

## A03-24・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

研究分担者 古瀬 順一  
宮崎大学教育文化学部 教授

## 研究目的

古典学研究のための、コンピュータを利用した新たな研究法の確立が本研究の目的である。特に、コンピュータに入力された古典テキストの文章の数量的特質（文長、品詞の出現頻度、単語の出現頻度等）の計量分析を中心に、従来の古典学の研究法とは異なる切り口で古典を分析する方法を確立し、古典学研究の新しい分野を開拓することを試みる。

## ●学術的な特色・独創的な点

古典に関する研究を進める上で、語彙、構文、文法、音韻等に関する情報の分析は基本的かつ重要な意味を持つ。しかし、これまでのこのような情報の分析を計量的な観点から行った研究は少ない。本研究では、古典に関する具体的問題の解決を通じて、古典研究における語彙、構文、文法、音韻等の計量的分析の有効性を示す。そのため、『源氏物語』及び関連文献、『梵文法華経』、『八千頌般若経』、『十地経』、井原西鶴全作品のデータベースを利用し、語彙、構文、文法、音韻等の計量的分析を行い、それらの文献の著者や成立などに関する疑問点の解明を試みながら、文章の計量的分析による古典研究という新分野を開拓する。

## ●国内外研究の中での当該研究の位置づけ

コンピュータを積極的に利用した文章の計量的な分析の研究は始まったばかりであるが、諸外国では古典の研究に計量的分析を用いた研究がいくつかみられる。しかし、日本の古典に関しては研究が非常に遅れており、研究代表者の一連の研究以外見当たらない。

## 研究計画

## ●基本方針

本研究の期間内で古典文献のデータベースを新たに作成し、それを用いて研究を進めることは、研究期間、研究経費からみて不可能に近い。したがって、すでにある程度完成している『源氏物語』、『西鶴作品』、『サンスクリット大乘仏典』等のデータベースを利用することを予定している。

研究は日本語の古典文献とサンスクリットの古典文献を中心に行う。

## ●研究計画

(1) 諸外国における、文献の計量分析に関する最近の研究動向に関する情報収集。

(2) 『源氏物語』のデータベースを中心に計量分析を進め、『源氏物語』の著者や54巻の成立順序に関する疑問の解明を試みる。この分析を通じて日本語古典文献の計量分析の問題点を検討する。また日本語のコンピュータ処理に関する最近の研究を調査し、すでに入力されているプレーンなデータ（文章を単に入力しただけで単語分割、品詞コードが付加されていないデータ）をどのように加工し計量分析につなげたらよいかを検討する。

## A03-25・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

研究分担者 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

武井 協三  
国文学研究資料館研究情報部 教授

松村 雄二  
国文学研究資料館研究情報部 教授

石塚 英弘  
図書館情報大学図書館情報学部 教授

柴山 守  
大阪市立大学学術情報総合センター 教授

アンドル・アーマー  
慶應大学文学部 教授

中村 康夫  
国文学研究資料館研究情報部 助教授

原 正一郎  
国文学研究資料館研究情報部 助教授

山田 哲好  
国文学研究資料館史料館 助教授

山田 奨治  
国際日本文化研究センター研究部 助教授

特定領域研究「古典学の再構築」の研究目的に従い、「情報処理」計画研究では古典学におけるコンピュータ利用の確立と普及を研究目的とする。とくに、古典の読解における諸作業の技法の確立が、本研究の主要

課題と位置づけられている。

したがって、本研究では電子化テキストの作成、処理、交換、流通までを総合的に扱い、また促進をはかるため、標準的な電子化テキストを実証的に研究し、具体的なテキストを作成し、また機能、処理システムを開発研究し、古典学におけるコンピュータ利用のあり方を研究する。

そこで、本研究では古典学テキストのデジタル化の推進と流通を具体的課題とし、従来2班で構成した計画研究を整理統合し、より効率的な研究を下記の4課題に集約し、進めることとしている。

