巻の成立順序について検討する。

『源氏物語』54巻に出現する助動詞は26種である。出現頻度の多い順に並べると、

す、む、たり、けり、なり、り、ぬ、き、べし、つ、る、

す、めり、さす、らむ、らる、じ、けむ、まじ、まし、

まほし、ごとし、らし、しむ、ます、むずとなる。助動詞の総出現数は43,617、出現率は0.116で、この出現率は動詞、助詞、名詞に次ぐ。

前記26種の助動詞のうち、出現頻度の少ない「ごとし」、「らし」、「しむ」、「ます」、「むず」の5語に関しては、これらの助動詞が出現する巻が54巻の半分にも満たないため、分析には用いないこととした。

図9は出現率の高い22種の助動詞の情報を用いて全54巻に主成分分析（相関行列）の結果である。AグループとCグループが重なり、BグループとDグループが重なっていることがわかる。

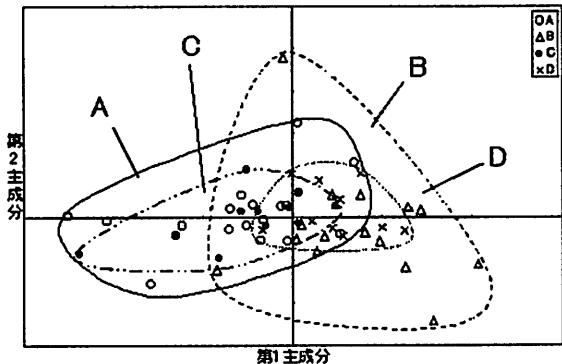


図9 54巻の主成分分析（相関行列）結果

図10は、A, Bの2グループの33巻のみで主成分分析を試みた結果である。Bグループの一巻を除き、A, Bグループにはほぼ二分されており、この二つのグループ間では助動詞の用い方に違いがあることがわかる。

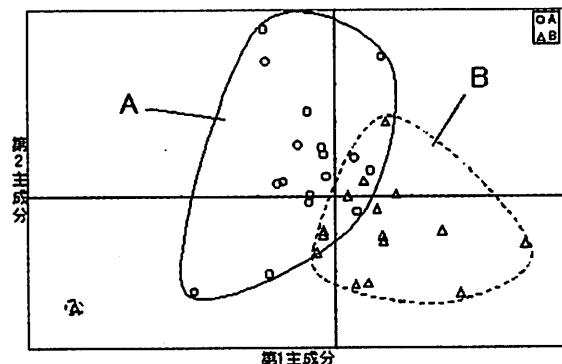


図10 A, Bグループ33巻の主成分分析結果

また図11はC, Dの2グループの21巻で主成分分析を試みた結果で、C, Dの二つのグループでも助動詞の用い方に違いがあることがわかる。

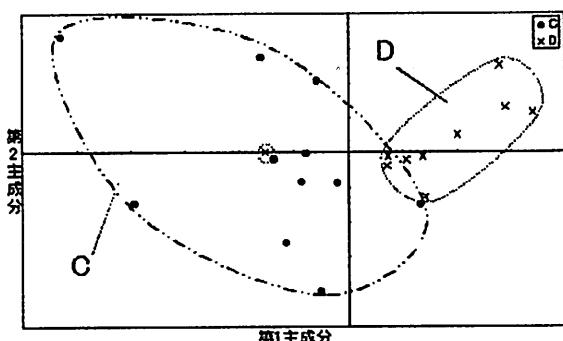


図11 C, Dグループ21巻の主成分分析結果

『源氏物語』の中では1~33巻が一つのまとまった話とされているが、この33巻は前述のA, Bの二つのグループに分けることができ、Bグループの巻が後に書かれたとする説がある。Bグループが後に書かれたかどうかは別にしても、図10は少なくともこのA, B分割説を支持しているように思える。また、著者別人説のあるDグループは、図5で見たとおり前半の44巻と平均出現率が大きく異なって、更に、図11に見られるようにDグループの前のCグループの11巻とも異なっている。しかしながら、図9を見るとBグループと重なっていることも指摘できる。

以上、図9~図11の結果を総合すると、もし54巻がすべて紫式部の手によるものであり、Aグループが最初に、Dグループが最後に書かれたとすると、54巻はA, C, B, Dの順に書かれた可能性が高いようと思われる。

### 3 井原西鶴作品の著者に関する研究

西鶴作品は短期間に大量の作品が出版されていることから、全ての作品が西鶴の手によるものではなく「西鶴工房」のようなものが存在したのではないかという説が出されている。この研究の目的は、西鶴の全散文作品を計量的に比較することによって全ての作品が西鶴の手によるものかどうかを調べることにある。

そのため、西鶴の全散文作品24作品のデータベース作成を試みた。以下の19作品

『好色一代男』、『諸艶大鏡』、『好色五人女』、『好色一代女』、『枕久一生物語』、『西鶴諸国はなし』、『本朝二十不孝』、『男色大鏡』、『武道伝来記』、『好色盛衰記』、『懷硯』、『日本永代蔵』、『色里三所世帶』、『武家義理物語』、『嵐無常物語』、『新可笑記』、『本朝桜陰比事』、『浮世栄花一代男』、『西鶴置土産』、

については、文章を単語に分割し、計量分析に必要な情報を付加した全文データベースを完成した(図12)。

このデータベースを用いて

『新編西鶴全集第1巻自立語索引篇』(上)(下)  
([好色一代男]、[諸艶大鏡]、[好色五人女]、[好色一代女]、[枕久一生物語])

『新編西鶴全集第2巻自立語索引篇』(上)(下)  
([西鶴諸国はなし]、[本朝二十不孝]、[男色大鏡]、[武道伝来記]、[好色盛衰記])

『新編西鶴全集第3巻自立語索引篇』(上)(下)  
([懷硯]、[日本永代蔵]、[色里三所世帶]、[武家義理物語]、[嵐無常物語]、[新可笑記]、[本朝桜陰比事])

を出版し（第3巻は平成15年2月出版予定），西鶴研究者に提供した。図9は第1巻，図10は第2巻の自立語索引の一部である。

残り5作品

「世間胸算用」，「西鶴織留」，「西鶴俗つれづれ」，  
「萬の文万古」，「西鶴名残の友」

の中で，「世間胸算用」，「西鶴織留」に関してはデータベースを作成中である。

西鶴作品の計量分析は24作品のデータベースの完成を待ってから入る予定であり，したがって本特定研究の研究期間内には西鶴作品の計量分析を行なうまでには至らなかった。

作品名	出現順位	丁	本文翻訳	本文古文	前方一致	後方一致	ルビ	本文翻訳	本文古文	前方一致	後方一致	ルビ	本文翻訳	本文古文	前方一致	後方一致	ルビ	本文翻訳	本文古文	前方一致	後方一致	ルビ	本文翻訳	本文古文	前方一致	後方一致	ルビ	本文翻訳	本文古文	前方一致	後方一致	ルビ
405	1528230	08-o	5	338	あさう	されど、萬子には手をかけず、至をあさう	△	神からひ。若も牛貝御玉子はありながら		あさし	4		形容詞	通用	一女	巻五																
201	470800	02-u	10	236	あさから	に渡り給ひ。女王と玉殿の御かたらいト	△	す。二度かへし給はぬなり。されば親子		あさし	4		形容詞	未然	詠艶	巻一																
1204	474710	08-u	12	107	あさから	といへるに、遂初しより	△	す。常よりしめやかなる、枕屏風を		あさし	4		形容詞	未然	永代	巻四																
401	1340530	15-o	3	240	あさから	我薄命の身ながら殿様の御情	△	すして。うれしく御枕をかはせしその		あさし	4		形容詞	未然	一女	巻一																
1001	2046380	05-o	6	29	あさから	捨、男にかたむくべ。此道のあさから	△	ぬ所を、あまねく弘法大師のひろめたまはぬ		あさし	4		形容詞	未然	男色	巻一																
1002	2182720	23-o	10	82	あさから	男は伴の市九郎とて、達駒町人、一念に若切のあさから	△	ぬすき人。此度の江戸心ざしも、堺町		あさし	4		形容詞	未然	男色	巻二																
103	121380	05-u	3	84	あさき	ふとく、すべらかしに、結びしき、廢切の、あさき	△	をいただきつれて、我からぬらす袂		あさし	4		形容詞	通用	逸体	一男	巻三															
1005	2392410	13-u	12	170	あさき	がたし。すこし群の座席の紅葉のあさき	△	脇額、みしに恋をもとめて、		あさし	4		形容詞	通用	逸体	男色	巻五															
1007	2527280	12-u	8	227	あさき	類ひなき事のなり。掲屋といひ内あさき	△	事にはあらず。		あさし	4		形容詞	通用	一男	巻七																
105	273490	17-o	1	148	あさく	冬映画のおぼつかなく、難か極の匂ひあさく	△	表にみえすき、女郎は浴衣姿の雑子		あさし	4		形容詞	通用	一男	巻五																
207	903290	03-u	9	425	あさく	長山人の氣を頂て、山の井のあさく	△	は人に見られず。動むるうちに		あさし	4		形容詞	通用	詠艶	巻七																
208	1003830	15-u	2	468	あさく	冬映画のおぼつかなく、難か極の匂ひあさく	△	唐房も色をむかしに埋み、露葉		あさし	4		形容詞	通用	詠艶	巻八																
303	1188030	04-u	7	162	あさく	に四つばかりのくげ紙を付て、顔自慢にあさく	△	かづき、ぬきあし中(け)ねりのありますがた是屋是し		あさし	4		形容詞	通用	五女	巻三																
401	1333010	12-o	11	237	あさく	歯並あらはとして面白。耳長みあつて諱あさく	△	身をはなれて櫻花見へすき。顔		あさし	4		形容詞	通用	一女	巻一																
403	1445180	12-o	11	293	あさく	てさまざまごかせども其甲斐なく、けふの體ひあさく	△	なりておもしろからず。愛に沙汰まつと		あさし	4		形容詞	通用	一女	巻三																
1102	84910	11-o	10	漫から	て、しおひて行通ふ互のこころ	△	アサす。裏つわににさる団子たつ		あさし	4		形容詞	未然	憶覗	巻二																	
107	408980	17-u	3	206	漫から	おもひしに。其二とせあまり世之介と	△	アサす。中立は前田の、或宿の		あさし	4		形容詞	未然	詠艶	巻三																
203	654350	09-o	4	314	漫から	我心玉にせよ。思入の漫から	△	ぬししを見せんと。まねけばうなづく		あさし	4		形容詞	未然	詠艶	巻三																
1403	752560	11-u	11	52	漫から	男と。二世のかたひらもしつて漫から	△	十。悦喜、からばわたくの景頃		あさし	4		形容詞	未然	詠艶	巻六																
206	841640	07-o	4	398	漫から	難別と云捨てて帰る仕女部の心入れ候久	△	唐房ひ追付江戸より此辺をさそりと申す		あさし	4		形容詞	未然	詠艶	巻七																
207	943540	16-o	11	441	漫から	とや申べし両方共に返事なしにいつとなく	△	ぬかんこはこれ時節をまつちこそうき世なれ		あさし	4		形容詞	未然	詠艶	巻七																
502	1071020		51	漫から	を見ぬ間をうらむ。是は皆	△	す。すれいかかる事なし。京		あさし	4		形容詞	未然	詠艶	巻六																	
304	1234250	04-u	3	185	漫から	難別と云捨てて帰る仕女部の心入れ候久	△	ぬ心からなるに、ふとあきそてて		あさし	4		形容詞	未然	詠艶	巻七																
305	1269470	02-o	9	205	漫から	漫から	△	す。すれひ追付江戸より此辺をさそりと申す		あさし	4		形容詞	未然	詠艶	巻七																
402	1401670	16-o	7	270	漫から	といへるにはじめりより命を捨て	△	ぬかんこはこれ時節をまつちこそうき世なれ		あさし	4		形容詞	未然	五女	巻四																
1001	2056800	08-o	10	33	漫から	を獲る獲ばざかりに入りわづ夜の闇の漫から	△	す。かづかずの申かはせはもあだになしき		あさし	4		形容詞	未然	一女	巻二																
1003	2230920	09-o	2	103	漫から	によつて、自然と若年にわきまへて、漫から	△	ぬ心しらず。すゑ光見麗けるに、遭理		あさし	4		形容詞	未然	男色	巻一																
1003	2247720	14-u	2	109	漫から	文の便しなつて、美道のかたらひ	△	あさから		あさし	4		形容詞	未然	明色	巻三																
1003	2279120	23-u	11	120	漫から	其中に是も采手と彼道を	△	ぬ中、人も見ゆらし侍る。有時		あさし	4		形容詞	未然	男色	巻三																
1004	2297090	06-o	4	131	漫から	申。いやいや隠し給ふな。我に執心	△	す申かせし。志賀左馬助めとがめて、人は		あさし	4		形容詞	未然	男色	巻三																
1004	2310830	10-o	9	136	漫から	のならはし。夫婦よしし、殊更に御海に	△	ぬ思し召す。見晴た。今となつて		あさし	4		形容詞	未然	男色	巻四																
1005	2385280	11-u	8	167	漫から	は御方へ預けらす。はいへは、六之瀧	△	す。よろこび、今夜の内にと進ける		あさし	4		形容詞	未然	男色	巻四																
1007	2532480	14-u	1	228	漫から	たてたがひに忘れじ。かしみがめならぬ	△	ぬ詞かはして、別れいに、露		あさし	4		形容詞	未然	男色	巻五																
601	2663280	08-o	9	291	漫から	まかなる平生せし。母よろぶ事	△	あさし		あさし	4		形容詞	未然	詠艶	巻七																
601	2680470	12-u	3	295	漫から	にあそばれ。おじ給はり。三年の契約	△	す。男子平生あり。名を市丸		あさし	4		形容詞	未然	詠艶	巻一																
601	2717030	22-u	2	309	漫から	働き、国内において是め。形相陰よろこひ	△	あさし		あさし	4		形容詞	未然	詠艶	巻一																
601	2725790	24-u	2	312	漫から	曲事をなしけり。上より御立願	△	す。抱も理なり。其日は吉例の		あさし	4		形容詞	未然	詠艶	巻一																
1104	2159550	18-o	3	366	漫き	いはんかたもなし。此度細木殿、漫から	△	ぬ縁情かはして、弓矢に、露		あさし	4		形容詞	未然	男色	巻五																
203	653030	08-u	9	313	漫き	の海の深きをみをして、漫き	△	あさし		あさし	4		形容詞	未然	詠艶	巻三																
603	2836030	13-o	7	356	漫き	道鉄が塵もまだるかに、川瀬の漫き	△	す。大きな傾向の如き。味噌汁の餅		あさし	4		形容詞	未然	詠艶	巻一																
607	3113900	06-o	12	462	漫き	くりかへし。後親翁の力を念せは、漫き	△	あさし		あさし	4		形容詞	未然	詠艶	巻一																
1001	2050320	07-o	8	31	漫く	それはそれは、世上のおもはくへり見るやな	△	事にあらず。其方さへははらずす		あさし	4		形容詞	未然	詠艶	巻七																
1007	2515170	09-o	2	222	漫く	に、松葉舟行レヒヘリ。山	△	、無用の探喰、人の縁子		あさし	4		形容詞	未然	男色	巻一																
603	2850170	17-u	8	362	漫く	名乗あひけれど、あらはにしては、心の漫く	△	あさし		あさし	4		形容詞	未然	詠艶	巻一																
604	2935620	21-u	8	393	漫く	また立出けるに、此女、心の漫く	△	あさし		あさし	4		形容詞	未然	詠艶	巻四																
204	718280	08-o	4	344	漫し	から情を出じやれ。井戸は運けられし、漫し	△	、駆轡油が入ば筋向ひにあり。		あさし	4		形容詞	未然	詠艶	巻六																
208	892720	22-e	6	418	漫ひ	死蛤の焼立られて、袖の庵の漫ひ	△	心に、深き所を貢て、ひさしく		あさし	4		形容詞	未然	詠艶	巻六																
1001	2112000	25-u	10	52	漫ひ	脇窓一縫となつて、思ひ川をこす。漫ひ	△	心にあらねば、瀬のはやき時		あさし	4		形容詞	未然	男色	巻一																

図12 西鶴作品全語彙データベースの内容

「あさし」の例

せ	ならべてなげわたし橋も物すごく下は いへば。鳥居幾度か越て	瀬(セ) のはやき浪もくだけてたましゐ	五女 5-3 10才11 486
せい	心立ちもかしこし。大橋は、 して踊出る。そもそも此男。	背(セ) のはげたる末社が申は。人に	諸艶 2-3 16才9 219
せい	てかへりし其様すさましやすぐれて 兎角は身遇が大事。明日から	せい 高くうるはしく、目つきすずやかに。	一男 6-2 05才5 117
	中より物こし去逆はやさしく。そち達	せい たんにて猪首にして。口中いやな事	諸艶 4-1 03ウ9 256
	に。或拾貨目の借銀有べし。	せい 高かしらは唐獅子のごとくちぢみ	五女 3-4 16才3 460
せい	親仁が念を入ると。 気を取事を得て。	情(セイ) を出じやれ。井戸は遠けれども	諸艶 4-2 08才3 260
	の玉を。つまぐりもせず 給ふといふ。我等は國元の	情(セイ) に入て。ひとしほはやかつてと	諸艶 5-5 18ウ2 291
	と。しばし見をくる事いかなる男も ば何の役にも立ぬ	情(セイ) 出しても。そんの行商ひ事也	諸艶 6-2 08才12 344
	此兩人栄花をきはめ。世間の	せい をいわるる。其石垣がくづいたら。	諸艶 2-5 23オ7 227
せいえもん	揚屋は江戸屋八幡屋。笠屋の の二の橋で窓時やうやうかけ付。	せい 。むかしにかはらず。是ひとつは半留	諸艶 6-3 10ウ7 283
	まで色人よりのおくり酒。昆布屋の	せい に見せ懸しも。湯出蓮さげたる參も	諸艶 6-5 19ウ11 292
せい	せいがいなみ 上には むらさきしぶりに。	せい ばかりなれば大夫てなくば望みなし。	一女 2-1 04オ3 526
	せいがんじあがるちょう 見しりて御幸町通	せい にて。作り口舌して重て噛しにくる事	一女 2-2 11ウ6 534
	せいがんじどおり に。忍びて人をつけける	せい に。鼻紙入より女郎の文出して大夫	一女 5-2 07オ4 590
せい	さしよせ。酒事にして。山市晴嵐	盛(ゼイ) をやめさせ。いよいよ諸わけ。	一男 6-7 21オ8 133
せいざえもん	三野のさかりに。高島屋の	清右衛門 に行て。此里の女郎見る事	諸艶 6-2 05ウ6 299
せいざん	開ける梅唇に春を覚え。	清右衛門 に様子を聞て。兎角は命が	諸艶 6-2 08オ3 301
せいし	て。然ばほれぬといふ。 女郎に掛たらば。はじめから	清右衛門 が座敷にして是より精進入りとて	枕久 1-6 373
	次第に薄みつちやもしほらしく。磨土の	青海浪(セイガイナミ) 紋所は、銀にて	一男 4-6 16ウ8 085
	おもはくの外京川原町の事十七歳	誓願寺上る町(セイグハンシア)	一女 3-1 02ウ3 546
	かはいがられ。にくからず。かための の伝七も。傾城二十三人の。	誓願寺通(セイグハンジトリ) のすへ	五女 3-1 04ウ1 439
	のために。こころの外の の団が諸分物語の事二	西湖(セイコ) の万景。此一景にまさらん	諸艶 2-2 08ウ7 213
	おもはぬかたさまなれば。此身 まだこず。女郎のめいわくがる。	清左衛門 梅への太夫。花月三笠を同じ心の	諸艶 6-1 08ウ3 297
	或時我を思ふとの。 子共手代の異見にも。彼	青山(セイザン) かはつて白雪の埋む時冬	一女 1-1 03ウ1 502
	して。科からさきへ逐るる	せいし を書といふ。それはまことで御座る	諸艶 1-2 08ウ1 187
	雲紙(セイシ) が。美なる面影も思ひ合。	せいし なども出て。二度目には門迄	諸艶 6-2 08ウ11 281
	雲紙(セイシ) のうるし判奈良木辻町の事	雲紙(セイシ) が。美なる面影も思ひ合。	諸艶 7-3 12ウ9 323
	雲紙(セイシ) のうるし判奈良坂や	雲紙(セイシ) のうるし判奈良坂や	一男 2-0 01オ8 025
	雲紙(セイシ) うるし判の。くちぬまで	雲紙(セイシ) うるし判の。くちぬまで	一男 2-4 10ウ1 035
	雲紙(セイシ) をつき集め。是も羽織にし	雲紙(セイシ) をつき集め。是も羽織にし	一男 6-7 20オ1 132
	雲紙(セイシ) 一枚書申候。則あのかた	雲紙(セイシ) 一枚書申候。則あのかた	一男 7-5 14ウ9 149
	雲紙(セイシ) は異見の種一江戸京大坂	雲紙(セイシ) は異見の種一江戸京大坂	諸艶 1-0 01オ8 179
	雲紙(セイシ) は異見のたね競べ物なき	雲紙(セイシ) は異見のたね競べ物なき	諸艶 1-2 06ウ1 184
	雲紙(セイシ) を仕るが。貴さまは何とと	雲紙(セイシ) を仕るが。貴さまは何とと	諸艶 1-2 07オ10 185
	雲紙(セイシ) を工めといふ。常の事は	雲紙(セイシ) を工めといふ。常の事は	諸艶 1-2 07ウ6 186
	雲紙(セイシ) をかけといふ。其儘書べき	雲紙(セイシ) をかけといふ。其儘書べき	諸艶 1-2 08オ4 186
	雲紙(セイシ) を取出し。悪所ぐるいにも	雲紙(セイシ) を取出し。悪所ぐるいにも	諸艶 1-2 09オ1 187
	雲紙(セイシ) を取て。うれしがるこそ	雲紙(セイシ) を取て。うれしがるこそ	諸艶 1-2 09オ6 187

図13 新編西鶴全集第1巻のKWICの一部

ああ	やどりの暗ともよめり。 に、花なり、紅葉なり。 かんじんの時肩にかみつき、 ば、是から辨れすべし。 黄金不買漢宮顔とこそ。	ああ 現かぬしか。新之助せめて霜ならば ああ 一度はちらぬ身か。子細を聞 ああ なんともならぬならぬ、こりや気がとをふ ああ 浮世のおもひ出と、大事の大夫殿に薬 嗟(アア) 不言不笑。それよ、馬を	男色 1-2 10才5 211 男色 2-5 29才6 257 盛衰 5-2 19才2 629 盛衰 5-5 19才9 700 男色 6-3 12才10 346 男色 3-2 10才3 270 盛衰 3-2 06才7 842 盛衰 3-4 15才7 651 盛衰 5-3 11才10 692 男色 5-2 09才5 317 盛衰 1-3 12才10 604 盛衰 3-1 01才7 637 盛衰 3-4 15才7 651 男色 8-3 11才12 394 伝来 3-3 15才13 451 伝来 3-1 05才6 463 伝来 3-2 10才3 468 伝来 1-3 19才10 431 諸国 2-4 12才2 37 男色 1-2 08才7 210 伝来 1-1 07才10 418 男色 4-3 11才8 298 男色 1-1 02才3 204 男色 2-1 06才5 234 伝来 5-1 03才10 508 伝来 5-1 03才11 606 男色 1-4 18才1 219 男色 6-2 07才1 341 男色 8-1 05才11 388 伝来 6-3 11才7 534 男色 5-2 09才12 318 男色 8-3 15才5 398 盛衰 3-4 16才1 651 盛衰 4-4 11才7 673 伝来 1-4 22才6 434 伝来 4-1 03才13 482 伝来 3-4 17才9 476 伝来 8-2 11才6 577 伝来 8-4 18才13 585 盛衰 1-3 12才5 604 伝来 1-4 21才1 435 諸国 2-6 16才10 42
ああら	なし、徳之助小総をうつて、	ああら 有かたの御吊やなど飜出して、下	男色 3-2 10才3 270
あい	此三人は、世わたりにかしこく、 今は長じて、闇の又 彼女、をつとつての撥音、 と、独習いふ。しづま鼻紙の 脣にすまたの手拍子、盃の 付をくりの銀、大隅が目鼻の やめず。今は長じて、	あい もをとらぬ分限、朝暮小判を溜る、 あひ、孫あひなどいひなして、酒にいたます あひ の手のよい程、うたは消て 間(アヒ) より、定札六七枚もづきしを取り出し、 間(アヒ) を頼めば、もとへもどさず、揮筆 間(アイ) 。あとさきかまはぬ物、闇の夜に 間(アヒ) の又あひ、孫あひなどいひなして、酒 あひあひ 申て数かさなり、偽りなしに醉出 相当る(一アター) なり。拙者ためには従弟みづから 相当れ(一アター) は、せひに念願晴さんと、思ひ あひうち に、兵之助、宇右衛門が左かいな打をとせば 相うち に切こまれ切込て、うき世の夢と 相生(アイオヒ) の松風うたひ立に、老人は住吉の 哀王(アイワウ) は竜陽君を念友に定まりて後、女乱 あひおとる べき心底にあらず。然も左京は采女 相思(アヒオモベ) と作れば、杜子美も又、今宵武州 相鍵(アイカギ) をこしらへ、闇にも星の林を 相鍵の三吉(アイカギー) 、目振間の虎蔵、貢穴の權	盛衰 3-4 15才7 651 男色 5-3 11才10 692 盛衰 1-3 12才10 604 盛衰 3-1 01才7 637 盛衰 3-4 15才7 651 男色 8-3 11才12 394 伝来 3-3 15才13 451 伝来 3-1 05才6 463 伝来 3-2 10才3 468 伝来 1-3 19才10 431 諸国 2-4 12才2 37 男色 1-2 08才7 210 伝来 1-1 07才10 418 男色 4-3 11才8 298 男色 1-1 02才3 204 男色 2-1 06才5 234 伝来 5-1 03才10 508 伝来 5-1 03才11 606 男色 1-4 18才1 219 男色 6-2 07才1 341 男色 8-1 05才11 388 伝来 6-3 11才7 534 男色 5-2 09才12 318 男色 8-3 15才5 398 盛衰 3-4 16才1 651 盛衰 4-4 11才7 673 伝来 1-4 22才6 434 伝来 4-1 03才13 482 伝来 3-4 17才9 476 伝来 8-2 11才6 577 伝来 8-4 18才13 585 盛衰 1-3 12才5 604 伝来 1-4 21才1 435 諸国 2-6 16才10 42
あいあい	の悪き盃の時呼出せば、	あひあひ 申て数かさなり、偽りなしに醉出	男色 8-3 11才12 394
あいあたる	千塚氏の何がし十三年忌に を悦び、此三月二十七日祥月命日に あいうち 爰に一命おします、ふんごみて ば、つるに宣人共にたたかひつかれ、	相当る(一アター) なり。拙者ためには従弟みづから 相当れ(一アター) は、せひに念願晴さんと、思ひ あひうち に、兵之助、宇右衛門が左かいな打をとせば 相うち に切こまれ切込て、うき世の夢と 相生(アイオヒ) の松風うたひ立に、老人は住吉の 哀王(アイワウ) は竜陽君を念友に定まりて後、女乱 あひおとる べき心底にあらず。然も左京は采女 相思(アヒオモベ) と作れば、杜子美も又、今宵武州 相鍵(アイカギ) をこしらへ、闇にも星の林を 相鍵の三吉(アイカギー) 、目振間の虎蔵、貢穴の權	伝来 3-3 15才13 451 伝来 3-1 05才6 463 伝来 3-2 10才3 468 伝来 1-3 19才10 431 諸国 2-4 12才2 37 男色 1-2 08才7 210 伝来 1-1 07才10 418 男色 4-3 11才8 298 男色 1-1 02才3 204 男色 2-1 06才5 234 伝来 5-1 03才10 508 伝来 5-1 03才11 606 男色 1-4 18才1 219 男色 6-2 07才1 341 男色 8-1 05才11 388 伝来 6-3 11才7 534 男色 5-2 09才12 318 男色 8-3 15才5 398 盛衰 3-4 16才1 651 盛衰 4-4 11才7 673 伝来 1-4 22才6 434 伝来 4-1 03才13 482 伝来 3-4 17才9 476 伝来 8-2 11才6 577 伝来 8-4 18才13 585 盛衰 1-3 12才5 604 伝来 1-4 21才1 435 諸国 2-6 16才10 42
あいき	而神不安也。胸痛 故、薦於中、而	* あい気(アイキ) 者、気虚、不能健運。故、 * あい気(アイキシ) 、或滞於上則為胸痛。以上	男色 8-3 11才12 394
あいきやく	同道にて参り、此外に	相容(アイキヤク) はなし。半弥殿はいまだ御若年の 愛敬(アイキヤウ) そなはりて有ける。是そよしや、	伝来 3-3 15才13 451 伝来 3-1 05才6 463 伝来 3-2 10才3 468 伝来 1-3 19才10 431 諸国 2-4 12才2 37 男色 1-2 08才7 210 伝来 1-1 07才10 418 男色 4-3 11才8 298 男色 1-1 02才3 204 男色 2-1 06才5 234 伝来 5-1 03才10 508 伝来 5-1 03才11 606 男色 1-4 18才1 219 男色 6-2 07才1 341 男色 8-1 05才11 388 伝来 6-3 11才7 534 男色 5-2 09才12 318 男色 8-3 15才5 398 盛衰 3-4 16才1 651 盛衰 4-4 11才7 673 伝来 1-4 22才6 434 伝来 4-1 03才13 482 伝来 3-4 17才9 476 伝来 8-2 11才6 577 伝来 8-4 18才13 585 盛衰 1-3 12才5 604 伝来 1-4 21才1 435 諸国 2-6 16才10 42
あいきょう	き。いづれの人にも、	相容(アイキヤク) はなし。半弥殿はいまだ御若年の 愛敬(アイキヤウ) そなはりて有ける。是そよしや、	男色 1-4 18才1 219 男色 6-2 07才1 341 男色 8-1 05才11 388
あいくち	ためし、誰かまふものもなく、	相口 はおとさぬやうに、頗ばかりの風	伝来 6-3 11才7 534 男色 5-2 09才12 318 男色 8-3 15才5 398 盛衰 3-4 16才1 651 盛衰 4-4 11才7 673 伝来 1-4 22才6 434 伝来 4-1 03才13 482 伝来 3-4 17才9 476 伝来 8-2 11才6 577 伝来 8-4 18才13 585 盛衰 1-3 12才5 604 伝来 1-4 21才1 435 諸国 2-6 16才10 42
あいこころう	書付をあぐべきよし、畏て	相心得(アヒコロヘ) 、いづれも、其頭頸に触て	男色 8-3 11才8 298
あいさつ	此親仁ちかふよびて、亭主の おどろかぬ男進て、人間に	あひさつ 迂もなし。こなた様には私に御	男色 1-1 02才3 204
	今にすたらぬ宿のかかの ながら、是ほど合点のゆかぬ	あひさつ するごとく、おのれ其身をして、若造	男色 2-1 06才5 234
	年比さばかりのよしみもなく、殊に 是非なく、色色思案めぐらし、蒙て	あひさつ よし。其比は色あそびの世盛、市橋	伝来 5-1 03才10 508
	酌るるもはづしく、よい程に	あひさつ はなし。広鶴の大臣ども、昼は世間	伝来 5-1 03才11 606
	共、清五郎無縫娘になりて、	愛撫(アヒサツ) よからぬ人の妻子を預り、かくし	男色 1-4 18才1 219
	は兄なれ共、様子あつて	愛撫(アヒサツ) 悪敷人もやと吟味せしに、	男色 6-2 07才1 341
あいじまん	首尾して、もつてひらきて	挨拶(アヒサツ) して帰り、其翌日たまり兼て、	男色 8-1 05才11 388
あいしらう	真旨の親のごとくいひをしへ、	挨拶(アヒサツ) そこそこにしらけたるに、首尾	伝来 6-3 11才7 534
あいす	うつくしき娘を盗て、二三日も	挨拶(アヒサツ) 悪敷、久久不通なりしが、此度立越	伝来 4-4 11才7 673
		相自慢(ージマン) 、後には、久七がむかしをのこす	伝来 1-4 22才6 434
		あひしらひ 紹へは、此娘いとかしこくも、今	伝来 4-1 03才13 482
		あいし ては又帰しぬ。是を不思議	伝来 3-4 17才9 476
			伝来 8-2 11才6 577
			伝来 8-4 18才13 585
			盛衰 1-3 12才5 604
			伝来 1-4 21才1 435
			諸国 2-6 16才10 42

図14 新福西鶴全集第2巻のKWICの一部

# 古典学のための情報処理

安永 尚志  
及川 昭文  
原 正一郎  
武井 協三  
松村 雄二  
中村 康夫  
山田 哲好  
アンドル・アーマー<sup>1</sup>  
柴山 守  
山田 稔治

## 【要旨】

情報処理計画研究班は、後期研究計画(2001年～2002年)の発足に当たり、前期(1999年～2000年)までの情報処理関連の2つの計画研究班の統合を行い、より効果的な研究推進を図った。2班の合併により、それまでの研究成果の分析評価を行い、研究課題の見直しと再構築を進め、さらに研究組織の再編成を行い、リスタートした。

4年間を通じての研究課題は古典学のための情報処理である。とくに、日本古典文学とその周辺に研究対象をしづり、標準的な電子化テキストの実証的研究とその高次活用のための流通方式の開発研究を目的としている。

研究課題とその実施は、以下の3点に要約できる。

- (1) 標準的な電子化テキストの作成
- (2) テキストデータベースの流通
- (3) コラボレーション研究

以下、順次概要を述べる。

- (1) 標準的な電子化テキストの作成  
古典学における電子化フルテキストの有効な構築法

を実証的に検討するため、実際に国際標準に基づいたマークアップによる電子化テキストを作成し、とくに後期を通じて実証実験を進め、評価した。すなわち、SGML/XMLに基づいた日本古典作品の電子化を行い、評価し、有用性を確認した。とりわけ、前期の研究成果であるKOKINルールに基づく日本古典文学本文データベースを基に、データ構造の変換法を確立し、テキストデータの共有化を促進した成果は大きい。さらに、原本イメージ情報とそのテキストとの比較や対応を含むマルチメディア処理の研究を進め、プロトタイプシステムによる実証実験を行った。

一方、古典テキストのデジタル化、データベース化を支援するためのソフトウェアの開発、さらに人文科学における様々なデータベースをWWW上で公開するための新しい手法BB-DB(Bare Born Database)を開発し、その実現を図った。

## (2) テキストデータベースの流通

第1に、自立型のデータファイルとして流通をはかった。光磁気ディスクやCD-ROMなどのパッケージメディアを利用し、データ編成・流通などの方式と応用システムについて、検討を加えた。一般に、人文科学における研究態様はパーソナル環境を重視するため、

身近な素材の提供という観点からはパッケージメディアによる方法は一定の評価はあるが、総合的かつ網羅的な研究資源とするには限界がある。

そこで、第2として、インターネットを前提としたデータベースの構築と流通をはかることが不可欠という認識の基に、システム環境の整備を含め実証実験を実施し、評価を加えた。とりわけ、後期においては日本古典文学本文データベースの実用版として、インターネットに完全に対応したCSS（クライアントサーバシステム）によるデータベースサービス環境を構築し、実用レベルの運用管理を実装し、提供している（試験公開）。

1999年より2ヶ年間、岩波書店刊行旧版「日本古典文学大系」の全100巻約600作品のフルテキストのデータベースの実験公開を行った。引き続き、2001年より、上述のシステム改善を加えた新バージョンの試験公開を行っている。これは現在も続けられているが、利用者も2,000人を超える、アクセス件数も10万件に及び、極めて有効であるとの評価が高い。

### （3）コラボレーション研究

複数の情報資源をネットワークで統合し、利用者に仮想的に1つの情報資源を提供し、共同研究を実施するための研究環境であるコラボレーションシステムの開発研究を開始し、成果を得た。すなわち、後期では分散情報資源の相互利用法（コラボレーション）による新しい古典学の形成の可能性を探る研究課題の基に、Dublin CoreによるメタデータとZ39.50による標準検索プロトコルの利用可能性を研究し、その有用性を確認した。実際にプロトタイプシステムを構築し、評価を得た。なお、このためにはさらに付加価値の高いテキスト処理機能と応用システムの開発研究が、不可欠であると考えている。

本報告書では、上記3点に焦点を絞り、研究経緯の概要を述べ、研究成果をまとめる。

### 【連携の成果】

特定領域研究に参加する研究者、および関連する研究者全員に対して、日本古典文学本文データベースの利用を承認し（知的財産権により利用規則を定めており、登録が必要）、研究に役立てていただいている。また、その利用経験などから、テキストの誤りなどの指摘を受け、随時データベースを修正している。さらに、システム機能の拡充の要請も多く寄せられ、システム開発に盛り込んでいる。

一方、1999年、特定領域研究「人文科学とコンピュータ」が完了し、その各種情報資源および研究成果の継承と継続的進展を目的の1つとして、本特定領域研究「古典学の再構築」に、「古典学のための情報処理」計画研究班が組織された経緯がある。このような言わば人文科学領域における永年の基盤整備の効果により、本計画研究班への理解と期待が深まり、極めて有効切な連携の体制が整ってきたことは、特筆に値するものである。

このような連携の中で、本計画研究班が抱える研究課題は多いが、具体的な研究成果も着実にあげることができたと考えられる。

より信頼度の高いデータ作成には、多くの専門家の協力が不可欠である。文学、語学はもとより、歴史学、民族学、民俗学、日本学など多くの人文科学の国内外研究者との連携を深め、コラボレーション研究を実施してきた。さらに、出来上がったあるいは開発途上の電子化情報資源の高次活用は、その専門領域の研究者にさせられる。そこからの様々な知見や要求仕様がなければ、研究に役立つ情報資源は成り立たない。

この意味で、本計画研究の研究課題の重点の1つはコラボレーション（電子的協調作業方式）研究であり、後期は中心的課題と位置づけ、推進している。

情報処理に対する人文科学領域の研究者の具体的なニーズの分析、開発されたソフトウェアの評価などにおいて、我々との忌憚のない意見交換を進めてきた。その成果の一つはDD-DB手法の開発である。これは、データベースやソフトウェアについての専門的な知識がなくても、自らが持つさまざまな資料（データ）をデータベースとして、WWW上で公開するための一連の処理手法をソフトウェアを含めて開発したものである。

### 【位置付け】

本計画研究は、日本文学を主とする古典テキストの国際標準に準拠したデジタル化の手法を研究し、またそれに基づき実際にデータ作成を行って、実利用環境において検証し、評価する。そのための研究プロセス全般の情報技術に関わる開発研究と位置づけている。

また、作成したデータベースはインターネット上に公開し、その流通性、高次活用性などを、人文科学領域の具体的研究テーマに即して検証する。さらに、古典研究者が古典テキストに対して情報処理を試みるのに必要な基本的ソフトウェアの整備、WWW公開のための手法開発を目指している。

後期の研究の柱の1つはコラボレーション研究であ

る。コラボレーション研究は文学研究を支援するための方法論の構築と位置づけている。研究者に原本やテキストだけではなく、多様な関連情報を簡便に提供することである。関連するあらゆる情報に容易にアクセス可能とすることである。この場合、単一のインターフェースで精度良くアクセスできなければならなければならないことは言うまでもない。

以下に、本研究の位置づけを5点に絞り、まとめておく。

- (1) 計画研究班の統合
- (2) 日本古典文学本文データベース（実験版）の試験公開
- (3) 日本古典文学本文データベース実用版の試験公開
- (4) SGML/XML
- (5) コラボレーション研究

以下、順次簡潔にまとめる。

### (1) 計画研究班の統合

前期計画研究（1999年？2000年）において、それまで2つの計画研究班（安永班と及川班と仮称）は、主にテキストデータのデジタル処理技術の開発研究を担当してきた。安永班は、デジタル化された古典テキストの公開および利用のための諸技術開発を中心とした。一方、及川班はテキストのデジタル化および表記のための基礎技術開発を中心に研究を進めた。

テキストデータ作成においては、いずれも当初の研究目標は達成し、その活用体制整備が今後の課題となっている。取り扱う研究対象としての分野が異なるが、データ記述文法は概ねSGMLを基本としている。それらの事情を踏まえ、またより効果的な情報資源の開発研究の推進と高次活用への展開を考慮し、両計画研究班の統合合併を実施した。

とくに、今後これらの諸技術を統合することにより、相乗的あるいは相補的な効果が期待される。そこで、まず後期に当たっては具体的にWWW上の古典テキストの公開と活用を目標に研究を進めていくことが肝要と認識された。

また、新しい試みとしてパラレルテキスト・データベース（例えば、日本語の古今和歌集と英訳された古今和歌集）の構築・利用手法の開発研究を進めていくには、両班で開発した技術やノウハウを活用する必要があり、そのためにも両班の統合が望まれた。

### (2) 日本古典文学本文データベース（実験版）の試験公開

前期の研究の範囲とその研究成果をまとめた。1999年4月15日から2ヶ月の予定で、懸案の日本古典文学本文データベース（実験版）の試験公開を開始した。日本古典文学本文データベース（実験版）は、岩波書店刊行の旧版「日本古典文学大系」全100巻の全作品の本文（テキスト）をデータベース化したものである。本文データベースは国文学研究資料館大型コンピュータで実験的に管理した。本文データベース検索システム（実験版）は、この本文データベースをインターネットに接続されたパソコンなどから、利用するための試験的なシステムである。国文学研究資料館のホームページからアクセス可能とした。全100巻中の約600作品から作品を選び、語彙を検索したり、全文を表示したり、ダウンロードすることなどができる。

試験公開としている理由と目的はいくつかあるが、2つあげておく。

第1は、より良いデータを作ることである。このデータベースの計画は1987年（昭和62年）にスタートしているから、試験公開に至るのに10年余を要している。ただし、実際にはデータ入力が終わった段階で中断の止む無きに至り、リスタートしたという経緯がある。リスタート時点での本文データベースは、データファイルであってデータベースではなく、またデータ校正も完了していなかった。したがって、本文データベース化を進め、同時に最低限のデータ校正を行う必要があった。とくに、従来の研究の範囲ではデータ校正に最善を期すというパワーを持ち得ず、試験公開による利用者のフィードバックに期すこととせざるを得ない。

第2は、出版社との知的財産権などに関わる問題である。印刷刊行された古典テキストの全文をデータベース化するには、出版社との契約が必要である。この他にも、翻刻、校訂者の著作権あるいはデータベースの著作権など多くの解決すべき課題がある。

試験公開では、本文データベースはブレーンテキストとKOKINルールによる符号化テキストの提供を行っている。システム機能も主としてダウンロードによる利用者自身の環境への資源提供を柱とする。語彙検索、語彙統計程度の必要最低限のシステム機能の提供しか行っていない。

ここでの本文データベース研究は、テキストデータ記述の研究、並びにそのデータベース化の研究である。KOKINルールは古典文学作品テキストのデータ記述

のための記述文法である。すなわち、マークアップ文法である。これはテキストの本文構造、レイアウトなどのメディアに関する表現構造を記述でき、また様々な表記（傍記と言う）を記述可能である。さらに、品詞や古文独特の掛詞などの表現や意味構造をも表現可能である。

電子化テキストはブレーンテキストでもよいかも知れないが、やはり符号化テキストが望まれる。また、データ記述されたまでのデータ流通が考えられるが、出来ればデータベース化したい。しかしながら、未だ一般に有効なフルテキストのためのDBMSはない。KOKIN ルールで記述されたデータから、関係モデルを利用したデータベースが開発されている。国文学研究資料館のホームページ経由での試験公開はこの方式である。論理レコードと言う概念を定義し、これは原本の本文の1行分に相当するが、これに属性として論理レコードの継続IDを持たせて対処している。

一方、DD-DB 手法の開発がある。これは、データベースやソフトウェアについての専門的な知識がなくても、古典学研究者が自ら持っている様々な研究資料（データ）をデータベースとして、WWWで公開するための一連の処理手法をソフトウェアを含めて開発したものである。

### (3) 日本古典文学本文データベース実用版の試験公開

後期（2001年度から現在）の研究範囲とその研究成果をまとめた。前期の日本古典文学本文データベース（実験版）の試験公開は、大型コンピュータに実装されたデータベースで運用を行った。国文学研究資料館の情報システムを CSS 分散管理システムに入れ替えたことで、本格的な WWWによる新たな日本古典文学本文データベースの実用版として、サーバコンピュータに再構築し、運用することとした。

ほとんど新たなシステム開発であったが、システム機能としては前期を完全に包含し、かつ使い勝手などはほぼ同等としている。ただし、性能改善を進め、追加、拡充した応用面での機能も多い。例えば、検索機能の1つに語彙索引検索機能があるが、1つの語彙の出現用例として、その語彙の前後（距離は自由に指定）はもとより、2語彙（語彙間の距離は自由指定）の共出による用例検索なども可能としている。これはまた複数の作品にわたって可能としている。すなわち、懸案の複数作品にわたる横断検索の実現である。

実用版のシステムであるが、出版社との知的財産権

の取り決めなど諸般の事情もあり、試験公開として運用している。そのため、利用者の登録を必要とする。本特定領域研究「古典学の再構築」の研究者は優先的に利用権を認めている。試験公開は2001年4月から、現在に至っており、本研究終了後も引き続き国文学研究資料館において運用を行う。

現在までに、海外の研究者を含め2,000名の登録があり、利用件数（主として作品のダウンロード件数）も、10万件に及んでいる。この種のデータベースでは、極めてよく利用されているデータベースと言える。

日本古典文学本文データベースを活用して研究を実施する場合の研究態様などについて調査研究した。すなわち、よく利用している利用者を選び、アンケート調査を実施し、データベースの使われ方、データベースが利用者の研究活動に与えた影響などについて、調査研究を行った。興味ある結果を得ているが、以下のような研究に与える効果が確認されている。

このデータベースは網羅性があり、また信頼度も高いので、これを前提とした研究が行われ始めている。さらに、研究のスピードをあげたり、生産性を高めたりする効果が出始めている。なお、詳しくは後述の研究成果の章でまとめているので、ここでのこれ以上の言及は割愛する。

後述のように、本データベースには研究成果の1つである SGML/XML 化した数10点の古典作品を追加している。作品数はデータ作成が整えば順次追加し、増やしていく予定である。それらのDTD 記述も含め、古典文学作品のデータ記述、データ構造を、我が国では初めて公開した。反響は大きい。

### (4) SGML/XML

日本古典文学作品の全文データベースの構築に着手した当時（1987年頃）、標準的なデータ記述文法、例えば SGML（Standard Generalized Markup Language：標準汎用マークアップ言語）は仕様検討の段階で、当然ながら日本語処理の可能な SGML 用のツールも存在していなかった。その後の進展を待ったが、最近まで普及せず、そのため独自のマークアップ規則を作成することになった。この独自の規則は、後ほど SGML との仕様比較の結果、その基本的なアイデアおよび概念は SGML と同じであった。このマークアップ規則を KOKIN ルール（KOKubungaku INformation Rules）と呼んでいる。KOKIN ルールは文学研究者が利用できるように、明快性と簡潔性を重視して設計されている。

KOKIN ルールの有効性を検証するために、多くの古典テキストの電子翻刻を試みた。これまでに、岩波古典大系（旧版）、嶽本大系、假名草子集成など、約150巻、約5,000万文字の電子化が終了している。その結果、KOKIN ルールは古典テキストを電子翻刻する上で、必要最小限の記述能力を有していることが確認され、有効適切なデータ記述言語の1つとの結論を得た。

しかしながら、KOKIN ルールはやはりローカルな仕様である。したがって、データ処理のためのツール、例えば構文解析ツールなどをすべて独自に開発する必要があった。DTD を明確に微細に定義していないため、作品ごとに現れる個別記述や構造を、言わば例外ルールとして取り入れざるを得ない。そのため、規則に揺らぎを生じてきた。

その後、テキストの電子翻刻やシステム間の電子的テキストデータ交換の手段として、SGML が採用されるようになってきた。このような状況を考慮した結果、KOKIN ルールに基づいて形成されてテキストデータを SGML に基づいた形式に変換し、電子テキストの効率的な管理と利用の促進を図ることにした。とくに、この場合データを再作成するのではなく、変換プロセスにより自動変換できれば極めて有効である。後述するが、KOKIN ルール化テキストデータを SGML テキストデータに変換する手法を研究し、成功した。この成果は大変に大きい。

日本古典文学作品の電子化テキストのデータ記述に SGML を中心とおくことは、以下のような観点で極めて有効適切と考えられる。

SGML で書かれたデータを取り扱う有効な DBMS は未だないが、例えば Open-Text などの文字列サーチ型フルテキスト検索システムが強力な手段となることが分かり、現在この方式が主流となっている。この場合 SGML は構造を記述しているので、単なる全文検索ではないより深い構造検索が可能となる。また、SGML から HTML, XML への変換は比較的簡単で、それにより直ちにインターネットで提供可能となる。さらに、SGML をベースとすることにより、コラボレーションシステムが現実解として期待される。本研究では、具体的にメタデータと標準の情報検索システムを用いるシステムの実証実験を開始している。メタデータは Dublin Core の適用を考えている。また、標準の情報検索プロトコルとして、Z39.50を考えている。

現在、SGML データから、XML データの形成に展開している。今やインターネットによる情報発信は当たり前のことであり、SGML から XML への展開は不

可欠なことである。とくに、SGML の仕様は非常に複雑であり、その仕様を完全に満たすツールは存在しない。例えば、DSSSL (Document Style Semantics and Specification Language : テキストスタイル意味指定言語) を完全に実現している市販ツールは皆無である。データ処理上の幾つかの問題点が存在している。そこで、上述のように作成あるいは変換された SGML データを、XML 互換の形式に変換することとした。

実際に、変換プログラムを開発研究し、日本古典文学大系中源氏物語、古今和歌集等々の約220作品ほどの XML 化データが作成され（現在も続いている）、上述のように試験公開されている（公開は現在20作品）。

## （5）コラボレーション研究

コラボレーション研究の目的は、文学研究を支援するための方法論の構築である。研究者に原本やテキストだけではなく、多様な関連情報を簡便に提供することである。例えば、古今和歌集の研究者に対して、和歌が詠まれた土地、民俗、風俗などの地理的情報、その時代の歴史的事象や事実、人口や経済などの統計情報、あるいは同時代の周辺地域の政治情勢など、関連するあらゆる情報にアクセス可能とすることである。この場合、単一のインターフェースで精度良くアクセスできなければならない。

具体的には、我々の所属する国文学研究資料館が有する日本文学研究資料および歴史研究資料を主要材料とした上で、次の2つのモデルを研究することにした。

- (1) 複数の機関あるいは個人が作成している多様な人文科学系データをデータベースとして、組織化するモデル。
- (2) 複数のデータベースをあたかも単一のデータベースであるかのように再組織化するようなモデル。

なお、コラボレーションシステムを考える上での必要十分条件は、あらゆる情報資源が標準記述されていることである。我々は SGML/XML に基づくデータを基本と考えることとする。したがって、これまでの情報資源の再定義が不可欠な要件となる。

## 【研究成果】

### 1. 日本古典文学本文データベース

#### 1.1 実験版（1999年～2000年）の運用実験

##### 1.1.1 利用者

日本古典文学本文データベース（実験版）の利用の状況を評価する。本節の内容は、総括班中間報告書に報告済みの内容に若干の手を加えたものである。以下の考察において、利用統計は全て1999年4月15日から2000年12月末までの21ヶ月間のものである。ただし、データベースは大型コンピュータであり、運用時間の制限があるため、実際にこの間に利用できた日数は397日である。全期間を通じての利用回数は42,000件を超える。1日当たり平均約100件強の利用があった。作品のダウンロードは約35,000件を数えた。

本文データベースの利用には利用登録が必要である。この期間に1,473名の登録があった。余り意味がないが、1人当たりの平均の利用回数は約28回である。なお、個人の最大利用回数は1,099回である。また、登録したものを利用のない人も多く、約45%である。実利用者は808名である。

一方、全体のうち、6%弱の海外の利用者（80名）がいる。欧米が主であるが、全世界（30ヶ国余）に渡っている。利用も全体の5～6%に渡っている。以下でも触れるが、この本文データベースは当時としては制限付きの試験公開でもあり、ほとんど宣伝をしていないが、海外でも関心が高いことを示している。日本文学の古典テキストを電子化テキストで入手し、様々な研究、教育活動などに活用するということを聞いている。海外の研究者とのコラボレーションの具体的な契機をつかむことに役立っている。主要な海外拠点として、ヴェネチア大学、フィレンツェ大学、オックスフォード大学、パリ大学、コレジュドフランス、UCLAなどとのコラボレーションが進んだ。

なお、ホームページへのアクセス回数は約34,000回である。利用回数との差異は、ホームページアクセス回数の統計採取開始は約1ヶ月遅れであること、同じアクセス中に複数の利用があることなどによる。

##### 1.1.2 利用状況

この試験公開版のシステムは高度な機能処理を提供していない。所望の作品テキストのダウンローディングを行って、後は利用者の利用環境で自在に使ってもらうことを主としている。

試験公開ではほとんど宣伝をしていない。ホームページにニュースがある程度である。したがって、利用の申し込みもホームページを見てからということになる。海外からの問合せ、アクセスもかなり多い。我々とのコラボレーションとしての共同研究者の利用もあるが、全般的に見ても海外における日本古典文学作品への関心の高さをうかがうことができる。

よく利用される文学作品の上位10位をあげてみる。上位から、今昔物語、日本書紀、源氏物語、栄華物語、宇津保物語、太平記、平家物語、万葉集（訓み下し）、古事記、萬葉集（白文）である。この順位は、初期からまた年間を通じてもほとんど不同である。さらに、後期の試験公開においてもこの傾向はほとんど変わらないので、面白い結果と言える。この傾向が何を物語るかについては興味のあるテーマであるが、ここでは詳細な検討は割愛する。ただし、一言付加するならば、これからは利用層の広がりと教育面での活用などをうかがうことができる。これらは大変メジャーな作品であるばかりでなく、量的にもかなり膨大な作品である。利用者は文学者だけではなく、広く人文科学全般に渡っていること、大学、大学院の教育ではほとんど必須の作品であることなどが、要因として考えられる。さらに、現在文学研究者の関心の高いものとして、説話文学の研究があるが、1,000余件のダウンロードがあり、これはその裏付けとも見える。

利用者から150余件の貴重な意見、問合せなどをいただいた。試験公開の目的は、使っていただきながら内容を正すということでもあり、大変貴重である。指摘していただいたテキストのエラーなどはその都度修正を行った。検索システムに関する問合せ、意見も多く、後期の実用版のシステム開発に大いに参考となつた。

##### 1.1.3 実証実験の成果

とくに、大型コンピュータを使っていることから来る運転時間の制約（平常半日運転、土日休止など）に対する要望が多かった（とくに、時差のある海外から）。国文学研究資料館において次期システムでの24時間運転の計画が進められており（2000年当時）、次期システム（現在すでに稼働中）で実現した。24時間運転は基より、指摘されたシステムの機能拡充などをはかり、2001年2月から、年度内の試行を行い、同4月からの本格稼働に備えた。

利用者からの強い要望は、フルテキストだけではなく、関連する原本画像データ、研究論文などの同時アクセスの実現である。この課題はまさしく日本文学

研究における情報資源のコラボレーションである。

一方、出版社との約束から、利用者の登録が必要である。利用目的も学術研究に限定している。多くの方々から、利用資格の緩和の要望がよせられている。現時点ではやむを得ない。現在、国文学研究資料館では、知的財産権などの協議も含め、一般公開のための諸準備が進められている。本研究班もこれに積極的に関わり、検討を進めた。

なお、後述するが、現在、2001年4月から新たな日本古典文学本文データベースの実用版としての試験公開が、国文学研究資料館のホームページからアクセス可能となっている。ホームページのアドレスをあげておく。

<http://www.nijl.ac.jp/>

以上の研究成果に基づき、後期の主要な研究テーマは人文科学におけるコラボレーションシステムの実装と、その利用による研究の新たな進展を狙うこととした。複数の情報資源をネットワークで統合し、利用者に仮想的に1つの情報資源を提供し、共同研究を実施するための研究環境であるコラボレーションシステムの開発研究である。海外の拠点を含め、実証実験を進める計画を進行している。

## 1.2 実用版（2001年－2002年）の実証実験

### 1.2.1 実用版システムの開発と運用

後期（2001年度から現在）の研究範囲とその研究成果をまとめる。前期における実験版の運用実験は多大の成果をあげることができた。その成果を引き継ぎ、大幅なシステム改善を行って、実用版の日本古典文学本文データベースを試験公開した。システム提供環境は、国文学研究資料館の情報システムのデータベースサーバとし、これは当該年度から情報資源の完全分散管理システムとして構築され、運用管理されている。

WWWを用いたCSS（クライアントサーバ・システム）方式による本格的な実用版日本古典文学本文データベースである。利用者はインターネットに接続されたパソコンなどから、WWWを経由しアクセスし、利用することができる。

ほとんど新たなシステム開発であったが、システム機能としては前期の諸機能は完全に包含し、かつ使い勝手などは一層のユーザフレンドリ方式がはかられている。性能改善を進め、新たに追加、拡充した応用面での機能も多い。例えば、文字列検索機能では1つの語

彙が選んだテキスト中のどこに現れるか（KWIC方式）においては、その語彙の前後の距離を自由に指定することができる。あるいは、2つの語彙がテキスト中に共出する用例検索では、2つの語彙間の距離も自由に指定することができる。これは、単一の作品のみならず、複数の作品に渡って可能としている。すなわち、懸案の複数作品にわたる横断検索が実現されている。

実用版のシステムであるが、前期と同様に、出版社との知的財産権の取り決めがある。加えて、諸般の事情もあり、現在試験公開として運用している。そのため、利用者の登録を必要とする。本特定領域研究「古典学の再構築」の研究者は優先的に利用権を認めている。試験公開は2001年4月から、現在に至っており、本研究終了後も引き続き国文学研究資料館において運用を行う。

現在までに、海外の研究者を含め2,000名の登録があり、利用件数（主として作品のダウンロード件数）も、10万件に及んでいる。この種のデータベースでは、極めてよく利用されているデータベースと言える。

### 1.2.2 利用の実態調査

#### （1）利用者

2002年2月の時点では、日本古典文学本文データベースを活用して研究を実施する場合の研究態様などについて調査した。その時点では、利用登録者数は約1,400名、1日平均約100件の利用があった。

よく利用している利用者約300名を選び、電子メールによるアンケート調査を実施した。新たなシステムは、2001年4月からあるから約10ヶ月間の利用のうち、10回を超える程度の利用実績のある利用者を対象としている。調査期間は2月20日から20日間ほどとした。データベースの使われ方、データベースが利用者の研究活動に与えた影響などについて調査することが目的である。

先に調査結果を述べれば、以下のような研究に与える効果が確認され、極めて興味ある結果を得た。このデータベースは網羅性があり、また信頼度も高いので、これを前提とした研究が行われ始めている。さらに、研究のスピードをあげたり、生産性を高めたりする効果が出始めている。

回答は約30%の100名弱であり、やゝ物足りないものの、自由記述欄を多く設けたので、貴重な意見を多くいただき参考とするところ大であった。回答者の分布は研究、教育に従事する者が60%，学生が30%，その他が10%であった。そのうち、利用目的あるいは研究分野では、日本文学とする者40%強、国語学とする

者30%弱、歴史学とする者10%強が主なものである。その他に若干の中国文学、宗教学などがあり、教育とする者も5%あった。なお、その他10%弱は不明である。この程度の統計でみても、日本古典文学本文データベースは日本文学、国語学を中心に様々な分野の研究と教育に使われていることが確認される。

## (2) 利用状況とその効果

利用方法に関して、次の設問を行った。すなわち、A：主に使用する機能と、B：使用する機能とその目的である。

A：主に使用する機能では、「全文表示・ダウンロード」と「語彙・文字列検索」が全体の9割以上を占めている。「全文表示・ダウンロード」では、それを自分自身で加工して利用する場合と、それ以外の目的の可能性が考えられる。同様に、「語彙・文字列検索」に関しては、特定の場所を探す場合と、用例収集を目的とする2つの利用形態が考えられる。一方、

B：使用的機能とその目的では、8割近くの回答者が用例収集のために語彙・文字列検索を行っている。同時に、特定の場所を読むための語彙・文字列検索と、本文データを加工するためのダウンロードも半数近くが行っている。

日本古典文学本文データベースの有効性とその具体的成果の設問では、回答者の8割以上が役に立ったと回答している。研究成果を得たとする回答者は34名あり、日本古典文学本文データベースを使用して作成した論文をあげた者もいるが、より具体的に何を調べたかについて多くの回答があった。興味深い事例としていくつか紹介する。

## (3) 事例

### 事例A：

古代の公文書中で紫菜と表記されていたアマノリが、平安時代に源順にとって甘苔と書かれ、次いで、うつほ物語の中で甘海苔と表記されて今日に至ったこと、同時代あるいはそれ以前に記録がないことが確認できた。

### 事例B：

例えば、日本で中世の自殺の研究をしているので、「切腹」だけではなく「捨身」「生害」「殺水」などの表現でこの現象が現れているので、関係のある表現を簡単に調べられることは、非常に役に立つ。

### 事例C：

「海鼠」という言葉が、どの程度、どこに出てくるのか、調べることができた。

次に、研究に及ぼす変化に対する設問で、データベースの利用により研究全体へ及ぼすと考えられる変化に対する主観的な感想を求めた。無変化とする1割の回答者を除き、何らかの変化を意識している。例えば、コンピュータの利用増加と、研究成果が出しやすくなつたと言う約4割の回答者があった。また、今までとは異なつたアプローチの研究ができるようになったとする積極的回答も約3割に上つた。

### 事例D：

量的な観点からみると、調査対象に劇的な増加を見、大きな流れをフォローしやすくなつた。

### 事例E：

引用が大変楽になった。また索引を引くより、語句検索をかけた方が楽にまた正確に抽出でき、重宝している。

### 事例F：

手元に所持しない作品や、所持していてもコピーなどの準備がしにくいテキストを、ダウンロードして加工すればすぐに教材として利用できるのは、大変便利であり、教育方法の改善の上で成果が上がつてゐる。

### 事例G：

とくに、近世のテキストは語彙索引が刊行されておらず、これまで用例の検索に多大の時間を要したが、データベースのおかげで近世の用例にあたりやすくなつた。

一方、良いこと尽くめではない。批判的あるいは慎重な意見もある。例えば、

### 事例H：

このデータベースから直接研究結果を出すのはためらわれる。語彙や用例検索に費やす時間は激減したが、見落としが無いか心配である。とくに本文が旧大系であるため、結局は再び確認のために検索をしている。

### 1.2.3 評価と課題

日本古典文学大系に限らず、新大系などの他の多くのシリーズや個別作品のデジタルテキストは、回答者のみならず大方の利用者の強い希望である。学術研究にとって、信頼できる多くの電子化テキストを入手したいという要望は極めて強い。これはひとえに費用と時間の問題でもあるが、それをかなり永続的に実施しうる体制や組織の問題でもあり、かつ最近では知的財産権の諸問題も含む。現在、インターネット上には様々な文学作品や歴史史料が公開されている。その他にパッケージメディアとして市販されている資料も多い。しかしながら、インターネット上で公開されるテキス

トデータの多くは、著作権などの関連から最新の研究成果を反映していない場合もあり得る。また、市販品は一般に極めて高価であり、個人が手にするのは容易ではない。このような状況の中で、研究に使用できる精確で信頼できる電子化テキストへの期待は大きい。この課題は本特定領域研究の終了後の資源の引き継ぎ、すなわち受け皿をどうするかと言う大きな問題を孕んでいる。

最後に、残った大きな研究課題について考える。すなわち、日本古典文学本文データベースは研究に影響を与えるのかという問い合わせである。この回答と考察はかなり難しい。回答からは、研究作業能率の向上や分析できるデータ量の増大は指摘できるが、研究に質的な影響を与えたかどうかは、研究者間の見解に相違がある。例えば、新しい用例の検索が容易になり、注釈に新たな知見が得られたとする評価がある。しかし、反面研究の根幹部への影響はないと言明する研究者も居る。ケース by ケースであるが、事例Aのケース（米国の歴史学者）、事例Bのケース（海苔加工の専門家）のように、日本古典文学本文データベースの存在を前提とした調査が行われていることはたいへん興味深い。

日本古典文学本文データベースの利用者は、日本文学、日本語学、日本史学の研究者が中心である。これらの主たる利用者において、研究活動に役立ったかの事例の収集はまだまだ少ない。さらに、これら以外の多くの異なる研究分野に対して、有用かどうかの評価は未だ分析に値する基本的データ自体が不足している。本調査研究はこれらの端緒をなすもので、今後の詳細な研究に期待せざるを得ない。日本古典文学本文データベースを「日本古典文学大系」の単なる電子版ではなく、研究者にとってより有用なデータベースとするためにも、今後の課題としたい。

なお、最後に付言するが、最近本データベースに研究成果の1つである SGML/XML 化した数10点の古典作品を追加し、公開した。作品数はデータ作成が整えば順次追加し、増やしていく予定である。このことは、古典文学作品の DTD 定義も含め、それらのデータ記述やデータ構造を我が国では初めて公開したものとして、高く評価される。反響は極めて大きい。

## 2. SGML/XML

### 2.1 研究経緯

#### 2.1.1 全文データベース

テキストに関する研究は主として語彙解析であり、

そのためには語彙索引を作らねばならない。語彙索引は対象となるテキスト中に現れる単語のデータベースであり、テキストを単語単位に分解し、ヨミ、品詞などの属性情報を付与したものである。欧米語のようにスペースなどの分離記号によって、単語の識別が容易な言語では、語彙解析ツールを利用して、語彙索引の作成を効率的に行うことができる。しかし、日本語テキスト、とくに古典テキストには、単語間に明確な分離記号がなく、さらに複合語を作る造語性などの問題がある。また、綴り字法は時代、ジャンル、作品により異なっている。そのため、単語の確定は研究者により差がある。

このような状況で、古典テキストを自動的に分かち書きするようなツールを望むことはできないので、語彙索引の作成は手作業が中心となる。また、語彙索引に求める内容は研究者により異なる。したがって、コンピュータを利用した語彙解析を行う準備として、まず総合的な全文データベースを作成し、そこに研究者の目的や方法に応じた多様な属性情報を付加する必要がある。

なお、全文データベースについては、さらに3.2.3で詳細にまとめる。

テキストデータ中に付加情報を埋め込むには、研究者の利便性とデータ処理の効率性を勘案したマークアップ規則を定める必要がある。まず、我々が独自に開発したマークアップ規則について、以下に簡単にまとめておく。

#### 2.1.2 KOKIN ルール（初期のマークアップ）

##### （1） 基本的考え方

我々が、全文データベースの構築に着手した1987年当時は、SGML は仕様検討が国際的に検討され始めた頃であった。したがって、SGML の普及はなく、また日本語処理の可能な SGML 用のツールも存在していなかった。そのため、我々独自のマークアップ規則を作成することになったが、最近の検証ではその基本的な考え方は SGML と同じであると確認されている。このマークアップ規則を KOKIN ルールと呼ぶ。KOKIN ルールは、日本文学研究者が容易に利用できるように、明快性と簡潔性を重視して設計されている。KOKIN ルールは、ルール #1 としてタグ規則、ルール #2 としてフラグ規則、およびルール #3 として付加値規則の 3 種類の規則として定義、構成されている。

以下、実際例として江戸期の小説を集成した東京堂

出版刊行「嘶本大系」を、説明例として用いる。これを原本というが、校訂された活字テキストである。原本には注釈、修正、漢字のヨミなど複雑なテキストデータ構造、並びに文書構造が見られる。

## (2) KOKIN ルール#1 (タグ規則)

テキストにはタイトル、節、章などの構造がある。以下では、このような文書の構造をテキストの論理構

造と呼ぶ。タグはテキストの論理構造を明示するための識別子である。マークアップは研究者のテキストの見方あるいは解析の視点を表現していると考える。我々の全文データベースでは、テキストの論理構造の定義は個々の研究者の判断に任せているが、データ交換などの便宜を考えて、タグの記述法を規則化している。これがタグ規則である。以下にタグ規則の概要を示す。

<Logical Record>	::= <Tag Begin><Tag><Tag End>   <TagBegin><Tag><Data><Tag End>
<Tag Begin>	::= 'Japanese-Yen-Mark'
<Tag End>	::= 'Star-Mark'
<Tag>	::= <Tag Symbol>   <Tag Symbol><Tag Attribute>
<Data>	::= <Line>   <Original Data>   <Repeating Symbol><Original Data>
<Line>	::= <Original Data>   <Serial Number><Original Data>
<Serial Number>	::= see Table 1
<Repeating Symbol>	::= ;
<Original Data>	::= see e.g. Figure 2
<Tag Symbol>	::= see Table 1
<Tag Attribute>	::= see Table 1

タグ規則の基本的な記述構文は [¥タグ記号 文字列 ★] である。タグ記号は全角アルファベットで表される。例えば、[¥T] はタイトル、[¥P] はページ、[¥G] は図表を表す。さらに、属性情報としての文字列が続く場合もある(表1参照)。タグ記号と [★] で囲まれた文字列が、そのタグ記号で示された論理領域となる。なお、[★] は省略可能である。

図1は、嘶本大系の論理構造の見方の1つを表したものである。ここでは「行」を論理

構造の基本と考えている(これを論理レコードと呼ぶ)。論理レコードがいくつか集まって「嘶」が構成される。嘶がいくつか集まって「小作品」が構成される。このように、テキストの論理構造は階層的であり、一般にツリー図として表すことができる。つまり、タグ規則のタグはツリー図のノード識別子であり、これはSGMLのエレメント名と同じ働きをしている。図1の論理構造に基づいて嘶本大系をマークアップした例を、図2に示す。

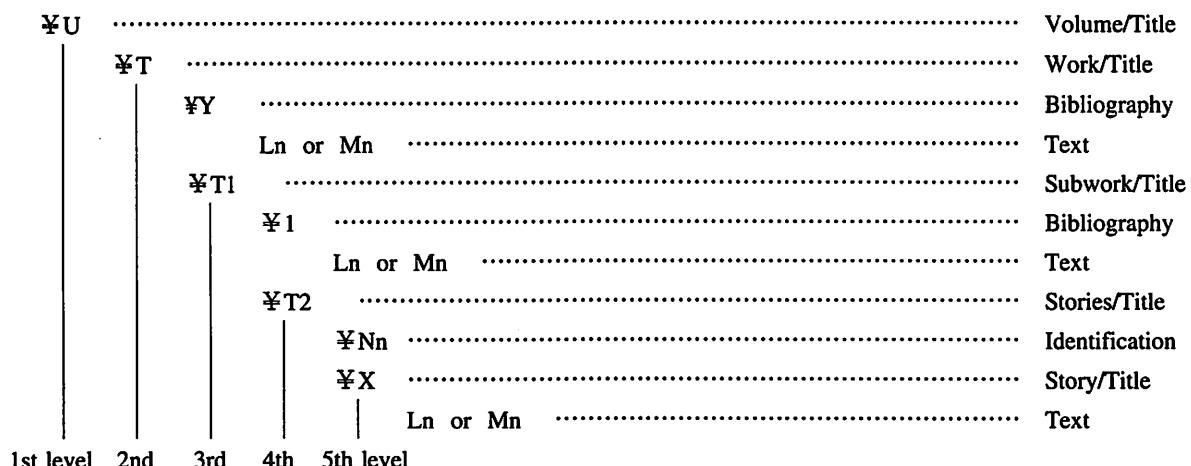


図1 嘶本大系の論理構造(例)

表1 タグの例とその役割

Tag	Attribute(n)	Role
U		Logical Recordset (Volume and its Title)
Tn	None	Logical Recordset (Work and its Title)
	1	Logical Recordset (Subwork and its Title)
	2	Logical Recordset (Group of Stories and its Title)
Yn	None	Logical Recordset (Bibliography of Work)
	1	Logical Recordset (Bibliography of Subwork)
Pn	<i>serial number</i>	Logical Recordset (Pages)
Nn	1	Logical Recordset (Bibliography of Subwork)
	<i>serial number</i>	Logical recordset (Serial Number of Story)
A		Logical Recordset (Author of Story)
B		Logical Recordset (Supplement about Story)
J		Logical Recordset (Keyword)
X		Logical Recordset (Story/Title)
Ln	<i>serial number</i>	Logical Record (Upper Column)
Mn	<i>serial number</i>	Logical Record (Lower Column)
Qn	None	Logical Recordset (Postscript of Work)
	1	Logical Recordset (Postscript of Subwork)
G		Logical Recordset (Picture)
gn	None	Logical Record (Title of Picture)
	<i>serial number</i>	Logical Recordset (Text in Picture)
H	None	Logical Record (Table)
hn	<i>serial number</i>	Logical Record (Title of Table)
		Logical Recordset (Text in Table)

00000270 ¥T 1 醒睡笑卷之一  
 00000275 ¥S 2  
 00000280 ¥T 2 謂被謂物之由来  
 00000290 ¥N 1 ¥J  
 00000300 ¥X  
 00000310M 5 △そらことをいふ物を，などうそつきとハいひならハせ  
 00000320M 6 し。されにや，うそといふ鳥，木のそらにとまりゐて／琴（こと）  
 00000330M 7 をひく／縁（えん）によせ，そらことをうそつきといふよし。  
 00000340 ¥N 2 ¥J  
 00000350 ¥X  
 00000360M 8 △いつれもおなし事なるを，／常（つね）にたくをハ／風呂（ふろ）といひ，  
 00000370M 9 たてあけの戸なきを／柘榴（しゃくろ）／風（ふ）呂とは，なんぞいふや。かゝ  
 00000380M10みいるとの心也。（3オ）〔1〕  
 00000390 ¥N 3 ¥J  
 00000400 ¥X  
 00000410M11△かいさうの／類（たくひ）にお／期（こ）といふ／藻（も）あり。かのおごもよく／食（しょく）  
 00000420M12をす、むる／功能（こうのう）あり。さてぞ／武家（ぶけ）の／台所（たいところ）に，／飯（めし）  
 をはからひ  
 00000430M13もり，人にす、むる／役者（やくしや）をおごとはいふならし。  
 00000440 ¥N 4 ¥J  
 00000450 ¥X  
 00000460M14△よろづ物のむさき事をきたないとハいかに。北は水の  
 00000470M15方なり。水なければ万物きよからず。しかるあひた，水な  
 00000480M16いといふになそらへ，きたないといふかや。  
 00000490 ¥N 5 ¥J  
 00000500 ¥X  
 00000510M17△／宗祇（そうぎ）／宗長（そうちやう）とつれたち，／浦（うら）の夕に立出あそ（3ウ）はれし  
 00000520M18に，漁人のあみに／藻（も）を引上たり。是はなにと名をいふぞと  
 00000530M19とハれたれハ，めとも申，も共申とこたふ。時に祇公，や  
 00000540 ¥P 5  
 00000550L 1 れ，是ハよい前句やとて，  
 00000560L 2 △△めともいふなりもともいふなり  
 00000570L 3 宗長に，つけられよとありければ，  
 00000580L 4 △△／引（ひき）つれて／野（の）かひのうしの帰るさに  
 00000590L 5 ／妻（め）牛ハうんめとなき，／男（お）牛ハうんもとなくなる。／祇公（きこう）／感（かん）せ  
 •  
 •

図2 KOKIN ルールに基づくマークアップ例（断本大系）

### (3) KOKIN ルール# 2 (フラグ規則)

古典テキストはテキストの本体部分と，本体部分の周辺に配置されている傍注や割書などの付加部分から構成されている。この意味で，テキストは2次元的な構造を持ったハイパーテキスト構造を持つと言える。フラグ規則は，これら付加的テキストの領域と，それが付属する本体部分との関係を記述するものである。

フラグ規則は，傍注などの付加的部分を本体部分に埋め込んで，2次元的なテキストのレイアウト構造を1次元の文字列に変換するための規則と言うこともできる。以下に，フラグ規則の概要を示す。

```

<Original Data> ::= <Flag Begin><Data Element><Flag End><Supplement> |
                   <DataElement><SpaceFlag><Supplement><Data Element> |
                   <Data Element>

<Data Element> ::= <String>

<Flag Begin> ::= '/'

<Flag End> ::= '/'

<Space Flag> ::= '/'

<Supplement> ::= <Right Supplement> | <Left Supplement> | <Bi-Supplement>

<Right Supplement> ::= <Supplement Begin><Supplement Element><Supplement End>

<Left Supplement> ::= <Left Supplement Begin><Supplement Element>
                     <Supplement End>

<Bi-Supplement> ::= <Supplement Begin><Supplement Element>
                     ' | '<Supplement Element><Supplement End>

<Supplement Element> ::= <Single Supplement> | <Double Supplement>

<Single Supplement> ::= <Supplement Element>

<Double Supplement> ::= <Supplement Element><Supplement Separator>
                      <Supplement Element>

<Supplement Begin> ::= '('

<Left Supplement Begin> ::= "(" | "

<Supplement End> ::= ')'

<Supplement Separator> ::= '#'

<Supplement Element> ::= <String> | <String><String Separator><String>

<String Separator> ::= '='

<String> ::= see e.g. Figure 2

```

フラグ規則の基本的な記述構文は[／本体部分／(傍記)]である。[／]で囲まれた文字列領域が、注釈などが付加される本体部分であり、[()と()]で囲まれた文字列領域が注釈などの傍記(付加部分)である。図3にフラグ規則による記述例を示す。

フラグ規則は Wittgenstein Archives プロジェクトで使われている MECS という記述法、あるいは TEI (Text Encoding Interchange) における<app>エレメントと同様の機能を持っていると考えられる。

Original Text	KOKIN Markup
トゴシ 戸越	/戸越/ (トゴシ)
ノ 戸越里村	/戸越/ (ノ) 里村
サガミノカミ 相模守 和泉守トモ	/相模守/ (サガミノカミ   和泉守トモ)
泰時ノ子 タカトキ 平高時	平/高時/ (泰時ノ子#タカトキ)

図3 フラグ規則によるデータ記述例

#### (4) KOKIN ルール # 3 (付加価値づけ規則)

前述のように、研究用電子化テキストの用意、つまり分かち書きを行い品詞情報やヨミなどの属性情報を付加するなどの作業は、研究者自身が行うことになっ

```
<String>           ::= words | <Value Added Begin> words <Value Added End><Value Added>
<Value Added>     ::= <Value Begin><Values><Value End>
<Values>          ::= <Value 1> | <Value 2> | <Supplement Value> | <Value 1>
                      <Binding Symbol><Value 2>
<Value 1>          ::= Pronunciation of Sino-Japanese Ideographs <Attribution 2 Begin>
                      Chinese Ideograph<Attribution End> | <Repeating Symbol><Value 1>
<Value 2>          ::= <Attribution 1 Begin><Variation><Attribution End><Attribution 2 Begin>
                      Information<Attribution End> | <Repeating Symbol><Element 2>
<Value Supplement> ::= Not Use
<Variation>        ::= Part of Speech | Name | Location | Position
<Value Added Begin> ::= ''
<Value Added End>  ::= ''
<Value Begin>       ::= '('
<Value End>         ::= ')'
<Attribution 1 Begin> ::= '['
<Attribution 2 Begin> ::= "["
<Attribution End>   ::= "]"
<Binding Symbol>   ::= '!'
<Repeating Symbol> ::= ';'
```

付加価値付け規則の基本的な記述構文は、[△文字列△ (属性情報)] である ([△] は空白を表す)。空白で囲まれた文字列領域が、付加価値情報(属性情報)が付けられる対象領域であり、付加価値情報(属性情報)を [() と []] 内に記述する。単語単位の確定や属性情報の種類などは研究者の目的などによって異なるため、全ての記述法をあらかじめ定義しておくことは不可能である。その意味で、付加価値付け規則は未完成である。

## 2.2 KOKIN ルールの評価と SGML 化の要件

### 2.2.1 検証

KOKIN ルールの有効性を検証するために、多くの日本古典文学作品テキストの電子化を試みた。これまでに、岩波書店刊行旧版「日本古典文学大系」、東京堂出版刊行「漸本大系」、同「假名草子集成」など、約150巻、約5,000万文字の電子化を行っている。そこから、KOKIN ルールは日本古典文学作品の全ジャンルに渡るテキストの電子化に対応可能であった。さらに、これら古典テキストを電子化する上で、有効適

ている。付加価値付け規則は、文字列を任意のサイズに分解し、そこに適切な属性情報を付加するための仕組みである。以下に、付加価値付け規則の概要を示す。

切なデータ記述文法機能を有しており、かつ必要最小限の記述能力を満たしていることが実証され、高い評価を得た。

引き続き、KOKIN ルールによる電子化テキストの有用性を検証するために、CD-ROM 化、SGML 化などを進め、いずれも成功し、有用性は実証された。例えば、KOKIN ルール化データをデータベースとして管理する方策として、通常の関係データベースを利用した 3 種類の全文データベースを作成し、データベースの形成、管理、利用についても、実績を得た。上述のように、前期並びに後期の日本古典文学本文データベースの試験公開は、関係データベースモデルによる全文データベースとして、既に WWW 上で公開され、多くの利用者により日々の研究活動に活用されている (<http://www.nijl.ac.jp/>)。本データベースでは、KOKIN ルールの階層性や要素の繰り返し出現を関係データベースモデルに適合させるため、KOKIN ルールによるデータ構造の正規化を行って対処している。

### 2.2.2 課題の検討

上記の検証により、KOKIN ルールの有用性は実証

できたが、いくつかの問題点も存在する。まず、KOKIN ルールは独自に開発されたデータ記述文法（あるいは言語として）であるので、あらゆるデータ処理のためのツールを作成しなければならない。例えば、基本的な KOKIN データの記号列の整合性を検証するための語彙解析プログラムなども自作せざるを得ない。とくに、あらゆるケースを想定した構文解析のような複雑な検証プログラムは、ほとんど開発不可能である。これなどは、KOKIN データを SGML 化する作業に着手するまでは準備できなかった例である。

SGML 化データを作る際に、KOKIN ルールからの自動変換を試みた。この場合に不可欠なツールはこの構文解析ツールである。これにより、問題点を一層明確化でき、KOKIN ルールの適応範囲を明確化することができた。KOKIN ルールは、日本古典文学作品のほぼ全ジャンルに渡るテキストのデータ記述文法である。しかしながら、これは翻刻された活字本すなわち活字印刷体の校訂本を原本としている。ここに盛られた言わば代表的な古典作品構造を検討した結果に基づいて設計されている。とりわけ、個々の作品独自な構文スタイルあるいは文書構造は、個々のルールとして個別に記述することを原則とした総合的な文法として定義、構築されている。個々のルールはローカルルールと呼ぶ。

したがって、実際にデータ記述を行ってみると、予想されていたことであるが、個々の作品ごとに独自規則が必要となった。ローカルルールが増えた。しかしながら、これらは KOKIN ルール #1（タグ規則、構造記述）に起因するものではなく、KOKIN ルール #2（フラグ規則、傍記記述）および #3（付加価値付け規則、語など確定記述）におけるものである。この結果、作品対応の KOKIN ルールの変更、拡張が施された。

なお、この事情は日本古典作品のデジタル化にとって、普遍的な課題であって、KOKIN ルールで記述できないという問題ではなく、同様に SGMLにおいても記述は困難で、根本的な課題と認識される。データ記述において、かなり工夫を要すと考えられる。

このため、前述の KOKIN データから関係データへの変換では、変換をタグ規則レベルに限定し、フラグ規則と付加価値付け規則に関わる文字列、記号列は、通常の文字列として扱うこととしている（つまり、データとして扱う）。これらをデータベース構造に追い込んで定義すると、かえって専門の利用に融通性が無くなるため、使う者の自由性に委ね、すなわち使う責任として扱うこととしている。

## 2.3 SGML によるデータ記述

### 2.3.1 要件の整理

前述のように、日本古典文学本文データベースにおけるデータは、独自に開発したマークアップ規則（KOKIN ルール）に基づいて作成されている。また、構文記述は主としてテキストデータにタグを付して（タグ規則）構造化したものである（明確な DTD（Data Type Definition）は定義しないが、タグリストで記述文書の冒頭に定義）。そこで、データベースにおける多様なデータ検索も、タグを目印とした単なる文字列探索と見なすことができれば、文字列検索システムに基づいたデータベースシステムの開発が可能である。この考え方は、小規模でありながら多彩なデータサービスを行おうとしている組織（国文学研究資料館）にとって、かなり有効なものと考えられる。

