漢字と情報

No. 8
2004・3

京都大学人文科学研究所　Documentation and Information Center for Chinese Studies（DICCS）
附属漢字情報研究センター　Institute for Research in Humanities, Kyoto University

- 江戸の殿様と北京の薬局
- N-gram による先秦文献の分類
- 人文研のアーカイブス(8)「大般若波羅蜜多經」

江戸の殿様と北京の薬局
―鎖国を越えた博物学趣味―

高津　孝

今から220年前，乾隆49年（1784），北京の前門（チエンメン）にある清朝御用達の薬種商同仁堂に二人の琉球人の若者が訪ねてきた。二人の若者，紅之誠，金文和は大型の彩色植物図50図を携えてきており，描かれた植物の鑑定を依頼してきたのである。植物図には，それぞれ標本が付され，簡単な生態についての説明が付いており，根や実も別の包みに収めてあった。同仁堂側で対応に当たったのは，周之良，鄧履仁，呉美山の三人であったが，見知らぬ植物も多く，回答しえたものはわずかで，中には薬とはならないものもあった。実はこの彩色植物図は薩摩藩の吉野植物園で作られたもので，江戸にいた藩主島津重豪（1745–1833）の命により，はるばる琉球を通じて，海路，福州に至り，陸路を辿って北京に運ばれ，清朝御用達の薬種商同仁堂に鑑定を依頼することになったものである。これは，日本本草学の抱えていた本質的な問題点，中国産の薬材を基本として成立している本草学の書物を異国である日本はどのように受容し理解すべきかという問題の解決を目指してのものである。この江戸の鎖国体制を突き破って行われた調査は，集大成され『質問本草』として残されている。天明５年（1785）島津重豪の時代に彩色写本として完成し，天保８年（1837）曾孫の島津斉彬（1809～1858）の時代に木版で出版された。

第25代薩摩藩主島津重豪は，江戸後期，薩摩の文化にとって最も重要な人物である。娘茂姫が将軍徳川家斉の御台所となり，将軍家の岳父という地位にもあり，島津重豪の時代，薩摩藩の社会的地位は急激に上昇する。薩摩藩の文化的事業の殆どはこの時代に集中し，それを推進したのは藩主の島津重豪であった。かれは蘭癖もあり，シーボルトとも会見している。また，博物学にとくに興味を抱き，多くの書物を編纂させた。一方，後年，薩摩藩が苦しむことになる500万両の藩債も彼の時代に遠因すると言われ，評価の難しい藩主である。

薩摩藩は，すでに万治２年（1659）山川に，貞享４年（1687）佐多に薬園を設置していたが，安永８年（1779）に鹿児島の北郊吉野台地に吉野薬園を新たに設け，薩摩藩内に見られる多くの薬草を集め栽培することにした。この吉野薬園で，天明元年（1781）から６年（1786）にかけて，精密な彩色植物図譜６冊が作成され，琉球を通じてはるばる福州，北京へと運ばれ調査が行われたのである。この彩色図譜は，その副本と推定されるものが，沖縄県立図書館に現存する。縦横 40×30 センチの大型の図譜３冊で，彩色植物図256図を含み，内容は天明２年から５年に作成された４冊

沖縄県立図書館本『本草質問』第３冊

鹿児島大学附属図書館所蔵
天明写本『質問本草』内篇巻二

分に相当する。美しく精密に描かれた植物図で，明・李時珍『本草綱目』や日本・貝原益軒『大和本草』の略画的な植物図に較べると，ボタニカル・アートとしてはるかに水準は高い。

中国での調査自体も極めて周到な準備のもとに行われている。植物図の横には標本が張り付けられ，植物名は記さないが，産地や開花時期などの情報を記載し，また，根や実は別に添付し，それでも不足する場合は，鉢植えを準備した。鑑定を依頼した相手は中国に渡った琉球人達のつてを辿った福州の医者，学者や北京の薬種商である。鑑定者の水準を試すための準備もされていた。真偽の明白に判明している柴胡・辛夷・楝・女貞などを加えたのは，その用意である。標準的薬材すら判別できない人物はこれによって排除される。相手と直接に対面して質問し得ない状況を前提として，実に色々な工夫が凝らされているのである。さらには，鑑定が一方に偏することを配慮して，長崎にやってくる清朝の商人，また薩摩に漂着する中国人にも鑑定を依頼しており，その結果は現行の『質問本草』に反映している。