#### (1) 古典学における電子化テキストの現状把握

古典学におけるテキスト処理を中心とし、コンピュータ利用法と電子化テキストの現状を調査研究する。電子化テキストの対象・内容・構造・文字セットなどに加え、システムさらに著作権に関わる諸問題を含むデータベース流通性などを調査研究する。

#### (2) 標準的な電子化テキストの作成

古典学における電子化テキストの有効な構築法を実証的に検討するため、実際に標準の電子化テキストを作成し、評価する。XML(またはSGML)に基づいた古典作品の電子化を進める。また、データ構造の変換法を確立して既存テキストデータの共有化も促進する。原本イメージ情報とテキストとの比較、対応を含むマルチメディア処理の研究、および古今集などを基礎とした日欧文によるパラレルテキストのデータベース構築とその利用法の研究を進める。

#### (3) テキストデータベースの流通

まず、自立型のデータファイルとして流通をはかる。MO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する。ついで、インターネット(WWW)を前提としたデータベースの構築と流通をはかる。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。

#### (4) 基本的なテキスト解析と処理システム

まず、基礎的なテキスト処理プログラムの開発研究を行う。例えばコンコーダンスとして必要な機能範囲を確定し、プロトタイプの開発と検討を行う。ついで、SGMLテキストデータに対する全文検索エンジン(OpenTextなど)の機能面での実証研究を行う。

総じて、本研究において、作成する電子化テキストは標準の言語仕様によるため、古典学分野における規範的なデータ記述の仕様となることが期待される。電子化のプロセスを含め、電子化テキストの処理技術を領域全体に促進することができる。また、データ流通

はCD-ROM等のパッケージ型を実用化し、またインターネットによる高次活用法を含め、国際流通化を促進する。

---

### A03-26・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 助教授

濱田(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

### 研究の目的

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、この言葉は、現在最も一般的に用いられているナーガリー文字および論文に用いられるローマ字転写の他に、グラントラ文字、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によって表記されてきた伝統を持っている。

インド文明圏においては、このような同一言語多書記法(script)表記や同一書記法(script)多言語表示が一般的であり、写本資料の正確な転記のような場合にはこうしたあり方を正確に反映できるような多言語文書処理システムが必要となる。

こうした需要に応えるために、研究代表者は、「古典学再構築」の前半2年間において、公募研究として、本研究と同一のテーマにおいて、ローマ字転写、ナーガリー文字、グラントラ文字、カンナダ文字、タミル文字、マラーヤラム文字に関して、多言語文書処理システムを開発した。

これによって、システムの汎用的拡張性が確認されたことから、後半の2年間において、その他のインド諸言語、東南アジアのタイ系文字・クメール文字およびチベット文字についての拡張を行うのが本研究の目指すところである。

インドに関しては、テルグ文字、ベンガル文字、オリヤー文字、グルムキー文字を追加して、インドの州

公用語のうちでグジャラーティーをのぞくすべてのインド系文字に対応する。

インドの外に関しては、新たに研究分担者として、アジア・アフリカ言語文化研究所の東南アジアとチベットの専門家に加わってもらい、タイ語の場合におけるパーリ語からの借用語、チベット語におけるサンスクリット語からの借用語などの従来の文字処理システムにおいては処理できないような問題を解決できるシステムを構築することを目的とする。

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十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 による諸文化に適應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と達するには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、インド系文字を用いているインドとチベットと東南アジアを包括的な一つの体系で処理できるようにすることの意義は大きい。

#### 研究計画・方法

インドの諸文字体系については、平成11～12年度に開発したシステムを拡張して、ベンガル語・オリヤー語・テルグ語・グルムキー文字を追加して、インドの主要公用語に用いられる10種の書記体系のうちの9種

までを統合して処理して文字出力できるようなシステムを開発する。

チベット語に関しては、サンスクリット語を含む古典チベット語および現代チベット語のテキストを入力して、文字の結合規則の相違に関するデータを収集して、チベット語処理において現代文と古典文の間での処理分岐規則を明らかにする。