一方、前述のように、KOKIN ルールは昨今の国際的な標準規約とは独立したものであり、システム的なサポートあるいは記述の標準化などの問題を抱えている。1990年半ばくらいから、テキストの電子化やシステム間の電子テキストデータ交換の手段として、SGML が出版業界などにおいて普及し始めた。このような状況を考慮し、我々の日本古典文学作品に対しても、SGML の採用の可否の実証的研究を開始した。

さらに、我々が開発研究し、蓄積をしてきている多種多様で大量のデータベースの必然的な統合あるいは共有化が不可欠であるとの状況もあり、これら全般的なデータを SGML にて記述する方策と利害得失の研究を開始した。この方策の一環として、KOKIN ルールに基づいて形成されてテキストデータを SGML に基づいた形式に変換し、電子テキストの効率的な管理と利用の促進を図ることが極めて重要と認識された。以下、KOKIN テキストデータを SGML テキストデータに変換する手法について述べる。文献目録データなどの SGML 化も同様の手法で実現している。概要を以下に示す。

- ①SGML DTD の作成（定義）
- ②KOKIN データの変換
- ③文字列検索システムを基盤としたデータベースシステムの開発
- ④SGML データを LaTeX 変換して冊子を作成する

### 2.3.2 DTD の定義

DTD（Data Type Definition：データ型定義）の骨格は、図 1、図 2 と同じである。本来、独立の KOKIN

ルール # 2 と # 3 は、併用する場合には曖昧な部分が残る。あるいは、# 3 は現在のところ実際に活用した例が無く、実用レベルのルールとは言えない。つまり、# 3 において曖昧性は残したママとなっている。例えば、記号 [ ] はフラグ規則においては <Supplement Begin> の意味で使っているが、同じ記号 [ ] は付加値付け規則では <Value Begin> の意味で使っている。<Supplement Begin> という意味の記号 [ ] は <Flag End> を表す [/] の後に現れるのに対して、<Value Begin> を表す [ ] は <Value Added End> を表す [ ] の後に現れるので、併存の場合にも区別することはできる。このままの併存の場合には、KOKIN ルールは文脈依存文法と言えなくもない。しかし、SGML は基本的には文脈自由文法のクラスなので、このような曖昧さは排除しなければならない。

そこで、KOKIN ルールの構造を、ER モデル (Entity Relation) などを用いて解析し、曖昧性の除去を行った。斬本大系の DTD はこの結果であり、DTD の整合性はパーサにより確認している。

一般に、日本語表記には表音文字である「カナ」と「かな」、表意文字である「漢字」、および「アルファベット」を含むいくつかの記号が用いられる。文字数は非常に多いため、符号化には 2 バイト以上が必要である。SGML 宣言中の SYNTAX 定義を修正する必要がある。付図 1 に SGML 宣言、付図 2 に DTD の例(斬本大系) を示す。

### 2.3.3 データ定義とデータ変換

#### (1) データ変換

KOKIN テキストデータから SGML テキストデータへの変換は、語彙解析と構文解析の 2 つのプロセスか

ら構成する。語彙解析プロセスでは、KOKIN ルールにおける [¥] に続くタグ文字列を、SGML における START-TAG(<)、GI(General Identifier) および END-TAG (>) に置き換える。このプロセスには、フラグ規則で指示されたテキスト領域を指示するための SGML 開始タグの生成も含まれる。この変換により、多くの省略タグ (Omit- Tag) を含んだ暫定的な SGML テキストデータが生成される。次に、構文解析プロセスでは生成されたデータの整合性を検証し、最後に DTD を参照しながら正規 (Canonical) SGML テキストデータに変換する。

なお、当面実用レベルにはないが、フラグ規則と付加値付け規則に存在している文脈依存的な部分を処理するための付加的なプログラムも用意した。このプログラムでは曖昧さの原因となっているタグ文字列を、一時的に別の記号列に置き換えることにより、タグの多義性を解消する。これらの変換過程は、KOKIN データの検証にもそのまま利用することができる。

斬本大系では、テキスト中に多くの注釈が存在している。研究者によっては、これらの注釈も研究上の重要なデータなので、電子化テキストにおいても、これらを保存し、必要に応じて画面上に再現したいという要求がある。フラグ規則は、この目的のために作られた規則である。KOKIN ルールの重要な機能であり、SGML 化においても興味深い部分である。以下では、SGML による典型的な注釈のマークアップ例を示す。

#### (2) 基本的な傍記の定義

基本的な傍記の例を図 4 に示す。これは本文中の漢字のヨミを表している。図の上段が SGML テキストデータであり、下段はその印字例である。

```
<小作品名>醒睡笑卷之一</小作品名><嘶><嘶名>謂被謂物之由来</嘶名>
<小嘶><小嘶番号 num= "1 "><本文レコード>
<行番号 pos="M" num= "5 ">△そらことをいふ物を、などうそつきとハいひならハせ
<行番号 pos="M" num= "6 ">し。さればにや、うそといふ鳥木のそらにとまりゐて
  <傍記素 fg="OFF">琴</傍記素><傍記>こと</傍記>
<行番号 pos="M" num= "7 ">をひく
  <傍記素 fg="OFF">縁</傍記素><傍記>ゑん</傍記>によせ、そらことをうそつきといふよし。
</本文レコード></小嘶>
```

図 4 - 1 基本的な傍記の SGML マークアップ例

醒睡笑卷之一  
 謂被謂物之由來  
 そらことをいふ物を、などうそつきとハいひならハセ  
 こと  
 し。されはにや、うそといふ鳥、木のそらにとまりゐて琴  
 あん  
 をひく縁によせ、そらことをうそつきといふよし。

図 4 - 2 基本的な傍記の SGML マークアップ例

SGML テキストデータでは、注釈を 2 つの方法で記述する。1 つは注釈の対象となる本文中の領域を示すもので、`<SuppElement>` と `</SuppElement>` で囲まれた部分である。これはフラグ規則の `<Flag Begin>` と `<Flag End>` に対応するものである。もう 1 つの部分は注釈そのものの領域を示すためのもので、`<Supp>` と `</Supp>` で囲まれた部分である。これはフラグ規則の `<Supplement>` に対応する。`<SuppElement>` には属性 “fg”

が定義されている。これは、注釈が複数行に跨っているか否かを示すフラグである。この例では注釈が複数行に跨っていないので、`fg = "OFF"` である。

### (3) 泣き別れ傍記

図 5 は、傍記が複数行に跨っている、所謂行と注釈の泣き別れの例である。そのため、タグ `<SuppElement>` の属性 “fg” は “ON” となる。

```

<行番号 pos="L" num="9">△<傍記素 fg="OFF">七歩</傍記素><傍記>しつほ</傍記>とぬるゝとハ何事そ。されハ尺迦<傍記素 fg="OFF">誕生</傍記素><傍記>たんしゃう</傍記>の時、阿<傍記素 fg="ON">難</傍記素><行番号 pos="L" num="10"><傍記素 fg="ON">陀竜</傍記素><傍記>なん</傍記><傍記>たりう</傍記>王ハ<傍記素 fg="OFF">湯</傍記素><傍記>ゆ</傍記>を<傍記素 fg="OFF">吐</傍記素><傍記>はき</傍記>、<傍記素 fg="OFF">難陀竜</傍記素><傍記>なんたりう</傍記>王ハ水を吐、此うぶ湯にぬれなが
  
```

しつほ	たんしゃう	なん
七歩とぬるゝとハ何事そ。されハ尺迦誕生の時、阿難		
たりう	ゆ	なんたりう
陀竜王ハ湯を吐、難陀竜王ハ水を吐、此うぶ湯にぬれなが		

図 5 泣き別れ傍記の SGML マークアップ例

### (4) 左右傍記

図 6 は、傍記が同一のテキストに対して、複数出現する例を示している。とくに、テキストに対して左右(縦書き)に傍記がある例である。

タグ `<SuppElement>` に続くタグ `<BiSupp>` が複数注釈領域の開始を示している。さらに、タグ `<RightSupp>`

は、注釈が本文の右側(図 6 では、横書きのため上側)にあることを示し、タグ `<LeftSupp>` は注釈が本文の左側(図 6 では下側)にあることを示している。

なお、付図 3 に、SGML による全文テキストの記述例を示す。これは図 2 に示した KOKIN テキストと同じ内容である。

```

<行番号 pos="L" num="6"><傍記素 fg="OFF">窮鼠</傍記素><傍記>きうそ</傍記>
<傍記素 fg="OFF">還</傍記素><傍記>かへつて</傍記>噛猫とあれは、にくむ処
<傍記素 fg="OFF">狂惑</傍記素>
<左右傍記><右傍記>きやうあく</右傍記><左傍記>まげまとふ</左傍記>
</左右傍記>なる<傍記素 fg="OFF">鼠</傍記素><傍記>ねつみ</傍記>
<傍記素 fg="OFF">根性</傍記素><傍記>こんじやう</傍記>なるか<行番号 pos="L" num="7">な。

```

きうそかへつて	きやうあく	ねつみこんじやう
窮鼠還	噛猫とあれは、	にくむ処狂惑なる鼠根性なるか
	まげまとふ	
		な。

図 6 左右傍記の SGML マークアップ例

## 2.4 XML によるデータ記述

### 2.4.1 基本的考え方

データの SGML 化により、データ作成からデータベースサービスあるいは電子出版への新しい手法が実現した。しかしながら、SGML には以下のようなデータ処理上の問題がある。

- (1) SGML の仕様は非常に複雑であり、その仕様を完全に満たすツールは存在しない。例えば、前述のテキストスタイル意味指定言語 DSSSL (Document Style Semantics and Specification Language) を完全に実現している製品ツールは皆無である。
- (2) SGML の仕様では多くの記述上の省略が許されているが（例えば、構造が不明確にならければ終了タグは省略できる），このためにデータ処理の手順が複雑になる。
- (3) SGML はハイパーリンクの機構を独自に定義していないため、WWW との親和性に弱い。

### 2.4.2 SGML から XML への変換

そこで、今までに作成されたあるいは変換された SGML データを、XML 互換の形式に変換することとした。図 7 に、データベース構築プロセスの全体的なフローの概要を示す。後述するが、図 7 はメタデータ・データベースシステムの構成を意図した概念を表している。

前述のように、テキストデータの SGML 化は、KOKIN ルールに従って電子化されているテキストの変換を主としたものであった。その意味では、KOKIN ルールを単に精密化したものと言える。また、この時点においては日本古典文学作品のための汎用 DTD は、未だ明確に定義していなかった。一方、後述のコラボレーション環境の構築を目標とする際に不可欠な要因は、日本古典文学作品のデータ構造の記述、すなわち DTD の定義である。

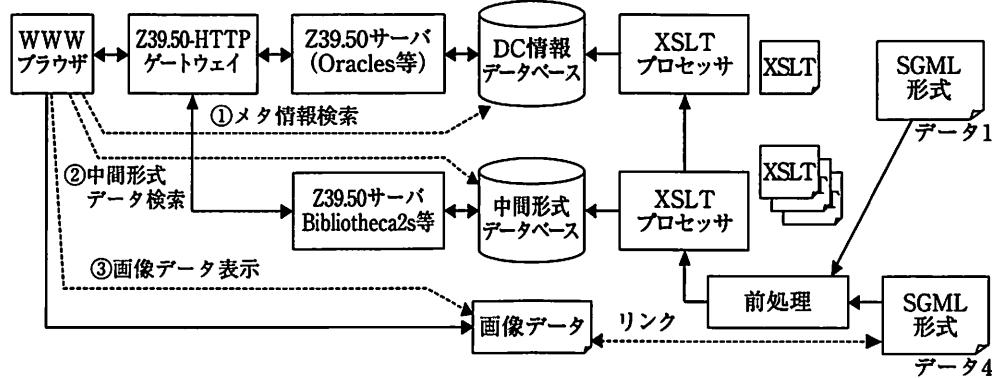


図7 SGMLからXMLへデータ変換フロー  
(コラボレーション環境データの形成)

そこで、以上の背景に基づき、XML化における実証実験においては、明確にDTDを定義することとし、その完成をみることができた。したがって、定義されたDTDは、今後のコラボレーションにおいて重要な要素になると考えられる。

ところで、海外に目を転じると、TEI(Text Encoding Interchange)と呼ばれる人文科学向けの汎用DTDが既に提案されている。これは、ACH(Association for Computer and Humanities), ACL(Association for Computer Linguistics), 並びにALLC(Association for Literary and Linguistic Computing)による国際共同作業として、1987年に開始し、1994年には最終報告書をとりまとめている。ここでは、文字、原稿、口語、音韻などの符号化あるいはハイパーテキストの扱いなど、人文科学領域のテキストのデジタル化において重要な多くの事項が検討されている。

KOKINルールは、TEIの活動とは無関係に開発されたものであるが、基本的な考え方は類似している。さらに、2.3.3で例示した様々な傍記(注釈)であるが、最近JIS4052あるいはほゞ同じ勧告がRuby Annotationとして、W3Cから発表されている。これによれば、ヨミのマークアップは以下のように表現される。

将来のデータ共有を考えると、これらの勧告などで示される論理構造と、個別に定義された論理構造の相互変換は重要な要件と考えられる。あるいは勧告の準拠は必要である。そこで、テキストデータについて国際的な規約との整合を試みている。この変換課程においては、XSLT(Extended Stylesheet Language Transformation, 拡張スタイルシート言語変換)を単純に適用できないので、補助的なプログラムや一部には手作業も必要とした。

例： 国文学研究資料館

```
<ruby>
<rb>国文学研究資料館</rb>
<rt>こくぶんがくけんきゅうしりょうかん</rt>
</ruby>
```

付図4に、日本古典文学作品テキストのXML DTD例を示す。また、付図5にXMLマークアップ例を示す。さらに、付図6にXMLテキストをHTMLに変

換するXSL(Extensible Stylesheet Language)定義の簡易版の例を示す。

### 3. コラボレーション

#### 3.1 コラボレーション研究の経緯

##### 3.1.1 研究の背景と課題

国文学研究資料館の情報化計画の1つに電子資料館システムの構築がある。日本文学研究に不可欠なあらゆる研究資料情報をデジタル化し、トータルな情報システムにより、利活用することが目的のシステムである。すでに、画像データ、動画データ、全文データ、目録データなどの多種多様で膨大なデータが形成されつつあり、これらの一一部はインターネット上で閲覧可能となっている。

しかしながら、これらのデータベースは極めて有用であるが、問題がない訳ではない。個々に、専門的目的を持って構築され、独自なものとして、内容や使い勝手が開発してきたものである。さらに、長い歴史的蓄積と経緯も持っている。データベースの目的の違い、あるいはメディアの種類による違い、開発時期の違いなどによる差異は大きい。すなわち、独立した個別のデータベースシステムと言っても過言ではない。

問題は、以下のような使い勝手として顕在化していく。

- (1) データベースごとに検索法が異なる。
- (2) 類似の資料が別々のデータベースに収容されているため、データベースの全体像や概要を把握していないと検索が困難である。
- (3) 関連した資料や研究成果を調べることが困難である。

この問題を解決する方策は、以下の2通りのアプローチしかないと考えられる。もちろん、よりキメの細かいガイドブックやヘルプシステムを作り、ユーザフレンドリな環境を提供することは相応に試みられている。

- (1) すべてのデータベースを一元化し、同一のDBMS、検索システムで運用管理する。
- (2) 別途、共通する（あるいは共有する）検索システムを構築し、検索機能を一元化する。

(1) の方法は最も有効と考えられるが、データベースの独自性や実現技術などからみて非現実的である。とくに、現在実用レベルにあって、かつ実績のある膨大かつ複雑な個々のデータベースを、共通のデータ構

造に変換することはほとんど不可能である。

したがって、(2) の方法しかないと考えられる。個々のデータベースの独自性を維持し、共通項による共通的な検索を実現することである。この場合、個々の持つ全機能並びに全データ項目の利用または検索はある程度捨象する。表層的利用または検索とする。ただし、その表層の規模の選定は重要である。

(2) の方法の研究に数年をかけ、実証実験を踏まえ結論に達している。すなわち、共通項にダブリン・コアによるメタデータを採用し、共通検索にZ39.50プロトコルの適用を図るものである。この方式は、我々のような人文科学系の複雑多岐にわたる複合データベースの管理運用については、国内外においても初めてのケースであり、成果が注目されている。

##### 3.1.2 コラボレーションシステムの意義

日本文学研究に関わるコラボレーションシステム（国文学研究資料館電子資料館システムに基づく）の開発研究を進めている。このコラボレーションシステムでは、国文学研究資料館の所蔵原本の目録と画像、主要作品の語彙や全文、研究論文の目録、歴史史料の所在と目録、OPAC（Online Public Access Catalog）などの個別データベースを基本とする。これらをダブリンコア・メタデータとZ39.50を利用して統合する計画である。

このシステムでは、以下のような典型的な利用を可能とする。例えば、国文学研究資料館の史料所在データベースから「伊能家」を検索すると、同マイクロ資料目録データベースから伊能忠敬の「日本経緯度実測」の所在書誌情報が、同画像データベースからその画像情報などが得られる。すなわち、関連するあらゆる情報にアクセスすることができる。しかも、操作は簡単かつ单一操作であり、加えて高精度で検索することができる。

上例は、国文学研究資料館単独のデータベース群の共通検索に関する単一利用であり、この方式の実現は大きな利便性を利用者に提供することは言うまでもない。一方、もしこの方式が関連するあらゆる研究機関に適用できるならば、その利便性は計り知れない。利用者が利用したいと思う現在活用できるデータベースは、全国の関連する研究機関などに分散されている。複数の図書館、博物館、文書館などに分散している資料を、一元的に検索することができれば、まさに学術研究の革命的とも言える環境が生まれる。したがって、ネットワーク上にデータを公開している研究機関

などが、我々と同様のコラボレーションシステムを導入することが可能ならば、研究機関を越えた情報検索が可能となる。

もちろん、最近学術情報コンテンツは増大化の一方であるが、利用に足るあるいは欲しいと思う学術情報の組織化はまだまだ不足であり、諸外国に比して遅れている。あらゆる分野での学術情報の組織化は一層進める必要がある。

### 3.2 コラボレーションシステム開発研究

#### 3.2.1 モデルの設定

コラボレーションシステム研究の目的は、日本文学研究を支援するための方法論（モデル）の構築である。本研究の目標は、関連する多岐に渡る研究材料や資源をデジタル化情報資源として、できるだけ簡便で、かつ同一のインターフェースで、精度よく入手し、研究を進めることである。そのためのシステム資源と環境の提供にある。

例えば、上記の伊能忠敬の例と同様に、「源氏物語」の研究者に対して、諸本（写本、版本）の画像やテキストは言うに及ばず、物語の舞台となった地域の地勢や地図的情報、その時代の歴史的事実と補間資料、人口や経済などの統計情報、民俗や風俗情報、あるいは同時代の周辺地域の政治情勢等々、関連するあらゆる情報を提供することである。しかも、単一のインターフェースで精度良くアクセスできることが最大の眼目である。

コラボレーションシステムを具体的に構築するに当たり、国文学研究資料館が有する日本古典文学および歴史学に関わる資料を主要材料とした。2つのモデルを考察し、実証実験を行った。すなわち、

- (1) 複数の研究機関あるいは個人が作成している多様な人文科学系データを、データベースとして組織化するモデル、および
- (2) 複数のデータベースをあたかも単一のデータベースであるかのように、再組織化するためのモデル

である。

#### 3.2.2 モデル設計

コラボレーション並びにモデルは、以下のように考える。

#### (1) コラボレーション

文学研究には原本、テキスト、目録などの多様な研究資料が必要であるが、1研究機関で全てを網羅することは不可能である。そこで複数の研究機関の情報システムをネットワークで結び、あたかも単一のシステムであるかのように協調してサービスを行うとよい。このようなシステム形態は電子的協調作業方式（コラボレーション）と呼ばれている。

したがって、コラボレーションに必要な諸機能（作成、蓄積、評価など）のモデル化を研究する必要がある。

#### (2) データモデル

従来の「標準化」の概念は、データ要素や構造を一元化する手続きであるが、多様な学術的背景を持つデータの不用意な一元化は無意味であり、むしろ危険と考えられる。しかしながら、国内外にあまねく分散している学術情報を有機的に相補的に利用しようとすれば、何らかの標準化は必要である。

本研究では、この相反するような命題をメタデータの導入と、その検索システムの標準化により、達成しようとするものである。メタデータはデータのデータで、つまりデータの内容を記述したデータである。メタデータの方式は2つのモデルを考える。検索用の汎用メタデータ・モデルと、研究領域ごとの専用メタデータ・モデルである。汎用メタデータ・モデルでは、インターネット検索用に開発された Dublin Core Metadata Element Set (DC) を基礎に、その拡張を図るものとする。

基本的なシステム概念設計として、既存のデータベースと利用者間に汎用メタデータベースを置く。利用者は既存のデータベースを直接検索するのではなく、汎用メタデータベースを検索する。ここで、汎用メタデータベースは一種の翻訳データベースと考える。利用者から個別データベースへ、また逆に個別データベースから利用者への翻訳である。その意味では、データベース設計における内部スキーマと外部スキーマのインターフェースである概念スキーマに相当する。汎用メタデータベース中のメタデータは、既存個別のデータベース中のキーワードなどに翻訳される。

そこで、汎用メタデータさえ標準化しておけば、異なるデータベースを同じ構造のデータベースと見なし、検索することができると考えられる。さらに、汎用メタデータベースをデータアクセスの仲介とすることにより、個別のデータベースの修正を最小限に留めることができる。

一方、専用メタデータ・モデルは、研究領域ごとの個別データを汎用メタデータに変換する際のガイドとして利用する。個別データの内容を汎用メタデータへ直接変換することも技術的には可能であるが、領域特有のデータを汎用データ構造へ展開することは実用的ではない。例えば、研究機関や研究者によるデータ解釈の差が大きいので、データ意味や品質にバラツキが生ずる。そこで、専用モデルを仲介させることで問題の軽減を図る。また、この専用モデルは新たにデータを作成しメタデータベースへ登録する際のガイドとしての役割も担う。

専用メタデータ・モデルの開発では、人文科学フルテキスト記述用の（TEI: Text Encoding Initiative）、史料記述用のEAD (Encoded Archival Definition)とISAD-G (International Standard of Archival Description-General)を基礎に置き、拡張を図ることとした。また、検索プロトコルの標準化は、全データベースを同じ方法で検索する手段を与える。具体的には、検索問い合わせの標準プロトコルであるZ39.50を考慮する。

### 3.2.3 研究経緯

我々（国文学研究資料館）が、公開あるいは公開準備中の各種目録データ、全文データ、原本画像および動画データ、歴史史料データなどを、コラボレーション研究の基本資源として、研究を進める。以下、まずこれらのデータベースの背景やデータ構造について、概観する。

#### （1）目録データベース

現在の出版物と異なって、日本古典籍の書誌項目、

構造は標準化されてはいない。例えば、書名は資料のいたる所に現れ、しかも記載が異なっていることも多い。また、古典籍の多くは個人や文庫のコレクションとして所蔵されている。これらの所蔵変遷は頻繁で、所在情報の把握は重要である。このような複雑な書誌構造を記述できる標準は存在しない。そのため、古典籍の目録データベースのレコード形式はLC-MARC (Library of Congress MACHine Readable Cataloguing) やJP-MARC (JaPan MACHine Readable Cataloguing)などの標準的な目録のレコード形式には準拠せず、独自のものである。

#### （2）原本画像データベース

文献資料の目録データベースでは、資料所在が判明しても、直接資料自体にアクセスすることはできない。とりわけ、地方や海外の研究者にとっては切実な問題となっている。そこで、文献資料そのものをデジタル化して、いわゆる画像データベースとして提供することができればよいと考えられる。デジタル画像は高精細でなければ価値がないという意見もあるが、国内外の多くの利用を望む研究者からはファクシミリ程度の画像でもよいという要望も強い。

原本画像データベースは以下のよう仕様で構築されている。現時点では、約700,000コマ（CD-ROMで約1,200枚）のデジタル化が終了している。これは、国文学研究資料館所蔵原本の約70%に相当する。画像データは、白黒2値、解像度をA3版用紙換算600DPIでデジタル化を行い、G4圧縮を行った上でTIFF形式によりCD-ROMへ蓄積している。TIFF情報の概要を表2に示す。

表2 画像データベースのTIFFの構造

Tag Name	Tag(hex)Type	Value
BitsPerSample	102	SHORT 1
Compression	103	SHORT 4 (G4 Fax)
Photometric	106	SHORT *1
DocumentName	10d	ASCII 函架番号*2
Make	10f	ASCII Name of Trader
SamplesPerPixel	115	SHORT 1
PageName	11d	ASCII Serial Page No. within a work*3
ResolutionUnit	128	SHORT 2(inch)
DateTime	132	ASCII "YYYY: MM"

〈注〉 \*1 ピュア上で文字部分が黒となるように設定すること。

\*2 EUC全角文字とする。

\*3 フラッシュ番号内での連続番号である。ファイル名から拡張子“.tif”を除いた文字列である（ターゲットを除く）。

画像データための国際的な標準レコード構造はないので、ここでは暫定的な構造を定義した。基本的な考え方は、マイクロフィルムのリール内の論理的構造と国文学研究資料館におけるマイクロフィルム管理簿の論理構造を用い、これらを SGML 化して定義する。

マイクロフィルム管理用 DTD の例を。図 8 のように定義する。そのタグの意味を表 3 に示す。なお、画像データベースのインデックスの例を、図 9 に示す。

```

<!ELEMENT index      ..      (record+)>
<!ELEMENT record     -O      (key, title, type, volume, book+)>
<!ELEMENT book       -O      (bno, rn, fno, begin, end)>
<!ELEMENT key        -O      (#PCDATA)>
<!ELEMENT frn        -O      (#PCDATA)>
<!ELEMENT title      -O      (#PCDATA)>
<!ELEMENT type       -O      (#PCDATA)>
<!ELEMENT volume     -O      (#PCDATA)>
<!ELEMENT bno        -O      (#PCDATA)>
<!ELEMENT rn         -O      (#PCDATA)>
<!ELEMENT fno        -O      (#PCDATA)>
<!ELEMENT begin      -O      (#PCDATA)>
<!ELEMENT end        -O      (#PCDATA)>

```

図 8 マイクロフィルム管理 DTD

表 3 画像データベースインデックスのタグの意味

タグ	項目名
<index>	・インデックス
<record>	・レコード
<key>	・キー
<title>	・書名
<volume>	・総冊数
<book>	・冊番号
<bno>	・作品
<rn>	・冊番号
<type>	・刊写別
<m>	・RN ナンバー
<fno>	・フラッシュ番号（ディレクトリ名に相当）
<begin>	・作品の開始コマのファイル名（ターゲットを含む）
<end>	・作品の終了コマのファイル名（ターゲットを含む）

---

```
<index>
<record>
<key>99 - 34</key>
<title>太平記</title>
<type>刊</type>
<volume>21</volume>
<book>
<bno> 1 </bno>
<rn>110 - 1 </rn>
<fno> 1 </fno>
<begin> 6 </begin>
<end>502</end>
</book>
. . . .
<book>
<bno>21</bno>
<rn>110 - 3 </rn>
<fno> 1 </fno>
<begin> 5 </begin>
<end>161</end>
</book>
</record>
<record>
<key>99 - 35</key>
<frn>110 - 3 - 2 </frn>
<title>休間抄</title>
<type>写</type>
<volume>54</volume>
<book>
<bno> 1 </bno>
<rn>110 - 3 </rn>
<fno> 2 </fno>
<begin>170</begin>
<end>200</end>
</book>
. . . .
<book>
<bno>54</bno>
<rn>110 - 8 </rn>
<fno> 1 </fno>
<begin> 5 </begin>
. . . .
</index>
```

---

図9 画像データベースインデックスの例

付図7に、画像データベースの検索例を示すように、画像データベースは目録データベースと連携している。利用者は、まず目録データベースを検索し資料存在を確認し、ついでデータベース間のリンクを辿って画像データベースへアクセスする。ここでは、データベース間のリンクにマイクロフィルムの請求番号を用いている。

この画像データベースでは、目録から画像という一方のリンクではなく、画像から目録へのリンクも可能としている点に特徴がある。具体的には、TIFFデータ構造仕様におけるタグ0x10d (Document Name) の内容を、請求番号で置き換えている。画像ビュアがこの情報を処理できれば、最初に画像データベースを眺めていて、興味のある画像を見つけたときにリンクを辿って目録情報を参照することが可能である。

以上のように、画像データベースはマイクロフィルムから作成している。したがって、従来の国文学研究資料館における各種データベースとの親和性を保存している。なお、上記画像データベースにおける原資料は館蔵であるため知的財産権などの問題はないことを付言しておく。

### (3) 演能演劇動画データベース

能などのいわゆる演芸も文学の研究対象である。演能演劇動画データベースは動画データベースの実験システムであり、DVDに蓄積された動画データをビデオ・オン・デマンド方式で配信するシステムである。

### (4) 全文データベース

全文データベースが試験公開されている。KOKINルールに基づき記述されたデータをデータベース化している。DBMSには関係データベースを用いている。現在のところ、全文データベース向きに開発されたDBMSはない。したがって、現在利用可能なDBMSを用いて、何らかの工夫をした上で全文データベースを管理運用するしかない。しかしながら、関係モデルはフラットな多くのデータを扱うには最適のシステムであるが、テキストが有する階層性、連続性などの複雑な構造を直接扱うことができない。また、SQL (Structured Query Language) やQBE (Query By Example)などの標準的な問い合わせ機能があるが、全文処理には充分と言えない。

テキスト処理向きのSGMLと同様な問い合わせ言語としてDQL (Document Query Language)の提案を行っている。DQLの記述モデルにはSQLのそれをそのまま利用したが、評価部分にはテキストの階層構造

や反復構造を処理できるような拡張を施している。DQLの問い合わせ記述能力は強力であったが、SQLの記述モデルを踏襲しているため、問い合わせ式が非常に複雑になってしまい、実用化には至っていない。

一方、最近では高速文字列検索システムやプログラムが利用できるようになり、日本語SGMLデータを扱える製品も市販されてきた。全文処理を文字列、用語や用例の検索が主となることを考慮すれば、高速文字列検索の方法は試みる価値がある。もちろん、いわゆるキーワード検索のように、概念を含む事項の情報検索には向きであるので、工夫する必要がある。

そこで、付図8に例示するように、文字列検索プログラム(OPEN TEXT)を利用した全文データベースシステムの開発を行っている。目録あるいはフルテキストは、文字型の不定長フィールドが一定の構造を持ったデータとみなすことができる。また、SGMLはテキストの構造を記述する能力を持つ国際標準規約であるから、検索を「テキストデータ中の文字列検索」と見なすことによって、文字列検索システムを利用したデータベースシステムと考えることができる。一般に、人文科学における情報検索では文字列に注目して検索することが多い。したがって、文字列検索システムを基本としたシステムが有効と考えられる。

最近、高速の文字列検索システムが開発され、市販されるようになった。それにより、高速の文字列検索ソフトウェアを用いたデータベースシステムを構築することが容易となってきた。以下では、OpenText社製の文字列検索エンジン(PAT)を用いて、SGMLによる斎本大系フルテキストデータベースを検索する参考例を示す。

以下の例は、“笑”という単一文字の含まれているテキストを検索する例である。なお、ゴシック体は入力文字列、“←”に続く文は説明、それ以外はコンピュータ出力である。

### 【例 1】

```
>> "笑"           ← "笑" という文字で検索を行う。
    2: 41 matches  ← 41ヶ所に"笑"という文字が見つかった。
>> pr            ← プリントアウトせよ。
</作品名><挿絵></挿絵><作品書誌><本文レコード>...
2026, ... 記素 fg="ON">笑</傍記素><傍記>せいすい</傍記><傍記>せう
</傍記> ...
513530, ... 記素 fg="OFF">笑</傍記素><傍記>わら</傍記>ふ。法師, なに事ぞ, 今...
<行番号 ... 記素 fg="OFF">笑</傍記素><傍記>わらい</傍記>しを, 大夫
<...5387, ...記素 fg="OFF">笑</傍記素><傍記>わらい</傍記>になせり。
954839, ...記素 fg="OFF">笑</傍記素><傍記>わらひ</傍記>て, われ
<傍記素 fg="...
.....<br/>
<行番号 pos="M" num= "...人笑けるを, 勝たる侍, わたさ
.....<br/>
</嘶名><小嘶>....<嘶名>頓作 笑卷之八
.....<br/>
759520, ...記素 fg="OFF">笑止</傍記素><傍記>しゃうし</傍記>におもひ,
<行番号 pos="M.. fg="OFF">笑止</傍記素><傍記>せうし</傍記>が
<行番号 po..記素 fg="OFF">笑止</傍記素><傍記>せうし</傍記>なよ,
```

例 1 では "笑" の出現する場所を指定していないので、正規表現レベルの検索（UNIX 系で使われる grep コマンドに相当）と同じである。嘶本大系は SGML

マークアップが施されているので、エレメントを指定した文字列検索が可能である。そこで、<傍記素>の領域にある "笑" を検索してみる。

### 【例 2】

```
>> region "傍記素" including "笑"  ← 傍記素の領域内にある"笑"という文字を検索する。
    7: 15 matches  ← 15箇所に"笑"という文字が見つかった。
>> pr            ← プリントアウトする。
2010, ... "M" num= "7"><傍記素 fg= "ON">笑</傍記素><傍記>せいすい</傍記><...
86003, ...する。知音の者<傍記素 fg= "OFF">笑止</傍記素><傍記>せうし</傍記>...
93796, ...num= "9">に,<傍記素 fg= "OFF">笑止</傍記素><傍記>せうし</傍記>...
.....<br/>
954822, ..>そわう</傍記><傍記素 fg= "OFF">笑</傍記素><傍記>わらひ</傍記>て ...
967351, ..">子つら+見て<傍記素 fg= "OFF">笑</傍記素><傍記>わらひ</傍記>て ..
985505, ..ば, 二人ともに<傍記素 fg= "OFF">笑</傍記素><傍記>わらひ</傍記>て ..
```

例 1 では、テキスト全体で41ヶ所に"笑"があったが、それらは<傍記素>、<嘶名>、<本文レコード>など、様々なエレメントの中に現れている。例 2 では、エレメントを<傍記素>に限定すると15ヶ所（ただし、同一の傍記素内に 2 回以上現れる場合も含む）であった。

このような検索は正規文法レベルの検索では不可能である。次に、"笑" という文字が含まれている<嘶>の<嘶名>を調べてみることにする。

### 【例 3】

```
>> region "傍記素" including "笑" ← 傍記素の領域内にある"笑"という文字を検索する。  
16: 15 matches ← 15ヶ所に"笑"という文字が見つかった。  
>> region "嘶" including % ← その傍記素を含んでいる"嘶"を検索する。  
17: 11 matches ← 11ヶ所の"嘶"が見つかった。  
>> region "嘶名" within % ← その嘶の領域内にある"嘶名"を検索する。  
18: 10 matches ← "嘶名"は10あった。  
>> pr.region."嘶名" ← プリントアウトする。  
72360, ..<嘶名>鈍副子</嘶名>...  
202141, ..<嘶名>賢たて</嘶名>...  
402944, ..<嘶名>唯有</嘶名>...  
.....  
721850, ..<嘶名>謳</嘶名>...  
757238, ..<嘶名>舞</嘶名>...  
776535, ..<嘶名>頓作</嘶名>...
```

嘶本大系の DTD によれば、エレメント<嘶>、<傍記素>、および<嘶名>の階層関係は、<嘶>⇒<傍記素> および<嘶>⇒<嘶名>である。そこで、最初に"笑"を含むエレメント<傍記素>を探し、次に探し出された各<傍記素>を含む上位エレメントである<嘶>を抽出し、最後に抽出された各<嘶>配下の<嘶名>を表示

するという手順を踏んでいる。この例で、<嘶>は11ヶ所なのに<嘶名>は10ヶ所となっている。これは、全ての<嘶>に必ずしも<嘶名>が付いているわけではないからである。実際、嘶本大系の DTD においても<嘶>以下の構造は、

```
<!ELEMENT 嘶 - O (嘶名 ?, (%para3;), 小嘶+)>
```

と定義されている。これは SGML の要素に注目した検索の典型例であるが、この方法は、他の目録検索などにも応用することができる。現在のコラボレーションを前提とした電子資料館システムの検索システムは、OPEN TEXT の機能を基盤として構築されている。

電子資料館システムでは、SGML 化された目録データベース、画像データベース、動画データベース、全文データベースなどから構成されている。このうち、目録データベースと画像データベースは前述のように関連づけられており、いわゆるマルチメディアデータベースシステムとなっている。

### 3.3 コラボレーションシステム開発

#### 3.3.1 Z39.50

Z39.50は、インターネット環境下において、情報検索における質問や結果、運用管理（課金、利用者認証など）を通じて標準化するものである。国際標準プロトコルとして、情報検索システムに必要な機

能を定義している。1970年代に、米国の議会図書館と書誌ユーティリティとの間で、コンピュータに蓄積されていた目録データを直接交換しようとする計画から始まった。

Z39.50は以下のような特徴を持つ。

- (1) データベースシステムのソフトウェアおよびハードウェアから独立したクライアントサーバ方式の規約である。そのため異種システム間で透過的な検索やレコードの送信が可能である。
- (2) 単一のインターフェースで異なるデータベースを利用できる。
- (3) WWW と異なり検索状態が保存される。
- (4) 書誌情報以外の情報検索にも利用できる。

データベースシステムのハードウェアやソフトウェアの実装に依存しないスキーマを実現するため、Z39.50ではアトリビュートセット（Attribute Set）と呼ぶ論理的なスキーマを定義している。アトリビュート

セットは目的に応じて何種類か提案されているが、大部分の先行的 Z39.50 システムでは Bib-1 という単一のアトリビュートセットのみを使用し、実現されている。

Z39.50 はクライアント・サーバ方式の検索規約である。すなわち、クライアント側の検索システムとサーバ側のデータベースシステムとが、Z39.50 の規約に従って情報交換を行う。クライアントであるユーザは、使い慣れた検索環境下で複数のデータベースにアクセスすることができる。欧米では早くから Z39.50 を用

いた検索システムが普及し、とくに図書館間における OPAC の相互検索用に多く利用されている。我が国では漸く注目され始めた段階であり、システムの構築例は多くない。

表 4 に、国文学研究資料館の Z39.50 サーバの仕様概要を示す。

項目	パラメータ
ホストアドレス	最大40 桁まで登録可能
ポート番号	ポート番号は数値で登録。最大 5 桁まで登録可能
データベース名	最大40 桁まで登録可能
レコードシンタックス	GRS-1あるいは SUTRS
漢字コード	ISO2022, EUC, ShiftJIS, ISOUCS2
認証フラグ	認証フラグは数値で登録。0：認証なし 1：認証あり

表 4 Z39.50 サーバの設定パラメータ（国文学研究資料館）

### 3.3.2 Dublin Core メタデータ

ダブリンコア・メタデータは、ネットワーク上で流通している様々な情報資源を効率的に探索するための技法である。メタデータとして、必要最小限の共通要素を定義し、それによる情報検索を行う。ところで、

YAHOO などに代表されるインターネット上の検索システムでは、タイトルや作者名などのキーワード項目（データ要素）を指定した検索ができない。これは、ネットワーク上の資源を網羅的に総ナメ的に検索する上では便利であるが、一般に検索ノイズが多くなる。また、図書目録などの検索システムでは、キーワード（検索要素）を適切に選択することにより、求める資料を効率的かつ正確に探し出すことができる。

しかし、図書館、博物館、文書館などで必要とされるデータ項目は、種類、書式、あるいはその意味づけなどにおいて必ずしも同じでない。同じ用語（キーワード）で、これらすべての関連するデータベースを検索した場合、得られる結果はかなり異なるに違いない。また、そもそも検索する方法も異なるから、新たな知見を得るような関連情報を見いだすことは容易ではない。

これに対して、ダブリンコア・メタデータはデータ検索における相互利用性の要件から生まれたものである。ダブリンコア・メタデータは、情報検索で必要と考えられる最小公倍数的なデータ要素（キーワード）のみを定義しているので、多様な情報検索システムの

検索項目との対応が比較的容易という特徴を持つ。すなわち、これによれば、目録、アーカイブなど異なる目的や構造を持った情報資源を効率的に検索することが可能となる。

このためには、メタデータをどのように定義し、作るかが極めて重要な要素となる。我々の方法は以下にも示すが、関連するデータベースのあらゆる情報項目を、ダブリンコア・メタデータという標準の情報項目に翻訳することである。

表 5 に、ダブリンコア・メタデータで定義されている検索要素の概略を示す。

表5 Dublin Core Metadata Element Set

(A) 情報資源の内容に関する要素
1) Title : 対象の名前
2) Subject : 内容のトピック
3) Description : 情報資源の内容に関する記述。アブストラクトなど。
4) Source : 情報資源の出所。
5) Language : 情報資源の内容を記述している言語
6) Relation : 他の情報資源との関係
7) Coverage : 場所や時間に関する情報資源の特性
(B) 情報資源を知的財産と見なした場合の要素
8) Creator : 情報資源の内容について責任を持つもの。著作者など。
9) Publisher : 情報資源を現在の形態にしたもの。出版社、機関など。
10) Contributor : 著者ではないが情報資源の作成に関わったもの。 編集者や翻訳者など。
11) Rights : 著作権、利用条件に関する記述へのリンク
(C) 情報資源の具現化に関する要素
12) Date : 現在の形で利用可能になった日付。
13) Type : 情報資源の型。ホームページ、テキストなど。
14) Format : 情報資源のデータ形式。PostScriptなど。
15) Identifier : 情報資源を一意に識別するための名称、番号

### 3.3.3 Z39.50とダブリンコア・メタデータの融合

ダブリンコア・メタデータの役割は、データベースの種類を超えた相互利用性の実現である。対象とする多くのデータベースから、適当な検索項目を抽出してダブリンコア・メタデータにマッピングする。これにより、ダブリンコア・メタデータを検索のゲートウェイとして、全てのデータベースを統合的に検索することが可能となる。例えば、国文学研究資料館における原本目録データベース、原本画像データベース、全文データベースなどの総合的情報検索が可能となる。一種の横断検索の実現といえる。

ところで、ダブリンコアは検索要素の定義のみであり、システムの実装については言及していない。したがって、ダブリンコア・メタデータベースシステムといつても、研究機関を超えた情報検索の場合には課題が残る。ある機関では、XML/SGMLのタグを利用した文字列検索システムとして実装し、他の機関では関係データベースシステムとして実装することが起こり得る。これは誤りではない。この実装を前提とした情報検索の仕組みを考えなければならない。例えば、ダブリンコア・メタデータにより、国文学研究資料館の

全情報資源が検索可能となっても、組織を越えた検索ゲートウェイとは一般に困難である。

そこで、これを解決する方法としては、

- (1) データクリアリングハウスの構築による
- (2) 検索手順についての標準規約を導入する

という2つの方法を考える。

データクリアリングハウスとは情報センターなどと解されるが、ネットワークを活用した情報の流通機構、すなわち情報の出所、入手方法などに関するデータを収集し、検索できるシステムを言う。インターネット上に情報資源を提供している機関は、その資源に関するアクセス情報（すなわち、メタデータ）をクリアリングハウスに登録する。利用者はクリアリングハウスを検索することにより、どこに、どのような情報が、どのような形式で存在しているかなどを知ることができる。

現在、このようなクリアリングハウスは増えつつある。例えば、地理情報クリアリングハウス・ゲートウェイ、人文科学ではECAI (Electronic Cultural Atlas In-

itiative) などがある。

一方、検索手順の標準規約（Z39.50）では、情報システムのハードウェアやソフトウェアに依存しない検索手順として、機関間のダブリンコア・メタデータベースシステムを結合した検索が可能である。前述のようなダブリンコア・メタデータベースシステムの実装とは無関係に実現することができる。現在、情報検索を目的とした世界的な標準交換規約としては、前述のZ39.50が挙げられる。

以上、2つの方法は互いに排他的なものではなく、補完的な手段であると考えられる。我々の試行は、後者（2）のZ39.50による解決法である。一般に、データクリアリングハウスでは専門領域のメタデータを収容する。つまりダブリンコア・メタデータに限定されず、対象分野の目的に応じた最適なメタデータを収容するシステムを構成できる。またデータの所在情報なども完備されるので、望ましい解決法であると考えられる。しかし、データクリアリングハウスを構築するためには、関連する機関などとの大幅な調整が必要であり、システムを維持、管理するためのコストも膨大となることが予想される。Z39.50による解決法の場合には、クリアリングハウスを構築する必要がないので実現が容易である。しかしながら、メタデータは機関ごとに管理されるので、少なくと所在情報の探索についてはユーザ側の作業となる問題がある。

### 3.3.4 メタデータベース・システムの構築

ダブリンコア・メタデータによるデータベースシステムの相互利用性と、Z39.50による複数のダブリンコア・メタデータベースシステムの透過的結合により、多様な情報資源を統一的に検索できるシステムの構築を目指す（以下、DC-Z39.50システムという。図7参照）。

DC-Z39.50システムでは、各データベースの要素をダブリンコア・メタデータへマッピングし、Z39.50のBib-1の要素をダブリンコア・メタデータへのアクセスポイントとして、検索できるようにする。これにより、OPACだけでなく、国文学研究資料館独自の書誌データベースや画像データベースなども検索できる。

DC-Z39.50システムは、データ生成部、メタデータ生成部、Z39.50サーバ、Z39.50-HTTPゲートウェイおよびデータベースから構成される。データ生成部は既存のデータをXML形式のデータに変換する。メタデータ生成部はXML形式に変換されたデータベースからダブリンコア・メタデータの要素を生成する。国文学研究資料館の殆どのデータはSGML化されている

ので、これらの変換は主にXSLTプロセッサによって行われる。Z39.50サーバはプロトコルを解釈し、その解釈に基づいて検索エンジンへパラメータを渡すと共にセッション関連の情報を管理する。Z39.50サーバは外部のZ39.50サーバあるいはZ39.50クライアントからの要求に応えることができる。Z39.50-HTTPゲートウェイは、WWWブラウザからの検索要求をZ39.50プロトコルに変換してZ39.50サーバに伝えると共に、Z39.50サーバからの応答をHTML文書に変換して利用者に返す。Z39.50-HTTPゲートウェイの特徴は、複数のZ39.50サーバと同時に通信できる点にある。これにより、複数のダブリンコア・メタデータベースの同時検索を実現している。

データベース（図7中では中間形式データベース）には、検索対象となるデータが蓄積されている。これらのデータベースは単独の検索システムとして機能すると共に、メタ情報検索の結果から、リンク情報を辿ってアクセスすることも可能である。

### 3.3.5 システムの評価

コラボレーションシステムを構築する際のマッピングの2つの課題、すなわち、

- (1) 各データベースから抽出すべき要素と、ダブリンコア・メタデータベースの要素間のマッピング
- (2) ダブリンコア・メタデータの要素と、Z39.50のBib-1アトリビュートセット間のマッピング

を解決する必要がある。

(1) の課題は、各データベースとダブリンコア・メタデータベースの要素を関連づけるガイドラインがないために、何らかの関連を新たに考慮する必要がある。そのため、現在のところマッピングはアドホックであり、一般的なOPACであっても、機関が異なればOPACの同じ要素がダブリンコア・メタデータの異なる要素へマッピングされる可能性がある。なお、各データベースから生成された要素とダブリンコア・メタデータベースの要素との関連は多対多である。

一方、(2)のダブリンコア・メタデータの要素と、Bib-1アトリビュートセットとのマッピングについては、以下の2つの方法が考えられる。

- (1) ダブリンコア・メタデータの15項目を、Z39.50のBib-1アトリビュートセットの内部にマッピングする方法。

## (2) ダブリンコア・メタデータ用に、Bib-1アトリビュートセットを拡張する方法。

現在の実証実験では、(1) の方法をとった。Bib-1アトリビュートセットに追加されたダブリンコア・メタデータ用の15項目をアクセスポイントに利用する。これらのアクセスポイントは、ダブリンコア・メタデータの要素と1対1対応であるため、マッピングが曖昧になる恐れがないためである。

付図9に、複数のダブリンコア・メタデータベースシステムを同時検索した例を示す。現時点では、国文学研究資料館が所有するデータベースのうち、マイクロ資料目録（館蔵マイクロフィルムの目録）、和古書目録（館蔵日本古典籍の目録）、論文目録（国文学年鑑による文学研究に関する論文目録）、史料所在目録（歴史史料の所在情報目録）、画像データベース（館蔵日本古典籍の画像データベース）、および動画データベース（演能関連の動画、音声などおデータ）の6つのデータベースが、コラボレーションシステムとして連携している。

コラボレーションシステムは、ダブリンコア・メタデータとZ39.50を併用した初期的な試みであり、機関内外の多様な情報資源への容易なアクセスを実現することを目指んでいる。コラボレーションシステムは漸く動き出した段階であり、他システムとの比較、評価などは今後の課題である。

なお、現時点で明らかになっている問題点について以下にまとめる。

### 3.3.6 今後の課題

ダブリンコア・メタデータについては、Dublin Core Simple (DCS) と Dublin Core Qualifier (DCO) という2つの考え方がある。Simple型の場合、15項目の基本要素をさらに細かく分けることはしない。これに対して、Qualifier型では基本要素を細かく分けようとする。ここではSimple型を採用している。しかし、他のデータクリアリングハウス、例えばECAIクリアリングハウスではQualifier型を採用しており、かつ独自の要素拡張を行っている。したがって、本コラボレーションシステムが他のデータクリアリングハウスとの協調を図る場合には、Qualifier型への拡張あるいは新たなマッピング法について検討を行う必要がある。

現在、Z39.50サーバを国内のいくつかのZ39.50サーバと接続し、正常に動作していることは確認されている。しかし、米国のZ39.50サーバとリンクさせる試験をUniversity of California San Diego大学図書館との

間で行ったところ、少なくとも2つの技術的な問題があった。1つはレコードシンタクスであった。レコードシンタクスは、Z39.50のスキーマによって変換された抽象データベースレコードを転送する際の物理構造について規定したものであり、汎用型(generic)レコードシンタクスと特定型レコード(content specific)シンタクスの2種類に分類される。汎用型にはGRS-1(Generic Record Syntax one)とSUTRS(Simple Unstructured Text Record)が、特定型には米国議会図書館のLC-MARKなどがある。我々のZ39.50サーバは国際的な利用を想定して、LC-MARKなどの特定レコードシンタクスには対応していないかった。しかし、米国のZ39.50サーバの多くがLC-MARKを採用しているため、検索結果を相互に変換することができなかった。

この問題については、我々のサーバを複数のMARCシンタクスに対応させることで、解決を図りつつある。

2つ目の問題は漢字コードであった。我々のZ39.50サーバはJIS,EUC,UNICODEに対応している。米国の場合、コンピュータの内部ではUNICODEを利用しているものの、通信の際にはEACCという米国標準の漢字コード(主に、図書館用)を使用している。このため、漢字データの変換が相互に実行できなかった。これについてはアメリカ側で対処する方向で検討を続けている。

## 付図1 SGML宣言 (EUC例)

---

```
<!SGML "ISO 8879-1986"

CHARSET
BASESET "ISO 646-1983//CHARSET
International Reference Version (IRV)//ESC 2/5 4/0"
DESCSET 0      256      0

CAPACITY PUBLIC "ISO 8879-1986//CAPACITY Reference//EN"

SCOPE DOCUMENT
SYNTAX

SHUNCHAR CONTROLS 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
18 19 20 21 22 23 24 25 26 27 28 29 30 31 127 255

BASESET "ISO 646 IRV"
DESCSET 0      256      0

FUNCTION RE    13
  RS     10
  SPACE 32
  TAB    SEPCHAR 9

NAMING LCNMSTRT
"&#160;&#161;&#162;&#163;&#164;&#165;&#166;&#167;&#168;&#169;&#170;&#171;&#172;
&#173;&#174;&#175;&#176;&#177;&#178;&#179;&#180;&#181;&#182;&#183;&#184;&#185;
&#186;&#187;&#188;&#189;&#190;&#191;&#192;&#193;&#194;&#195;&#196;&#197;&#198;
&#199;&#200;&#201;&#202;&#203;&#204;&#205;&#206;&#207;&#208;&#209;&#210;&#211;
&#212;&#213;&#214;&#215;&#216;&#217;&#218;&#219;&#220;&#221;&#222;&#223;&#224;
&#225;&#226;&#227;&#228;&#229;&#230;&#231;&#232;&#233;&#234;&#235;&#236;&#237;
&#238;&#239;&#240;&#241;&#242;&#243;&#244;&#245;&#246;&#247;&#248;&#249;&#250;
&#251;&#252;&#253;&#254,"

UCNMSTRT
"&#160;&#161;&#162;&#163;&#164;&#165;&#166;&#167;&#168;&#169;&#170;&#171;&#172;
&#173;&#174;&#175;&#176;&#177;&#178;&#179;&#180;&#181;&#182;&#183;&#184;&#185;
&#186;&#187;&#188;&#189;&#190;&#191;&#192;&#193;&#194;&#195;&#196;&#197;&#198;
&#199;&#200;&#201;&#202;&#203;&#204;&#205;&#206;&#207;&#208;&#209;&#210;&#211;
&#212;&#213;&#214;&#215;&#216;&#217;&#218;&#219;&#220;&#221;&#222;&#223;&#224;
&#225;&#226;&#227;&#228;&#229;&#230;&#231;&#232;&#233;&#234;&#235;&#236;&#237;
&#238;&#239;&#240;&#241;&#242;&#243;&#244;&#245;&#246;&#247;&#248;&#249;&#250;
&#251;&#252;&#253;&#254,"

LCNMCHAR "-"
UCNMCHAR "-"

NAMECASE GENERAL NO
ENTITY NO

DELIM GENERAL SGMLREF
SHORTREF SGMLREF
NAMES SGMLREF
QUANTITY SGMLREF
ATTCNT 100
NAMELEN 18
LITLEN 4000

FEATURES
MINIMIZE DATATAG YES      OMITTAG YES      RANK      YES      SHORTTAG
YES
LINK      SIMPLE NO      IMPLICIT YES      EXPLICIT NO
OTHER      CONCUR NO      SUBDOC NO      FORMAL      NO
APPINFO NONE>
```

---

## 付図2 嘸本大系のDTD定義例

```

<!DOCTYPE 原本 [
<!-- *****
<!-- * 揃絵・表のタイプ表記法宣言
<!-- *****
<!NOTATION DUMMY-GTYPE SYSTEM "dummy graphic notation type "
<!NOTATION DUMMY-TTYPE SYSTEM "dummy table notation type "
<!-- *****
<!-- * 外部エンティティ宣言
<!-- *****
<!ENTITY % graph-ent SYSTEM "entity#grpah.ent"
<!ENTITY % table-ent SYSTEM "entity#table.ent"
%graph-ent;
%table-ent;
<!-- *****
<!-- * マーク区間用エンティティ
<!-- *****
<!ENTITY % v1 "INCLUDE" -- 揃絵・表の対処検討時用 --
<!ENTITY % v2 "IGNORE" -- 揃絵・表の対処決定時(正式版)用 --
<!-- *****
<!-- * パラメータエンティティ
<!-- *****
<!ENTITY % para1
 "#PCDATA|傍記素|傍記|左傍記|左右傍記|左右二重傍記|右二重傍記|左二重傍記|間傍記|
 四角囲み|割書き|本文挿絵|本文表">
<!ENTITY % para2 "挃絵|表|作品書誌|小作品書誌|ファイル名">
<!ENTITY % para3 "作者事項*&典拠事項*>
<!ENTITY % para4 "挃絵|表|ファイル名">
<!ENTITY % para5 "#PCDATA|四角囲み|割書き|本文挿絵|本文表">
<!ENTITY % para6 "挃絵|表|ファイル名行番号">
<!-- *****
<!-- * 要素宣言
<!-- *****
<!ELEMENT TAG OMIT GROUP -->
<!ELEMENT SE -->
<!-- ***** 原本 ***** -->
<!ELEMENT 原本 O O (原本名,作品+)
+(ファイル名)>
<!ELEMENT 原本名 - O (%para1;)+ -(ファイル名)>
<!-- ***** 作品 ***** -->
<!ELEMENT 作品 - O (作品名,(%para3;),小作品+,作品奥書*)
+(作品書誌|挃絵|表)>
<!ELEMENT 作品名 O O (%para1;)+ -(%para4;|作品書誌)>
<!ELEMENT 作者事項 - O (%para1;)+ -(%para2;)>
<!ELEMENT 典拠事項 - O (%para1;)+ -(%para2;)>
<!ELEMENT 作品書誌 - O (作品書誌名?,本文レコード?)-(作品書誌)>
<!ELEMENT 作品書誌名 - O (%para1;)+ -(%para4;)>
<!ELEMENT 作品奥書 - O (作品奥書名?,本文レコード?)-(作品書誌)>
<!ELEMENT 作品奥書名 - O (%para1;)+ -(%para4;)>
<!-- ***** 小作品 ***** -->
<!ELEMENT 小作品 - O (小作品名?,(%para3;),嘶+,小作品奥書*)
-(作品書誌)+(小作品書誌)>
<!ELEMENT 小作品名 - O (%para1;)+ -(%para4;|小作品書誌)>
<!ELEMENT 小作品書誌 - O (小作品書誌名?,本文レコード?)-(小作品書誌)>
<!ELEMENT 小作品書誌名 - O (%para1;)+ -(%para4;)>
<!ELEMENT 小作品奥書 - O (小作品奥書名?,本文レコード?)-(小作品書誌)>
<!ELEMENT 小作品奥書名 - O (%para1;)+ -(%para4;)>
<!-- ***** 嘸 ***** -->
<!ELEMENT 嘸 - O (嘶名?,(%para3;),小嘶+)
-(小作品書誌)>
<!ELEMENT 嘸名 - O (%para1;)+ -(%para4;)>
<!-- ***** 小嘶 ***** -->

```

```

<!ELEMENT 小嘶      -O  (小嘶番号,キーワード+,小嘶名?,(%para3;),本文レコード)>
<!ELEMENT 小嘶番号 -O  EMPTY>
<!ELEMENT キーワード -O  (キー+)
<!ELEMENT キー      -O  (#PCDATA)>
<!ELEMENT 小嘶名    -O  (%para1;)+           -(%para4;)  >
<!-- ***** 本文レコード ***** -->
<!ELEMENT 本文レコード -O  (%para1;)+           +(行番号)>
<!-- ***** 本文挿絵・本文表 ***** -->
<!ELEMENT 本文挿絵   -O  EMPTY>
<!ELEMENT 本文表     -O  EMPTY>
<!-- ***** 挿絵・表 ***** -->
<!ELEMENT 握絵       -O  (題名?,後書?,本文*)      -(%para2;行番号)>
<!ELEMENT 表         -O  (題名?,後書?,本文*)      -(%para2;行番号)>
<!ELEMENT 題名      -O  (%para1;)+>
<!ELEMENT 後書      -O  (%para1;)+>
<!ELEMENT 本文      -O  (本文通番,本文題名?,本文レコード)>
<!ELEMENT 本文通番  O O  (%para1;)+>
<!ELEMENT 本文題名  -O  (%para1;)+>
<!-- ***** ファイル名 ***** -->
<!ELEMENT ファイル名 -O  EMPTY>
<!-- ***** 行番号 ***** -->
<!ELEMENT 行番号    -O  EMPTY>
<!-- ***** フラグ／制御記号 ***** -->
<!ELEMENT 傍記素    --  (%para5;)+           -(%para6;)>
<!ELEMENT 傍記      --  (%para5;)+           -(%para6;)>
<!ELEMENT 間傍記    --  (%para5;)+           -(%para6;)>
<!ELEMENT 左右傍記  --  (右傍記+,左傍記+)  -(%para6;)>
<!ELEMENT 右傍記    --  (%para5;)+           -(%para6;)>
<!ELEMENT 左傍記    --  (%para5;)+           -(%para6;)>
<!ELEMENT 左右二重傍記 -- (右二重傍記,左二重傍記) -(%para6;)>
<!ELEMENT 右二重傍記 -- (右側傍記+,左側傍記+) -(%para6;)>
<!ELEMENT 左二重傍記 -- (右側傍記+,左側傍記+) -(%para6;)>
<!ELEMENT 右側傍記  --  (%para5;)+           >
<!ELEMENT 左側傍記  --  (%para5;)+           >
<!ELEMENT 四角囲み   -O  EMPTY>
<!ELEMENT 割書き    --  (行+)
<!ELEMENT 行        --  (部分+|(%para1;+))  -(割書き) >
<!ELEMENT 部分      --  (%para1;)+           >
<!-- **** 属性定義 **** -->
<!-- * -->
<!-- ***** 属性定義 ***** -->
<!ATTLIST ファイル名 num  CDATA  #REQUIRED >
<!ATTLIST 小嘶番号  num  CDATA  #REQUIRED >
<!ATTLIST 行番号    pos  (L|M)   L
<!--[ %v1; [
<!ATTLIST 傍記素    fg   (ON|OFF) OFF >
<!ATTLIST 割書き    fg   (ON|OFF) OFF >
<!ATTLIST 四角囲み  fg   (S|E)   S >
]>
<!ATTLIST 本文挿絵  name ENTITY #IMPLIED >
<!ATTLIST 本文表    name ENTITY #IMPLIED >
<!ATTLIST 握絵      name ENTITY #IMPLIED >
<!ATTLIST 表        name ENTITY #IMPLIED >
]>
<!ATTLIST 本文挿絵  name ENTITY #REQUIRED >
<!ATTLIST 本文表    name ENTITY #REQUIRED >
<!ATTLIST 握絵      name ENTITY #REQUIRED >
<!ATTLIST 表        name ENTITY #REQUIRED >

```

---

### 付図3 SGMLによるマークアップ例

---

```
<小作品名>醒睡笑卷之一</小作品名>
<嘶>
<嘶名>謂被謂物之由来</嘶名>

<小嘶><小嘶番号 num= '1'><キーワード><キー></キー></キーワード>
<本文レコード>
<行番号 pos='M' num= '5'>△そらことをいふ物を， などうそつきとhaiひならハセ
<行番号 pos='M' num= '6'>し。されはにや， うそといふ鳥， 木のそらにとまりゐて<傍
記素 fg='OFF'>琴</傍記素><傍記>こと</傍記>
<行番号 pos='M' num= '7'>をひく<傍記素 fg='OFF'>縁</傍記素><傍記>ゑん</傍記>
によせ， そらことをうそつきといふよし。
</本文レコード>
</小嘶>

<小嘶><小嘶番号 num= '2'><キーワード><キー></キー></キーワード>
<本文レコード>
<行番号 pos='M' num= '8'>△いつれもおなし事なるを， <傍記素 fg='OFF'>常</傍記
素><傍記>つね</傍記>にたくをハ<傍記素 fg='OFF'>風呂</傍記素><傍記>ふろ</傍記>
といひ，
<行番号 pos='M' num= '9'>たてあけの戸なきを<傍記素 fg='OFF'>柘榴</傍記素><傍
記>しやくろ</傍記><傍記素 fg='OFF'>風</傍記素><傍記>ふ</傍記>呂とは， なんぞい
ふや。かゝ
<行番号 pos='M' num= '10'>みいるとの心也。 (3才) [1]
</本文レコード>
</小嘶>

<小嘶><小嘶番号 num= '3'><キーワード><キー></キー></キーワード>
<本文レコード>
<行番号 pos='M' num= '11'>△かいさうの<傍記素 fg='OFF'>類</傍記素><傍記>たく
ひ</傍記>にお<傍記素 fg='OFF'>期</傍記素><傍記>こ</傍記>といふ<傍記素
fg='OFF'>藻</傍記素><傍記>も</傍記>あり。かのおごもよく<傍記素 fg='OFF'>食</
傍記素><傍記>しよく</傍記>
<行番号 pos='M' num= '12'>をすゝむる<傍記素 fg='OFF'>功能</傍記素><傍記>こう
のう</傍記>あり。さてぞ<傍記素 fg='OFF'>武家</傍記素><傍記>ぶけ</傍記>の<傍記
素 fg='OFF'>台所</傍記素><傍記>たいところ</傍記>に， <傍記素 fg='OFF'>飯</傍記
素><傍記>めし</傍記>をはからひ
<行番号 pos='M' num= '13'>もり， 人にするむる<傍記素 fg='OFF'>役者</傍記素><
傍記>やくしや</傍記>をおごとはいふならし。
</本文レコード>
</小嘶>

<小嘶><小嘶番号 num= '4'><キーワード><キー></キー></キーワード>
<本文レコード><行番号 pos='M' num= '14'>△よろづ物のむさき事をきたないとhai
かに。北は水の
<行番号 pos='M' num= '15'>方なり。水なければ万物きよからず。しかるあひた， 水
な
<行番号 pos='M' num= '16'>いといふになそらへ， きたないといふかや。
</本文レコード>
</小嘶>
```

---

•  
•

#### 付図4 古典テキスト用 XML-DTD 例

```
<?xml version="1.0" encoding="Shift_JIS" ?>
<!-- ## 国文学研究資料館 ## -->
<!-- ## 岩波日本古典文学大系用 XML DTD ## -->
<!-- ##
<!-- ## 作成日 : 2001.10.2 ## -->
<!-- ## 更新日 : 2001.10.2 ## -->
<!-- ## 更新日 : 2001.10.20 ## -->
<!-- ## 更新日 : 2001.12.1 ## -->
<!-- ## 更新日 : 2001.12.21 ## -->
<!-- ## 更新日 : 2002.02.09 ## -->
<!-- ## VERSION : 2.3 ## -->
<!-- ## -->
<!-- <!DOCTYPE classics [ -->

<!-- =====
-->
<!-- parameter entity definition -->
<!-- =====
-->

<!ENTITY % para1 "#PCDATA | gaiji | ruby | between | square | divideline" >
<!--
<!ENTITY % para2 "#PCDATA | square | divideline" >
-->
<!ENTITY % para3 "illust | filename | lineno" >
<!ENTITY % para4 "illust | filename" >
<!ENTITY % para5 "illust" >
<!ENTITY % cm.fileDesc "#PCDATA" >
<!ENTITY % cm.encodingDesc "#PCDATA" >
<!ENTITY % cm.profileDesc "#PCDATA" >
<!ENTITY % cm.revisionDesc "#PCDATA" >
<!ENTITY % comp.kobanashi "kobanashi-illust |
    kobanashi-illust-name |
    kobanashi-illust-text |
    kobanashi-table-name |
    kobanashi-org-table-name |
    kobanashi-table-text" >

<!-- =====
-->
<!-- element definition -->
<!-- =====
-->

<!ELEMENT classics (
    header?,
    text
) >
(
    fileDesc,
    encodingDesc*,
    profileDesc*,
    revisionDesc?
)
```

```

<!ELEMENT fileDesc
      >
      (
        %cm.fileDesc;
      )>
      (
        %cm.encodingDesc;
      )>
      (
        %cm.profileDesc;
      )>
      (
        %cm.revisionDesc;
      )>
      (
        front?,
        body,
        back?
      )>
<!ELEMENT front
      (
        %para1;
      )*>
      (
        body.head* |
        body.bibl* |
        (div0+ |
          p |
          %para4;
        )* |
        body.back)* >
<!ELEMENT body.head
      (
        %para1;
      )*>
<!ELEMENT div0
      (
        div0.head* |
        div0.bibl* |
        (div1 |
          p |
          waka-group |
          kobanashi-group |
          %para3;
        )* |
        div0.back)* >
<!ELEMENT div1
      (
        div1.head* |
        div1.bibl* |
        (div2 |
          p |
          waka-group |
          kobanashi-group |
          %para3;
        )* |
        div1.back)* >
<!ELEMENT div2
      (
        div2.head* |
        div2.bibl* |
        (div3 |
          p |
          waka-group |
          kobanashi-group |
          %para3;

```

```

      )* |
    div2.back)* >
  (
  div3.head* |
  div3.bibl* |
  (div4 |
    p |
    waka-group |
    kobanashi-group |
    %para3;
  )* |
  div3.back)* >
  (
  div4.head* |
  div4.bibl* |
  (div5 |
    p |
    waka-group |
    kobanashi-group |
    %para3;
  )* |
  div4.back)* >
  (
  div5.head* |
  div5.bibl* |
  (div6 |
    p |
    waka-group |
    kobanashi-group |
    %para3;
  )* |
  div5.back)* >
  (
  div6.head* |
  div6.bibl* |
  (div7 |
    p |
    waka-group |
    kobanashi-group |
    %para3;
  )* |
  div6.back)* >
  (
  div7.head* |
  div7.bibl* |
  (
    p |
    waka-group |
    kobanashi-group |
    %para3;
  )* |
  div7.back)* >
  (
  %para1;
  )* >
  (
  %para1;
  )* >
  (
  %para1;

```



```

<!ELEMENT div4.back          /*>
(
  %para1;
)*>
(
  %para1; |
  %para3; |
  waka-group
)*>
(waka-number |
waka-foreword* |
waka-author? |
waka |
waka-back* |
waka-endword* |
%para3;
)*>
(
  kobanashi-number |
  kobanashi-name |
  kobanashi-keywords |
  kobanashi |
  kobanashi-author |
  kobanashi-supplement |
  %para3;
)*>
  EMPTY >
pos  (L|M)      "L"
id   CDATA       "0"      >
(
  %para1; |
  %para3;
)*>
  EMPTY >
pos  (L|M)      "L"

```

```

      id      CDATA      "0"      >
<!ELEMENT kobanashi-name      (
      %para1; |
      %para3;
    )* >
<!ELEMENT kobanashi-keywords  (
      %para1; |
      %para3;
    )* >
<!ELEMENT kobanashi           (
      %para1; |
      %para3; |
      %comp.kobanashi;
    )* >
<!ELEMENT kobanashi-author    (
      %para1; |
      %para3;
    )* >
<!ELEMENT kobanashi-supplement(
      %para1; |
      %para3;
    )* >
<!ELEMENT kobanashi-illust      (#PCDATA) >
<!ELEMENT kobanashi-illust-name  (#PCDATA) >
<!ELEMENT kobanashi-illust-text   (#PCDATA) >
<!ELEMENT kobanashi-table-name   (#PCDATA) >
<!ELEMENT kobanashi-org-table-name (#PCDATA) >
<!ELEMENT kobanashi-table-text    (#PCDATA) >
<!ELEMENT gaiji                 EMPTY >
<!ATTLIST gaiji                set (mojikyo |
                                         kokubunken) #REQUIRED
                                         code CDATA #REQUIRED >
<!ELEMENT ruby                  (
      (rb,
       rt) |
      (rbc,
       rtc+)
    )>
<!ATTLIST ruby                 fg (ON|OFF) "OFF" >
<!ELEMENT rbc                  (
      rb+
    )>
<!ELEMENT rtc                  (
      rt+
    )>
<!ATTLIST rtc                  position (right1|right2|left1|left2) #IMPLIED >
<!ELEMENT rb                   (
      %para1;
    )* >
<!ELEMENT rt                   (
      %para1;
    )* >
<!ATTLIST rt                  rbspan CDATA #IMPLIED >
<!ELEMENT between               (
      %para1;
    )* >
<!ELEMENT square                EMPTY >
<!ATTLIST square               fg (S|E) "S" >
<!ELEMENT divideline            (
      line+

```

```
<!ATTLIST divideline      )>
    fg (ON|OFF) "OFF" >
    (
        %para1;
    )* >
<!ELEMENT filename        EMPTY >
<!ATTLIST filename       id CDATA #REQUIRED >
    (
        %para1; |
        %para5;
    )* >
<!ATTLIST lineno          pos (L|M) "L"
    id CDATA "0" >
    EMPTY >
<!ELEMENT illust          id CDATA #IMPLIED >
<!ATTLIST illust
```

---

付図5 古典テキストの XML マークアップ例

```
<?xml version="1.0" encoding="Shift_JIS" ?>

<!DOCTYPE classics SYSTEM "../dtd/xml-n-classics2.dtd" [
]>
<?xml:stylesheet type="text/xsl" href="../xslt/classic/honmon_limit.xsl" ?>
<!--
<?xml:stylesheet type="text/xsl" href="../xslt/kokin/kokin.xsl" ?>
-->
<classics>
  <text>
    <body>
      <div0>
        <div0.title>源氏物語</div0.title>
        <div1>
          <div1.title><ruby fg="OFF"><rb> 桐 壺 </rb><rt> き り つ ぼ
</rt></ruby></div1.title>
          <filename id="p27" />
        </div1>
        <div1>
          <div1.title>きりつぼ</div1.title>
          <lineno pos="L" id="15">△いづれの御<ruby fg="OFF"><rb>時</rb><rt>とき
</rt></ruby>にか。女御・更衣あまたさぶらひ給ひけるなかに、いと、やむ</lineno>
          <lineno pos="L" id="16">ごとなき<ruby fg="OFF"><rb>際</rb><rt>きは
</rt></ruby>にはあらぬが、すぐれて<ruby fg="OFF"><rb>時</rb><rt>とき</rt></ruby>
めき<ruby fg="OFF"><rb>給</rb><rt>たま</rt></ruby>ふありけり。</lineno>
          <lineno pos="L" id="17">△はじめより、「われは」と、<ruby
fg="OFF"><rb>思</rb><rt>おも</rt></ruby>ひあがり給へる御かた※※、めざましき者に
お</lineno>
          <lineno pos="L" id="18">としめそねみたまふ。おなじ程、それより下<ruby
fg="OFF"><rb>臍</rb><rt>ら ふ</rt></ruby>の更衣たちは、まして、<ruby
fg="OFF"><rb>安</rb><rt>やす</rt></ruby>から</lineno>
          <lineno pos="L" id="19">ず。あさゆふの<ruby fg="OFF"><rb>宮</rb><rt>み
や</rt></ruby>づかへにつけても、人の<ruby fg="OFF"><rb>心</rb><rt>こゝろ
</rt></ruby>をのみ<ruby fg="OFF"><rb>動</rb><rt>う ご</rt></ruby>かし、<ruby
fg="OFF"><rb>恨</rb><rt>う ら</rt></ruby>みを<ruby fg="OFF"><rb>負</rb><rt>お
</rt></ruby>ふつもり</lineno>
          <lineno pos="L" id="110">にやありけむ、いと、あつしなりゆき、もの心
ばそげに<ruby fg="OFF"><rb>里</rb><rt>さと</rt></ruby>がちなるを、いよ</lineno>
          <lineno pos="L" id="111"> + + 「あかずあはれるもの」に<ruby
fg="OFF"><rb>思</rb><rt>お ぼ</rt></ruby>ほして、人の<ruby fg="OFF"><rb>謗
</rb><rt>そし</rt></ruby>りをも、えはゞからせ給は</lineno>
```

<lineno pos="L" id="112">ず、世の<ruby fg="0FF"><rb>例</rb><rt>ためし</rt></ruby>にもなりぬべき御もてなしなり。</lineno>

<lineno pos="L" id="113">△<ruby fg="0FF"><rb>上達部</rb><rt>かむだちめ</rt></ruby>・<ruby fg="0FF"><rb>上</rb><rt>うへ</rt></ruby>人なども、あいな<, <ruby fg="0FF"><rb>目</rb><rt>め</rt></ruby>をそばめつゝ、「いと、まばゆき、人の</lineno>

<lineno pos="L" id="114">御おぼえなり。<ruby fg="0FF"><rb>唐土</rb><rt>もろこし</rt></ruby>にも、かゝる、事の<ruby fg="0FF"><rb>起</rb><rt>おこ</rt></ruby>りにこそ、世も<ruby fg="0FF"><rb>亂</rb><rt>みだ</rt></ruby>れ<ruby fg="0FF"><rb>惡</rb><rt>あ</rt></ruby>しかりけれ」</lineno>

<lineno pos="L" id="115">と、やう++、<ruby fg="0FF"><rb>天</rb><rt>あめ</rt></ruby>の<ruby fg="0FF"><rb>下</rb><rt>した</rt></ruby>にも、あぢきなう、人のもて<ruby fg="0FF"><rb>惱</rb><rt>なや</rt></ruby>みぐさになりて、楊貴妃</lineno>

<lineno pos="L" id="116">の<ruby fg="0FF"><rb>例</rb><rt>ためし</rt></ruby>も、ひき<ruby fg="0FF"><rb>出</rb><rt>い</rt></ruby>でつべうなりゆくに、いと、はしたなきこと<ruby fg="0FF"><rb>多</rb><rt>おぼ</rt></ruby>かれど、かたじ</lineno>

<filename id="p28" />

<lineno pos="L" id="11">けなき御心ばへの、たぐひなきを<ruby fg="0FF"><rb>頼</rb><rt>たの</rt></ruby>みにて、まじらひたまふ。<ruby fg="0FF"><rb>父</rb><rt>ちゝ</rt></ruby>の大納言はな</lineno>

<lineno pos="L" id="12"><となりて、<ruby fg="0FF"><rb>母</rb><rt>はゝ</rt></ruby><ruby fg="0FF"><rb>北</rb><rt>きた</rt></ruby>の<ruby fg="0FF"><rb>方</rb><rt>かた</rt></ruby>なむ、いにしへの人の、<ruby fg="0FF"><rb>由</rb><rt>よし</rt></ruby>あるにて、<ruby fg="0FF"><rb>親</rb><rt>おや</rt></ruby>うち<ruby fg="0FF"><rb>具</rb><rt>く</rt></ruby>し、さし<ruby fg="0FF"><rb>當</rb><rt>あた</rt></ruby></lineno>

<lineno pos="L" id="13">りて世の<ruby fg="0FF"><rb>思</rb><rt>おぼ</rt></ruby>え<ruby fg="0FF"><rb>花</rb><rt>はな</rt></ruby>やかなる御かた※※にも<ruby fg="0FF"><rb>劣</rb><rt>おと</rt></ruby>らず、<ruby fg="0FF"><rb>何</rb><rt>なに</rt></ruby>事の<ruby fg="0FF"><rb>儀式</rb><rt>ぎしき</rt></ruby>をも、もてなし給</lineno>

<lineno pos="L" id="14">ひけれど、とりたてゝ、はか※※しき<ruby fg="0FF"><rb>後見</rb><rt>うしろみ</rt></ruby>しなければ、「<ruby fg="0FF"><rb>事</rb><rt>こと</rt></ruby>」あるときは、なほ</lineno>

・・・・・・・・・・・・・  
・・・・・・・・・・・・・

付図6 古典テキスト用 XSLT 例（XML から HTML への変換用）

```
<?xml version="1.0" encoding="Shift_JIS" standalone="no"?>
<!-- ##### -->
<!-- ## -->
<!-- ##      国文学研究資料館 -->
<!-- ##      岩波日本古典文学大系用 X S L T -->
<!-- ## -->
<!-- ## -->
<!-- ## -->
<!-- ## -->
<!-- ##      作成日 : 2002.01.16 -->
<!-- ##      VERSION : 0.1 -->
<!-- ## -->
<!-- ## -->
<!-- ## -->
<!-- -->
<!-- -->
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
    xmlns:xlink="http://www.w3.org/1999/xlink"
    xmlns="http://www.w3.org/1999/xhtml"
    version="1.0">

<xsl:variable name="title">国文学研究資料館 岩波日本古典文学大系</xsl:variable>
<xsl:variable name="touchu-width">30%</xsl:variable>
<xsl:variable name="honmon-width">70%</xsl:variable>
<!--
<xsl:variable name="touchu-file-name">okinwaka_toucyu.xml</xsl:variable>
<xsl:variable name="kouchu-file-name">okinwaka_koucyu.xml</xsl:variable>
<xsl:variable name="waka-touchu-file" select="document('okinwaka_toucyu.xml')"/>
<xsl:variable name="kouchu-file" select="document('okinwaka_koucyu.xml')"/>
-->

<xsl:output method="html"
    encoding="Shift_JIS"
    indent="no"
    omit-xml-declaration="yes" />
<xsl:strip-space elements="*"/>

<!-- waka -->
<xsl:include href=".//waka_limit.xsl" />

<!-- kobanashi -->
<xsl:include href=".//kobanashi.xsl" />

<!-- renga -->
<xsl:include href=".//renga.xsl" />

<!-- gaiji -->
<xsl:include href=".//gaiji.xsl" />

<!-- ===== -->
<!-- root -->
<!-- ===== -->
<xsl:template match="/">
    <html>
        <head>
            <meta http-equiv="content-type" content="text/html;
charset='Shift_JIS'" />
```

```

        <meta content="text/css" http-equiv="Content-Type">
            <link rel="stylesheet" type="text/css"
        href="../../css/classic.css" />
        </meta>
        <title>
            <xsl:value-of select="$title" />
        </title>
    </head>
    <body>
        <h1 align="center">
            <xsl:value-of select="$title" />
        </h1>
        <xsl:apply-templates />
    </body>
</html>
</xsl:template>

<!-- ===== -->
<!-- classics -->
<!-- ===== -->
<xsl:template match="classics">
    <xsl:apply-templates />
</xsl:template>

<!-- ===== -->
<!-- header -->
<!-- ===== -->
<xsl:template match="header">
    <xsl:apply-templates />
</xsl:template>

<!-- ===== -->
<!-- fileDesc -->
<!-- ===== -->
<xsl:template match="fileDesc">
    <xsl:apply-templates />
</xsl:template>

<!-- ===== -->
<!-- encodingDesc -->
<!-- ===== -->
<xsl:template match="encodingDesc">
    <xsl:apply-templates />
</xsl:template>

<!-- ===== -->
<!-- profileDesc -->
<!-- ===== -->
<xsl:template match="profileDesc">
    <xsl:apply-templates />
</xsl:template>

<!-- ===== -->
<!-- revisionDesc -->
<!-- ===== -->
<xsl:template match="revisionDesc">
    <xsl:apply-templates />
</xsl:template>

<!-- ===== -->

```

```

<!-- text -->
<!-- ===== -->
<xsl:template match="text">
<!--
-->
    <hr />
    <table border="0" width="100%" class="contents">
        <tr>
            <td>
                <h1>目次</h1>
            </td>
        </tr>
        <tr>
            <td valign="top">
                <table border="0" width="100%">
                    <xsl:for-each select="body/div0/div0.head |
                        body/div0/div1/div1.head | 
                        body/div0/div1/div2/div2.head">
                        <tr>
                            <td>
                                <p>
                                    <xsl:if
test="name()='div1.head'">
                                        <xsl:text>      </xsl:text>
                                    </xsl:if>
                                    <xsl:if
test="name()='div2.head'">
                                        <xsl:text>      </xsl:text>
                                    </xsl:if>
                                    <a href="#{generate-id()}">
                                        <xsl:apply-templates />
                                    </a>
                                </p>
                            </td>
                        </tr>
                    </xsl:for-each>
                </table>
            </td>
        </tr>
    </table>
<!--
-->
    <hr />
    <xsl:apply-templates />
</xsl:template>

<!-- ===== -->
<!-- front -->
<!-- ===== -->
<xsl:template match="front">
    <div>
        <xsl:apply-templates />
    </div>
</xsl:template>

<!-- ===== -->
<!-- body -->
<!-- ===== -->

```

```

<xsl:template match="body">
    <div>
        <xsl:apply-templates />
    </div>
</xsl:template>

<xsl:template match="body.head">
    <p>
        <xsl:apply-templates />
    </p>
</xsl:template>

<xsl:template match="body.bib">
    <p>
        <xsl:apply-templates />
    </p>
</xsl:template>

<xsl:template match="body.back">
    <p>
        <xsl:apply-templates />
    </p>
</xsl:template>

<!-- ===== -->
<!-- back -->
<!-- ===== -->
<xsl:template match="back">
    <div>
        <xsl:apply-templates />
    </div>
</xsl:template>

<!-- ===== -->
<!-- div -->
<!-- ===== -->
<xsl:template match="div0">
    <div>
        <xsl:apply-templates />
    </div>
</xsl:template>

<xsl:template match="div1">
    <div>
        <xsl:apply-templates />
    </div>
</xsl:template>

<xsl:template match="div2">
    <div>
        <xsl:apply-templates />
    </div>
</xsl:template>

<xsl:template match="div3">
    <div>
        <xsl:apply-templates />
    </div>
</xsl:template>

```

```

<xsl:template match="div4">
    <div>
        <xsl:apply-templates />
    </div>
</xsl:template>

<xsl:template match="div5">
    <div>
        <xsl:apply-templates />
    </div>
</xsl:template>

<xsl:template match="div6">
    <div>
        <xsl:apply-templates />
    </div>
</xsl:template>

<xsl:template match="div7">
    <div>
        <xsl:apply-templates />
    </div>
</xsl:template>

<!-- ===== -->
<!-- div head -->
<!-- ===== -->
<xsl:template match="div0.head">
    <table class="div0_title" border="0" width="100%">
        <tr>
            <td>
                <h1 id="{generate-id()}">
                    <xsl:apply-templates />
                </h1>
            </td>
        </tr>
    </table>
</xsl:template>

<xsl:template match="div1.head">
    <table class="div1_title" border="0" width="100%">
        <tr>
            <td>
                <h2 id="{generate-id()}">
                    <xsl:apply-templates />
                </h2>
            </td>
        </tr>
    </table>
</xsl:template>

<xsl:template match="div2.head">
    <table class="div2_title" border="0" width="100%">
        <tr>
            <td>
                <h3 id="{generate-id()}">
                    <xsl:apply-templates />
                </h3>
            </td>
        </tr>

```

```

        </table>
</xsl:template>

<xsl:template match="div3.head">
    <h4 id="{generate-id()}">
        <xsl:apply-templates />
    </h4>
</xsl:template>

<xsl:template match="div4.head">
    <p id="{generate-id()}">
        <xsl:apply-templates />
    </p>
</xsl:template>

<xsl:template match="div5.head">
    <p id="{generate-id()}">
        <xsl:apply-templates />
    </p>
</xsl:template>

<xsl:template match="div6.head">
    <p id="{generate-id()}">
        <xsl:apply-templates />
    </p>
</xsl:template>

<xsl:template match="div7.head">
    <p id="{generate-id()}">
        <xsl:apply-templates />
    </p>
</xsl:template>

<!-- ===== -->
<!-- div bib -->
<!-- ===== -->

<xsl:template match="div0.bib">
    <p>
        <xsl:apply-templates />
    </p>
</xsl:template>

<xsl:template match="div1.bib">
    <p>
        <xsl:apply-templates />
    </p>
</xsl:template>

<xsl:template match="div2.bib">
    <p>
        <xsl:apply-templates />
    </p>
</xsl:template>

<xsl:template match="div3.bib">
    <p>
        <xsl:apply-templates />
    </p>
</xsl:template>

```

```
<xsl:template match="div4.bib">
  <p>
    <xsl:apply-templates />
  </p>
</xsl:template>

<xsl:template match="div5.bib">
  <p>
    <xsl:apply-templates />
  </p>
</xsl:template>

<xsl:template match="div6.bib">
  <p>
    <xsl:apply-templates />
  </p>
</xsl:template>

<xsl:template match="div7.bib">
  <p>
    <xsl:apply-templates />
  </p>
</xsl:template>

<!-- ===== -->
<!-- div back -->
<!-- ===== -->
<xsl:template match="div0.back">
  <p>
    <xsl:apply-templates />
  </p>
</xsl:template>

<xsl:template match="div1.back">
  <p>
    <xsl:apply-templates />
  </p>
</xsl:template>

<xsl:template match="div2.back">
  <p>
    <xsl:apply-templates />
  </p>
</xsl:template>

<xsl:template match="div3.back">
  <p>
    <xsl:apply-templates />
  </p>
</xsl:template>

<xsl:template match="div4.back">
  <p>
    <xsl:apply-templates />
  </p>
</xsl:template>

<xsl:template match="div5.back">
  <p>
    <xsl:apply-templates />
  </p>
</xsl:template>
```

```

        </p>
</xsl:template>

<xsl:template match="div6.back">
    <p>
        <xsl:apply-templates />
    </p>
</xsl:template>

<xsl:template match="div7.back">
    <p>
        <xsl:apply-templates />
    </p>
</xsl:template>

<!-- =-->
<!-- p -->
<!-- =-->
<xsl:template match="p">
    <xsl:apply-templates />
    <br />
</xsl:template>

<!-- =-->
<!-- ruby -->
<!-- =-->
<xsl:template match="ruby">
    <ruby>
        <xsl:apply-templates />
    </ruby>
</xsl:template>

<xsl:template match="rbc">
    <xsl:apply-templates />
</xsl:template>

<xsl:template match="rtc">
    <xsl:apply-templates />
</xsl:template>

<xsl:template match="rb">
    <rb>
        <xsl:apply-templates />
    </rb>
</xsl:template>

<xsl:template match="rt">
    <rt>
        <xsl:apply-templates />
    </rt>
</xsl:template>

<!-- =-->
<!-- between -->
<!-- =-->
<xsl:template match="between">
    <span class="between">
        <span class="parenthesis">
            <xsl:text>(</xsl:text>
        </span>

```

```

<xsl:apply-templates />
<span class="parenthesis">
    <xsl:text>)(</xsl:text>
</span>
</span>
</xsl:template>

<!-- =-->
<!-- square -->
<!-- =-->
<xsl:template match="square">
    <span class="square">
        <xsl:apply-templates />
    </span>
</xsl:template>

<!-- =-->
<!-- divideline -->
<!-- =-->
<xsl:template match="divideline">
    <span class="divideline">
        <xsl:if test="not(parent::between)">
            <span class="parenthesis">
                <xsl:text>(</xsl:text>
            </span>
        </xsl:if>
        <xsl:apply-templates />
        <xsl:if test="not(parent::between)">
            <span class="parenthesis">
                <xsl:text>)(</xsl:text>
            </span>
        </xsl:if>
    </span>
</xsl:template>

<!-- =-->
<!-- line -->
<!-- =-->
<xsl:template match="line">
    <span class="line">
        <xsl:apply-templates />
        <xsl:if test="not(position()=last())">
            <span class="separator">
                <xsl:text> | </xsl:text>
            </span>
        </xsl:if>
    </span>
</xsl:template>

<!-- =-->
<!-- filename -->
<!-- =-->
<xsl:template match="filename">
    <p class="filename" align="right">
        <xsl:choose>
            <xsl:when test="preceding::div1.head[position()=1]">
                <xsl:choose>
                    <xsl:when
                        with(ancestor::div1/child::div1.head[position()=1], '古今和歌集')>
                            <xsl:value-of
                                test="starts-

```

```

select="translate(substring-after(preceding::div1.head[position()=1], '古今和歌集'), '△', '　') />
        <xsl:text>　</xsl:text>
    </xsl:when>
    <xsl:otherwise>
        <xsl:value-of
select="translate(ancestor::div1/child::div1.head[position()=1], '△', '　') />
        <xsl:text>　</xsl:text>
    </xsl:otherwise>
</xsl:choose>
</xsl:when>
<xsl:otherwise>
</xsl:otherwise>
</xsl:choose>

<xsl:choose>
<xsl:when test="ancestor::div1/child::div2/child::div2.head and
preceding-sibling::div2.head">
    <xsl:value-of
select="translate(preceding::div2.head[position()=1], '△', '　') />
    <xsl:text>　</xsl:text>
    </xsl:when>
    <xsl:when test="ancestor::div1/child::div2/child::div2.head and
not(preceding-sibling::div2.head)">
        <xsl:when>
        <xsl:otherwise>
        </xsl:otherwise>
    </xsl:choose>
<!--
<a href="{concat($touchu-file-name, '#', @id)}" target="toucyu" id="@id">
-->
    <xsl:value-of select="substring-after(@id, 'p')"/>
    <xsl:text>ページ</xsl:text>
<!--
-->
</p>
</xsl:template>

<!-- ===== -->
<!-- lineno -->
<!-- ===== -->
<xsl:template match="lineno">
    <div>
        <xsl:attribute name="class">
            <xsl:choose>
                <xsl:when test="string-length(substring-after(@id, 'l'))=1">
                    <xsl:text>lineno1</xsl:text>
                </xsl:when>
                <xsl:when test="string-length(substring-after(@id, 'l'))=2">
                    <xsl:text>lineno2</xsl:text>
                </xsl:when>
                <xsl:when test="string-length(substring-after(@id, 'l'))=3">
                    <xsl:text>lineno3</xsl:text>
                </xsl:when>
                <xsl:when test="string-length(substring-after(@id, 'l'))=4">
                    <xsl:text>lineno4</xsl:text>
                </xsl:when>
                <xsl:otherwise>
                </xsl:otherwise>
            </xsl:choose>
    </div>

```

```

</xsl:attribute>
<span class="lineno_no">
    <xsl:value-of select="substring-after(@id, 'l')"/>
    <xsl:text>行</xsl:text>
</span>

<xsl:apply-templates />

<!--
<xsl:choose>
    <xsl:when test="concat(preceding-
sibling::filename[position()=1]/@id, @id)=$kouchu-file//pageline/@idref">
        <span>      <a href="{concat($kouchu-file-name, '#',
preceding-sibling::filename[position()=1]/@id, @id)}" target="koucyu">●校注</a></span>
    </xsl:when>
-->
<!--
<xsl:otherwise>
    <span>      </span>
</xsl:otherwise>
-->
<!--
<xsl:choose>
-->

</div>
<!-- test test -->
</xsl:template>

<!-- ===== -->
<!-- illust -->
<!-- ===== -->
<xsl:template match="illust">
    <xsl:apply-templates />
</xsl:template>

<!-- ===== -->
<!-- △の削除、全角スペースへの置き換え -->
<!-- ===== -->
<xsl:template match="text()">
    <xsl:choose>
        <xsl:when test="starts-with(current(), '△△△△')">
            <xsl:value-of select="translate(current(), '△', '　')"/>
        </xsl:when>
        <xsl:when test="starts-with(current(), '△△△')">
            <xsl:value-of select="translate(substring-after(current(), '△△△'), '△', '　')"/>
        </xsl:when>
        <xsl:when test="starts-with(current(), '△△')">
            <xsl:value-of select="translate(substring-after(current(), '△△'), '△', '　')"/>
        </xsl:when>
        <xsl:otherwise>
            <xsl:value-of select="translate(current(), '△', '　')"/>
        </xsl:otherwise>
    </xsl:choose>
</xsl:template>

</xsl:stylesheet>

```

付図7 画像データベースの検索例



付図8 全文データベースの利用の画面例



付図9 DC-Z39.50システムの検索画面例

The screenshot shows a Microsoft Internet Explorer window displaying a search interface for a DC-Z39.50 system. The title bar reads "詳細検索 - Microsoft Internet Explorer". The address bar shows the URL "http://base2nij.ac.jp/~z3950/dc.cgi-bin/z3950.cgi". The main content area is titled "md@ 詳細検索". A message at the top says "プラウザの戻るボタンは使用しないで下さい。検索終了時は [検索先選択] をクリックして下さい。" Below this is a table titled "検索状態" showing port numbers and connection status:

状態	ホストアドレス	ポート番号	データベース名	参考
<input type="radio"/>	17216.642	211	dc-mcrs	Connection OK
<input type="radio"/>	17216.642	212	dc-wako	Connection OK
<input type="radio"/>	17216.642	213	dc-renbin	Connection OK
<input type="radio"/>	17216.642	214	dc-history	Connection OK
<input type="radio"/>	17216.642	215	dc-images	Connection OK

Below the table is a "検索条件入力" section with fields for "検索対象" (DC-Title, DC-Creator), "キーワード" (田舎源氏), and "項目内条件" (AND). The "検索履歴" section shows a single entry: "表示 集合番号 ヒット件数 検索内容" followed by the query "@attr User=DC-Title \"田舎源氏\"". The bottom status bar indicates "ページが読み込まれました" and the address bar shows "検索結果 - Microsoft Internet Explorer".