現在残されている刊本『質問本草』は，その著者を琉球・呉継志とする。その巻頭には，呉継志の名前での質問状と鑑定に当たった多くの中国人医師や学者からの回答の手紙が添付されている。鑑定者達は，質問者が琉球人であること，鑑定の対象となった植物が琉球の産物であることを疑ってはいない。また，日本においても，天明５年写本の存在が明らかになるまでは，『質問本草』は琉球人の著作として知られてきた。架空の人物琉球人呉継志を設定することは，薩摩藩の支配下にありながら，独立国として清朝と朝貢貿易を行っている近世琉球の現状を隠蔽するために必要な措置であった。『質問本草』は，鎖国体制にある江戸時代において，琉球を介して中国の福州，北京まで出向き薬草鑑定調査を行ったという特異な成果としての意味ばかりでなく，その存在自体に複雑な国際関係を反映した博物学著作なのである。

（鹿児島大学法文学部教授）

N-gram による先秦文献の分類

山田崇仁

先秦を対象とした歴史研究は今日続々と発見される考古資料に注目が集まりがちだが，決して既存の文献の重要性が低下したわけではない。その中でも特に春秋戦国期の歴史イメージは諸子百家の残した文献に大きく規定されていると言えるだろう。これら諸子百家の文献は誰か一人の手によるものではなく，戦国から漢代にかけての長期間にわたって重層的に成立してきた学派のアンソロジー的な書物であるが，個別の文献のみならず各篇の成書年代となると各研究者の間で説が分かれ，現在でも諸子百家の思想や学派の歴史的展開を描く事の難しさの要因となっている。

従来の先秦文献の成書に関しては，目録学・思想・言語学などの手法を駆使して研究が行われてきた。この内目録学的な手法は，いわば文献を外からの材料によって分析する手法と言える。これに対し思想や言語学的な手法は，文献の文章そのものを問題解明の手がかりとする中からの分析手法と言う事が出来る。これらの方法は現在までに多くの業績を挙げてきているものの，それぞれの方法論の違い故に文献の成書について矛盾をきたす事が少なくない。そのため，諸子百家の歴史的展開についても議論が分かれる事になってしまい，未だ多くの文献で研究者が共有する年代観が確立されているとは言い難いのである。

筆者はこの問題に対し，従来の研究成果を尊重しつつも，それとは少し異なるアプローチを試みる事にした。それがここで述べる N-gram を利用した統計解析的な手法である。

漢字文献と N-gram

まず，N-gram について簡単に説明しておこう。N-gram は情報理論の創始者として名高いクロード・エルウッド・シャノン（Claude Elwood Shannon）が提唱した確立統計的な言語処理の手法である。N-gram 自体は「あるテキストの総体を前から順に任意の N 個の文字列または単語の組み合わせで分割したもの」を意味する。例えば，『論語』の冒頭「子曰，學而時習之，不亦説乎。」を対象に二文字単位で分割すると，「子曰」「曰學」「學而」「而時」「時習」「習之」「之不」「不亦」「亦説」「説乎」となる。N-gram では「子曰」「學而」のような個々の文字列または単語の組み合わせを「共起関係」と呼び，N 個の数（gram）に応じてそれぞれ「1-gram, 2-gram, 3-gram …」と呼ばれる（上記の場合は 2-gram）。またテキスト全体での任意の「共起関係」を集計した結果は「共起頻度」と呼ばれている。

漢字文献に対し N-gram を利用した研究については，石井公成氏や師茂樹氏を中心に行われてきた。石井は主に複数版本の系統調査の手段として N-gram を利用した NGSM（N-Gram based System for Multiple document comparison and analysis）を提示され，師氏は N-gram の共起頻度生成プログラムである morogram を開発するのみならず，N-gram を用いた研究環境・手法の発展を目指しメーリングリストを主催するなど活発に活動されている[1]。

では，実際に漢字文献を対象に N-gram を利用して研究を行う場合，どの様なメリット・デメリットがあるのだろうか。

N-gram では，対象となる文字列を一次元の遡上に載せて先頭から任意の N-gram 単位で切り分ける方法上，テキストが内包する様々な情報から「枇字列の並び」にだけ着目した手法と言える。そのため，漢字文献で多用される対句や一字違いの用例の比較には不向きであり，また gram 数を長くすればするほど「ノイズ」と呼ばれる無意味な共起の大量発生が避けられない。更に NGSM を利用して文献の先後を判断する場合，0 頻度問題（共起頻度が 0 であった場合に，その共起が「存在しない」のか「たまたま含まれなかっただけ」なのかという問題）に対する判断が重要とな