タイ語に関しては、パーリ語を含む仏教タイ語および現代タイ系言語のテキストを入力して、文字の結合規則の相違に関するデータを収集して、タイ語系言語処理において現代文とパーリ語等からの借用語との間での処理分岐規則を明らかにする。

さらに、パーリ研究の点からも重要なクメール文字の処理システムも開発する。

チベット語処理システムおよびタイ語処理システムの実装のために、チベット文字およびタイ文字のフォントを開発し、両言語の処理システムを開発する。

言語処理において重要なソーティングを実現するために、タイ語およびチベット語のソートシステムを開発する。

このようにローマ字転写、ナーガリー文字、グラントタ文字、カンナダ文字、タミル文字、マラヤーラム文字、ベンガル文字、オリヤー文字、テルグ文字、グルムキー文字、チベット文字、タイ文字、クメール文字を印刷出力とすることの出来るシステムを完成させる。

研究の成果は、出版物 (CD-R パッケージ) と WebPage の形で公表する予定である。

---

---

#### A03-27・計画研究

### 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

研究分担者 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

本計画研究では、以下に挙げる三つのテーマに関して研究を行う。

(1) デジタル技術の急速な進展に伴い、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、本計画研究では昨年度、マハーバーラタの哲学篇である『解脱法品』(12.168 - 171) を対象にインターネットでの共同

研究を開始した。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巡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訂してインターネットに公開するというものである。昨年度の研究でこの共同作業のプロセスをほぼ確立することができたので、今年度からの二年間は関係資料を収集しつつ、このプロセスに従って171章に続く章の訳注を行い、さらに、『解脱法品』に関わる叙事詩以外の文献の研究も行うつもりである。また、その成果はウェブページに掲載し、活字で学術雑誌にも発表していきたい。なお、昨年度の研究成果は『京都大学文学部研究紀要』30号に“An Annotated Translation of MBh 12.168 - 171”として発表した。

(2) 文献そのものを電子化することにより、古典学においてもコンピュータ支援の下での様々な研究が可能となってきた。ここで問題となるのは電子化に際しての表記の問題である。従来は、特定の目的毎に専用のフォーマットや表記上の取り決めを定め、一度限りの利用のために多大な労力を払うことが多かったが、このようにして作られたデータはそのままの形では他所で利用することが難しい。近年、WWWの隆盛により、一部のデータはHTMLの形式で交換されるようになったが、HTMLは本来、WWWでの表示を指向したマークアップ言語であり、表現能力そのものは乏しい。そのため、最近では、表示指向から内容指向への発想の転換と、再利用可能な汎用性のある表記法の必要性が認識されつつある。このような状況を踏まえて、本研究では、古典文献の研究で再利用可能で、かつ、汎用性のある表記方法をXMLアプリケーションとして規定し、このアプリケーションを使ってデータを電子化するとともに、様々な形で再利用可能なデータベースを実現することを目指している。なお、このデータベースは(1)に挙げたオンライン共同研究の公開に応用していきたい。

(3) 前二年間の本計画研究でなされた文字コード体系の研究成果を引継ぎ、文字コードシステムに関する最近の動向に関する情報の収集をさらに続けていきたい。

## A03-28・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

研究分担者 小西 貞則  
九州大学大学院数理学研究院 教授

室城 秀之  
白百合女子大学文学部 教授

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、もっ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、し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。

前期における具体的な作業は、前田本を底本とする「うつほ物語」の品詞情報付きデータベースの作成であったが、市販の校訂本ではなく原本影印に基づく作業であったために多大の労力と時間を、本文入力、単語分割等の基礎作業に取られ、数量化3類の分析に供しうる完成データの作成は、俊陰、蔵開の2巻にとどまったことは遺憾であった。

したがって、前期の報告書で紹介した数量化3類の分析については、その分析結果が暫定的なものであることを認めざるをえない。後期において全巻のデータベースの完成を急ぎ、さらに多様な分析データを蒐集し、考察を試みたい。

と同時に、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とも保証のきざりではない。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。