# 古典テキストのデジタル化と データベース構築・利用支援システムの開発

及川 昭文

## 【要旨】

古典テキストのデジタル化、データベース化を支援するためのソフトウェアの開発、及びデータベースをWeb上で公開するための新しい手法「BBDB」を開発し、その実現を図った。「BBDB」は Bare Bone Database の略で、従来の DBMS (Database Management System) をまったく利用しないで、同等の機能、性能を実現したものである。

具体的には以下のような項目について研究開発を進めた。

- 1) 古典テキスト原典の PDF 化に関するフィージビリティ研究
- 2) データベース構築・利用支援ソフトウェアの開発
- 3) 日本語・ギリシャ語混在表記ツールの開発
- 4) BB-DB 手法の開発及びその実現
- 5) 古今集のパラレル・テキストデータベースの作成

## 【他領域との連携による成果】

本研究班は古典テキストの情報処理を中心テーマとしているが、具体的な開発にあたっては、古典研究を専門とする人文系の研究者との密接な連携を必要とした。すなわち、情報処理に対する研究者の具体的なニーズの分析、開発されたソフトウェアの評価等において、忌憚のない意見交換を進めてきた。その成果の一つは DDDB 手法の開発である。これは、データベースやソフトウェアについての専門的な知識がなくても、自らが持つさまざまな資料（データ）をデータベースとして、Web で公開するための一連の処理手法をソフトウェアを含めて開発したものである。

## 【位置付け】

本研究は、直接的に古典テキストをデータベース化したり、分析することではなく、そのための基礎的な

ソフトウェアの開発、Web 公開のための手法開発である。すなわち、古典研究者が古典テキストに対して情報処理を試みるのに必要な基本的ソフトウェアの整備を目指したものである。

## 【平成11年度～12年度の研究成果】

### 1) 古典テキスト原典の PDF 化に関するフィージビリティ研究

スキャナで古典テキストを取り込み、PDF (Portable Document Format) 型式で蓄積する方法について検討した。PDF 型式のデータは、エクスプローラやネットスクープなどのブラウザがあれば、インターネットを介して、どこからでも簡単に利用できるもので、非常に汎用性の高い型式である。スキャナで読み込むときの最適な解像度や、利用上の問題点など、基礎的なことがらについて分析し、古典テキスト原典のデジタル化として PDF が適しているかどうかのフィージビリティ・スタディを行った。

### 2) データベース構築・利用支援ソフトウェアの開発

これまでに開発を続けてきたいたプログラム群を集大成し、ホスト・コンピュータ上でデータベースの編集やエラー修正、冊子体資料の作成等を容易に行えるプログラム・パッケージを開発した。現在、このパッケージのサーバへの移植を進めている。

市販されている、あるいはメーカー等の提供するデータベース関連ソフトウェアには、一般的に構築を支援するプログラムは含まれていない。すなわち、機械可読化されているいないに関わらず、データはすべてエラーのないクリーンな状態にあることを前提としており、データの均質性の保持は利用者側で対応すべきものとなっている。そこで、いろいろな種類のデータベース構築を行う過程で、汎用性を持った支援プログラ

ム群 UPS (Utility Package for Standard Data Format File) の開発を継続して行ってきた。この UPS 開発の主な目的は、①データの物理的、論理的なエラー・チェックの自動化と、②豊富な冊子体資料の作成の 2 つである。

①は人手と時間を要し、かつ単純作業の連続であることにより、その省力化・効率化のみならず、データの品質を一定の基準に維持することをめざすものである。②は構築されたデータベースをオンラインで利用するだけでなく、冊子体資料として活用できるようにすることをめざすものである。とくに考古学においては、オンラインで情報検索をするだけでなく、データベースに基づいて分布図やグラフ、あるいは地名表などの 3 次、4 次資料を作成することが多く、そのためにはいろいろな種類の冊子体資料が不可欠なものとなってくる。

UPS は1980年代の中頃から開発を始め、新種のデータベースや新しい問題に対応できるように、プログラムのバージョン・アップや追加が続けられてきており、現在では約100本のメインプログラムと200以上のサブルーチンから構成されている。約20のサブルーチンを除いて、すべて FORTRAN 言語で書かれており、非常に高い可般性と汎用性を有している。

UPS 全体としての構成は次のようになる。

- ・データチェック
- ・データ変換（自動修正を含む）
- ・レコード変換
- ・マージ処理（項目及びレコード単位）
- ・ソート処理
- ・リスト出力
- ・KWIC（Keyword In Context）リスト作成処理
- ・統計処理
- ・分かち書き処理
- ・その他

UPS で処理できる文字は 1 バイト系の EBCDIC コードと 2 バイト系の JEF コードであるが、 $\hat{a}$   $\hat{e}$   $\hat{i}$   $\hat{o}$   $\hat{u}$  等の文字も 1 バイト系文字として取り扱うことができる。また、EBCDIC コードではカタカナとアルファベットの小文字が同じコードに割り振られているが、これを異なるコードに割り振り、混在したテキストの処理も可能としている。

### 3) 日本語・ギリシャ語混在表記ツールの開発

西洋古典学を専門としない研究者であっても、人文科学系の研究に従事していると、時として古典ギリシ

ア語の単語を表記する必要に迫られることがある。ギリシア語に特有のアクセントや氣息記号を無視し、単にギリシア文字だけを表記したいのであれば、TeX の数式モードで一語ずつギリシア文字を記述することによって、一応の目的をはたすことはできる。しかし、アクセントや氣息記号などを遗漏なく表記しようと思えば、この方法ではほとんど絶望的である。

多言語環境を備えた一部のワープロソフトを使用すれば、ある程度このような要求を満たすことはできる。しかしこの場合には、フォントライセンスの関係から、PDF ファイルなどに出力したとき、ギリシア語フォントを埋め込めないといった問題の発生がある。これは一般に公開することを目的とした PDF の使用において、きわめて大きな問題となる。さらにプリントアウトの品質や、文献データベースとの動的連携などを考慮すると、できる限り TeX 上で日・希混在環境を実現することが望ましいとも言える。

日本語 TeX 環境を実現する事実上の標準は、pLaTeX2e であると言ってよい。pLaTeX2e 上で日・希混在環境を実現する方法には二つある。一つは GNU Emacs/Meadow 上でギリシア語を使用するための CGreek パッケージを導入し、さらに入力・編集結果を pLaTeX2e 上で処理するために、Ibycus4 パッケージと Levy フォントを導入するというものである。この方法については CGreek プロジェクトの詳細な解説がある。もう一つの方法は、LaTeX 2e に多言語処理機能を付加するものとして有名な Babel パッケージを利用するというものである。

CGreek パッケージを利用する方法には、二つの大きな利点がある。一つは、GNU Emacs/Meadow によるテキスト入力時点において、日・希（欧）の混在文が視覚的に確認できるという点である。さらに CGreek パッケージは、TLG が制作・配布するギリシア古典を収録した CD-ROM の利用環境を統合しているため、極めて完成度の高いギリシア古典データベースを活用しながらテキストの執筆が可能であるという特徴をもっている。したがって、西洋古典学を専門とする研究者は、迷わずこの方法を選択すべきであろう。

しかしこの方法は、エディタに GNU Emacs/Meadow が使用されていることを大前提としているため、特に西洋古典学を専門とするなどの特別な理由がない場合、日ごろ使い慣れているエディタのままで、日・希（欧）混在環境を実現したいと思う研究者も多いことだろう。この場合に有効なのが上記第二の方法、すなわち LaTeX2e に Babel パッケージを導入するというものである。またこの場合、一般に TeX 上で多

言語処理を行うためには Babel パッケージを使用するという方法が多くとられているところから、すでに Babel パッケージを利用している研究者にあっては、これまでの環境を維持したままギリシア語処理機能を付加できるという利点もある。

しかしながら、Babel パッケージは、本来、欧文 LaTeXe 用に開発されたものであり、そのままでは日本語 pLaTeXe 上で使用することはできない。Babel を pLaTeXe 上で使用するためには Babel および pLaTeXe システムに対して種々の調整を加える必要がある。よって本研究では、Babel および pLaTeXe のシステム構成を探査し、両者の間に互換性を実現するための方法を確立することによって、日・希混在環境の構築をめざしている。

#### 4) BBDB 手法の開発及びその実現

ここで DBMS を利用することのメリット、デメリットを論じる余裕はないが、BBDB はそれぞれの DB に対応した小規模なプログラム群から構成される各種ツールを用意し、DBMS を利用することなく同等の機能を実現しようとするものである。

BBDB の最大の特徴は、1R1H、すなわち1 record を1 HTML file としてサーバ上に格納するところにある。このことによって従来の DBMS 以上の柔軟性と拡張性を実現している。例えば、ある特定のレコードに関して例外的な処理を必要としても、その処理を行

うページにリンクを張るだけでよい。そのリンクもあるディレクトリに例外処理のページを格納しておけば、自動的に対象とする遺跡のページとリンクを張れるようなツールを用意している。

BBDB は、現在開発中のシステムであり、その詳細を体系的に示すことはできないが、大まかな概念は図 1 のとおりである。ここでは DB 管理者、DB エンジニア、サーバエンジニアの 3 つの役割が必要になる。もちろん 1 人が全てを兼ねる場合もあるが、通常は DB 管理者 (=DB エンジニア) とサーバエンジニアという構成になる。そして、DB 管理者 (=DB エンジニア) が公開したい DB を持っている研究者ということになる。いったんシステムが確立されれば、その運用に必要なのは DB 管理者 (=研究者) のみとなる。運用は全て Web 上の運用管理ページから行えるようになっており、特別な処理や新しい機能の追加等を除けば、DBMS の素人である研究者でも容易に運用していくことが可能である。

このようなことが可能になっているのは、BBDB が simple is best, small is beautiful という基本哲学のもとでシステム開発を進めているからである。すなわち、処理プログラムは可能な限り単純に、そして小さな単位で行うようになっており、保守が容易で拡張性に富んだものとなっている。

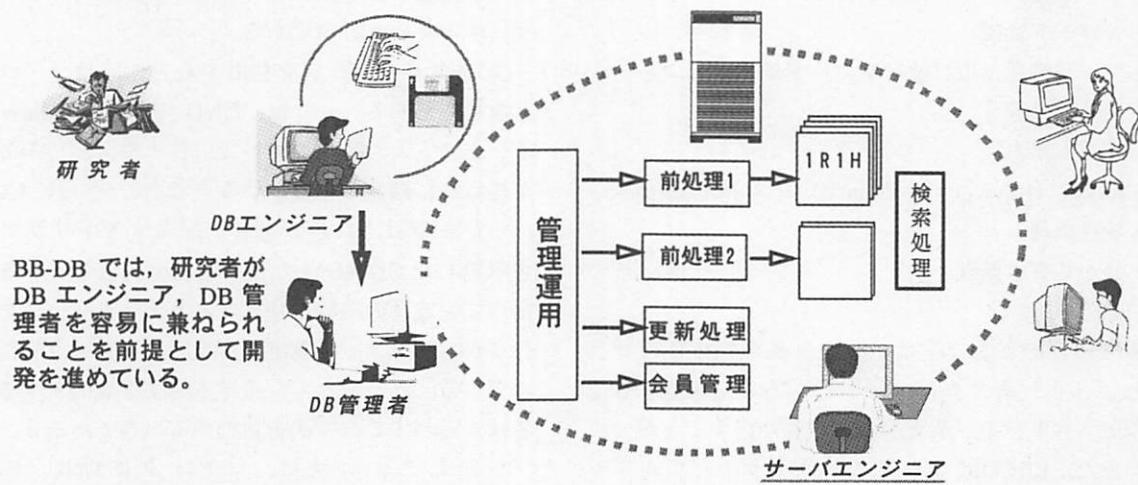


図 1 BBDB システムの概念図

#### サーバへの実装上の基本原則

ここでは DB をサーバへ実装のための基本原則について解説する。実装のための処理は、大きく DB 公開とシステム管理の 2 つの側面がある。DB 公開には、

前処理、CGI によるプログラム処理、システム管理には、更新システム、ユーザ管理システム、また、開発過程においては、DB の内容を担当する DB エンジニアリングと DB をサーバで公開するサーバエンジニ

アリングの2つの仕事が発生することになる。

前処理と場合分け：公開するデータは、可能な限りあらかじめ html に変換する。しかし、一つひとつのファイルを管理していたのでは、大変なので、これらは公開データとし、直接編集することはない。管理すべきファイルは、サーバにはおかれず、管理は個人パソコンの表ソフト等で管理更新を行う。サーバにはテキスト形式のファイルを転送し、そこからそれぞれのレコードの html ファイルを生成する。これを前処理と呼ぶ。

前処理では、公開に際して予測されるユーザの行動を考慮して、静的部分、動的部分に大きくわけ、静的部分については前処理で、html を出力し、動的部分については cgi を用意する。例えば、県別インデックス（47県）のように数の固定されているものについては、静的処理として、前処理で ken.html を生成する。また、頻繁に検索が実施されると考えられるデータについては、あらかじめ前処理でテーブルを抽出しておき、そのテーブルを対象に検索を行う。さらに、対象によって、将来的に更新される部分と更新されない部分でプログラムを書き分ける。

プログラム開発：前処理の実行は大きな CPU 負荷を伴ったとしても、頻繁に行われる作業ではないので、できるだけシンプルに、それぞれのステップの入れ替えや書き換えが容易に行えるよう明示的に記述する。過度に高級な技術、あるいは処理のための処理といったメタ処理はできるだけ使わない。単純な技術の組合せを重視し、プログラムの可読性を高くし、プログラムはできるだけ小さい単位で作成する。また、運用中の機能の追加、機能別に独立したテスト等が行えるように、システム全体に影響を与えるようなプログラムの開発はしない。高速に、確実に、生のデータを直接アクセスすることで、管理効率のよいシステムを構築する。

ユーザインタフェース：クライアントの CPU やブラウザにパフォーマンスが依存しないように、サーバから転送するデータは html を基本とし、転送に負荷となるグラフィックスや java はできるだけ使わない。ユーザがそのまま html ファイルを保存した際、論文等に引用しやすいよう、それぞれ独立したレコードとして出力する。それぞれわかりやすいように、インデックス別、機能別にリンクを記述したトップページを用意する。

管理者インタフェース：管理者は、トップページの html ファイルの編集と利用者の管理、DB 全体の更新を行う。トップページは一般的な index.html であり、このトップページは、DB 管理者が関連情報をリンクしたり、おしえ、利用マニュアルを追加できるように、単純な html で提供する。一方、利用者の管理、DB 全体の更新は、別のディレクトリに置かれたページから行う。http 標準実装の認証処理を行い、管理者以外はアクセスできない。このページには、管理作業別にリンクが施され、どういう手順で更新すればよいかは、プログラム実行中に明示され、作業に間違いのないよう確認メッセージを出す。

DB エンジニアとサーバエンジニア：DB エンジニアとサーバエンジニアの役割と目的について説明する。DB エンジニアは、データシートのデザイン、データの入力、データの修正、データの表記一貫性の確保、またそのチェックプログラムの開発とデータチェックの実施、コンピュータで制約のある文字表現等の問題とその解決、原本と出典、1 次データとの関連位置づけ、データシートの管理を行い、DB の目的、DB 利用方法、利用規定の明確化、ユーザとの情報交換等の作業を行う。

サーバエンジニアとは、データシートのデザイン及び対象 DB の特性をよく検討し、前処理の実施方法、検索の方法を検討する。また、プログラミングを行うだけでなく、データ出力のテストを行いつつ、対象の内容を十分に考慮した、集計方法、可視化の方法等を検討し、実装する。その上で、ユーザはどのような検索を望んでいるのか、どのような情報を必要とするか、どの程度の頻度及び範囲で更新を行うか、追加情報はどのように取り扱うか、ユーザのデータ読み出しアクセスはどこまで許可するのか、ユーザのデータ書き換えを認めるか等を、DB エンジニアとサーバエンジニアの間であらかじめ話し合い、動的データ、静的データ、検索の範囲、データ更新の方法を決定する。さらに、品質を向上させるためには、利用者を増やすためにはどのようにすればよいかも検討し、ワークショップや研究発表等についても検討している。

5) 古今集のパラレル・テキストデータベースの作成  
古典文学においては解釈や吟味が研究の多くを占める中、近藤（「古今和歌六帖の歌語データベース化によって見た歌語の位相」『歌ことばの歴史』笠間書院1998）のような作者の性別により、和歌を計量分析し、和歌におけるジャンダを発見する試みや村上（「文

章分析と統計学」、特集／知としての統計学、数理科学 11、サイエンス社、1995) らのように作品の真質、作者の推定を多変量解析による判別する研究が近年行われはじめた。これらの試みは、統計学やプログラミング、データベースなどの計算機技術が人文科学においても科学的分析の方法として取り入れられている事実を示している。本研究では、このような古典文学の計量分析を支援するためのデータベースを開発することを目的としている。今回は、BBDB によるデータベース化を試みた。

「古今集」とはいっても文化・歴史があるだけにその研究分野は幅広い。「古今集」をキーワードに国立国語研究所国語学研究文献総索引データ第1.02版を検索、集計した結果、表のように文章・文体、文法に関する研究が多いことがわかる。また、古語にもかかわらず、音声・音韻に関する研究が少なからずあることから、日本語音韻の歴史的変遷を分析する上で重要な意義があるのであろう。

国語学研究文献総索引による「古今集」の研究分野の集計

順位	頻度	分類	順位	頻度	分類
1	52	文章・文体	9	7	音声・音韻
2	21	古典の注釈	10	5	年鑑単行本
3	20	国語史	11	4	国語資料
4	18	大辞典	12	2	文字・表記
5	18	語彙・用語	13	2	国語教育
6	16	白紙台紙	14	2	国語学一般
7	14	文法	15	1	書評・紹介
8	11	追補	16	1	敬語・丁寧語

日英パラレルテキスト：人手による翻訳が存在している場合、それと原文とを並べあわせたデータベースを作成しておけば、原文を利用した文法的解析に加えて、翻訳を利用した意味的解析も同時にを行うことができる。本研究では、古今和歌集を題材に原文（国文学研究資料館開発）のデータと Laurel Rasplica Rodd 翻訳の英文を電子化したデータをあわせ持つ、パラレルテキストデータベースを開発した。設計思想の第一は、上記のように、文法的特徴の抽出だけでなく、意味的特徴の抽出の可能のこと、第二に、単なる検索・参照の道具としてではなく、計量的目的を可能にする構造を備えていることである。

- ・調査目的の語が五七五七七のどこにあるのか、容易に計算できること
- ・作者名など異なる記述（例えば、きのつらゆき／つらゆき／貫之／紀貫之など）が複数あっても、数えあ

げられること

- ・注記、校異、異本、枕詞、係結び、切れ字、掛詞、物名などの参照が容易なこと
- ・字あまりの句が容易に抽出できること
- ・撥音仮名表記／ム表記／ン表記／無表記の計量が容易にできること
- ・「いなおほせどり」(208=読み人知らず、307=忠岑) のような同一語句の発見が容易にできること (N-gram の実装)
- ・「かきつばた」のような折句の発見が容易にできること
- ・体言止め句の参照が容易にできること
- ・語彙量／延べ語数／異なり語数などの計算が容易にできること
- ・品詞ごとの計算が容易にできること
- ・2語以上の任意の語句が接続している場合の計算が容易にできること
- ・上記におけるそれぞれの計算や抽出が原文だけでなく翻訳文においても同様にできること
- ・ある単語の翻訳が容易に参照でき、その翻訳の異なりが容易に抽出できること（たとえば、「花」は234首で、その内，“flower”は46首，“blossom”は88首，“bloom”は43首。）

BBDB による和歌のパラレルテキストデータベース検索システムの実装を行った。検索の種類としては、原表記、読み表記、ローマ字表記、英語翻訳、作者標準表記による検索を実装した。検索システムは以下のファイルを対象に検索種別にアクセスする。

- genhyoki.db... 和歌原表記データ
- wakayomi.db... 和歌読み表記データ
- wakaroma.db... 和歌ローマ字表記データ
- wakaeng.db... 和歌英語翻訳データ
- wakatag.db... 品詞タグつき英語翻訳データ
- authors.db... 作者標準表記データ

また、okinmain.db には、古今集全ての情報を 1 和歌に対し、1ID をつけて管理したデータベースで、これを前処理にて、1 和歌1HTML ファイルを生成し、検索結果からハイパーリンクにより参照できるファイルを作成した。

それぞれの検索結果は検索文字列を赤で示し、原表記検索および読み表記検索、ローマ字表記検索の場合には英語翻訳を、英語翻訳検索の場合には原表記を、作者標準表記検索の場合には原表記および英語翻訳表

記をそれぞれ2段に配置して表示するようにした。

それぞれの作者名をクリックすると、それぞれの歌の情報一覧が別の.htmlのページに表示されるようにした。

頻度情報として、検索該当数を表示するだけでなく、原表記検索および読み表記検索の場合には、57577のいずれの句に検索語が出現したかを集計した。

英語翻訳検索の場合には、品詞による検索もできるように、品詞タグをつけた検索ファイルを用意した。このタグづけ作業には、Brill's Taggarを用いて行った。Brill氏自身のホームページのURLは、<http://www.cs.jhu.edu/~brill/>である。

## 和歌の検索

計算機を和歌検索に用いる際の利点としては、瞬時に条件にあう和歌を抽出すること、あるいはその頻度を正確に集計するなどがあげられる。

あるキーワードとその前後のコンテキストを調べ、キーワードとその共起関係にある語を抽出し、頻度を調べることで、作者論の検証を行うなどの作業が可能になる。

たとえば、歌とは直接的なつながりがない枕詞は、意味による関連や音による関連があり、通常は5音節からなる。つまり57577の5の句にくるはずである。はじめの5にくるか、なかの5にくるかの頻度調査は容易なはずである。具体的に「あしひきの」は一般的に、「やま」の枕詞であるが、古今集では、どういうことばの枕詞となっているのかなどは、データベースを検索することによって、検証することができる。

実際に「あしひきの」を検索してみると、読人不知(11)、貫之(4)、躬恒(1)、滋蔭(1)、友則(1)

の18件が検索され、57577分布は、／8／0／10／0／0／となり、第1句、第3句ということがわかる。また、古今集では18件のすべてが「あしひきの／やま」という使われ方であることが瞬時にわかる。

## 今後の課題

現在のメインデータベースファイルは、検索および計算を容易にするために独自の形式で入力作業をすすめたが、パラレルテキストとしての汎用性を考える上には国文学研究資料館形式へのコンバータを作り、日本語のファイル、英語のファイルのいずれの更新が行われてもよいようにすべきである。

現在の実装では、任意の語を検索する方式であるが、和歌を閲覧するだけでは、あまり問題にはされないが、頻度を計算するにはやや問題がある。たとえば、原表

記で検索した場合には、かなで書かれてあったり、漢字で書かれてあったり、校異が存在したりで、一意に集計できない。一方、読み表記で検索した場合、検索語のひらがなが1文字あるいは2文字などのように短い場合には望む語以外もヒットしている可能性が高い。たとえば、係の「ぞ」を調べる場合には、「そ」では「こそ」や「そめて」などにもヒットしてしまう。今回の開発では、ローマ字検索を実装し、“zo”（スペース「zo」スペース）を行って、計算することでこの問題を解決しているが、厳格に行う場合には、品詞情報をそえたデータが必要となろう。

つぎに既知の和歌の技巧、たとえば、掛詞、本歌取り、歌枕、見立て、物名など、詞書や注釈などに記述されている情報から、データすることは手作業にて行うことはできても、多変量解析を利用した未知の規則技巧の発見には困難が予想される。区切れや体言止めについても同様で、それぞの情報を併記したデータ、古語辞典を利用したデータの構築も今後の作業として必要になろう。

## 謝辞

今回の古今集パラレルテキストデータベース構築にあたっては、さまざまな研究者の支援を得た。国文学研究資料館の中村康生先生には国文学研究資料館データベース二十一代集の利用を快諾していただいたばかりでなく、CDROMを貸与してくださいました。また、国文学研究資料館の安永尚志先生には、パラレルテキストと国文学資料について御指導いただきました。青山学院大学の近藤泰弘先生計算機利用の和歌研究の可能性について、千葉大学の近藤みゆき先生には和歌研究作家論について、御指導いただきました。コロナ大学ボーラー校のローレルラスプリカロッド先生には、御著書、“Kokinshū”を寄贈していただいたばかりなく、本研究のためのデータベース化および公開について、快諾いただきました。ここより感謝いたします。

# Web公開のためのデータベース・エンジニアリング\*

及川 昭文, 山元 啓史

現在ホームページ上で、一般的な DBMS の類を利用することなく貝塚データベースを公開している。これは、データベース中の各レコードを 1 つの HTML ファイルとして格納し、検索や表示のための一群のプログラムを開発、実装することによって、従来の DBMS と同等の性能、機能を実現したものである。これまでのように使用する DBMS についての詳しい知識がなくても、ワープロ・ファイルやエクセル等の表形式ファイルを、簡単な手続きと手順で、Web 上にデータベースとして格納し、利用することができる仕組みを作り上げた。また、HTML の持っているリンク機能等を活用して、従来の DBMS では実現が困難であった機能も実現することができた。これらの一連の手法、手順、機能について貝塚データベースを例にして報告する。

The Kaizuka (Shell Mound) Database, which doesn't require DBMS, is now available to the public on the web. A user can retrieve and display any record from this database with almost the same performance and functions of an ordinary DBMS. We have designed these features so that each record arranges as one HTML file, and a series of tiny programs makes it possible to access simple tables for user search requests. Anyone with little knowledge about DBMS can upload files, which are recorded as a word-processor file or a spreadsheet file, to the web server as a unit of the database. We have also made external access, which is not possible under an ordinary DBMS, possible by using hyper-links which the HTML files contain. In this paper, a series of techniques, procedures, and features of the Kaizuka Database will be shown.

## はじめに

コンピュータが計算する機械としてだけではなく、情報を処理する機械へと進化するとともに現れてきた DBMS という概念が大きく変わろうとしている。それはインターネットに代表される一連のネットワーク技術と、html という情報処理言語の出現によって引き起こされつつある。DBMS の進化をたどれば、まずホストコンピュータ、やがてワークステーション、そしてパソコンと、そのハードウェア環境はダウンサイジングされてきており、同時に運用も集団から個人へ、専門家から非専門家へとよりユーザーフレンドリーなものへと変化しつつある。以前は専門的な知識と十分な経験がなければ、DBMS を使いこなせなかつたのが、ある程度の知識や技術があれば何とか Access

等に代表されるデータベース（以下、DB）ソフトを使えるような状況になってきている。

ただ、このある程度が問題である。筆者の勝手な憶測であるが、ワープロソフトにしろ DB ソフトにしろ、90%以上のユーザは、これらのソフトが有している機能の 20%も利用していないのではないかと思われる。すなわち、これらのソフトをフルに使いこなそうと思ったら、以前と同じように十分な知識と経験、そして技術を身につけなければ、それは叶わないというのが現実であろう。

初心者は初心者なりに使えるということは、確かに進歩といえるが、それはあくまで自分だけが利用する場合である。DB をサーバに格納し、一般に広く公開し、日常的な保守も行おうとしたら、DB ソフトだけでなく、サーバやネットワークについての十分な知識

や技術が必要になり、そのようなユーザはほとんどいなくなってしまう。自分の持っているDBを広く公開したくてもできないいるユーザでも、簡単な手続きと手順でWeb上にDBを公開できるBB-DB Engineeringを提案する。

## 1. BBDB Engineering ——その概念——

BBとはBare Boneの略で、simple resourceすなわち、できるだけ単純な形態を保ちながらDB化することである。また、ここでいうエンジニアリングとは、情報工学的なことだけを指すのではなく、1次資料からDBを作成し、Webで公開するに至る過程に必要な情報技術、技法（手法）、手順等を総称したものとして使っている。

### 1.1 何が問題なのか

多くの有用な学術コンテンツが存在しているにもかかわらず、なかなか研究室から発信されてこない大きな理由としては、3つのことが考えられる。

技術の問題：永らく紙と鉛筆の世界で過ごしてきた人文系研究者にとって、ワープロは何とか使うことができても、DBまでは手がないことがある。あるいは、何とかDB化はできても、それをWebで公開するだけの知識も技術もないし、そこまで学習する時間も意欲もないことがある。

人文系学問固有の問題：多くの人文系学問の場合、1次資料（史料）の収集、データシートの作成に膨大な時間と人手を要し、苦労して作ったDBをそう簡単に公開したくないという心情的な要素があり、なかなかDB公開に至らないという側面は否定できない。

質と量の問題：DBを公開する以上は、それに値するだけの質、量が備わっていなければならない。とくにデータのQuality Controlは、その利用価値を左右するものであり、十分な配慮が必要である。しかし、一般的なDBMSはQuality Controlについて無関心で、全ては作成者の責任になっている。この負担は研究者にとっては非常に大きいものがある。

これらの問題は、Web公開に至るまでの2つの障壁として存在することになる。すなわち、まず第一はデータをDB化できない。第二はDB化までは何となるが、自ら使うのみで公開はできない。この2つの壁を取り除くためには、情報技術の開発だけではなく、それらの技術を活用するためのエンジニアリングが必要である。

### 1.2 何が必要なのか

多くの研究者は研究に必要な資料やデータをカードにしたり、図表にしたり、あるいはファイルフォルダーにまとめたりして整理している。DB化とはこれらの資料やデータをコンピュータ上の仮想空間に移し替える作業に他ならない。言いかえれば、それはコンピュータにとって理解しやすい、処理しやすい、管理しやすい形態への変換作業である。このことは、その形態がどのようなものを熟知していないと、DB化に失敗する恐れが大きいことを意味しており、そのことがDB化を難しいものにしている。したがって、その変換作業が研究者が日常的に行っている資料やデータの整理作業の延長として行えれば、DB化は困難なものでなくなくなる。すなわち第一の壁は消滅する。

いまやDBの検索は、Webのページ上で行うのが主流となりつつあるが、そこでは多くの場合DBMSが使用されている。したがって、利用者はDBMSに習熟している必要はないが、提供する側にはそれらのDBMSを熟知している技術者を必要とする。一方、ホームページ作成は、さまざまなアプリケーションソフトやツールが開発されており、容易な作業となっている。したがって、ホームページ作成と同じような手順で、DBをサーバ上に構築し、DBMSと同じような検索や表示が可能なツール群が用意されれば、DBMSを知らないても、WebでのDB公開が比較的簡単に可能になる。第二の壁の消滅である。

### 1.3 BBDBとは

ここでDBMSを利用することのメリット、デメリットを論じる余裕はないが、BBDBはそれぞれのDBに対応した小規模なプログラム群から構成される各種ツールを用意し、DBMSを利用することなく同等の機能を実現しようとするものである。

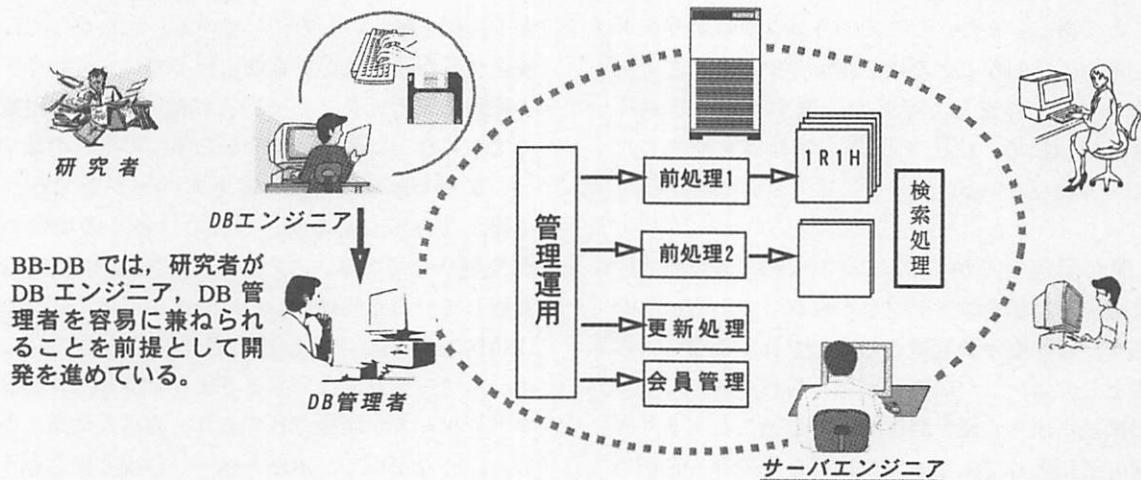
BBDBの最大の特徴は、1R1H、すなわち1 recordを1 HTML fileとしてサーバ上に格納するところにある。このことによって従来のDBMS以上の柔軟性と拡張性を実現している。例えば、ある特定のレコードに関して例外的な処理（貝塚DBの場合、図表や写真の項目は設定していないが、ある遺跡についてのみ、写真等を追加したいといった処理）を必要としても、その処理を行うページにリンクを張るだけでよい。そのリンクも、あるディレクトリに例外処理のページを格納しておけば、自動的に対象とする遺跡のページとリンクを張れるようなツールを貝塚DBでは用意した。

BBDBは、現在開発中のシステムであり、その詳細を体系的に示すことはできないが、大まかな概念は

図1のとおりである。ここではDB管理者、DBエンジニア、サーバエンジニアの3つの役割が必要になる。もちろん1人が全てを兼ねる場合もあるが、通常はDB管理者(=DBエンジニア)とサーバエンジニアという構成になる。そして、DB管理者(=DBエンジニア)が公開したいDBを持っている研究者ということになる。いったんシステムが確立されれば、その運用に必要なのはDB管理者(=研究者)のみとなる。運用は全てWeb上の運用管理ページから行えるよう

になっており、特別な処理や新しい機能の追加等を除けば、DBMSの素人である研究者でも容易に運用していくことが可能である。

このようなことが可能になっているのは、BBDBがsimple is best, small is beautifulという基本哲学のもとでシステム開発を進めているからである。すなわち、処理プログラムは可能な限り単純に、そして小さな単位で行うようになっており、保守が容易で拡張性に富んだものとなっている。



## 2. BBDB ——その適用例——

今回われわれの提唱するBBDBを貝塚DBに適用した。以下、サーバでの具体的な処理手順について述べる。なお、貝塚DBのURLは<http://aci.soken.ac.jp/>である。

### 2.1 サーバへの実装上の基本原則

ここではDBをサーバへ実装のための基本原則について解説する。実装のための処理は、大きくDB公開とシステム管理の2つの側面がある。DB公開には、前処理、CGIによるプログラム処理、システム管理には、更新システム、ユーザ管理システム、また、開発過程においては、DBの内容を担当するDBエンジニアリングとDBをサーバで公開するサーバエンジニアリングの2つの仕事が発生することになる。

前処理と場合分け：公開するデータは、可能な限りあらかじめhtmlに変換する。しかし、一つひとつのファイルを管理していたのでは、大変なので、これらは公開データとし、直接編集することはない。管理すべきファイルは、サーバにはおかれて、管理は個人パソコンの表ソフト等で管理更新を行う。サーバにはテキ

スト形式のファイルを転送し、そこからそれぞれのコードのhtmlファイルを生成する。これを前処理と呼ぶ。

前処理では、公開に際して予測されるユーザの行動を考慮して、静的部分、動的部分に大きくわけ、静的部分については前処理で.htmlを出力し、動的部分についてはcgiを用意する。例えば、県別インデックス(47県)のように数の固定されているものについては、静的処理として、前処理でken.htmlを生成する。また、頻繁に検索が実施されると考えられるデータについては、あらかじめ前処理でテーブルを抽出しておき、そのテーブルを対象に検索を行う。さらに、対象によって、将来的に更新される部分と更新されない部分でプログラムを書き分ける。

プログラム開発：前処理の実行は大きなCPU負荷を伴ったとしても、頻繁に行われる作業ではないので、できるだけシンプルに、それぞれのステップの入れ換えや書き換えが容易に行えるよう明示的に記述する。過度に高級な技術、あるいは処理のための処理といったメタ処理はできるだけ使わない。単純な技術の組合せを重視し、プログラムの可読性を高くし、プログラムはできるだけ小さい単位で作成する。また、運用中

の機能の追加、機能別に独立したテスト等が行えるよう、システム全体に影響を与えるようなプログラムの開発はしない。高速に、確実に、生のデータを直接アクセスすることで、管理効率のよいシステムを構築する。

ユーザインターフェース：クライアントのCPUやブラウザにパフォーマンスが依存しないように、サーバから転送するデータはhtmlを基本とし、転送に負荷となるグラフィックスやjavaはできるだけ使わない。ユーザがそのままhtmlファイルを保存した際、論文等に引用しやすいよう、それぞれ独立したレコードとして出力する。それぞれわかりやすいように、インデックス別、機能別にリンクを記述したトップページを用意する。

管理者インターフェース：管理者は、トップページのhtmlファイルの編集と利用者の管理、DB全体の更新を行う。トップページは一般的なindex.htmlであり、このトップページは、DB管理者が関連情報をリンクしたり、おしらせ、利用マニュアルを追加できるように、単純なhtmlで提供する。一方、利用者の管理、DB全体の更新は、別のディレクトリに置かれたページから行う。http標準実装の認証処理を行い、管理者以外はアクセスできない。このページには、管理作業別にリンクが施され、どういう手順で更新すればよいかは、プログラム実行中に明示され、作業に間違いのないよう確認メッセージを出す。

DBエンジニアとサーバエンジニア：DBエンジニアとサーバエンジニアの役割と目的について説明する。

DBエンジニアは、データシートのデザイン、データの入力、データの修正、データの表記一貫性の確保、またそのチェックプログラムの開発とデータチェックの実施、コンピュータで制約のある文字表現等の問題とその解決、原本と出典、1次データとの関連位置づけ、データシートの管理を行い、DBの目的、DB利用方法、利用規定の明確化、ユーザとの情報交換等の作業を行う。

サーバエンジニアとは、データシートのデザイン及び対象DBの特性をよく検討し、前処理の実施方法、検索の方法を検討する。また、プログラミングを行うだけでなく、データ出力のテストを行いつつ、対象つまり本稿における例では貝塚DBの内容を十分に考慮した、集計方法、可視化の方法等を検討し、実装する。その上で、ユーザはどのような検索を望んでいるのか、どのような情報を必要とするか、どの程度の頻度及び範囲で更新を行うか、追加情報はどのように取り扱うか、ユーザのデータ読み出しアクセスはどこまで許可

するのか、ユーザのデータ書き換えを認めるか等を、DBエンジニアとサーバエンジニアの間であらかじめ話し合い、動的データ、静的データ、検索の範囲、データ更新の方法を決定する。さらに、品質を向上させるためには、利用者を増やすためにはどのようにすればよいかも検討し、ワークショップや研究発表等の計画も行う。

## 2.2 プログラムとツール群

前処理ツール：DBエンジニアからサーバエンジニアに引き渡されるファイルは表1にある4種で、それぞれのオリジナルは、DBエンジニアのコンピュータで管理され、サーバにはコピーが置かれ、それをもとに前処理を行う。したがって、DBMSでは、それがクラックされた場合には、オリジナルファイルが危機に曝されるが、本システムにおいてはその心配はない。

表2の前処理プログラムにより、それぞれのファイルを生成するが、これら生成されたファイルは公開形式であって、最も重要なオリジナルではない。これらの更新プログラムは管理者ディレクトリに置かれ、管理者ページからWEBブラウザを通して、実施される。どれがどの作業を行うかは、ページに記述してあるので、使用方法を迷うことはない。

表1 前処理用ファイル

ファイル名	数	サイズ	内 容
kai.db	5678	1.4MB	貝塚メインファイル
bib.db	2664	0.7MB	文献メインファイル
table.euc	146	2.6KB	時代遺跡遺構コードテーブル
kencode.tab	47	0.6KB	県コードテーブル

表2 前処理用プログラム

プログラム名	行数	サイズ	内 容
makekai.pl	200	kai.db	遺跡レコードファイル KZxxxxxx.html
makebib.pl	104	bib.db	文献レコードファイル KBxxxxxx.html
makeiko.pl	146	kai.db	遺構別インデックスファイル iko.html
makeiseki.pl	144	kai.db	遺跡別インデックスファイル iseki.html
makejidai.pl	143	kai.db	時代別インデックスファイル jidai.html
makeken.pl	134	kai.db	県別インデックスファイル ken.html
search_format.pl	119	kai.db	遺物検索テーブル search_item.tab.html

実行処理プログラム：ユーザが利用するCGIは、遺物検索のみである。貝塚DBでは、該当遺物の一覧は「完全一致」「部分一致」「かつ」「あるいは」の組合せにより、固定ではないため、その一覧生成にCGIを用いた。動的出力なので同時に該当件数の総数計算、県別集計、県別頻度色分け地図の出力を実装した。そ

のため、414ステップスと他のプログラムよりも少々長いが、それでも9KB程度である。

- ・kz\_search\_f.pl:414 (検索プログラム：検索し、結果をhtmlで出力する。出力には、各レコード.htmlへのリンクが出力される。また、外部プログラムmapをアクセスし、遺物頻度地図を出力する。)

公開アクセス管理ツール：利用は、検索、遺跡別一覧、遺構別一覧、県別一覧は、ユーザ制限なしでアクセスできるようにし、それぞれの遺跡レコードあるいは文献レコードのアクセスにはユーザ制限をもうけた。ユーザのアクセス制限はhttp標準実装の認証を利用した。これらレコードにアクセスするためには、会員申請をする。ユーザが会員申請の受付フォームを使って会員申請を行うと、フォームの内容をregist.datにアpendする。

また、同じ内容をDB管理者に、メールで自動的に送り、会員申込があったことを知らせる。データ更新・訂正報告フォームは、DBを利用している会員がデータの訂正を見つけた場合、それを報告するフォームで、同様に報告が行われると、その内容は、DB管理者にメールで報告される。この場合、公開DBのどこにも報告の内容は記録されない。会員アクセス許可あるいは削除、会員名簿更新ツールはDB管理者のみが使えるプログラムで、会員アクセス許可プログラムは、許可実行とともに自動的にIDとパスワードが送られる。プログラムと生成ファイルは表3、4のようになる。

導入及び更新手順：サーバが立ち上がったら次の手順でDBの導入を行う。

#### 1. 基本データの転送

(kai.db,bib.db,ken.tab,table.euc)

#### 2. 前処理の実施

- ・貝塚アップデータ（前処理／kai\_update.pl:115）
- ・索引アップデータ（前処理／idx\_update.pl:143）
- ・文献データアップデータ

（前処理／bib\_update.pl:105）

#### 3. 検索システムの導入 (kzsearch\_f.pl:414)

#### 4. 外部プログラムの導入 (jgrep,map)

#### 5. アクセスシステムの導入

#### 6. テスト

### 2.3 検索実行速度のテスト

DBMSを利用するとバイナリツリーあるいはハッシュ表等の検索テーブルを作成し、検索実行速度をあげる仕組みがあり、これを導入すると確実に飛躍的な改善が可能となる。しかしながら、その分プログラムの構造は複雑になり、処理のための処理、デバック、

追加機能の実装にコストがかかる。そこで、本研究におけるシステムでは、どの程度の時間がかかっているかを計算し、プレーンテキストの検索が実際的であるかどうかを検討した。

具体的には、登録されている5678(1382KB)レコードに対し、遺物とレコードIDだけを抽出したファイル、search\_item.tabを前処理として作成した。これは4424(839KB)レコードになった。この4424レコードを対象に検索した場合のCPU時間を計算する。検索に利用したプログラムはjgrep(GNUgrep version2.0+multi-byte extension 1.04)で、CGIで外部呼び出しをしているプログラムと同じプログラムである。計測に利用したプログラムは、Gnutime version 1.7で、実施したCPUは、Intel-MMX133MHz.OSは、Linux kernel2.0.34。総メモリは96MBでいずれの実行環境においても、Swapは検出されなかった。出力は結果のテキスト量に依存するので、出力はいずれも/dev/nullに送った。

表3 会員管理プログラム

プログラム	行数	内 容
kzregist.pl	388	会員申請受付フォーム
kzmember.pl	140	会員名簿公開プログラム
kztouroku.pl	329	会員アクセス許可ツール
kz_delete.pl	241	会員アクセス削除ツール
kz_correct.pl	219	会員名簿更新ツール
kzrepo.pl	313	データ更新・訂正報告フォーム ニュース・メイン転送システム

表4 会員管理用ファイル

ファイル名	内 容
htaccess	アクセスできるファイル、ディレクトリ、グループファイル、パスワードファイルのパスを記述。
group	貝塚グループでこのファイルに会員IDが記述されてはじめて、レコードにアクセスできる。
passwd	各会員のencryptパスワードが格納されている。
regist.dat	会員申請受付フォームから送られてきた内容がアpendされるファイル。氏名、所属、電子メールアドレス、連絡先、自己紹介などが収集される。
id 2 email.dat	会員IDと電子メールの関係テーブル

それぞれ50回試行し、ユーザCPU時間、システムCPU時間、実CPU時間の3種類の平均時間計算（単位は秒）を求めた。また、該当件数、部分一致、完全一致、「かつ、あるいは」等の条件によって変動が認められるか、サイズ10倍の疑似ファイルの場合はついで検討した。表5はその計測結果の一覧である。

「あるいは」以外いずれの条件においても、実時間は、0.12-0.18秒程度であった。「あるいは」はjgrep

を2回呼び出しているので、やや多いが、それでも、0.2秒程度である。search\_item.tabを10回コピーしたファイルを作成し、それに対し、CPU時間を計算したところ、単純にCPU時間は10倍となった。

search\_item.tabを10回コピーしたファイルにおける結果でわかるように、5万件のデータを仮定した場合、それに要する時間は1.2秒程度である。この速度は、htmlのデータをクライアントのリクエストにより、通信転送する時間を考慮した場合、この検索時間よりページの転送時間の方が大きくなるものと思われる。133MHzのCPU速度でこの程度であるので、最近の1.3GHzとなると、さらに問題とならない速さといえよう。現在の貝塚DBは、5千件のデータであるが、現在作成中の遺跡DBでは、10万件に及ぶものとされている。その場合でも、CPU時間はかなり小さいものといえよう。

表5 会員管理プログラム

*U=user, S=system, R=real*

部分一致「シジミ」該当件数 1877

<i>U</i>	Min = 0.100	Mean = 0.136	Mac = 0.170	SD = 0.015
<i>S</i>	Min = 0.020	Mean = 0.044	Mac = 0.080	SD = 0.015
<i>R</i>	Min = 0.170	Mean = 0.173	Mac = 0.190	SD = 0.006

完全一致「シジミ」該当件数 143

<i>U</i>	Min = 0.040	Mean = 0.069	Mac = 0.090	SD = 0.012
<i>S</i>	Min = 0.000	Mean = 0.023	Mac = 0.070	SD = 0.013
<i>R</i>	Min = 0.080	Mean = 0.095	Mac = 0.280	SD = 0.028

完全一致「ウシ」該当件数 236

<i>U</i>	Min = 0.070	Mean = 0.093	Mac = 0.110	SD = 0.012
<i>S</i>	Min = 0.010	Mean = 0.029	Mac = 0.050	SD = 0.012
<i>R</i>	Min = 0.120	Mean = 0.122	Mac = 0.180	SD = 0.010

部分一致「ウシ」該当件数 4240

(search\_item.tab の10倍ファイル)

<i>U</i>	Min = 0.900	Mean = 0.977	Mac = 1.080	SD = 0.042
<i>S</i>	Min = 0.150	Mean = 0.248	Mac = 0.330	SD = 0.042
<i>R</i>	Min = 1.210	Mean = 1.223	Mac = 1.280	SD = 0.011

完全一致「ウシ」かつ「ウマ」該当件数 149

<i>U</i>	Min = 0.060	Mean = 0.096	Mac = 0.130	SD = 0.016
<i>S</i>	Min = 0.000	Mean = 0.030	Mac = 0.060	SD = 0.014
<i>R</i>	Min = 0.170	Mean = 0.178	Mac = 0.250	SD = 0.011

完全一致「ウシ」あるいは「ウマ」該当件数 587

<i>U</i>	Min = 0.140	Mean = 0.180	Mac = 0.210	SD = 0.016
<i>S</i>	Min = 0.010	Mean = 0.041	Mac = 0.080	SD = 0.017
<i>R</i>	Min = 0.210	Mean = 0.219	Mac = 0.230	SD = 0.005

## 2.4 BBDBの利点

ここでは、DBのレコード表現にhtmlを採用した利点、外部プログラムとの組合せによる利点、通信プロトコルとしてhttpを用いた利点について述べる。

htmlの利点：まず、出力形式をhtmlとすることで、クライアントの画面サイズを気にすることなく、レコード出力を行うことができる。また、前処理出力時に、外部htmlの参照を指定することで、レコードに外部の情報を附加することもできる。それらはとくにテキストである必要はなく、リンクに用いられるものであれば、何でもよい。実際に関連情報として、PhotoCDのファイルをリンクしているページもある。それぞれのレコードはhtmlファイルとして独立しており、外部のサーバ上にある、htmlからも参照することができる。

静的なhtmlではなく、DBMSによってその都度出力するシステムではこういうことはできない。さらに、他のサーバから参照してもらいたいときには、タグを使い、サーチエンジンロボットに明示的に覚えさせれば、外部検索エンジンからも利用できる。

外部プログラムとの組合せによる利点：まず、UNIXにインストールされている、あるいはインストール可能なプログラムとの連携が簡単にとれることがあげられよう。例えば、貝塚DBでは、貝塚の遺物検索結果の県別頻度表示の可視化手法としてMapOfJapan<sup>(1)</sup>のライブラリ一式を利用している。また、http標準実装の認証を利用したユーザ管理システムを保有しており、同時にユーザ管理ファイルの電子メールアドレスを使って、sendmailによりMLを構成したり、ニュースを一括送信したりすることができる。これらは、特にUNIXの技術としては新しいものでも特別なものでもないが、一般的なDBMSだけを使っていたのではこのようなことはできない。いずれにしても、DBMSだけを操作するだけでは、インターネットの利点を十分にいかすことはできない。

httpの利点：まず、プロトコルが広く使われているhttpであり、レンタルサーバを利用する場合でも、それがどのようなものであれ、httpであるはずなので、DBは必ず公開できる。httpはファイル名をアクセスするプロトコルであり、レコードのアクセススピードは即ファイルシステムのファイルアクセススピードで

(1) MapOfJapan (<http://aoiki2.si.gunma-u.ac.jp/PseudoFTP/UNIX/MapOfJapan/>) は青木繁伸氏（群馬大学社会情報学部）作成の日本地図描画システムである。Cプログラムソースコード及びデータ表現のマニュアルが添付されているので、データの塗り分けの方法、矩形の表現方法等の変更が可能である。一番興味深いのは、プログラムもさることながら、作者が地図データの更新を懸命に行っていることである。作者の努力に対して感謝します。

ある。つまり、DBMS 起動による CPU 負荷、そのための余計なメモリ負荷は皆無ということになる。また、アクセス許可はディレクトリあるいはファイル毎で可能で、それに用いられる認証システムは、http 標準のものが用いられるので、あらたに認証システムを用意する必要はない。さらに、http 標準のアクセスログファイルにより、ユーザがどのファイルを何回アクセスしたかを知ることができる。現在の貝塚 DB では、検索キーワードのログの集計も実装されている。

### 3. 課題と将来構想

すでに述べたように、BBDB はまだ開発途上のシステムであり、今後検討していかなければならない課題も少なくない。具体的にいくつか挙げてみる。

**更新頻度の高い DB**：貝塚 DB は、新規レコードの一括追加や誤り訂正以外には、日常的な更新はほとんどない。学術研究用 DB には、この種のものが多いが、一方、頻繁な更新を必要とする DB も少なくない。このような DB に対して BBDB が対応可能かどうかについては、まだ十分に検討していない。今後具体的な DB を選んで、必要なツール等の開発を進めていくことを予定している。

**Quality Control**：DB の内容に関する品質管理は、ほとんどの場合作成者に任せられており、実際の DB 構築においてその負担は大きいものがある。コンピュータを使ってどのような品質管理が可能かは、対象とする DB に依存する場合が多いが、標準的な品質管理のためのツール群を開発することは十分に可能であると考えている。

**マニュアル整備とツール群の体系化**：現在 BBDB を適用しているのは、貝塚 DB と小松左京コーパス (SF 作家小松左京氏の全作品のフルテキスト DB) の 2 つである。後者の URL は <http://castelj.soken.ac.jp/groups/komatsu> で、すでに一般に公開している。今後他の DB についても適用していく予定で、そこにおいても多くのツールが開発されると想定される。これらのツール群の体系化と、BBDB を利用したい研究者向けのマニュアル作成が緊急の課題としてある。

**CCDB 構想**：現在その実現を目指して検討を進めているのが、CCDB 構想である。CCDB とは Collaborative Creation of Database のことである。これまでの DB 作りは、1 人または少人数の研究者がこつこつと地味な作業を続けるというのが一般的であったが、これを 100 人あるいは 1000 人が Web 上での共同作業を通じて、大規模 DB を構築するという試みである。その

ためには体系的な作業手順、参加者をマネジメントするための仕組み、データの品質管理等についての研究開発が必要である。

### おわりに

BBDB は、まだ生まれたばかりのシステムである。今後どのように発展していくのか、あるいは意図したような汎用性、拡張性、柔軟性を保持できるのか、未確定な要素が多くある。今しばらく開発を続け、CCDB を含めた体系化ができた時点で、改めて報告をしたい。

最後になるが、貝塚 DB について簡単に説明する。これは日本全国で発見されている遺跡のうち、動物遺存体を出土している遺跡を集め DB 化したもので、レコード数は約 5,600 である。この DB 作成に協力してくれた多くの方々に深く感謝する。

#### \* 利用刊行物：

- 情報処理学会研究報告「Web 公開のためのデータベース・エンジニアリング」(及川昭文、山元啓史) (2001-CH-49) より転載

# ACI-japan -It's Perspectives for the Future-

AKIFUMI OIKAWA †

Even in the field of humanities, some research methods using databases has become very popular. It, however, is still necessary for researchers to have some specific knowledge and skills of database when creating and handling it. ACI-Japan project is designed to improve this situation. ACI means "Academic Contents Initiative" and the project will provide software and hardware for researchers who want to create database by themselves. In this paper, we discuss the problem of database creation, the philosophy of database and the database supporting project mentioned above in detail.

## *Databases in the Academic World*

IT (Information Technology) has brought many changes in various ways, in not only our daily life, but also the academic world. Personal Computers have become the main tools for our academic activities, and they may replace the traditional tools such as "paper and pencils." In every academic field, various data are being created everyday in electric media and this means that many databases are opened for the academic researches.

Many research institutions have set up and proceeded with database projects vigorously and systematically with a big budget. The following are some examples: a) Large-scale corpus of classical Japanese literary works by The National Institute of Japanese Literature; b) Database of the museum's large collection by National Museum of Ethnology; c) Electronic publishing of various information in Japanology by International Research Center for Japanese Studies; and d) Database about Japanese history and archaeology by National Museum of Japanese History.

The Ministry of Education and other ministries have the policy of encouraging database projects by giving them grants. Because of this policy, large-scale databases have been steadily developed. Along with high popularity of the Internet, the importance of contents has been greatly acknowledged and useful contents in various academic fields have begun to appear.

It, however, is still difficult for individual researchers to start developing their own databases. Once we begin to work with data such as word-processor files, spreadsheet data files, also sometimes yet in the shape of some paper, etc., we can find soon that the office is filled with a large amount of data.

Many portions of such precious data can be used more effectively and repeatedly, if they are compiled as a database. Many researchers, nevertheless, feel awkward and think that it is very difficult to create databases of their own data, and very few of them try such a task in fact.

## *Issues of Database Creation*

Apart from budget-related problems, there may be two major issues: one is researchers' skill, and the other is quality and quantity of data.

---

† The Graduate University for Advanced Studies

### (1) Researchers' skill:

Generally researchers in the humanities, who have long used only paper and pencils, become able to use PCs, but may not manage well to build databases by themselves. Even if they have developed databases, it seems to be difficult to make those databases widely available via WWW. Because they may not have enough knowledge/skills, besides no time to deal with their database and no motivation to learn how to realize it.

### (2) Quality and quantity of data:

Useful databases have to contain high standards, in other words, quality and quantity of data. Data should be controlled very carefully, because low-quality databases, for example, incompatibility of data, lack of consistency, typos, misinterpretation, etc., are fatal defects, and have very little value. Unfortunately, DBMS generally has not so much consideration for these quality control issues. These are considered to be the responsibility of each database owner or creator, despite the quality control management improve database's value greatly.

As a natural consequence, many data which have been collected with great efforts are left as raw data, not organized as database. We have some experiences that we met many valuable data that were accumulated by professors and researchers, and those accumulated data have gone somewhere at the moment of the their retirement. That is so regrettable and also would be a great loss for the future study.

### *ACI-japan Project*

ACI-Japan is a project to find ways to solve these problems. Its objectives are 1) to support individual researchers to create databases of useful academic contents; and 2) to assist them to spread their databases widely though Internet.

Although this project did not start with research grants, it has been mainly supported and maintained by volunteers of each study area. Basically, as our policy, researchers who want to join the project are responsible to make their own databases and to open them on the Web in order that other researchers can use them freely. All ACI-Japan can do is to offer the opportunities, that means, to provide hardware and software to support the database development and their activities.

So far, for examples of software support, ACI-Japan has two software packages: BBDB (Bare Bone Database) and BBQC (Bare Bone Quality Control), both of which were solely developed in this project.

BBDB has been developed for researchers who can use merely word processors, and enables them to build their own databases easily and to open the databases into the Internet world painlessly.

BBQC is for database quality control. As the name "Bare Bone" suggests, both packages of software doesn't have any bells and whistles and have been designed as simple as possible for novice users because we consider that the easiness of database operation is important.

ACI-Japan is now developing new software called "BBCC (Bare Bone Collaborative Creation of Database Contents)." This program will give support to gigantic database development in collaboration with several people, or maybe, dozens of people. It also have some potential of collaborative database creations beyond space-time.

ACI-Japan has just started. It consists of now only 10 members and developed 8 databases. We are now actively expanding this project, but cannot foresee how it develops in the future. It is our responsibility to pass down intellectual property to future generations. I hope ACI-Japan will contribute to this important assignment.

---

ACI-japan URL: <http://aci.soken.ac.jp>

Databases in ACI-japan

- Shellmound in Japan
- Bibliography for Shellmound in Japan
- Sakyo Komatsu Corpus
- Kokin Database
- Illustrations for Japanese
- Illustrated Sentence for Japanese
- Sentence & Vocabulary Test for Japanese
- High Resolution Molecular Spectroscopy

## 古今和歌集パラレルデータベースと公開システム\*

山元 啓史 †      及川 昭文 ‡

† カリフォルニア大学サンディエゴ校

‡ 総合研究大学院大学

### 要旨

数理解析を目的として、古今和歌集（以下、古今集）データベースとそれらを編集、公開、集計するためのマネージメントシステムを開発した。データベースは標準表記に従う1,111の和歌データのほか、英翻訳データベース、作者データベース、品詞解析データも作成し、それぞれ並行に登録し、検索、計算できるようにした。データベースマネージメントシステムは、BBDBというデータベース公開システムとBBQCという品質管理システムで構成される。このシステムを利用して、上記古今集DBの要素を構築し、公開、検索だけでなく、著者毎、和歌毎、品詞毎、等の集計が行えるようにした。本論文では、古今集の数理解析の予備段階として、1) データベースの構築方法、2) マネージメントシステムの設計と仕様、および、3) データベースの利用の実際、について述べる。

## Kokin Waka Shū Pararell Database and Management Systems

Hirofumi Yamamoto†      Akifumi Oikawa‡

†University of California, San Diego

‡Graduate University for Advanced Studies

### Abstract

We have developed the Kokin Waka Shū database (Kokinshū), the database of the collection of Japanese classic poems by Imperial order, and database management system in order to analyze Japanese classic poems. The database contains not only the general information of 1,111 Japanese classic poems in the original, but also the translations in English, and the parts of speech of each word in both Japanese and English. The database management system consists of two components: a database publishing system called "Bare Bone Database (BBDB)" and a database quality control system called "Bare Bone Quality Control (BBQC)." Using this management system, all the elements of the Kokinshū DB have been combined systematically, and users can not only search the information they want, but also calculate the number of authors, poems, words by the parts of speech, and so forth.

In this paper, a preliminary report on the Kokinshū project describes 1) The process of building a database, including the selection of categories; 2) The design and the development of management systems; and 3) The use of pararell Kokinshū database.