る。

以上の様に，N-gram は一定の方法的限界を抱えている事は確かである。しかし用例（共起）を機械的且つ網羅的に収集するという N-gram 的手法は，文献の中から特定の用字パターンを抽出するのに大変有効な方法である。従来ならば特定の用字に絞って文献を渉猟するという手法が一般的であったが，N-gram 的手法では初めに全ての組み合わせ例を収集するので，あらかじめ指定したキーワードを抽出するのも効率よく済むし，従来ならば注目されてこなかった情報をも拾える可能性をも秘めているのである。

N-gram とクラスター分析

―『韓非子』を事例として―

筆者もこれまで N-gram を利用した研究を幾つか発表しているが[2]，これらは N-gram の網羅的集計結果から有為な用例を人間が抽出して分析を行うタイプの研究であった。本稿では『韓非子』を事例に，N-gram の共起頻度データ全体を統計学的な手法で分析する方法について紹介してみたい。

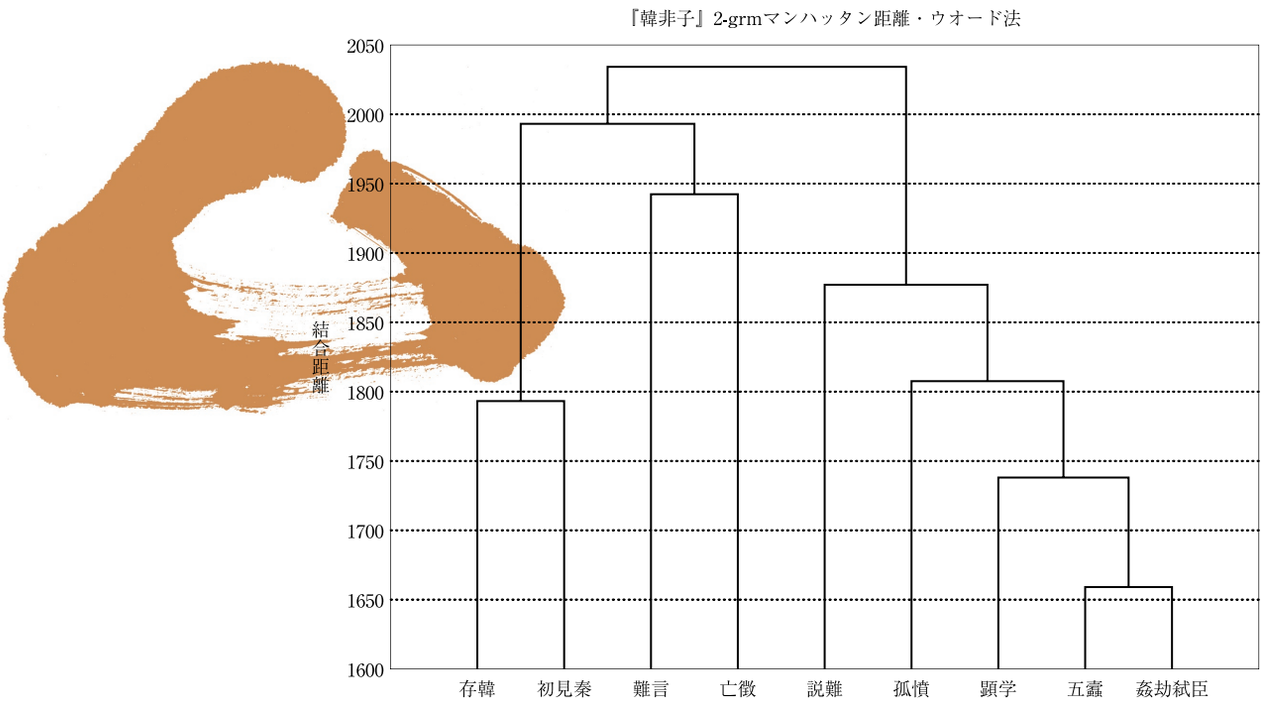
ご存じの通り『韓非子』は，韓非を開祖とする法家（韓非学派）の著作集だが，どの篇が韓非の自作であるかについては，『史記』以来様々な説が提示されてきた。今回は，従来の研究で韓非の自著もしくはそれに近いとされる諸篇と，韓非後学の著作とされる諸篇とを何篇か選び，moro-gram を用いて2-gram・頻度１以上の共起頻度を集計し[3]，多変量解析の分野で利用される統計的手法であるクラスター分析を行って各篇がどのようなグループに分かれるかを調べてみた。