## A03-28・公募研究

### 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

研究分担者 豊島 正之  
東京外国語大学アジア・アフリカ言語文化研究所 助教授

#### 研究目的

この研究は、中世日本語資料である「抄物」(しょうもの)中、屈指の量と資料性に聞こえる『玉塵抄』と、その中国原典『韻府群玉』のそれぞれの内部構造、及びそれらの注釈・被注釈関係をネットワークとして捉え、それをXML(Extensible Markup Language)による表現形式で構造データ化し、データベースとして実装・運用して、日本の抄物である『玉塵抄』が、中国原典『韻府群玉』を如何に受容したかについての実証的な研究の基礎とする。又、一般的な注釈と原典との注釈・被注釈関係という文献学的情報の表現形式の策定とそのデータベース化を試みる。

『玉塵抄』は、韻書の発展形である『韻府群玉』中の小見出し(熟語・成句)に対する注釈書であり、その読解には被注文献『韻府群玉』の内部構造の意識が常に必要であるが、『韻府群玉』小見出しの構造的検索が困難なため、項目相互が関連付けにくく、著しく難解である。又、被注原典に対する抄物側の施注の精疎は、従来の研究では必ずしも十分には追究し得なかった処であるが、『韻府群玉』のような原典に於ては、これこそが解読の要となる情報である。被注文献とその注釈の階層関係の追跡には、情報の相互関係をネットワークとして表現し、それを随意に検索できるデータベース化が不可欠である。

この研究では、『韻府群玉』の小見出し(熟語・成

句)部分をテキストデータ入力し、過年度の研究で入力済の『玉塵抄』加注部分との相互参照関係(リンク)をXMLの構造・リンク情報として表現し、そこからデータベースを(半)自動構築して、『玉塵抄』本文と関係させた注釈・被注釈データベースへと発展させる。この結果、『玉塵抄』の加注・非加注だけでなく、参照・非参照の精疎の実態が初めて統計的にも明らかになり、日本の抄物である『玉塵抄』が、中国原典『韻府群玉』を如何に受容したかについての実証的な研究が可能になる。又、注釈・被注文献間の引用・階層関係のXMLによる表現形式の策定と、そこからのデータベース化の試験的な実装を行うことで、一般的な注釈と原典との注釈・被注釈関係という文献学的情報のデータ表現形式に関する提言を行う。

従来のテキストデータ形式は、文献の内部構造自体の表現への提言は少なくないが、当該文献外への参照(引用・注釈・被注)情報の表現方法が貧弱で、抄物とその原典のような注釈・被注釈関係を表現する形式を欠いている。この研究は、この解決を主眼とし、XMLによる表現として一般性・可搬性を損なわない形でデータ形式を策定するだけでなく、そこからデータベースを(半)自動構築することで、運用にも便である実用的な形式を策定する事を目指す。これによって、注釈・被注釈関係にある文献群に応用可能な、柔軟且つ洗練されたデータ表現方法が獲得されるものと期待される。

#### 研究計画・方法

1. 『韻府群玉』小見出し部分の入力『韻府群玉』の諸本の原本調査・撮影を経て、『韻府群玉』のうち、韻から直接検索できない小見出し部分をテキストデータ入力する。『玉塵抄』が注釈対照としない部分(韻府群玉巻七平声庚韻以降)も、『玉塵抄』中に引用されるので、全て入力する。

2. データ形式の策定と既入力済『玉塵抄』のデータ形式の変換『玉塵抄』のうち、過年度の研究で既に入力済の部分のデータ形式を、今回の研究で策定するXML(Extensible Markup Language)形式へ変換する。

3. 『玉塵抄』から『韻府群玉』への参照情報のXMLによる表現『韻府群玉』小見出しと『玉塵抄』の関連箇所からの参照データを入力し、リンクをXMLで表現して『韻府群玉』への参照情報として整備する。

4. 注釈・被注釈データベースの試験運用とデータ表現形式の洗練上記で作成したXMLデータ形式からSQL系データベース(mysqlを予定)へ(半)自動変換を行い、相互参照データベースとして試験運用し