\*本研究は、文部科学省科学研究費補助金特定領域研究(A)118「古典学の再構築」の助成を得た。

# 1 はじめに

本論の目的は、古今集和歌集の英語翻訳を利用した日英パラレルデータベースの開発とそのマネージメントシステムの開発について報告することである。

前者であるが、このような古典文学の日英対訳のデータベースは、一般的に海外の日本文学研究者あるいは海外で行われている日本研究とりわけ、和歌の教育に役立つものと思われる<sup>1</sup>が、筆者らは特に、古今集という共通の材料を通して、外国人のもつ日本文化観が日本人のそれとどのように違うか、比較分析することを第一の目的とし、その基礎資料としてのデータベース開発を進めている。

ところが、データベースの開発を進める過程で古今集収録の和歌だけでなく、技巧のデータ、品詞データ、作者データ、複数からなる翻訳データ、統計分析のための前処理データなどが発生し、それらをまとめて管理する仕組みが必要となった。

そこで、後者の目的であるが、管理だけでなく単独あるいは複数を同時に公開、検索、計算するためのシステムの設計と開発を行った。設計には、筆者らが使うだけでなく、さまざまな、場合によっては、情報処理の技能・知識を必ずしも有していないユーザによる利用やデータベースの共同開発の可能性も考慮した。

以下において、コンテンツ開発として古今集データベースの開発、システム開発としてマネージメントシステムの開発を述べた上で、システムの利用の概略を説明する。

## 2 古今集データベースの開発

和歌を中心とする古典文学領域のデータベース化はかなり活発であり[1, 2]、国内だけでなく、海外においても、日本研究に焦点をあてたサイト<sup>2</sup>が活動を行っている。また、データベースが整備されると同時に、数理的手法による研究が増え、より客観的視点に基づいた議論が行われるようになってきた[3, 4, 5, 6]。

筆者らも数理的手法のための基礎とするデータ

<sup>1</sup> 日本研究・日本文学に関する学会が国際会議として海外で定期的に開かれるようになった。たとえば、EAJS (European Association for Japanese Studies) の国際会議は 1976 年以来 3 年に 1 度開催されている。中でも 2000 年フィンランドでの開催には、33 カ国から約 400 名の研究者が参加し、何とそのうち半数以上の約 270 名が外国人研究者であった。日本研究の研究者は日本人とは限らない時代となっていることがわかる。日本研究が論文という形式だけでなく、これら外国人研究者による翻訳データベースという形式も広く流通する可能性もある。

<sup>2</sup> たとえば、Japanese Text Initiative : バージニア大学エレクトロニック・テキスト・センターとピッツバーグ大学東アジア図書館が共同で進めているプロジェクトで、日本古典文学の電子テキストが WWW で利用できるようになっている。

ベース開発を進めてきたが、研究を進めるうちに、原典や翻訳のデータベースはもとより、歌の修辞技巧データ（たとえば、序詞、掛詞のある歌番号とその種類、分類、係結びの歌番号と係結びの種類、縁語）や注釈のデータや地名、人名のデータ、品詞タグ付データ、数理計算をするために行った前処理データ、分析結果をまとめたデータなど、多変量解析に利用するさまざまなデータが発生した。しかしながら、これらのデータは、個人的に数値処理を行って結果を出し、論文を書くだけでなく、データベースとして管理、公開しておけば、他の領域を同じくする研究者にとっても有用なはずである。

そこで、1) 原典データの加工、既存の資料の電子化【基礎】、2) 原典の情報をさまざまな視点からサポートあるいは詳細化するデータの開発【詳細】、3) 数値処理が行える形式への前処理【前処理】を軸としてデータベース化を進めた。

具体的には、1) 国文学研究資料館開発のデータベース[2]を基礎データとして利用<sup>3</sup>、2) 英翻訳本[7, 8, 9]を電子化し、翻訳データを作成、3) 品詞タグつきデータ、4) 修辞技巧タグつきデータ、5) 近藤みゆき氏提供のジェンダーデータ<sup>4</sup>を追加し、後述するシステムで WEB 公開できるようにした。以下では、そのうち翻訳データ、品詞タグつきデータの開発について述べる。

### 2.1 翻訳データ

素材：英翻訳データは 3 種類電子化し、データベース化した。まず、はじめに許諾を得、Rodd[7] の翻訳を電子化し、これを公開用データとすることとした。つぎに、McCullough[8] の翻訳も翻訳者の違いが結果に及ぼす影響を検討するため、同様に電子化し、タグ付を行った（非公開）。最後に、日本人による英語翻訳が上記 2 翻訳とどのように異なるかを比較するために、Honda[9] についても同様の作業を行った。

構造：表 1 に示すようにデータは 1 首 1 レコードとし、ID を与えた。Rodd[7]、McCullough[8] はできるだけ 5 7 5 7 7 に対応するよう翻訳してあるが、必ずしも日本語の各行とは一致してはいない。Honda[9] は、まったく 5 7 5 7 7 を意識せず翻訳している。3 翻訳とも原文通りに入力し、改行のある箇所をスラッシュで示した。ただし、Honda[9] は古語英語を隨所に用いており、別フィールドを用意し、現代英語表記に改めたデータも追加した。最後に翻訳中の英単語に品詞タグをつけ、英翻訳データ

<sup>3</sup> 同研究プロジェクトによる標準表記をはじめ、おおむね、異本、校訂に関する取り扱いは同じである。

<sup>4</sup> 男性歌人、女性歌人、読人不知のタグ付データ

ベースでの検索でも品詞検索が行えるようにした。またこれにより、日本語にはない事物の単数形複数形の異なりを計算することができるようになった。ただし、現時点では各単語は基底形（動詞なら原型、名詞なら単数形）へ変換されていないので、意味上の語彙の頻度集計は行えない。これは今後の課題としてデータを追加する。

表 1 Rodd の英翻訳のデータ：古今和歌集卷一春哥上一ドル記号+アルファベット1文字+縦棒をタグとしている。\$A| は通し番号で体系本番号に相当、\$B| は作者名、\$C| は前文、\$D| は和歌、\$E| は品詞タグデータ（以下の略語解説はここにあるもののみ掲載）

---

```
$A|000001
$B|Ariwara no Motokata
$C|Written when the first day of spring came
within the old year.
$D|spring is here before /
year's end when New Year's Day has /
not yet come around /
what should we call it is it /
still last year or is it this
$E|spring/NNS is/VBZ here/RB before/IN
year's/JJ end/NNS when/WH New/JJ
Year's/JJ Day/NNS has/VBZ not/NEG
yet/RB come/VB around/IN what/HW
should/MD we/PRP call/VB it/PRS is/VBZ
it/PRS still/RB last/JJ year/NNS or/CC
is/VBZ it/PRS this/DTG
```

---

NNS=普通名詞単数形、NNP=普通名詞複数形、VB=動詞原形、VBZ=動詞3人称現在単数形、RB=副詞、IN=前置詞、JJ=形容詞、WH=wh語、NEG=否定語、MD=助動詞、PRS=代名詞単数、PRP=代名詞複数、CC=等位接続詞、DTG=決定詞/代名詞

## 2.2 品詞タグつきデータ

**単語分割の意義：**古典文学を計算機で分析する研究には、近藤(n-gram)[5]、竹田(LCS)[6]のように計算アルゴリズムによって、数値化し、単語分割作業を必要としないものと、宮島ら[10]、村上ら[11]、村田ら[12]のように語彙の計量を目的とし、そのため単語分割作業を前提とするものに大きく分かれる。

前者のような単語分割作業を必要としない方法論は単語分割や品詞づけの作業に労力を要すること、専門家間でも品詞の解釈<sup>5</sup>や文法の取り扱いに基本

<sup>5</sup> 「はるくれば かりかへる なり(古今 30)」の伝聞推定の「な

的な考え方の違いがあること<sup>6</sup>、複合語の認定については揺れがあること<sup>7</sup>、などの問題点を解決するものとして注目される。

一方、後者のように語彙の計量を研究目的とする場合には、単位切り、単位認定は避けられぬ作業となる。また、作家の特徴を抽出しようとしたとき、なるべく文章の内容と関連性の薄い要素を用いた方がよい（金[15]）ことが報告されている。このような場合には、作品内容に依存した意味的特徴よりも、文法のような内容に依存しない特徴抽出を用いるほうが妥当であり、そのためには単語を品詞ごとに分けて、統計処理を実施したほうがよいと思われる。

長期的視点という要素はあるが、データベースのための単語分割あるいは品詞データ作成作業を行う意義について、次のように考えた。

1. 確かに労力はかかるが、一度データベースを作成すればずっと使える【苦の解決】
2. 品質管理を丁寧に行えば、いずれ実用に耐えられるようになる【質の解決】
3. 古典作品は今後ともどんどん増え続けるデータとは異なり、作業は有限。共同開発すれば、作業時間も短縮できる。【量の解決】

本研究の目的は、和歌に含まれる特徴や文化を抽出し、それが異なる文化を背景とする研究者のそれとの差について数量解析するものである。したがって、筆者らは、宮島ら[10]の分割単位にならい、語彙の計量ができるよう、各和歌を単位分割し、品詞タグをつける作業を開始した。

**方法：**古語辞書を作成し、その辞書を使って置換を行うプログラムを作成した。これで作業のあらかたを行い、細かな修正は人手によって行った。ただし、プログラムは文脈ルール、いわゆる文法解析は行わず単純に最長一致文字列置換のみの機能とした<sup>8</sup>。

具体的な手順は次のようになった。まず、宮島ら[10]のデータおよびATOK8/9用古文入力変換率向上支援単語ファイル<sup>9</sup>を加工して置き換え用古文單

り」をめぐって、「ラ変型活用語には連体形に、四段動詞には終止形につく[13](p.24)」とあるが、一方で「かりかへるなり..「なり」はすべて連体形につく、終止形につくというのは誤解[14](p.88)」とも。

<sup>6</sup> 岩波古語辞典では(1)動詞は終止形見出しではなく、連用形見出しとし、(2)形容動詞は認めない方針が貫かれている。したがって、(古今 562)の「けに」は岩波では形容動詞ではなく副詞であり、宮島ら[10]では形容動詞を品詞として認めているため「け、異、形動」となっている。

<sup>7</sup> 近藤[5]は「一語をどう認定するかは、その基準の立て方にも様々な立場があり、從来から多くの研究がなされてきた。そもそも単語をめぐる基準からして、一通りではない」と述べ、複合語処理の難しさを説明している。

<sup>8</sup> 一般的に語および活用形の文法を用いると多くの情報を利用することになり、変換効率がよくなるように考えられるが、古文法はいつも同じではなく時代によって異なる。

<sup>9</sup> Ver. 1.1 太田瑞穂氏作、ワープロ古文入力のための辞書。動詞・助動詞・形容詞の活用形などのエントリを持つ。一般的に、語彙計量では動詞活用形のそれぞれを集計したり、助動詞

語辞書を作成した。岩波体系本表記のデータを作成し、最長文字列一致で、単語[品詞-詳細]形式のデータに置き換えられるプログラムを作成した。学習参考書を利用し、動詞助動詞など頻繁に用いられる述部パターンを登録した<sup>10</sup>。

以上、機械的手続きによる大雑把な作業で、以降は手作業で、品詞タグのついていない箇所のタグづけ作業、係結びタグの付与、読み、活用形の誤りの修正、品詞とは関係ないが、序詞、掛詞のフィールドを追加し、その記述を行った。

その際、辞典および各種文法解説書[13]を利用した。また、諸説の見解ができるだけ考慮するため、片桐[16]、久曾神[14]を参考にした。片桐[16]の現代語訳はその解説にもあるように、できるだけ語を省略せず、各語を忠実に翻訳していることから、単語や品詞の認定、結びの省略など、明示的でない要素の記述に役立った。

さらに、著作権上公開はできないが、品詞詳細タグを付すために参考とした現代語訳として、久曾神[14](全部)および片桐[16](一部)の現代語訳の入力も行った。品詞および活用タグは表2のように略号で記述した。

表2 品詞のデータ化：現代語訳と一部に序詞および掛詞が作成作業を通して追加された。Aは歌番号、Bは品詞データ、品詞データ中の字句は全てひらがな。括弧内は品詞が略号で示され、ハイフンで活用形、漢字表記、助動詞・助詞の用法、種類など詳細が記述されている。Cは久曾神[14]現代語、Dは片桐[16]現代語、Eは修辞技巧。データ中の改行は任意。

\$A|000113  
\$B|はな [名] の [格助]、いろ [名] は [係助] /  
うつり [ラ四-用] に [完-用] けり [詠-終]  
な [終助-感] /いたづらに [形動ナリ-用] /  
わ [代] が [格助] み [名]、よ [名] に [格助]、  
ふる [ハ下二-体-経る/ラ四-体-降る] /  
ながめ [名-詠め/名-長雨]、せ [サ変-未]  
し [過-体]、ま [名] に [格助] /  
\$C|美しい花の色はいつしか色褪せてしまった  
ことよ。いたずらに長雨が降り続いているうち  
に(私の容色も衰えてしまったことよ。なすべ  
きこともなく、世を過ごす物思いをしていた間に)。  
\$D|花の色は褪せてしまっていたのだなあ、虚し  
くも。長雨に降りこめられ、また、我が身が世に  
暮らしてゆく上での物思いに耽っていた間に。  
\$E|序詞=わがみよにふる-ながめ、  
掛詞=ふる(降る)-ふる(経る),  
掛詞=ながめ(眺め)-ながめ(長雨)

の計量は行わないで、本来はこれら辞書エントリは作成する必要があった。

<sup>10</sup> 例) 「てなりけり」→て [接助] なり [断-用] けり [詠-終]

### 3 マネージメントシステムの開発

#### 3.1 設計仕様

データベースのモデルのなかで、いくつかの情報を統合する方法としては、次のようなものが考えられる。1) 原典にすべての情報をマークアップする方法(SGML, XML), 2) 個々のファイルは小さな単位としてリレーションデータとして持つ方法(SQL), 3) 単なるフラットなファイルとして持ち、相互にリンクする方法。

3) の方法は、一番データベースらしくない方法であり、単なる何の構造もマークアップもないテキストの寄せ集めである。しかし、これはフルテキストデータベースなどの呼び名で使われており、実際にサーチエンジンなどのデータベースはこれにあたる。極端ではあるがファイルであれば何でもよいゆえ、そのまま、CSVデータを入れて、縦計算をする、HTMLを入れて相互にリンクする、pdfをデコードして版下データベースとする、などいろいろ応用しやすい。

筆者らは、最もシンプルな3)の方法を基礎に、特別なソフトウェアやまたその知識を必要とせず、定義、タグ付が簡単でわかりやすい方法を検討した。また、データベース操作においては、研究者の行動の一部として、アップロード、公開、利用、ダウンロード、更新、アップロードといったサイクルが円滑に容易にできる方法を検討した。そこで、表3にしめすような仕様とした。

表3 システムの概要

1. データの作成に特別なソフトウェアを必要としないこと。
2. 公開が簡単であること。
3. 公開と同時に簡単に数値の計算・集計ができること。
4. スクリプトあるいは、データ構造、計算および検索の定義が簡単にできること。
5. 複数のデータベースが登録できること。
6. 同じ定義であれば、複数のデータベースをまたがって検索、計算ができる。いいかえれば、データベース間の通信ができるような仕組みを持つこと。
7. 複数のデータベースが相互に参照できること。
8. データの訂正やデータの記述方法の変更が発生した場合、それを一意に変更できること。
9. 分析の基となったデータがデータベースオーナーだけでなく、ユーザニヨリもダウンロードでき、それを加工して新たなデータが作成できるようにすること。

以上から、システムを1)データベースの公開を支援するシステム(BBDB)と2)作成や品質の管理を支援するシステム(BBQC)の2部で構成することにした。以下に詳細を報告する。

### 3.2 検索・公開システム (BBDB)

**概要:** データベースの公開・検索を容易に実施できるように設計されたシステムを BBDB(Bare Bone Database)<sup>11</sup> という。

BBDB は、総合研究大学院大学で公開されている貝塚データベース、小松左京コーパスなどの公開システムをもとに、プログラミングレスでデータベース管理者が簡単にブラウザからアップロードするだけで、公開・検索できるようにしたものである<sup>12</sup>。

BBDB によってデータベースを公開するには、定義部とデータ部を含むファイルを記述する。これらは、簡単なテキストファイルなので、ワープロ、エディタで記述できるものである。

**機能:** 管理者用の WEB 公開マネージャ、ユーザ用の検索、集計、ダウンロードなどの機能をがあり、WEB 公開マネージャとして、登録データベース一覧、データベース登録、削除、ディレクトリ認証管理、アピアランス管理、データベース概要ページ生成、検索指定画面生成、ダウンロード指定画面生成、検索として、一般語句検索、前後語句ソートつき KWIC 検索、数値比較検索、絞り込み検索、AND/OR 検索、集計として、頻度集計(正逆順位ソート付)、基礎統計(個数、平均、最大、最小、標準偏差)、ダウンロードとして、ファイル形式変換(Tab, CSV, BBDB)、テーブル略号変換、文字コード変換、ダウンロード必要フィールドの指定などの機能が利用できる。すべての操作はブラウザから行う。

**定義部:** 古今集 DB を例として表 4 に示す。ただし、簡単のため、これは一般公開用に作成された単純データベースのものであり、複数の相互参照の定義は含まれていない。定義部ではデータの構成・属性のほかに検索や計算の指定が記述されている。情報はすべて \$\$DB\_NAME のように【ドル2つ+定義文字列+】で示されるタグで、定義される。

**データ部:** ファイルのデータ部の開始箇所には、\$\$DATA を置き、それ以降にレコードを書くことができる。BBDB 形式の 1 レコードは、表 5 に示す形式である。各フィールドは、[\$A] のように、ドル1つ+アルファベット1文字+縦棒の3文字で始まる。デフォルトでは改行は無効で、各行の適当なところで改行をいれてよい。各フィールドのタグはわずか3文字なので、冗長性も少ないゆえに、誤入力も起こりにくい。閉じタグはない。制約は、1 レ

<sup>11</sup> Bare Bone は、直訳すれば「骨むき出し」の意だが、筆者らは、飾りなど一切ないが、シンプルであるがゆえに誰にでも使いやすく、かつ重要なものを、意図している。

<sup>12</sup> 2001 年、及川が筑波大学で行っている数理考古学の授業において BBDB はデータベース作成の演習に利用され、受講生は夏休みの宿題として、各 50 レコード以上の考古学関連データベースを作成した。

表 4 古今集 DB の定義ファイル (一部)

```
$$DB_ID|KW
$$DB_VER|2.0
$$DB_NAME|古今和歌集データベース
$$DB_OWNER|Hiroyuki Yamamoto
$$DB_EMAIL|yanagen@ucsd.edu
$$DB_ABST|古今和歌集 1,111 のデータベースで、歌一つが 1 レコードで収録されている。歌の原表記、仮名表記、岩波仮名表記、ローマ字表記、英語翻訳、品詞解析データ、作者名、作者名標準表記、作者性別などが収録されており、作者別の集計や性別による歌の分類、分析ができる。
$$HEADER|
$A|歌番号=歌につけられたユニークな番号 (6 桁)
$B|作者=本文に見られる作者名
$C|作者標準=作者の標準表記
$D|性別=作者の性別
( 検索は m = 男 / f = 女 / n = 読人不知 を指定すること )
$E|作者英文=作者の英文表記
$F|題=各歌の題
$G|題仮名=題の仮名表記
$H|題英語=題の英語表記
$I|歌=歌の標準表記
$J|歌仮名=歌の仮名表記
$K|歌岩波=歌の岩波体系による仮名表記
$L|歌品詞=歌の品詞分類
$M|歌ローマ=歌のローマ字表記
$N|歌英語=歌の英語翻訳
$O|解釈=英文 (Rodd, L.R.) による解釈
$P|リンク=歌人データベースへのリンク
$$REPLACE|
$D|
$$FIELD|
```

コード最大フィールド数が A-Z の 26 であること<sup>13</sup>、A フィールドは必ず 6 桁のレコード ID であること、の 2 点だけである。

**公開手順:** 定義部データ部の記述されたデータファイルが完成したら、次は、公開処理を行う。すべての公開処理はブラウザを使って行われる。簡単にいえば、ブラウザからデータファイルを転送し、その確認をしたら、ボタンを押してデータ保存する。もう一度ボタンを押して、HTML を生成して終りである。

はじめてアップロードする場合は、定義部とデータ部の全部もしくは一部が記載されたファイルをアップロードする必要がある。追加アップロードする場合には、定義部は不要であり、データ部相当を記述したファイルとして、Tab 形式、CSV 形式のファイルをアップロードすることができる。

公開手順を実行することによって、データベースの概要や検索、計算、ダウンロードの各種設定が、自動的に行われる。データベースの概要が生成された画面を図 1 に示す。1 レコードは 1 HTML 変換<sup>14</sup>し、公開すべきすべてのデータの生成を終了する。このようにあらかじめ設定ファイルから表示用のファイルまで、ボタン一つですべて作成してしまう。

<sup>13</sup> 26 以上使用しても使えないでも、それよりもまず、1 レコードが本当に 26 の要素を持たなければならないものなのかを、検討するべきであろう。

<sup>14</sup> これを我々は 1R1H と呼んでいる。

表 5 古今集 DB の 1 レコード (000007) : 一般公開版につき、作者の公開許諾のないデータはここには掲載されていない。各フィールドの詳細は、表 4 の HEADER 定義にしたがっている。

```
$A|000007
$B|よみ人しらす
$C|読人不知
$D|m@
$E|Anonymous
$F|題しらず
$G|たいしらず
$H|Topic uknown.
$I|心さし/ふかくそめてし/おりければ/きえあへぬ雪の/花とみゆらん@
$J|こころざし/ふかくそめてし/おりければ/きえあへぬゆきの/はなとみゆらむ@
$K|こころざし/ふかくそめてし/をりければ/きえあへぬゆきの/はなとみゆらむ@
$L|こころざし-名@/ふかく-形ク-用@, そめ-マ下二-用@て-接助@し-副助-強@/をり-ラ四-用@けれ-詠-已@ば-接助
@/きえ-ヤ下二-用@消える@, あへ-ハ下二-未@ぬ-消-体@, ゆき-名-雪@の-格助@/はな-名@と-格助@, みゆ-マ上一
終@らむ-現推-体@/@
$M|kokorozashi / fukaku someteshi / orikereba / kiaeenu yuki no / hana to miyuran /
$N|so longingly have I / awaited the fresh flowers / of spring that they have
/ dyed my soul and I see snow / as clustered blooms on branches /
$O|The former Chancellor was Fujiwara no Yoshifusa
```

### 3.3 品質管理システム (BBQC)

**概要 :**公開前の入力ミスなどチェック、公開後の更新、追加などこれまで多くのユーザが手作業で行っていたさまざまな品質管理の省力化、自動化を目的とし、実施できるように設計されたシステムを BBQC(Bare Bone Quality Control) という。標準的な機能の一覧は表 6 に示す通りである。BBQCによる品質管理処理は、ユーザが BBDB 形式のファイルを転送し、サーバ上で処理した後、ダウンロードもしくは BBDB に転送し、完了する。

**機能 :**ファイルをサーバに転送するとまず、定義部の文法チェックが行われる。その後、定義部にしたがって品質管理のための処理が行われる。表 6 の C1-10 は品質管理のためのチェックポイントである。文字種のチェックや不要な空白、数値のタイプ、範囲などが機械的にチェックされる。また、重複 ID、欠番 ID のチェックの機能、チェックにより判明した不都合をサーバ上で修正する機能(一括項目名変更、手作業修正)、また、その修正を確認するためのファイル内容表示、ファイルの追加、結合、不要になったファイルの削除機能、ユーザが自分のパソコンの表計算ソフトで作成した、csv, tab 形式のファイルを bbdb 形式に変換する機能、文字コード変換、データ項目一覧作成機能、などを持っている。

システム内部では、それぞれの仕事に応じたフィルタプログラムが定義部にしたがって実行する。仕組みとしては簡単であるが、実際、これらすべてのチェックポイントについて手作業で行うのは簡単ではない。

表 6 データベース品質管理操作メニュー

- A. upload definition file
- B. upload data file
- C. preset check points
  - 1. 1-byte char
  - 2. 2-byte char
  - 3. unnecessary space deletion
  - 4. integer
  - 5. float
  - 6. unknown table code
  - 7. mandatory
  - 8. field duplication
  - 9. max. number
  - 10. min. number
- D. process data
  - a) quality control
    - 1. quality check (checking point)
    - 2. duplicating/missing id check
    - 3. item authorization check
  - b) data modification
    - 4. replace values
    - 5. modify record by field item
  - c) file handlings
    - 6. delete file
    - 7. cat file
    - 8. append file to file
  - d) file code and format conversion
    - 9. tab -> bbdb
    - 10. bbdb -> tab
    - 11. csv -> tab
    - 12. tab -> csv
  - e) item list generation
    - 13. in alphabetical order
    - 14. frequency in use of a word
- E. modify record on the web
- F. delete record on the web

## 4 データベースの利用の実際

実際には、原典となる古今集 DB、3種の翻訳データ、品詞データ、作者データの4点からなるデータベースが作成されているが、著作権上公開できないものがあるので、実際に公開しているサイトを例に、BBDBと古今集 DB の利用の実際について説明する。

古今集 DB のページ<sup>15</sup>に行くと、図 1 に示すように定義部に記述した内容が表示され、データベースの概要を知ることができます。上の検索リンクを選んで、検索を行う。図 2 にあるように、語句検索 (and, or), KWIC 検索、数値比較検索を混在させた絞り込み検索ができる。

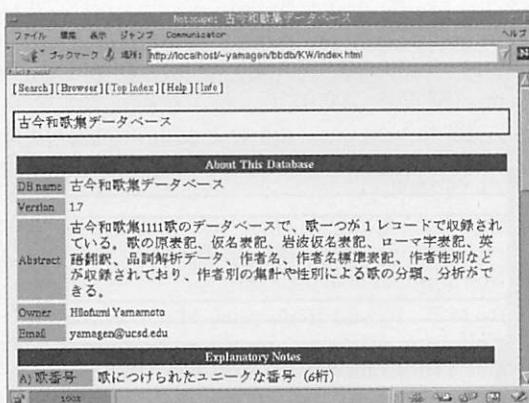


図 1 生成された概要ページ

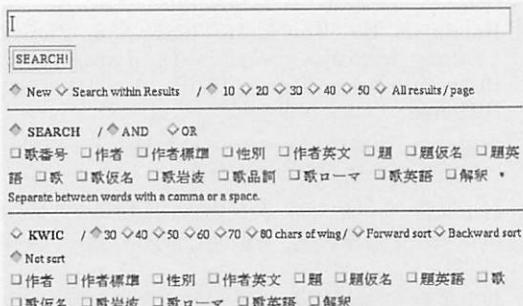


図 2 古今集 DB の検索指定画面

たとえば、和歌の中で「鳥」を詠んでいる歌を検索したいとする。すると結果表示画面 3 では、30 件ヒットしたことがわかる。この画面には、それぞれのレコード (HTML) へのリンクと次の 10 件を表示するためのリンクとダウンロード用のリンク [download], 統計計算のためのリンク [stat] が用意されている。リンク KW000143 をクリックすると、素性の歌のページへ行く。もちろん、これはあらかじめ生成さ

<sup>15</sup> <http://aci.soken.ac.jp/kokinshu/KW/>

れた静的なページであるので、BBDB 以外からもリンクすることができる。ここで、[download] をクリックすると、「鳥」30 件のデータがダウンロードできる。その隣の [stat] をクリックすると、大雑把ではあるが頻度と基礎統計の計算ができる。たとえば、作者標準の FREQ を ON にして、[stat] ボタンを押す (図 4) と、表 7 のように 27 の歌の作者分布がわかる<sup>16</sup>。さらに歌品詞を指定し、集計した結果が表 8 である。これによると、鳥 30 羽のうち 12 羽が「ほととぎす」であることがわかる。

search [鳥] 30 results found.
Page 1 [1 - 10]
1 KW000137 読人不知 さきまつ／山時鳥／うちはふき／いまもなかな／こ
2 KW000141 読人不知 さきなき／いまた旅なる／時鳥／花たち花に／やとは
3 KW000143 ▲素性 時鳥／はつ聲きけは／あちきなく／ぬしさたまらぬ／戀
4 KW000147 読人不知 時鳥／なかなか里の／あまたあれは／猪うとまれぬ／思
5 KW000152 ▲三国町 やよやまで／山時鳥／ことつてん／我世中に／すみわひ
6 KW000153 ▲友則 さみたれに／物思ひをれは／時鳥／夜ふかく鳴て／いつ
7 KW000155 ▲千里 やとりせし／花橋も／かれなくに／なと時鳥／鶯たえぬ
8 KW000159 読人不知 こそこの夏／鳴ふるしてし／時鳥／それがあらぬか／聲の
9 KW000160 ▲貫之 五月雨の／空もとゝろに／時鳥／かにをうしとか／よた
10 KW000162 ▲貫之 時鳥／人まつ山に／なくなれば／独うちつけに／戀ま

Page 1 2 3 [Download] [Stat]

図 3 「鳥」の検索結果

Sort	normal	reverse	case-sensitive
by item	◆	◆	◆ Yes ◆ No
by value	◆	◆	
Item	Type	on off	
作者標準	FREQ	◆ ◆	
性別	FREQ	◆ ◆	
歌	FREQ	◆ ◆	
歌仮名	FREQ	◆ ◆	
歌岩波	FREQ	◆ ◆	
歌品詞	FREQ	◆ ◆	

図 4 基礎統計データ出力指定画面

表 7 基礎統計データ出力結果「作者標準」

0	27	100.00%	TOTAL
1	13	48.15%	読人不知
2	3	11.11%	友則
3	2	7.41%	貫之
4	2	7.41%	忠岑
5	1	3.70%	業平
6	1	3.70%	三国町
7	1	3.70%	敏行
8	1	3.70%	寵
9	1	3.70%	千里
10	1	3.70%	躬恒
11	1	3.70%	素性

<sup>16</sup> 結果出力で、「30 found」と出ていたのは、30箇所に「鳥」が見つかったの意味。

表 8 統計データ出力結果「歌品詞」(一部)

0	397	100.00% TOTAL
1	21	5.29% の-格助
2	15	3.78% に-格助
3	12	3.02% ほととぎす-名
4	10	2.52% も-係助
5	8	2.02% を-格助
6	8	2.02% ば-接助
7	8	2.02% と-格助
8	7	1.76% が-格助
9	6	1.51% て-接助
10	5	1.26% は-係助

## 5 おわりに

本論<sup>17</sup>では、古今集DBの品詞データと英翻訳データの開発について報告した。品詞データには、まだ誤りや活用形の未整理の箇所があり、今後もデータを見直し、修正を加えていく必要があるが、品詞や活用形の認定、複合語の取り扱いが研究者によって異なり一意に決められず、データ化する上で問題となることを述べた。

また、公開や品質管理を効率的に行うシステムを開発し、その目的、利用方法、利用例について報告し、定義ファイルとデータファイルをアップロードするだけで、検索、集計処理の指定まで一括して行え、簡単な計算処理まで行えることを示した。

今後の課題として2つのことがある。ひとつは古今集DBのコンテンツの充実とその精度の向上、もうひとつはBBDB、BBQCの公開と普及である。

前者は、現在のDBを広く公開することによって、利用者である研究者からデータベース中の誤りを指摘してもらったり、新たな意見や提案をデータベースに反映することによって実現していくことが可能である。そのためのツール群もBBDBの中に組み込んでいく予定である。

後者については、すでに公開のためのプラットフォームとなるサーバを総合研究大学院大学の図書館に設置しており、マニュアルの作成と並行してその準備作業を進めている。BBDB、BBQCを広く普及していくためには、まず、研究者が手軽に利用できるプラットフォームが必要で、今後研究者への呼

<sup>17</sup> 国文学研究資料館の中村康夫先生には国文学研究資料館データベース二十一代集の利用を快諾していただきましたばかりでなく、CDROMを貸与してくださいました。国文学研究資料館の安永尚志先生には、パラレルテキストと国文学資料について御指導いただきました。青山学院大学の近藤みゆき先生には、品詞データ作成の際、国文法についてご指導いただきました。実践女子大学の近藤みゆき先生には、ジェンダーデータつき単位切りデータを提供していただきました。コロラド大学ボルダー校のローレルラスプリカロッド先生にはご著書をいただきましたばかりでなく、データベース化および公開について、快諾いただきました。お礼申し上げます。

びかけやワークショップの開催などを通じて、その実現を図っていきたいと考えている。

## 参考文献

- [1] 佐竹昭廣, 立川美彦: 重層型情報時代に対応する国文学高機能情報形成手法の開発とその実用化に関する研究, 技術報告, 国文学研究資料館 (1998). 平成7年度～平成9年度科学研究費基盤研究(A)(2)研究成果報告書(課題番号 07401014).
- [2] 中村康夫, 立川美彦, 杉田まゆ子: 国文学研究資料館データベース 古典コレクション『二十一代集』(正保版本) CD-ROM, 岩波書店 (1999).
- [3] 村上征勝: 文章分析と統計学, 数理科学 特集 知としての統計学, Vol. 11月号, No. 389, pp. 27-33 (1995).
- [4] 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).
- [5] 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).
- [6] 竹田正幸, 福田智子, 南里一郎, 山崎真由美, 玉利公一: 和歌データからの類似歌発見, 統計数理, Vol. 48, No. 2, pp. 289-310 (2000).
- [7] Rodd, L. R. and Henkenius, M. C.: *Kokinshū - A Collection of Poems Ancient and Modern*, Cheng and Tsui Company, Boston MA USA (1984).
- [8] McCullough, H. C.: *Kokin Wakashū, The first Imperial Anthology of Japanese Poetry Translated and annotated by Helen Craig McCullough with Tosa Nikki and Shinsen Waka*, Stanford University Press, Stanford, CA, USA (1985).
- [9] Honda, H.-H.: *The Kokin Waka-Shū; The 10th Century Anthology*; edited by The Imperial Edict, Hokuseido Press, Tokyo (1970). Translated by H. H. Honda.
- [10] 宮嶋達夫, 中野洋, 鈴木泰, 石井久雄: フロッピー版古典対照語彙表, 笠間書院 (1989).
- [11] 村上征勝, 今西祐一郎: 源氏物語の助動詞の計量分析, 情報処理学会論文誌, Vol. 40, No. 3, pp. 774-782 (1999).
- [12] 村田菜穂子, 岩田俊彦: 平安時代の文学作品における形容動詞対照語彙データベースの構築とそれを用いた語彙論的研究, 情報処理学会研究報告, Vol. CH45-10, pp. 73-80 (2000).
- [13] 小林和彦: 古典新釈シリーズ 5 古今和歌集, 中道館 (1978).
- [14] 久曾神昇: 古今和歌集 全訳注 (1)-(5), 講談社学術文庫 (1982).
- [15] 金明哲: 自然言語における統計手法を用いた情報処理, 統計数理, Vol. 48, No. 2, pp. 271-287 (2000).
- [16] 片桐洋一: 古今和歌集全評釈 上・中・下, 日栄社編集書 全3巻, 講談社 (1998).

# 古典文献データベースの表記体系確立

徳永 宗雄  
山田 篤

## 【要旨】

古典学研究分野に於けるコンピュータの導入は、古典文献および研究資料群をデジタルデータとして保有することを可能にし、検索・分析を容易かつ多様にするのみならず、研究成果公開やその他の情報交換の補助手段として一般化するに至った。しかし、コーディング方式や文書フォーマット等、文献をデジタル化する際の画一的な方法が定まらず、各研究者が独自の形式でデータを蓄積してきたため、従来情報交換の障壁となってきた。また、XMLの文字レパートリであり、近年多くのオペレーティング・システムによって採用されているUnicodeは、世界中の多くの言語に含まれる文字を一つのコード体系で表現しようとする意欲的な試みであるが、当初の16ビットで表現可能な文字数は所詮65536字であり、特に東洋の言語研究においては、この文字セットに含まれていない文字も少なくなかったため、学術研究に用いるに妥当でない場合がある。このような状況を踏まえて、本計画研究前期二年では、文字コードシステムの乱立を抜本的に解決する方法として、バイト数による固定した通し番号ではなく、動植物の分類に見られるように、文字に「学名」を付けて処理する方式を提案した。その研究成果は、第一期情報処理調整班研究成果報告書において圧縮ファイル (GSCSm. tar. gz) で提出した。

・デジタル技術の急速な進歩に伴い、今後、古典学・人文学研究はオンラインでの研究に比重を移していくものと考えられる。この観点から、本計画研究では、マハーバーラタの哲学篇である「解脱法品」を対象に、インターネットでの共同研究を試行した。これは、古典学・人文学では先駆的な試みである。この共同研究は、研究班の中心メンバーが作成したドラフトをめぐって研究班員が電子メールで意見を交換し、その成果を取り込んでドラフトを改定してインターネットに公

開するというものである。しかし、この種の共同研究を本格的な軌道に乗せるためには、オンラインでのXMLを用いた情報交換システムの開発が不可欠となる。そこで本計画研究後半の二年間では、オンライン共同研究の基盤となるべき古典学 XML 環境の構築を目指した研究を主に行った。また、この環境の構築の観点から、文字コードにかかる問題を新たに検討しながらした。この研究は研究分担者山田篤（京都高度技術研究所情報メディア研究室長）によるものであり、その研究成果の詳細は本報告書の後半に付す。

## 【本計画研究における、他領域との連携による成果】

本研究は、オンライン古典学共同研究を情報工学の知識で実現するところにあり、その成果は人文学と情報工学の連携によるものである。

## 【本計画研究の位置付け】

オンライン共同研究はすべての古典文献研究のために企画されたものであり、特定の古典文献との関連で位置づけられる性質のものではないので、「位置づけ」については省略する。

## 【平成13年度－14年度の主な研究成果】

(研究分担者山田篤執筆)

オンライン共同研究の基盤となる古典学のためのXML環境の構築を目指して、

- 1) データの XML 化に関する検討
- 2) テキスト出力に関する検討
- 3) データベースに関する検討
- 4) ウェブアプリケーションに関する検討
- 5) 文字コードに関する検討
- 6) オンライン共同研究支援システムに関する検討

を行った。以下でそれぞれの点について概要を述べる。

## 1. データの XML 化に関する検討

テキスト出力形式の多様化の実状に対応すべく、再利用可能で汎用性のある表記方法として XML を採用することとした。拡張可能なマーク付け言語である XML (<http://www.w3.org/XML/>) は1998年に W3C (World Wide Web Consortium) 勧告となった標準であり、現在さまざまな用途で用いられている。その用途は大別すると、メタデータの記述とテキストへのアノテーションに分けることができるが、我々の目的であるオンライン共同研究においては研究対象となるテキスト（具体的にはマハーバーラタ）へのアノテーションの記述が主となる。

このアノテーションの付与を最詳レベルまで一度にやってしまおうとすると往々にして破綻するため、今回は比較的粒度の粗い

- a) 原テキストの構造の記述
- b) 原テキストに対する訳注

を対象とした。原テキストの構造としてはマハーバーラタの場合、`book>chapter>verse` という階層関係があり、`verse` の中ははじめに `speaker` がきて次に複数の `line` が続くという構成が一般的である。この各行毎に對訳を割り当て、注釈は `verse` 単位でつけられるようにした。設計したタグセットを DTD (Document Type Definition) として表現したものを図 1 に示す。また、このタグセットを用いて実際にデータの電子化を行った。このタグセットを用いたアノテーションの一例を図 2 に示す。

アノテーション付与作業は原テキストの電子化とそれに対するタグ付けという作業になる。アノテーションの付与されたテキストをドラフトと呼ぶ。ドラフトはオンライン共同研究の過程で隨時更新されるものである。更に、ドラフトの作成は粗い部分から徐々に詳細なタグをふっていくという段階的詳細化の考え方方が適用できる。例えば今回の DTD で記述されていない語レベルの情報（固有名詞に対するマーク付け等）は、すでにマーク付けされたドラフトに対する新たなタグの追加という形で記述可能である。

```
<!ELEMENT book (title?, pre*, chapter+)>
<!ATTLIST book
          num      CDATA  #IMPLIED>
<!ELEMENT title (#PCDATA)>
<!ATTLIST title
          sub      CDATA  #IMPLIED>
<!ELEMENT pre  (#PCDATA)>
<!ELEMENT chapter (verse+)>
<!ATTLIST chapter
          num      CDATA  #IMPLIED>
<!ELEMENT verse (text, comment?)>
<!ATTLIST verse
          num      CDATA  #IMPLIED>
<!ELEMENT text (speaker?, line+)>
<!ELEMENT speaker (#PCDATA|trans)*>
<!ELEMENT line (#PCDATA|trans)*>
<!ELEMENT trans (#PCDATA|n|name|q|qu ote)*>
<!ELEMENT comment (#PCDATA|n|name|q|quote|rightarrow)*>
<!ELEMENT n   (#PCDATA)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT q   (#PCDATA)>
<!ELEMENT quote (#PCDATA)>
<!ELEMENT rightarrow (#PCDATA)>
```

図 1 XML 化に用いた DTD

```

<book num="12">
<chapter num="170">
<verse num="1">
<text>
<speaker><trans>YudhiSThira said —</trans>yudhiSThira uvaaca.</speaker>
<line><trans>Whether rich or poor, happiness and suffering fall on them who live
independently.</trans>dhanino vaadhanaa ye ca vartayanti svatantriNaH</line>
<line><trans>Grandfather, who
[obtains happiness and suffering] and how?</trans>sukhaduHkhaag amas teSaAM kaH kathaM
vaa pitaamaha</line>
</text>
<comment><rightarrow/>NiilakaNTha provides a specific interpretation for <quote>
svatantriNaH</quote>: <quote>svazastraanusaariNaH</quote>. I propose
*<quote>svayantritaH</quote> 'instigated by themselves' as the original
reading. (cf. svayantriNaH in Dn1, M5). — I do not take <quote>
kaH</quote> as the correlative of <quote>sukh aduHkhaagamas</quote>, for <quote>
kaH</quote> would then mean a strong negation of the experience of happiness
and suffering.</comment>
</verse>

```

図2 XMLによるアノテーションの例

## 2. テキスト出力に関する検討

次いで、このようにして作成した XML 文書から多様なテキスト出力形式を生成する実験として、'one source, multi use' の考え方に基づき同一の XML 文書からウェブページと印刷物を生成する方法について検討した。これには XML 文書の変換のための XSLT (XSL Transformations (<http://www.w3.org/TR/xslt>)) を用いる。XSLT は1999年に W3C 勧告となった XML 文書を他の XML 文書に変換するための言語である。我々の場合は、変換元が先に作成した XML 文書であり、変換先としてウェブページと印刷物の 2 種類を考えることになる。

ウェブページの場合は HTML を変換先として設計すればよいが、印刷物の場合は若干の検討を要する。XSL (Extensible Stylesheet Language (<http://www.w3.org/TR/xsl/>)) は XML 文書用のスタイルシート言語として策定され、2001年に W3C 勧告となった。これを用いると、XML 文書をフォーマッティングしてページイメージ（例えば PDF 等）を作成可能であるが、現状ではまだ処理系が不完全である。処理系の問題は将来的には解決されると考えられるが、別の問題としてページイメージの一部を論文原稿の一部として利用したいといった再利用性に関する要求がある。これに対しても、ページイメージとしてレンダリングされる一段手前のものが必要になる。このため、印刷物に関しては、我々は論文等が LaTeX で作成されるという

前提で、XML 文書からの変換先として LaTeX のソースを生成することとし、ページイメージは LaTeX ソースのコンパイルによって得るという 2 段階の方法を探った。

元の XML 文書には表示方法に関するタグは含まれておらず、表示に関する指示はすべて変換時にタグ毎に行うことになる。元の XML 文書から HTML に変換する XSLT 記述と、LaTeX ソースコードに変換する XSLT 記述の 2 種類を作成し、それぞれの中で HTML タグ及び LaTeX コマンドを用いて表示方法を指定した。変換のための XSLT 記述は膨大になるため省略するが、HTMLへの変換結果の表示例を図3に示す。

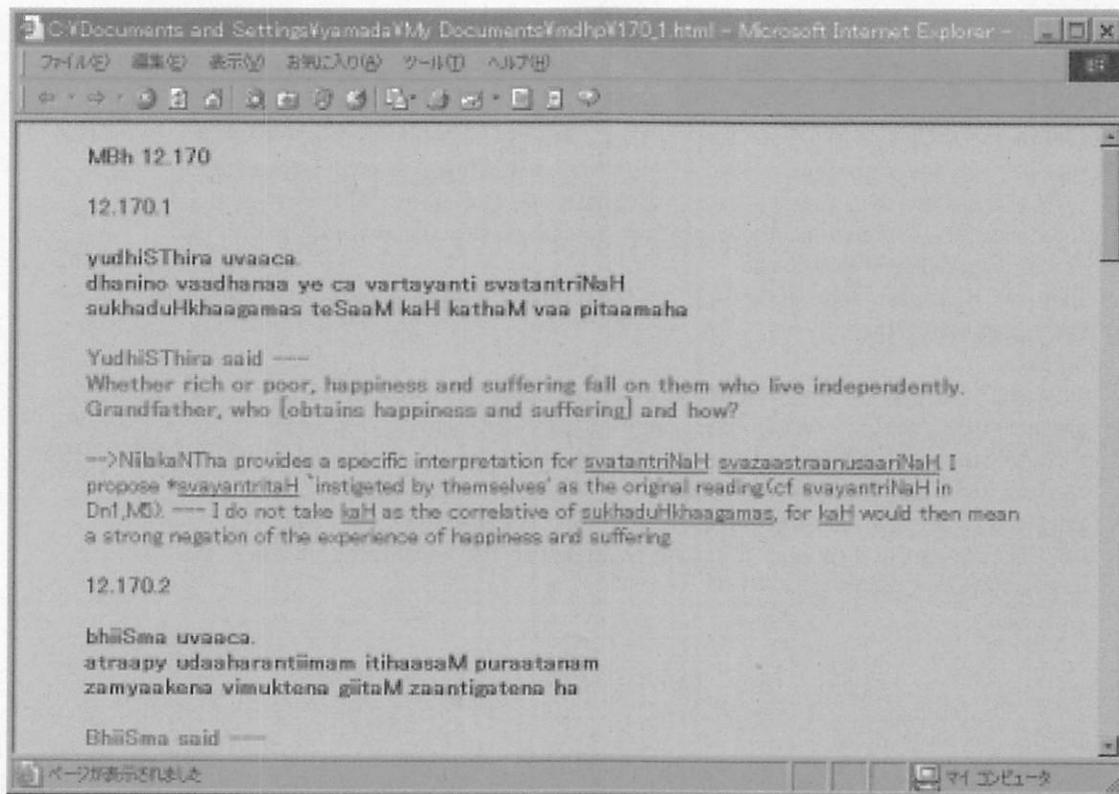


図 3 HTML への変換

XML 文書を XSLT 記述によって変換するための処理系としては Apache XML プロジェクトの Xalan (<http://xml.apache.org/xalan-j>) を用いた。Xalan は Apache Software License に基づき配布されているオープンソースのソフトウェアである。なお、HTML への変換については、MicroSoft Internet Explorer 5.5 以降ないし Mozilla 1.0 以降であれば、変換元の XML 文書にスタイルシート宣言を含めておけば、XML 文

書を読み込むと変換後の HTML 文書のレンダリング結果が表示される。

次に LaTeX のソース文書に対しても同様の考え方で XSLT 記述を作成した。LaTeX の場合は原テキストの文字列に対し LaTeX コマンドを用いて文字修飾を付与する変換も組み込んだ。変換結果の LaTeX ソースの例を図 4 に示す。

```
{#bf Chapter 170}#vspace{1.7ex}##
YudhiSThira said ---#medskip##Whether rich or poor, happiness and suffering fall on
them who live
independently. Grandfather, who
[obtains happiness and suffering] and how? (1)#vspace{1.7ex}
#footnote{#hspace*{-3ex}#raisebox{0.65ex} {#scriptsize#bf170.1}#hspace{1ex}}
#rightarrow$NiilakaNTha provides a specific interpretation for {#em
svatantriNaH}: {#em svaz#=a} str#{=a} nus#=a} riNaH}. I propose
*{#em svayantritaH} 'instigated by themselves' as the original
reading. (cf. svayantriNaH in Dn1, M5). --- I do not take {#em
kaH} as the correlative of {#em sukhaduHkh#=a} gamas}, for {#em
kaH} would then mean a strong negation of the experience of happiness
and suffering. }
```

図 4 LaTeX への変換例

以上で、元の XML 文書さえ管理すればウェブ公開用の HTML 文書、印刷用の LaTeX ソース文書を自動生成できるようになった。修正や更新が生じた場合も HTML 文書、LaTeX 文書を個々に書き換えるのではなく、元の XML 文書のみ書き換えればよいので、修正の手間やミスは減少することが期待できる。また、修正者は表示方法を指定する HTML タグや LaTeX コマンドに煩わされることなく、内容のみに集中することができます。

### 3. データベースに関する検討

次に XML 文書の管理方法に関する検討を行った。これまで説明してきた XML 文書はコンピュータ上ではファイルとして存在する。これらをコンピュータのファイルシステムを用いてそのまま管理することも不可能ではないが、XML 文書の数が膨大になると管理が難しくなる。また、XML 文書を対象にした検索を考えた場合も効率的な管理方法が必要になる。そこでデータベースシステムの利用を検討した。

現状では XML 文書をデータベースに格納する方法として、RDB (Relational Database) を用いる方法と XML DB を用いる方法が考えられる。RDB は XML DB に比較して技術的にも熟成しており、オープンソースのソフトウェアとしても PostgreSQL や MySQL 等、利用可能なものが多数ある。しかしながら、XML 文書を RDB に格納する際には、RDB のテーブル形式へのデータのマッピングが不可欠となる。言い換えば XML 文書を XML 文書のまま RDB に格納することは原則としてできず、RDB へのデータの格納時に XML 文書を分解し、取り出し時に再構築する必要が生ずる。このため、我々は XML 文書をそのままの形でデータベースに格納する方式を採用した。このような DB を XML DB と呼ぶ。XML DB はまだ発展途

上の技術ではあるが、W3C でも XML 文書に対する検索用言語として XML Query (<http://www.w3.org/XML/Query>) の仕様が策定中である。また、例えば Apache XML プロジェクトからは Xindice (<http://xml.apache.org/xindice>) というオープンソースのソフトウェアが公開されている。Xindice では検索用の言語として XPath (<http://www.w3.org/TR/xpath>) を、DB 更新用の言語として XML:DB XUpdate (<http://www.xmldb.org/xupdate>) を用いる。XML DB は非常に巨大な少量の XML 文書を管理する場合には向いていないが、それほど大きくない多数の文書を管理する場合に向いている。我々の場合は、行単位の管理はあまり現実的ではないとしても、XML 文書を chapter ないし verse 単位で管理することにより、XML DB を効果的に用いることができる。

### 4. ウェブアプリケーションに関する検討

以上で、XML 文書を XML DB に格納し、必要に応じて取り出し HTML ないし LaTeX ソースに変換して提供する仕組みができた。次にこれをウェブアプリケーションとして構成し、データの管理もウェブを介してできるような仕組みを考えた。ウェブアプリケーションであるので、中心にウェブサーバを置く。ユーザは管理者と閲覧者に分かれ、ともにウェブブラウザを用いてサーバにアクセスする。管理者はウェブサーバにアクセスして XML DB に格納されているデータを修正したり、新たなデータを追加・格納できるようにする。閲覧者はウェブサーバにアクセスし、XML DB に格納されているデータを取り出して、HTML に変換された結果の表示を閲覧することができる。

システムの構成を図 5 に示す。

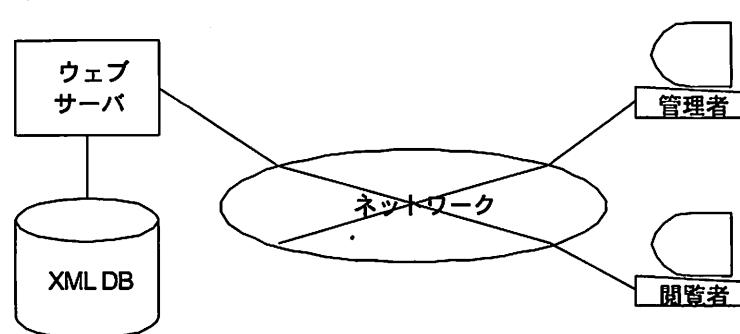


図 5 ウェブアプリケーションシステムの構成

XML DBとしては先に述べた Xindice を用いる。Xindice に対しては XML:DB API (<http://www.xmldb.org/xapi>) を用いてアクセスすることができる。この API を用いるために Java で専用の Servlet を作成することも考えられるが、ラピッドプロトタイピングのために我々はウェブサーバとしては横浜ベイキットの BayServer (<http://www.baykit.org/bserv/>) を用い、XML コマンド Xi (<http://www.baykit.org/xi/>) を用いてシステムを構築することとした。ウェブサーバ上には管理者向けのページと閲覧者向けのページを準備し、ユーザ認証によって管理者を識別する。閲覧者が既存の XML 文書を chapter 及び verse 番号によって指定すると、ウェブサーバは XPath を使って対応する XML 文書を取り出し、XSLT による変換を施して、結果の HTML 文書を閲覧者のコンピュータに送る。管理者もほぼ同様の手順でシステムを利用することができますが、これに加えて修正・追加ボタンを押下することにより、XML DB 上のデータの修正や更新ができる。このとき、ウェブサーバは管理者が入力したデータを受け取り、XUpdate を用いて XML DB に格納する。

## 5. 文字コードに関する検討

これまで述べた XML 文書で用いる文字は原則として ASCII の範囲で、この範囲外の文字は変換時に LaTeX コマンドを用いて表現するという方法を用いていた。この場合は XML 文書の charset としては ISO-8859-1 を指定すればよい。一方、XML の文字コードの省略時解釈は Unicode (UTF-8) である。現状ではまだ過渡期と考えられるものの、ユーザサイドの環境としても徐々に Unicode が使えるようになってきている。例えば Internet Explorer や Mozilla といったウェブブラウザでも最近のバージョンでは Unicode で記述されたページを表示できるようになっており、Unicode が使用可能な環境は今後整っていくものと考えられる。このため、Unicode の表現能力や構成には一部疑問の余地もあるが、Unicode を利用可能にしておくことは必須である。

当初16ビットの文字コードであった Unicode も現在 Unicode 3.2 が公表されており、21ビットスペースまで拡張されている。一方で文字コードに関する国際標準としては ISO 10646 で定められている UCS (Universal Character Set) がある。ISO 10646 は31ビットのコードセットであり、231文字まで表現可能であるが、Unicode と互換になっており、両者で規定されている文字は同じコード、同じ名称を持っている。両者

の対応関係は Unicode 1.1 が ISO 10646-1:1993 に、Unicode 3.0 が ISO 10646-1:2000 に、Unicode 3.2 が ISO 10646-2:2001 に対応している。なお、UTF-8 は ISO 10646-1:2000 Annex D で定義されているエンコーディング方式である。これらは既存の文字コードで表されていた文字を包摂する形で策定が進められているため、この中で規定されている文字については原則としてそのコードを利用することした。しかしながら、古典学の分野で用いる文書の中には ISO 10646 で規定されていない文字を使う必要性が依然として残る。これらについては、XML ベースで文書を管理する利点を生かし、独自の XML タグを用いて表現することとする。

また、これらの表示については、先に述べた XSLT による変換の仕組みを用いて、ビットマップ画像ないしひベクトル画像等を用いて変換し、視覚化する。ここでも管理する文書とその表示形態を分離しておく手法が有効に機能する。

ここで、先に述べたウェブアプリケーション構築時の UTF-8 エンコーディングの取り扱いについて述べる。ユーザサイドの表示系については先に述べたように、最新のウェブブラウザを用い、ユーザ側に対応するフォントがインストールされれば、表示は可能である。サーバサイドの XML DB である Xindice はバージョン 1.0 では UTF-8 の取り扱いが不完全であったが、バージョン 1.1 (現状ではまだ CVS バージョン) では改善され、UTF-8 エンコーディングの文書の格納、取り出しが可能となっている。

ユーザサイドからサーバサイドへのデータの受け渡しについては、現在 W3C で XForms (<http://www.w3.org/MarkUp/Forms/>) が勧告候補案として検討されているが、現状ではまだ HTML フォームを用いる方法が一般的である。この場合、ユーザがフォームに入力した文字をサーバ側で正しく受け取り処理することができるかどうかはシステムの構成に依存する。まず、ブラウザ側は最新のものであれば UTF-8 でエンコーディングされたウェブページ内のフォームからの submit は UTF-8 で送ることができる。よって、サーバ側で受け取った文字列を UTF-8 としてデコードすればとりあえずの目的は達成される。ただし、現状では特に日本語環境の場合、UTF-8 のみに特化するのではなく、必要に応じて Shift\_JIS 等の従来から用いられているものと UTF-8 を使い分けられることが望ましい。サーバ側の構成として Apache, Tomcat, BayServer でテストしたところ、BayServer を単独でウェブサーバとして用いた場合に UTF-8 と Shift\_JIS のペ

ージを使い分けてそれぞれ正しくデータを受け取れることが確認できた。

## 6. オンライン共同研究支援システムに関する検討

オンライン共同研究の推進には、研究者同士が意見を交換できるような仕組みと、意見交換の結果作成された研究成果を公刊する仕組みが必要となる。そこで、これまでに述べたウェブアプリケーションを拡張して、複数の研究者が同一のドラフトに対してネットワークを介して互いに意見・注釈をつけ、その結果を互いに参照することができるような仕組みについて検討を行った。

この場合、次の3種類のユーザが生じる。

- a) 管理者：ドラフトの登録その他システムの管理を行う
- b) 共同研究者：登録されたドラフト、他の研究者の意見等に対して自分の意見を書き込む
- c) 閲覧者：ドラフト、意見の閲覧のみ

場合によっては、参加者全員がb)に属し、c)が存在しないこともあります。このうち、a)とc)は今までに述べてきたウェブアプリケーションで実現できるため、新たに生じたb)に対して更に検討を加える。

ウェブ上における意見交換システムとしては掲示版システムや、最近ではWiki等のシステムがある。掲示板ではスレッドを使って一連の議論の流れを作ることはできる。スレッドのルートとしてドラフトを設定すれば、要請を擬似的に実現することは可能と考えられるが、スレッドの動的な形成は難しく、単一の見方しかできなくなる可能性がある。議論の流れを様々な形で見ようすると、各自の意見を個々に格納しておき、必要に応じて表示対象や表示形式を決めるという動的な編成が必要になると考えられる。動的な編成を考えると、XLink (<http://www.w3.org/XML/Linking>)

等のロケーションモデルを用いて、それが何に対する意見かといった情報を個別に持つほうが都合がよい。また、電子データの印刷出力にあたっては、単独著作として自分が執筆した部分のみを取り出す場合と、共同著作として複数の研究者の執筆部分を取り出す場合を考えられる。更に編集著作の場合はこの後に編集作業が入る。これらの場合もドラフトに対して特定の研究者の意見のみを取り出したり、ある話題に対する複数の意見を取り出すといった任意の見方を導入できるようにするために、それぞれの意見が独立に操作できることが望ましい。

このために、はじめに設計したXML DTDではコメントは管理者のみが記入できるものとしていたが、これを任意の共同研究者が自らの意見を記述できるようにするとともに、ドラフトのXML文書とは別に管理し、「誰」が「何」に対して付与したコメントかを管理することとする。

共同研究者が意見を書き込む際の処理の流れは次のようになる。

- 1) 意見を付与する対照となるドラフト、ないし他の意見を選択する
- 2) 自分の意見を書き込む
- 3) 入力された意見が、1)で選択された場所を示す情報、及び入力者、入力日時の情報とともにXML DBに格納される

閲覧の際にはドラフトを元にした表示、研究者毎の表示といった様々な見方が可能となるようになる。これらに共通した処理の流れは次のようになる。

- 1) ユーザがどのような見方をしたいかを選択する
- 2) XML DBから関連する文書（ドラフト、意見）が検索される

3) 検索された文書群から表示イメージを構成されるウェブアプリケーションの場合、ここで構成される表示イメージはHTMLベースで、ボタン等のインターフェースを用いてインタラクティブに動作させることもできる。

印刷物の生成の場合も処理の流れは同様で、3)において構成される表示イメージが、検索されたコンテンツをすべて含んだページイメージとなる。

ここで示したオンライン共同研究支援システムは先に述べたウェブアプリケーションシステム上に実装することができる。今後は表示方法の更なる改善を行うとともに、FOP (<http://xml.apache.org/fop/>) 等のXSL処理系を用いた印刷イメージの直接生成にも取り組みたい。

以上

# 古典学のための多言語処理システムの開発

高島 淳

## 1 【要旨】

- 古典研究のような一般には用いられることの少ない言語や文字の処理に関しては、今後相当の期間にかけて、コード体系の統一や汎用的な多言語処理の方法が確立される可能性が存在しない。従って、将来的にはいかなるコード体系にも変換可能であるような多言語多書記法処理系を当面構築する必要がある。
- この目的のために、データ自体は任意に規定された ASCII 転写法で保持しつつ、印刷等の出力においては TeX に基づく多様な文字表記を可能とするシステムを開発した。
- 当面の実装は、ローマ字転写、ナガリー文字、グランタ文字、カンナダ文字、タミル文字、マラヤーラム文字、テルグ文字、ベンガル文字、オリヤー文字、グルムキー文字、チベット文字、クメール文字、タイ文字に関して達成した。

イ文字およびクメール文字の処理を可能としたことで、幅広い領域の研究を統合した成果を挙げたといえる。

## 3 【位置付け】

古典研究においては、当然のことながら、現在においては用いられていない言語や文字を使用しなければならないことが多い。研究代表者が専門とするインドの場合では、代表的な古典語はサンスクリット語であるが、ローマ字転写の場合でもコードが統一されているとは言えず、多くの処理上の問題点を抱えている。最古期のヴェーダ語アクセントの問題も一例であるし、写本資料の正確な転記の観点からは、ローマ字転写ではなく、ナガリー文字、グランタ文字、テルグ文字、カンナダ文字、テルグ文字、ベンガル文字、オリヤー文字などの多様な文字を用いて表記したいという要望も強い。

本研究においては、こうした要望に応えるべく、現在研究者が使用している多様な転写入力方法に対応していると同時に、その一つの入力方法からローマ字転写、ナガリー文字、グランタ文字、カンナダ文字、タミル文字、マラヤーラム文字、テルグ文字、ベンガル文字、オリヤー文字、グルムキー文字などを印刷出力とすることの出来るシステムを構築することを目的とする。また、ローマ字転写においては、語の構成要素に出来る限り従うようなハイフネーション区切りの方式も開発する。

Unicode の制定によってコンピュータ上の多言語処理が可能になったように言われることがあるが、Unicode は一般的な商用利用に限定したコード化を行っている

## 2 【他領域との連携による成果】

本研究におけるシステムの開発は、当初からすべての分野の古典研究に対応できるような汎用的システムを目指しているから、すべての分野の研究者の文書処理に関する視点が参考とすべきものである。

本研究においては、研究代表者が専門とする古典サンスクリットに関する処理を出発点として、古典タミル語を含む大部分のインド諸語の処理を可能にしたにとどまらず、サンスクリット転写を含むチベット文字、東南アジア研究およびパーリ語文献のために必要なタ

ため、将来 ISO が新たなコード体系を制定するのを待つ間は、いまだコード化されていないような文字や言語を使用する古典研究のためには独自の多言語処理体系を開発する必要がある。特に Unicode は、言語と文字種が同一であるとの前提に立っているため、上述したような同一言語多書記法 (script) 表記のような場合に全く対応できないのである。本研究で開発するシステムは、同一言語多書記法 (script) 表記や同一書記法 (script) 多言語表示を可能とすると同時に、研究者各人による容易なカスタマイズを可能にすることで、現在のような過渡期において充分に実用的なシステムとなると言え、古典研究者にとって強力なツールとなるであろう。

現在、一方では ISO による諸文化に適応した標準化の試みや、他方ではとりあえず Unicode を使用して可能な限りの多言語環境の構築の試み（たとえば Omega Project）が行われているが、前者が実用的なレベルに到達するには 20 年近くを必要とすることが確実であるし、後者は Unicode の問題点の故に、2~3 年以内に実用化されたとしても、古典研究には不十分である。その点で、本研究は短期間で実現可能であり、将来新しい標準が策定された場合にも、容易にコンバート可能である。

また、どのようなプラットフォームの上でも運用可能なシステムとするために Perl と TeX を用いた汎用的なシステムを開発することから、広範囲の研究者の利便となると同時に、インド以外の（特に同じインド系文字を使用している東南アジア）古典研究にも容易に拡張可能である。

## 4 研究の経過

本研究は、当初、「古典学再構築」の公募研究として始め、ついで計画研究として実施したものである。その内訳は以下の通りである。

### 第一期

研究種目：特定領域研究 (A)(2)(公募研究)

研究期間：平成 11 年度～平成 12 年度

課題番号：11164223

研究課題：古典学のための多言語処理システムの開発

研究組織：

研究代表者：高島 淳（東京外国語大学アジア・アフリカ言語文化研究所助教授）

研究経費：	平成 11 年度	1,400 千円
	平成 12 年度	1,400 千円
	計	2,800 千円

### 第二期

研究種目：特定領域研究 (A)(2)(計画研究)

研究期間：平成 13 年度～平成 14 年度

課題番号：13018211

研究課題：古典学のための多言語処理システムの開発

研究組織：

研究代表者：高島 淳（東京外国語大学アジア・アフリカ言語文化研究所教授）

研究分担者：峰岸 真琴（東京外国語大学アジア・アフリカ言語文化研究所教授）

研究分担者：星 泉（東京外国語大学アジア・アフリカ言語文化研究所助手）

研究経費：	平成 13 年度	4,000 千円
	平成 14 年度	3,800 千円
	計	7,800 千円

## 5 研究の成果

### 5.1 多言語処理用 preprocessor の開発

古典学のように多種多様な言語をプレーンテキストに基づいた形で処理するために、適当なタグ付けを行おうとする際に考えられるシステムとしては、現在なら当然 XML に準拠した方式となるであろう。

しかしながら、本研究の元となるシステムの開発は 1992 年から行っていたという事情もあり、また、使用者にできるかぎり負担をかけないシステムを開発するという方針から、入力タグは、XML に類似しながらも、独自の形式となっている。

すなわち、XML においては、開始タグと終了タグが <xxx /xxx> という形で、明確に対応付けされていなければならないのに対して、本システムにおいては、終了タグは > の一種類のみに限定して、対応付けは位置によって判断されるものとした。終了タグのインプットの手間を省くのが、このようにした最大の目的である。実際の利用においては、<sanskrittransliteration some texts in

transliteration <nagari some texts in Devanagari> some texts in transliteration <grantha some texts in Grantha> some texts in transliteration> (この例において下線部は実際にはもっと入力しやすい短いものを用いるが、わかりやすくするために長いタグで表示した)のような2段構成がもっとも一般的で、何重もの入れ子構造を取るような場合はほとんど想定できないため、終了タグの対応は充分に明確であると判断したためである。

実装はPerl5を用いて行っているが、Perl4の時代から開発しているため、わずかの変更でPerl4でも動作するようにすることができる。

開始タグについては、プリアンブル部において、<n>がnagari、<grn>がgrantha、<k>がkannada、<m>がmalayalam、<tml>がtamil、<tr>がtransliterationなどと定義しているのを変更するだけで、使用者が自由に設定可能である。

転写入力方法への対応としては、サンスクリットに関する限り、KH方式・町田方式・折衷方式に対応しているが、要望次第で他のどのような入力方式にも対応可能である。国外の研究者のために、ナーガリー文字についてのみはCSX(Computer Sanskrit Extended)方式への対応も実装してある。

今回の実装で対応しているのは、ローマ字転写、ナーガリー文字、グランタ文字、カンナダ文字、タミル文字、マラヤーラム文字、テルグ文字、ベンガル文字、オリヤー文字、グルムキー文字、チベット文字、クメール文字、タイ文字であり、インド国内で使われている文字については、グジャラーティー文字、アッサム文字(ベンガル文字と異なるのは1文字のみ)を除いて公用語についてはほぼ対応したことになる。今後は、東南アジアのインド系文字を実装していく予定である。

## 5.2 サンスクリット語ローマ字転写表記のPostScriptフォント

現在のインターネット上での電子出版における標準的方法はPDF形式であるが、通常のMetafontによるフォントによっては、埋め込まれるフォントがbitmap方式となり、ファイルサイズが大きくなり、拡大縮小の自由がなくなるなどの多くの問題が生じる。

これに対処するために、TeX出力がPostScriptフォントを利用できるようにする必要がある。今回対応したすべての文字について、Textraceというプログラムを用いて、本来よりはサイズが大きいという難点はあっても、PostScript Type1フォントに変換する作業を行い、pdfに埋め込むことを可能にしている。

ローマ字転写表記に関しては、TeXのT1エンコー

ディングと下部128についてはまったく同じで、上部128について、すべての南アジア諸語を表記できるエンコード体系を作成して、Jörg KnappenのECフォントのソースを組み合わせた形で、表1のようなフォントを作成した。

しかしながら、ローマ字転写表記に関しては、ごく一部の要素を除くと、Acrobat付属のフォントへマッピングする仮想フォント(virtual font)を作成することが可能で、この場合には、フォントの埋め込みがまったく不要になるので、そうした実装も行うこととした。三つのグリフのみがこちらではサポートされない点については、表2を見ていただきたい。

上記のローマ字転写用フォントと同じエンコード体系であって、実際に使用するフォントをTimes Romanの中に含まれるアルファベットとdiacritic記号との組み合わせにするマッピングを行ったvirtual fontを作成することによって、転写記号表記が国際的互換性を持った形のPDFとして制作できるようになった。

古典学との関連で特記しておくと、このエンコード体系では、ヴェーダ語のアクセント表記に関しても、二重アクセントを最初から実装していて、詠唱用の特殊アクセントをのぞいた通常のアクセントは容易に処理できるようになっている。

## 5.3 サンスクリット語ローマ字転写表記のハイフネーション区切り規則

サンスクリット語においてはきわめて長大な複合語が存在するために、散文を引用する場合人間が一つ一つ指示しないかぎり正しいハイフネーション区切りが行われない。これは、論文を書く場合に著者に無用の努力をせまるものであり、論文の生産効率を著しく下げるものであった。

ナーガリー文字等の場合には、音節の切れ目で任意に区切りを挿入することが可能であるが、ローマ字転写の場合には、接頭辞や接尾辞を考慮しながら切れ目を入れることが一般的である。

TeXのhyphenation patternの作成は、区切り位置がマークされている辞書が存在する場合には、patgenというプログラムによって自動的に生成することが可能であるが、そのような辞書は存在していない。そこで、pra, prati, praty, antarなどの接頭辞を直接にパターンとして記述するとともに、r+mなどの区切りとされる可能性の高い子音連続を定義することにした。

これだけでは、充分な定義とならないため、ナーガリー文字の場合に準拠して、音節単位の区切りも許可することにした。これは、不正確な区切りを増やすこ

0-31	'	'	^	~	..	~	.	~	~	-
	'	,	'	,	<	>	"	"	"	«
»		-	—	o	1	J	ff	fi	fl	
॥										
32-47 (記号)	!	"	#	\$	%	&	,	( )		
	*	+	,	-	.	/				
48-57 (数字)	0	1	2	3	4	5	6	7	8	9
58-64 (記号)	:	;	<	=	>	?	@			
65-90 (大文字)	A	B	C	D	E	F	G	H	I	J
	K	L	M	N	O	P	Q	R	S	T
	U	V	W	X	Y	Z				
91-96 (記号)	[	\	]	^	_	'				
97-122 (小文字)	a	b	c	d	e	f	g	h	i	j
	k	l	m	n	o	p	q	r	s	t
	u	v	w	x	y	z				
123-127 (記号)	{		}	~	-					
128-255	K	Ḿ	N	Ŷ	Ž	Z	á	í	ú	é
	ó	á	í	ú	ƒ	ѣ	ł	ṁ	ḥ	ḥ
	ï	ü	ü	ě	ö	ł	ń	ř	ń	,
	ø	ʃ	k	ṁ	ñ	ÿ	ž	ż	ł	ł
	ſ	z	k	č	ğ	đ	š	ć	ý	ǵ
	ą	ä	o	ö	þ	đ	g	ł	z	ń
	ṁ	~	=	-	Ā	Ā	Ā	ī	ī	ī
	Ū	Ū	Ū	Ē	Ē	Ō	Ō	Ŗ	Ŗ	Ŗ
	Ł	R	Ŗ	L	Ł	M	H	Ń	Ń	T
	D	N	Ś	S	R	L	ā	ā	ā	ī
	ĩ	ĩ	ū	ū	ū	ē	ē	ō	ō	ř
	ř	ł	ł	r	ř	ł	ł	m	ḥ	ń
	ñ	t	d	n	s	s	r	ł		

表 1: サンスクリットのローマ字転写用フォント aarm のグリフの一覧表

0-31	'	'	^	-	"	~	°	ˇ	ˇ	-
	'	,	,	<	>	"	"	"	"	«
	»	-	—		■	1	■	ff	fi	fl
	ffi	ffl								
32-47 (記号)	„	!	"	#	\$	%	&	,	(	)
	*	+	,	-	.	/				
48-57 (数字)	0	1	2	3	4	5	6	7	8	9
58-64 (記号)	:	;	<	=	>	?	@			
65-90 (大文字)	A	B	C	D	E	F	G	H	I	J
	K	L	M	N	O	P	Q	R	S	T
	U	V	W	X	Y	Z				
91-96 (記号)	[	\	]	^	-	‘				
97-122 (小文字)	a	b	c	d	e	f	g	h	i	j
	k	l	m	n	o	p	q	r	s	t
	u	v	w	x	y	z				
123-127 (記号)	{		}	~	-					
128-255	Ķ	Ļ	Ņ	Ŗ	Ŗ	Z	ā	ī	ū	é
	ó	á	í	ú	í	í	ł	ł	ł	ł
	ő	ü	ü	ě	ö	ł	ł	ł	ł	ł
	đ	-	k	ṁ	n	ÿ	ž	ż	ż	ż
	š	z	k	č	g	đ	š	’	ý	ǵ
	ä	ä	o	ö	b	d	g	t	z	ñ
	ń	ń	*	-	Ā	Ā	Ā	ī	ī	ī
	Ū	Ū	Ū	Ē	Ē	Ō	Ō	Ŗ	Ŗ	Ŗ
	Ļ	Ŗ	Ŗ	Ļ	Ļ	Ŗ	Ŗ	Ŗ	Ŗ	Ŗ
	Đ	Ņ	Ś	Ś	Ŗ	Ŗ	ā	ā	ā	ī
	ĩ	ĩ	ū	ū	ū	ē	ē	ō	ō	ō
	ř	ř	ř	ř	ř	ř	ř	m	h	ň
	ñ	ñ	đ	n	ś	ś	ř	ř	ř	ř

表 2: サンスクリットのローマ字転写用 virtual フォント ptmr8s のグリフの一覧表

とになり、不満を感じる人が多いかもしれないが、自動的に挿入されるハイフン記号を通常のハイフン記号より短くかつ低く出力することで、著者の責任ではないことが明確になるように考慮した。

完全に満足のいくハイフネーション区切りを行わせたい場合には、音節区切りを許可している部分のパターンを削除して、自動区切りが成功しない場合には人間の手によって入力する方式に頼るべき事になるが、この場合でも労力は大幅に軽減されるはずである。

英語等の他の言語の部分とサンスクリット語の部分で、区切り規則を変更するための基本システムとしては、J. L. Braams の制作による Babel システムを用いて、そのサブパッケージとして動作するように実装を行った。

#### 5.4 カンナダ文字システム

カンナダ文字 TeX システムは、本研究において開発したものではないが、その後の出発点となったものであるので、簡単にその開発経緯を述べておく。

このシステムの基本となるカンナダ文字の Metafont プログラムは、1995 年度に、B.B.Rajapurohit 教授が AA 研の客員教授として来日した際に、福井令氏の協力を得て作成されたものを元としている。基本字形以外の下付き文字などの追加、全体としてのプログラム上のマクロ機能の追加による統一、タイプライタ字形に影響されていた部分の修正などを、家元太郎氏の助言などに基づきながら高島が行ったものが、現在使用しているものである。

256 文字の Metafont の 1 フォントの中に、基本文字、下付き文字、母音 i および e と結合した字形、ā および virāma(母音なしを示す記号) と結合するときの字形、母音記号、その他の記号、余った領域にいくつかの欧文記号、と納められている。

カンナダ文字の構成は、基本文字と母音記号の組み合わせだけを考える場合には、母音記号が基本子音字の左側にくることがないので、TeX と Metafont による ligature 機能によって、前処理 (preprocessor) なしの入力が可能なように思われるかもしれないが、子音結合がある場合には、母音記号の後に下付き文字が来るので、preprocessor の使用が不可欠である。

ハイフネーションのために、音節（子音群+母音+アヌスヴァーラあるいはヴィサルガ）単位によるハイフネーションを Babel システムを用いて実装してある。

カンナダ文字の転写との対応を示す音節表を表 3 に挙げたので、参照していただきたい。

#### 5.5 グランタ文字システム

グランタ文字は、南インドにおいて用いられているサンスクリット表記用の文字で、膨大なサンスクリット資料の表記とともに、南インド刻文資料の表記に不可欠なものである。

しかしながら、それを用いるバラモンたちの数が少なかったことや、欧米の学者たちでこの文字を読める人の数が少なかったことから、ナーガリー文字によってサンスクリット表記を行う慣習が普及してしまった結果、いまだにコンピュータ上でこの文字を処理するシステムはインド国内においてすら開発されていなかった。

また、この文字による印刷は、第 2 次大戦以前にはそれなりに行われたものの、戦後の例は少数のあまり質の良くない出版物に限られており、印刷術としてのあり方も確立されたものとは言えない。結合文字のリストを調べようとしても、わずか数ページのパンフレット的な教科書以外に参考となる書籍が存在していないのが現状である。

そこで、本研究においては、まず戦前において出版された本を仔細に検討することからはじめ、印刷本に必要なグランタ文字の結合文字の分布、3 文字以上の結合の場合の下付き処理のあり方を収集分析する作業を第一に行なった。

こうした分析に基づいて、グランタ文字のフォントを METAFONT プログラムによって制作した。文字の形体としては、あまり印刷術の影響を受けてこなかった歴史を反映して、比較的写本自体に現れる形体と類似したものを探用した。

結合文字に関しては、256 文字までしか一つのフォントセットに入れられないという TeX の限界にあわせて、一つの形体 (sca) については一文字としては実装せずに、ca の下付文字として現れるようにした。この形体は、調査した印刷本のうちのいくつかには現れず、出現頻度が低かったためである。この点を除いては、充分に 20 世紀前半の印刷規範を満たす印刷が可能になったものと思われる。

このフォントの制作によって、従来からあったタミル文字フォントにおいてグランタ文字からの借用文字が表記できずにいたという問題点も、これを導入した新たなタミル文字フォントを制作することで解決することができた。

また、数字に関しては、0 から 9 までのタミル数字を実装したが、100 や 1000 は実装しなかった。これは、20 世紀前半の印刷物においては、10 進法表記が一般的となっていて、これらは用いられていないからである。しかしながら、写本を読む能力を教育するという視点



からは、これらの知識も必要であり、今後の教科書用に異体字リストなどとともに追加のフォントを作成する必要があると思われる。

入力方式としては、ナーガリー文字などと同様にしてシステムに統合したので、タグを変更するだけで、おなじ文章を多様な文字で表記できるようになったが、転写方法に関しては、いくつかの問題が残されている。ドラヴィダ系の言語においては、多くの流音があるため、その表記を容易にしようと配慮した入力転写方式では、ヒンディー語等の反り舌由来の流音と抵触して、どちらの音も覚えやすい位置に配置することが困難なのである。

現在の所は、南インドも北インドも同じ比重で研究しているような研究者はほとんどないので、大部分は共通であっても、南インド中心と北インド中心とで、2種類の入力方法を使い分けるように設定してある。必要とあれば、いつでも両方式の間での変換ができるようなツールを制作してあるので、これで問題が生じることはない。

今回のグラントア文字印字システムの開発は世界で初めてのもので、南インドの碑文研究において困難を感じていた研究者から大いに歓迎されるものとなり、すでにインド人研究者からの提供要請を受けて、配布を開始している。

グラントア文字の転写との対応を示す音節表を表4に挙げたので、参照していただきたい。

## 5.6 ナーガリー文字システム

ナーガリー文字システムのベースとしては、広く用いられている、Frans J. Velthuisによるものを用いた。字形自体はほとんど改変していないが、できる限りpre-processorによる動作の部分を減らしてligature機能に置き換える、コードマッピングの変更、などを行い、フォント名もと変更している。

preprocessor部分の実装においては、子音結合が単独のグリフ（実際に表現される字形）によるものか、部分文字形（半字形）との結合によるものかの、それらの組み合わせによるものかの判断が主要な要素となっている。Velthuisのpreprocessorによっては処理されない、頻度の低い子音結合のいくつかについての処理を追加してある。

ここでも、ハイフネーションのために、音節（子音群+母音+アヌスヴァーラあるいはヴィサルガ）単位によるハイフネーションをBabelシステムを用いて実装してある。

ナーガリー文字の転写との対応を示す音節表を表5に

挙げたので、参照していただきたい。

## 5.7 タミル文字システム

タミル文字システムのベースとなるMetafontとしては、Washington Universityによって開発されたWashington Tamil (WNTML10.MF) を用いた。

今回開発したシステム全体との整合を図るために、コードマッピングの変更、virāma点付き文字の基本字形への採用、ligatureの広範囲の採用、歐文記号の取り込み、基本字体を斜体に変更、いくつかの文字形の修正などを行っている。

音節（子音群+母音+アヌスヴァーラあるいはヴィサルガ）単位によるハイフネーションの実装も一応行つたが、タミル文字においては、ハイフネーション記号が用いられないため、改行位置は、単語あるいは助詞の意味のある切れ目においてしか行ってはいけない。そのため、今後、語末の助詞の連続に関する規則などを用いて、有効な規則の実装が可能になるまで、自動ハイフネーション機能の動作は行わないよう設定している。

タミル文字の転写との対応を示す音節表を表6に挙げたので、参照していただきたい。

## 5.8 マラヤーラム文字システム

マラヤーラム文字システムのベースとしてはJeroen Hellingmanによるものを用いた。当面、正書法改革による現代の比較的結合子音の数の少ない方式にのみ対応することとして、コードマッピングの変更等を行った。

ここでも、ハイフネーションのために、音節（子音群+母音+アヌスヴァーラあるいはヴィサルガ）単位によるハイフネーションをBabelシステムを用いて実装してある。

マラヤーラム文字の転写との対応を示す音節表を表7に挙げたので、参照していただきたい。

## 5.9 ベンガル文字システム

ベンガル文字システムに関しては、Abhijit Dasによるメタフォントソースを用いた。これは、Anshuman PandeyによるBengali TeXに含まれているものであるが、preprocessorに関しては、これを特に参考とせず、ナーガリー文字と共に仕様で実装した。

ベンガル文字によるサンスクリット語テキストを想定していないので、ハイフネーション区切りは今のところ実装していない。

ベンガル文字の転写との対応を示す音節表を表8に挙げたので、参照していただきたい。



表5: ナガリー文字の音節表

Table of Devanāgarī syllables

			a	आ	आ का खा गा घा डा चा छा जा झा अा रा ना ता था एा नी पा फा वा भा मा या रा ला वा शा सा हा डा छा झा
			ā	आ	आ का खा गा घा डा चा छा जा झा अा रा ना ता था एा नी पा फा वा भा मा या रा ला वा शा सा हा डा छा झा
			o	ओ	ओ को खो धो ठो ञो झो
			au	औ	औ को खो धो ठो ञो झो
			ai	ऐ	ऐ के ले गे वे ले ते ले ए ले
			e	ए	ए के ले गे वे ले ते ले ए ले
			i	ई	ई के ले गे वे ले ते ले ए ले
			ū	ऊ	ऊ कु लु गु वु चु छु जु खु अु रु टु
			ū	ऊ	ऊ कु लु गु वु चु छु जु खु अु रु टु
			ī	ई	ई कु लु गु वु चु छु जु खु अु रु टु
			ū	ऊ	ऊ कु लु गु वु चु छु जु खु अु रु टु
			ī	ई	ई कु लु गु वु चु छु जु खु अु रु टु
			ā	आ	आ का खा डा चा छा जा झा अा रा ना ता था एा नी पा फा वा भा मा या रा ला वा शा सा हा डा छा झा
k <sup>h</sup>	g <sup>h</sup>	n	c <sup>h</sup>	j <sup>h</sup>	अ का खा डा चा छा जा झा अा रा ना ता था एा नी पा फा वा भा मा या रा ला वा शा सा हा डा छा झा
s	ś	r	t <sup>h</sup>	ṭ	सा शा डा चा छा जा झा अा रा ना ता था एा नी पा फा वा भा मा या रा ला वा शा सा हा डा छा झा
p	p̄	b	d <sup>h</sup>	ḍ	पा फा वा भा मा या रा ला वा शा सा हा डा छा झा
m	m̄	v	d̄	वा शा सा हा डा छा झा	मा या रा ला वा शा सा हा डा छा झा
y	ṛ	s	ñ	रा ला वा शा सा हा डा छा झा	या रा ला वा शा सा हा डा छा झा
h	ṝ	s̄	ñ̄	हा डा छा झा	हा डा छा झा
l	ks				ला छा झा

Table of Tamil syllables

	a	ā	i	ī	u	ū	e	ē	ai	o	ō	au
k	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஓ	ஔ	ஓள்
க	க	கா	கி	கீ	கு	கூ	கெ	கே	கை	கொ	கோ	கெள்
ந்	ங	ஙா	ஙி	ஙீ	ஙு	ஙூ	ஙெ	ஙே	ஙை	ஙொ	ஙோ	ஙென்
c	ச	சா	சி	சீ	சு	சூ	செ	சே	சை	சொ	சோ	சென்
ந்	ஞ	ஞா	ஞி	ஞீ	ஞு	ஞூ	ஞெ	ஞே	ஞை	ஞொ	ஞோ	ஞென்
t	ட	டா	டி	டீ	டு	டூ	டெ	டே	டை	டொ	டோ	டென்
ந்	ண	ணா	ணி	ணீ	ணு	ணூ	ணெ	ணே	ணை	ணொ	ணோ	ணென்
t	த	தா	தி	தீ	து	தூ	தெ	தே	தை	தொ	தோ	தென்
n	ந	நா	நி	நீ	நு	நூ	நெ	நே	நை	நொ	நோ	நென்
p	ப	பா	பி	பீ	பு	பூ	பெ	பே	பை	பொ	போ	பென்
m	ம	மா	மி	மீ	மு	மூ	மெ	மே	மை	மொ	மோ	மென்
y	ய	யா	யி	யீ	யு	யூ	யெ	யே	யை	யொ	யோ	யென்
r	ர	ரா	ரி	ரீ	ரு	ரூ	ரெ	ரே	ரை	ரொ	ரோ	ரென்
l	ல்	லா	லி	லீ	லு	லூ	லெ	லே	லை	லொ	லோ	லென்
v	வ	வா	வி	வீ	வு	வூ	வெ	வே	வை	வொ	வோ	வென்
ழ	ழ	ழா	ழி	ழீ	ழு	ழூ	ழெ	ழே	ழை	ழொ	ழோ	ழென்
ன்	ன	னா	னி	னீ	னு	னூ	னெ	னே	னை	னொ	னோ	னென்
ற	ற	றா	றி	றீ	று	றூ	றெ	றே	றை	றொ	றோ	றென்
ஷ	ஷ	ஷா	ஷி	ஷீ	ஷு	ஷூ	ஷெ	ஷே	ஷை	ஷொ	ஷோ	ஷென்
s	ஸ்	ஸா	ஸி	ஸீ	ஸு	ஸூ	ஸெ	ஸே	ஸை	ஸொ	ஸோ	ஸென்
j	ஜ்	ஜா	ஜி	ஜீ	ஜு	ஜூ	ஜெ	ஜே	ஜை	ஜொ	ஜோ	ஜென்
ஹ்	ஹ	ஹா	ஹி	ஹீ	ஹு	ஹூ	ஹெ	ஹே	ஹை	ஹொ	ஹோ	ஹென்
க்ஸ்	க்ஸி	க்ஸா	க்ஸி	க்ஸீ	க்ஸு	க்ஸூ	க்ஸெ	க்ஸே	க்ஸை	க்ஸொ	க்ஸோ	க்ஸென்

Note: Recently all irregular ā, o, ō and ai forms are reformed to regular styles. For example, ஞா is now ணா, likewise ஞை is ஞொ, ஜை is ஜை. This reform concerns ண, வ, ன and ஞ.