クラスター分析とは，各項目の距離と結合方式（何れも複数の種類がある）にもとづいて対象をグループ分けし，何らかの方法でそれを図示する方法である。まず，各篇の長さに起因する共起頻度総数のバラツキをならすために一定の正規化を行ったものを使用した。距離はノイズデータや極端に値が偏っている変数の影響を最小限に抑える事を考慮に入れてマンハッタン距離（市街地距離）を[4]，結合方式には一般的に有効性が高いとされるウオード法を採用し[5]，最後に各クラスターの結合結果を樹状図（デンドログラム）にした。

結合距離1950で各クラスターを分割すると，所謂韓非の著書と目される五篇とそれ以外の四篇との大きく二つのクラスターに分かれる。この結果は従来の『韓非子』諸篇の分類に関する学説とも矛盾しないため，N-gram とクラスター分析とい

共起	合計	姦劫弑臣	亡徴	初見秦	孤憤	五蠹	顯學	難言	存韓	説難
以為	69	7	3	4	0	7	24	13	2	9
天下	68	10	1	24	0	15	4	0	14	0
之所	64	23	0	0	4	15	14	1	3	4
不可	64	13	0	1	7	9	24	0	6	4
而不	62	9	11	5	3	9	12	6	3	4
人主	58	14	4	0	19	11	8	0	0	2
所以	53	15	0	1	4	5	16	2	7	3
亡也	52	0	48	1	3	0	0	0	0	0
可亡	48	0	48	0	0	0	0	0	0	0
者可	47	0	47	0	0	0	0	0	0	0

共起頻度集計上位10位まで



『韓非子』を対象とした2–gram の樹状図

う手法は，基本的には有効だと言えるだろう。

但し，現在この手法を用いた文献の分析はまだ試行錯誤の段階である。樹状図を描く前の段階である共起頻度の正規化の手法・ノイズの排除・クラスター手法（距離・結合方法）等，日々問題の洗い出しと検討を行っている最中であり，今後の研究の進展によって本稿で試みた方法以外のものを採用する可能性もある。

また，仮にある複数の文献（篇）が同一・若しくは近いクラスターに属したからと言って，直ちにそれらが近縁関係にあると即断するのは危険だろう。確かにこの方法だと，同じ文字列の組み合わせを同程度使用しているものほど近くなるのは確かだが，一方の文献が他方の模倣であった場合でも共起頻度の関係から同一のクラスターになる可能性もある。そのため，同一クラスターに属する篇同士の関係については，他の統計手法を試みたり個別の文献の中身に立ち返って精査したりする必要があろう。

しかし，N–gram ＋クラスターによる分類で同一のクラスターに属した場合，少なくとも双方を比較検討する価値はあると言える。そこで両者の関係が明らになり，それを積み重ねる事によってより広範囲の文献の相互関係が明らかとなるだろう。ただ漠然と文献を比較するのではなく，一種の機械的フィルタリングとしてこの方法を利用する価値は認められても良いのではないだろうか。

（独立行政法人日本学術振興会特別研究員）

1　morogram に関しては，以下の Web サイトを参照。http://sourceforge.jp/projects/morogram/

2　「『國語』韋昭注引系譜資料について― N–gram 統計解析法による分析―」『立命館史学』22号　2001「歴史記録としての『春秋』―N–gram モデルと統計解析法による分析―」『中国古代史論叢』立命館大学東洋史学会中国古代史論叢編集委員会編　2004

3　テキストデータは，台湾の中央研究院漢籍電子文献のものを使用した。外字などは特に訂正せずそのまま利用したが，句読点や括弧類などは全て削除してある。

4　それぞれの変数の差の絶対値を求め，それを合計したものを 2 点間の距離とする方法。

5　クラスター内の平方和を最も小さくするという基準でクラスターを形成する。最小分散法とも呼ばれる。

人文研のアーカイブス（8）

大般若波羅蜜多經　巻第四百十七

唐釋玄奘譯

金刊本

梶浦　晉

この『大般若經』は，金代に刊行された大蔵経の零本である。刊本の漢訳大蔵経は，北宋官版《開宝蔵》以来，清の官版《龍蔵》まで官版・私版十数蔵が確認されている。金代に大蔵経が雕造されたことは，ながく知られていなかったが，民国22年（1933）に山西省趙城県広勝寺でほぼ一蔵が発見された。金版大蔵経が《趙城蔵》あるいは《広勝寺蔵》などと呼ばれるのは発見の由来によるものである。しかしながら，発見されたのは広勝寺であるが，記録や刊記等によると皇統年間から大定年間にかけて山西省南部の解州天寧寺で雕造され，のち中都（北京）に板木がうつされ，元代にいたるまで補刻を行いながら印刷されていたことが判明しており，《天寧寺版大蔵経》あるいは単に《金版大蔵経》と称するのが適切であろう。現存のものは，元代の補刻を経たものである。