て、データ表現形式を吟味する。XML形式から運用可能なデータベースへの(半)自動変換は、近時例が増えて来たため、大きな困難はないものと見込んでいる。尚、回線事情が許せば、インターネットからのデータベース検索も可能にする予定である。

## A04「古典の世界像」

### A04-30・計画研究 東アジアの科学と思想

研究代表者 川原 秀城  
東京大学大学院人文社会系研究科 教授  
研究分担者 梁 一模  
東京大学大学院人文社会系研究科 助手

#### 研究目的

本研究の目的は、東アジアの前近代の科学関連資料・科学古典を収集整理し、それを通して当時の科学と科学思想を、東アジア文明総体の中に位置付けるところにある。具体的には分析の時代を18から19世紀に限定して、中国明清期の漢訳西洋科学書と伝統の暦算学書と医学書、朝鮮時代の天文数学書と医学書を読解・比較分析し、東アジアの科学の流れとその総合的な特徴を考察する。方法論的には、総合的視点をもって東アジア科学の全体を分析するところが新しく、目論んでいることといえば、現在、世界的に見て研究が手薄である朝鮮科学の正当な評価にほかならない。

今回、朝鮮18、19世紀に研究の重点を移動する予定であるが、それは11、12年度の研究を通して、朝鮮時代の漢籍や韓国 OS 用の CD ROM 収集にもめどが立ち、17世紀朝鮮数学や19世紀中国の西学輸入について斬新な研究成果をえることができたからである。当初の予想を越えて朝鮮文献の収集がうまくいっているのは、韓国研究者の援助と、研究分担者が韓国学界の事情に精通していることに負うところが大きい。

#### 研究計画・方法

研究代表者と研究分担者の間の研究分担に関する基本原則(基本合意)は、研究代表者が東アジア科学古典の学説の分析を中心とするのにたいし、研究分担者は外的アプローチを試み、思想や文化との関係に注目して中国と朝鮮の科学を分析するところにある。

13年度は11、12年度の研究成果を踏まえて、大きく

二つのことを追求する。一つは東アジア、主に朝鮮の科学関連基本資料を収集することである。日本には朝鮮科学史関連図書の数が絶対的に不足しており、後世の人文研究に支障が生じないためにも、その収集整理は必要不可欠といわねばならない。他の一つは文献解読であり、後半期については18～19世紀の朝鮮科学と科学思想の文献の分析に時間と精力を集中する。具体的には、研究代表者は17世紀朝鮮数学の分析につづいて18～19世紀の東算書の読解を試み、研究分担者は19世紀中国の科学思想の分析をつげながら、朝鮮19世紀の西学輸入の問題に新たな展開を試みる。また本年度以降については、訪韓にくわえて11、12年度にあまりできなかった訪中を実施し、中国科学院自然科学史研究所の研究者と学術交流を果たしたいと思っている。

### A04-31・計画研究

#### 中国における制度と古典 科举制度と言語史・文学史の相関から

研究代表者 平田 昌司  
京都大学文学研究科 教授  
研究分担者 田口 一郎  
新潟大学人文学部 助教授

#### 研究目的

中国の言語史(ないし言語規範意識の歴史)は、文学史・制度史を始めとするさまざまな領域に大きな関わりを持ちつづけている。

問題の所在を最も顕著に示すのは、官僚有資格者選抜試験としての科举にほかならない。その出題方法の変化、試験規則の改定の裏には、往々にして言語の史的变化や方言対立(地域間対立)がひそんでいる。科举史と言語史とを結びつけた研究をおこなうことにより、中国において権威的な地位を付与された言語が、いかなる機能を果たしてきたか、6～20世紀という長期間にわたる観察が可能になる。

11～12年度の研究においては、対象を主に11世紀までとした。13～14年度は、重点を南宋・金・元・明代(12～16世紀)に移し、(1)11世紀までにほぼ完成をみていた科举制度の言語規範は、非漢民族(遼・金・元)のもとでどのように運用されたか、(2)皇帝が非漢人から漢人へとかわった元・明の交替期に、モンゴルの残した言語規範意識はどのように継承されたか、