表 6: タミル文字の音節表

表7：△△△—△△△文字の音韻表

### Table of Malayalam syllables

Table of Bengālī syllables

	a	ā	i	ī	u	ū	়	e	ai	o	au	am	ah
k	অ	আ	ি	ই	উ	ু	়	এ	়াই	ও	ঔ	়াম	়াহ
k <sup>h</sup>	ক	খ	ি	ি	ু	ু	়	ে	কে	থে	থো	ক্ৰ	থ়
g	গ	ঘ	ি	ি	ু	ু	়	ে	গে	গৈ	গো	গ্ৰ	গ়
g <sup>h</sup>	ঁ	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
n	ন	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
c	চ	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
c <sup>h</sup>	ছ	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
j	জ	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
j <sup>h</sup>	ঁ	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
ñ	ঁ	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
t	ত	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
t <sup>h</sup>	ঁ	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
d	দ	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
d <sup>h</sup>	ঁ	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
n	ন	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
p	প	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
p <sup>h</sup>	ঁ	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
b	ব	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
b <sup>h</sup>	ঁ	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
m	ম	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
y(j)	য়(জ)	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
ঁ	ঁ	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
r	র	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
l	ল	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
v(b)	ব	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
s	শ	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
s <sup>h</sup>	ঁ	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
h	হ	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
r <sup>h</sup>	ঁ	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ
ks	ক্ষ	ঁ	ি	ি	ু	ু	়	ে	়ে	়ৈ	়ো	়্ৰ	়ঁ

表 8: ベンガル文字の音節表

## 5.10 オリヤー文字システム

Jeroen Hellingman によるオリヤー文字メタフォントを用いて、ベンガル文字と同様の仕様で実装した。現在のところ、ハイフネーション区切りは実装していない。

オリヤー文字の転写との対応を示す音節表を表 9に挙げたので、参照していただきたい。

## 5.11 テルグ文字システム

Lakshmi V. S. Mukkavilli による TeluguTeX を基に実装した。テルグ文字の子音結合などについての多くの TeX マクロも、彼によるものをそのまま採用している。

テルグ文字の転写との対応を示す音節表を表 10に挙げたので、参照していただきたい。

## 5.12 グルムキー文字システム

Anshuman Pandey による Gurmukhi for TeX に含まれている Hardip Singh Pannu 氏による True Type フォントから TTF2MF によってコンバートされた pun10.mf をそのまま用いて実装した。

そのため、結合文字などの処理をすべて *preprocessor* の側で処理している。子音結合が 13 種、子音重複 (addak) が 23 種、鼻音子音重複 (tippii) が 2 種、さらに母音記号のカーニングなどを *preprocessor* で処理している。

グルムキー文字の転写との対応を示す音節表を表 11に挙げたので、参照していただきたい。

## 5.13 タイ文字システム

研究分担者の峰岸の指導の下に *metafont* を作成した。タイ文字の概要および入力方法に関する問題等については、別稿「クメール語およびタイ語の翻字について」に記述したので、技術的な問題についてのみ概略を述べる。

タイ文字においては、下付き文字が存在しないので、処理上の問題は、上部に付加される母音記号と声調記号のみである。これは当初 TeX のマクロを用いて処理していたが、最終的には、高さや位置の異なる複数の記号をフォントの中に用意することによって *ligature table* だけで処理することを可能とした。

*preprocessor* の処理としては、21 の母音が少数のアルファベットで表記される問題と、前置される母音記号の転置などが中心で、比較的容易に行えるものである。

## 5.14 クメール文字システム

クメール文字に関しては、当初の研究計画には含めていなかったが、Yannis Haralambous によるシステムが存在するので、それを統合することにした。

*Metafont* および、複雑な文字構成を実現する TeX マクロについては、Haralambous のものをそのまま用い、入力方法を研究分担者峰岸の方法（「クメール語およびタイ語の翻字について」参照）に変更して、*preprocessor* をそれに適応する形で新規に作成した。

Table of Oriyā syllables

	a	ā	i	ī	u	ū	ṛ	ṝ	l	ī	e	ai	o	au	am̄	ah̄
k	ଅ	ଆ	ଇ	ଇ	ଉ	ଉ	ର୍	ର୍ଣ	ଲ	ଈ	ଏ	ାଇ	ଓ	ଓଇ	ଆମ	ଆହ
k <sup>h</sup>	କ	କା	କି	କି	କୁ	କୁ	କୁ	କୁ	କୁ	କି	କେ	କେଇ	ଲୋ	କୋ	କଂ	କଃ
g	ଖ	ଖା	ଖି	ଖି	ଖୁ	ଖୁ	ଖୁ	ଖୁ	ଖୁ	ଖି	ଖେ	ଖେଇ	ଖୋ	ଖୋ	ଖଂ	ଖଃ
g <sup>h</sup>	ଗ	ଗା	ଗି	ଗି	ଗୁ	ଗୁ	ଗୁ	ଗୁ	ଗୁ	ଗି	ଗେ	ଗେଇ	ଗୋ	ଗୋ	ଗଂ	ଗଃ
n̄	ଞ	ଞା	ଞି	ଞି	ଞୁ	ଞୁ	ଞୁ	ଞୁ	ଞୁ	ଞି	ଞେ	ଞେଇ	ଞୋ	ଞୋ	ଞଂ	ଞଃ
c	ଚ	ଚା	ଚି	ଚି	ଚୁ	ଚୁ	ଚୁ	ଚୁ	ଚୁ	ଚି	ଚେ	ଚେଇ	ଚେ	ଚେ	ଚଂ	ଚଃ
c <sup>h</sup>	ଛ	ଛା	ଛି	ଛି	ଛୁ	ଛୁ	ଛୁ	ଛୁ	ଛୁ	ଛି	ଛେ	ଛେଇ	ଛୋ	ଛୋ	ଛଂ	ଛଃ
j	ଜ	ଜା	ଜି	ଜି	ଜୁ	ଜୁ	ଜୁ	ଜୁ	ଜୁ	ଜି	ଜେ	ଜେଇ	ଜୋ	ଜୋ	ଜଂ	ଜଃ
j <sup>h</sup>	ଜ	ଜା	ଜି	ଜି	ଜୁ	ଜୁ	ଜୁ	ଜୁ	ଜୁ	ଜି	ଜେ	ଜେଇ	ଜୋ	ଜୋ	ଜଂ	ଜଃ
ନୀ	ନ	ନା	ନି	ନି	ନୁ	ନୁ	ନୁ	ନୁ	ନୁ	ନି	ନେ	ନେଇ	ନୋ	ନୋ	ନଂ	ନଃ
t	ତ	ତା	ତି	ତି	ତୁ	ତୁ	ତୁ	ତୁ	ତୁ	ତି	ତେ	ତେଇ	ତୋ	ତୋ	ତଂ	ତଃ
t <sup>h</sup>	ତ	ତା	ତି	ତି	ତୁ	ତୁ	ତୁ	ତୁ	ତୁ	ତି	ତେ	ତେଇ	ତୋ	ତୋ	ତଂ	ତଃ
d	ଦ	ଦା	ଦି	ଦି	ଦୁ	ଦୁ	ଦୁ	ଦୁ	ଦୁ	ଦି	ଦେ	ଦେଇ	ଦୋ	ଦୋ	ଦଂ	ଦଃ
d <sup>h</sup>	ଦ	ଦା	ଦି	ଦି	ଦୁ	ଦୁ	ଦୁ	ଦୁ	ଦୁ	ଦି	ଦେ	ଦେଇ	ଦୋ	ଦୋ	ଦଂ	ଦଃ
n̄	ନ	ନା	ନି	ନି	ନୁ	ନୁ	ନୁ	ନୁ	ନୁ	ନି	ନେ	ନେଇ	ନୋ	ନୋ	ନଂ	ନଃ
p	ପ	ପା	ପି	ପି	ପୁ	ପୁ	ପୁ	ପୁ	ପୁ	ପି	ପେ	ପେଇ	ପୋ	ପୋ	ପଂ	ପଃ
p <sup>h</sup>	ପ	ପା	ପି	ପି	ପୁ	ପୁ	ପୁ	ପୁ	ପୁ	ପି	ପେ	ପେଇ	ପୋ	ପୋ	ପଂ	ପଃ
b	ବ	ବା	ବି	ବି	ବୁ	ବୁ	ବୁ	ବୁ	ବୁ	ବି	ବେ	ବେଇ	ବୋ	ବୋ	ବଂ	ବଃ
b <sup>h</sup>	ବ	ବା	ବି	ବି	ବୁ	ବୁ	ବୁ	ବୁ	ବୁ	ବି	ବେ	ବେଇ	ବୋ	ବୋ	ବଂ	ବଃ
m	ମ	ମା	ମି	ମି	ମୁ	ମୁ	ମୁ	ମୁ	ମୁ	ମି	ମେ	ମେଇ	ମୋ	ମୋ	ମଂ	ମଃ
y(j)	ଯ	ଯା	ଯି	ଯି	ଯୁ	ଯୁ	ଯୁ	ଯୁ	ଯୁ	ଯି	ଯେ	ଯେଇ	ଯୋ	ଯୋ	ଯଂ	ଯଃ
ୟ	ର	ରା	ରି	ରି	ରୁ	ରୁ	ରୁ	ରୁ	ରୁ	ରି	ରେ	ରେଇ	ରୋ	ରୋ	ରଂ	ରଃ
ଲ	ଲ	ଲା	ଲି	ଲି	ଲୁ	ଲୁ	ଲୁ	ଲୁ	ଲୁ	ଲି	ଲେ	ଲେଇ	ଲୋ	ଲୋ	ଲଂ	ଲଃ
v(w)	ଲ	ଲା	ଲି	ଲି	ଲୁ	ଲୁ	ଲୁ	ଲୁ	ଲୁ	ଲି	ଲେ	ଲେଇ	ଲୋ	ଲୋ	ଲଂ	ଲଃ
b̄	ବ	ବା	ବି	ବି	ବୁ	ବୁ	ବୁ	ବୁ	ବୁ	ବି	ବେ	ବେଇ	ବୋ	ବୋ	ବଂ	ବଃ
s̄	ଶ	ଶା	ଶି	ଶି	ଶୁ	ଶୁ	ଶୁ	ଶୁ	ଶୁ	ଶି	ଶେ	ଶେଇ	ଶୋ	ଶୋ	ଶଂ	ଶଃ
s	ଶ	ଶା	ଶି	ଶି	ଶୁ	ଶୁ	ଶୁ	ଶୁ	ଶୁ	ଶି	ଶେ	ଶେଇ	ଶୋ	ଶୋ	ଶଂ	ଶଃ
ହ	ହ	ହା	ହି	ହି	ହୁ	ହୁ	ହୁ	ହୁ	ହୁ	ହି	ହେ	ହେଇ	ହୋ	ହୋ	ହଂ	ହଃ
ର̄	ର	ରା	ରି	ରି	ରୁ	ରୁ	ରୁ	ରୁ	ରୁ	ରି	ରେ	ରେଇ	ରୋ	ରୋ	ରଂ	ରଃ
ତ̄	ତ	ତା	ତି	ତି	ତୁ	ତୁ	ତୁ	ତୁ	ତୁ	ତି	ତେ	ତେଇ	ତୋ	ତୋ	ତଂ	ତଃ
କ̄	କ	କା	କି	କି	କୁ	କୁ	କୁ	କୁ	କୁ	କି	କେ	କେଇ	କୋ	କୋ	କଂ	କଃ
କ୍ଷ	କ୍ଷା	କ୍ଷି	କ୍ଷି	କ୍ଷି	କ୍ଷୁ	କ୍ଷୁ	କ୍ଷୁ	କ୍ଷୁ	କ୍ଷୁ	କ୍ଷି	କ୍ଷେ	କ୍ଷେଇ	କ୍ଷୋ	କ୍ଷୋ	କଂ	କଃ

表9: オリヤー文字の音節表

Table of Telugu syllables

	a	ā	i	ī	u	ū	ṛ	ṝ	!	ī!	e	ē	ai	o	ō	au	am	ah
k	క	କ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
k̄	କ	କ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
g	ଗ	ଗ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
ḡ	ଗ	ଗ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
c	ଚ	ଚ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
c̄	ଚ	ଚ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
l̄	ଲ	ଲ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
j̄	ଜ	ଜ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
j̄̄	ଜ	ଜ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
p̄	ପ	ପ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
b̄	ବ	ବ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
m̄	ମ	ମ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
ȳ	ଯ	ଯ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
r̄	ର	ର	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
l̄	ଲ	ଲ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
v̄	ଵ	ଵ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
n̄	ନ	ନ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
s̄	ଶ	ଶ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
h̄	ହ	ହ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
l̄̄	ଲ	ଲ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
v̄̄	ଵ	ଵ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
n̄̄	ନ	ନ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
s̄̄	ଶ	ଶ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ
h̄̄	ହ	ହ	ି	ି	ୁ	ୁ	ର	ର	ା	ା	େ	େ	ାଇ	ୋ	ୋ	ାୟ	ାମ	ାଃ

表 10: テルグ文字の音節表

## Table of Gurumukhī syllables

表 11: グルムキー文字の音節表

## 5.15 チベット文字システム

研究分担者星の指導の下に、metafontを作成するとともに、チベット文字の転写方法として最も一般的なワイリー方式(サンスクリット転写に関しては拡張ワイリー方式)の入力方法に対応するpreprocessorを実装した。

実装したものを以下に表で示す。

チベット語の30基字

<i>ka</i>	<i>kha</i>	<i>ga</i>	<i>nga</i>
ཀ	ཁ	ဂ	ኅ
<i>ca</i>	<i>cha</i>	<i>ja</i>	<i>nya</i>
ཅ	ཆ	ጃ	ኋ
<i>ta</i>	<i>tha</i>	<i>da</i>	<i>na</i>
ታ	ߠ	ኅ	ኅ
<i>pa</i>	<i>pha</i>	<i>ba</i>	<i>ma</i>
ፖ	ፖ	ባ	ማ
<i>tsa</i>	<i>tsha</i>	<i>dza</i>	<i>wa</i>
ཙ	ཙ	ኅ	ዋ
<i>zha</i>	<i>za</i>	' <i>a</i>	<i>ya</i>
ڙ	ڙ	አ	ያ
<i>ra</i>	<i>la</i>	<i>sha</i>	<i>sa</i>
ڒ	ڸ	ሻ	ሻ
<i>ha</i>	<i>a</i>		
ܹ	ܻ		

チベット語の母音記号

<i>a</i>	<i>i</i>	<i>u</i>	<i>e</i>	<i>o</i>
ܻ	ܿ	ܼ	ܵ	ܽ

語頭の母音

<i>a</i>	<i>i</i>	<i>u</i>	<i>e</i>	<i>o</i>
ܻ	ܿ	ܼ	ܵ	ܽ

サンスクリット表記

<i>ā</i>	<i>ī</i>	<i>ū</i>	<i>r̥</i>	<i>l̥</i>	<i>ī̥</i>	<i>ai</i>	<i>au</i>	<i>am̥</i>	<i>ah̥</i>
ܻ	ܿ	ܼ	ܰ	ܲ	ܿ	ܾ	ܻ	ܻ	ܽ

*gha dha bha jha*

ܻ ܼ ܻ ܿ

*ta tha da dha na sa kṣa*

ܰ ܼ ܰ ܿ ܻ ܻ ܺ

ya 有足字(7種)

<i>kya</i>	<i>khya</i>	<i>gya</i>	<i>pya</i>	<i>phyā</i>	<i>bya</i>	<i>m̥ya</i>
ܻ	ܿ	ܻ	ܻ	ܿ	ܻ	ܻ

ra 有足字(14種)

<i>kra</i>	<i>khra</i>	<i>gra</i>	<i>tra</i>	<i>thra</i>	<i>dra</i>	<i>nra</i>
ཀ	ཁ	ግ	ጥ	ທ	ດ	ኅ
<i>pra</i>	<i>phra</i>	<i>bra</i>	<i>mra</i>	<i>shra</i>	<i>sra</i>	<i>hra</i>
ፖ	ፖ	ብ	ሙ	ሃ	ሮ	ኩ

la 有足字(6種)

<i>kla</i>	<i>gla</i>	<i>zla</i>	<i>bla</i>	<i>sla</i>	<i>rla</i>
ཀ	ግ	ዘ	ብ	ሸ	ሩ

wa 有足字(16種)

<i>kwa</i>	<i>khwa</i>	<i>gwa</i>	<i>cwa</i>	<i>tswa</i>	<i>tshwa</i>	<i>nywa</i>	<i>zhwa</i>
ཀ	ཁ	ግ	ቁ	ቃ	ቃ	ኋ	ቍ
<i>shwa</i>	<i>twa</i>	<i>dwa</i>	<i>zwa</i>	<i>swa</i>	<i>lwa</i>	<i>rwa</i>	<i>hwa</i>
ሂ	ጥ	ወ	ወ	ሮ	ሮ	ሁ	ወ

ra 有頭字(12種)

<i>rka</i>	<i>rtsa</i>	<i>rta</i>	<i>rga</i>	<i>rja</i>	<i>rdza</i>
ሩ	ሩ	ሩ	ሩ	ሩ	ሩ
<i>rda</i>	<i>rba</i>	<i>rnga</i>	<i>rnya</i>	<i>rna</i>	<i>rma</i>
ሩ	ሩ	ሩ	ሂ	ሩ	ሁ

la 有頭字(10種)

<i>lka</i>	<i>lca</i>	<i>lta</i>	<i>lpa</i>	<i>lga</i>
ሉ	ሉ	ሉ	ሉ	ሉ
<i>lja</i>	<i>lda</i>	<i>lba</i>	<i>lnga</i>	<i>lha</i>
ሉ	ሉ	ሉ	ሉ	ሉ

sa 有頭字(11種)

<i>ska</i>	<i>stsā</i>	<i>sta</i>	<i>spa</i>	<i>sga</i>	<i>sda</i>
ሂ	ሂ	ሂ	ሂ	ሂ	ሂ
<i>sba</i>	<i>sngā</i>	<i>snyā</i>	<i>sna</i>	<i>sma</i>	
ሂ	ሂ	ሂ	ሂ	ሂ	

有頭有足字(16種)

<i>rgwa</i>	<i>rtswa</i>	<i>rtya</i>	<i>rkyā</i>	<i>rmya</i>
ሩ	ሩ	ሩ	ሩ	ሩ
<i>skra</i>	<i>sgra</i>	<i>snra</i>	<i>smra</i>	<i>spra</i>
ሂ	ሂ	ሂ	ሂ	ሂ

<i>sgya</i>	<i>skyā</i>	<i>smyā</i>	<i>spyā</i>	<i>sbyā</i>
ሂ	ሂ	ሂ	ሂ	ሂ

## 2 有足字(2種)

*grwa* *phywa*

ግርዋ ፎይዋ

## チベット語の句読点

· | ॥ ॥ ॥ ॥ ॥ ॥ ॥ ॥

## チベット語の数字

0 1 2 3 4 5 6 7 8 9  
၀ ၁ ၂ ၃ ၄ ၅ ၆ ၇ ၈ ၉

以上の表において示されていない要素として、以下のようなものがある。

### 1. 音節の強制分割

*gyang* と書くと གྙ ཉ ཉ ཉ となるが、 གྙྙ ཉ ཉ ཉ としたい場合には *g-yang* と書けばよい。

### 2. サンスクリットの結合子音

通常のチベット語に現れないサンスクリットの結合子音を表記するには、 + を用いる。

*bud+dha* *a'nan+da* *in+dra*  
ବୁଦ୍ଧା ଅନନ୍ତ ଇନ୍ଦ୍ରା

### 3. tsheg の禁止

チベット文字の音節分割を示す *tsheg* 、 ་ は、 空白等の前で自動的に現れるように設定してあるので、字母表などでその出現を禁止するためには、 . を後ろにつける必要がある。 *ka* では ອ ອ となるが、 *ka*. とすれば ອ ອ となって、 *tsheg* は現れない。

## 6 多言語処理システム(**T<sub>E</sub>X Preprocessor**)の使い方

### 6.1 はじめに

**T<sub>E</sub>X**は、当初数学用に開発されたものであるが、柔軟な実装形態と高度な組み版への対応から、多言語対応の応用が数多くなされてきた。インド系文字の実装に関しては、母音記号が前置される問題等の解決が必要で、大部分の場合前処理用のプログラム(preprocessor)を媒介として行われている。これは、多くの言語を処理しようとする場合にはきわめて面倒となる。一つのpreprocessorで多言語の処理を可能にするシステムも存在するが、プログラムに固定されているため、個々の要求にあわせてカスタマイズするには困難が伴っている。

本システムは、preprocessorにPerlのスクリプトを使うことによって柔軟かつ拡張性の高い処理を可能にするものである。処理速度がそれほど高くないという問題点はパソコンの速度の向上によって、現在ではほぼ解消されたと言えるであろう。

### 6.2 インストール

本システムは、**T<sub>E</sub>X**とPerlさえ動けば、どのようなプラットフォームの上でも動作する。Mac OSに関しては、日本語対応のMacPerlが、64Kbytesまでのスクリプトしか読み込めないという制限を持っているため、すべての機能をもったバージョンは動作しない。Mac OS Xは実質がBSD Unixであるから、Unixと同様に動作するが、使用した経験がないので具体的な使用方法等の指示は専門の参考書などに仰ぎたい。Unixの場合にはほとんどのシステムで**T<sub>E</sub>X**もPerlもインストール済みであるはずであるから、以下ではWindowsでのインストールについて簡単に解説する。Windows3.1あるいはDOSでの使用も可能であるが、一般的とは言えないので、ここでは割愛する。

なお、以下の奥村氏のサイトが様々な情報の入手の入り口としては最適であろう。

<http://www.matsusaka-u.ac.jp/~okumura/texfaq/>

#### 6.2.1 **T<sub>E</sub>X**のインストール

現在Windows対応の日本語**T<sub>E</sub>X**で常に最新のバージョンを維持していただいているのが、近畿大学九州工学部の角藤亮氏のものである。以下の氏のホームページに必要な情報とパッケージが揃っている。

<http://www.fsci.fuk.kindai.ac.jp/~kakuto/win32-ptex/web2c74.html>

同じパッケージは以下のサイトでミラーされている。

<ftp://akagi.ms.u-tokyo.ac.jp/pub/TEx/win32/>  
<ftp://ftp.t.ning.gr.jp/pub/text/TEx/ptex-win32/current/>  
<ftp://ftp.u-aizu.ac.jp/pub/tex/ptex-win32/current/>  
<ftp://ftp.eng.kagoshima-u.ac.jp/pub/TEx/>  
<ftp://ftp.eng.kagoshima-u.ac.jp/pub/TEx/ptex-win32/current/>

氏の説明を読めば、インストールはきわめて簡単であるが、氏のパッケージ群のうちの何をインストールしなければいいのかについて以下に説明する。

氏のパッケージ群は、以下のものからなる。

#### (a) 最小インストール

※ latex.tar.gz  
※ platex.tar.gz  
※ mftools.tar.gz  
※ ptex-3.0.5-w32.tar.gz  
※ web2c-7.4.4-lib.tar.gz  
※ web2c-7.4.4-w32.tar.gz

#### (b) 標準インストール

(a)に以下を付け加えます。

※ dvipsk-w32.tar.gz  
※ ltxpkgs.tar.gz  
※ txpx-pazofonts.tar.gz  
timesnew.tar.gz  
manual.tar.gz  
※ makeindex-w32.tar.gz  
oldformat.tar.gz  
oldinputs.tar.gz  
pdftex-w32.tar.gz  
※ t1fonts.tar.gz  
※ dvipdfm-w32.tar.gz  
※ vf-a2bk.tar.gz

#### (c) フルインストール

(a), (b)に以下を付け加えます。

context.tar.gz  
context-doc.tar.gz  
※ dvi2ps-3.2j-w32.tar.gz  
etex-2.1-w32.tar.gz  
jdv2kps-4.0-w32.tar.gz  
jtex-1.9.1-w32.tar.gz  
omega-1.23.2-w32.tar.gz  
omegaj-w32.tar.gz  
plain2-2.53-w32.tar.gz  
t1utils-w32.tar.gz  
texinfotools-w32.tar.gz  
tiff2png-w32.tar.gz

ttf2pk-w32.tar.gz  
txtutil.tar.gz  
vf-n2bk.tar.gz  
cweb-w32.tar.gz  
dvitoools-w32.tar.gz  
nts-loader-w32.tar.gz  
tex4htk-w32.tar.gz  
psutils-w32.tar.gz

上記のうちで、※印をつけたものが、一応必須でインストールしなければならないものである。dvi2psを必須としているのは、dvipsの動作において文字間隔が乱れる場合があるからである。

これらを、インストールした後に、私のホームページから、必要なパッケージを取ってきて、適切なフォルダにコピーすれば良いが、私の配布するCD-ROMから、展開されたイメージをそのままCドライブのルートにコピーするのがもっとも簡単であろう。

## 6.2.2 Ghostscript のインストール

Ghostscriptとは、フリーのPostscriptインターフリタであるが、 $\text{\TeX}$ でPostscriptフォントを利用しようとするときには必須である。そのインストールについては以下のurlにある説明が詳しくて有用である。

<http://auermath.aichi-edu.ac.jp/~khotta/ghost/index.html>

現時点での最新バージョンは、8.00であるが、通常の使用のためなら、6.0以降のどのバージョンでも問題なく動くであろう。

要点は、

<ftp://ftp.t.ring.gr.jp/pub/text/CTAN/support/ghostscript/aladdin/current/gs800/gs800w32/exe>  
<ftp://ftp.t.ring.gr.jp/pub/text/CTAN/support/ghostscript/ghostgum/gsv43w32.exe>

あるいは、私のパッケージの\$gsフォルダにあるgs800w32.exeとgsv43w32.exeとを順に起動して指示に従ってインストールすることである。

日本語化されたバージョンは、

<ftp://ftp.t.ring.gr.jp/pub/text/\TeX/ptex-win32/gs/gs800-j-wapi.zip>

および私のパッケージの同じフォルダにあり、gs800の上に、上書きすればよい。

<ftp://ftp.ascii.co.jp/pub/\TeX/ascii-ptex/dos/unz512xN.exe>

を実行して得られる、unzip.exeあるいは私のパッケージのtoolsフォルダに入っているunzip.exeを用いる場

合、gs800-j-wapi.zipをc:\$gsフォルダにコピーしてから、そのフォルダにおいてunzip -o gs800-j-wapiを実行すればよい。

## 6.2.3 Dviout for Windows のインストール

大島利雄氏による、非常に高性能なDVIドライバであるDviout for Windowsは、私のパッケージでは\$usr\$local\$dvioutに納められている。Dvioutは、常に改良が続けられているので、<http://akagi.ms.u-tokyo.ac.jp/dvitest.html>から最新版入手するのが好ましい。

インストールについては、

[http://akagi.ms.u-tokyo.ac.jp/tex\\_dvioutw.html](http://akagi.ms.u-tokyo.ac.jp/tex_dvioutw.html)

に詳しく述べられているが、私のパッケージを使う場合、必要な設定を保存したinstall.parファイルが同梱されているので、最初に起動したときに、

Install fundamental parameters?

と尋ねられる際に、noと答え、起動後に、

Option -> Read parameters

を選択して、install.parファイルを読み込めば完了する。

## 6.2.4 WinShell のインストール

WinShellは、統合 $\text{\TeX}$ 環境を可能とするエディタを中心としたフリーのプログラムである。 $\text{\TeX}$ に関するすぐれた入門書である『[改定版] LaTe $\chi$ 2ε 美文書作成入門』(奥村晴彦著 技術評論社、2000年)が推奨しているところから、私のシステムもこの上で簡単に操作することが可能なようにしてある。

同書所収のものよりも新しいbeta2.2.2版を\$winshellに納めてある。最初にwinshell221.exeを起動して指示に従ってインストールする。続いてws222beta.zipを解凍して、すべてのファイルをwinshellをインストールしたフォルダに上書きコピーする。一度winshellを起動すると、“Choose Language”と訊いてくるので、英語を選択する。そのまま終了してから、\$winshell\$custumにある、winshell.iniとwinshelltools.bmpをC:\$WINDOWS\$Application Data\$WinShell(MEの場合)などのフォルダ(システムによって異なるので不明の場合はwinshell.iniを検索すればよい)に上書きコピーする。これでインストールは完了である。

## 6.2.5 Perl のインストール

ShiftJIS環境で普通に日本語を用いる場合にはいまだActivePerl5.22を用いる方が容易であると思われる

<http://www.activestate.com/Products/ActivePerl/download.html>

の旧版の場所 (Need an older version of ActivePerl?) すなわち

<http://www.activestate.com/download/ActivePerl/>  
APi522e.exe

から入手するか、私のパッケージの ActivePerl フォルダから起動して適当なディレクトリ（例えば c:\perl）にインストールする。

日本語化されたバージョンは

<http://www.shonanblue.ne.jp/~kipp/perl/jperl/>  
index.html

から、jperl522.zip として入手できる。私のパッケージには、これと、それを解凍した、jperl522.tar.gz と jinstall.bat が同じフォルダに入っているので、そのフォルダで jinstall.bat を実行すると日本語版がインストールされる。

Windows 95/98/ME で初めてインストールする場合はそれぞれのインストールで Windows 95/98/ME を restart させる必要があるが、デフォルトのフォルダにインストールする予定で、最初に次のサブセクションの環境設定を行っている場合にはその必要は、最初の一回の再起動だけになる。

### 6.2.6 環境変数等の設定

以上のプログラムをいつでも実行できるようにするために、c:\autoexec.bat に以下のような設定を追加する必要がある。

```
PATH C:\perl\bin\MSWin32-x86-object;  
C:\perl\bin;c:\usr\local\bin;c:\gs\gs8.00\bin;  
c:\gs\gs8.00\lib
```

そのほかの環境変数の設定としては

```
set TMP=c:\tmp
```

として、c:\tmp というフォルダを作成しておくことが windows95 では必要なことがある。

Ghostscriptについては、c:\gs 以外の場所にインストールした場合は以下の環境変数の設定が必要である。

```
SET GS_LIB=c:\gs\gs8.00  
    \lib;c:\gs\gs8.00\kanji;c:\gs\fonts
```

古いバージョンでは、C ドライブ以外に TeX をインストールする場合は以下の環境変数の設定が必要であった。

```
set TEXMFMAIN=d:/usr/local/share/texmf
```

最新バージョンでは、不要となっている。

また、perl による preprocessor にワイルド・カードで \*.txtなどを処理させたい場合には、私のパッケージの activePerl にある wild.pm を c:\perl\lib にコピーした

上で

```
set PERL5OPT=-Mwild
```

も追加する必要がある。

私のパッケージをまるごとコピーしたのではなく、私のサイトから必要なパッケージを入手して使用する場合には、標準の L<sup>A</sup>T<sub>E</sub>X のパッケージに対して以下の変更を加える必要がある。

#### 1. babel の上書き

```
texmf/tex/generic/babel/hyphen.cfg
```

は、たとえば hyphen\_o.cfg などに名前を変え、私のパッケージのものと置き換える。Babel の中核的部分であるこのプログラムは、ハイフネーションパターンファイルの読み込みの際に、単なるグルーピングをして読み込んでおり、改行によって終了したパラグラフ以外の、パラグラフの途中での異なったハイフネーションパターンを処理しないために、この変更が必要である。

また、最近の pTeX の配布は、誤った babel のハイフネーションの読み込みを避けるため

```
texmf/ptex/platex/config/hyphen.cfg
```

が存在しているので、このファイルの名前も適当に変更しておく必要がある。

```
texmf/tex/generic/config/language.dat
```

も、たとえば language\_o.dat などに名前を変えて、私のパッケージのものと置き換える。

これらを行った上で、

```
tex --ini latex.ltx
```

```
ptex --ini platex.ltx
```

を行って、latex.fmt および platex.fmt を新たに作成して

```
texmf\web2c
```

にコピーしてハイフネーションが有効になるようにする必要がある。

fmtutil.exe を使用する場合は、

```
texmf/web2c/fmtutil.cnf
```

で、platex についても language.dat を使用するように設定を追加することを行い、fmtutil を用いて latex.fmt platex.fmt を更新する。

また、もしも C ドライブ以外にインストールした場合には

```
texmf/web2c/texmf.cnf
```

の中の c: をそれにあわせて変更する必要にも注意すること。

現在のプリンタの解像度からすると

```
C:\usr\local\share\texmf\dvi2ps\config
```

```
define mode cx
```

を

```
define mode canonex  
に変更する。  
C:¥usr¥local¥share¥texmf¥dvips¥config¥  
                 config.ps
```

の  
\* Default resolution.  
D 300

を  
\* Default resolution.  
D 600

に変更する等も美しく印刷するためには必要である。  
dvipsv を使って日本語フォントを埋め込んだ pdf を  
作成したい場合には

```
C:¥usr¥local¥share¥texmf¥dvips¥vfontcap  
の
```

c¥:/winnt/fonts/msmincho.ttc:  
を

c¥:/windows/fonts/msmincho.ttc:  
等に変更することも、 windows 95/98/ME の場合には  
必要である。

## 6.2.7 多言語システムのインストール

私のパッケージをそのままコピーした以外の場合は、  
<http://www3.aa.tufs.ac.jp/~tjun/software/tex/>  
以下の個別パッケージを入手して、必要な場所に置く  
必要がある。metafont ファイルは、fonts/source/public  
以下に適当な名前を付けたフォルダに置く。pk フォン  
トについては、mkpk.awk という awk のスクリプトが  
添付してあるので、付属する .def ファイルを各自の環  
境に合わせてカスタマイズしてから実行すれば、一挙  
にフォントを生成することができる。dviout のフォン  
ト作成の自動実行によっても可能であるが、メタフォ  
ントのソースファイルの存在するフォルダで行うと正  
しく働かないことに注意する必要がある。

スタイルファイルは、tex/latex/の下ならどこでもよ  
いが、私の場合は、tex/latex/babelx というフォルダの  
下に個別の名前を付けて置いてある。

preprocessor である to\_tex.pl については、バッチファ  
イルによって起動してもよいし、角藤氏のパッケージ  
に含まれている perlexec.exe を使い、それを to\_tex.exe  
という名前に変更してから to\_tex.pl と同じフォルダで  
path の通ったところに置いておくのでもよい。

## 6.3 多言語システムの使い方

### 6.3.1 基本的な使い方

以下に述べる入力方法等はすべてカスタマイズ可能  
であるが、とりあえず筆者の使っている方法について

の概略を説明することにする。

多言語の切り替えは、<xx > という形のタグ（し  
るし）をつけることによって行う。基本的な設定は以  
下のようになっている。

ローマ字転写	<tr >
ナーガリー	<n >
タミル	<tml >
グラント	<grn >
カンナダ	<k >
マラヤーラム	<mm >
テルグ	<tl >
ベンガリー	<bn >
オリヤー	<or >
グルムキー	<gm >
チベット	<tib >
タイ	<tai >
クメール	<kh >
英語	<e >
フランス語	<f >
ドイツ語	<d >

入る記号 <tr <n 等の後には必ず空白を置かなければ  
ならない。必要に応じてどのような新しい言語や文字  
に対応できるようにするための仕様であり、この空白  
は出力には現れない。閉じる記号は文字列に直接続い  
てかまわない。<tr input strings> のように入力すれば  
よい。< や > 自体を入力したいときは、その前に ¥ を  
つけてやればよい。¥< ¥> のようにである。ただし、  
¥記号自体の二重チェックを行っていないので、「号」  
のような第2バイト目とエスケープ記号が漢字として  
意味を持つ文字になってしまふ漢字の後ではこれが有  
効でない場合があるので、そうした場合には空白を追  
加したりする必要がある。

そのほかに、デフォルトの設定を行うことが出来  
る。もっとも良く使うモードを <のみで使用するよう  
に定義できる。通常はローマ字転写としてあるが、カ  
スタマイズの項目で説明するやり方で設定すればど  
のようなモードを設定しても良い。

また、<txcmd > というモードが TeX の命令を入力す  
るために用意してある。たとえば、¥footnote という  
命令は、<tr > の中では、oo が オ に変換されるためそ  
のままでは使えないでの、そうしたときに使用すれば  
よい。もちろん、¥footnote のように良く使う命令は、  
¥let¥fn ¥footnote などで命令の方を変更しておく方が  
良いことが多い。

こうした <xx> によるモードの変更は、何重にでも入れ子にすることが出来る。たとえば、<f Text du <mahaabhaarata> par tel quel> のようにである。

標準の設定では、<tr <f 等はイタリック体を出力する。立体を出力するためには、<tr: のようにコロンを付ける。このカスタマイズについても後述するが、起動オプションに -i を付けることによって通常が立体で <tr: <f: 等がイタリック体になるように意味を反転させることが出来る。

注意しなければならないことは、<> のスコープが行単位であるということである。<xx> で始まったモードは改行以前に > で終了させなければならない。人間というのには必ず間違えるものであるから、<> の非対応のエラーについては、入力テキストの拡張子を .err にえたファイルを生成するようになっている。タグジャンプが可能なエディタによって元のファイルに飛べる形式のファイルである。このファイルは上書きモードで生成されるので、そういう名前のファイルを自分から作ってはならない。

xxxx.txt 383 No closing >!

xxxx.txt 383 Number of Inmark and Outmark not matching!!

とじる>がないときには、上記のようなメッセージが出力されるが、同じ一つのエラーに対する警告である。

Unkown Inmark <zzz in line 10

if you want to use < character use ¥<

この警告は、<n と書くつもりで <m などの未定義の入力記号を使ったり、<n などの後の空白を忘れたりしたときに表示される。

too many closing > in line

if you want to use > character use ¥>

この警告は、余分の < があるときに出力されるが、上記のエラーがあると <> の対応が取れていないので、同時に表示されることになる。

複数行の一括指定の命令等については、つぎのセクションで解説する。

また、標準的な入力方法は以下の表のようになっている。

### 北インド関係を主要な対象とする人々へのローマ字転写方法

#### 母音

a	i	u	R(.r)	L
aa	ii	uu	RR(.r.r)	LL
e	ai	o	au	

H(.h) [Visarga] M(.m) [Anusvāra]

#### 子音

k	kh	g	gh	G("n)
c	ch	j	jh	J(~n)
T(.t)	Th(.th)	D(.d)	Dh(.dh)	N(.n)
t	th	d	dh	n
p	ph	b	bh	m
y	r	l	v	
z("s")	S(.s)	s		
h	.l(cerebral l)			
_r [dravidian r]	Z [Tamil l]	_n [Tamil two loop n]	_k [Tamil āydam]	

.a(@) [avagraha]

.o [OM]

ヒンディー語の鼻母音は、A AA のように大文字を用いる。また、ヌクターのつく子音は、

.R [dot D] .Rh [dot Dh] f [dot ph]  
q [dot k] .kh [dot kh] .g [dot g] .z [dot j]

のような方法による。

マラーティー語、ネパール語の

~o [english o] ~a [english a] ~r [Marathi and Neplali r] ような表記もサポートしている。

記号類は、

/ danda // double danda | candrabindu  
.. abbreviation # elliptical dot . period  
のようになっている。

特別な意味を持つ入力要素として, + は k+h あるいは a+i のように用いて、独立した文字であることを示す。

また、ローマ字転写において大文字とするときは、&a などのように、& を直前に付加してやる。

以上に見るように、この入力方式は、Velthuis 方式、

KH 方式、町田方式の折衷的な方式であるので、折衷方式ととりあえず呼んでいる。

しかしながら、この入力方式では南インドのドラヴィダ諸語を用いている人には、いささか入力が面倒なので、以下の表のような入力方式も実装している。

南インド関係を主要な対象とする人々へのローマ字転写方法

母音				
a	i	u	^(r)	.L
aa	ii	uu	^(rr)	.LL
e	ee	ai		
o	oo	au		
H(.h) [Visarga] M(.m) [Anusvāra]				
子音				
k	kh(k')	g	gh(g')	G("n")
c	ch(c')	j	jh(j')	J
T(t)	Th(.th T')	D(d)	Dh(.dh D')	N(.n)
t	th(t')	d	dh(d')	n
p	ph(p')	b	bh(b')	m
y	r	l	v	
z(\$)	S	s		
h	L(l)			
R(_r) [dravidian r] Z(_l) [Tamil l] _n [Tamil two loop n] _k [Tamil āydam]				

### 6.3.2 補助機能

以上のほかに、行頭のコロン ":" による命令がいくつか存在している。

複数行にわたるモードの指定については、以下のようにになっている。

これらの命令は行頭においてのみ認識され、また、:nin 等の文字列以降改行までには空白かタブ記号のみが許される。例外として、:tin のみは :tin: という形式も存在し、:tin ではイタリック体で、:tin: では立体で出力される。この意味も -i オプションによって反転する。また、自動的にインデントされる設定になっている。

インデントに関しては、英文等の引用のために、:qin :qout という命令が用意されている。

そのほかに論文を書くために最低限必要な TeX の命令を単純化したものが、同様の行頭のコロンによる命令として用意されている。

- :t タイトル
- :a オーサー
- :s セクション
- :ss サブセクション
- :sss サブサブセクション

いずれも、:s 空白 (タブ) から行末までを、\$section{} の引数として変換する。但し、タイトル (:t) のみでは、なにも出力されず、オーサー (:a) があって始めてタイトル表題が出力される。

	入る	出る
転写	:tin	:tout
ナーガリー	:nin	:nout
ベンガリー	:bnin	:bnout
オリヤー	:orin	:orout
グルムキー	:gmin	:gmout
テルグ	:tlin	:tlout
カナダ	:kin	:kout
タミル	:tmlin	:tmlout
マラヤーラム	:min	:mout
グラント	:grnin	:grnout
チベット	:tibin	:tibout
クメール	:khmin	:khmout
タイ	:taiin	:taiout

その他の特別の処理としては、ピリオド三つの連続を TeX の \\$ldots コマンドに変換するように設定してある。

### 6.3.3 Preprocessor の起動

拡張子が .tex 以外の、元原稿ファイルを上記のようなやり方で作成したら、まず preprocessor である to\_tex.pl を、

```
jperl to_tex.pl Infile
```

のように起動して処理する。その結果、元ファイルの拡張子を .tex に変更した出力ファイルと .err に変更したエラーファイルとが生成される。エラーファイルが空でない場合は、エラーがなくなるまで、元ファイルを修正する必要がある。

拡張子が .tex である出力ファイルには、先頭に \\$input{my\_pre}, 末尾に \\$input{my\_post} という行が含まれている。論文の種類などにあわせて適当な記述を含んだ my\_pre.tex ファイルをフォルダごとに用意しておくことが必要である。添付されているファイルを元に適切に変更してほしい。

この出力ファイルに platex あるいは latex をかけて処理することで、容易に多言語出力を得ることができる。

Winshell を使用する場合には、上部の左から、 to\_tex, platex, dviout の三つのボタンが並んでいるので、左から順にボタンをクリックすることで画面出力までを得ることができる。

### 6.3.4 カスタマイズの方法

まず、入力モードの切り替えの記号については、

```
$inmark = "<";  
$inmarklength = 1; # < = 1 letter  
$outmark = ">"; # this must be different from $inmark  
$outmarklength = 1; # > = 1 letter
```

とある部分を変更することで変更可能であるが、これ自体の変更はあまりお勧めできない。TeX 等の記号と重ならずに用いることができて、なおかつ他の通常使うことの多い記号と重ならないものがほとんど思いつかないからである。上記の設定から解るように、2 文字以上を使うことが可能なので、その場合には問題ないが、入力の手間が増えてしまうからである。

ここでのモードの判別記号については、全く自由に変更することが可能である。

```
$convert{'N'} = "nagari";
```

という設定の 'N' が <N という部分に対応しているから、

```
$convert{'hn'} = "nagari";
```

```
$convert{'sn'} = "nagari";
```

と追加すれば、<hn でヒンディー語のナガリー表記を、<sn でサンスクリットのナガリー表記を示すと

いうように区別でき、印字に於いては同一でも文字と言語を明確に規定したデータベースを構築していくことができる。

あるいは、

```
$convert{'N'} = "indiantrans";
```

と書き換えるだけで、テキスト自体には何の変更も加えずに、印字出力だけをナガリー文字から、ローマ字転写に変更することができる。

入力方法については、csx, norman, velthuis, kh, dravida (折衷方式南インド用), mixsouth (折衷方式南インド用), mix (折衷方式北インド用) を起動オプションとして用意してあるが、実際に完全にサポートしているのは上述した折衷方式の北インド版と南インド版のみである。

方式	単字オプション	引数オプション
csx	-c	-tc
norman	-n	-tn
velthuis	-v	-tv
KH	-k	-tk
dravida(mixsouth)	-d	-td
		-tmixs
mix	-m	-tm

なにも指定しない場合のデフォルトは、 mix (折衷方式北インド用) である。

デフォルトモードの変更は

```
$convert{'default'} = "indiantrans"; # default mode(just "<"), transliteration
```

を indiantrans から nagari や no\_convert などに変更するだけよい。

入力方式を独自のものに変更したい場合には、以下のように変更する。

たとえばナガリーモードは、

```
sub nagari {
```

で呼ばれているから、その冒頭での入力方式に対応する部分

```
if ($conv_mode eq "csx") {
```

の最後の

```
else {
```

を

```
elsif ($conv_mode eq "mix") {
```

に変更して、みずから作成した subroutine を

```
else {
```

```
&my_new_subroutine;
```

```
}
```

と呼ぶように変更する。

この subroutine 自体の書き方については、省略するが、`mix_nagari` 等の既存の subroutine を見れば理解できるであろう。sed の置換を少しでも行ったことのある人ならば、perl の正規表現の違いを把握するだけで、容易に修得できると思う。

イタリック体の出力でなくする記号を変更するには、  
`$noitamark = ":";`  
となっているところを適当に変更すればよい。

また、この記号の意味を逆転させるオプションをつければ、同等の動作をさせたい場合は、

`$noitareverseflag = 0;`

を

`$noitareverseflag = 1;`

に変更すればよい。

なお、その他の動作変更用の設定として

`$doublecr = 0;`

となっているのを、1 に変更すると、改行が2重化されるので、`TEX` のパラグラフの終了形態にそのまま対応できるようになるが、エラーのチェックの際に行数が異なってしまうのであまりお勧めできない。

## クメール語およびタイ語の翻字について

峰岸 真琴  
高島 淳

### 要旨

東南アジア大陸部において、インド起源の文字は、ビルマ語、タイ語、ラオス語、クメール(カンボジア)語といった孤立語タイプの言語で用いられている。そのうちの多くの言語では母音の数が多く、その言語の一部は声調言語である。この論文では、クメール語とタイ語を例にして、東南アジアのほとんどの言語が5つ以上の母音音素を持つにも関わらず、どのようにして、ローマ字のキーボードのような、5つの母音を想定する母音表記法を持つインド起源の書記システムを受け入れたのか、を示すことを目的とする。

以下では、まずクメール語の複雑な母音体系を示す。一般に、言語に比べて書記体系は変化しづらい。クメール語の正書法は、過去の音韻体系、破裂音の無声化あるいは母音の系統だった分歧といった歴史的変化をはっきりと示している。さらに、タイ語の音韻体系の概要を示す。タイ語の文字体系においては、組み合わせ母音記号の追加や声調記号の付加といった、音韻体系に応じた正書法上の工夫が行われている。

南アジアから東南アジアにわたるインド系文字の使用範囲は、同時にインド起源の宗教的文献が受容された範囲でもある。小論では、これらインド起源の文字を共有する人々が、パーリ語の文書のような共有の知的遺産を利用する際に、文字受容以降に各言語に生じた音韻変化によって相互理解を妨げられることなく、語源と歴史的な関係をよりよく理解するために有用な共通転写法を提案する。

### 7 はじめに

#### 7.1 文字転写法の必要性

現代のインド人が東南アジアを旅行することがあるとすれば、東南アジア大陸部において現在使用されている、カンボジア、タイ、ラオス、ミャンマーといった地域の文字は一見してインド起源であることに、おそらく気がつくだろう。それでもほとんどのインド人は、カンボジア語で‘Mekong’「メコン」と書かれるチベットに源を発し、中国、ラオス、カンボジアからベトナムに至る川の名称が「ガンジス川」に由来することには気づかないであろうし、タイ語の“ラック(Lak)”がインドの人名ラクシュマナを表すことにも気がつかない

いであろう。

小論では、他の言語の文字システムを受容し、自身の音韻体系に合わせて修正を加えた例として、クメール語(カンボジア語)とタイ語の文字を取り上げる。

クメール語の書記体系は、同言語においてインド系文字を受容した後に生じた有声閉鎖子音の無声化および子音変化に伴う母音体系の変化といい音韻の史的変化を反映している。その結果、現代クメール語の書記体系は、インド系文字の原音と、実際の発音がかけ離れてしまった。一方、タイ語の書記体系はインドの書記体系受容に関する、クメール語とはまた異なる例で

あって、タイ語におけるインド系文字の受容の後に、有声破裂子音の無声帯気化とともに、声調体系がどのように複雑化したかを示している。

これらの言語におけるインド起源の文字体系の受容と独自の修正によって、学術的分野で必要とされる、特に多言語テキストデータを処理・使用する場合に、ローマ字転写の問題が引き起こされている。例えば、カンボジア人の学者は、パーリ語のテキストに関心を持ったとしても、タイ文字およびクメール文字で書かれたパーリ語テキストの比較などはしたがらないであろう。

このような障壁を軽減するためには、クメール人とタイ人との学術的交流だけでなく、インド起源の文字を使用する共同体の中で、共通の転写法を工夫して用いることが有用であると考えられる。

小論はまず現代クメール語とタイ語の音韻体系についてその概要を述べる。前者は複雑な母音体系を持ち、後者は典型的な声調言語である。これらの言語のインド系文字受容期の音韻体系と、文字受容後の音韻の通時的な変化は、本来のインド系文字のもっていたと推定される体系内の価値と、現代におけるその実際の発音の比較から推論することができる。小論ではさらに、インド起源の文字を共有する人々の学術上の交流のために利便を図り、かつローマ字配列キーボードを用いた多言語テキストデータの入力を簡便に行うために、これら言語の転写方法を提案する。

## 7.2 歴史的背景

クメール語とタイ語の歴史的な関係は、いくぶんか朝鮮語と日本語との関係に似ている。前者は外からの、一インドと中国それぞれからの—文化的影響を直接に受け、後者は前者から間接的に外来高文化を受容した。

中国の資料によれば、クメール土着の女王がインドから来た王子と結婚し、扶南の王朝の創始者となった、という言い伝えがクメール人にあるとされる。また、西暦658年のチャンパーの碑文によれば、最初のクメール王朝は、ナーガ(龍王)の娘であるソーマと結婚したバラモンのカウンディンヤによって建てられたとされ

る。こうした伝説は、インド文明の影響を受けた、土着の母系制の権威がかつて存在したことを示唆している。

また、東南アジアの古代文化の考古学は、西暦以前から、東南アジアとインドとの間に交流があったことを物語っている。古クメール語で書かれた最も古いクメールの碑文(K.600)は、西暦611年という時代にさかのぼるとされている、だが時代の明記されていない碑文もあり、また、さらに時代が古いサンスクリット語で書かれた碑文も存在する。タイ文字の最も古い碑文は、13世紀のスコータイ王朝ラームカムヘーン王のものである。西暦1292年のラームカムヘーン王碑文によれば、王が、南インド系の文字に起源を持つクメール文字おそらく参考にして、タイ文字を発明したとされている。したがって、タイ文字は、その元となつたクメール文字と同様に、南インド系文字グループに分類される。

## 7.3 言語的背景

東南アジアの歴史学の開拓者であるセデス(Cœdès)は、東南アジアはその早い時代に、バラモン教、社会システム、農耕技術、そしてもちろん文字、を受け入れることによって「インド化」された、と述べている(参考文献1を参照のこと)。言語の系統関係の観点からは、東南アジアの大陸部は、モン・クメール、タイ・カダイ、シナ・チベット、ミャオ・ヤオ諸語が集中している中国の南部分とともに、ひとつのまとまった言語地域を形成している。それらの言語の中で、クメール(カンボジア)語はオーストロアジア(オーストリック)語族の中のモン・クメール語族に属し、そしてタイ(シャム)語はタイ・カダイ語族に属する。

両言語の系統的な違いにも関わらず、クメール語とタイ語は多くの言語学的特徴を共通に持っている。これは、お互いに長い接触の歴史を持つことがその一因でもあるが、どちらも屈折も格標示も無い「孤立語」という言語類型に分類される言語であることもまたその原因である。

# 8 クメール語の音素体系

## 8.1 クメール語の音節構造

以下にクメール語の音韻体系を示す。クメール語に固有の単語は単音節語あるいは二音節語である。二音節語は、二音節目が必ず主音節である。クメール語においては、固有の単音節語、二音節語のほかに、サンスクリット語、パーリ語、フランス語、英語起源の複音

節の借用語が多く用いられている。以下では、Cを用いて子音を表し、Vは母音を表すものとし、丸括弧で括られたものは随意的な要素を表すものとすると、クメール語の単語は次のように表すことが可能である。

$$(1) (C_4(C_5)V(C_6))C_1(C_2)V(C_3)$$

短母音 V には、語尾子音  $C_3$  が必ず後ろに置かれる。つまり、短母音は閉音節にしか現れない。これに対して長母音  $\tilde{V}$  には、 $C_3$  が後ろに置かれても、置かれな

くともかまわない。つまり、閉音節にも開音節にも現れる。また、第一音節に語頭子音群  $C_4C_5$  がある場合には、その音節の語末子音  $C_6$  が落ちる。

## 8.2 クメール語の子音体系

クメール語の子音音素は以下の表 1 に示される、必要な場合には、かぎ括弧の中に国際音声記号を表示してある。

表 1 クメール語の子音音素表

調音法	唇歯	両唇	歯	硬口蓋	軟口蓋	声門
無気閉鎖音		p	t	c [tʃ]	k	?
帶氣閉鎖音		ph	th	ch [tʃh]	kh	
有声閉鎖音	b [b]	d [d]				
鼻音		m	n	p	ŋ	
摩擦音	(f)		s			h
流音			r, l			
接近音	v			j		

サンスクリットの子音字の原音との比較から、以下の特徴が観察される。

- (2) クメール語には、有声有氣閉鎖音 /bh, dh, jh, gh/ がない。
- (3) クメール語には有声閉鎖音の一部が欠けている；例えば、/g, j/。
- (4) 有声閉鎖音は、/b, d/ のように「入破音」の特徴を備えているさらに、それらは無声閉鎖音のように振る舞い、無声子音に伴う種類の母音がそれらに

後続する。

- (5) クメール語には、/t, th, d, dh, n, s/ のような、そり舌 (retroflex) 子音はない。一般に、そり舌子音は東南アジアの言語には存在しない。
- (6) 無声無気音、無声有氣音と有声閉鎖音との区別は、音節頭あるいは母音間に現れるが、音節末の位置には、[p', t', k'] のような破裂を伴わない内破音だけが現れる。
- (7) (f) は借用語にしか現れない。

## 8.3 クメール語の母音体系

クメール語の母音は表 2 に示される、必要な場合には、かぎ括弧の中に国際音声記号を表示してある。

表 2 クメール語の母音音素表

	前舌	中舌	後舌	二重母音
狭母音	i, ii	ɯ,ɯɯ	u, uu	iɛ, ɯɛ, uo
半狭母音	e, ee, è [e̥], èe [e̥e̥]	ə, əə, ð [ð̥], ðə [ð̥ə̥]	o, oo, ðo [ð̥o̥]	ɛa, iø, ða
広母音	ɛ, ð [ð̥], ðð [ð̥ð̥]	a, aa	ɔ, ɔɔ, ð [ð̥ɔ̥], ðð [ð̥ð̥ɔ̥]	ae, aø, ao

13 個の長母音、13 個の短母音、9 個もの二重母音がある。クメール語の母音の数は普通のインド系の文字が想定しているよりも多いことは明らかである。

## 8.4 クメール語のレジスター (段: 二つの平板声調) の体系

表 2 に示したように、クメール語には 2 系列の母音がある。一つはアクセント記号を伴わない ‘A’ 系列の

母音であり、もう一つはグレーヴアクセント記号で表示される ‘O’ 系列の母音である。

二つの系列の違いは、舌の高さ位置が、[e̥] と [e̥e̥] のように、‘A’ 系列では ‘O’ 系列よりも低い。また、‘A’

の系列は二重母音 [e:i] のように発音されることもあるが、これに対して ‘O’ 系列ではそのような発音はない。先行研究の一部には、参考文献の 2 のように、この二つの系列の 1 番目は頭声（ヘッドレジスター）であり、2 番目は気息性の母音を伴った胸声（チェストレジスター）であるというように、発声 (phonation) の型の違いとして記述するものもある。類似のレジスターの対立はモン・クメール語族であるミャンマーのモン語やタイのクーイ(スウェイ)語にも存在する。筆者自身の観察によれば、気息性の発音の対立は現代の標準クメール語では失われている。したがって、上記の母音の対立は調音上の対立と考えられる。

発音に比べて、保守的で変化しづらいクメール文字は、現行の複雑な母音体系を説明する手がかりとなる。

正書法が示すように、かつてクメール語の閉鎖音には、無声音 (aghosa) および有声音 (ghosa) の違い、すなわち /\*kā/ と /\*gā/ のような音素対立があったと考えられる。

その後、有声子音に後続する母音はその音質が変化した。しかし、この段階では \*kaa と \*gea のような子音間の音素対立は存在しており、母音の発声上の違い

は音韻的には関与的ではなかった。

さらにその後、子音の区別が失われ、/ka/ と /kəə/ のような母音の音質の違いが音素対立と担うようになつた。このような歴史的な音韻変化にも関わらず、現代クメール語の正書法はいまだに伝統的な体系を保存しているのである。

その結果として、/ka:/ と /kəə/ のように、同じ初頭子音を持つ母音の音韻的な区別は、次の表 3 のā のように同じ母音記号を持つ、無声子音と有声子音の違いとして表記される。また、プロンペンの標準的な発音では、[əə] は [ia] と同化している。

表 3 クメール語の書記と発音

書記	kā	gā
発音	[ka:]	[kiə]

これは、現代標準クメール語の正書法において、母音の発音は母音記号によってのみ同定されるのではなく、先行する子音文字と母音記号との組み合わせによって同定されることを意味する。

## 9 クメール語の転写法

小論で提案されるクメール語の転写は以下の原則に基づいている。

(8) 文字の順番は  $C(-C)(M)V$  である。子音 (C)，その付加記号 (肩付き文字)(M)，母音 (V) は発音される音素の順を反映しているのであって、字形

(glyph) の順ではない。

### 9.1 クメール語の子音の転写

表 4 にクメール語の子音の転写法を示す。以下「インド系対応音」を Ind. eq で、「転写」を Translit で、「文字」を Char で、「発音」を pron で表すこととする。

表 4 クメール語子音の転写法

No.	Ind. eq.	Translit.	Char.	pron.	No.	Ind. eq.	Translit.	Char.	pron.
1	k	k	ក	kɔ:	21	p	p	ឃ	bɔ:
2	kh	k'	ខ	kho:	22	ph	p'	ឃ	pho:
3	g	g	ក	kɔ:	23	b	b	ឃ	pɔ:
4	gh	g'	ឃ	khɔ:	24	bh	b'	ឃ	phɔ:
5	ñ	G	ឃ	ŋɔ:	25	m	m	ឈ	mɔ:
6	c	c	ឃ	co:	26	y	y	ឃ	yɔ:
7	ch	c'	ឃ	cho:	27	r	r	ឃ	rɔ:
8	j	j	ឃ	cɔ:	28	l	l	ឃ	lɔ:
9	jh	j'	ឃ	chɔ:	29	v	v	ឃ	vɔ:
10	ñ	J	ឃ	ŋɔ:	30	s	s	ឃ	so:

11	t̪	T	ତ	dɔ:	31	h	ହ	ତୁ	hɔ:
12	ʈh	T'	ଥ	thɔ:	32	l̪	L	ଲୁ	lɔ:
13	ɖ	D	ଡ	dɔ:	33	a	q	କ	?ɔ:
14	ɖh	D'	ଧ	thɔ:	(27a)	r̪	RI	ୟ	ru:
15	ɳ	N	ଣ	nɔ:	(27b)	ର୍ର	RII	ୟୁ	ruru:
16	t̪	t̪	ତ	tɔ:	(28a)	ି	LI	ୟ	lu:
17	th	t'	ଥ	thɔ:	(28b)	ି	LII	ୟ	lu:
18	d̪	d̪	ଡ	tɔ:					
19	dh	d'	ଧ	thɔ:					
20	n̪	n̪	ଣ	nɔ:					

以下の点について注意していただきたい。

(9) そり舌子音は大文字で示される。

(10) 鼻音子音はそり舌の ɳ を除いて、対応する有声の破裂音として、同じ調音点を持つ大文字で表す。例えば、G と g, J と j。

(11) 帯気音は引用符「'」ひとつで標示される。

(12) 有声破裂音は本来のインド系文字と同じ文字を用いて表す。

(13) いわゆる内在母音記号 ‘a’ が、ダミーの子音字(ゼロ子音)として用いられる場合には、‘q’ を用いて表す。

## 9.2 クメール語の子音群の転写

音節の最初の位置に二つ以上の子音が子音群として現れる場合には、2番目(あるいは3番目)の子音は「足」(ଚାନ୍ଗ 'cā:ŋ') と呼ばれる下付の異形態(異グリフ)で表される。1番目と2番目の子音には80通りの組み合わせがある。

小論の転写方法では、同じ転写方法が1番目の子音

と2番目の子音に適用される。2番目の子音は1番目の子音に、ハイフンネーションや特別な結合記号を用いずずに、直接に後続する。

次の表5では代替子音文字 ‘q’ が音節頭子音として使われている。

表5 クメール語子音連続の転写法

No.	Ind. eq.	Translit.	Char.	pron.	No.	Ind. eq.	Translit.	Char.	pron.
1	-k	qk	କ	kɔ:	21	-p	qp	ପ	bɔ:
2	-kh	qk'	କ୍ଷ	khɔ:	22	-ph	qp'	ଫ	phɔ:
3	-g	qg	ଗ	kɔ:	23	-b	qb	ବ	pɔ:
4	-gh	qg'	କ୍ଷ	khɔ:	24	-bh	qb'	ବ୍ର	phɔ:
5	-ɳ	qG	ଣ	ɳɔ:	25	-m	qm	ମ	mɔ:
6	-c	qc	ଚ	cɔ:	26	-y	qy	ଯ	yɔ:
7	-ch	qc'	ଚ୍ଛ	cho:	27	-r	qr	ର	rɔ:
8	-j	qj	ଜ	cɔ:	28	-l	ql	ଲ	lɔ:
9	-jh	qj'	କ୍ଷ	chɔ:	29	-v	qv	ବ	vɔ:
10	-ñ	qJ	ଣ	ɳɔ:	30	-s	qs	ଶ	sɔ:

11	-t̥	qT	ঃ	dɔ:	31	-h	qh	ঃ	hɔ:
12	-th	qT'	ঃ	thɔ:	32	-	-	ঃ	-
13	-d̥	qD	ঃ	dɔ:	33	-a	qq	ঃ	?ɔ:
14	-dh	qD'	ঃ	thɔ:					
15	-n̥	qN	ঃ	nɔ:					
16	-t	qt	ঃ	tɔ:					
17	-th	qt'	ঃ	thɔ:					
18	-d	qd	ঃ	tɔ:					
19	-dh	qd'	ঃ	thɔ:					
20	-n	qn	ঃ	nɔ:					

### 9.3 その他のクメール語の子音記号

表 6 にあるように、 その他の ヴィサルガ (visarga), アヌスヴァーラ (anusvāra) は、 クメール語においては、 それぞれ音節末の/h, m/ の表記に用いられる。アヌス

ヴァーラは 母音 /a ā, u/ のみに後続する。したがって ‘aM, aaM, uM’ という転写は ঃ, ঃ, ঃ を表す。

表 6 クメール語記号の転写法

No.	Ind. eq.	Transliteration	Char.	Pronunciation
1	ঃ	x	:	h (visarga)
2	ঃ	M	ঃ	m (anusvāra)

### 9.4 クメール語の母音記号の転写記号

一般に南インド起源のインド系文字の場合にそうであるように、 クメール語の /e, ae, ai/ の母音の記号は、 音節頭子音の左に書かれ、 /o, au/ の母音は、 子音字の左右に分かれて書かれる。しかし、 ここで提案する転写は発音の順番に従う。母音は先行する子音の後ろに書かれる。グリフの位置は考慮しない。

以下の表 7 では、 代替子音文字 ‘q’ が語頭子音として用いられる。表 7 中の 1a はタイ語からの借用語の中で短母音 /a/ を明示的に表すために用いる。また、 1b は インド系の借用語において /a/ はいかなる音節末子音を伴わないことを明示するために用いられる。

表 7 クメール語母音記号の転写法

No.	Ind. eq.	Transliteration	Char.	Ser. 'A' pron.	Ser. 'O' pron.
1	a	qa	ঃ	a	ɛə
1a	a	qa^	ঃ	a	ɛə
1b	a	qQ	ঃ:	a	ɛə
2	ā	qaa	ঃ	a:	iə
3	i	qi	ঃ	e	i
4	ī	qii	ঃ	əi	i:
5	—	qw	ঃ	ə	u

6	—	qww	ጀ	ə:	ɯ:
7	u	qu	ጀ	o	u
8	uu	quu	ጀጀ	oo	uu
9	—	quo	ጀጀ	uo	uo
10	—	qae	ጀጀጀ	aə	ə:
11	—	qwa	ጀጀጀጀ	mə	mə
12	—	qia	ጀጀጀጀጀ	iə	iə
13	e	qe	ጀጀጀጀጀጀ	e:	ɛ:
14	ee	qee	ጀጀጀጀጀጀጀ	ae	ɛ:
15	ai	qai	ጀጀጀጀጀጀጀጀ	ai	ɛi
16	o	qo	ጀጀጀጀጀጀጀጀጀ	o:	u:
17	au	qau	ጀጀጀጀጀጀጀጀጀጀ	av	əv

## 9.5 クメール語の母音文字の転写

語頭子音を持たない母音の場合、それは独立母音字として表記される。クメール語には /a, ă/ の単独の形式は無い。前者は時々声門閉鎖音 [?] として発音される「ゼロ」子音としての役割を持ち、後者は初期の碑文にさえ用いられない。

表 8 に、上記の /a, ă/ を除いた、単独の母音の転写を示す。これらの母音文字は今日ではしだいに用いられなくなっている。ここで提案されている転写では、単独の母音は大文字で表記される。

表 8 クメール語独立母音字の転写法

No.	Ind. eq.	Transliteration	Char.	Ser. 'A' pron.	Ser. 'O' pron.
3	i	I	ጀ	e	i
4	ī	II	ጀጀ	əi	i:
7	u	U	ጀ	o	u
8	uu	UU	ጀጀ	oo	uu
8+v	uv	UV	ጀጀጀ	əv	əv
13	e	E	ጀ	e:	ɛ:
15	ai	AI	ጀጀ	ai	ɛi
16	o	O	ጀጀጀ	o:	u:
16+y	oy	OY	ጀጀጀጀ	aoy	
17	au	AU	ጀጀጀጀጀ	av	əv

## 9.6 クメール語の付加記号

クメール語は声調言語ではない。しかし、音節の質、母音の音量を変化させる上付き文字がある。

ここで表 9 に提案する転写において、これらの付加記

号は先行する子音の属性を変化させると考えられている。したがって、それらの付加記号は子音の直後に表記される。

表9 クメール語の変音記号の転写法

No.	Ind. eq.	Translit.	Char.	No.	Ind. eq.	Translit.	Char.
1	—	q1	៥	5	—	q5	៥
2	—	q2	៥	6	—	k8	៥
3	—	q3	៥	7	—	q0	៥
4	—	q4	៥	8	r	qR	៥

No.1 Ḃ 母音短縮符号兼母音削除記号：先行する音節の母音が短いことを表示する。例えば、mႤ (kā + k 母音短縮記号) [kak]

No.2 ḃ 母音系列変換符号: ja ḃ, ja ḡ, ma ḃ, ya ḡ, ra ḃ, va ḃ のような ‘O’ 系列の子音文字の上に置かれ、それらを ‘A’ (aghosa) 系列に変換する、したがって後続する母音は ‘A’ 系列で読まれる。‘pa’ ḃ [boɔ] の場合、この符号は無声化の機能も担う。‘pa’ ḃ は [b] だが、‘p2’ ḃ は [p] である。

No.3 ḃ 「上付きの treisap」また別の母音系列変換符号: 子音文字の上に置かれ、‘pa’ ḃ, ‘sa’ ḃ, ‘ha’ ḃ and ‘a’ ḃ のような ‘A’ 系列文字を ‘O’ 系列 (ghosa) 文字に変換する。

No.4 Ḃ 声調標示: おそらくタイ語からの第四声調記号の借用である。強勢イントネーションを標示する。

No.5 ḃ 「下付の treisap」もう一つの母音系列変換符号: 子音文字の下に置かれ、‘O’ 系列を ‘A’ 系列に変換する。また逆も行う。

No.6 ḃ 2語の単音節の単語にのみ用いられる記号。すなわち ‘ka’ と ‘Ta’, ḃ と ḃ。

No.7 Ḃ 母音削除記号: 母音記号を発音させないようにする。‘kertōi’ ḃ ḃ

No.8 ḃ: レーパは、いくつかの場合にのみ子音文字に付加される。例えば、‘d’amR’ ḃ ḃ (dharma)。

## 10 タイ語の音素体系

### 10.1 タイ語の音節構造

クメール語と同様、タイ語固有の単語は単音節あるいは二音節である。二音節の単語では、2番目の音節が必ず主音節である。タイ語固有の単音節語、二音節語のほか、パーリ語、サンスクリット語、英語起源の借用語の複音節語が用いられる。

C を用いて子音を表し、V は母音を表すものとする、丸括弧で括られたものは随意的な素を表し、T は声調を表す。タイ語の単語は以下のように表すことが可能である。

$$(14) (C_4(C_5)V(C_6))C_1(C_2)V(C_3)/T$$

クメール語と同様に、タイ語の短母音 Ḃ は必ず音節

末子音  $C_3$  が後続する。これに対して、長母音 ḃ 必ずしも  $C_3$  が後続しなくともかまわない。

タイ語の場合、 $C_1$  と  $C_2$  (あるいは  $C_4$  と  $C_5$ ) との可能な組み合わせは非常に限られている。すなわち /kr-, khr-, kl-, khl-, kw-, khw-, pr-, phr-, pi-, phl-, tr-, (thr-)/ という組み合わせしかない。特に最後のひとつは借用語にしか現れないことを指摘しておく。

### 10.2 タイ語の子音体系

表10に、タイ語の子音音素の目録を示す。クメール語の子音音素表と較べると、タイ語には口蓋音の鼻音 /ŋ/ が存在しないことと唇歯音の摩擦音 /f/ が存在すること以外はほとんど同じである。

表 10 タイ語の子音体系

調音法	唇歯	唇	歯	硬口蓋	軟口蓋	声門
無気閉鎖音		p	t	c [tç]	k	?
帶氣閉鎖音	ph	th	ch [tʃh]		kh	
有声音	b	d				
鼻音	m	n			ŋ	
摩擦音	f		s			h
流音			r, l			
接近音	w		y			

タイ語とクメール語の違いのひとつは、 $C_1$  と  $C_2$  との組み合わせが、前者は 12 通りの組み合わせしかない

のに対して、後者が 80 通りの組み合わせがあることである。

### 10.3 タイ語の声調体系

タイ語とクメール語のまた別の違いは、表 11 に示してあるように、前者が 5 種類の声調対立を持つのに対

して、後者は声調を持たないことである。

表 11 タイ語の 5 声調体系

Tone Number	1	2	3	4	5
Contour	mid level	low level	falling	high level	rising
Symbol	aa	àa	âa	áa	ăa

### 10.4 タイ語の母音体系

表 12 に示したように、タイ語には 9 つの短母音、9 つの長母音および 3 つの二重母音がある。

表 12 タイ語の母音体系

	前舌	中舌	後舌	二重母音
狭母音	i, ii	ɯ,ɯɯ	u, uu	ia, iɯa, ua
半狭母音	e, ee	ə, əə	o, oo	
広母音	ɛ, ɛɛ	a, aa	ɔ, ɔɔ	

### 10.5 タイ語の転写

この論文で提案しているタイ語の転写は、クメール語において(8)で与えられたのと似た、以下の原則に基づく。

(15) 文字の順番は  $C - (C)(M)V$  とする。子音 (C)、その (上付きまたは下付きの) 付加記号 (M)、母音 (V) の順番は発音の順番を反映し、字形 (グリフ) の順番は反映しない。

その違いは以下のようである。

タイの転写法では、子音連続は明示的にハイフン (-) によって標示される。タイ語には子音連続を表示する

記号は無い。全ての子音は完全なグリフで表記される。一方、クメール語では、2 番目あるいは 3 番目の子音に特別に使われる下付きの付加記号 (coeung 脚と呼ばれる) があり、子音とのつながりを表示するいかなる表示記号も必要なかった。

もう一つの違いは、タイ語には声調を標示する 4 つの上付の付加記号があるが、下付きの変音付加記号は無いことである。

### 10.6 タイ語の子音字の転写法

表 13 はタイ語の子音字の転写法である。

表 13 タイ語の子音字の転写法

No.	Ind. eq.	Translit.	Char.	pronunciation	ex. word	meaning of ex.	spelling of ex.
1	k	k	ก	kɔ̄o	kài	chicken	ไก่
2	kh	k'	ຂ	khɔ̄o	khài	egg	ไข่
3	g	g	ກ	khɔ̄o	khwaai	buffalo	ควาย
4	gh	g'	ຂ	khɔ̄o	rakhaŋ	bell	ระฆัง
5	ñ	G	ງ	ŋɔ̄o	ŋuu	snake	งู
6	c	c	ຈ	cɔ̄o	caan	dish	จาน
7	ch	c'	ຈ	chɔ̄o	chìŋ	small cymbal	ฉิ่ง
8	j	j	ຈ	chɔ̄o	cháaŋ	elephant	ช้าง
9	—	z	ຊ	sɔ̄o	sôo	chain	โซ่
10	jh	j'	ຜ	chɔ̄o	kachæə	tree	กະຜວມ
11	ñ	J	ຢ	yɔ̄o	yǐŋ	woman	หญิง
12	—	D"	ຝ	dɔ̄o	chadaa	crown	ຫຼາມ
13	t	T	ຝ	tɔ̄o	patàk	goad	ປັກ
14	th	T'	ຝີ	thɔ̄o	thăan	base	ສ້ານ
15	đ	D	ໜ	thɔ̄o	monthoo	Mañdo	ມັນໂທ
16	đh	D'	ໜີ	thɔ̄o	phûu thâu	old man	ຜົ່າກ
17	ñ	N	ໝ	nɔ̄o	neen	Nera	ເນຣ
18	—	d"	ດ	dɔ̄o	dèk	child	ເຕີກ
19	t	t	ດ	tɔ̄o	tàu	turtle	ເຕິກ
20	th	t'	ດ	thɔ̄o	thູງ	bag	ດູງ
21	d	d	ຖ	thɔ̄o	thahăan	soldier	ທ່າງ
22	dh	d'	ຖີ	thɔ̄o	thoŋ	flag	ອົງ
23	n	n	ນ	nɔ̄o	nŭu	mouse	ໜຸງ
24	—	b"	ບ	bɔ̄o	bai mái	leaves of tree	ໄບໄມ້
25	p	p	ປ	pɔ̄o	plaa	fish	ປລາ
26	ph	p'	ຜ	phɔ̄o	phûuŋ	honeybee	ຜົງ
27	—	f	ຝ	fɔ̄o	fää	cover	ຝາ
28	b	b	ພ	phɔ̄o	phaan	base	ພານ
29	—	F	ຝ	fɔ̄o	fan	tooth	ພື້ນ
30	bh	b'	ກ	phɔ̄o	sămphau	junk	ສຳກາ
31	m	m	ນ	mɔ̄o	máa	horse	ນັກ
32	y	y	ຢ	yɔ̄o	yák	giant	ຢັກໜີ
33	r	r	ຮ	rɔ̄o	rwa	boat	ເຮືອ
34	l	l	ລ	lɔ̄o	liŋ	monkey	ລືງ
35	v	v	ວ	wɔ̄o	wăen	ring	ແຫວນ

36	š	X	ສ	s̥ɔ̄	sǎalaa	arbor	សាលា
37	ş	S	ຜ	s̥ɔ̄	rūis̥ii	ascetic	ពុជ្រ
38	s	s	ສ	s̥ɔ̄	sǔa	tiger	លីវ
39	h	h	ឃ	h̥ɔ̄	hiip	box	ទិប
40	!	L	ធម៌	l̥ɔ̄	cùlaa	kite	កុដា
41	a	q	ឬ	?ɔ̄	?àaj	bowl	ឧង
42	—	Q	ឯ	h̥ɔ̄	nók h̥uuk	owl	នកម្ពុក

転写の原則はほとんどクメール語のものと同じである。

タイ語にはクメール語に比べてより多くの文字がある。理由として、インド系の言語の音韻論には存在しない音素を表すために文字を追加したということもある。また、タイ語では、クメール文字では現在区別しない /š, ş, s/ の綴り字上の違いを保存しているというこ

ともある。

インド系文字の受容後に生じた音韻変化のために、タイ語には多くの同音の文字がある。

結果として、同音文字を区別するために、タイ語では表 13 に示したような、(ニワトリ (kay) の K のような) 日常的な語彙を用いる。

## 10.7 タイ語の母音の転写法

タイ語の母音記号の転写法を表 14 に示す。

表 14 タイ語の母音記号の転写法

No.	Ind. eq.	Translit.	Char.	pron.	No.	Ind. eq.	Translit.	Char.	pron.
1	va	ua	ໂຟ້າ	ua	16	—	eu	ເອົວ	ø
2	a	O	ໂອ(ໂ)	ɔ̄	17	—	eH	ເອະ	e?
3	ah	aH	ໂະ	a?	18	au	ເອຸ	aw	
4	a	a	ໂອ	a	19	—	auH	ເອຸຫະ	ɔ?
5	ava	Cua	ໂຟ້າ	a	20	—	@	ເອີ່ມ	ø-
6	ā	aa	ໂາ	a:	21	—	iaH	ເອີບະ	ia?
7	am	aM	ໂຳ	am	22	—	ia	ເອີຍ(ໂ)	ia(-)
8	i	i	ີ	i	23	—	IaH	ເອີວະ	wa?
9	ī	ii	ີ້	i	24	—	Ia	ເອີ່ວ(ໂ)	wa(-)
10	—	I	ີ	ɯ	25	—	E	ເວ	e:
11	—	II	ີໍໂ/ີໍວ	ɯ:	26	—	E8	ເວີ	e-
12	u	u	ູ	u	27	—	EH	ເວວະ	e?
13	ū	uu	ູ້	u:	28	o	ໂ	ໄ	o:
14	e	e8	ເີ້ວ	e-	29	—	aI	ໄວ	ay
15	e	e	ເີ	e	30	ai	ໄວ	ໄ	ay

ັ, ໂ, ແ (y, v, a) のような子音文字は母音を表示する組み合わせ文字の一部として用いられる。

## 10.8 タイ語の付加記号

タイ文字では、表 15 に示されているような上付きの付加記号によって声調と母音の長さの変更を表す。

表 15 タイ語の修正記号の転写法

No.	Ind. eq.	Transliteration	Char.	
1	—	q1	៥	Tone mark 1
2	—	q2	៥៥	Tone mark 2
3	—	q3	៥ᬁ	Tone mark 3
4	—	q4	ᬁᬁ	Tone mark 4
5	—	q8	ᬁᬁᬁ	(vowel shortener)
6	—	q0	ᬁᬁᬁᬁ	(vowel killer)

2つの語頭子音連続を持つ語の場合、声調記号は2番目の子音字の上に置かれる。

## 11 結論

この論文では、クメール語とタイ語の音韻論の輪郭を述べた。どちらもインド系の文字体系が表すことを想定していなかったような複雑な母音体系を持っている。小論ではこうした言語の転写法を提案した。

この転写法の特徴は:

1. 現在の発音とは関わらず、閉鎖音の本来音価を保存している。

2. グリフとして現れる位置、順番ではなく、音素が発音される順番に従っている。

3. C(C)MV という連續の順序: 付加記号 (M) は、V ではなく、C を変更していると考える。

インド起源の、例えばパーリ語の文書といった共有の知的遺産を共有する学術的な社会において用いるためにこの転写法は工夫されている。

## 参考文献

1. G. Coedès. *The Indianized states of Southeast Asia*. Kuala Lumpur, 1968.
2. J. M. Jacob. *A concise Cambodian-English Dictionary*. London: OUP, 1974.
3. Comrie, Bernard (ed.). *The major languages of East and South-East Asia*. London: Routledge, 1990.
4. Institute bouddhique. *vacanānukrama khmeera. [Dictionnaire Cambodgien]*. Phnom Penh, 1967.
5. 峰岸 真琴. 「現代クメール語の正書法と諸方言」, 『言語研究』第 88 号, pp.41–67. 日本言語学会 1985.
6. Noss, Rochard. B. *Thai reference grammar*. Washington, D.C., 1964.

# チベット文字における結合文字の使用に関する調査

星 泉  
高島 淳

## 0. はじめに

本研究では、チベット文字における結合文字の使用に関する調査を行った。

チベット文字は、ラーフミー文字を共通の祖とするインド系文字の仲間で、音節文字の一種である。チベット文字においても、他の多くのインド系文字と同様に結合文字が用いられる。チベット文字の場合は子音連続の一部を結合文字によって表すが、それは縦に子音字を重ねる形で表記される。これらの結合文字にはいったいどれだけの種類があるのか、ということについては、これまで十分には明らかにされてこなかった。そのため、これまでに開発されたチベット文字表記システムでは、表示できない結合文字が多く発生する結果となっている。

こうした問題点を解決するために、本研究では、特にサンスクリットと古代チベット語における結合文字に注目し、これらを網羅的に収集する作業を行った。古代チベット語の結合文字まで視野に入れた同種の研究は未だ存在しない。完成したリストは、学術的な利用にたえるチベット文字表示システムを作るための基礎資料としての活用が期待される。

チベット文字によって表される言語は、チベット語をはじめ、サンスクリット、ゾンカ（ブータンの公用語）、ラダック語（インド）などがあるが、本研究では主としてチベット語とサンスクリットを調査の対象とした。

以下に用いる転写記号等の理解のために拡張ワイリー転写法と対応したチベット文字の表を最初に載せておくことにする。

チベット語の 30 基字

<i>ka</i>	<i>kha</i>	<i>ga</i>	<i>nga</i>
ཀ	ཁ	ڳ	ڻ
<i>ca</i>	<i>cha</i>	<i>ja</i>	<i>nya</i>
ཅ	ཆ	ڃ	ڽ
<i>ta</i>	<i>tha</i>	<i>da</i>	<i>na</i>
ߕ	ߠ	ڏ	ڻ
<i>pa</i>	<i>pha</i>	<i>ba</i>	<i>ma</i>
પ	ફ	બ	મ
<i>tsa</i>	<i>tsha</i>	<i>dza</i>	<i>wa</i>
ત્શ	શ	ઝ	વ
<i>zha</i>	<i>za</i>	' <i>a</i>	<i>ya</i>
જ્શ	જ	ા	ય
<i>ra</i>	<i>la</i>	<i>sha</i>	<i>sa</i>
ર્શ	લ	શ	સ
<i>ha</i>	<i>a</i>		
હ	ા		

サンスクリット表記

<i>a</i>	<i>i</i>	<i>u</i>	<i>r</i>	<i>l</i>	<i>l̥</i>	<i>ai</i>	<i>au</i>	<i>am</i>	<i>ah</i>
ા	િ	ુ	ર	લ	લ્લ	ઐ	ાઉ	અમ	ાહ
<i>gha</i>	<i>dha</i>	<i>bha</i>	<i>jha</i>						
ગ્ઘા	ધા	ભા	જા						
<i>ta</i>	<i>tha</i>	<i>da</i>	<i>dha</i>	<i>na</i>	<i>sa</i>	<i>kṣa</i>			
ત્શ	થા	દ	ધા	ના	શ	ક્ષા			

# 1. チベット文字の結合文字

チベット文字では、子音連続の一部が子音字の縦結合の結合文字によって表されるとしたが、結合文字のあり方は、対象とする言語によって異なる。そこで以下では言語別に解説する。なお、ここでいう子音連続とは、あくまで文字表記上の子音連続を指すものであり、実際の発音での子音連続とは異なるものである。

また、インド系文字におけるいわゆる有声有氣音の系列は、チベット文字では単独の文字ではなく、有声子音字と子音字 ha の結合文字として書かれるので、これらは結合文字として扱う。

*a*, *i*, *u*などの長母音を表す記号については、単独の母音記号は存在せず、子音字 *h* を結合させることによって表す。例えば、*॥* の長音は *॥h*, *ஓ* の長音は *ଓh*, *ୟ* の長音は *ୟh*, などである。従って、これらも子音結合の一種として扱うことにする。

## 1.1 チベット語の表記

チベット語の表記においては、発音上の音節を単位に文字表記し、音節の切れ目にはツェクと呼ばれる音節の切れ目を表す点を挿入する。ここでいう音節とは、まさにこれを指す。

子音連続には、音節初頭子音連続と音節末子音連続があり、それぞれ分けて説明する。

### 1.1.1 音節初頭子音連続

音節初頭子音連続のうち、初頭子音字が *g-*, *d-*, *b-*, *m-*, *-'* の 5つであり、かつ、後続の子音字が *'*, *y*, *r*, *l*, *v* 以外である場合は、子音字を横に並べるだけで、特に結合のための記号等は使用しない。

一方、初頭子音が *r*, *l*, *s*-の 3つである場合は、子音字を縦に結合させる。

また、初頭子音連続のうち、末尾が *-'*, *-y*, *-r*, *-l*, *-v* の 5つである場合は、直前の子音字の下に縦結合させる。ただし、初頭子音連続が *g* と *y* の組み合わせである場合は、単語によって横に並べる場合と縦に結合させる場合があるので注意を要する。これらについては、転写の際にそれぞれ *g-ya*, *gya* と書き分けることによって対処しているのが現状である。

### 1.1.2 音節末子音連続

音節末子音連続の場合は、末尾が *-s* のケースと *-d* のケースがあるが、いずれの場合も横に子音字を並べるだけで、特に結合のための記号等は使用しない。

### 1.1.3 古代チベット語における留意事項

古代チベット語においては、-gs などの末子音連続が、縦に重ねられて書かれることがあるが、これらについては、今回は結合文字としては扱わない。また、dang དང་など、音声的には CVC 連続である場合にも、文字の上では da と nga の文字を縦に重ねて表記する (5) という例があるが、これらについても、今回は結合文字としては扱わない。

### 1.1.4 チベット語における略字の問題

チベット語の文献においては、略字体が見られる。これらについてはサンスクリットのリストにも存在しない結合文字があり、別途調査の必要があるが、今回は十分な調査を行うことが出来なかつたため、リストにはほとんど掲載されていない。

一つだけ挙げている用例は、*ଘ* という結合文字であるが、これは『藏漢大辞典』によると、*ଘଘ* または*ଘଘ* という単語に現れる結合文字で、これは *ଘଘଘ* という語の略字である。

## 1.2 サンスクリットの表記

サンスクリットの表記は、あらゆる子音連続を縦に順に重ねて表記する。

チベット語とサンスクリットが混在している場合は、同じ子音の連続であっても、横に並べるチベット語と縦に結合させるサンスクリットの間で、混乱が生じるケースがある。具体的には下記の子音連続において、こうした問題が起こることが明らかになっている。

*g-*: gta, gda, gna, gsha, gsa

*d-*: dka, dga, dgya, dgra, dpa, dba, dbra, dma

*b-*: bta, bda, bdra

*m-*: mn̄ga, m̄tha, mna

## 1.3 中国語など、外来語の表記

中国語におけるチベット文字表記でよく用いられるものとして、チベット語にない *f* 音を、*h* と *ph* の子音結合として表す表記法がある (5)。hpha という結合はチベット語にもサンスクリットにも存在しないものである。

現代チベット語やゾンカにおける外来語表記の調査を通じて明らかになったところでは、英語などの表記の際には、音節末子音連続の表記は、子音字を横に並べるという形で対処しており、特に縦結合は用いていない。

## 2. 調査方法

本研究では、チベット文字によって表される言語のうち、チベット語（古代チベット語を含む）とサンスクリットについて調査を行い、どのような結合文字が使用されるのかを一覧できるようにした。

調査のために利用した資料は下記の通りである。

サンスクリット－チベット語辞典、チベット語－サンスクリット辞典：

*Dictionnaire tibétain-sanscrit / par Tse-ring-ouang-gyal; reproduction phototypique publiée par J. Bacot.* (Paris: P. Geuthner, 1930).

『梵藏対照詞典』四川省アバ羌族自治州蔵文編訳局編纂、(甘肅民族出版社, 1989).

古代チベット語文献：

*Choix de Documents Tibétains conservés à la Bibliothèque Nationale*, III (Paris: Bibliothèque Nationale, 1990), IV (Tokyo: ILCAA, 2001).