広勝寺本のほかには，チベットのサキャ北寺で発見された555巻（もと元大都の大宝集寺にあったもの）以外には，まとまった違例は確認されていない。広勝寺本は，抗日戦争時に，日本の掠奪をふせぐため，八路軍が寺から運びだし大行山脈中に疎開をさせ，戦後に北京図書館（現中国国家図書館）にうつされ，現在も該館に収蔵されている。

近年，『中華大蔵経（漢文部分）』の主たる底本として影印され，その全貌を容易にうかがえるようになった。なお，『中華大蔵経』では，広勝寺本が欠けている箇所で大宝集寺本がある場合は，これを底本としている。

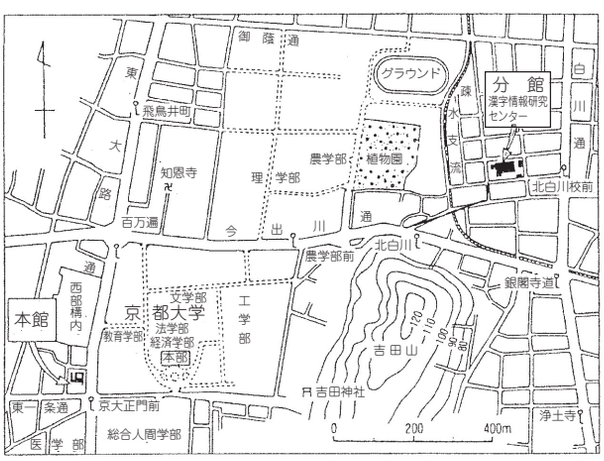
広勝寺で発見された後，民間に流出したものも若干あり，現在，日本では約40巻の所蔵が確認されており，本所にはこの『大般若經』等十巻を所蔵している。

掲出本は，松本文三郎氏より寄贈されたものであるが，印記により，かつて中国の著名な蔵書家であった周叔弢氏のもとにあったことが知られる。この巻第四百十七は，零本とはいえ，『中華大蔵経』では大宝集寺本が影印されており，広勝寺本と大宝集寺本とが比較できる点において貴重なものである。

（センター助手）

HP・TOPICS

今回は，人文研の「所員のホームページ」にある麥谷邦夫教授のＨＰのコンテンツを紹介することにします。道氣社という別称で名高い老舗のサイトなのですが，お勧めは何と言っても道教関係資料の「漢籍テキスト・データベース檢索」です。「電子版漢籍文庫」もそうですが，テキスト校勘がきちんとなされていて，信頼度は抜群です。「中國思想研究者のためのインターネット資源簡介」は役に立つコメント付きリンク集で，「イメージ・ギャラリー」には，ご自身で撮影された道教の名山のお宝写真があります。



【DICCS NEWS】

・2004年度の漢籍担当職員講習会は，初級は10月４日（月）～10月８日（金）に，中級は11月８日（月）～11月12日（金）に予定している。

　また，それらに加えて新たに「TOKYO 漢籍 SEMINAR」の開催を文部科学省との協議によって決定した。このセミナーは，場所を東京に移して，広く社会一般の人々に漢籍に対する関心と理解を深めることを目的として，講演会形式で行おうとするものである。開催時期は，毎年３月中旬を予定している。

・2004年３月12日（金）14時より全国漢籍データベース協議会第４回総会を学術総合センターで開催した。本年度は，中国国家図書館副館長の陳力先生，中国国家図書館分館副館長の孫学雷先生を招き，陳力氏の「中国古籍デジタル化の現状と展望」と題する講演を行ったほか，関西大学学術資料課の鵜飼香織氏が参加機関からの報告として関西大学の場合の現状報告を行った。

・従来「東洋学文献類目」の編纂には学術情報メディアセンターの汎用機を用いていたが，2001年度版から守岡知彦助手が開発した本センター内のシステムに移行し，最新のデータをすぐに WWW 上で公開できるようにした。冊子体の主な変更点は，版面をＢ５版からＡ４版にしたこと，漢字字体を原書のままにしたこと（従来は繁体字に正規化）などである。

・最新のセンター刊行物

「東洋学文献類目」2001年度（2004年３月）

「漢籍目録を読む」（東方学資料叢刊第12冊，井波陵一編，2004年３月）

発行日　2004年３月20日

発行所　京都大学人文科学研究所附属
漢字情報研究センター

〒606-8265 京都市左京区北白川東小倉町47

電話 075-753-6997　FAX 075-753-6999

http://www.kanji.zinbun.kyoto-u.ac.jp/