古典チベット語、および現代チベット語については、網羅的な調査は行っていないが、筆者が収集した辞典類や新聞、雑誌などの印刷物に基づいて調査した結果を示した。

他に、中国が2002年12月にUnicodeコンソーシアムに提案したチベット文字表"BrdaRten"（བྲ୍ଦାର୍ତେନ）には、結合文字のリストが掲載されていたので、これも参考にした。"BrdaRten"の入手先は下記の通りである。

<http://std.dkuug.dk/jtc1/sc2/wg2/docs/n2558>

## 3. 結合文字リスト

調査の結果、子音字の単純な結合（母音記号をのぞく）のみで、総計1,093の結合文字が用いられていることが明らかになった。

現在このリストは最終的な校正の段階であり、後述するような誤記の可能性について検討中である。できる限り早く、web上で公開の予定であるので、本稿ではその形式についてのみ述べ、ここでは敢えて掲載しないこととする。

古典チベット語と現代チベット語のリストでは、サンスクリットの表記は除いてある。hphaなどの外来語表記は、現代チベット語のリストに記載している。

左端の数字は、結合文字の数を表すもので、1～6まである。

左から順に、総合リスト、サンスクリット(SKT)、古代チベット語(OT)、古典チベット語(CT)、現代チベット語(MT)、BrdaRtenのリストである。

なお、空欄は、調査した資料に該当する結合文字が存在しなかったことを示すものである。

また、サンスクリットの結合文字については、頻度も合わせて調査した。1,093件のうち、頻度が5以上のものは519件、4以上のものは572件、3以上のものは627件、2以上のものは777件であった。参考までに、サンスクリットの語彙を含む約14万語のデータベース（町田和彦氏による「ヒンディー語の辞書見出し語形検索」[http://www3.aa.tufs.ac.jp/~kmach/mrd/mrd\\_top\\_j.htm](http://www3.aa.tufs.ac.jp/~kmach/mrd/mrd_top_j.htm)）によると、結合文字の数は682に留まっており、これは今回の調査で出た頻度3以上のものと数がほぼ一致する。従って、頻度が2以下のものについてはチベット人によるサンスクリットの誤記である可能性もあり、注意を要する。

## 4. 考察

### 4.1 サンスクリットのチベット文字表記の問題点

本研究で調査した結合文字と「ヒンディー語の辞書見出し語形検索」（サンスクリットの語彙を含む）およびAPTEサンスクリット電子辞書（見出し項目数51352）をつきあわせてみたところ、チベット文字によるサンスクリット辞書には存在するが、データベースには存在しないという結合文字が存在することが分かった。例えば下記のような例である。

pth lkka lmm rnq rgbhya

最初のpthの例を調べてみると、すべてsthの誤記であることが明らかになった。これは、sとpの字形が類似している古い時代のネパーリー書体の文献から筆記して写したこと意味していると思われる。

lkka, lmmaなどは、kやmが二重化しているだけで、発音における子音の二重化がそのまま書記されるという、サンスクリット写本にもしばしば現れる現象であることが確認できた。

rnqの例は、parñiがparṇqiと誤記されたもので、長音を示すংが筆写の過程で読み間違えられたものであろう。

rgbhyaの例は、今回の調査において発見されたもっとも長い結合文字である。サンスクリットでは4子音の結合にすぎないが、チベット文字表記では6つの要素の結合である。サンスクリットの「力」を意味する単語ūrjの双数具格・与格・奪格形ūrgbhyāmの一例のみである。この形は、サンスクリットの辞書見出しで

は出現しないのが当然であるが、初学者の便を考えて採録されたのであろうか？

逆のケースとして、skh という結合は、APTE 辞書の見出し語で 11 例ほど存在するにもかかわらず、本研究で利用したデータの中には存在しなかった。チベットにもたらされたサンスクリット文献は仏典を中心があるので、語彙の偏りが大きいためにこういった現象が見られると考えられるが、この点に関しては、チベット文献におけるサンスクリットの語彙に関する調査を深める必要がある。

その他に、ndy のように明らかに誤記であるが、即座に元となっているサンスクリットを特定できない例がいくつもある。出典が辞書であるために、文脈が存在しないので、どのような誤記であるのかが明確でないものである。

今後は、辞書ではなくオリジナルのデータからこうした調査を継続していく必要がある。

## 4.2 古代チベット語の文字表記の問題点

本研究で調査した古代チベット語文献には、通常の古典、現代のチベット語表記には見られない結合文字

が存在する。今回の調査で明らかになった結合文字のリストは下記の通りである。

jra dzra dla pyva phva myva rkra rca rcha rtsha rha  
rhya lcha ltsha ltha skhya sca sja sphra sphra sya  
ssha hyva

古代チベット語の表記を可能にするためには、こうした結合文字も表示できなければならないし、またソートされる仕組みも用意しなければならない。

これら 23 文字のフォントについては、現在開発中であり、まもなく実装する予定である。また、ソートシステムについても、2 音節以上の場合にはサンスクリット語彙であると見なして、自動的にチベット文字順とサンスクリット順とを混在できるシステムを開発しているが、単音節の場合には手作業によるマークアップが必要な状態に留まっている。

これについても、チベット文献に用いられるサンスクリット語彙のリストを網羅すれば、自動化が可能になるが、膨大な作業量を要する作業であり、今後のチベット語辞書作成の過程の中で地道に継続していく予定である。

# 平安時代物語文の比較計量的研究

今西 裕一郎  
室城 秀之  
小西 貞則

## 【要旨】

本研究は、「源氏物語」および、それにならぶ平安時代の長編物語「うつほ物語」の、それぞれの文章について計量分析を行い、両者の比較を通して、「仮名」という我が国固有の文字で書かれた平安時代物語文の特質を明らかにしようとするものである。

本研究で対象とする二つの物語については、従来、「源氏物語」が平安時代の仮名文の完成を示す、もつとも完成度の高い文体を有する作品であるのに対し、その1世代前の「うつほ物語」は、いまだ発展途上にある未完成な仮名文によって書かれた作品であるという評価が下されている。

研究代表者（今西）は、これまで統計数理研究所の村上征勝氏との共同研究によって「源氏物語大成」に基づく「源氏物語」の本文データベースを作成し、それを用いた品詞別の計量分析によって「源氏物語」の文章の特性を解明しようと試みてきた。本研究はそれを承けて、同様の方法を「うつほ物語」にも適用し、まず「うつほ物語」の本文データベースを作成し、かかる後に種々の統計データを蒐集、分析する。そして「源氏」「うつほ」両作品に対する分析結果の比較検討により、平安時代物語から看取される仮名文生成発展の具体相を明らかにしようとするものである。

## 【位置付け】

### ① 「うつほ物語」とは

11世紀初頭、「源氏物語」が一条天皇の後宮に発し、傑作として広く宮廷に認知され、話題となる以前、宮廷社会で「源氏物語」のような位置を占めていたのは

「うつほ物語」であった。

全20巻からなり、全54巻の「源氏物語」には及ばないまでも、平安時代の作り物語としてそれに次ぐ規模を有する「うつほ物語」は、著者も成立年次も未詳であるが、成立はおおよそ天禄-長徳（970-999）頃かと推量され、作者の一人には和漢の文芸に通じた源順（みなもとのしたがう）が想定されてきた。

一条朝屈指の知識人藤原公任の家集「公任集」には、

円融院の御時にや、うつほの涼、仲忠とい  
づれまされり  
と論じけるに、しのなはしはすし（ママ）  
か方にやありけん、女  
一宮は仲忠が方におはしけるにや、いづれ  
をいるるなど  
あるに、ものないひそとおほせられければ、  
ともかくもい  
はでおはしけるをいひにおこせ給うければ  
沖つ波吹上の浜に家居してひとりすすしと思ふべ  
しやは  
（530）

という歌が収められており、その詞書には意味不明な点もあるものの、一条朝の前々朝、円融天皇の宮廷において、「うつほ物語」を代表する二人の作中人物、涼と仲忠の優劣といった話題が取り上げられていたことが知られる。

また、「源氏物語」の作者（紫式部）に先だって一条朝の後宮に仕えた清少納言の著「枕草子」にも、「うつほ物語」に関する記事は散見する。

まず、「ものがたりは」という段で、

ものがたりは、すみよし、うつほ。とのうつり、  
くにゆづりはにくし。(以下略)

と、「住吉物語」のつぎに「うつほ物語」が挙げられ、  
また、「かへる年の二月二十よ日」の段では、

御前に人々いとおほく、上人などさぶらひて、物語のよきあしき、にくき所などをぞ定め言ひそしる。涼、仲忠などがこと、御前にも劣りまさりたるほどなど仰せられける。「まづこれはいかに。  
とくことわれ。仲忠が童生ひのあやしさを、せちに仰せらるるぞ」などいへば、「何かは。琴なども天人の下るばかり弾きいで、いとわるき人なり。帝の御女やは得たる」といへば、仲忠が方人ども所を得て「さればよ」などいふに………。

のように、清少納言が仕えた中宮定子の前で定子その人も加わって、前に見た『公任集』と同様の、「涼、仲忠優劣論」が交わされた様子を、「うつほ物語」の内容に触ながら『公任集』以上に詳しく述べている。従来「後宮サロン」と称されることもあった後宮社会とは、かくも物語を玩ぶことに熱心な場所であった。「うつほ物語」は、「源氏物語」出現以前のそのような平安朝宮廷社会における物語のベストセラーであり、いわば『源氏物語』以前の『源氏物語』だったのである。

## ② 『うつほ物語』と『源氏物語』

そのような「うつほ物語」の後に世に出た『源氏物語』は、「うつほ物語」の影響なしにはその物語を始めることができなかつたように見受けられる。

『源氏物語』は、その第1巻桐壺巻で、主人公光源氏の誕生と、元服、結婚までの成長を語る。そこでは主人公源氏のたぐいまれな資質、美質がのべられ、余人には及びもつかない運命が暗示されるのであるが、その一々はそれだけを取り上げれば驚くほど「うつほ物語」的である。

「うつほ物語」の始めの3巻、「俊蔵」、「藤原の君」、「忠こそ」は、その順に物語が進行するのではなく、それぞれが独自の物語の発端となっている。つまり、「うつほ物語」には発端(書き出し)が、三つあるのであって、その後で三つの発端が融合するような形で物語は展開するのである。



興味深いのは、「うつほ物語」の、この三つの発端に語られる、それぞれの主人公にかかる設定が、光源氏のそれを彷彿させるようなものである、という点である。

第1巻俊蔵巻の書き出しは、主人公仲忠の幼時を次のように記す。

むかし、式部大輔、左大弁かけて清原の大君、皇女腹におのこ子ひとり持たり。その子、心のさときこと限りなし。父母「いとあやしき子なり。生ひ出でむやうを見む」とて書も読ませず、言ひ教ふることもなくて生ほし立つるに、年にも合はず丈たかく心かしこし。七歳になる年、父が高麗人にあふに、この七歳なる子、父をもどきて、高麗人と文を作り交はしければ、おほやけ聞こし召して、「あやしうめづらしきことなり。いかで試みむ」と思すほどに、十二歳にて冠しつ。

(第1巻 俊蔵)

これは、光源氏の、

七つになり給へば読書(ふみ)始などせさせたまひて、世に知らずさとうかしこくおはすれば、あまりおそろしきまで御覧す。

という設定、さらに、

そのころ高麗人のまゐれるがなかに、かしこき相人ありけるを聞こし召して、宮の内に召さんことは宇多のみかどの御試めあれば、いみじう忍びてこの御子を鴻臚館につかはしたり。

という、高麗の相人との対面の場面における、

文など作りかはして、今日明日帰り去りなんとするに、かくありがたき人に対面したるよろこび、かへりては悲しかるべき心ばへをおもしろく作りたるに、御子もいとあはれなる句を作り給へるを、限りなうめでたてまつりて、いみじき贈り物どもをささげたてまつる。

というくだりを想起させないではおかしい。

次いで第2巻「藤原の君」の次に掲げる冒頭では、

むかし、藤原の君ときこゆる、一世の源氏おはしまけり。童より名高くて顔かたち、心たましひ、身の才、人にすぐれ、学間に心入れて、遊びの道にも入り立ち給へるときに、見る人「なほかしこき君なり。帝となり給ひ、国知り給はましかば、天の下豊かなりぬべき君なり」と世界ござりて申すときに、よろずの上達部、親王たち、「婿に取らむ」と思はずなかに、時の太政大臣の一人娘に、御冠し給ふ夜、婿取りて、限りなくいたはりて住ませ奉り給ふほどに………。(第2巻 藤原の君)

まず主人公の「一世の源氏」という設定が光源氏と共通し、「学間に心入れて、遊びの道にも入り立ち給へる」という一文も、光源氏についての「わざとの御学問はさる物にて、琴笛の音にも雲居を響かし、すべて言ひ続ければことことしうたてぞなりぬべき人の御様なりける」という叙述と軌を一にする。さらに「国知り給はましかば、天の下豊かなりぬべき君なり」という世評は、光源氏に対する前掲高麗人の相人の予言、

國の親となりて帝王の上なき位に昇るべき相おはします人の、そなたにてみれば乱れ憂ふことやあらむ。おはやけのかためとなりて天下をたすくる方にて見れば、またその相たがふべし。

を連想させるし(内容は『源氏物語』の方が格段に複雑であるが)、加えて元服の夜、太政大臣の一人娘の婿になるという段取りも、

引き入れの大臣の皇女腹に、ただ一人かしづき給ふ御むすめ、春宮よりも御けしきあるを、おぼしわづらふ事ありける、この君にたてまつらんの御心なりけり。内にも御けしき給はらせ給へりければ、「さらばこの折の後見なかもるを、添ひ臥しにも」ともよほさせ給ひければ、さおぼしたり。

という、元服時の光源氏と左大臣の一人娘(後世、「葵上」と呼ばれる)との婚姻と同巧である。

最後に第3巻「忠こそ」の冒頭では、

かくて、また、嵯峨の御時に、源忠経ときこゆる左大臣おはしけり。また、右大臣橘千蔵と申すおはしけり。(中略) 御妻には一世の源氏、かたち

きよらなる名取り給へるが十四歳なるを得給ひて住み給ふほどに、十六歳といふ年の五月五日に、玉光り輝きたる男の、いとをかしげなるを生み給へり。名をば忠こそといふ。その子をまた思ふ人なく類なく限りなき御仲にて、これもかれもかたみに心ざし深くたまひ契りて経給ふほどに、忠こそ、生ひ出でくるままにかたちきよらなること限りなし。三つになるに心のさとくらうらうじきこと限りなし。父母撫で養ひ給ふこと限りなし。(中略) 忠こそ五つになる年の三月に、母君にはかに隠れ給ひぬべし。 (第3巻 忠こそ)

主人公忠こそ誕生の叙述「玉光り輝きたる男の、いとをかしげなるを生み給へり」が、光源氏の「世になくよらなる玉のをのこ御子さへ生まれ給ひぬ」に通ずることは言うまでもなく、さらに幼くして母に先立たれるという設定も、3歳で母桐壺更衣を失う源氏に共通する。

以上、『うつほ物語』冒頭3巻の設定に見られる光源氏との類似点、共通点のなかには、事柄の性質上、影響を受けなくてもおのずから似通う、その時代の共同幻想という面もないとは断言できないが、全体的に見て、『源氏物語』の出発に際しての『うつほ物語』の大きな感化は否定することはできないであろう。

『源氏物語』の作者が『うつほ物語』を読み、その内容にも共感を抱いていたことは、作り物語の優劣の論議を主題とした『源氏物語』第17巻「絵合」で、『うつほ』の「俊蔭」と『竹取物語』とを並べた上で、

俊蔭ははげしき波風におぼほれ、知らぬ国に放たれしかど、なほ指して行きける方の心ざしもかなひて、つひに人の朝廷(みかど)にも我が国にもありがたき才のほどを弘め、名を残しける古き心をいふに、絵の様も唐土(もろこし)と日本(ひのもと)とをとり並べて、おもしろき事どもなほならびなし。

という理由で、『うつほ物語』に軍配を上げていることからも、十分に察せられる。

### ③『うつほ物語』の退場

『源氏物語』の登場によって、円融朝以来一条朝初期まで後宮の文芸(物語)愛好者の人気の的であった『うつほ物語』は、急速にその地位を新参の『源氏物語』に譲らざるを得なくなつたようである。それほど

「源氏物語」は画期的な魅力あふれる作品であったということなのであろうが、たんにそれだけではなかった。

「源氏物語」は、「うつほ物語」を清少納言とともにこよなく愛した一条天皇の中宮定子の没落後、あらたに一条朝後宮の主となった、中宮彰子のもとで、紫式部によって構想され執筆された物語である。

ということは、「うつほ物語」を凌駕する傑作「源氏物語」は、たまたま中宮彰子に仕える紫式部によって書かれたという風に、単純に考えて事足りるような作品ではないということである。

「源氏物語」の作者の手になる「紫式部日記」を繙けば、「源氏物語」の制作に中宮彰子の父で当時の齊衡権力者、藤原道長が相当に関与し、読みようによつては「源氏物語」制作の最終段階は彼の指示によって動いていたように見えるほどである。

「紫式部日記」寛弘五年十一月の記事によれば、ある程度の完成を見た（それに「宇治十帖」が含まれていたかどうかは不明であるが）「源氏物語」は、中宮彰子のもとで清書され、同年十一月十七日の彰子の宮中還御の際の贈り物として、一条天皇に献上された。そして、一条天皇はその「源氏物語」に接し、その作者紫式部について「この人は日本紀をこそ読みたるべきれ。まことに才あるべし」と、賞讃の言葉を口にしたという。

うちの上の、源氏の物語、人に読ませ給ひつつ聞こし召しけるに、「この人は日本紀をこそ読みたるべきれ。まことに才あるべし」とのたまはせけるを……。

最高権力者の後援によって後宮で制作され、さらに収覧に供されるという形で物語が流布するということになれば、その物語はたんなる文芸愛好の産物ではなく、後宮政策という政治的所産であるといってよい。

我が娘彰子と一条天皇との間にすでに二人の皇子を儲け、外戚として不動の地位を確立した道長にとって、次に目指すべきは彰子を筆頭とする後宮の文化の振興であったと思われる。ライバル定子はすでに崩御し、その後見であった中関白家も没落した時点にあってなお、「枕草子」からその片鱗が窺われるような、定子の卓越した才知によって形成された後宮文化の余韻は色濃く残っていたはずである。中宮定子の後宮文化に取って代わる中宮彰子の後宮文化を創造すること、道長の後宮政策の次の目的はそこにあった。

このことは、定子に清少納言のような女房が仕えていたことに対抗するかのように、彰子に紫式部をはじ

め、赤染衛門、和泉式部等々、文芸的才知の持ち主を女房として抱え込んだことからも、容易に察せられる。紫式部が「源氏物語」を執筆したのは、このような環境においてであった。

ということであってみれば、「源氏物語」の登場は、その出来映えゆえに結果としてそれまでの人気作品「うつほ物語」を後宮から追い出したというよりも、そもそも「源氏物語」制作の意図自体が定子後宮の人気作品「うつほ」の追い落としにあった、と考えてもよいのではないか。

「源氏物語」はその内容の卓越もさることながら、藤原道長の政治的配慮、後宮政策に沿って、一条朝の宮廷に姿を現した作品ではなかつたか。とすれば「源氏物語」を読むということは、中宮彰子、ひいてはその父道長への忠誠の証という、政治的意味合いをも持つていた、ということになろう。

実際、その家集に円融朝での「涼、仲忠優劣論」を記した藤原公任が、時移って中宮彰子の第二子敦成親王誕生五十日の宴席で、「このわたりに若紫やさぶらふ」と、『源氏物語』を踏まえた冗談を紫式部に言いかけた（『紫式部日記』）という有名なエピソードも、従来この公任の言が中宮彰子方への追従と解されているように、「源氏物語」を読み、話題にすることが、文芸的であるよりも政治的なふるまいであったという事情を雄弁に語っている。

「悪貨は良貨を駆逐する」という喩えはこの場合適切ではないが、あたかもそのような勢いで、『源氏物語』は後宮から「うつほ物語」を追い出したのではなかつたか。

一条朝の後宮を制覇した「源氏物語」は、以後、宮廷社会で多くの人々によって読まれ、書写されて、今日なお鎌倉時代の古写本が少なからず伝存する。それに対して、影の薄くなった「うつほ物語」は、伝本の数も少なく、何よりも鎌倉時代はもとより室町時代の写本すら今日に残されていないという惨状である。この古写本の欠落という状況は、「うつほ物語」研究の大きな障害であり、「源氏物語」研究に比して「うつほ」研究が大きく遅れているのもこのせいであるといつても過言ではない。

実は、本研究における「うつほ物語」データベースの作成においても、劣悪な本文の処理に、「源氏物語」の場合とくらべて格段の、多くの時間と労力を費やすこととなつた。

#### ④ 『うつほ物語』と『源氏物語』の違い

平安時代以前、『万葉集』の編纂された奈良時代には、まだ後世にいう意味での「仮名」は存在しなかった。もちろん、日本語を表記するためにすでに漢字を「仮名」的に使用していたが、その字体は漢字そのままであり、それ故に、『万葉集』で用いられた漢字の仮名的用法は「万葉仮名」といわれる。

平安時代になって、漢字の草書体を仮名に用いるようになり、さらに漢字の原型を留めないまでに崩した字形が仮名として用いられるようになる。前者は「草(そう)の仮名」と呼ばれ、後者は「女手(おんなで)」と呼ばれた。

「女手」において、初めて「連綿体」(續け書き)が可能となり、字体の簡略化と相俟って書記の利便性が向上し、仮名文の発達を促すことになる。

このような仮名の発達と雁行して、仮名文は綴られるようになるのは当然の成り行きである。比較的初期の仮名文に属するものとして、『古今和歌集』の仮名序が挙げられるが、これは、本邦初の勅撰和歌集の序文という特殊な文章なので、今は考察の対象から外す。

『源氏物語』によって完成の域に達したと見なされる平安時代仮名文の、現存作品で古いものは『土佐日記』、『竹取物語』、ついで『うつほ物語』、『かげろふ日記』であり、その後に『枕草子』、『源氏物語』が続く。『土佐』、『竹取』を初期仮名文と位置づけるとすれば、『うつほ』、『かげろふ』は中期、『枕』、『源氏』を完成期の仮名文ということが出来よう。これらの諸作品の中でも、『うつほ』と『源氏』は膨大な言語資料であり、計量分析にも堪える言語量を有する作品である。

ところで、この二つの物語は、同じ平安中期の作り物語というジャンルに属しながらも、はなはだ異なった印象を与える作品である。

その相違は、作者の性別、すなわち男性作者の手になった物語と女性作者の手になった物語との差異と解されており、事実それが最大の要因であることは疑いない。

たとえば、『源氏物語』第25巻「螢」には、『うつほ物語』に触れた一節があり、そこで『うつほ物語』を代表する女性の登場人物「貴(あて)宮」(藤原の君のむすめ)が、次のように評されている。

うつほの藤原の君のむすめこそ、いと重りかには  
かばかしき人にて、あやまちなかめれど、すくよ  
かに言ひ出でたることもしわざも、女しき所なか  
めるぞ、ひとやうなるや。  
(螢 819頁)

この一文の趣旨は、玉上琢弥『源氏物語評釈』の言を借りれば、

貴宮は、多くの姫君たちの中でも特別に美しく、才たけて、しかも現代的であったので、早くから多くの求婚者が殺到した。それぞれに意をこめ、金を尽くして求めるのであるが、貴宮は平静そのもの。そして結局は東宮に入内する。確かにこの点では、貴宮は「おもりかにはかばかしき人にて」あやまちのない人である。

しかし、この『源氏物語』の作者は、この貴宮の生き方を非難する。あまりに賢さが過ぎて女らしくない。「心浅げなる人まね」をする女と同様、女としては好ましくない、困ったものであると言う。

ということであるが、それは換言すれば、男性作者が描き出した理想的女性像に対する、女性作者からの異議である。

『うつほ』、『源氏』両物語の違いは、物語られる素材、内容の違いとなって現れる面もあれば、類似の素材、内容を扱ってもその扱い方の違いとして現れる面もある。そしてそれらの差異は、それらを表現する仮名文の差異としても発現しているのではないか。

このような観点に立って、本研究は、まだ仮名文の歴史が一世紀に満たない時期の仮名文生成の様態を、『うつほ物語』と『源氏物語』との文章の比較研究を通して明らかにすることを目指した。

#### 【連携の成果】

「うつほ物語」本文データベース作成、およびそれに基づく計量分析を行うための、情報学との連携を行った。

#### ① 『うつほ物語』本文の選定

かつて『源氏物語』の本文データベースを作成するに際しては、底本に池田龟鑑著『源氏物語大成』校異篇に翻字されている大島本を採用した。それは今日もっとも信頼できる『源氏物語』のテキストとして定評があり、また「大成」索引篇の単語分割に従ってデータベースを作成する上で好都合であったからである。『うつほ物語』のデータベース作成においても、最初になすべきは底本の選定であった。

しかしながら、『源氏物語』と違って、本文の損傷著しい近世期の写本しか伝わらない『うつほ物語』には、すでに静嘉堂文庫蔵紀氏旧蔵本(図1)、関戸家

#### ④ 『うつほ物語』と『源氏物語』の違い

平安時代以前、『万葉集』の編纂された奈良時代には、まだ後世にいう意味での「仮名」は存在しなかった。もちろん、日本語を表記するためにすでに漢字を「仮名」的に使用していたが、その字体は漢字そのままであり、それ故に、『万葉集』で用いられた漢字の仮名的用法は「万葉仮名」といわれる。

平安時代になって、漢字の草書体を仮名に用いるようになり、さらに漢字の原型を留めないまでに崩した字形が仮名として用いられるようになる。前者は「草(そう)の仮名」と呼ばれ、後者は「女手(おんなで)」と呼ばれた。

「女手」において、初めて「連綿体」(続け書き)が可能となり、字体の簡略化と相俟って書記の利便性が向上し、仮名文の発達を促すことになる。

このような仮名の発達と雁行して、仮名文ば綴られるようになるのは当然の成り行きである。比較的初期の仮名文に属するものとして、『古今和歌集』の仮名序が挙げられるが、これは、本邦初の勅撰和歌集の序文という特殊な文章なので、今は考察の対象から外す。

『源氏物語』によって完成の域に達したと見なされる平安時代仮名文の、現存作品で古いものは『土佐日記』、『竹取物語』、ついで『うつほ物語』、『かげろふ日記』であり、その後に『枕草子』、『源氏物語』が続く。『土佐』、『竹取』を初期仮名文と位置づけるとすれば、『うつほ』、『かげろふ』は中期、『枕』、『源氏』を完成期の仮名文ということが出来よう。これらの諸作品の中でも、『うつほ』と『源氏』は膨大な言語資料であり、計量分析にも堪える言語量を有する作品である。

ところで、この二つの物語は、同じ平安中期の作り物語というジャンルに属しながらも、はなはだ異なった印象を与える作品である。

その相違は、作者の性別、すなわち男性作者の手になつた物語と女性作者の手になつた物語との差異と解されており、事実それが最大の要因であることは疑いない。

たとえば、『源氏物語』第25巻「螢」には、「うつほ物語」に触れた一節があり、そこで『うつほ物語』を代表する女性の登場人物「貴(あて)宮」(藤原の君のむすめ)が、次のように評されている。

うつほの藤原の君のむすめこそ、いと重りかには  
かばかしき人にて、あやまちなかめれど、すぐよ  
かに言ひ出でたることもしわざも、女しき所なか  
めるぞ、ひとやうなるや。  
(螢 819頁)

この一文の趣旨は、玉上琢弥「源氏物語評釈」の言を借りれば、

貴宮は、多くの姫君たちの中でも特別に美しく、才たけて、しかも現代的であったので、早くから多くの求婚者が殺到した。それぞれに意をこめ、金を尽くして求めるのであるが、貴宮は平静そのもの。そして結局は東宮に入内する。確かにこの点では、貴宮は「おもりかにはかばかしき人にて」あやまちのない人である。

しかし、この『源氏物語』の作者は、この貴宮の生き方を非難する。あまりに賢さが過ぎて女らしくない。「心浅げなる人まね」をする女と同様、女としては好ましくない、困ったものであると言う。

ということであるが、それは換言すれば、男性作者が描き出した理想的女性像に対する、女性作者からの異議である。

『うつほ』、『源氏』両物語の違いは、物語られる素材、内容の違いとなって現れる面もあれば、類似の素材、内容を扱ってもその扱い方の違いとして現れる面もある。そしてそれらの差異は、それらを表現する仮名文の差異としても発現しているのではないか。

このような観点に立って、本研究は、まだ仮名文の歴史が一世紀に満たない時期の仮名文生成の様態を、『うつほ物語』と『源氏物語』との文章の比較研究を通して明らかにすることを目指した。

#### 【連携の成果】

「うつほ物語」本文データベース作成、およびそれに基づく計量分析を行うための、情報学との連携を行った。

#### ① 『うつほ物語』本文の選定

かつて『源氏物語』の本文データベースを作成するに際しては、底本に池田龟鑑著「源氏物語大成」校異篇に翻字されている大島本を採用した。それは今日もっとも信頼できる『源氏物語』のテキストとして定評があり、また「大成」索引篇の単語分割に従ってデータベースを作成する上で好都合であったからである。『うつほ物語』のデータベース作成においても、最初になすべきは底本の選定であった。

しかしながら、『源氏物語』と違って、本文の損傷著しい近世期の写本しか伝わらない『うつほ物語』には、すでに静嘉堂文庫蔵紀氏旧蔵本(図1)、関戸家

旧蔵俊景本（図2）、板本（図3）等、複数の伝本が影印やマイクロフィルムで公開されているが、『源氏物語』における大島本に相当するような善本がなく、『源氏物語』に較べて本文の状況がはなはだ劣悪である。やむをえず、その中ですでに刊行された索引や校注書の底本として採用されて比較的善本であるという評価を有し、かつ紙焼き写真の入手可能な尊経閣文庫蔵前田本（図4）をデータベースの底本に採用することにした。

図1

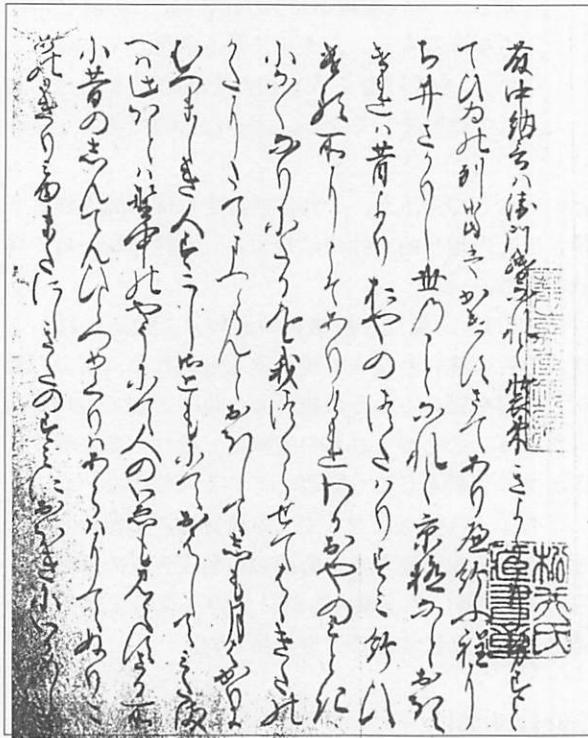


図2

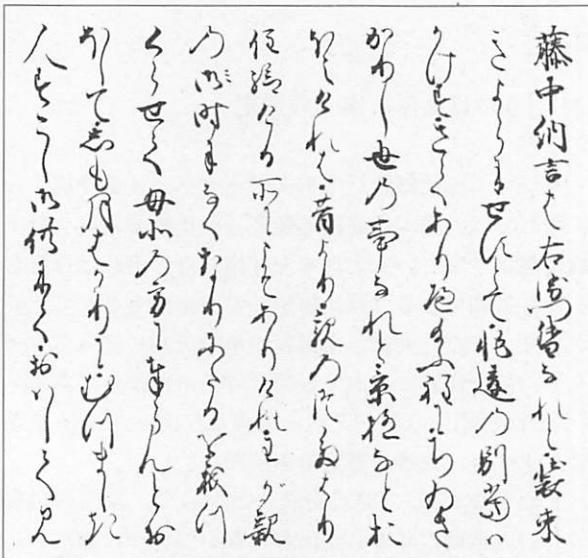


図3

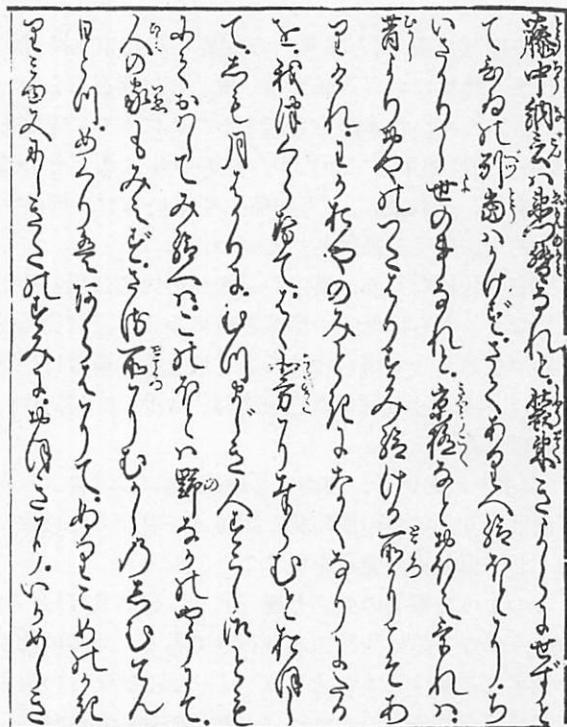
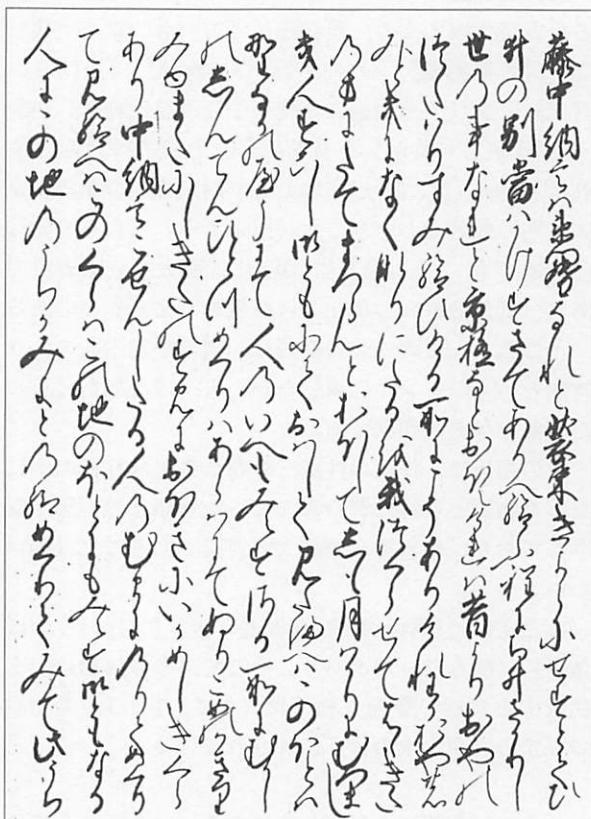


図4



## ② 前田本の翻字と単語分割

まず、研究分担者（室城）の協力により、前田本の翻字を行い、単語分割に進むためのデータベースの基礎とした（図5）。

ついで、翻字本文に対する単語分割を、『源氏物語』本文データベース作成の際に用いた『源氏物語大成』索引篇に基づく単語辞書によりコンピューターで行った（図6）。

図5

\*467藤中納言は衛門督なれと装束きよらにせずとてひゐの別当はかけすさてありへ給ふ程￥にちゐさかりし世の事なれと京極などおほえければ昔よりおやのつたはりすみ給ひける所にこそありけれわかおやのみときになくなりにたるを我つくらせてはゝきたの方にたてまつらんと￥おほしてしも月はかりにむつましき人すこし御ともにておはして見たまへはこのほとは野なかのやううにて人のいへもみえすぎる所にむかしのしんでんひとめくりはあらはにてぬりこめのかきりみゆまたにしきのすみにおぼきにいかめしきくらあり中納言こせんしたる人のむまにのりてめくれり見給へは￥このくらはこの地のほとにもみえす御ともなる人にこの地のうちかみよとの給めくりてみて此うち」一オなりと申ちかくよりてみたまへはくらのめくれりに人のかはねかすしらすあり￥おそろしと見つゝなをうちよりて見たまへはよになくいかめしき上かけたりその上のうへは￥かねをよりかけてふんしたりそのふんのむすびめにご治部卿のぬしのみなりしよりつけたり中納言み給ておとろきてこれはふみともならんむかしるいたいのはかせのいゑなりけるを一まいふみみ￥えすそのみちならぬきむなどとによの中にもちりこゝにものこりたる物をこれあけさせん￥とおぼすほとにかはらのほとよりとし九十はかりにてゆきをいたゝきたるやうなる女おきなはいには￥いきてまつこゝさらせ給へくとなくなそかく申とてみすいしんと￥へはなをまつこゝさせ給へおぼくの人とりころしつる」一ウくらなりまつ御らんせよこゝら￥の人のかはねをさせ給なん時ありやうは申さんといへはあやしかりてうちさりてたち￥まひたり＊468さてこれらか申やうこのむらはいみしくさかへて侍し所なりことしはたとせあまり￥みそとせにはまたらぬほとになんかくぼろひて侍そのゆへはむかしひとりこをもろこしにわたし給￥へりし人の御殿になんありしそのこをえまちえ給はてうせ給てのちにそのこかへりいま￥したりしきてこの殿をいときようにつくりてすみ給しほとにみむすめ一人なんもち給へりし￥そのむすめのちゐさくいますかりし時よりよにきこえぬをんしやうかくこゑなんたえさりしその御しやうかくを￥きく人はみなきもこゝろさかえてやまある物なく」二オなりおいたる物もわからくなりしかはきやうのうちの人は￥めくりてうけ給はりしそのむすめとつき時にり給しかはみかとをさして人かよはさてありしに大王￥みこみやとのゝはらのみよはひのつかひはあけたてはたちめくりてこともえつせてそ侍しし￥かありしほとにそのゝちはゞかくれ給にしかはかのみすめはきこえ給はすなりにきさりしかは￥この殿はかはら人さと人いりこたちてこほちはてゝたゞ一二ねんにかくなり侍にきやと￥もよろつの物ともとりしかこともなかめりしかはこのくらはかりは物とも侍らんと￥てまかりよる物はやかてたふれておぼくの人しに侍ぬよるは人にも見えはへらてむまにのりてきつゝゆつるうちをしつゝよめくりするやうになん侍かく」二ウ

図6

*467藤中納言	
は	し世の事なれ
×	なれ
衛	なと京×極
門	などおほえ
督	けれ
なれ	は昔
と	より
装束	おや
きよらに	の
せ	つたはり
す	すみ
とて	給ひ
ひゐ	けかる
の	所
別	に
當	こそ
はか	あり
けす	けれ
さて	わか
ありへ	おや
給	のみとき
ふ	に
程	なくなり
￥に	に
ち	たる
ゐ	
さ	
かり	

図7

*467藤中納言	
は	と京極
衛門督	などおほえ
なれ	けれ
と	は昔
装束	より
きよらに	おや
せ	の
す	つたはり
とて	すみ
ひゐ	給ひ
の	けかる
別	所
當	に
はか	こそ
けす	あり
さて	けれ
ありへ	わか
給	おや
ふ	のみとき
程	に
￥に	なくなり
ち	に
ゐ	たる
さ	
かり	

しかし、コンピューターによる分割は、機械的な作業であることに加え、

(1) 「うつほ物語」の本文の乱れ（判読不可能な文字列の存在）

(2) 「源氏物語」に存在しない語が少なからずすること

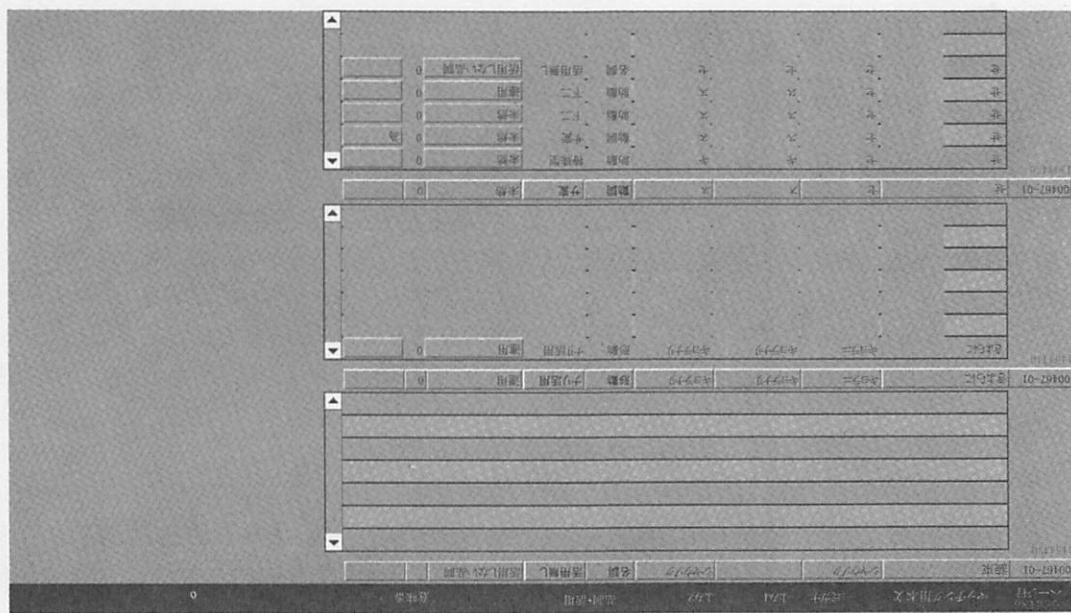
等の原因で、単語分割の精度は5割前後にとどまり、分割原稿の修正にかなりの時間と労力を要した。

修正作業は、コンピューター分割の初步的誤りを手作業で直し（図7）、その上で各単語を、前後の文脈付のKWIC形式で50音順に配列したデータ（図8）によって行った。

zたてまつらまほしけに 00520-13 れ/も/かれ/も/さて/あら/せ/ <zたてまつらまほしけに> おもは/れ/たる/を/見/たまへ/つ/へ  
 き/も/いと  
 zときのかうい 00506-07 /なと/は/このころ/こ/うみ/給へ/つる/ <zときのかうい> の/御もと/に/たてまつり/給へ/り/御ふみ/  
 は/我/かか  
 zなみをはまさて 00486-17 /より/おいならひ/つつ/ひめまつ/は/ <zなみをはまさて> ちよ/は/すき/なん/藤大納言/いは/の/う  
 へ/に/いま  
 zには 00509-03 しまさ/め/おんな/か/つくりこと/きこえさする/ <zには> なを/見/たてまつり/給へ/かし/それ/を/かの/御かた  
 /の  
 zよになかり 00491-10 とろく/こと/かきりなし/これ/は/また// <zよになかり> つる/て/かな/いかに/し/つる/こと/そ/や/宮  
 あ  
 zわれは人か 00509-01 け/さはかり/に/や/おはします/らん/宮/ <zわれは人か> かの/きみ/は/いと/いみしき/物を/きん/の/う  
 るし/の  
 z宮かた 00494-10 /は/よく/おはす/と/のみ/そ/見ゆる/おとと/ <z宮かた> は/すつし/中納言/いと/あはれ/と/思/きこえ/たり/  
 みところ  
 z又ふたつ 00526-08 り/なと/す/とう宮/の/こそ/いみしか/なれ/ <z又ふたつ> 人/ある/物/と/もしり/給は/て/としころ/に  
 あ 00472-04 かり/の/ほと/に/なん/うへ/さるへき/事/に/こそ/ <a> なれ/さる人/を/は/かねて/より/いたはり/なと/こそ/す  
 れ  
 あ 00476-13 かさ/せ/給へ/と/きこえ/給へ/は/はしたなけに/そ/ <a> める/きみ/なかたか/か/ため/に/は/これ/に/まさる  
 あ 00477-10 おほむせうそこ/あり/めつらしき人/の/たいらかに/ <a> なる/も/ありかたき/こと/の/さまさま/ものせ/らるる/  
 なる/を  
 あ 00478-01 いかて/と/おもひ/し/こと/を/さまさまに/そこ/に/ <a> なる/を/いと/うらやましく/その/わたり/の/こと/を/も  
 あ 00493-12 しろめたく/や/と/そ/いてや/なを/こそ/きこえ/も/ <a> なれ/うち/に/も/ありかたく/めつらしく/し/給/こと/の  
 あ 00494-13 のみ/こそ/は//たひたひ/ある/御ふみ/を/みれ/は/ <a> めれ/おとと/そこ/を/こそ/いかに/見/給/らん/よき人  
 あ 00510-10 /さ/なり/な/すた/さし/たまう/このむ/ねすみ/も/ <a> なり/まことを/は/おそろしき/物/は/彈正の宮/こそ/おは  
 す/めれ  
 あ 00510-12 /とう宮のわか宮たち/こそ/おそろしき/物/は/よに/ <a> めれ/いかやうに/おひいて/給は/ん/と/す/らん/ゆくさ  
 き/の  
 あ 00512-04 は/この/こ/を/うれし/と/思ふ/に/こそ/は/ <a> なれ/て/つたへ/む/と/や/おもふ/らん/おとと/さ  
 あ 00512-17 /おほく/なん/宰相のきみ/あやしき/御心/に/こそ/ <a> なれ/宮/は/御心/御さえ/も/猶/ことに/は/あひおはしま  
 す  
 あ 00513-15 しむし/に/たる/に/や/これ/も/おはせ/と/のみ/ <a> めれ/と/かくて/のみ/みえ/給/は/いまー/人/に  
 あ 00524-03 り/つる/いみしく/おひいて/ぬ/へき/もの/に/こそ/ <a> めれ/宮/の/そ/かやう/に/なり/しか/と/これ  
 あ 00526-07 はう/に/すへ/す/と/いふ/はかり/に/こそ/は/ <a> めれ/又/こと人/の/いふ/されと/これ/は/時々/人  
 あ 00527-09 は/こもり/待/ぬる/うへ/よき/事/に/こそ/は/ <a> なれ/かくもん/なと/心+b/に/いれ+c/て/ものせ/らるる/  
 は  
 あ 00529-04 ん/に/きこゆれ/は/宮/そや/右大将/の/御せうそこ/ <a> めり/や/とて/つけおとろかし/たてまつり/給へ/は/君/  
 としころ/ちかまもり  
 あいきやうつき 00489-05 と/ちいさく/ひちちかに/ふくらかに/ <あいきやうつき> たまへ/り/御とし/廿/ひとりのおほす/に/そ  
 /おはする/れいの/けちす  
 あいきやうつき 00494-08 /むつかし/と/うちむつかり/給/こゑ/ <あいきやうつき> 左大将/ちかく/て/おかし/と/きき/給/て/き  
 たのかた/に  
 あいきやうつき 00502-06 と/あてに/ものものしく/きよう/に/て/ <あいきやうつき> 給へ/り/御とし/五十四/されと/いと/わ  
 かく/みえ/給/左大将  
 あえなん 00471-10 つ/きか/せ/給/人/ちかく/きか/さら/ん/は/ <あえなん> とて/たまふ/その/みなみ/に/これ/より/は/ちいさ  
 き/所  
 あえ物 00472-12 /て/とこ/を/あまた/たいらかに/もたまへ/る/ <あえ物> は/そこ/に/も/けしうはあら/しか/し/と/の給へ/は/  
 まかて  
 あおき 00505-13 ん/の/よせきり/たり/しかみ/に/ひとつみ/ <あおき> しきしとも/に/つつみ/て/こえう/に/つけ/て/たてま  
 つり/給へ  
 あか 00469-12 かり/を/し/つづ/くら/を/あけ/させ/給/さらに/ <あか> す/そこ/に/二三日/おほく/の/人/を/ひきみ/て  
 あか 00469-17 れい/ひらか/せ/給へ/と/いのり/たまふ/されと/ <あか> す/人/の/申/やう/天下/いかに/いふ/とも/この  
 あかいろ 00498-05 まいる/おとな/四人/わらは/四人/おとな/は/ <あかいろ> の/かうきぬ/あや/の/すりも/あやかいねり/の/う  
 ちき/き/たり  
 あかいろ 00524-17 また/源中納言のきたのかた/の/御もと/より/ <あかいろ> の/おり物/の/からきぬ/からも/すりも/れう/の/  
 ほそなか/三へかさね  
 あかき 00481-12 /あき/給へ/り/しろき/御そ/の/はり/たる/に/ <あかき> か/うち/たる/たてまつり/て/御ゆか/のはしのかた/  
 に/みさりいり  
 あかき 00483-04 て/かなしけれ/は/ゆかしく/て/をき/つ/さて/ <あかき> うすやう/ひとかさね/に/御ふみ/たまはる/へき/人/  
 は/また/め  
 あかき 00497-08 つ/おなし/ひさく/して/しろき/御かゆ/一/をけ/ <あかき> 御かゆ/一/をけ/しろかね/の/たたいふ/やつ/に/御か  
 ゆ/の/あはせ  
 あかきみ 00523-03 し/その/御心/を/うしなは/せ/給へ/る/こそ/ <あかきみ> は/いと/うれしく/おほえ/給へ/こと人/は/かく/  
 おもひけた/せ  
 あかく 00470-08 /かた/に/よき/ほと/の/はしら/はかり/に/て/ <あかく> まろきもの/つみをき/たり/たた/くちもと/に/もくろ  
 く/を/かき/たる  
 あかつき 00510-14 /まことの/女御のきみ/を/さはかしかり/し/ <あかつき> に/見/たてまつり/し/は/や/いと/よく/に/たてま  
 つり  
 あかつきかた 00531-06 つら/ん/と/あそぶ/こと/かきりなし/ <あかつきかた> に/みな/少将/より/はしめて/かんたちめ/か  
 つき/給ふ/かんたちめ  
 あかみ 00502-09 /かゆをけ/の/ふた/に/は/すすし/の/いと/の/ <あかみ> たる/しりふた/と/いふ/もの/の/やうに/しなして/  
 おほひ  
 あかむたひ 00484-03 いれ/て/-おりひつ/みそ/と/かきつけて/ <あかむたひ> すこし/しろき/きぬ/を/ぬいめ/は/なく/て/そ  
 くひ/など  
 あからかなる 00472-17 て/たまひかる/やうに/みえ/給/御そ/は/ <あからかなる> からあや/の/うちき/の/御そ/ひとかさね/た  
 てまつり/て/御けうそく/に



10



6 

9)。大块儿同音異體的場合這個問題上，該當考究數的品詞指標來附加句子為 (圖 10) ⑩乙，而決定該數的品詞指標來表示乙句子 (圖 10) ⑩乙，而決

手裏拿着一本《道德經》。老子說「天地萬物生於有，有生於無」，「天地萬物生於有，有生於無」，老子說「天地萬物生於有，有生於無」，「天地萬物生於有，有生於無」。

③ 分割本文字以對象品詞情報的附加

④ 古典籍における誤写・誤読の痕跡をデータベースに残す

単語分割の精度に関しては、分割する本文にあらかじめ校訂を加え、本文表記の標準化を施しておけば、その精度はかなり向上することが予想される。研究分担者（室城）は、すでに活字校訂本『うつほ物語 全』（おうふう、平成7年）を出版しており、当初はその本文を使用して、データベースを作ることも考慮したが、しかしその本文には室城の解釈による校訂が施されているのみならず、読者の便宜を考えて、仮名表記の原文に対して多くの漢字が宛てられている。本研究においては、「うつほ物語」のデータベースを、単に今回の研究用にのみ有効なものとしてではなく、古典籍の伝本における誤写・誤読の実態をも探りうる多面的なデータベースとして構築するために、既製の活字本文ではなく、あらたに翻字本文を作成し、あえて前田本の本文の実態に即した作業を行った。

一般に、長い年月に渡って転写を経てきたテキストは、意識的、無意識的を問わず本文が改変され、あるいは本文の脱落ないしは誤写を蒙ることが避けられない。

就中、鎌倉、室町期の古写本がなく、近世の写本しか伝存しない作品の場合は、その弊著しいものがある。平安時代の作品中、『蜻蛉日記』、『枕草子』と並んで、『うつほ物語』はその最たるものである。

本文の改変や脱落という事態に対しては、あらたに原本もしくはそれに近い古写本が出現するような場合を除いて、原型は想像する以外に手の打ちようはないが、誤写・誤読による本文の転化の場合は、仮名の字形の類似を想定することで、ある程度の原型復元が可能である。

この方法を仮名文の本文解説に体系的に適用した最初の研究は、18世紀後期、天明年間（1781-1789）に出版された、坂徵の『かけろふの日記解環』であった（図11）。

『蜻蛉日記』の本文批判は、国学の祖、僧契沖の版本書き入れに始まり、江戸期の学者はみなその契沖説を吟味することから『蜻蛉日記』本文研究を出発させた。坂徵もまた然りであったが、彼はさらに次のように言う。

契沖諸本ニ直シノ無所ニ及デハ、止事ヲ得ズシテ憶説ヲモテアナグリ求メキ。ソノ求ムベキ手ヨリハ万ノカナノ転訛セルニヨリオシハカリテ、ヤ、本ニ復サレシヨリ外ニ術ナキコト治定セルニヨリテナリ。

そして、次のような具体例に言及する。

図11

ソノ轉ノ大例ヲイハ。うう らる れ 等ハ互ニテ  
ヤマリヤスシ。或ハにノ篇旁ヲ脱シテコト成。はヲニタヘ。  
はハひニ轉シ。ウハセニ交ジ。也ガ又世ニ化ヌ。カヤウノ類カツフル  
イトマナシ。甚シクシテハモトニ一字ガラんニ字ニシテ。ソクニ字ガ  
ヲ乃ノ二字ニアセ。或ハモトニ字カレハ二字ニベナシ。モく字ガ  
るとノ二字ニ成。又ハモトニ字ニモ誤ルコレラノルイヌスクナカラヌ  
サレドソレラハ猿イマシ。猿ノ字カモトニ二字ノカナニナリ也  
框ヲカナニウツゼルガ又轉シテナリヒツリナガレ。  
或ハモトニ字ワフるノ三字ニ轉セシ如キモ有。カク大ニ轉  
ズトハイヘトモ。只一轉ノ誤ハヨク又思ヒメグラシテサカノ  
ボリモトムルニ。猿タヨリナキニアラズ。

近代になって、この方法を体系的に記述したのは、池田亀鑑『古典の批判的処置に関する研究』（昭和16年、岩波書店）であり、『土佐日記』を対象とした本文復元研究の一環として、第二部「国文学に於ける文献批判の方法論」の第十五章には、「日本古典作品に於ける本文転化の諸類型とその実例」と題して、百数十頁にわたってその実例を例示し、解説を加える（図12）。

图12

う わ か の き う う あ う う う あ の か 示 う う  
一 一 一 一 一 一 一 一 一 一 一 一 一 一 一  
よ シ ふ う す シ サ 互 コ 画 コ 画 ケ ク キ 互 キ 互 オ ウ ミ ウ ミ ア  
チ タ ソ シ ウ ス シ サ 互 コ 画 コ 画 ケ ク キ 互 キ 互 オ ウ ミ ウ ミ ア  
う ろ か の 互 か か う う か の 互 み か か 示 が  
一 一 一 一 一 一 一 一 一 一 一 一 一 一 一  
め ん 互 み ま か ふ う か は の の 互 よ か か 示 が  
ミ ム ミ ミ マ ホ フ ハ ヒ ハ ハ ハ ハ ハ ハ ハ ハ ハ ハ ハ ハ ハ ハ ハ ハ ハ ハ

(一)かたかるへき

かたあるへき

(二)をむなかたをとこかた

おとこうた女うわ

(三) そのゝもなかに

そのことなきに

(四)たてしまかとそ

たてしまことそ

(五) ふたかりて

ふなさりて

(六) いかにかまたれまし

いかにとまたれまし

(七)はつかりかねそ

はつかりのねそ

(八)いそのかみもろたか

いそのかみもろたふ

九  
た  
ち  
ま  
か  
ひ

徒然草	静嘉堂文庫藏本
元祐集	本願寺本三十六人集その他
小大君集	宮内省圖書寮藏異木三十六人集傳小大君筆御藏切
堤中納言物語	群書類從本歌仙家集本他
異木赤染衛門集	東京文理大藏横山山清舊藏本
竹取物語	桂宮御舊藏本紫式部日記歌本
倭漢朗詠集	近衛家藏傳行成筆本その他の
木歌合卷	京都帝大附屬圖書館藏三手文庫藏今非似閑舊本
永承五年祐子内親王家歌合	十卷

近代に入ってからの日本文学研究でこの方法を積極的に採用したのは、前掲『かげろふの日記解環』の注釈姿勢を踏襲した柿本獎『蜻蛉日記全注釈』上下2巻(昭和42年、角川書店)であり、下巻に付載された「誤

「写一覧」(図13)は「底本に見える誤写の諸相のすべてが一覧できるようにした」労作で、『蜻蛉日記』の本文を考える際の貴重な情報である。

图13

「うつほ物語」の本文もまた、このような本文転化の観点から、精密に再検討されなければならない本文であり、そのことに鑑み、今回作成の「うつほ物語」本文データベースでは、単語の使用比率を考察の対象とする当面の研究に必須の要件ではないにもかかわらず、あえて転化本文の痕跡を残すこととした。

すなわち、校訂なしでは意味不明で読めない本文、単語に対しては、前掲室城秀之校訂の「おうふう」版によって読み仮名を施し、検索は校訂本文によって行う。しかし、検索結果は底本前田本の表記のまま表示されるというシステムである。

たとえば、俊蔭巻の

かたちのきようにはえのかしこきことさらにたとふへきかたなし。

という本文中、「きよう」は明らかに「きよ（清）ら」とあるべきで、前田本本文は「ら」と「う」の仮名字体の類似から「きよう」に誤られたものである。

今回のデータベースでは、「きよら」の語を検索すると、「きよう」の原表記のまま、この例が掲出されることになる。

この種のデータは、それを集計したり分類したりすることによって、当該写本の筆写の癖や精密度を知ることができ、その結果を本文批判に還元して、より高度な本文批判を可能ならしめることになるであろう。

ただし、本研究に必須の要件ではない情報を付け加えるという余分な作業を行ったため、「うつほ物語」本文データベースの作成に予想以上の時間と労力を要したことは、反省しなければならない。

### 【研究成果】

上記の要領で「うつほ物語」12巻までのデータベースの作成を終え、種々のデータを得ることが出来たので、それに基づき、「源氏物語」との比較を行った。

#### ① 統計的数値による「うつほ」と「源氏」の比較

##### 「うつほ物語」巻12までの品詞の比率

図14

		度数	パーセント	有効パーセント	累積パーセント
有効	名詞	28551	24.7	24.8	24.8
	代名	1217	1.1	1.1	25.9
	名副	1148	1	1	26.9
	代副	150	0.1	0.1	27
	名代	30	0	0	27
	動詞	20277	17.6	17.6	44.7
	補動	3832	3.3	3.3	48
	形容	4121	3.6	3.6	51.8
	形動	1134	1	1	52.8
	副詞	3403	2.9	3	55.5
	接続	258	0.2	0.2	55.7
	感動	136	0.1	0.1	55.8
	連体	1153	1	1	56.8
	連語	1915	1.7	1.7	58.5
	枕詞	7	0	0	58.5
	助詞	36424	31.5	31.7	90.2
	助動	11098	9.6	9.6	99.9
	接尾	14	0	0	99.9
	未詳語	154	0.1	0.1	100
	ナシ	1	0	0	100
合計	115024	99.8	100		
欠損値	システム欠損箇				
	435	0.4			
	合計	115459	100		

①-1 「うつほ物語」各巻における品詞使用率

図15-1

巻数		品詞						
		名詞	代名	名副	代副	名代	動詞	
俊陰	度数	4023	180	136	19	6	3221	
	巻数の %	23.10%	1.00%	0.80%	0.10%	0.00%	18.50%	
	総和の %	3.50%	0.20%	0.10%	0.00%	0.00%	2.80%	
藤原	度数	2928	108	89	5	3	1919	
	巻数の %	26.70%	1.00%	0.80%	0.00%	0.00%	17.50%	
	総和の %	2.50%	0.10%	0.10%	0.00%	0.00%	1.70%	
忠二	度数	1354	71	71	8	2	1277	
	巻数の %	19.90%	1.00%	1.00%	0.10%	0.00%	18.80%	
	総和の %	1.20%	0.10%	0.10%	0.00%	0.00%	1.10%	
春日	度数	1318	37	43	6	1	823	
	巻数の %	27.60%	0.80%	0.90%	0.10%	0.00%	17.20%	
	総和の %	1.10%	0.00%	0.00%	0.00%	0.00%	0.70%	
嵯峨	度数	2543	121	119	19	1	1854	
	巻数の %	23.30%	1.10%	1.10%	0.20%	0.00%	17.00%	
	総和の %	2.20%	0.10%	0.10%	0.00%	0.00%	1.60%	
祭使	度数	2446	73	96	18	5	1682	
	巻数の %	26.80%	0.80%	1.10%	0.20%	0.10%	18.40%	
	総和の %	2.10%	0.10%	0.10%	0.00%	0.00%	1.50%	
菊宴	度数	2778	91	69	8	7	1451	
	巻数の %	31.30%	1.00%	0.80%	0.10%	0.10%	16.30%	
	総和の %	2.40%	0.10%	0.10%	0.00%	0.00%	1.30%	
吹上	度数	1258	33	42	2	1	874	
	巻数の %	27.30%	0.70%	0.90%	0.00%	0.00%	18.90%	
	総和の %	1.10%	0.00%	0.00%	0.00%	0.00%	0.80%	
吹下	度数	2986	104	137	19		2202	
	巻数の %	24.00%	0.80%	1.10%	0.20%		17.70%	
	総和の %	2.60%	0.10%	0.10%	0.00%		1.90%	
あて	度数	1489	55	66	5	2	930	
	巻数の %	27.70%	1.00%	1.20%	0.10%	0.00%	17.30%	
	総和の %	1.30%	0.00%	0.10%	0.00%	0.00%	0.80%	
内侍	度数	4153	288	214	33	2	3208	
	巻数の %	22.10%	1.50%	1.10%	0.20%	0.00%	17.10%	
	総和の %	3.60%	0.30%	0.20%	0.00%	0.00%	2.80%	
沖白	度数	1275	56	67	8		836	
	巻数の %	25.90%	1.10%	1.40%	0.20%		17.00%	
	総和の %	1.10%	0.00%	0.10%	0.00%		0.70%	
合計		度数	28551	1217	1149	150	30	20277
		巻数の %	24.80%	1.10%	1.00%	0.10%	0.00%	17.60%
		総和の %	24.80%	1.10%	1.00%	0.10%	0.00%	17.60%

図15-2

補動	形容	形動	副詞	接続	感動	連体	連語	枕詞
473	646	215	543	29	30	283	309	
2.70%	3.70%	1.20%	3.10%	0.20%	0.20%	1.60%	1.80%	
0.40%	0.60%	0.20%	0.50%	0.00%	0.00%	0.20%	0.30%	
379	330	68	289	34	7	129	176	1
3.50%	3.00%	0.60%	2.60%	0.30%	0.10%	1.20%	1.60%	0.00%
0.30%	0.30%	0.10%	0.30%	0.00%	0.00%	0.10%	0.20%	0.00%
329	286	79	197	10	3	74	98	
4.80%	4.20%	1.20%	2.90%	0.10%	0.00%	1.10%	1.40%	
0.30%	0.20%	0.10%	0.20%	0.00%	0.00%	0.10%	0.10%	
140	157	39	129	6	3	38	66	
2.90%	3.30%	0.80%	2.70%	0.10%	0.10%	0.80%	1.40%	
0.10%	0.10%	0.00%	0.10%	0.00%	0.00%	0.00%	0.10%	
459	457	124	418	41	18	130	232	
4.20%	4.20%	1.10%	3.80%	0.40%	0.20%	1.20%	2.10%	
0.40%	0.40%	0.10%	0.40%	0.00%	0.00%	0.10%	0.20%	
269	324	73	188	21	9	46	154	2
2.90%	3.50%	0.80%	2.10%	0.20%	0.10%	0.50%	1.70%	0.00%
0.20%	0.30%	0.10%	0.20%	0.00%	0.00%	0.00%	0.10%	0.00%
235	280	90	167	27	3	66	109	
2.60%	3.20%	1.00%	1.90%	0.30%	0.00%	0.70%	1.20%	
0.20%	0.20%	0.10%	0.10%	0.00%	0.00%	0.10%	0.10%	
106	164	29	92	14	2	39	59	
2.30%	3.60%	0.60%	2.00%	0.30%	0.00%	0.80%	1.30%	
0.10%	0.10%	0.00%	0.10%	0.00%	0.00%	0.00%	0.10%	
450	447	101	316	40	20	77	221	3
3.60%	3.60%	0.80%	2.50%	0.30%	0.20%	0.60%	1.80%	0.00%
0.40%	0.40%	0.10%	0.30%	0.00%	0.00%	0.10%	0.20%	0.00%
220	213	41	139	23	6	30	70	
4.10%	4.00%	0.80%	2.60%	0.40%	0.10%	0.60%	1.30%	
0.20%	0.20%	0.00%	0.10%	0.00%	0.00%	0.00%	0.10%	
553	647	229	816	2	31	209	334	1
2.90%	3.40%	1.20%	4.30%	0.00%	0.20%	1.10%	1.80%	0.00%
0.50%	0.60%	0.20%	0.70%	0.00%	0.00%	0.20%	0.30%	0.00%
219	170	46	109	11	4	32	87	
4.40%	3.40%	0.90%	2.20%	0.20%	0.10%	0.60%	1.80%	
0.20%	0.10%	0.00%	0.10%	0.00%	0.00%	0.00%	0.10%	
3832	4121	1134	3403	258	136	1153	1915	7
3.30%	3.60%	1.00%	3.00%	0.20%	0.10%	1.00%	1.70%	0.00%
3.30%	3.60%	1.00%	3.00%	0.20%	0.10%	1.00%	1.70%	0.00%

図15-3

助詞	助動	接尾	未詳語	ナシ	合計
5606	1717		5		17441
32.10%	9.80%		0.00%		100.00%
4.90%	1.50%		0.00%		15.20%
3439	1034		8		10946
31.40%	9.40%		0.10%		100.00%
3.00%	0.90%		0.00%		9.50%
2164	766		2		6791
31.90%	11.30%		0.00%		100.00%
1.90%	0.70%		0.00%		5.90%
1542	434		2		4784
32.20%	9.10%		0.00%		100.00%
1.30%	0.40%		0.00%		4.20%
3373	1012		12	1	10934
30.80%	9.30%		0.10%	0.00%	100.00%
2.90%	0.90%		0.00%	0.00%	9.50%
2920	791		18		9135
32.00%	8.70%		0.20%		100.00%
2.50%	0.70%		0.00%		7.90%
2736	753	1	4		8875
30.80%	8.50%	0.00%	0.00%		100.00%
2.40%	0.70%	0.00%	0.00%		7.70%
1382	495	1	23		4616
29.90%	10.70%	0.00%	0.50%		100.00%
1.20%	0.40%	0.00%	0.00%		4.00%
4170	1106	1	29		12429
33.60%	8.90%	0.00%	0.20%		100.00%
3.60%	1.00%	0.00%	0.00%		10.80%
1624	451	3	6		5373
30.20%	8.40%	0.10%	0.10%		100.00%
1.40%	0.40%	0.00%	0.00%		4.70%
5993	2014	7	36		18770
31.90%	10.70%	0.00%	0.20%		100.00%
5.20%	1.80%	0.00%	0.00%		16.30%
1475	525	1	9		4930
29.90%	10.60%	0.00%	0.20%		100.00%
1.30%	0.50%	0.00%	0.00%		4.30%
36424	11098	14	154	1	115024
31.70%	9.60%	0.00%	0.10%	0.00%	100.00%
31.70%	9.60%	0.00%	0.10%	0.00%	100.00%

①-2 「うつほ物語」品詞別の頻出語

図16-1

名詞 語数 パーセント	代名 語数 パーセント		助詞 語数 パーセント		助動詞 語数 パーセント	
	コト	ヒト	モノ	ホト	オト	キミ
1225 4.29	コレ	252 20.71	ニ	6504 17.86	ス	1927 17.36
916 3.21	ソレ	190 15.61	テ	5385 14.78	ム	1523 13.72
657 2.30	ココ	189 15.53	ノ	4920 13.51	ナリ	1014 9.14
297 1.04	ワレ	106 8.71	ハ	4169 11.45	タリ	984 8.87
271 0.95	ソコ	60 4.93	ト	3364 9.24	キ	884 7.97
249 0.87	カレ	58 4.77	ヲ	2819 7.74	ヌ	782 7.05
249 0.87	タレ	55 4.52	モ	2203 6.05	リ	722 6.51
246 0.86	コナタ	41 3.37	ナト	859 2.36	ケリ	567 5.11
236 0.83	イツレ	38 3.12	ナム	858 2.36	ツ	507 4.57
226 0.79	ゾ	35 2.88	ヨリ	669 1.84	ヘシ	492 4.43
220 0.77	カシコ	27 2.22	ヤ	846 1.77	ル	259 2.33
217 0.76	アナタ	20 1.64	コソ	592 1.63	ラム	246 2.22
211 0.74	イツク	19 1.56	カ	551 1.51	ラル	240 2.16
201 0.70	オノレ	16 1.31	ゾ	441 1.21	シ	177 1.59
178 0.62	ナンチ	16 1.31	カナ	316 0.87	メリ	172 1.55
171 0.60	イツコ	14 1.15	ハカリ	246 0.68	マシ	161 1.45
157 0.55	マロ	13 1.07	トモ	238 0.65	サス	151 1.36
157 0.55	ソナタ	9 0.74	ツツ	194 0.53	コトシ	91 0.82
153 0.54	コレラ	7 0.58	タニ	163 0.45	ケム	76 0.68
148 0.52	ソレラ	6 0.49	シテ	159 0.44	シム	38 0.34
144 0.50	ワ	6 0.49	ノミ	159 0.44	マホシ	26 0.23
138 0.48	イツチ	5 0.41	マテ	151 0.41	ムス	8 0.07
137 0.48	カナタ	5 0.41	カシ	150 0.41	ラシ	8 0.07
135 0.47	イツカタ	4 0.33	ナ	89 0.24	ケラシ	2 0.02
133 0.47	タレタレ	4 0.33	シ	84 0.23	サイシャウノチュウ	2 0.02
129 0.45	アレ	3 0.25	ヘ	75 0.21	アシタツ	1 0.01
126 0.44	オノラ	3 0.25	ナカラ	72 0.20	アシハラ	1 0.01
115 0.40	アシコ	2 0.16	ヤハ	72 0.20	ウナキ	1 0.01
114 0.40	クソ	2 0.16	サヘ	67 0.18	オロシ	1 0.01
112 0.39	コ	2 0.16	モノヲ	39 0.11	カスカ	1 0.01
111 0.39	コチ	2 0.16	カハ	29 0.08	カラノオンソ	1 0.01
109 0.38	キンチ	1 0.08	テシカナ	29 0.08	クロムラサキ	1 0.01
102 0.36	クヤツ	1 0.08	モノカラ	24 0.07	コトマチオモテ	1 0.01
102 0.36	セウシヤウ	1 0.08	ハヤ	18 0.05	サイシヤウ	1 0.01
102 0.36	チユウシヤウ	1 0.08	ナント	12 0.03	サシム	1 0.01
97 0.34	ナニカシ	1 0.08	ヨ	10 0.03	サヨフケカタ	1 0.01
90 0.32	ヤヤ	1 0.08	カラ	7 0.02	サエモンノカミノキ	1 0.01
90 0.32	ランナ	1 0.08	モノカ	3 0.01	ジ	1 0.01
89 0.31			モノカハ	3 0.01	シコロクノミコ	1 0.01
89 0.31			コトシ	2 0.01	シシウテンノニヨウ	1 0.01
88 0.31			シモ	2 0.01	シラハリ	1 0.01
86 0.30			ナトカハ	2 0.01	ズ	1 0.01
84 0.29			ナテカ	2 0.01	ス」	1 0.01
84 0.29			ニシカナ	2 0.01	タタトホ	1 0.01
83 0.29			モタウフ	2 0.01	タヒスミ	1 0.01
82 0.29			モノノ	2 0.01	チノナミタ	1 0.01
79 0.28			イ	1 0.00	チユウナコン	1 0.01
79 0.28			カテラ	1 0.00	ツミシロ	1 0.01
79 0.28			カテラモ	1 0.00	テウフク	1 0.01
78 0.27			カモ	1 0.00	ト	1 0.01

図16-2

	動詞		形容詞		形容動詞		副詞	
	語数	パーセント	語数	パーセント	語数	パーセント	語数	パーセント
アリ	1378	6.80	ナシ	527	12.79	イカナリ	142	12.52
ス	979	4.83	オホシ	169	4.10	キヨラナリ	77	6.79
キユウ	709	3.50	アヤシ	167	4.05	アハレナリ	72	6.35
オモフ	641	3.16	ヨシ	166	4.03	コトナリ	69	6.08
ノタマフ	580	2.86	オナシ	154	3.74	セチナリ	59	5.20
ミル	548	2.70	イミシ	120	2.91	オホキナリ	41	3.62
イフ	529	2.61	カキリナシ	114	2.77	イタツラナリ	33	2.91
ナル	386	1.90	カシコシ	109	2.64	ニハカナリ	26	2.29
マキル	370	1.82	オモシロシ	99	2.40	ホノカナリ	26	2.29
オホス	324	1.60	フカシ	85	2.06	イササカナリ	23	2.03
タテマツル	288	1.42	メテタシ	83	2.01	オホイナリ	23	2.03
ハヘリ	268	1.32	カナシ	75	1.82	マメヤカナリ	22	1.94
モノス	245	1.21	ヒサシ	72	1.75	キヨケナリ	20	1.76
ツカウマツル	206	1.02	チカシ	66	1.60	メツラカナリ	18	1.59
シル	204	1.01	ウレシ	54	1.31	ハルカナリ	15	1.32
サフラフ	201	0.99	カヒナシ	51	1.24	ココロコトナリ	14	1.23
タツ	197	0.97	ヲカシ	46	1.12	ユタカナリ	14	1.23
イル	193	0.95	オホツカナシ	45	1.09	カヤウナリ	13	1.15
ミュ	174	0.86	タカシ	44	1.07	アカラサマナリ	12	1.06
オハシマス	163	0.80	メツラシ	41	0.99	オロカナリ	12	1.06
キク	163	0.80	イカメシ	40	0.97	サカリナリ	12	1.06
ツク	147	0.72	カタシ	40	0.97	サヤウナリ	12	1.06
マウス	137	0.68	アリカタシ	38	0.92	タタナリ	11	0.97
タマフ	134	0.66	ハヤシ	38	0.92	マレナリ	11	0.97
イツ	133	0.66	イトボシ	37	0.90	オホロケナリ	10	0.88
カク	128	0.63	ワカシ	37	0.90	サマサマナリ	10	0.88
オホユ	118	0.58	ニナシ	36	0.87	シツカナリ	10	0.88
カヘル	112	0.55	シロシ	35	0.85	ヲカシケナリ	10	0.88
ヒク	110	0.54	ヒトシ	35	0.85	ススロナリ	9	0.79
オハス	104	0.51	クルシ	34	0.83	サスカナリ	8	0.71
スム	101	0.50	フビシ	33	0.80	ウツクシケナリ	7	0.62
ハシム	100	0.49	ツラシ	32	0.78	サカサマナリ	7	0.62
コラヌス	95	0.47	アシ	29	0.70	サラナリ	7	0.62
タマハル	94	0.46	ハガナシ	25	0.61	タヒラカナリ	7	0.62
オモホユ	92	0.45	オソロシ	24	0.58	ノトカナリ	7	0.62
オモホス	86	0.42	チヒサシ	24	0.58	カタハナリ	6	0.53
ヌ	81	0.40	ヒロシ	23	0.56	コトワリナリ	6	0.53
ワタル	79	0.39	ススシ	22	0.53	シメヤカナリ	6	0.53
アソハス	76	0.37	コトモナシ	21	0.51	フサンナリ	6	0.53
ツクル	76	0.37	ヤンコトナシ	21	0.51	フヒンナリ	6	0.53
フ	76	0.37	ヨロシ	21	0.51	ツカナリ	6	0.53
キル	75	0.37	ハツカシ	20	0.49	イマサラナリ	5	0.44
ウケタマハル	74	0.36	アサマシ	19	0.46	ココロホソケナリ	5	0.44
マカツ	74	0.36	ナカシ	19	0.46	スミヤカナリ	5	0.44
マサル	73	0.36	ネタシ	19	0.46	ハツカナリ	5	0.44
イテク	72	0.36	ウルハシ	18	0.44	ムケナリ	5	0.44
ナス	69	0.34	コヨナシ	18	0.44	アサヤカナリ	4	0.35
アソフ	68	0.34	ツレナシ	18	0.44	アタナリ	4	0.35
キコシメス	64	0.32	クチラシ	17	0.41	アラハナリ	4	0.35
メス	62	0.31	ココロウシ	17	0.41	アルカナキカナリ	4	0.35

「うつほ物語」巻ごとの出現上位の語

図17

第一巻		第二巻		第三巻		第四巻					
	語数 パーセント		語数 パーセント		語数 パーセント		語数 パーセント				
ニ	1030	5.91	二	608	5.5	ニ	394	5.8	二	296	6.2
テ	1025	5.88	テ	532	4.8	テ	348	5.1	ノ	239	5
ノ	723	4.15	ノ	508	4.6	タマフ	280	4.1	テ	221	4.6
ハ	695	3.98	ハ	397	3.6	ト	276	4.1	ハ	175	3.7
ヲ	548	3.14	ト	380	3.5	ハ	261	3.8	ヲ	150	3.1
ト	539	3.09	タマフ	341	3.1	ス	206	3	ト	121	2.5
ス	423	2.43	ス	273	2.5	ノ	191	2.8	タマフ	119	2.5
タマフ	389	2.23	ヲ	247	2.2	ヲ	189	2.8	ス	91	1.9
モ	360	2.06	モ	196	1.8	モ	145	2.1	モ	78	1.6
ム	227	1.30	ム	149	1.4	コト	126	1.9	キ	55	1.1
キ	221	1.27	キユウ	137	1.2	ム	124	1.8	アリ	53	1.1
ヒト	204	1.17	アリ	129	1.2	ナリ	90	1.3	ム	45	0.9
アリ	193	1.11	ナリ	125	1.1	アリ	77	1.1	タリ	43	0.9
コト	186	1.07	タリ	99	0.9	ヒト	72	1.1	ハナ	41	0.9
ナリ	173	0.99	コト	96	0.9	キ	70	1	ナリ	40	0.8
ヌ	154	0.88	ヒト	93	0.8	ナム	55	0.8	ナム	39	0.8
コノ	150	0.86	モノ	92	0.8	ハヘリ	55	0.8	カク	36	0.8
タリ	136	0.78	リ	85	0.8	モノ	54	0.8	ミル	36	0.8
イフ	113	0.65	ナム	80	0.7	ヌ	51	0.8	ナト	33	0.7
ミル	101	0.58	ノタマフ	78	0.7	オトト	50	0.7	ヤ	33	0.7

第五巻		第六巻		第七巻		第八巻					
	語数 パーセント		語数 パーセント		語数 パーセント		語数 パーセント				
ニ	586	5.1	二	530	5.8	ニ	556	6.2	二	284	6.1
ハ	468	4.3	テ	457	5	ノ	534	6	テ	256	5.5
テ	435	4	ノ	409	4.5	テ	397	4.5	/	190	4.1
ノ	422	3.8	ハ	276	3	ハ	282	3.2	ヲ	128	2.8
タマフ	384	3.5	タマフ	260	2.8	ス	205	2.3	ハ	120	2.6
ト	320	2.9	ス	255	2.8	タマフ	201	2.3	ト	119	2.6
ス	262	2.4	ト	238	2.6	ヲ	196	2.2	ス	115	2.5
モ	221	2	ヲ	213	2.3	ト	138	1.5	タマフ	105	2.3
ヲ	208	1.9	モ	163	1.8	タリ	118	1.3	モ	58	1.2
コト	178	1.6	アリ	120	1.3	モ	105	1.2	ム	57	1.2
ム	155	1.4	タリ	98	1.1	ナト	97	1.1	キ	56	1.2
アリ	129	1.2	ナリ	82	0.9	ム	87	1	アリ	47	1
イト	129	1.2	ヨリ	81	0.9	アリ	79	0.9	コト	45	1
ナト	128	1.2	ヌ	73	0.8	キ	76	0.9	タリ	41	0.9
ナム	101	0.9	キユウ	70	0.8	ナリ	75	0.8	ヌ	40	0.9
ナリ	97	0.9	ヒト	69	0.8	リ	71	0.8	ナリ	37	0.8
キユウ	92	0.8	ム	68	0.7	ナム	61	0.7	ハヘリ	37	0.8
オモフ	89	0.8	ナム	66	0.7	ヒト	57	0.6	ナカタタ	35	0.8
タリ	78	0.7	ナト	63	0.7	ハヘリ	53	0.6	ヨリ	35	0.8
ヤ	78	0.7	カク	59	0.6	キミ	51	0.6	ミカト	34	0.7

第九巻		第十巻		第十一巻		第十二巻					
	語数 パーセント		語数 パーセント		語数 パーセント		語数 パーセント				
ニ	671	5.4	二	279	5.2	二	1022	5.4	二	273	5.5
テ	552	4.4	テ	276	5.1	テ	820	4.3	テ	210	4.2
ノ	520	4.2	ノ	276	5.1	ノ	794	4.2	タマフ	192	3.9
ハ	448	3.6	タマフ	199	3.7	ハ	742	3.9	/	191	3.9
タマフ	397	3.2	ト	155	2.9	ト	568	3	ハ	165	3.3
ト	391	3.1	ハ	147	2.7	タマフ	485	2.6	ス	132	2.7
ス	362	2.9	ヲ	119	2.2	ス	483	2.6	ト	128	2.6
ヲ	310	2.5	ス	118	2.2	ヲ	414	2.2	ヲ	111	2.2
モ	305	2.4	モ	105	1.9	モ	392	2.1	モ	96	1.9
キユウ	136	1.1	リ	65	1.2	ム	387	2.1	キ	61	1.2
ム	134	1.1	コト	57	1.1	アリ	323	1.7	コト	57	1.1
アリ	131	1	ナト	55	1	コト	237	1.3	ム	56	1.1
ナト	128	1	ヌ	55	1	ナム	189	1	ナム	55	1.1
ヤ	115	0.9	タリ	51	0.9	ナト	175	0.9	アリ	53	1.1
コト	112	0.9	アリ	45	0.8	ナリ	168	0.9	タリ	48	1
ナム	101	0.8	キユウ	43	0.8	コソ	165	0.9	リ	48	1
リ	101	0.8	モノ	42	0.8	タリ	164	0.9	ヌ	46	0.9
ヌ	98	0.8	カク	40	0.7	ヤ	162	0.9	ナト	41	0.8
オモフ	97	0.8	ヨリ	37	0.7	イト	149	0.8	ヨリ	41	0.8
ヒト	96	0.8	オモフ	36	0.7	ヒト	141	0.7	ノタマフ	37	0.7

①-3 「源氏物語」と「うつぼ物語」の品詞の使用率

『源氏物語』各巻における使用語上位20語一覧

図18-1

	01桐 語数 4804	02蒂 語数 9383	03空 語数 2187	04夕 語数 9564
	パーセント	パーセント	パーセント	パーセント
ニ	258 5.37	431 4.59	104 4.76	463 4.84
タマフ <sub>a</sub>	219 4.56	407 4.34	101 4.62	459 4.8
ノ	206 4.29	327 3.49	75 3.43	329 3.44
テ	189 3.93	321 3.42	66 3.02	328 3.43
モ	157 3.27	298 3.18	63 2.88	276 2.89
ヲ	138 2.87	243 2.59	56 2.56	252 2.63
ト	121 2.52	212 2.26	50 2.29	233 2.44
ハ	99 2.06	186 1.98	46 2.1	189 1.98
イト	70 1.46	161 1.72	42 1.92	147 1.54
ズ	66 1.37	157 1.67	41 1.87	141 1.47
バ	66 1.37	140 1.49	33 1.51	132 1.38
ス	58 1.21	132 1.41	33 1.51	130 1.36
ケリ	56 1.17	120 1.28	30 1.37	122 1.28
コト	48 1	108 1.15	28 1.28	108 1.11
ヒト	42 0.87	103 1.1	27 1.23	100 1.05
ヌ	39 0.81	100 1.07	23 1.05	97 1.01
リ	39 0.81	90 0.96	23 1.05	97 1.01
ナド	38 0.79	90 0.96	21 0.96	90 0.94
キ	36 0.75	ハベリ <sub>a</sub> 90 0.96	オモフ 19 0.87	ヌ 90 0.94
タリ	36 0.75	ヒト 89 0.95	ナド 18 0.82	キ 88 0.92

	05紫 語数 9406	06末 語数 6138	07賀 語数 5558	08宴 語数 2009
	パーセント	パーセント	パーセント	パーセント
ニ	431 4.58	282 4.59	259 4.66	100 4.98
テ	430 4.57	252 4.11	253 4.55	94 4.68
タマフ <sub>a</sub>	384 3.87	227 3.7	252 4.53	75 3.73
ト	304 3.23	217 3.54	198 3.53	タマフ <sub>a</sub> 71 3.53
モ	292 3.1	タマフ <sub>a</sub> 204 3.32	ノ 184 3.31	ト 65 3.24
ノ	286 3.04	モ 183 2.66	モ 160 2.88	ハ 50 2.49
ヲ	200 2.13	ハ 151 2.46	ハ 138 2.48	モ 49 2.44
バ	197 2.09	ヲ 143 2.33	ヲ 138 2.48	ヲ 41 2.04
ハ	187 1.78	タリ 112 1.82	バ 91 1.64	バ 35 1.74
ズ	147 1.56	ズ 105 1.71	ズ 77 1.39	タリ 29 1.44
ム	130 1.38	ム 97 1.58	リ 70 1.26	ズ 28 1.39
タリ	129 1.37	バ 96 1.56	イト 66 1.19	ナリ 27 1.34
イト	128 1.34	ケリ 77 1.25	タリ 64 1.15	ム 26 1.29
リ	113 1.2	イト 68 1.11	ナド 58 1.04	イト 25 1.24
ナド	111 1.18	アリ 65 1.06	ケリ 57 1.03	リ 25 1.24
ナリ	108 1.13	リ 57 0.93	ヒト 56 1.01	ヌ 21 1.05
ハベリ <sub>a</sub>	86 0.91	ヌ 54 0.88	ヌ 52 0.94	アリ 17 0.85
アリ	85 0.9	ナド 52 0.85	ム 52 0.94	コト 17 0.85
ヒト	85 0.9	ナリ 52 0.85	ナリ 50 0.9	ド 17 0.85
ケリ	83 0.88	オモフ 50 0.81	アリ 49 0.88	ベシ ヤ 16 0.8

	09葵 語数 9165	10賀 語数 9662	11散 語数 724	12須 語数 8390
	パーセント	パーセント	パーセント	パーセント
ニ	421 4.59	ニ 487 5.04	ノ 37 5.11	ニ 434 5.17
タマフ <sub>a</sub>	392 4.28	タマフ <sub>a</sub> 455 4.71	ニ 36 4.97	テ 368 4.39
テ	390 4.26	テ 404 4.18	テ 31 4.28	ノ 327 3.9
ノ	348 3.78	ノ 398 4.1	モ 29 4.01	タマフ <sub>a</sub> 317 3.78
モ	311 3.39	モ 337 3.49	タマフ <sub>a</sub> 28 3.59	モ 270 3.22
ト	282 3.08	ヲ 248 2.57	ハ 17 2.35	ト 207 2.47
ヲ	228 2.49	ト 219 2.27	ト 16 2.21	ヲ 192 2.29
ハ	224 2.44	ハ 199 2.08	ナリ 15 2.07	ハ 169 2.01
ズ	142 1.55	バ 150 1.55	ヲ 14 1.93	ズ 114 1.36
バ	141 1.54	コト 144 1.49	ケリ 11 1.52	バ 113 1.35
ナド	97 1.06	ズ 137 1.42	バ 11 1.52	コト 98 1.17
リ	97 1.06	リ 128 1.32	キ 10 1.38	リ 95 1.13
コト	95 1.04	イト 99 1.02	ズ 10 1.38	ム 90 1.07
タリ	94 1.03	ナド 97 1	コト 9 1.24	ケリ 82 0.98
ム	92 1	オボス 94 0.97	ヌ 8 1.1	キ 79 0.94
アリ	89 0.97	アリ 93 0.96	ベシ 8 1.1	ナド 79 0.94
イト	89 0.97	ナリ 93 0.96	タリ 7 0.97	ナリ 78 0.93
ナリ	85 0.93	ム 93 0.96	ナシ 7 0.97	ヌ 78 0.93
ヌ	84 0.92	ヌ 89 0.92	ヒト 7 0.97	ヒト 78 0.93
ヒト	80 0.87	ド 79 0.82	ル 7 0.97	イト 77 0.92

図18-2

13明	7862	14滑	6296	15蓋	4605	16闇	934
	語数パーセント		語数パーセント		語数パーセント		語数パーセント
ニ	417 5.3	ニ	338 5.37	ニ	226 4.91	ニ	53 5.67
テ	345 4.39	タマフ a	288 4.57	テ	197 4.28	テ	41 4.39
ノ	320 4.07	テ	278 4.42	ノ	195 4.23	ノ	38 4.07
ヲ	249 3.17	ノ	237 3.76	モ	185 3.58	モ	25 2.68
モ	245 3.12	モ	205 3.26	タマフ a	157 3.41	ト	22 2.36
ト	208 2.62	ヲ	186 2.95	ヲ	132 2.87	タマフ a	20 2.14
タマフ a	195 2.48	ト	143 2.27	ト	102 2.21	ハ	19 2.03
ハ	180 2.04	ハ	121 1.92	ハ	96 2.08	キ	18 1.93
コト	101 1.28	コト	103 1.64	ズ	75 1.83	ケリ	18 1.93
ズ	100 1.27	ズ	80 1.27	タリ	62 1.35	ヲ	17 1.82
タリ	96 1.22	バ	80 1.27	バ	60 1.3	バ	14 1.5
ム	91 1.18	ケリ	73 1.16	ケリ	55 1.19	コト	13 1.39
バ	73 0.93	イト	64 1.02	ナド	53 1.15	ナド	13 1.39
ヌ	71 0.9	ム	63 1	ム	53 1.15	アリ	12 1.28
キ	68 0.86	タリ	62 0.98	キ	52 1.13	ズ	12 1.28
リ	68 0.86	ナド	62 0.98	ヒト	46 1	ゾ	12 1.28
ナド	67 0.85	オボス	61 0.97	ナリ	41 0.89	ヌ	12 1.28
ヒト	64 0.81	リ	56 0.89	アリ	40 0.87	ド	11 1.18
アリ	63 0.8	アリ	54 0.86	ヌ	40 0.87	タリ	10 1.07
イト	63 0.8	キ	53 0.84	イト	38 0.83	コノ	9 0.96

17絵	3655	18松	4030	19薄	6023	20朝	3993
	語数パーセント		語数パーセント		語数パーセント		語数パーセント
ニ	190 5.2	ニ	203 5.04	ニ	304 5.05	ニ	195 4.88
ノ	182 4.98	テ	196 4.86	テ	245 4.07	タマフ a	169 4.23
テ	162 4.43	ノ	151 3.75	ノ	245 4.07	テ	160 4.01
タマフ a	142 3.89	タマフ a	130 3.23	タマフ a	227 3.77	ノ	148 3.71
ハ	109 2.98	ヲ	115 2.85	モ	207 3.44	モ	146 3.66
ヲ	106 2.9	ト	112 2.78	ト	153 2.54	ト	117 2.93
モ	92 2.52	モ	100 2.48	ヲ	145 2.41	ヲ	108 2.7
ト	84 2.3	ハ	83 2.06	ハ	143 2.37	ハ	92 2.3
コト	49 1.34	タリ	61 1.51	コト	134 2.22	ズ	53 1.33
ズ	44 1.2	バ	61 1.51	ズ	92 1.53	ナド	53 1.33
バ	43 1.18	ヌ	53 1.32	ム	78 1.3	バ	47 1.18
リ	41 1.12	ズ	50 1.24	バ	73 1.21	キ	46 1.15
イト	40 1.09	ケリ	47 1.17	ナド	67 1.11	タリ	43 1.08
タリ	35 0.96	キ	45 1.12	ヌ	64 1.06	ム	43 1.08
ナリ	35 0.96	ム	44 1.09	アリ	57 0.95	リ	39 0.98
キ	34 0.93	ナド	42 1.04	タリ	56 0.93	コト	38 0.95
アリ	33 0.9	コト	41 1.02	イト	54 0.9	ヒト	37 0.93
ス	31 0.85	ベシ	37 0.92	ベシ	52 0.86	オボス	35 0.88
ケリ	30 0.82	ナリ	33 0.82	ナリ	49 0.81	ド	35 0.88
ヌ	29 0.79	イト	31 0.77	ヒト	49 0.81	ヌ	33 0.83

21少	10038	22玉	8185	23初	2687	24朝	4040
	語数パーセント		語数パーセント		語数パーセント		語数パーセント
ニ	505 5.03	ニ	413 5.05	ノ	141 5.25	/	197 4.88
テ	464 4.82	テ	348 4.25	ニ	123 4.58	ニ	189 4.68
ノ	431 4.29	ト	292 3.57	タマフ a	109 4.08	タマフ a	164 4.06
タマフ a	404 4.02	ノ	285 3.48	テ	100 3.72	テ	146 3.61
モ	301 3	タマフ a	218 2.64	モ	86 3.2	モ	139 3.44
ト	267 2.86	ヲ	212 2.59	ハ	63 2.34	ハ	107 2.85
ヲ	254 2.53	モ	209 2.55	ト	54 2.01	ヲ	102 2.52
ハ	244 2.43	ハ	203 2.48	ヲ	54 2.01	ト	93 2.3
バ	152 1.51	バ	134 1.84	リ	45 1.67	ズ	69 1.71
コト	139 1.38	タリ	129 1.58	タリ	42 1.58	イト	58 1.44
ズ	139 1.38	ム	124 1.51	ズ	36 1.34	バ	58 1.44
タリ	109 1.09	ケリ	111 1.36	ナド	35 1.3	タリ	54 1.34
イト	107 1.07	ズ	108 1.3	アリ	29 1.08	アリ	50 1.24
ム	106 1.06	イト	95 1.16	イト	25 0.93	ナド	44 1.09
リ	101 1.01	ヒト	94 1.15	ケリ	25 0.93	ベシ	43 1.08
アリ	98 0.98	ヌ	92 1.12	バ	24 0.89	ム	42 1.04
ケリ	97 0.97	キ	83 1.01	ナリ	21 0.78	リ	39 0.97
ナド	93 0.93	ナリ	77 0.94	ヒト	20 0.74	ヒト	36 0.89
ナリ	91 0.91	アリ	72 0.88	ム	19 0.71	コト	35 0.87
オモフ	83 0.83	ナド	71 0.87	ゾ	18 0.67	ケリ	30 0.74
				ヌ	18 0.67		

図18-3

25類	3767	26常	4350	27筆	653	28野	3510
	語数パーセント		語数パーセント		語数パーセント		語数パーセント
ニ	178 4.73	ニ	196 4.51	ニ	44 6.74	ニ	177 5.04
ノ	182 4.3	テ	181 4.16	テ	31 4.75	ノ	188 4.79
タマフ <sub>a</sub>	155 4.11	ノ	174 4	ノ	31 4.75	テ	155 4.42
テ	134 3.58	ト	149 3.43	タマフ <sub>a</sub>	23 3.52	タマフ <sub>a</sub>	139 3.98
モ	110 2.92	タマフ <sub>a</sub>	141 3.24	ト	23 3.52	モ	97 2.78
ハ	99 2.63	モ	130 2.99	モ	18 2.76	ト	83 2.36
ト	93 2.47	ハ	115 2.64	ズ	12 1.84	ハ	77 2.19
ヲ	91 2.42	ム	102 2.34	ナリ	12 1.84	ヲ	62 1.77
タリ	64 1.7	イト	73 1.68	ヲ	12 1.84	バ	57 1.62
バ	64 1.7	ズ	73 1.68	イト	11 1.68	ズ	46 1.31
ズ	60 1.59	ヲ	70 1.61	バ	10 1.53	イト	44 1.25
コト	55 1.46	タリ	59 1.36	タリ	9 1.38	ナド	44 1.25
ナド	53 1.41	ナリ	58 1.33	ハ	9 1.38	タリ	43 1.23
ケリ	49 1.31	バ	56 1.29	コソ	7 1.07	リ	43 1.23
アリ	48 1.27	アリ	53 1.22	ヒト	7 1.07	ケリ	36 1.03
ヒト	45 1.19	コト	51 1.17	ココロ	6 0.92	アリ	33 0.94
ナリ	44 1.17	ヤ	49 1.13	コト	6 0.92	ム	31 0.88
イト	43 1.14	ヒト	46 1.06	ヤ	6 0.92	ナリ	30 0.85
リ	34 0.91	リ	43 0.99	アリ	5 0.77	ミル	28 0.8
ヌム	33 0.88	ナド	38 0.87	ケリ	5 0.77	カゼ	27 0.77
				ホド	5 0.77	ヌ	27 0.77
				ム	5 0.77		

29行	5233	30持	2795	31真	7232	32梅	3637
	語数パーセント		語数パーセント		語数パーセント		語数パーセント
ニ	263 5.03	ニ	127 4.54	タマフ <sub>a</sub>	347 4.8	ノ	204 5.61
タマフ <sub>a</sub>	222 4.24	モ	110 3.94	テ	323 4.47	ニ	189 5.2
テ	207 3.96	ノ	108 3.79	ニ	308 4.26	タマフ <sub>a</sub>	187 4.59
ノ	178 3.4	テ	104 3.72	モ	244 3.37	テ	157 4.32
ト	144 2.75	タマフ <sub>a</sub>	93 3.33	ノ	236 3.26	モ	100 2.75
モ	144 2.75	ト	79 2.83	ト	210 2.9	ヲ	94 2.58
ハ	137 2.62	ヲ	79 2.83	ヲ	190 2.63	ト	88 2.42
ヲ	137 2.62	ハ	69 2.47	ハ	176 2.43	ハ	74 2.03
コト	104 1.99	バ	47 1.68	ズ	109 1.51	リ	55 1.51
ム	77 1.47	ム	46 1.65	ム	108 1.49	アリ	48 1.26
バ	75 1.43	イト	44 1.57	コト	103 1.42	タリ	48 1.26
ズ	70 1.34	ズ	42 1.5	バ	96 1.33	ナド	44 1.21
ハペリ <sub>a</sub>	60 1.15	ナリ	41 1.47	アリ	89 1.23	ズ	42 1.15
イト	58 1.11	コト	38 1.29	イト	82 1.13	バ	42 1.15
ケリ	57 1.09	ケリ	35 1.25	ヒト	77 1.06	ム	34 0.93
ナド	57 1.09	アリ	32 1.14	ナド	71 0.98	コト	31 0.85
アリ	55 1.05	タリ	28 1	リ	70 0.97	ナリ	31 0.85
ベシ	55 1.05	ベシ	25 0.89	ナリ	69 0.95	イト	29 0.8
ナリ	48 0.92	ナム	24 0.86	ケリ	65 0.9	ベシ	28 0.77
ナム	47 0.9	ハペリ <sub>a</sub>	24 0.86	タリ	64 0.88	ヤ	27 0.74
リ	44 0.84	リ	24 0.86	ド	60 0.83	ヒト	28 0.71

33裏	4430	34上	20196	35下	20222	36柏	7925
	語数パーセント		語数パーセント		語数パーセント		語数パーセント
ニ	207 4.67	ニ	1064 5.27	ニ	998 4.94	ニ	354 4.47
ノ	208 4.65	テ	859 4.25	テ	882 4.36	タマフ <sub>a</sub>	338 4.26
テ	194 4.38	ノ	830 4.11	ノ	813 4.02	テ	329 4.15
タマフ <sub>a</sub>	176 3.97	タマフ <sub>a</sub>	789 3.91	タマフ <sub>a</sub>	790 3.91	モ	295 3.72
モ	138 3.12	モ	638 3.18	モ	633 3.13	ノ	280 3.53
ヲ	123 2.78	ヲ	565 2.8	ヲ	551 2.72	ト	229 2.89
ト	109 2.48	ハ	527 2.61	ハ	524 2.59	ハ	188 2.37
ハ	94 2.12	ト	426 2.11	ト	499 2.47	ヲ	188 2.37
ズ	56 1.26	ズ	305 1.51	ズ	305 1.51	バ	129 1.83
リ	55 1.24	コト	284 1.41	バ	282 1.39	ズ	124 1.56
ナド	51 1.15	バ	262 1.3	コト	265 1.31	コト	108 1.36
ナリ	49 1.11	アリ	258 1.28	イト	251 1.24	イト	97 1.22
バ	46 1.04	ム	224 1.11	ム	240 1.19	ム	88 1.11
イト	41 0.93	ナド	211 1.04	ケリ	218 1.07	ナド	85 1.07
ヌ	41 0.93	イト	205 1.02	タリ	216 1.07	アリ	83 1.05
ケリ	40 0.9	タリ	202 1	アリ	209 1.03	キ	75 0.95
コト	39 0.88	ナリ	180 0.89	ナリ	187 0.92	ケリ	73 0.92
タリ	39 0.88	ケリ	175 0.87	ヌ	189 0.84	ヌ	72 0.91
アリ	35 0.79	リ	175 0.87	キ	162 0.8	ナリ	70 0.88
ヤ	33 0.74	ヌ	168 0.83	ヒト	158 0.78	タリ	68 0.86
ム	32 0.72	ベシ	165 0.82	ナド	148 0.73	ヒト	62 0.78

図18-4

	37横	3694	38縦	2748	39縦	14017	40御	3719
	語数/パーセント		語数/パーセント		語数/パーセント		語数/パーセント	
ニ	189	5.12	ノ	180	5.82	ニ	622	4.44
タマフ <sub>a</sub>	183	4.95	ニ	143	5.2	タマフ <sub>a</sub>	585	4.17
テ	173	4.68	テ	130	4.73	テ	555	3.98
ノ	158	4.22	タマフ <sub>a</sub>	129	4.69	ノ	512	3.85
モ	115	3.11	モ	80	2.91	モ	481	3.43
ハ	93	2.52	ヲ	71	2.58	ト	431	3.07
ヲ	93	2.52	ハ	63	2.29	ヲ	385	2.8
ト	92	2.49	ト	50	1.82	ハ	323	2.3
バ	52	1.41	リ	38	1.38	ズ	242	1.73
イト	51	1.38	コト	36	1.31	バ	198	1.41
ズ	48	1.3	ナド	35	1.27	ム	181	1.29
ム	43	1.18	イト	29	1.06	コト	177	1.26
ナド	40	1.08	タリ	28	1.02	イト	174	1.24
コト	38	0.97	ケリ	25	0.91	アリ	155	1.11
ナリ	35	0.95	ズ	25	0.91	ヌ	145	1.03
リ	35	0.95	ナリ	21	0.76	ベシ	141	1.01
タリ	34	0.92	バ	20	0.73	リ	141	1.01
キ	33	0.89	ス	19	0.69	ナリ	135	0.98
ケリ	29	0.79	ヌ	16	0.58	ヒト	126	0.9
ヌ	28	0.78	ベシ	16	0.58	ケリ	121	0.86
アリ	27	0.73	ナム	15	0.55			オボス
ハベリ <sub>a</sub>	27	0.73	キ	14	0.51			タメツ
ヒト	27	0.73	ム	14	0.51			ミル

	41幻	42匂	2898	43紅	2517	44竹	8065	
	語数/パーセント		語数/パーセント		語数/パーセント		語数/パーセント	
ニ	229	5.34	ニ	153	5.68	ニ	138	5.48
ノ	201	4.69	ノ	148	5.49	テ	118	4.81
テ	172	4.01	タマフ <sub>a</sub>	115	4.27	タマフ <sub>a</sub>	115	4.57
モ	165	3.85	モ	104	3.88	ノ	90	3.58
タマフ <sub>a</sub>	147	3.43	テ	99	3.67	ハ	78	3.1
ハ	111	2.59	ヲ	83	3.08	ト	72	2.86
ヲ	108	2.52	ハ	74	2.74	ヲ	82	2.46
ト	99	2.31	ト	50	1.85	モ	52	2.07
ナド	70	1.63	ズ	36	1.34	バ	39	1.55
ズ	59	1.38	ヒト	33	1.22	ナド	34	1.35
キ	55	1.28	リ	28	1.04	ズ	33	1.31
コト	52	1.21	ケリ	27	1	キ	30	1.19
ケリ	46	1.07	ナド	27	1	イト	29	1.15
タリ	46	1.07	イト	26	0.98	ム	28	1.11
ベシ	45	1.05	ベシ	24	0.89	タリ	27	1.07
ム	43	1	タリ	23	0.85	ナリ	27	1.07
イト	41	0.98	バ	23	0.85	リ	26	1.03
リ	40	0.93	ム	23	0.85	アリ	24	0.95
ヌ	38	0.89	アリ	21	0.78	ヒト	23	0.91
バ	38	0.84	コト	21	0.78	キコユ <sub>a</sub>	20	0.79
ミル	36	0.84				ド	20	0.79
						ヤ	20	0.79

	45橋	7299	46椎	7289	47続	17436	48早	3557
	語数/パーセント		語数/パーセント		語数/パーセント		語数/パーセント	
ニ	382	5.23	ニ	344	4.72	ニ	794	4.55
テ	304	4.18	テ	300	4.12	テ	781	4.48
ノ	294	4.03	タマフ <sub>a</sub>	293	4.02	タマフ <sub>a</sub>	765	4.39
タマフ <sub>a</sub>	249	3.41	ノ	283	3.88	モ	615	3.53
モ	230	3.15	モ	268	3.85	ノ	600	3.44
ヲ	198	2.71	ト	179	2.48	ト	528	3.02
ト	177	2.42	ヲ	175	2.4	ヲ	470	2.7
ハ	142	1.95	ハ	172	2.36	ハ	404	2.32
ズ	118	1.59	ズ	122	1.67	バ	273	1.57
イト	87	1.19	ム	111	1.52	ズ	258	1.48
ナド	86	1.18	ナド	102	1.4	ム	257	1.47
タリ	84	1.15	バ	99	1.38	イト	195	1.12
バ	83	1.14	タリ	91	1.25	コト	191	1.1
ケリ	82	1.12	ナリ	80	1.1	リ	177	1.02
ハベリ <sub>a</sub>	76	1.04	イト	74	1.02	ナリ	160	0.92
ム	76	1.04	コト	72	0.99	アリ	158	0.91
キ	75	1.03	リ	68	0.93	ナド	155	0.89
ヌ	70	0.98	アリ	66	0.91	ヒト	153	0.88
ベシ	66	0.9	ケリ	65	0.89	オモフ	144	0.83
コト	64	0.88	ヌ	64	0.88	ヌ	143	0.82
						ベシ	32	0.9

図18-5	49宿	18835	50東	12996	51浮	14413	52婧	11797
		語数/パーセント		語数/パーセント		語数/パーセント		語数/パーセント
ニ	875	4.65	二	806	4.66	二	879	4.71
テ	730	3.88	テ	585	4.5	テ	593	4.11
タマフ <sub>a</sub>	662	3.51	ノ	450	3.46	ト	497	3.45
モ	657	3.49	ト	421	3.24	モ	460	3.19
ノ	639	3.39	モ	398	3.08	ノ	449	3.12
ハ	555	2.95	タマフ <sub>a</sub>	349	2.89	タマフ <sub>a</sub>	433	3
ト	512	2.72	ヲ	325	2.5	ハ	378	2.61
ヲ	427	2.27	ハ	310	2.39	ヲ	359	2.49
ズ	299	1.59	バ	240	1.85	ム	257	1.78
バ	292	1.55	ズ	201	1.55	タリ	227	1.57
ム	270	1.43	タリ	199	1.53	バ	219	1.52
アリ	248	1.32	ム	189	1.45	ズ	202	1.4
ナド	244	1.3	イト	189	1.3	ヒト	201	1.39
ケリ	219	1.18	ヒト	159	1.22	イト	188	1.29
コト	213	1.13	コト	155	1.19	コト	183	1.27
イト	198	1.04	オモフ	143	1.1	ナド	150	1.04
リ	186	0.99	ナリ	143	1.1	アリ	148	1.01
ナリ	185	0.98	ケリ	137	1.05	ナリ	148	1.01
タリ	184	0.98	アリ	126	0.97	ヌ	137	0.95
ベシ	165	0.88	ナド	120	0.92	ベシ	123	0.85
53手	14219	54夢	3555					
		語数/パーセント		語数/パーセント		語数/パーセント		語数/パーセント
ニ	718	5.05	二	204	5.74			
テ	600	4.22	テ	159	4.47			
ト	523	3.88	ト	123	3.48			
ノ	480	3.38	ノ	107	3.01			
モ	416	2.93	タマフ <sub>a</sub>	105	2.95			
ヲ	384	2.7	モ	92	2.59			
タマフ <sub>a</sub>	376	2.64	ハ	83	2.33			
ハ	353	2.48	ヲ	82	2.31			
ヒト	237	1.67	ム	62	1.74			
ム	219	1.54	バ	61	1.72			
タリ	211	1.48	コト	56	1.58			
バ	209	1.47	ハベリja	56	1.58			
ズ	202	1.42	キ	53	1.49			
キ	200	1.41	ズ	49	1.38			
コト	180	1.27	ナド	46	1.29			
イト	168	1.18	イト	44	1.24			
ケリ	165	1.18	タリ	43	1.21			
アリ	149	1.05	ヒト	39	1.1			
ナド	147	1.03	アリ	37	1.04			
ハベリja	145	1.02	ナム	37	1.04			

「源氏物語」品詞別頻出語

図19-1

名詞	65697		動詞		62897		形容		22263		形動		9063	
	語数	パーセント			語数	パーセント			語数	パーセント			語数	パーセント
コト	4496	6.84	アリ		3686	5.86	ナシ		1338	6.01	アハレナリ		797	8.79
ヒト	3463	5.27	オモフ		2357	3.75	イミジ		689	3.09	イカナリ		615	6.79
ホド	1571	2.39	オボス		1838	2.92	オホシ		550	2.47	コトナリ		402	4.44
ココロ	1483	2.26	米尔		1806	2.87	ヲカシ		536	2.41	サスガナリ		279	3.08
サマ	1348	2.05	ス		1665	2.65	アヤシ		508	2.28	カヤウナリ		200	2.21
モノ	1073	1.63	キコユ		1339	2.13	ヨシ		482	2.17	サマダマナリ		154	1.7
ヨ	965	1.47	ノタマフ		1124	1.79	フカシ		402	1.81	コトワリナリ		151	1.67
カタ	874	1.33	イフ		995	1.58	イタシ		362	1.63	サヤウナリ		141	1.56
ミヤ	723	1.1	ナル		930	1.48	イトホシ		339	1.52	ホノカナリ		138	1.52
ヒトビト	696	1.06	ミユ		889	1.41	チカシ		318	1.43	アナガチナリ		111	1.22
アリサマ	694	1.06	オボユ		673	1.07	オナジ		294	1.32	ヲカシゲナリ		111	1.22
ケシキ	644	0.98	オハス		667	1.06	ハカナシ		293	1.32	マメヤカナリ		103	1.14
ミ	630	0.96	マキル		660	1.05	ココログルシ		288	1.29	ニハカナリ		102	1.13
キミ	508	0.77	シル		532	0.85	クチヲシ		286	1.28	コマヤカナリ		89	1.09
トコロ	452	0.69	キク		529	0.84	カナシ		283	1.27	ラウタゲナリ		98	1.08
ケハヒ	355	0.54	モノス		525	0.83	ワカシ		241	1.08	ハナヤカナリ		95	1.05
ムカシ	348	0.53	ココチス		492	0.78	クルシ		237	1.06	キヨラナリ		92	1.02
ワザ	318	0.48	ハベリ		460	0.73	ウレシ		213	0.96	オロカナリ		80	0.88
ヲリ	314	0.48	オハシマス		441	0.7	メデタシ		211	0.95	アテナリ		79	0.87
オトド	311	0.47	サプラフ		378	0.6	ハヅカシ		203	0.91	コマカナリ		79	0.87
ナカ	291	0.44	ワタル		374	0.59	アサマシ		201	0.9	ノドヤカナリ		74	0.82
フミ	286	0.44	イヅ		328	0.52	ツラシ		198	0.89	ノドカナリ		69	0.76
トシゴロ	281	0.43	モテナス		316	0.5	ナツカシ		196	0.88	シヅカナリ		66	0.73
ナニゴト	269	0.41	スグス		258	0.41	ウシ		194	0.87	ウツクシゲナリ		65	0.72
ランナ	268	0.41	マサル		253	0.4	ココロウシ		191	0.86	シノビヤカナリ		65	0.72
ココチ	266	0.4	カク		247	0.39	コヨナシ		185	0.83	ツレヅレナリ		65	0.72
キン	251	0.38	オモヒイヅ		243	0.39	メヅラシ		169	0.76	ハヅカシゲナリ		65	0.72
ヨノナカ	247	0.38	フ		212	0.34	カギリナシ		166	0.75	キヨゲナリ		64	0.71
カギリ	231	0.35	シノブ		206	0.33	ココロボソシ		165	0.74	メヅラカナリ		64	0.71
コエ	230	0.35	マウス		201	0.32	ココロヤスシ		163	0.73	ココロコトナリ		62	0.68
ウチ	227	0.35	イル		199	0.32	フリナシ		161	0.72	コトサラナリ		60	0.66
カタチ	218	0.33	スグ		196	0.31	ヤンゴトナシ		156	0.7	ナカナカナリ		58	0.64
スヂ	214	0.33	ツカウマツル		194	0.31	オモシリシ		142	0.64	イマサラナリ		57	0.63
ノチ	212	0.32	タテマツル		191	0.3	カタジケナシ		132	0.59	モノアハレナリ		56	0.62
hana	202	0.31	ゴランズ		183	0.29	オボツカナシ		130	0.58	オモヒノホカナリ		54	0.6
ココロバヘ	198	0.3	オボシイヅ		176	0.28	コヒシ		127	0.57	ハルカナリ		54	0.6
ツイデ	197	0.3	ソフ		169	0.27	ワヅラハシ		125	0.56	エンナリ		52	0.57
トノ	189	0.29	アク		162	0.26	アリガタシ		121	0.54	ズロナリ		50	0.55
オマヘ	187	0.28	カハル		159	0.25	ウシロメタシ		120	0.54	トリドリナリ		49	0.54
トキ	187	0.28	カヘル		152	0.24	カヒナシ		120	0.54	アラハナリ		48	0.53
ツミ	181	0.28	ウチトク		151	0.24	ウツクシ		118	0.53	セチナリ		48	0.53
ナミダ	176	0.27	ウラム		150	0.24	コトゴトシ		114	0.51	シメヤカナリ		47	0.52
ソラ	173	0.26	ヰル		150	0.24	ツツマシ		113	0.51	ネンゴロナリ		47	0.52
ヤウ	169	0.26	キコユ		143	0.23	オソロシ		106	0.48	ココロヒツツナリ		46	0.51
ココロザシ	168	0.26	ナク		143	0.23	カタシ		106	0.48	ムゲナリ		46	0.51
アタリ	166	0.25	ソフ		142	0.23	ヒサシ		106	0.48	クルシゲナリ		45	0.5
モノガタリ	164	0.25	キコシメス		137	0.22	メヤスシ		105	0.47	ケザヤカナリ		45	0.5
イロ	156	0.24	ニル		133	0.21	ハシタナシ		104	0.47	サカリナリ		45	0.5
オヤ	155	0.24	ミス		133	0.21	ハカバカシ		102	0.46	マホナリ		45	0.5
カゼ	152	0.23	オモヒシル		131	0.21	ヲサナシ		102	0.46	サマコトナリ		44	0.49

『源氏物語』『うつほ物語』における  
主要品詞の使用頻度比較

図19-2

副詞	15616		助詞		1E+05		助動		43617	
	語数	パーセント	語数	パーセント	語数	パーセント	語数	パーセント	語数	パーセント
イト	4224	27.05	ニ	18331	15.32	ズ	5576	12.78		
カク	1054	6.75	テ	16054	13.41	ム	4761	10.92		
ナホ	818	5.24	ノ	14307	11.95	タリ	4312	9.89		
エ	659	4.22	モ	12096	10.11	ケリ	3637	8.34		
タダ	626	4.01	ト	10670	8.92	ナリ	3536	8.11		
スコシ	482	3.09	ヲ	9608	8.03	リ	3378	7.74		
ゲニ	458	2.93	ハ	9000	7.52	ヌ	3219	7.38		
マタ	438	2.8	バ	5506	4.6	キ	2997	6.87		
イトド	338	2.16	ナド	3849	3.22	ベシ	2895	6.64		
カウ	290	1.86	ド	2430	2.03	ツ	1485	3.4		
マシテ	284	1.82	ヤ	1994	1.67	ル	1453	3.33		
イカガ	253	1.62	コソ	1869	1.56	ス	1267	2.9		
ママニ	246	1.58	ナム	1802	1.51	メリ	978	2.24		
サラニ	238	1.52	ゾ	1698	1.42	サス	686	1.57		
ナカナカ	233	1.49	ヨリ	1346	1.12	ラム	670	1.54		
マダ	229	1.47	ツツ	799	0.67	ラル	646	1.48		
オノヅカラ	203	1.3	ノミ	738	0.62	ジ	521	1.19		
アマタ	197	1.26	カナ	670	0.56	ケム	479	1.1		
イカデ	187	1.2	バカリ	593	0.5	マジ	472	1.08		
マヅ	172	1.1	カシ	587	0.49	マシ	398	0.91		
ハタ	157	1.01	ダニ	586	0.49	マホシ	204	0.47		
カクテ	154	0.99	デ	576	0.48	ゴトシ	36	0.08		
マコトニ	150	0.96	マデ	507	0.42	ラシ	5	0.01		
ヤウヤウ	135	0.86	カ	503	0.42	シム	3	0.01		
ヤガテ	135	0.86	ガ	462	0.39	マス	2	0		
ウタテ	130	0.83	ナガラ	419	0.35	ムズ	1	0		
カナラズ	116	0.74	シ	378	0.32					
ツネニ	115	0.74	サヘ	324	0.27					
イヨイヨ	104	0.67	トモ	251	0.21					
ワザト	104	0.67	ヨ	206	0.17					
ツヒニ	84	0.54	カハ	173	0.14					
ヲサヲサ	84	0.54	モノヲ	163	0.14					
スペテ	82	0.53	ヤハ	157	0.13					
トカク	80	0.51	モノカラ	139	0.12					
イササカ	76	0.49	シテ	128	0.11					
フト	75	0.48	ヘ	93	0.08					
サテ	67	0.43	ゾ	92	0.08					
カタミニ	66	0.42	ナ	91	0.08					
シカ	66	0.42	ドモ	90	0.08					
ナベテ	63	0.4	バヤ	86	0.07					
モシ	63	0.4	ナ	65	0.05					
ヨニ	63	0.4	ナ	43	0.04					
オノオノ	59	0.38	ナム	40	0.03					
シヒテ	56	0.36	ガナ	28	0.02					
トミニ	56	0.36	テシカナ	26	0.02					
ヤヲラ	53	0.34	シテ	23	0.02					
モロトモニ	49	0.31	カラ	18	0.02					
ナド	48	0.31	ツツ	14	0.01					
カネテ	44	0.28	モノノ	14	0.01					
セメテ	42	0.27	ニシカナ	13	0.01					

図20

うつほ	源氏
助詞	31.5
名詞	24.7
動詞	17.6
助動	9.6
形容	3.6
副詞	4.1

## ② 数量化3類による分析結果

すでにデータを入力済みの『源氏物語』の数量化3類の分析表に重ねあわせる形で、「うつほ物語」12巻

分のデータを、形容詞、形容動詞、助動詞、助詞、副詞（各上位20語）の5品詞について数量化分析を試みた。

図21

源氏とうつほ 形容詞上位20語 2003/02/01

◆源氏  
×うつほ  
●形容詞

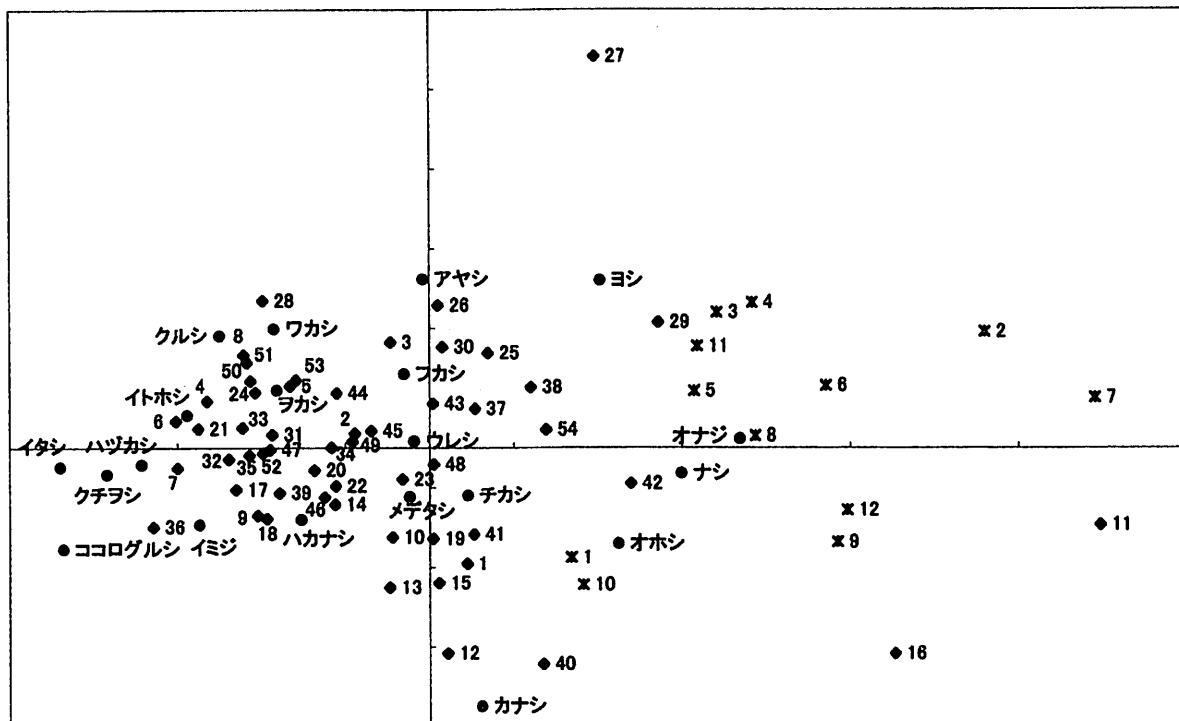


図22

源氏とうつほ 形容動詞上位20語 2003/01/25

◆源氏  
×うつほ  
●形容動詞

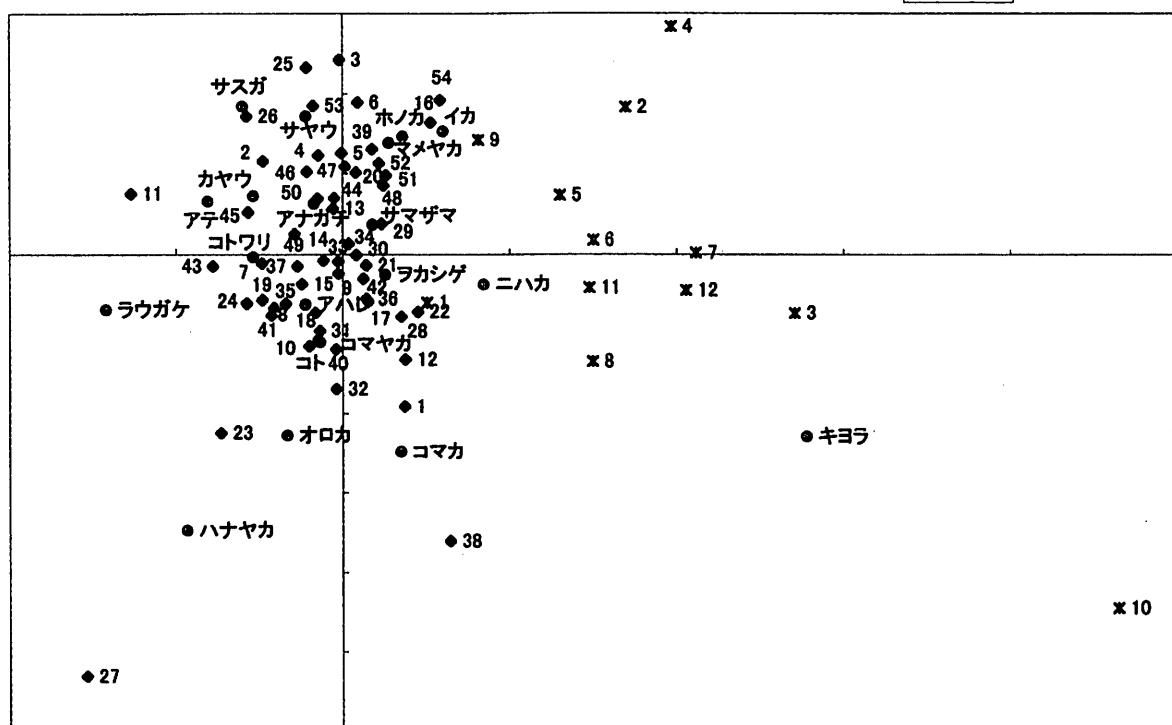


図23

源氏とうつほ 助詞上位21語 2003/02/01

- ◆源氏
- ✗うつほ
- 助詞

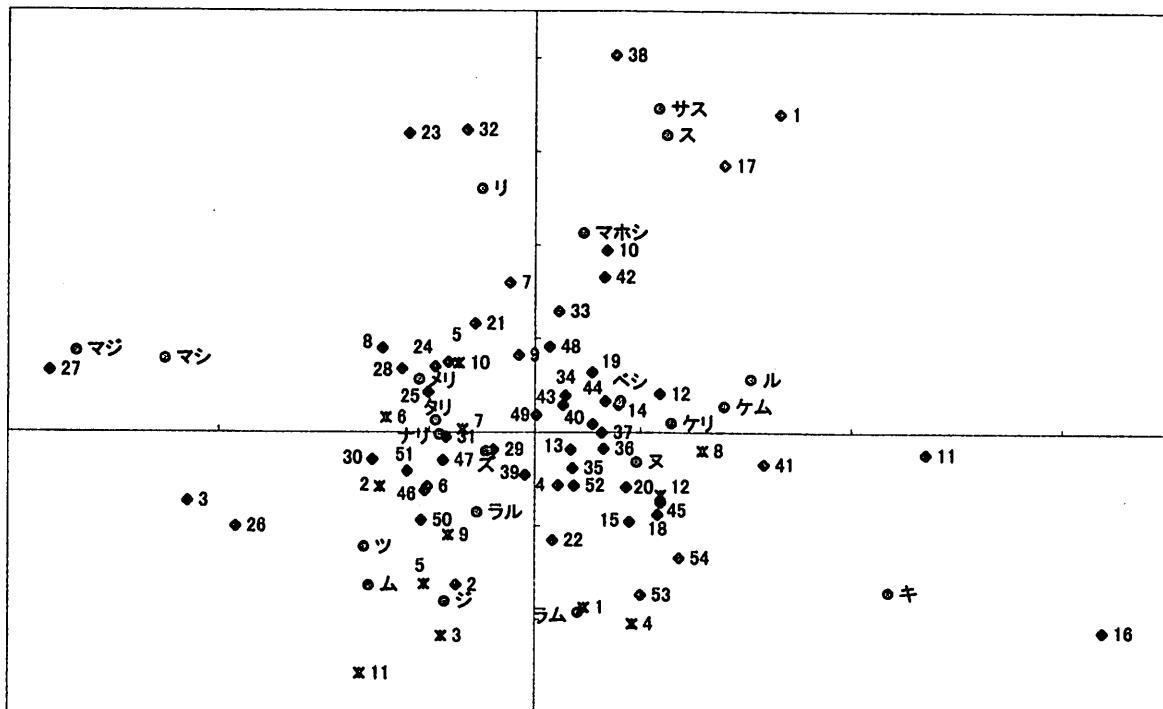


図24

源氏とうつほ 助詞上位30語 2003/01/25

- ◆源氏
- ✗うつほ
- 助詞

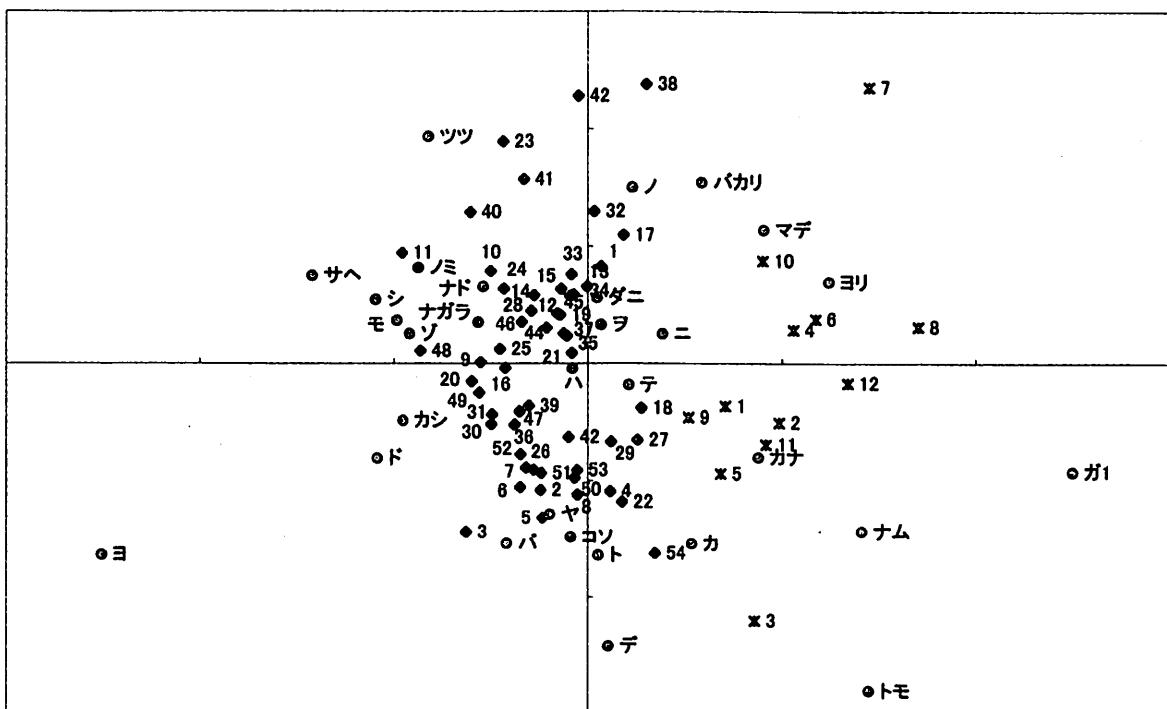


図25

源氏とうつほ 副詞上位20語 2003/01/25

- 源氏
- ✗うつほ
- 副詞

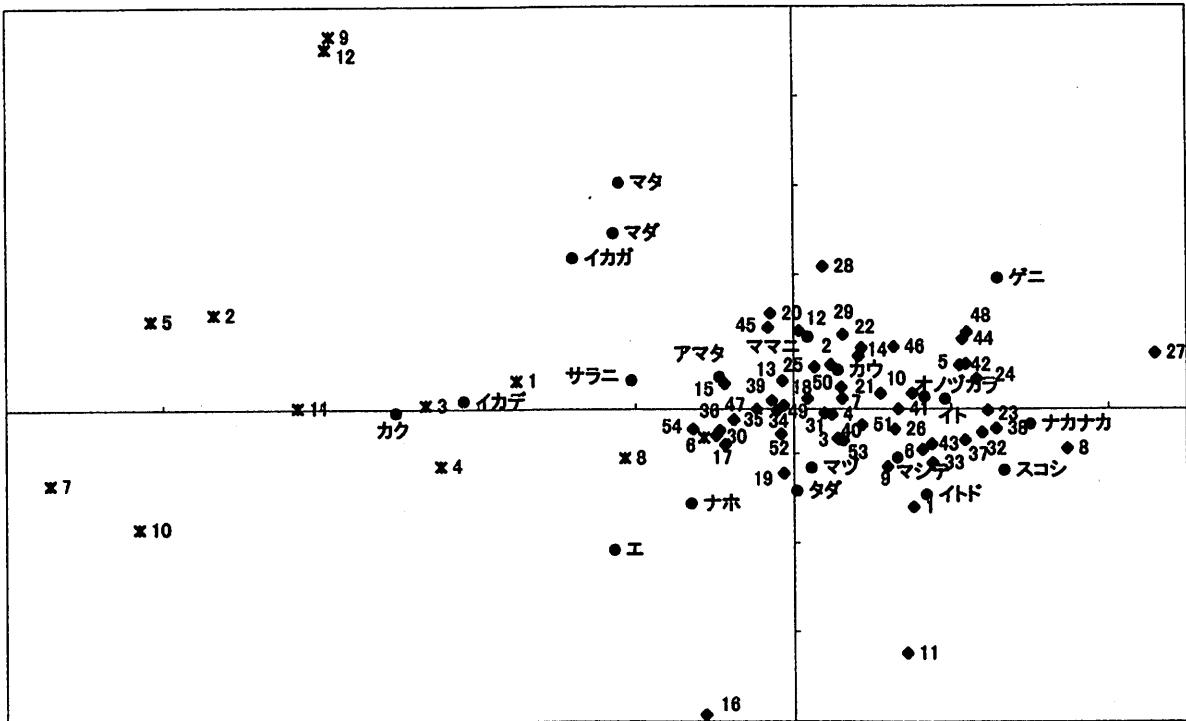


図21～25のいずれにおいても、×印で表示した「うつほ物語」の各巻は「源氏物語」の巻群と異なる集合を示している。その中でも特に図21（形容詞）、図22（形容動詞）、図25（副詞）の場合は、「源氏物語」の各巻が各品詞の語彙群と接近して塊を形成するのに對して、「うつほ物語」の各巻が語彙群から距離を置いて（したがって「源氏物語」の各巻からも距離を置いて）かたまるという顯著な分布が見られる。これは「源氏物語」の各巻がいずれも豊富な語彙を駆使した文章からなるのに対し、「うつほ物語」の文章が「源氏物語」ほど多様な語彙からなっていない、つまり単純、單調な文章であることを意味するであろう。

たとえば、図21の形容詞の場合、「うつほ物語」の各巻は図の右面に集中し、「よし」、「おなじ」、「なし」、「おほし」といった基本的、形式的な形容詞との親近を示すのに対し、「源氏物語」の各巻は図の中心から左面に密集し、多様な情意的形容詞との密接な相関を示す。

同じことは図25の副詞についてもいえる。「うつほ物語」が「かく」、「いかで」、「さらに」といった基本的、形式的な副詞と親近性を示すのに対し、「源氏物語」は様々な副詞と満遍なく関連をもち、かつ「うつほ物語」と疎遠な「なかなか」、「おのづから」、「げに」

といった情意性豊かな副詞との濃い相関を示すのである。

「うつほ物語」の文章が単純単調であり、「源氏物語」の文章が複雑微妙な陰影に富むことは、それらを読み比べた読者の等しく看取るところであるが、それは、品詞の度数に基づく数量化3類の分析によっても、明らかに裏付けられるのである。

### ③ 今後の課題と展望

数量化3類の分析結果にたいしては、もっと精密な読解を試みる必要がある。引き続き、「うつほ物語」を全体として「源氏物語」と比較するにとどまらず、複雑な成立過程が想定される「うつほ物語」の巻ごとの分析も併せて試み、成立過程の反映の有無をも検証したい。

また、本報告書では、作業の不手際から、主要品詞のうち、「名詞」、「動詞」のデータを提出できなかつたこと、および「うつほ物語」全20巻中12巻分のデータしか提示できなかつたこと、を遺憾とする。

さらに、分析方法の原点に立ち返って、品詞の使用頻度という尺度が、古典の本文の比較分析にどの程度

有効性を持ちうるかという点の吟味も要求されるであろう。そもそも品詞という単位、あるいは概念は、日本語にとって必ずしも本来的な要素ではなく、古人の言語意識に沿ったものであるかどうかを保証のかぎりではない。品詞に代わる、より効果的な尺度はないか。たとえば人物呼称、あるいは品詞の単位を越えた文字列等の探索が必要である。

# インド古典天文学書の研究と伝統暦プログラムの改良

矢野 道雄

## 【要旨】

インド古典文献や碑文に見られる年代を西暦に変換するには当時の天文学・暦法に基づいて暦を再現する必要がある。現代の天文学に基づいたインド暦のプログラムはすでに存在しているが、この種のプログラムは精密な現代の天文学理論を取り入れているために、その結果として復元される暦はかえって実際の古代の暦とは一致しない場合が多い。このような観点から、私は伏見誠氏の協力を得てインド古典天文学書の中で最も影響力が強かった『スールヤ・シッダーンタ』に基づいたインド暦プログラムを開発してきた。MSDOS 版の version 2 は 1993 年に京都産業大学の FTP server で公開し、すでに多くのインド学関係者に用いられてきた。

しかし最近では MSDOS のユーザは少なくなり、Windows や Mac OS 上で実行できるようなプログラムが要求されるようになったので、再度伏見氏の協力を得て、web page で容易に実行できるようにプログラムを書き換え、さらにいくつかの点でバージョンアップすることにした。バージョンアップに際して、『スールヤ・シッダーンタ』のサンスクリットテキストを電子テキストとして利用できるように入力しながら、もういちど読み直し、この古典テキストのアルゴリズムとプログラムのアルゴリズムを比較検討し、インドの数理天文学の特徴をよりよく理解するべくつとめた。さらに、インドの伝統的な太陽暦の要素も加えるために、インドで識者の意見を求めるに至った。

現代の最先端の道具を用いながら、できるだけ古典天文学書に忠実であろうとする本研究は、まさに「古典学の再構築」という名にふさわしいものであると思う。

## 【位置付け】

天文学・暦法はコンピュータがもっとも活躍する分野のひとつであり、無数の便利なソフトが有料あるいは無料で公開されている。インドの暦についてもすでにいくつかの web page で参照することができるが、それらが提供しているのはいずれも現代の西暦とインド暦の相互変換であり、現代の天文学の理論に基づいている。しかし古代の暦日は古代天文学の方法で作成された暦に基づいているのであるから、現代天文学の方法で逆算して得られる暦の日付と一致するとはいえない。とくに『スールヤ・シッダーンタ』では月の「出差」(evection) を考慮に入れないと、毎月の上弦または下弦あたりで現代天文学との相違が大きくなり、そのため日に相違が生ずることがある。本研究の最大の特徴は、紀元後 6 世紀ごろから近世にかけてインドで最も広く流布していた天文学書である『スールヤ・シッダーンタ』の理論とアルゴリズムにほぼそのまましたがってインド古代の暦を復元しようとしている点にあり、このようなプログラムは世界で初めてである。

しかし本研究開始以前は、このプログラムはまだユーザーにとって十分に馴染みやすかったわけではなく、ある程度の予備知識がないと活用しにくいものであった。そのためか、2000 年 2 月 29 日に私の webpage 上で実行できるようにした暫定版を公開した直後は、それほど大きな反響はなかった。したがって、web page のレイアウトや説明文書もふくめて、このプログラムの利用の仕方をいっそユーザーフレンドリーにする必要があることを知った。本研究テーマの科研費申請は平成 12 年度で終わったが、それ以後もユーザからの質問を参考にしながら次第にバージョンを上げていった。

## 【他領域との連携による成果】

インドの古典文献や歴史資料を扱っている研究者からしばしばインドの日付について質問を受けることがある。とくに MSDOS 版の PANCANGA を公開して以来、e-mail によって碑文や写本の年代の確認を求められることが多くなった。そのような場合私は喜んでプログラムを走らせ、自分でも結果を確かめることにしてきた。それによってプログラムの信頼性を評価することができるからである。すでに内外の多くの研究者の質問に答えてきたが、ほとんどすべての場合このプログラムの結果が有効であることがわかった。したがって、プログラム自体はもうこれ以上改良する余地はほとんどなく、むしろ「使いやすい」ものにするための工夫と、そのための親切なドキュメントを準備することの方が大切であるということが分かった。実際にインドの識者と議論したときも同様の印象を得た。

## 【研究成果】

ハーバード大学に留学中の伏見誠氏に依頼していた、TurboPascal 用のプログラム PANCANG2.PAS を CGI で実行できるような Perl のスクリプトに書き換えるという作業は完了した。2 種類のプログラムを走らせた結果の比較、およびインドで入手した暦との比較については、報告論文「インド暦プログラム PANCANGA について」(『古典学の現在』第 1 号(平成 12 年 3 月、35-52 頁)として発表した。またプログラムの構造を十分に理解しなければいけないので、私自身も Perl によるプログラミングの勉強を始めた。さらにその後のバージョンアップに関しては、『第 I 期 公募研究論文集』(平成 13 年 8 月 13 日)の「インド古典天文学書の研究と伝統暦プログラムの改良」で詳しく述べた。

これと平行して、まずこのプログラムが『スールヤ・シッダーンタ』の理論にどこまで忠実であるかを確認するために、『スールヤ・シッダーンタ』のサンスクリット原典をもう一度丁寧に読み直す必要があると感じたので、このテキストを電子化し、和訳しながらプログラムとの比較を行った。とくに『スールヤ・シッダーンタ』には、太陽、月、および惑星の補正計算において周転円の大きさが象限によって変化するという独特なアイデアがみられるが、われわれのプログラムではこれをいちおう無視しうるものとみなして平均値を採用したので、この点に関して考察した。この特異な方法が実際の計算にどの程度影響を与えていたかを追跡する必要があるように思われた。インドの識者との議論でもこのことは話題になったが、古代の暦作成者たちが平均値を用いていた可能性も否定できないと言わ

れた。古代の暦算者が実際にどのように計算していたかは不明である。

また学派による天文定数の相違を考慮に入れられるように、一連の定数を別に準備しておいてオプションで選択できるようにすることも考えられるが、学派が異なると暦計算のアルゴリズムも異なる場合が多いので、ひとつのプログラムですべてをカバーすることは困難であることがわかった。

インドにおけるイスラーム暦も無視できない要素なので、イスラーム天文学・暦法に関する文献も収集し、研究してきたが、これも取入れてひとつのプログラムするとなるとサイズが大きくなり、またかえって利用者に混乱を与える恐れがあるので、すでに公開されているイスラーム暦プログラムにリンクを張るなどして利用してもらう方が効果的であることがわかった。

さらに南インドでは独特的な太陽暦が用いられており、これがネパールの太陽暦とほとんど同じであるという興味深い事実が、1999 年にシンガポールで手にいれた南インドのタミル語の暦と、2000 年のネパールの暦とを比較検討することによってわかった。このようなインドの伝統的太陽暦についてさらに調査する必要があることを痛感したので、2000 年 9 月 11 日より 2 週間にわたってインドで現地調査を行った。まず独立後にインド新政府が発足したときにできた「改暦委員会」(Calendar Reform Committee) の最初からの委員である Commodore S.K. Chatterjee 先生(当時 84 歳)に面会し、貴重な参考意見を聞き、最近の著書や論文をいただいたことが、最大の収穫であった。また当時国際科学史学会の会長であったスッバラヤッパ(Subbarayappa) 博士により、マイソールで先祖代々伝統暦を作成している R. Kumar 氏を紹介していただいたおかげで、氏の自宅を訪問し、1 時間にわたってインタビューすることができた。さらにスッバラヤッパ博士が所属するバンガロールの National Institute of Advanced Studies で、私のプログラムについて講演をする機会が与えられたので、この研究所のすぐれた研究者たちから意見を求めることができた。またバンガロール在住のインド天文学史研究者であり、近著 *Indian Astronomy* の著者である Balachandra Rao 博士の訪問を受け、意見を交わすことができた。

さらに、オリッサ地方の太陽暦には他にはない特徴があるということは知っていたが、今までの調査ではこの地方の暦を入手することができなかつたので、今回はじめてブナーシュバルとブーリーに赴いて、いくつかの暦を収集することができた。とくにインド全域にわたって私がもっとも広く調査した年である 1991

年の当地の暦を手に入れることができたのは大きな収穫であった。

コンピュータの出現によって伝統的な暦計算の技術は急速に忘れられていくであろうことは、すでに1991の調査において予測できたが、上記報告書に述べたような多様性がこれからどれくらい生き続けることができるのか、これからも確認していく必要があると思った。

この調査旅行で得た知見に基づいて、プログラムのなかにインド式太陽暦をとり入れることにした。その基本は次の通りである。まず太陽が新しい宮に入る「サンクランティ」の瞬間は従来のプログラムのなかでも必要なので、これは関数として用いることができる。次に「サンクランティ」が当日の日没までにあればその日を太陽月の第1日とし、もし日没以後であれば翌日を第1日とすることにした。

なお、バンガロールは「インドのシリコンバレー」と呼ばれ、インドのみならず、世界のソフトウェア産業の中心になり、日本でもにわかに注目をあびるようになった。その重要性は、私が訪問した1週間ほど前に森首相が当地を訪れたことからもわかる。とくに最近はインド人の頭脳流出が話題になっているが、このような文化的な背景は伝統的な天文計算法にも存在すると考えられる。このような方面についても考察するならば、本研究は古典研究の一貫でありながら、すぐれて現代的な視点を提供するものもあるといえるだろう。そもそも「アルゴリズム」という現代のコンピュータ用語そのものが、「インド式計算法」をヨーロッパ世界に紹介したイスラームの天文学・數学者アル・フワリズミーの名前に由来している。このことからもわかるように、インドの天文学書に見られる「解へ至るための手続き」はまさにアルゴリズムの世界であるといえる。したがってインド天文学書の論理的な分析はイ

ンド人の思考様式を知るために重要な手がかりを与えると思われる(産経新聞2000年7月12日夕刊に筆者の談話が掲載されている)。

本プログラムではウッジャイニー(Ujjainī, 現在のUjjain)の子午線を基準としている。ウッジャイニーよりも西あるいは東に経度差がある土地の場合にはそれ対応する時間(時差)だけ日出は前後する。しかし経度差の影響は、任意の土地における日の出の瞬間のカリ・ユガの初めからの積日を、ウッジャイニーにおけるそれと近似的に等しいとみなしていることによるだけである。インド亜大陸は南北に広がっており、西の端近くのカルカッタはウッジャイニーから15度も離れていない。つまり経度差による積日数の差は1/24日以下であり、その間の月の太陽からの離角の増加はおよそ30分である。したがって、カルカッタのような土地では、朔やティティの変わり目の瞬間が日出の前後1時間以内にあるときには日付が1日ずれる可能性があるが、ウッジャイニーからの経度差があまり大きくなない土地ではそれほど影響はない。それより緯度や上に述べた「出差」の影響のほうがよほど大きい。したがって当初はこのプログラムでは経度差を考慮に入れる必要がないと判断した。じつさい、現代のシンガポールでインド人たちが用いている暦がマドラスから輸入されたものであるように、古代においても経度差を考慮しないまま遠くから手に入れた暦を用いることもあったはずだ。

しかし『第I期研究成果報告』において述べたように、土地の経度を考慮すべきではないかというユーザーの質問があったので、version3.10からは経度も入力できるように変更した。その際インドの代表的な都市をリストにし、そこから緯度と経度を選択できるようにした。現在のところ最新のバージョンは2002年3月に公開した3.11である。

# XML の骨抜き利用法\*

アジア・アフリカ言語文化研究所データベースの例

豊島 正之

## 目次

<b>1 tag は有用・構造記述は無用</b>	1
1.1 SGML は、XML になって良くなったか . . . . .	1
1.2 アジア・アフリカ言語文化研究所のデータベースでの骨抜き XML . . . . .	2
1.2.1 tag が欲しい実例 . . . . .	2
1.2.2 構造記述が邪魔な実例 . . . . .	4
<b>2 XML ドキュメントの交換可能性</b>	6
2.1 「互換性」が確保する「互換」とは何か . . . . .	6
2.2 attribute の付与する意味は、交換可能ではない . . . . .	7
2.2.1 attribute の詳細化と交換可能性は相容れない . . . . .	7
2.2.2 attribute のスコープの問題 . . . . .	7
<b>3 交換可能性の無さは XML のせいではない</b>	7
<b>4 検索に望まれる技術</b>	8
4.1 お仕着せ検索の問題 . . . . .	8
4.2 手作業の現状・望まれる技術 . . . . .	8
<b>5 引用文献</b>	9

## 1 tag は有用・構造記述は無用

### 1.1 SGML は、XML になって良くなったか

XML は、SGML を単純化したものだから、SGML の欠陥の多くを継承している。

XML が SGML を改善したのは、処理プログラムの負荷の軽減が主であって、文書作成・交換に関わる問題は、本質的に SGML の持つ問題を解決していない。

(1) 文書は、階層的な要素によって構成されるとする前提。

要素の境界によって文書の境界を兼用する「タイル貼り方式」の文書モデルで、異解釈、欄外注記 (marginaria)、誤写・虫損・欠損 (lacunae)、掛詞、渡りゼリフなど、解釈によって異なる構成を共存させる事が、困難 (SGML[CONCUR])、又は不可能 (XML)。

又、empty element は、全てこの構造の例外となるが、実際の文献翻刻にはページ付けは必須の要素であるし、他に、挿画等の非構成的要素も頻繁に現われる。

そもそも、ページ付け等は、本来は「ページ」「丁」(Octavo 等)の「折」といったブロックを成す element のマークである。しかし、これらの外形のブロックは、「section」「paragraph」といった内容のブロック element と(殆どの場合)階層構造をなさない別系列のブロックとなり、XML の前提とする「唯一一本の木構造」では表現出来ない。このため、外形ブロックの記述の替りに「ブロックの境界」という empty element を立てているだけであり、この様な empty element の利用法は、欺瞞である。

又、クロストークの様な非線条的 (serial) なデータ構造の表現には、empty element で文書中にアンカーを打ち、そこに ID attribute を振って相互参照するという <a name="ID"> <href a="ID"> 型の表現

\* 本稿は、2001 年 10 月 27 日 九州大学文学部において XML pro/con – XML で書く文献学的データというテーマで開催された「古典学の再構築」情報処理 (A03) 班主宰研究集会において発表されたものである。

「抄物の原典参照データベースの構築」に関する直接の研究成果は、B02 班の報告書に掲載することになったため、ここにはデータベース関連の技術的问题に関する成果の一部のみを掲載することとした。(情報処理班報告書編集担当による注)

が用いられるが、これは、empty element の abuse である。

(2) 意味は形式(構造)が構造内の位置(価値 valeur)によって付与するものであり、ラベルや内容の持つ意味には関わらないとする、構造主義的な形式・意味分離モデルの本質的な問題。

(3) attribute に構造が存在しない。

(4) element タグの意味交換が不可能。(後述)

(5) 文書の入れ子が出来ない。

上位文書に読み込まれるためには、独立文書として適格・well-formed であってはならない。

(6) 名前に局所性・スコープが無い。

といった問題<sup>1</sup> は、そのまま XML に継承されている。

SGML の欠陥のうち、名前の局所性は、XML は「名前空間」を持ち込んで解決しようとした。これは、文書の element/attribute にユニークな前綴を付けることで全てを固有名詞化して衝突を避けようとしたものである。通常のプログラム言語での「名前空間」は、当該ブロック内での identity を保証して情報共有を許すものの筈であるが、XML の「名前空間」は、衝突を避ける余り異化 differentiation を最優先事項とし、identity の機能は捨て去って仕舞った。(後述する様に、SGML の attribute は複数 element 間での情報共有が出来ないが、名前空間の適切な設計の下では、これを可能に出来たかも知れないのである)。この結果、XML 文書の validation は時に直観に反する結果を与え、実質的にはほぼ不可能になった。Bourret(2001) は、namespace を使って valid 且つ conforming な XML 文書を書く為には

- Use one prefix per XML namespace.
- Do not use the same prefix for more than one XML namespace.
- Use at most one default XML namespace.

The latter three guidelines guarantee that prefixes are unique.

せよとするが、折角 namespace 仕様を設定して置きながらこうした「自主規制」が必要になるとは、仕様の失敗としか思われない。namespace 仕様が最優先事項とした identifier のユニーク性すら、実は保証出来ない様である。

## 1.2 アジア・アフリカ言語文化研究所のデータベースでの骨抜き XML

アジア・アフリカ言語文化研究所(以下「AA 研」)で公開中のデータベースの内、豊島が管理している下記は、ソースは全て XML で記述してある。これは、tag の well-formedness の検証、要素の抽出(DOM)等に既存の XML ソフトウェアが便利である、というだけの理由であり、文書構造の検証等は一切意図していない。

そもそも、文書構造による構造的意味の付与自体、こうした文献学的な翻刻では殆ど無意味である。DTD 等による構造的意味の付与とは、既に理解された構造に基づく意味をパターン化し、複数の文献をそのパターンに基づいて範列化して把握しようとするものである。一方、文献学的研究とは、(正面切って言うのも恥かしいが)その文献の構造自体が研究対象の一つであって、データ化に「構造が既に理解されている」という前提を描く事自体が誤っている。

以上から、文書構造による構造的意味の付与は行わない。従って、(文書構造の検証の役にしか立たない)DTD も作成しない。文献学的な翻刻で DTD が必要になるのは、element の相互関係が余程複雑になって検証困難になるケースか、attribute が極めて多数ある場合等に限られるのではなかろうか。AA 研のデータベースの設計は単純を旨としているので、このどちらにも該当しない。

簡単に言えば、AA 研の下記のデータベースでは、XML の「構造的文書」特性を全て骨抜きにし、ソフトウェアのスケルトンが準備された便利なタグ記述言語、という側面のみを利用している。

### 1.2.1 tag が欲しい実例

#### (1) 三省堂「言語学大辞典」データベース

<http://irc.aa.tufs.ac.jp/SLEDB/>  
版元三省堂の全面的な協力を仰ぎ、科学研究費補助金を得て、AA 研情報資源利用研究センターが進めている全文検索マルチメディア・データベース。現在、主として権利関係から、ごく一部を公開中。

データベースの素材の性質上、音素、形態素等、語彙が短いものを検索したい事が多いが、単純な検索では、これは通常語の「ゴミ用例」の中に埋没して仕舞う。そこで、元データのコーディングに際して「eadem lingua (当該言語)」タグを設定し、記述対象となっている言語のみが選定出来る様にした。

<sup>1</sup> これらに就ては、出雲・豊島(1993)、豊島(1992)、豊島(1994)に詳述した。

eadem lingua (当該言語) タグで「q」を検索する指定。

SLE

mtoyo@aa.tufs.ac.jp, 2000

## SLE 「三省堂言語学大辞典」本文検索

last update, 2000/09/11

本データベースは、株式会社三省堂の許諾を得て、東京外国语大学アジア・アフリカ言語文化研究所 情報資源利用研究センターが作成・管理しているものです。各項目・記事の著作権は、それぞれの項目・記事の著者にあり、データベースの著作権は東京外国语大学アジア・アフリカ言語文化研究所 情報資源利用研究センターにあります。

このデータベース又はその検索結果の全部又は一部は、学術目的での私的使用に限って、複数が許諾されています。

一般的な文字  
general keywords [例(正規表現)  
可)]

当該言語の用  
eadem lingua [例(正規表現)  
可)]

空白で区切ると AND, !(横線) で OR, !(感嘆符) で否定。  
例 クメール 脱(脱族) 「クメール」と「脱族」又は「脱族」を含む記事  
例 接頭(接頭)|接(接頭詞) 「接頭語」「接尾語」「接辞」「接頭詞」…を検索

[100] 件まで結果を表示(0で無制限)

検索開始 やり直す

mtoyo@aa.tufs.ac.jp

<http://aa.tufs.ac.jp/SLEDB/> [2001/10/23 9:18:45]

「q」を音素目録に持つ言語は、(公開中のデータベースには)一つ

しか無かった。

SLE

search result for (SL(q))

1 例中 1 例表示

1. TEXT/I-1430L1 クア語^LCu (坂本英輔) q

結果の子音に、後に声門閉鎖音を伴う /wq, -yq/、後に気音を伴う /wh, -yh, -lh/ がある。

SLE project : sle@aa.tufs.ac.jp

## (2) 「国語学」全文データベース

<http://jcs.aa.tufs.ac.jp/SJLDB/>

国語学会データベース委員会が、科学研究費補助金を得て進めている機関誌「国語学」の全文検索データベース構築。AA 研情報資源利用研究センターが、技術的側面から支援している。現在、主として権利関係から、「展望」記事の一部を公開中。

国語学では、資料名が、そのまま研究術語名や(当

単純に「q」を検索すると、ゴミが多く、音素、形態素等の用例は埋没する。

SLE search result for (text(q))

62 例中 62 例表示

1. TEXT/I-1035L1 オーストロアジア語族 英 Austroasiatic (坂本英輔) q

Logan, James Richardson (1856), "Ethnology of the Indo-Pacific Islands", Journal of Indian Archipelago, New Series I (Singapore)Forbes, C. J. F. Smith (1878), "On the connexion of the M&#0097;s of Pegu with the Koles of Central India", Journal of the Royal Asiatic Society of Great Britain and Ireland, New Series 10 (London)Keane, Augustus Henry (1882-83), "Des rapports ethnologiques et linguistiques des races indo-chinoises et Indo-pacifiques", Annales des Expositions Orient 5 (Paris)Terrien de Lacouperie, Albert Etienne Jean Baptiste (1885-87), The languages of China before the Chinese (Trans.)

2. TEXT/I-1035L1 オーストロアジア語族 英 Austroasiatic (坂本英輔) q

N. F. (Brasselschweig) Hevesy, Wilhelm von (1930-32), "On W. Schmidt's Munda-Mon-Khme comparison (Does an 'Austro' family of languages exist?)", Bulletin of School of Oriental Studies 6 (London)Haudcourt, André (ed.): G. (1954), "De l'origine des tons en vietnamien", Journal Asiatique CCXLI (Paris)Pithnow, Heinz-J. & u00f6;gen (1959), Versuch einer historischen Lautlehre der Khmer-Sprache (Otto Harrassowitz, Wiesbaden)…音韻研究の最大成ともいべき大作。

3. TEXT/I-1117R1 オン語 Ong (坂本英輔) q

[参考文献] Ferlus, Michel (1974), "La langue ong, muta-dions consonantiques et transphonologiques dions", Asie du Sud-Est et Monde Indo-lindien V, 1 (CeDRASEMI, Paris) [参照] カトゥ語族 オーストロアジア語族(坂本英輔)

4. TEXT/I-1133L1 カオ語 Cao (坂本英輔) q

[参考文献] Ferlus, Michel (1974), "Découplage des groupes/linguistiques austroasiatiques dans le centre indo-chinois", Asie du Sud-Est et Monde Indo-lindien V, 1 (CeDRASEMI, Paris) [参照] オーストロアジア語族 カトゥ語族(坂本英輔)

5. TEXT/I-1133L2 カオ語族 Kho (坂本英輔) q

北ベトナムのソンラ(Son La), ライチャウ(Lai Chau), 北ラオスのポンサリ(Phong Salt), 上ナニタ(Haut-Nam Tha)などに分布している。サー・スクア(Xa Xuak), カー・クアン・リム(Kha Quang Lim), カー・ビット(Kha Bit), サー・カオ(Xa Ao; Kiao)などの方言からなる語群で、オーストロアジア語族のモン・クメール語族・クム語族に属する。

<http://aa.tufs.ac.jp/SLEDB/SLEDBSLE-s.cgi> [2001/10/23 9:22:24]

該分野の専門雑誌名になる事が多いため、単純な検索では、これらが混雜して、「ゴミ用例」が増える。又、「展望」(サーベイ)記事では、展望記事の著者名・論文名等と、展望(論評)対象の原著者名・原論文名等のメタな用例が混じる恐れがある。このため、「雑誌名」(当該論文の)「著者名」等のタグを施したデータコーディングを行い、タグを意識した検索が可能にしてある。

「万葉」を検索、但し雑誌「万葉」（「萬葉」）を除く。

SJS : Study In Japanese Linguistics Survey DB search

mtoyo@aa.tufts.ac.jp, 2000/11/27

## 『国語学』：「展望」記事検索

国語学会データベース委員会  
最新更新日付 2000/11/27

これは、『国語学』の「展望」のうち、著者許諾済みのものについて、ダウンロード用の記事と記事検索を公開するものです。

### 権利関係

1. この『国語学』「展望」データベースに含まれる「展望」記事の著作権は、それぞれの著者に属します。
2. それぞれの記事本文を、この「展望」データベースによって検索し、検索結果を私的に複製する事、又はその全文(画像を含む)をダウンロードして私的に複製することは、それぞれの記事の著者によって許諾されています。つまり、これらの記事は、自由に検索・私の複製して構いませんが、あくまで私の複製の範囲でご利用下さい。
3. それぞれの記事本文を、この「展望」データベースによってデータベース化し、wwwサーバーに置いて一般の読者・ダウンロードに供する事は、それぞれの記事の著者によって国語学会データベース委員会に対して許諾されています。
4. この『国語学』「展望」データベースのデータベース著作権は、国語学会データベース委員会にあります。

検索文字列:   
正規表現可:

検索例  
空白で区切ると AND, |(横棒)で OR, !(感嘆符)  
で否定

[万葉]篇	假名漢字	→	「万葉」又は「萬葉」且つ「假名」又は 「漢字」を含む記事を検索
[平上去入]声	→	「平声」「上声」「去声」「入声」を検索	
(伏韻)作用性	→	「状態性」又は「作用性」を含む記事を検索	
状態)作用性	→	「状態」又は「作用性」を含む記事を検索	
接(四)尾)?(辞尾)	→	「接頭辞」「接尾辞」「接定語」「接尾 語」「接辞」「接語」を検索	

[http://jcs.aa.tufts.ac.jp/cgi/DB/SJS/SJS-e.htm \(1/2\) \[21/10/23 10:10:16\]](http://jcs.aa.tufts.ac.jp/cgi/DB/SJS/SJS-e.htm (1/2) [21/10/23 10:10:16])  
SJS : Study In Japanese Linguistics Survey DB search

**[万葉]篇** **|集** → 「万葉」又は「萬葉」で、「集」は含まれない記事を検索

詳細  
タグ名 検索語(正規表現可、空白で区切ると AND, |(横棒)で OR, !(感嘆符)で否定  
定  
著者 <author>   
題名 <title>   
刊載誌名 <serial>   
年紀 <date>   
出版 <publisher>

例: [万葉]篇 → 雑誌「万葉」ではない記事を検索

100 件まで結果を表示 (0 で無制限)  
 検索開始  やり直し

- 最初のページへ
- 戻るへ
- 国語学会ホームページへ
- 『国語学』目次・著者索引へ

国語学会データベース委員会  
連絡先 国語学会事務室 平113-0033 東京都文京区本郷1-13-7 日吉ハイツ404号  
電子メール jimusitu@blue.ocn.ne.jp  
委員長 山口住吉

[http://jcs.aa.tufts.ac.jp/cgi/DB/SJS/SJS-e.htm \(2/2\) \[21/10/23 10:10:16\]](http://jcs.aa.tufts.ac.jp/cgi/DB/SJS/SJS-e.htm (2/2) [21/10/23 10:10:16])

### 1.2.2 構造記述が邪魔な実例

原本自体に構造無視がある場合は、構造記述をスキップした検索が要求される。

この極く単純な例は、版本の版心・頁付け(ノンブル)等である。これらは、原本の構造記述自体から除外す

### 検索結果

SJS search result for text(万葉), not\_serial([萬葉])

2001/10/23 10:11:17 generated by 'SJS-fmt.cgi', do not edit

## SJS 検索結果 `text(万葉), not\_serial([萬葉])`

全 11 例中 11 例表示

1. 097-11 文字・表記 [野村烈助] 原本圖像(PDF) 「国語学」 97(1974-06-30) 万葉

大原隆の「万葉「仮名」から仮名へ」(国語学93号、昭48・6)は、その題目から予想されるような、かな字形の底面に関するものではなく、「systemとしての万葉仮名が内包していたところの、それも母体として片仮名・平仮名が生まれるべき必然性を叙述」したものである。すなまち、ここで述べられるのは、万葉がなにに内包される。筋形表示の確実さへの要求と、表現・伝達に致命的な支障がおきない範囲での簡略化への指向という二つのペルトルのうち、筋字圖の底辺への拡大という象徴の中で、後者がしだいに強くなり、漢字本体の機能についての「価値ある忘却」がおこなわれ、しだいに「摸字隠れて」いく過程である。

2. 129-03 文法(史的研究—古代) [山口住吉] 原本圖像(PDF) 「国語学」 129(1982-06-30) 万葉

[129-18上] その他、内田賀徳「動詞直復形態の述語」(『帝國山学院大学日本文学研究』11、昭55・2)、信太知子「上代語における連体形態体法について... 万葉集を中心にク語法との関連など」(『川嶋和夫博士追憶記念国語学論集』大修館書店、昭56・7)、山口住吉「形容動詞の成立」(『国語と国文学』58・5、昭56・5)、同「タリ型形容動詞の成立」(『国語国文』50・2、昭56・12)などがあった。

3. 129-03 文法(史的研究—古代) [山口住吉] 原本圖像(PDF) 「国語学」 129(1982-06-30) 万葉

[129-18上] その他、内田賀徳「動詞直復形態の述語」(『帝國山学院大学日本文学研究』11、昭55・2)、信太知子「上代語における連体形態体法について... 万葉集を中心にク語法との関連など」(『川嶋和夫博士追憶記念国語学論集』大修館書店、昭56・7)、山口住吉「形容動詞の成立」(『国語と国文学』58・5、昭56・5)、同「タリ型形容動詞の成立」(『国語国文』50・2、昭56・12)などがあった。

4. 129-03 文法(史的研究—古代) [山口住吉] 原本圖像(PDF) 「国語学」 129(1982-06-30) 万葉

[129-18上] その他、内田賀徳「動詞直復形態の述語」(『帝國山学院大学日本文学研究』11、昭55・2)、信太知子「上代語における連体形態体法について... 万葉集を中心にク語法との関連など」(『川嶋和夫博士追憶記念国語学論集』大修館書店、昭56・7)、山口住吉「形容動詞の成立」(『国語と国文学』58・5、昭56・5)、同「タリ型形容動詞の成立」(『国語国文』50・2、昭56・12)などがあった。

[http://jcs.aa.tufts.ac.jp/cgi/DB/SJS/SJS-fmt.cgi \(1/2\) \[21/10/23 10:11:18\]](http://jcs.aa.tufts.ac.jp/cgi/DB/SJS/SJS-fmt.cgi (1/2) [21/10/23 10:11:18])

SJS search result for text(万葉), not\_serial([萬葉])

は、仮名萬葉氏が万葉集の「ある」と「をり」の用法に基づいて、前者が「存在の客観的記述」、後者が「存在の主体的描寫」を表すとしたのを承け、この区別が妙物の「ある」と「をり」についても有効であることを主張している。近藤泰弘「丁寧語のアスペクト的性格—中古語の「はべり」を中心に」(『辻村紀念』平成4)は、さまざま「はべり」の用法の中に、「あり」や「り・たり」の対応としての状態性の意味を持つものと、アスペクト的には中立で、敍述的な意味を与えたためだけに動詞等に付加されるものの二種類があることを証明した。TAKEUCHI, Lone "Long-term developments in the Japanese aspectual system-a case of 'linguistic drift'" (Acta Orientalia School of Oriental and African Studies, London, 1993)は、広範囲の現象に目配りの効いたテンス・アスペクト史の概観になっている。

11. 185-14 敷理的研究(豊島正之) 原本圖像(PDF) 「国語学」 185(1996-06-29) 万葉

[加藤95] 加藤浩司(1995)古代語形容副詞活用・シク活考... 万葉集和歌における用法上の差異に着目して.../名古屋大学国語国文学 76(九五--一七)

SJS DB project

[http://jcs.aa.tufts.ac.jp/cgi/DB/SJS/SJS-fmt.cgi \(4/4\) \[21/10/23 10:11:18\]](http://jcs.aa.tufts.ac.jp/cgi/DB/SJS/SJS-fmt.cgi (4/4) [21/10/23 10:11:18])

るのが常識的であるが、これ以外に、原本に挿絵・そのキャプション等が割込む場合を例に掲げる。

### • 古典文庫・上方狂言本 DB

<http://irc.aa.tufts.ac.jp/KotenBunko/>

AA 研情報資源利用研究センターが、古典文献の翻刻からのデータベース作成実験として、著者全てからの了承の下、実験構築中のデータベース。

KotenBunko : simple DB search  
mtoyo@aa.tufs.ac.jp, 2000

## 古典文庫「上方狂言本」「江戸板狂言本」検索見本

東京外国语大学アジア・アフリカ言語文化研究所  
情報資源利用研究センター  
最新更新日付 2000/09/05

これは、「古典文庫」の「上方狂言本」一~九、「江戸板狂言本」一~四のデータベース検索の見本として、取り敢えず、「上方狂言本 七」(土田盛)を示すものである。

### 参考

1. 「古典文庫」電子化計画検索
2. 「古典文庫」目録(MS-DOS Shift-JIS)

## 検索見本

検索文字列(正規表現可)

検索例  
空白で区切ると AND, | (並び) で OR, !(感嘆符) で否定

せんじ ひらりき → 「せんじ」と「ひらりき」両方を含む  
|一二三四五五六七八九|つ → 「一つ」、「二つ」、「...」、「九つ」  
[\*あんじ]東 → 「...東」で、「君東」や「(こしも)と東」等ではないもの  
南門 → 八文字 → 「南門」を含むが「八文字」を含まないもの

詳細指定 タグ名 検索語。正規表現可、空白で区切ると AND, | (並び) で OR, !(感嘆符) で否定

役人(役) <CastingActor>   
役名(役) <CastingRole>   
役名(役) <CastingRole>   
例: こしもと → 一役人名中の「こしもと」を除外

挿絵中の <Figure>   
字句   
欠損欄 <Lacuna>   
書籍事項 <BookItem>

http://rc.aa.tufs.ac.jp/KotenBunko/[12].201/1023 11:13:47  
KotenBunko : simple DB search

件まで結果を表示 (0で無制限)

## 権利関係

「上方狂言本 七」(古典文庫 409、土田盛 1980)の電子化・複製・複製・公衆送信・送信可能化について  
は、古典文庫 主人吉田卓一博士、著者土田盛様よりの貢献を頂戴している。

東京外国语大学アジア・アフリカ言語文化研究所情報資源利用研究センターは、学術目的での個人利用に限って、複製を許諾する。

## 本文ファイルのダウンロード

409-1.txt 古典文庫409 上方狂言本 七(土田盛) (2000/09/03更新)

- [本文ファイルのデータ形式に就て\(2000/09/05\)](#)

豊島正之 / mtoyo@aa.tufs.ac.jp

原本 3 才が「取」～4 ウ「かはした事」と続き、その間の 3 ウ・4 才に挿絵が入る。翻刻本もこの分断を踏襲している。挿絵・キャプションをまたいで「取かはし」を検索したい。

KotenBunko search result for 'text(取かはし)'  
2001/10/23 11:13:37 generated by 'KBD.s.cgi' do not edit

## KBD 検索結果 'text(取かはし)'

### 1 例表示

1. 409 上方狂言本 七(土田盛) PDF 「古典文庫」409(1980,昭和55年10月20日) 128((四ウ)) 取かはし

かくて、さだの左宿門やしきには、むこ主ぜんのからう、とね川与一左衛門は、大平七郎右衛門にむかい、「此間お心づかいにあづかり悉し、則今日主ぜん國へへれば供を申、おないとまう」と云ば、七郎右衛門聞、「口、それはさんねんに存る、波音とうりうの間に申合し路、かまいてさうい音へからず」と云ば、「何が祝たがいに、しんもんを取かはした事、何時てもけさう△△△△△△付ん」と△△△むこ君花□がき主ぜん立出給へば、

KotenBunko project : (temporary mailto mtoyo@aa.tufs.ac.jp)

```

<PAGENOTE page="126"/>
へぞ上りける.<lf/>
<lf/>
かくて、さだの左衛門やしきには、むこ主ぜんのからう、とね川与一
左衛門は、大平七郎右衛門にむかい、「此間お心づかいにあづかり忝
し、則今主ぜん國へかへれば供を申、おいとま申」と云ば、七郎右
衛門聞、「<Lacuna> □そ </Lacuna> れはざんねんに存る、初御とうりうの間に申合し義、か
まいてさうい有べからず」と云ば、「何が初、たがいに、しんもんを
取 <FOLIA fol=(三才)"/><lfPreFigure/>
```

<Figure>

挿絵 <tab/> 第一図 <origFOLIA fol=(三ウ)" /><lf/>
挿絵 <tab/> 第二図 <origFOLIA fol=(四オ)" /><lf/>

```

<PAGENOTE page="127"/>
第一図 (上段)<lf/>
『みうら介六、松へ上り』<lf/>
『やつこ八平、馬引る』<lf/>
『きつね共、さうれいする所』<lf/>
(下段)<lf/>
『こしもとおしげ』<lf/>
『井の上新三郎、かけ付る』<lf/>
『はるひめ、おそれ入給ふ』<lf/>
『かづさがいもとおきく』<lf/>
『ざとう久いち、にぐる』<lf/>
『きつね、介六 <atRight> 二 </atRight> ばけ来る』<FOLIA fol=(三ウ)" /><lf/>
<lf/>
第二図 (上段)<lf/>
『井上新三郎』<lf/>
『大殿さだの左衛門』<lf/>
『若殿いおり』<lf/>
『むこ主せん殿』<lf/>
『かづさが女房ねすが』<lf/>
『大平七郎右衛門』<lf/>
『主せんからう与一左衛門』<lf/>
(下段)<lf/>
『左衛門、りつふく』<lf/>
『七郎右衛門、うつてに行んと云』<lf/>
『新三郎、せうこに立』<lf/>
『おすぐ、おしとめる』<lf/>
『かづさいもとおきく』<lf/>
『かづさ、せんぎを聞』<lf/>
```

<PAGENOTE page="128"/>

(上段)<lf/>
『かづさの介九郎』<FOLIA fol=(四オ)" /><lf/>
</Figure>

かはした事、何時てもけさう <Lacunae> △△△△△ </Lacunae> 付ん」と <Lacunae> △△△ </Lacunae> むこ君花 <Lacuna>
□/が </Lacuna> き  
主せん立出給へば.<lf/>  
若殿いおり出、おいとまごいの給ふ.<lf/>  
所へ、大殿左衛門出給ひ、「主せんかへり給 <Lacuna> □/ふ </Lacuna> か、此度、身かきしよ  
くに付見まひと有て、はる」  
「参らるゝだん、まんぞくに存る、本ぶ  
くいたしてれば、おつ付はる姫をおくるでござらふ程に、ふびんをく  
はへて下され。」「其だんはお心やすう思召ませ。」「おううれしう存  
る、むこ殿おたちじや、かづさふうふはいづくにゐるぞ」との給ふ所  
へ.<lf/>

(古典文庫 409 土田衛編「上方狂言本七」より)

## 2 XML ドキュメントの交換可能性

### 2.1 「互換性」が確保する「互換」とは何か

SGML/XML で文書の「互換」を確保する為には、次  
の様なものが試みられてきた。

#### 1. 同一 DTD を使う

→ 拡張性皆無で、実際には実施不可能。

同一 DTD が強制出来る程に統制が取れた文書作  
成集団なら、そもそも「互換性」等を気にする必  
要は無いのではなかろうか

#### 2. 極力構造性の無い DTD にし、共通辞書（「ボキヤ ブライ」）又はタグ名対応辞書を作る

→ ebXML 等、「業界標準」作成の模索で使われ  
た方式。共通辞書が作れる位なら、文書標準位、  
軽いものだろう。

#### 3. DTD をパーティ化し、タグ名は決め打ち、部分詳 細化は parameter entity で行う。

→ 一見、拡張性がある。TEI を始め、過去の多  
くの SGML 大アプリケーションが取った方法で  
あるが、SGML/XML の文法的制約、特に名前ス  
コープの無さから、重厚長大複雑怪奇な DTD に  
なり、メンテナンスが困難なばかりか、名前の衝  
突など、ユーザには解決困難な問題を引き起  
こす。しかも、タグ名を決め打ちした処で、タグの  
表現内容に就ての合意が無ければ無意味である。  
これは TEI の破綻の最大の要因となった。

4. DTD とは別のスキーマを設け、タグの構造的意味付与を、スキーマ文法記述で行う。

→ XML スキーマの取る方法。スキーマ自体が乱立の状況で、安定して使えない。又、タグへの構造意味付与も、表現内容の合意が無ければ無意味である点は同じ。

## 2.2 attribute の付与する意味は、交換可能ではない

attribute は意味付与以外の機能を持たないから、element 名よりも、遙かに意味に直結している。(意味に直結していなければ、attribute を振る意義がそもそも無い)。

### 2.2.1 attribute の詳細化と交換可能性は相容れない

attribute は、詳細化すればする程、交換不能になる宿命を持つ。DTD の詳細化は共通 DTD 一つであるのに、実際の文書が要求する詳細化の枠組みは、文書ごとに異なるからである。

例えば、「日本電子出版協会電子出版交換フォーマット」JepaX (<http://x.jepa.or.jp/jepax/>) は、element による書籍の構造記述は殆ど放棄して <div> 一辺倒とし、その代りに <div> に attribute を振って意味付与しようとして、attribute に下記を定義している。

"表紙"、"あらすじ"、"抄録"、"著者紹介"、"とびら"、"謝辞"、"献辞"、"序文"、"凡例"、"目次"、"図版目次"、"部"、"篇"、"章"、"節"、"項1"、"項2"、"項3"、"項4"、"項5"、"項6"、"あとがき"、"解説"、"付録"、"索引"、"用語集"、"年表"、"関連書籍"、"奥付" "引用"、"詩"、"プログラムリスト"、"参考文献"、"囲み" …

「将来の仕様更新時に追加されていく可能性がある」と明記されている(確かに「項 n」はどんどん追加されそうである)が、これだけでも「序文に献辞が含まれていたら?」「参考文献と関連書籍の差は?」「項1の下に更に項目列挙があったら?」「引用がプログラムリストだったら?」「プログラムリストが詩だったら?」<sup>2</sup>等々、実際のコーディングでは様々な疑問が湧いて来る。送り手のコーダ当人に疑問が湧く様では、受け手はとても安心して受け取れない。

### 2.2.2 attribute のスコープの問題

attribute は、element をスコープに持つ。(これが、SGML では唯一の名前スコープであった)。従って、

global attribute というものが存在しない。

TEI(<http://www.tei-c.org/>) は、この解決のために、全ての attribute の後ろに entity reference で global attribute を埋め込むという挙に出たが、これは徒労である。何故なら、異なる element の attribute の意味が同じである保証が全く無いからである。attribute は element local なので、異種の element に同じ attribute を振っても、それらの属性が同一である事の表現にはならない。

```
<thesis type="manuscript">
<book type="manuscript">
```

等と振る事自体は勿論自由だが、SGML/XML の仕様からすれば、これは振られた attribute がたまたま同一の manuscript という文字列だったというだけの事で、両者にそれ以上の如何なる関係も推定出来ないから、勿論両者の同定も出来ない。實際

```
<temparature status="high">
<tide status="high">
<quality status="high">
```

を同定されては困るだろう。

この様な attribute の仕様は、同一 attribute を inherit しながら順次詳細化していく polymorphism が、element 間で行う事が出来ない事を意味する<sup>3</sup>。これは、電子文書のオブジェクト化には全く反する性質である。XML の namespace 仕様は、この様な local 仕様を element にも適用しようとするものである。

## 3 交換可能性の無さは XML のせいではない

こうした XML 文書の意味の面での互換性の無さは、そもそも XML の問題ではない。

タグによる情報交換のためにはタグの指示する概念に就ての共通理解が必要であるが、それが存在しないのが、真の問題である。

TEI 書誌記述を例に取れば、<docTitle> というタグは、TEI 仕様書は形式を <titlePart main="...", sub="...", desc[ription]="..."> と規定するだけで、title とは外題か内題か、外題にしても題簽か扉か、2 冊目のみに原題簽が残る時の他の冊の打付け書きとの区別はどうするか、角書は含むか、等の基本的な書誌事項の規定が何も無い。つまり、<title> タグ概念の詳細に合意が取れな

<sup>2</sup> 世の中には、perl で書いた時 (Larry Wall 作) というものも存在する。

<sup>3</sup> 尚、TEI-P3 は、こうしたオブジェクト指向化を element class という概念を立てて行なっているが、実はそこで行なっている inheritance は、上位の attribute 記述を手作業で(!) コピーして来る (3.7.1) といった体のものである。これも、TEI のせいというよりは、SGML の記述力不足とすべきであろう。

いのであるが、これは、TEI の不備というより、そもそも書誌学に於て言語・時代・分野を超えての「title」という共通理解が存在しないからに他ならない。(言語・時代・分野を超なければ、TEI の存在意義はそもそも無い)。つまり、情報交換規格の不備ではなく、書誌記述に対する共通理解形成の不備であり、書誌学が、実は交換可能な概念を持っていない事が、ここに露呈しているのである。

事は書誌学に限らない。HTML の <p> で頻用される「paragraph」自体、定義は困難で、W3C の HTML specification (4.1)<http://www.w3.org/TR/html4/struct/text.html#edef-P> は、「Authors traditionally divide their thoughts and arguments into sequences of paragraphs.」と tradition に責任を押し付けている。しかし、これは正しい態度であって、早く檜山(1998)の指摘がある。「段落や見出しは、… 文書の記述様式である。国語教育や読み書きの経験で習得され、社会や特定集団内で共有される規範である。この規範を写し取るには、目的の文書文化(読み書きの習慣と形式の総体)のなかに身を置き、経験するしかないだろう。」

こうした「文化的に定義される術語」という側面以外に、研究対象として見た場合、title や author などの曖昧な用語は、曖昧であるからこそ、概念のバスケットとして使えるのであり、研究を進展させるものである。(例えば title page の研究 Smith(2000) を title の厳密定義から始める訳には行かない)。「short-title catalogue」と断った目録が編まれるのは、title に一貫した特徴がないからである。交換可能な程に精密化された概念を求めていたのでは、研究の余地が無い。title や author といった概念語は、記述表現(「これが title である」)ではなく、encoding を行った者の解釈表現(「私はこれを title と呼ぼう」「title に相当するものとして、私はこれを擧げる」)であり、一つ一つの用例に則して解釈を理解するしかなく、そもそも「交換」には向いていないのである。こうした意味からも、文書の構造を出来るだけ精密に定義して構造的意味によって文献を記述するという発想は、全く文献学的研究向きではないと言えよう。

以上から、文書構造を SGML/XML 風の hierarchy で記述する事には、文献学的研究に取っては何の意味も見出せない。少なくとも、文書構造自体の「交換」が必要な局面も、有用な局面も存在しないと考える。TEI の様な規準が役に立つとしても、せいぜい(作業者間で合意を見ているタグ付け規約での)タグ名・attribute 名統一等、コーディング担当者のマニュアル・心覚え程度であろう。

実際、TEI conformant なドキュメントから、<title>

タグを拾う事で、何が主張出来るだろうか? 抽出したデータから書誌目録が編纂出来るだろうか? 実際は、当該文書のタイトルに加え、関連文書のタイトル、その文書の注釈の為に引用された書・用語の解説書のタイトルまでも <title> であるので、<title> で収集されるのが玉石混交であるだけでなく、本来拾いたいものが、<title> でマークされずに取り落されている可能性もある。

## 4 検索に望まれる技術

XML のマークアップ校正を繰り返したテキストに対して、tag を剥ぎ取った形での頒布を求められるのは、大変空しいが、しばしばある事である。これは、XML データの検索技術等、(今の処)誰も信用していないからであろう。

### 4.1 お仕着せ検索の問題

Langland, William(ed. Adams, Robert, et al (2000)) は、中世英語彩色写本を「TEI-P3 full conformant」でマークアップし、ピュア(browser)兼検索ソフトウェアを添えたものであるが、その Readme は、次の様に述べて、このソフトウェアを実質見限っている。

However, since the browser does not permit full display of every feature of the manuscript that we have recorded in SGML markup, you may find it useful for some purposes to search the \*.sgm text files with an ASCII editor rather than search with the browser.

...

Readers should, therefore, expect the search facility on the browser sometimes to be less thorough and useful than that on a word processor or text editor operating on the plain ASCII text.

TEI-P3 full conformant なのはマークアップであって、ソフトウェアではないのである。TEI-P3 を活かした汎用の検索ソフトウェアが如何に難事かを伺う事が出来よう。

### 4.2 手作業の現状・望まれる技術

#### (1) well-formedness の検証・エディタ

XML エディタ数種を試したが、TagEditor(アンテナハウス <http://www.antenna.co.jp/>) に落ち着いた。(因みに、試した中で最も安価であった)。

a) well-formed でない文書、作成中の文書を読ませても、最後まで読む。

呆れた事に、世の中には「well-formed ではない」とエラーメッセージを出して読み込みを中止する XML エディタも、売り物になっている。これでは、既存の文書を読み込んで XML 化する等は、不可能である。ゼロから XML 文書を手で書く以外には使えないエディタというデザインのセンスは、XML の利用局面を全く理解していないと思う。

b) 文書の encoding 指定を無視して読み込める。

文書が自分自身の encoding を指定するのは、クレータのパラドクスであって無意味である<sup>4</sup>。これは、特に XML 文書が SMTP 等で転送された時に然りであるが、こうした「自分自身を裏切る」指定をした文書を読ませると、「encoding が違う」と言って止まる御粗末なエディタすら売り物になっている。

c) 文書を排他的に開かないで、他からのアクセスも許す。

XML 文書の宿命として一旦全文を slurp せざるを得ないが、排他的に開き続ける事がないので、別のソフトウェアで更新を掛ける事も出来る。他からの更新が掛かると、TagEditor はそれに気付き、読み直すか否かを問い合わせて来る。slurp 後も開き続け、他からの読み込みすら sharing violation を引き起こすエディタもある。

## (2) 検索

上述の AA 研のデータベースでは、perl 経由の DOM (Level 1, <http://www.w3.org/DOM/>) で各階層のオブジェクトを取り出し、それを階層ごとに結

合したものを perl の DBI (Database interface) 経由で SQL 系 DBMS (mySQL) のテーブルに格納し、これを同じ DBMS で検索・提供している。これだと一旦「下準備」が済めば高速であるが、タグの種類が多くなって来ると「下準備」に手間が掛るし、DOM の利用も実質的には各データベース毎に補助的なルーチンを補わねばならず、手作業に近い。ここを一般化しようとすると、SQL のテーブル定義から可変にしなくてはならず、これは perl の DBI・java の JDBC 経由等で可能であるが、余りに大事になるので諦めた。尚、XSLT は、構文が美しくない事<sup>5</sup>、動作に今だ不審な点が残る事等から、利用していない。

オブジェクトの階層を意識した検索が、もう少し簡単に出来ると有り難い。

期待する技術としては、on-the-fly で DOM を意識したフィールドを取り出して検索出来る事が第一である。これ自体は perl や java で可能だが、その場合、インデクシングや検索自体まで perl や java で書く破目になる。これは避けたいので、効率の良い格納・検索等の処理は、背後に DBMS が位置して処理して欲しい。現在、こうした関係を綺麗に解決して呉れるプロダクトを探している処である。例えば、SQL 系で、DOM による view を可能にするか、select 句に DOM 呼び出しを書いたかの様に振る舞うプロシージャがあれば、かなり楽な筈であるが、実例には、未だお目に掛かれない。

XML native Database には大変興味があるが、今の処、とても文系の一研究所が試しに購入出来る価格ではないのは残念である。

## 5 引用文献

1. Bourret,Ronald(2000) XML Namespaces FAQ (<http://www.rpbourret.com/xml/NamespacesFAQ.htm>)
2. Langland, William(ed. Adams, Robert, et al (2000)) Corpus Christi College, Oxford MS 201 (The Piers Plowman Electronic Archive, Vol.1, The University of Michigan Press, [CD-ROM], ISBN 0-472-00275-9)
3. Smith, Margaret(2000) The title page – its early development 1460-1510 (Oak Knoll press)
4. 出雲朝子・豊島正之 (1993) 『玉塵抄』と計算機 II (文部省科学研究費研究補助金「『玉塵抄』の計算機処理に就ての発展的研究」研究成果報告書)
5. 豊島正之 (1992) TEI から見た SGML のはなし (情報処理語学文学研究会会報 12, <http://jcs.aa.tufts.ac.jp/mtoyo/TEI/JALLC-12-TEI.pdf>)

<sup>4</sup>豊島 (1994) 参照。

<sup>5</sup>その必然性も無いのに、無理やり XML で記述する事にしてあるからであろう。

6. 豊島正之 (1994) TEI-P3 から見た SGML のはなし (情報処理語学文学研究会会報 15、  
<http://jcs.aa.tufts.ac.jp/mtoyo/TEI/JALLC-TEIP3.pdf>)

7. 榎山正幸 (1998) 文書型定義 (DTD) とその設計  
([http://www.saiensu.co.jp/ct\\_Fresource/199809/On-DTD.htm](http://www.saiensu.co.jp/ct_Fresource/199809/On-DTD.htm))

# Web を用いたコラボレーションに関する試み\*

永崎 研宣

## 0. はじめに

インターネットの登場は、これまでコンピュータとそれほど関連を持たずにいられた人文科学の領域においても大きな変革をもたらしつつある。インターネットによって提供されるサービスである World Wide Web (以下、Web)，電子メール，メーリングリスト等は、人文科学分野においても研究情報の流通や共有において画期的な手段として用いられるようになっていている。<sup>(1)</sup>

なかでも、当初から様々なコンピュータ環境を前提としていた Web は、Windows や Macintosh, UNIX 等、コンピュータ環境が必ずしも統一されているとは限らない研究者同士での共同作業を行うには適しているメディアであるといえる。また、Web は、情報提供者が一方的に情報を垂れ流すだけの単なる一方向のメディアというだけでなく、CGI (Common Gateway Interface) に代表される双方向の情報流通の仕組みが用意されており、それを生かした多種多様なサービスが日常生活にまで浸透してきている。こうした双方向の情報流通システムは、比較的容易に構築できるようになっており、そのことが Web をはじめとするインターネットを普及させた強力な要因となっていることは言うまでもない。そこには、技術面での容易さ、ハードウェア調達の容易さ、オープンソースソフトウェアの本格化、といった背景がある。

まず、技術面では、システムの構築にあたって必要

なプログラミング言語として、Perl や PHP といった、比較的習得の容易な高級スクリプティング言語が利用できるようになっている、また、それらを習得・利用するためのドキュメントが書籍や Web サイトといった形で広く流通してきていることから、専門家でなくともある程度のシステムの構築が可能となってきている。

ハードウェアに関しては、従来、かなり高額であったサーバ機器が安価に入手できるようになり、一方で、一般に利用されるようなパーソナルコンピュータの信頼性が高まったためにそうした機器をサーバ用途に使う場合も出てきている。したがって、サーバ機器を用意する必要があるような Web サイトであっても、ハードウェアの調達に関しては以前と比べてかなり容易になってきている。

他方で、フリーの OS である Linux やフリーの Web サーバソフトである Apache に代表されるように、いわゆるフリーソフトウェア、あるいはオープンソースソフトウェアが、高機能と高安定性を実現できるようになってきており、かなり本格的に利用できるものとなってきたという点も見逃せない。こうした状況により、人文科学分野においても、双方向性を生かした様々なコンテンツが登場してきている。

本稿では、上記のような Web の特性を利用した、インド仏教文献の電子テキスト校訂を Web 上で共同作業として行うための試みについて報告したい。

(1) 人文科学分野におけるコンピュータの応用については、インターネットが普及する以前にも「人文学と情報処理」や情報処理学会人文科学とコンピュータ研究会において様々な報告がなされており、また、人文科学分野におけるコンピュータ教育についてもごく基本的なものは行われてきている。しかし、その状況はインターネットの普及によって一変する。これについては、永崎 [1] において詳述しているので興味のある方は参照されたい。

## 1. 電子テキスト校訂システムの考え方

校訂テキストは、エディターの手により書籍の形で出版されるのが通例であった。しかしながら一方、コンピュータの普及により、CD-ROMによる電子テキストの形で出版されるものもあらわれてきていている。さらに、インターネットの普及により、直接インターネット上で電子テキストとして配布される場合も出てきている。

インターネットによる配布の場合には、書籍やCD-ROMの形で出版されたテキストとは異なり、配布しているテキストを更新することが極めて容易である。<sup>(2)</sup>そして、CGI等を利用してすることで、校訂の作業そのものを、世界中のどこからでも、共同で行うことが可能となる。また、インド仏教文献の場合、通常のコンピュータ環境では表示できない文字があり、常にその点が問題となってきたが、Webにおいては、多言語情報の共有に関する様々な試みがなされており、こうした問題についてもある程度の解決が可能である。

したがって、ここでは、こうした特性を利用して構築した、Web上でインド仏教テキストの校訂作業を共同で行うためのシステムについて報告したい。

まず、一般的に、テキスト校訂を共同作業として行う場合には、以下のような問題が考えられる。<sup>(3)</sup>

- ・最新版の確定の困難さ
- ・作業の重複の可能性
- ・相互の意思の疎通の手間
- ・中心になる人への負担

すなわち、作業状況を相互に把握するのが難しいということは問題の少なからぬ部分を占めているということである。しかしながら、こうした問題点の多くは、Webの双方向性・即時性を利用して解消できる可能性がある。Web上に置かれたテキストを共同で校訂していく場合、作業記録を残すようにできれば、誰がどういう作業をやっているか余所からで

も瞬時に把握でき、作業の結果としての最新版のテキストが常にどこかに公開されるようにしておけば、最新版を共有することも容易であり、また、こうした作業は、インターネットに接続されたコンピュータさえあれば、いつでもどこからでも可能となる。

一方で、Webを利用した場合でも、いくつかの問題点がある。利用者のコンピュータ環境が皆完全に同じとは限らず、特に、特殊文字への対応等については、多様な環境にあわせられるようにするか、利用者の環境を統一してしまうか、という選択を迫られることになる。また、Webは基本的にはどこからでも誰からでもアクセスできるものである。そのことは、最新版のテキストを共有するには適した特性だが、一方で、テキストの校訂という局面では、そのままにしておくと誰でもテキストの書き換えをできてしまうということになってしまう。それでは大きな混乱を招くことになりかねない。したがって、テキストの書き換えができるのはあらかじめ決められた校訂作業者のみにしておく必要があろう。さらに、校訂作業者間の見解の相違をどう扱うかという問題は依然として残ることになる。

このようなことから、今回、Web上で構築するテキスト校訂システムは、以下のよう仕様を目指すこととした。

- ・修正結果はすべて残る
- ・修正者、修正日付の記録が残る
- ・修正個所の修正履歴を一覧できる
- ・最新版は常にテキストとしてダウンロード可能
- ・修正者は修正理由をつけられる
- ・データは行ごとに管理
- ・コンピュータ環境に左右されない多言語環境

すなわち、どの個所に、誰が、いつ、どんな理由で修正を加えたのかを一通り記録し、一覧できるようにしておくことで、いつでも、どこからでも、校訂作業者はこれまでの作業の文脈を踏まえつつ作業の継続ができるなどを目指した。

(2) 更新しやすいということは、依拠するテキストがどの時点のものであるかを常に意識しなければならないということである。

(3) これ以外に、テキスト校訂を共同作業で行おうとする場合のそもそもの問題として、複数の校訂作業者間でそれぞれに異なる見解があり得るにも関わらず、共同作業である以上、テキストのすべてについて、暫定的にであれ、一致した見解を出さなければならぬという点がある。この問題については、基本的には、最終的な決定者を決めて委ねるということになると思われるが、それでも見解の相違が解消できない場合に、システム上、こうした見解の相違をうまく反映できるような仕組みを検討しているところである。

## 2. 電子テキスト校訂システムの構成

システムの構成としては、ハードウェアは、Pentium III 450MHz, DRAM 640MB, HDD 40GB の PC/AT 互換機を用いている。OS は、安定性を重視し、Vine Linux2.1CRを中心とした構成となっている。

テキストデータ<sup>(4)</sup>は、フリーのリレーショナルデータベースである PostgreSQL のバージョン7.1に、テキスト1行を1レコードとして入力している。<sup>(5)</sup>データベースを用いることで、データのより高い安全性の確保と効率のよい検索性が可能となる。このデータベース上のテキストデータは、Web サーバからの入出力ができるように設定しており、入出力に際してのデータベースサーバとの通信は、フリーの Web サーパソフトである Apache 1.3、および動的 Web サイト構築用に設計されたフリーのスクリプティング言語である PHP 4を利用して行っている。

校訂作業者を限定するため、HTTP の Basic 認証を利用したアクセス制限をかけている。校訂作業者ごとに ID とパスワードを発行し、テキストを修正する際には、この認証用の ID を記録することで修正者の記録を行うようにしている。

## 3. 電子テキスト校訂システムの実際

このシステムを校訂作業者が利用する場合には、まず、トップページにアクセスすることになる。トップページへのアクセスの際、ID とパスワードを入力することになる。(図 1)

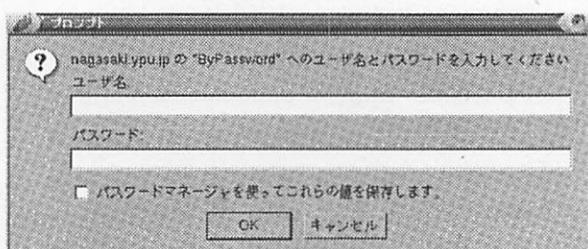


図 1

認証を通ると、トップページが表示される。(図 2)

図 2

トップページでは、テキスト全体がダウンロードできるリンクが用意されており、それ以外に、編集用テキストとして 3 種類の表示機能が用意されている。3 種類の表示機能は、それぞれ、表示させたいページ数を入力して「実行」ボタンをクリックすると、テキスト修正モードで指定されたページのテキストが表示されるようになっている。

(4) このテキストデータは、筑波大学小野基先生がインターネット上の Web サイト小野基 [4] 「Dharmakīrti E-text」<<http://www.logos.tsukuba.ac.jp/~nagasaki/dharmakirti/e-text.html>>にて公開されている「Pramāṇavārttikasvavṛtti」を利用させていただいていることを感謝とともに記しておく。

(5) この「1行」というのは、ここで用いている電子テキストが依拠している刊本校訂テキスト (Pramāṇavārttikasvavṛtti: R. GNOLI, The Pramāṇavārttikam of Dharmakīrti, the first chapter with the autocommentary. Roma 1960.) の行に基づいている。

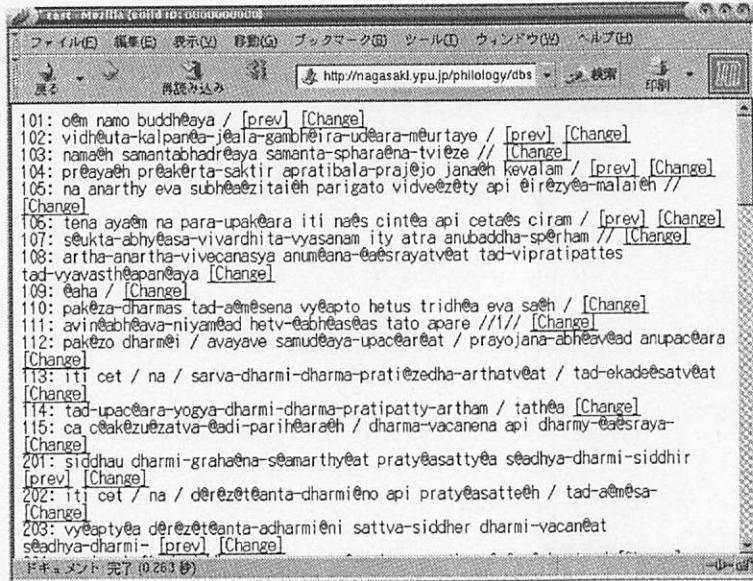


図 3

一番上のものは ASCII 文字だけを用いたテキスト表示であり、動作は軽快である。ASCII 文字の転写方式は、利用させていただいた電子テキストの方式をそのまま採用している。(図 3)

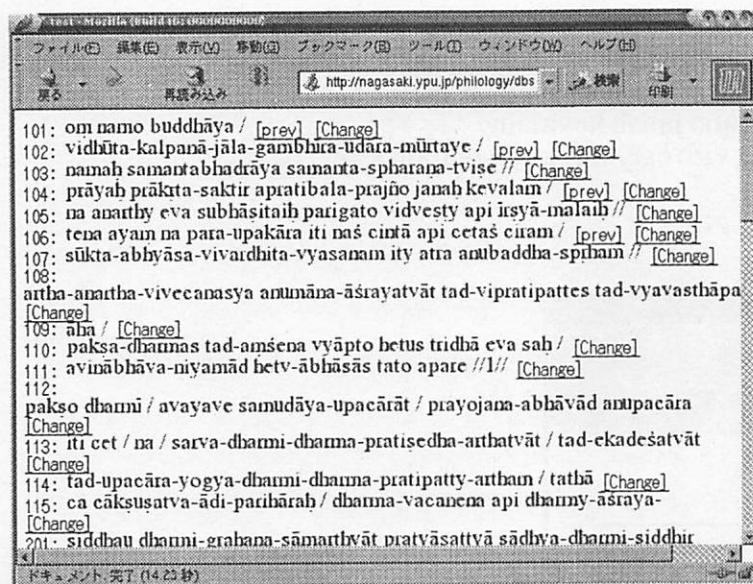


図 4

二番目のは、テキストをダイヤクリティカルマークつきのローマ字転写で表示させるようになっている。ただし、この種の文字はすべてのコンピュータ環境で表示できるとは限らないため、「文字焼き」というシステムを用いて文字を一度画像に変換してから表示させている。<sup>(6)</sup> (図 4)

(6) この文字画像表示システム「文字焼き」は、ASCII 転写文字等をその言語のフォントを用いて画像として表示させるシステムである。詳しくは永崎 [2]、永崎 [3] を参照。なお、ここでこの文字を表示させるためのフォントについては、山口県立大学鈴木隆泰先生が公開している TS font を利用させていただいている。

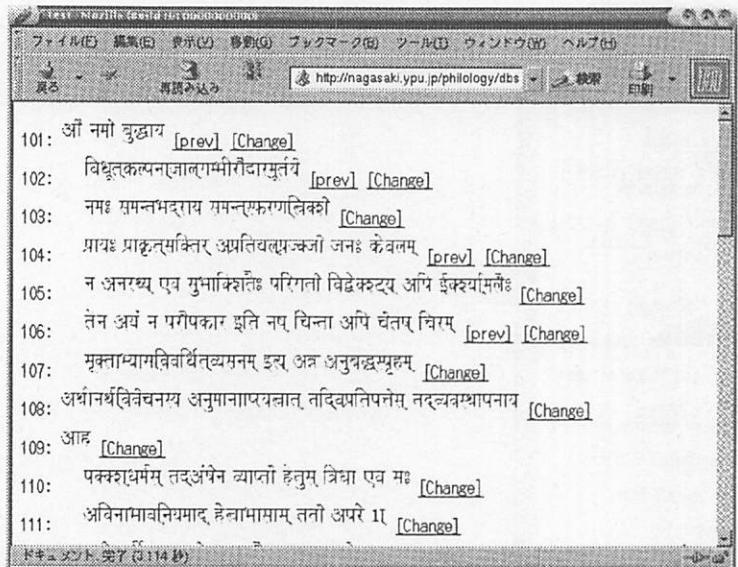


図 5

三番目の中のものは、テキストをデーヴァナーガリーで表示させるようになっている。デーヴァナーガリーについても、「文字焼き」を用いて文字を一度画像に変換してから表示させている。(図 5)

101: om namo buddhāya / [prev] [Change]  
 102: vidhūta-kalpanā-jāla-gambhīra-udāra-mūrtaye / [prev] [Change]  
 103: namah samantabhadrāya samanta-spharana-tvise // [Change]  
 104: prāyah prākṛita-saktir apratibala-prajño janah kevalam / [prev] [Change]  
 105: na anarthy eva subhāsitaib parigato vidvesty api iṣyā-malaib // [Change]

図 6

テキストを変更する際には、「change」のところをクリックする。(図 6) そうすると、

図 7

テキスト修正画面が表示される。(図7) このテキスト修正画面は、テキストそのものを変更する以外に、修正者の名前、理由、優先順位が入力できるようになっている。<sup>(7)</sup>

101: om namo buddhāya / [prev] [Change]  
102: vidhūta-kalpanā-jāla-gambhīra-udāra-mūrtaye / [prev] [Change]  
103: namah samantabhadrāya samanta-spharana-tvis / [Change]  
104: prāyah prākṛita-saktir apratibala-prajño janah kevalam / [prev] [Change]  
105: na anarthy eva subhāṣitaiḥ parigato vidvesty api īrsyā-malaiḥ // [Change]

図8

修正結果は即座にテキストに反映されるようになっている。そして、修正前のテキストは図8に示している「prev」をクリックするともう一つのウインドウが開いて確認できるようになっている。(図9)

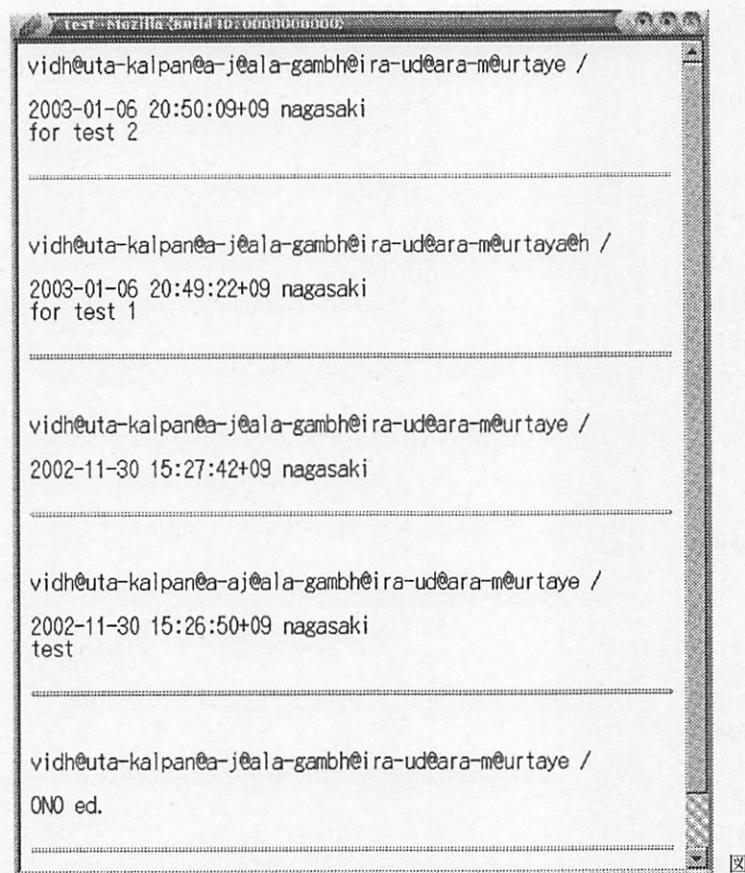


図9

修正前のテキストは、それぞれ、テキスト、日付、修正者、修正理由が表示されるようになっている。

(7) 修正日付は自動的に入力されるようになっている

#### 4. 終わりに

今後、このシステムは本格的な運用に入ることになる。それと並行しつつ、他にも様々なテキストをこのシステムに組み込んでいく予定である。また、校訂作業者間の見解の相違をうまく反映できるようなシステムも検討している。そうしたことについては、また折りをみて報告していきたい。

参考文献等：

永崎 [1]：「文科系学生に何を学んでもらうのか～「情報発信のためのリテラシー」の可能性」『平成14年度情報処理教育研究集会講演論文集』(2002年10月), pp. 389-392.

永崎 [2]：「インド系文字による Web 環境での情報の共有」『人文科学とコンピュータシンポジウム論文集 Vol. 2001, No. 18』(社)情報処理学会 (2001年12月), pp. 213-220.

永崎 [3]：「Web パブリッシングにおけるアジア諸言語の文字情報 —アジア諸言語語学教育 CAI のための文字表示システム開発を通じて」『科研費報告書 アジアの文字と出版・印刷文化及びその歴史に関する調査・研究 —デジタル化移行の基礎として—』(2001年4月), pp. 97-109.

小野基 [4]：Dharmak@irti E-text  
<<http://www.logos.tsukuba.ac.jp/~nagasaki/dharmakirti/e-text.html>>

(\*) 本稿は、2002年11月30日に山口県立大学において行われた、「古典学における情報技術の利用」研究会(情報処理調整班主宰)において発表されたものである。

研究協力者でありながら草稿を寄せられた永崎研宣氏(山口県立大学国際文化学部講師)に感謝の意を表したい。(情報処理班報告書編集担当による注)

---

---

## **発表成果一覧**

---

## 古典文献の計量的分析

研究代表者 村上 征勝  
統計数理研究所領域統計研究系 教授

分担者 古瀬 順一  
宮崎大学教育文化学部 教授

### 【発表成果一覧】

#### [著書]

1. 村上征勝, 「文化を計る—文化計量学序説ー」, 朝倉書店, 2002

#### [編集]

1. 浅野晃, 篠原進, 富士昭雄, 村上征勝, 谷脇理史, 小川武彦, 西島孜哉, 「新編西鶴全集第一巻」, 勉誠出版, 2000
2. 浅野晃, 富士昭雄, 谷脇理史, 西島孜哉, 村上征勝, 岡本勝, 花田富二夫, 江本裕, 染谷智幸, 井上和人, 「新編西鶴全集第二巻」, 勉誠出版, 2002
3. 浅野晃, 富士昭雄, 竹野静雄, 村上征勝, 大久保順子, 井上敏幸, 広島進, 杉本好伸, 中嶋隆, 「新編西鶴全集第三巻」, 勉誠出版, 2003 (予定)

#### [CD-ROM]

1. 源氏物語本文研究データベース, 勉誠出版, 2000

#### [論文]

1. 村上征勝, 今西祐一郎 (1999) 「源氏物語の助動詞の計量分析」, 情報処理学会論文誌, Vol. 40, No. 3, 774-782

#### [研究報告, 解説記事など]

1. 村上征勝 (1999) 「『源氏物語』の計量分析—「字治十帖」と他の巻との文体の違いについてー」, 人文学と情報処理 No. 20, 81-86
2. 村上征勝, 今西祐一郎 (1999) 「『源氏物語』の計量分析」, 人文科学における数量的分析 (4), 61-66
3. 村上征勝 (2000) 「著者を推理する」, ESTRELA, No. 70, 81-84

4. 村上征勝 (2000) 「古典文献の新たな研究法」, 古典学の再構築, No. 5, 44-45
5. 村上征勝 (2000) 「源氏物語を数字で読む」, ESTRELA, No. 71, 76-79
6. 村上征勝 (2000) 「文章の計量分析—その歴史と現状ー」, 計測と制御
7. 村上征勝 (2000) 「古典文献の計量分析」, ESTRELA, No. 72
8. 村上征勝 (2000) 「源氏物語 科学で読み解く」, 日本経済新聞 文化欄
9. 村上征勝 (2000) 「源氏物語本文の統計分析」, データマイニング・シンポジウム論文集17-26
10. 村上征勝 (2001) 「文章の計量分析—源氏物語を中心ー」, 統計数理とデータマイニング, 発見科学2, 統計数理研究所共同研究リポート142, 25-36
11. 村上征勝 (2002) 「著者を探る古文書の計量分析」, 電子情報通信学会誌, Vol. 85, No 3, 158-161
12. 村上征勝 (2002) 「文化を計る—『源氏物語』, 古代寺院, 浮世絵の計量分析ー」, 情報科学, 第22号, 1-10
13. 村上征勝 (2002) 「文章を計る—新たな文献研究法ー」, 人工知能学会誌, vol. 17, No. 3, 331-334
14. Murakami, M. (2002) 「Classical Texts, Real or Fake?」, Pacific Neighborhood Consortium Abstracts, A-1-A-4

#### [学会発表・講演など]

1. 「サンスクリット大乗仏典の計量分析の方法と問題点」, 1999年9月, 日本宗教学会学術大会
2. 「文章を科学する—総合的な学習への一つの提案ー」, 1999年11月, 総合学習学会第1回大会
3. 「統計学の応用領域の開拓と統計の普及—人文学の領域における試みー」, 1999年12月, シンポジウム「21世紀へ向けての統計教育」
4. 「データで読み解く源氏物語」, 2000年3月, 生命ソフトラボラトリーリsearch会
5. 「コンピュータで探る名作の謎—源氏物語を中心ー」, 2000年6月, 生命ソフトラボラトリーリsearch会
6. 「若紫やさぶらう—文章の統計分析で『源氏物語』の著者を探るー」, 2000年9月, 国際日本文化センター共同研究会
7. 「源氏物語を科学する」, 2000年9月, 東京電力館科学ゼミナール

8. 「サンスクリット法華経の計量分析の展望」, 2000年9月, 日本印度学仏教学会第51回学術大会
9. 「源氏物語における地の文と会話文の比較分析」, 2000年10月, 日本行動計量学会第28回大会
10. 「『源氏物語』本文の統計分析」, 2000年10月, データマイニング研究会
11. 「文章の計量分析—『源氏物語』を中心に—」, 2000年11月, 研究集会「統計数理, データマイニングと発見科学」
12. 「“ことば”新研究—統計分析への誘い—」, 2001年7月, 人工知能学会ことば工学研究会
13. 「日本語文献の計量的研究」, 2001年10月, NII 定例研究会
14. 「文化を計る—『源氏物語』, 浮世絵, 考古学データの計量分析—」, 2002年3月, 札幌学院大学講演会
15. 「文学を科学する—『源氏物語』の数量分析を中心にして」, 2002年5月, 日本音響学会特別講演
16. 「『源氏物語』の文頭文末の計量分析」, 2002年9月, 日本行動計量学会第30回大会
17. 「コンピューターで探る名作の謎—シェイクスピア作品, 「静かなドン」, 「源氏物語」を中心に」, 2002年9月, 2002年度統計関連学会連合大会市民講演会
18. "Classical Texts, Real or Fake?", 2002年9月, Pacific Neighborhood Consortium.

**[新聞記事, 放送, 雑誌などの紹介]**

1. 「データで読み解く源氏物語」, 読売新聞, 1999年7月8日
2. 「データベースポテンシャル」, CX・PAL, Vol. 43, 3 - 5, 2000年1月
3. 「源氏物語 科学で読み解く」, 日本経済新聞, 2000年6月28日
4. 「統計手法で文章のクセを読む」, Glaxo Wellcome New Science, No. 38, 2000年7月
5. 「源氏1000歳の輝き」, 東京新聞, 2000年10月26日
6. 「デジタル国文学」, 日本経済新聞, 2001年9月8日

**2 A03班・計画研究**

---

## 古典学のための情報処理

研究代表者 安永 尚志

国文学研究資料館研究情報部 教授

分担者 及川 昭文

総合研究大学院大学教育研究情報資料センター 教授

分担者 武井 協三

国文学研究資料館研究情報部 教授

分担者 松村 雄二

国文学研究資料館研究情報部 教授

分担者 石塚 英弘

国文学研究資料館研究情報部 教授

分担者 柴山 守

大阪市立大学学術情報総合センター 教授

分担者 アンドル・アーマー

慶應大学文学部 教授

分担者 中村 康夫

国文学研究資料館研究情報部 助教授

分担者 原 正一郎

国文学研究資料館研究情報部 助教授

分担者 山田 哲好

国文学研究資料館史料館 助教授

分担者 山田 奨治

国際日本文化研究センター研究部 助教授

**【発表成果一覧】**

安永 尚志

- (1) S. Hara and H. Yasunaga: Academic Digital library and Contents in Japan, J. of ALLC, Vol.14, No.1, pp.131-145, 1999
- (2) 安永尚志: 日本古典文学本文データベース (実験版) の試験公開, 学術月報, pp. 51 - 54, Nov. 2000
- (3) H. Yasunaga: Il Giappone in Italia alle soglie del XXI secolo: bilanci e prospettive di studio (21世紀のイタリアにおける日本学研究への課題), AISTUGIA: Associazione Italiana per gli Studi Giapponesi (イタリア日本学会) 特別講演, AIS-

- TUGIA-14, 2000
- (4) 安永尚志：日本文学研究とコンピュータ，文学，2000年7.8月号，pp.85-93, 2000
- (5) 安永尚志：情報国文学について考える，情報知識学会第8回研究報告会論文集，pp.33-44, 2000
- (6) 安永尚志：国文学研究資料館の国際コラボレーション計画 (Il Progetto di Collaborazione Internazionale del Kokubungaku Kenkyu Shiryokan), AISTUGIA-15, pp.11-14, 2001
- (7) S. Hara and H. Yasunaga: Resource Sharing System for Humanity Researches, LREC(3rd International Conference on Language Resources and Evaluation), LREC2002, pp.51-58, 2002
- (8) 山西史子，安永尚志：データベースは研究に影響を与えるか—日本古典文学本文データベース利用者調査—，情報知識学会第10回研究報告会講演論文集，pp. 1 - 4, 2002
- (9) 安永尚志：日本文学研究情報組織化のための国際コラボレーション計画，情報知識学会第10回研究報告会講演論文集，pp. 57 - 60, 2002
- (10) H. Yasunaga: International Collaboration Research for Japanese Literature, 2002 PNC Conference (Pacific Neighborhood Consortium) Abstracts, pp.18-23, 2002
- (11) H. Yasunaga: A feasibility study on the International sharing and standardization of information resources for Japanese literature research and education (日本文学教育研究資料情報の国際共有と標準化), AIDLG(Associazione Italiana Didattica Lingua Giapponese), II-2002 (印刷中), 2002
- 原 正一郎
- (1) S. Hara: OCR for CJK Classical Text, Proc. PNC/ECAI 2000, 2000
- (2) 原正一郎，安永尚志：国文学電子資料館システム—マルチメディアデータベースへのSGMLの適用—，国文学研究資料館紀要，No. 26, pp. 25-54, 2000
- (3) 原正一郎，安永尚志：国文学と電子資料館，情報処理学会研究会報告，2000-CH-46/2000-MUS-35, Vol. 2000, No. 49, pp. 85-88, 2000
- (4) 原正一郎，安永尚志：メタデータを利用したデータベースの統合，人文科学とコンピュータシンポジウム論文集, vol. 2000, pp. 97-102, 2000
- (5) 原正一郎，安永尚志：メタデータによるマルチメディアデータ統合の試み，情報処理学会研究報告，2001-CH-51, vol. 2001, No. 67, pp. 47-54, 2001
- (6) 原正一郎，安永尚志：国文学研究支援のためのデータベース統合の試み，人文科学とコンピュータシンポジウム論文集, vol. 2001, No. 18, pp. 125-132, 2001
- (7) 原正一郎，安永尚志：文学研究のためのデータベースシステムの諸問題，日本語学, vol. 20, pp. 48-60, 2001
- (8) 原正一郎，安永尚志：国文学研究支援のためのSGML/XML データシステム—国文学データ共有のための標準化—情報知識学会論文誌, Vol. 11 No. 4, pp. 17-34, 2002
- (9) S. Hara, H. Yasunaga: Resource Sharing System for Humanity Researches, LREC(3rd International Conference on Language Resources and Evaluation), 3. LREC2002, pp.51-58, 2002

#### 及川 昭文

- (1) 及川昭文：貝塚データベース—インターネットによる公開とコラボレーション—，日本情報考古学会第10回大会, pp. 13-18, 2000
- (2) 及川昭文，山元啓史：Web 公開のためのデータベース・エンジニアリング，情報処理学会研究報告, CH-49, pp. 49-56, 2001
- (3) 及川昭文：シミュレーションによる遺跡分布の推定，電子情報通信学会誌, 85-3, pp. 181-185, 2002
- (4) Akifumi Oikawa: Copyright in Japan and Distribution of the CASTEL/J Database, Kazuko Nakajima Ed. 'Learning Japanese in the Network Society', University of Calgary Press, pp.173-189, 2002
- (5) 坂谷内勝，山元啓史，及川昭文：CASTEL/J CD-ROM ミレニアムバージョンの開発，水谷修，李德奉「総合的日本語教育に向けて」, pp. 602-609, 2002
- (6) 山元啓史，及川昭文：古今和歌集データベースの開発と和歌の数理解析，情報処理学会研究報告, CH-54, pp. 29-36, 2002
- (7) 山元啓史，及川昭文：古今和歌集パラレルデータベースと公開システム，人文科学とコンピュータシンポジウム2002, pp. 265-272, 2002

- 石塚 英弘
- (1) 宇陀則彦, 江草由佳, 高久雅生, 石塚英弘: Z39.50による日本語書誌データ検索システム, 情報知識学会誌, Vol. 9, No. 2, pp. 1 – 15, 1999
  - (2) 伊東千夏, 宇陀則彦, 石塚英弘, 藤原譲: 意味関係抽出手法統合による概念の体系化, 情報知識学会誌, Vol. 9, No. 4, pp. 38 – 48, 2000
- 松村 雄二
- (1) 松村雄二: とばすがたりのなかの中世—ある尼僧の自叙伝, 臨川書店, 1999
  - (2) 松村雄二: 源氏物語歌と源氏取り, 源氏物語研究集成14「源氏物語享受史」, 風間書房, 2000
- 武井 協三
- (1) 武井協三: 若衆歌舞伎・野郎歌舞伎の研究, 八木書店, 2000
  - (2) 武井協三: 江戸芝居町図屏風(ボストン美術館蔵)一景観年代について—『歌舞伎・淨瑠璃稀本集成』下巻, 八木書店, pp. 519 – 531, 2002
- 中村 康夫
- (1) 中村康夫: 実践女子大学文芸資料研究, 実践女子大学文芸資料研究所電子叢書, 1999
  - (2) 中村康夫: 古典研究のためのデータベース, 臨川書店, 2000
- 柴山 守
- (1) 柴山 守: 古文書の文字切出しを考える, 人文学と情報処理, Vol. 18, pp. 57 – 63, 1998
  - (2) 柴山 守: 大坂町触のSGML/XML化と全文検索, 京都大学大型計算機センター第62回研究セミナー報告, pp. 3 – 15, 1999
  - (3) 柴山 守: ビデオ動画像による古文書画像入力システムの実現, 画像ラボ, Vol. 10, No. 10, pp. 36 – 40, 1999
  - (4) 柴山 守, 山田寛治, 加藤 寧, 原正一郎ほか5名: 古文書翻刻支援システム開発プロジェクト報告(1) 情報処理学会研究報告2000–CH–45, Vol. 2000, No. 5, pp. 1 – 8, 2000
  - (5) 和泉勇治, 加藤 寧, 根本義章, 山田寛治, 柴山守, 川口 洋: ニューラルネットワークを用いた古文書個別文字認識に関する一検討, 情報処理学会研究報告2000–CH–45, Vol. 2000, No. 5, pp. 9 – 15, 2000
  - (6) 尾崎浩司, 柴山 守, 荒木義彦: 古文書画像のレイアウト認識と標題抽出, 京都大学大型計算機センター第64回研究セミナー報告, pp. 59 – 73, 2000
  - (7) 尾崎浩司, 柴山 守, 荒木義彦: 古文書画像のレイアウト認識と標題抽出, 情報処理学会研究報告, 2000–CH–47, pp. 47 – 54, 2000
  - (8) 柴山 守, 藤川和利, ベンカテッシュ・ラガワン: 「日本経済史資料」のカラーマイクロ化とWWWによる画像検索システム, 大阪市立大学学術情報総合センター紀要, Vol. 1, pp. 32 – 37, 2000
  - (9) 富田浩章, 柴山 守, 荒木義彦: 古文書画像の2値化レベル制御による対話型文字分割とその評価, 電気学会論文誌C, Vol. 118–C, No. 4, pp. 503 – 509, 2000
  - (10) 山田寛治, 柴山 守: n-gramによる古文書証文類翻刻支援の検討, 人文科学とコンピュータシンポジウム論文集, Vol. 2000, No. 17, pp. 185 – 192, 2000
  - (11) 尾崎浩司, 柴山 守, 荒木義彦: 古文書画像の標題文字セグメンテーション, 人文科学とコンピュータシンポジウム論文集, Vol. 2000, No. 17, pp. 279 – 286, 2000
  - (12) 柴山 守: 「日本経済史資料」の大型古絵図撮影とデジタル化の事例, 月刊IM, 第39巻, 第12号, 2000
  - (13) 柴山 守: 大規模史料のカラーマイクロ化とWWWによる検索システム—歴史史料の〈保存〉と〈提供〉の両立を図る—, 人文学と情報処理, Vol. 30, pp. 97 – 106, 2001
  - (14) 後藤真, 高山典史, 柴山守: 正倉院文書のXML化による構造化を復元の検討, 情報処理学会研究報告 2001–CH–51, vol. 2001, No. 67, pp. 31 – 38, 2001
  - (15) 柴山 守, 吉井良邦: 近世資料アーカイブズのためのバーチャル図書館, 人文科学とコンピュータシンポジウム論文集, Vol. 2001, No. 18, pp. 109 – 116, 2001
  - (16) 柴山 守, 吉井良邦: 近世資料データベースとZ39.50標準による統合検索, 大阪市立大学学術情報総合センター紀要, Vol. 3, pp. 41 – 49, 2002
  - (17) 後藤 真, 柴山 守: 正倉院文書復原過程のXML/XSLTによる記述, 情報知識学会誌, Vol. 11, No. 4, pp. 2 – 16, 2002
  - (18) 山田寛治, 柴山 守: 古文書を対象にした文字認識の研究, 情報処理, Vol. 43, No. 9, pp. 950

– 955, 2002

- (19) 後藤 真, 柴山 守: 正倉院文書研究資料の XML/XSLT による記述と統合, 人文科学とコンピュータシンポジウム論文集, Vol. 2002, No. 13, pp. 209 – 216, 2002

#### 山田 瑩治

- (1) 山田 瑩治, 柴山 守, 原正一郎他: 古文書翻刻支援システム開発プロジェクト報告, 情報処理学会研究報告, 2000 – CH – 45, pp. 1 – 8, 1999
- (2) 和泉 勇治, 山田 瑩治, 柴山守他: ニューラルネットワークを用いた古文書個別文字認識に関する一検討, 情報処理学会研究報告, 2000 – CH – 45, pp. 9 – 16, 1999
- (3) 尾崎浩司, 柴山 守, 荒木義彦: 古文書レイアウト画像のピラミッド型抽象化と標題の自動抽出, 平成11年度電気関係学会関西支部連合大会論文集, 1999
- (4) 柴山 守, 星野 聰: ビデオ撮影による古文書画像の効率的画像入力法と自動接続処理, 情報処理学会論文誌, 40(3), pp. 885 – 893, 1999
- (5) 山田 瑩治: 浮世絵研究における数量的手法の導入—顔表情の分類と識別—, 画像ラボ, Vol. 10, No. 11, pp. 30 – 33, 日本工業出版, 1999
- (6) 山田 瑩治: 人文科学とコンピュータの10年, 情報処理学会研究報告, Vol. 99, No. 43, pp. 1 – 6, 1999
- (7) 山田 瑩治, 金子裕之: 文字認識手法による古筆分類の一方法, 人文科学とコンピュータシンポジウム論文集, 情報処理学会シンポジウムシリーズ, Vol. 99, No. 13, pp. 45 – 52, 1999
- (8) 山田 瑩治: 文化資料と画像処理, 勉誠出版, 2000
- (9) 山田 瑩治, 加藤寧, 川口洋, 原正一郎, 石谷康人, 柴山守, 笠谷和比古, 小島正美, 梅田三千雄, 山本和彦: 古文書翻刻支援システム開発プロジェクト報告 (1) —プロジェクト概要—, 情報処理学会研究報告, Vol. 2000, No. 8, pp. 1 – 8, 2000
- (10) 山田 瑩治, 柴山守: n-gram による古文書証文類翻刻支援の検討, 人文科学とコンピュータシンポジウム論文集, 情報処理学会シンポジウムシリーズ, Vol. 2000, No. 17, pp. 185 – 192, 2000
- (11) 山田 瑩治, 早川聞多: デジタル画像による浮世絵研究の試行, 情報処理学会研究報告, Vol. 2000, No. 100, pp. 25 – 32, 2000
- (12) 山田 瑩治, 加藤寧, 並木美太郎, 川口洋, 原正

一郎, 石谷康人, 柴山守, 笠谷和比古, 小島正美, 梅田三千雄, 山本和彦: 古文書翻刻支援システム (HCR) 開発プロジェクト報告 (2), 情報処理学会研究報告, Vol. 2001, No. 51, pp. 9 – 16, 2001

- (13) 山田 瑩治, 八村広三郎, 竹田正幸, 馬場章: 人文科学とコンピュータの学際的研究とは, 情報処理学会研究報告, Vol. 2001, No. 51, pp. 47 – 54, 2001
- (14) 山田 瑩治, 早川聞多, 塙原和郎: 情報科学的浮世絵研究の試み, 京都大学大型計算機センター第67回研究セミナー報告, pp. 62 – 69, 2001
- (15) 山田 瑩治: 日本文化の模倣と創造—オリジナリティとは何か—, 角川書店, 2002
- (16) 山田 瑩治, 早川聞多, 村上征勝, 塙原和郎: 浮世絵における顔表現の科学, 日本研究, 第25集, pp. 13 – 49, 2002
- (17) 山田 瑩治, 和泉勇治, 加藤寧, 柴山守: 類似文字検索機能をそなえた電子くずし字辞典の開発, 情報処理学会研究報告, Vol. 2002, No. 52, pp. 43 – 50, 2002
- (18) 山田 瑩治, 柴山守: 古文書を対象にした文字認識の研究, 情報処理, Vol. 43, No. 9, pp. 950 – 955, 2002
- (19) 山田 瑩治: 浮世絵の顔表現, 総研大ジャーナル, No. 1, pp. 24 – 29, 2002

#### 近藤 泰弘

- (1) 近藤泰弘: インターネットの変貌—メディアからソースへ—, 文学, 2001年9月号, pp. 207 – 209, 2000
- (2) 近藤泰弘: 文化資源としてのデジタルテキスト, 国語と国文学, 924, pp. 127 – 139, 2000

#### [ホームページ]

日本古典文学本文データベース (試験公開),

<http://www.nijl.ac.jp/>

国際コラボレーション計画,

<http://world.nijl.ac.jp/~kiban-s/>

## 古典文献データベースの表記体系確立

研究代表者 徳永 宗雄  
 京都大学大学院文学研究科 教授  
 分担者 山田 篤  
 財団法人京都高度技術研究所情報メディア研究室 室長

### 【発表成果一覧】

徳永 宗雄

Muneo Tokunaga (in print) Japanese Studies in the Epico-Puranic Literature (1990-2000). In: The Way to Liberation -Indological Studies in Japan, Vol. 2. Japanese Studies on South Asia, No. 5.

徳永宗雄（印刷中）「The Six Teachers of Bodhya-MBh 12. 171. 60-61をめぐって—『神子上恵生教授頌寿記念論文集』」龍谷大学文学部。

徳永宗雄「『マハーバーラタ』第12巻の成立に関する覚書」『印度学仏教学会』第51巻第1号、日本印度学仏教学会。67-69 (458-456)。

徳永宗雄（2002）「平安の巻」と水供養——「マハーバーラタ」第12巻の形成過程を探る——」『東方学』104輯、財団法人東方学会、1-15 (169-155)。

Muneo Tokunaga (2001) 「An Annotated Translation of MBh 12. 168-171」『京都大学文学部研究紀要』No. 40, 1-25.

Muneo Tokunaga (2001) 「An Annotated Translation of MBh 12. 172-174」『インド思想史研究』京都：インド思想史学会、13: 63-79.

徳永宗雄編（2001）『週間朝日百科 インドの文学I（世界の文学 115）』朝日新聞社。

Muneo Tokunaga (2000) Shift of Caesura in the Tris-

tubh of the Mahabharata. In: Japanese Studies on South Asia No.V: The Way to Liberation, Tokyo: Association of Asian Studies,

Muneo Tokunaga (2000) A Review of the Epic Subjunctive Form 『戸崎宏正博士古稀記念論集』九州大学出版会。517-532.

Muneo Tokunaga (2000) Portable Mahabharata (CD-ROM with verse-search program). 文部省科学研究費補助金特定領域研究 (A) 118 「古典学の再構築」計画研究研究成果。

徳永宗雄（2000）「オンライン古典学共同研究について」『古典学の再構築』No. 6. 文部省科学研究費特定領域研究 (A) 118 「古典学の再構築」28.

徳永宗雄（2000）「古典文献データベースの表記体系確立」文部省科学研究費特定領域研究 (A) 118 「古典学の再構築」研究成果中間報告書。66-68.

徳永宗雄（2000）「情報処理」文部省科学研究費特定領域研究 (A) 118 「古典学の再構築」研究成果中間報告書。12-13.

Muneo Tokunaga (2000) The Metrically Processed Text of the Mahabharata. Digital text uploaded on [http://tiger.bun.Kyoto-u.ac.jp/mtokunag/skt\\_texts](http://tiger.bun.Kyoto-u.ac.jp/mtokunag/skt_texts) で公開中（電子テキスト）。

Muneo Tokunaga (2000) The Metrically Processed Text of the Ramayana. Digital text uploaded on [http://tiger.bun.Kyoto-u.ac.jp/mtokunag/skt\\_texts](http://tiger.bun.Kyoto-u.ac.jp/mtokunag/skt_texts) で公開中（電子テキスト）。

Muneo Tokunaga (2000) The Fifteen Principal Upanisads for Word-search. Digital text uploaded on [http://tiber.bun.kyoto-u.ac.jp/mtokunag/skt\\_texts/Upanisads](http://tiber.bun.kyoto-u.ac.jp/mtokunag/skt_texts/Upanisads).

Muneo Tokunaga (2000) Japanese Studies of Indian Philosophy and Religion. In: Towards Understanding Each Other. Fifty Years' History of India-Japan Mutual Studies. Ed. by Ch. Nakane and M. Naito. Committee for Japan-India Academic Exchange. 26-30.

Muneo Tokunaga (1999) Passages Related to the Brhaddevata Skandasvamino's Commentary on the Rgveda. インド思想史研究11 (インド思想史学会), 104 – 112.

徳永宗雄 (1999) 「ラーマーヤナとヒンドゥー教」『ラーマーヤナの宇宙』春秋社, 49 – 56.

徳永宗雄 (1999) 「人文科学と情報処理」『古典学の再構築』No. 3. 文部省科学研究費特定領域研究 (A) 118 「古典学の再構築」43.

#### 4 A03班・計画研究

## 古典学のための多言語文書 処理システムの開発

研究代表者 高島 淳

東京外国语大学アジア・アフリカ言語文化研究所 教授

分担者 峰岸 真琴

東京外国语大学アジア・アフリカ言語文化研究所 助教授

分担者 濱田(星) 泉

東京外国语大学アジア・アフリカ言語文化研究所 助手

### 【発表成果一覧】

高島淳, 「絶対者とヒンドゥー教」, 阿部美哉・園田稔編『世界の宗教を学ぶ人のために』世界思想社, (印刷中).

Takashima, Jun, "Pratistha in the Saiva Agamas," Studies in the Pratistha, ed. S. Einoo (Delhi: Manohar), in press.

高島淳・澤田英夫, 「インド系文字について」, 「アジア文字曼陀羅～インド系文字の旅」実行委員会編『「アジア文字曼陀羅～インド系文字の旅」ガイドブック』, 東京外国语大学アジア・アフリカ言語文化研究所, pp. 3 – 6, 2002.

高島淳, 「ヒンドゥー教儀礼のマニュアル」, 「アジア文字曼陀羅～インド系文字の旅」実行委員会編『「アジア文字曼陀羅～インド系文字の旅」ガイドブック』, 東京外国语大学アジア・アフリカ言語文化研究所, pp. 40 – 41, 2002.

高島淳・山部順治, 「ヴィシヌ神の化身を描いた絵入り写本」, 「アジア文字曼陀羅～インド系文字の旅」実行委員会編『「アジア文字曼陀羅～インド系文字の旅」ガイドブック』, 東京外国语大学アジア・アフリカ言語文化研究所, pp. 54 – 55, 2002.

高島淳, 「チャイタニヤ派ヒンドゥー教の写本」, 「アジア文字曼陀羅～インド系文字の旅」実行委員会編

「『アジア文字曼陀羅～インド系文字の旅』ガイドブック」，東京外国語大学アジア・アフリカ言語文化研究所，pp. 58–59, 2002.

高島淳，「ヴィシュヌ神の称名集」，「アジア文字曼陀羅～インド系文字の旅」実行委員会編「『アジア文字曼陀羅～インド系文字の旅』ガイドブック」，東京外国語大学アジア・アフリカ言語文化研究所，pp. 60–61, 2002.

高島淳，「19世紀サウラーシュトラ文字の写本」，「アジア文字曼陀羅～インド系文字の旅」実行委員会編「『アジア文字曼陀羅～インド系文字の旅』ガイドブック」，東京外国語大学アジア・アフリカ言語文化研究所，pp. 62–63, 2002.

高島淳，「『マラバル・グランタあるいはサンスクリット文字』と題されたマラヤーラム文字の教本」，「アジア文字曼陀羅～インド系文字の旅」実行委員会編「『アジア文字曼陀羅～インド系文字の旅』ガイドブック」，東京外国語大学アジア・アフリカ言語文化研究所，pp. 78–79, 2002.

Minegishi, M. and Takashima, J., "Issues in Transliteration for Khmer and Thai Scripts," The Indo-European Conference on Multilingual Technologies (IEMCT), ed. by R. K. Arora, Mahesh Kulkarni, Hemant Darbari, New Delhi:Tata McGraw-Hill Publishing Co.Ltd. pp. 161-175, 2002.

高島淳，「インドはわからない？」，平凡社東洋文庫編集部編『東洋文庫ガイドブック』，平凡社，pp. 101–115, 2002.

高島淳，「島 岩+坂田貞二〔編〕『聖者たちのインド』」，『南アジア研究』第13号，pp. 242–245, 2001.

高島淳，「古典学のための多言語文書処理システムの開発」，『第Ⅰ期公募研究論文集』文部科学省科学研究費補助金特定領域研究(A) 118「古典学の再構築」(平成10年度～平成14年度)「古典学のための多言語文書処理システムの開発」，pp. 174–184, 2001.

高島淳，「旧フランス国立印刷局の活版印刷の伝統と現状」，『アジアの文字と出版・印刷文化及びその歴史に関する調査・研究——デジタル化移行の基礎とし

て——』平成11年度～平成12年度科学研究費補助金基盤研究(A) (2) 研究成果報告書(研究代表者 町田和彦)，pp. 21–38, 2001.

高島淳，「グランタ文字」，『華麗なるインド系文字』(町田和彦編著)白水社，pp. 198–201, 2001.

高島淳，「古典学のための多言語文書処理システムの開発」，『第Ⅰ期研究成果報告』文部省科学研究費補助金特定領域研究(A) 118「古典学の再構築」，pp. 162–165, 2001.

高島淳，「小倉泰著『インド世界の空間構造——ヒンドゥー寺院のシンボリズム——』」，『宗教研究』第74巻第2輯325号，pp. 187–192, 2000.

高島淳，「サンスクリット語の機械可読辞書の開発とバーザへの適用」「インド諸言語のための機械可読辞書とバーザの開発」平成9年度～平成11年度科学研究費補助金基盤研究(A)(2)研究成果報告書(研究代表者 ペーリ・バースカララー)オ，pp. 73–105, 2000.

#### 峰岸 真琴

Minegishi, Makoto, "Morphological typology from Southeast Asian viewpoint", Papers from XIth SEALS Conference, Arizona State Univ., in press.

峰岸真琴，「翻訳評価の通言語的評価の試み」，『機械翻訳——多言語機械翻訳システムの評価研究』，pp. 42–52, 2002.

峰岸真琴，「形態類型論の形式モデル化」，『アジア・アフリカ言語文化研究』64号，pp. 101–128, 2002.

Minegishi, M. and Takashima, J., "Issues in Transliteration for Khmer and Thai Scripts," The Indo-European Conference on Multilingual Technologies (IEMCT), ed. by R. K. Arora, Mahesh Kulkarni, Hemant Darbari, New Delhi: Tata McGraw-Hill Publishing Co.Ltd. pp. 161-175, 2002.

峰岸真琴，「タイの彩色仏教経典」，『アジア文字曼陀羅～インド系文字の旅』実行委員会編「『アジア文字曼陀羅～インド系文字の旅』ガイドブック」，東京外国語大学アジア・アフリカ言語文化研究所，pp. 22–23, 2002.

峰岸真琴, 「パーリ語の仏教經典」, 「アジア文字曼陀羅～インド系文字の旅」実行委員会編『「アジア文字曼陀羅～インド系文字の旅」ガイドブック』, 東京外国語大学アジア・アフリカ言語文化研究所, pp. 30 - 31, 2002.

峰岸真琴, 「エモニエ『フランス語カンボジア語辞典』」, 「アジア文字曼陀羅～インド系文字の旅」実行委員会編『「アジア文字曼陀羅～インド系文字の旅」ガイドブック』, 東京外国語大学アジア・アフリカ言語文化研究所, pp. 76 - 77, 2002.

峰岸真琴, 「シンタグマから見た動詞」, 『日本言語学会第124回大会予稿集』, pp. 16 - 23, 2002.

峰岸真琴, 「類型分類の再検討——孤立語の視点から」, 『アジア・アフリカ言語文化研究』63号, pp. 1 - 36, 2002.

Minegishi, Makoto, "Southeast Asian Languages --- A Case for the Caseless?", Working Papers: International Symposium on Non-Nominative Subjects, pp. 452-464, 2001.

峰岸真琴, 「クメール文字」, 『言語学大辞典 世界文字辞典』三省堂, pp. 349 - 357, 2001.

峰岸真琴, 「バハウ・フモン文字」, 『言語学大辞典 世界文字辞典』三省堂, pp. 747 - 749, 2001.

峰岸真琴, 「言語記述の科学性を巡って——メタ理論的考察」, 『ELPR 2002』(特定領域研究「環太平洋の「消滅に瀕した言語」にかんする緊急調査研究」実績報告書), pp. 55 - 67, 2001.

## 星 泉

### [図書]

星泉, 「現代チベット語ラサ方言動詞辞典」, アジア・アフリカ基礎語彙叢42, 東京外国語大学アジア・アフリカ言語文化研究所, pp. 500, (印刷中)

### [論文等]

星泉, 「古代チベットの王家の谷で——撮影現場レポート & その後——」, 『古典学の再構築』第12号, pp. 66

- 67, 2002.

星泉, 「チベットの歌舞劇の物語本」, 「アジア文字曼陀羅～インド系文字の旅」実行委員会編『「アジア文字曼陀羅～インド系文字の旅」ガイドブック』, 東京外国語大学アジア・アフリカ言語文化研究所, pp. 32 - 33, 2002.

星泉, 「革表紙のチベット常用経典」, 「アジア文字曼陀羅～インド系文字の旅」実行委員会編『「アジア文字曼陀羅～インド系文字の旅」ガイドブック』, 東京外国語大学アジア・アフリカ言語文化研究所, pp. 34 - 35, 2002.

星泉, 「シッキムの手写本」, 「アジア文字曼陀羅～インド系文字の旅」実行委員会編『「アジア文字曼陀羅～インド系文字の旅」ガイドブック』, 東京外国語大学アジア・アフリカ言語文化研究所, pp. 42 - 43, 2002.

星泉, 「玄中寺碑文拓本」, 「アジア文字曼陀羅～インド系文字の旅」実行委員会編『「アジア文字曼陀羅～インド系文字の旅」ガイドブック』, 東京外国語大学アジア・アフリカ言語文化研究所, pp. 74 - 75, 2002.

星泉, 「『タングートあるいはチベットの文字』と題されたチベット文字の教本」, 「アジア文字曼陀羅～インド系文字の旅」実行委員会編『「アジア文字曼陀羅～インド系文字の旅」ガイドブック』, 東京外国語大学アジア・アフリカ言語文化研究所, pp. 82 - 83, 2002.

永崎研宣・星泉, 「Web 環境での多言語表示」, ITU ジャーナル(財団法人日本 ITU 協会), pp. 26 - 29, 2002.

星泉, 「現代チベット語の名詞化接辞を持つ述語の意味——名詞述語から動詞述語への拡張と展開——」, 『東京大学言語学論集』20号, pp. 261 - 274, 2001.

### [図書]

Hoshi, Izumi, Imaeda, Yoshiro et alii, Choix de Documents Tibétains conservés à la Bibliothèque Nationale, Tome IV Corpus Syllabique, ILCAA, pp. 1143, 2001.

### [講演]

星泉, 「世田谷セミナリオ「アジアに広がるインド系文字」(1) インド系文字の広がりとしくみ」世田

【口頭発表】

星泉、「多言語処理技術の基盤整備」、新エネルギー・  
産業技術総合開発機構主宰「平成14年度 NEDO 研究  
助成事業研究成果報告会」、2002.12.2.

## 平安時代物語文の比較計量的研究

研究代表者 今西 裕一郎

九州大学大学院人文科学研究院 教授

分担者 小西 貞則

九州大学大学院数理学研究院 教授

分担者 室城 秀之

白百合女子大学文学部 教授

### 【発表成果一覧】

村上征勝、今西裕一郎 源氏物語の助動詞の計量分析  
「情報処理学会論文誌」40巻3号 1999 pp.774-782.

室城 秀之

「うつほ物語の和歌総合研究」[単著]

私家版（平成11年度白百合女子大学研究奨励）2000  
-01

「うつほ物語」—祝祭と共食の物語 [単著]

週刊朝日百科世界の文学26 「竹取物語 伊勢物語」  
(朝日新聞社) 2000-01

[論文]

「うつほ物語」調度関係語彙—照明具 [単著]

「国文白百合」31 2000-03 pp.39-47. 9

[論文]

「うつほ物語」の三条院について—「源氏物語」の  
六条院との比較を通して [単著]

論集平安文学5「平安文学の想像力」(勉誠出版)2000  
-05

[論文]

「うつほ物語」飲食関係語彙総覧—飲食器・調理器  
編一 [単著]

「白百合女子大学研究紀要」36 2000-12

[著書]

「古今和歌六帖」自立語索引—第一帖— [単著]

私家版（平成12年度白百合女子大学研究奨励）2001  
-01

[論文]

「うつほ物語」飲食関係語彙総覧—飲食行為・調理  
行為編一 [単著]

[論文]

「うつほ物語」の様 [単著]

「国文学解釈と教材の研究」46-5 2001-04

[著書]

「うつほ物語の総合研究2 古注釈編」[共著]

〔勉誠出版〕2002-02

[論文]

鯉と雉—「うつほ物語」の飲食表現— [単著]

白百合女子大学言語・文学研究センター「言語・文学研究論集」2 2002-03

[論文]

料理する男たち—「うつほ物語」の飲食表現— [単著]

「国文白百合」33 2002-03 pp. 21-29. 9

[論文]

The Current State of Research on the Utsufo Monogatari [単著]

ACTA ASIATICA 002-09 pp.18-32.

## インド古典天文学書の研究と 伝統暦プログラムの改良

研究代表者 矢野 道雄  
京都産業大学 教授

### 【発表成果一覧】

[論文]

Michio Yano, 'La Cina e le Zone Limitrofe', Storia Della Scienza (Enciclopedia Italiana's International Encyclopaedia of the History of Science), Volume II Cina, India, Americhe , Edited by K. Chemla, Rome: Instituto della Enciclopedia Italiana, 2002, pp. 269-275.

矢野道雄, 「顕在化する大規模国際会議の課題」「学術の動向」2002.5, 日本学術会議, 2002年5月。

Michio Yano, 'Yabuuti Kiyosi as a Historian of Exact Sciences', East Asian Science, Technology, and Medicine, No. 18, (2001), pp. 13-19.

Michio Yano, 'Kiyosi Yabuuti (1906-200)', Archives internationales d'histoire des science, Vol. 51 (2001), pp. 155-157.

Michio Yano, S.B. Rao, Indian Astronomy: An Introduction , [Indian Journal of History of Science, 36.1-2 (2001), pp. 65-73.] (書評論文)

矢野道雄, 「インド古典天文学書の研究と伝統暦プログラムの改良」「文部省科学研究費補助金特定領域研究(A) 118「古典学の再構築」第Ⅰ期公募研究論文集」2001年8月, pp. 165-173.

矢野道雄, 「インド暦プログラム PANCANGA について」(『古典学の現在 I』2000年, pp. 35-52.

Michio Yano, 'The First Equation Table for Mercury in the Huihui Li', Memoirs of the International Institute

for Linguistic Sciences, Kyoto Sangyo University, No. 1 (1999), pp. 233-244.

Michio Yano, 'Tables of Planetary Latitude in the Huihui li (I)', Current Perspectives in the History of Science in East Asia, ed. by Yung Sik Kim and Francesca Bray, Seoul National University, 1999 (June 30), pp. 307-315.

#### [講演]

「インドの天文学と宇宙論」, 国際基督教大学アジア文化研究所主催

ミニシンポジウム「アジアの宇宙観」, 2002年12月21日。

「インドの科学とことば」, 京都大学人文科学研究所文化創生フォーラム 第2回「科学のことばと宗教のことば」, 2002年12月19日。

「インドにおける科学の口承と文献伝承」, 国際基督教大学科学史フォーラム, 2002年10月30日。

'Oral and Written Transmission of the Exact Sciences in India', International Workshop on Asian Contribution to the Formation of Modern Science, International Institute for Asian Studies, Leiden, September 20, 2002.

'Mathematical Science in India and China—Transmission and Comparison', Internatinal Colloquium on the History of Mathematics, Northwest University, Xi'an, China, August 15, 2002.

'Medicine and Divination in India', Workshop on Divination in China: a comparative perspective, Needham Research Institute, December 7, 2001 'Yabuuti Kiyosi as a Historian of Exact Sciences', 21st International Congress of History of Science, Mexico City, July 9, 2001.

'Indian astronomy and Pancanga Program', National Institute of Advanced Studies, Bangalore, Septermber 15, 2000.

'Buddhism as a Vehicle of Astrology, History of Astrology Lectures at Warburg Institute, University of London, Summer Term, May 3, 2000.

## 抄物の原典参照データベースの構築 —『韻府群玉』と『玉塵抄』を例として—

研究代表者 出雲 朝子  
青山学院女子短期大学 教授

分担者 豊島 正之  
東京外国语大学アジア・アフリカ言語文化研究所 助教授

#### [発表成果一覧]

出雲朝子, 「文語体抄物における主格助詞「が」—抄物文の近代語的性格—」, 『国語と国文学』, 平成14年1月特集号, 東京大学国語国文学会, pp. 96-106, 2002.

豊島正之, 「音韻を計る」, 上野善道編『音韻』朝倉日本語講座3, 朝倉書店, (印刷中).

豊島正之, 「キリストン文献の漢字整理について」, 『国語と国文学』79卷11号 (平成14年11月「近代語」特集号), 東京大学国語国文学会, pp. 47-59, 2002.

豊島正之, 「漢字による情報交換と漢字制限」, 『漢字文化圏の諸相』筑波大学東西言語文化の類型論特別プロジェクト研究成果報告書平成13年度別冊, 筑波大学, pp. 37-54, 2002.

豊島正之, 「「ぎやどべかどる」解説」, 尾原悟編『ぎやどべかどる』(『キリストン研究』38), 教文館, pp. 354-392, 2001.

#### [口頭発表]

清水康行・豊島正之, 「一世紀前の日本語の音声」, 『一世紀前の日本語の音声—音源アルヒーフの1900-1901年の録音資料から—GICAS PHONARCプロジェクト研究発表会』, アジア・アフリカ言語文化研究所COE拠点 GICAS, 2003. 2. 18.

豊島正之, 「言語普遍の系譜—セッションの序として」, 『東京外国语大学アジア・アフリカ言語文化研

究所 AA 研棟竣工記念シンポジウム「誰が世界を翻訳するのか」, 東京外国語大学アジア・アフリカ言語文化研究所, 2002. 6. 29.

豊島正之, 「XML の骨抜き利用法——アジア・アフリカ言語文化研究所データベースの例」, 科学研究費特定領域研究「古典学の再構築」情報処理班主宰研究会「XML pro/con——XML で書く文献学的データ」, 九州大学文学部, 2001. 10. 27.

豊島正之, 「漢字による情報交換と漢字制限」, 筑波大学「東西言語文化の類型論」特別プロジェクト主宰国際シンポジウム「日本・韓国における漢字受容の過去・現在・未来」, 筑波大学国際会議場, 2001. 7. 27.

豊島正之, James William Breen, 「キリスト教版「ぎやどべかどる」の活字字体についての計量的報告」, 近代語研究会第184回研究発表会, 神戸山手大学講堂, 2001. 5. 18.

## 論集「情報処理」「古典学の再構築」研究成果報告集 IV

---

平成15年3月20日発行

編集者 高島 淳

編集所 文部科学省科学研究費補助金 特定領域研究

「古典学の再構築—20世紀後半の研究成果総括と  
文化横断的研究による将来的展望」

A03「情報処理」調整班

〒183-8534 東京都府中市朝日町3-11-1

東京外国语大学アジア・アフリカ言語文化研究所

発行者 中谷英明

発行所 「古典学の再構築」総括班

〒651-2180 神戸市西区伊川谷町有瀬518

神戸学院大学人文学部

印 刷 株式会社シンクス

---

३  
विश्वामीति योगीत्वा त्रिविक्षुपा  
त्रिविक्षुपा योगीत्वा योगीत्वा  
योगीत्वा योगीत्वा योगीत्वा  
योगीत्वा योगीत्वा योगीत्वा  
योगीत्वा योगीत्वा योगीत्वा

1998 - 2002



古典学の再構築