

オントロジとは何か

—和漢古典学のオントロジ—

What is ontology?

— Ontology of the classic study about Japan and China —

相田 満

1.オントロジ概論—情報学的見地から—

1.1.はじめに—情報学と文学研究—

コンピュータが人間の生活や文化の「社会基盤」として不可欠な存在となった今、蓄積される「情報資源」とその運用・処理のための研究、すなわち「情報学」は、物質の従来の学問分野には見られないほどの実用性・具現性をともなって発展している。

すでに個人がコンピュータの世界全てに通暁することが不可能であるのは、ウィーナーが、あらゆるものは「物質とエネルギーと情報からなる」とまで喝破したような、「情報学」自体の持つ普遍的かつ総合的特性に由来する。急速に発展してきた当該学問は、既存諸科学との整合性を未だ十分に見いだせぬままにも、あらゆる事象・分野にわたって根底的な変革を与えるほどの影響を及ぼし続けている。

文学研究という歴史の古い学問領域でも、コンピュータとの関わりは、無視し得ない、大きな変化として意識されつつある。研究の質の変化という観点からは、その最たるものは、扱うデータ量の巨大化という点においてであろう。典拠研究に際して、コンピュータを使用した「検索」を行うことは、すでに当然の手續きとなっているのである。

その結果、これまでの研究で使用されたデータ量とは桁違いに膨大な情報を駆使するだけでは、その独創性・新規性を評価することもできなくなりつつある。さまざまな情報をいかに解釈し、組み立て、選別するか、そこに研究者の技量が問われる時代となろうとしている。

1.2.情報過多の時代

情報過多の状況を打開する切り札的存在であった電子情報化技術も、逆に情報の増殖と多様化を助長せしめる結果をもたらしつつある。「いかにテキストを作成・構築するか」というこれまでのパラダイムはサイバー空間における環境問題へと変化し、「すでにある情報からいかに有用な情報を汲み取るか」という段階へと移行してきたということは、衆目の一致する所である。

博学の度合いや情報価値の正当性を問う時、「辞書にない」「教科書にない」という言葉がよく使われてきた。しかし、新たに生み出される情報群の大半が電子化情報へとシフトしつつある状況では、ネットを通してアクセス可能なサイバースペースにおける情報の存在も無視できなくなっている。むしろ日常生活レベルにおける資源・評価的価値の比重は逆転の方向に向かおうとしている。

ところが、Web の検索エンジンを使用しての検索では、あふれかえる情報をいかに組織化し、有用な情報を得て、効果的な運用をはかるかという問題の方が重要な問題となりつつある。しかし、増殖の一途をたどる情報群に対して、そのな運用のための基盤資源の不足のために、十分な方途を見いだせないままなのである。

その根底には、人間の認識系の変容に対処すべき分類のあり方という大問題が横たわっている。多様化するマルチメディアコンテンツと、それらをつなぐハイパertext情報を統御すべき新たなパラダイムの導入は、氾濫する情報群に対する緊急課題となっているのである。

1.3. 情報「資源のための資源」の必要性

情報の統御と組織化を行うためには、情報を蓄積する側が自ら情報の組織化を行う場合と、すでにある情報資源から情報を引き出す際に、何らかの組織化を施した検索行為を行う場合の両方が考えられるが、実際には、双方を融合させたシステムが実装されることが望ましい。

いずれの手法にも、バックボーンに、シソーラスに代表される「統制語」という概念に沿って構築された言語資源の存在が不可欠である。つまるところ、効果的な機能を実現させるための「技術」は、充実した「分類用情報資源」による下支えがなければ、有効な機能は期待できない。

1.4. 統制語の限界

統制語とは、同義語、類義語、慣用語などの類を一つだけそのカテゴリーにまとめたものである。

1851年にイギリスのロジェ(Roget)が刊行した事典名に由来するシソーラスは、この考えを進めることにより、コンピュータを使用した様々な検索システムとともに発達し、現在、さまざまな分野で関係各領域の用語が調査・集積され、構造化と関係性を付加されつつ蓄積が進んでいる。

しかし、一般的には、処理技術の進展に対して、分類用の資源の蓄積と整備の進み方はきわめて遅い。

さらに困ることは、「分類のための情報資源」の構築には大変な労力がかかることである。そのためか、この種のデータが無料または廉価の機械可読式テキストとして公開されることは稀で、その数も多くない。

データにプロテクトをかけて、特定のシステムに依存して使用される仕様となっていたり、たとえ販売されてはいても、高額なロイヤリティが要求されることも少なくないのである。

このように、現在、「分類用資源」は、情報検索を支援するための重要な手続きである「情報資源の組織化」のために必須の資源であるにもかかわらず、いまだ十分な蓄積と流通の方途を得ぬままである。

困ったことに、代表語の認定には恣意性が避けられず、しかも必ずしもその効力の永続性が保証されるものではない。時には、専門家が長い時間を掛けて制定した語が存外に広まることなく、多少の意味のずれを包含しながらも一般的に言い習わされた俗語に駆逐されることも少なくない。

さらに、現代語においては、それらの歴史的蓄積を捨てて、一語一語に新たな概念を造語として付加することが繰り返されてきた。しかも、付加された概念を統一的な世界観、つまり一意の世界に収斂する分類体系に収斂させようとする努力が重ねられてきたのである。しかし、そうした試みも、一世代、およそ30年以上という長期の経年変化を重ねると、新たな古典的概念体系を一項目増やすに過ぎないという危険性もある。生々流転の激しい現代語を一つの枠に長期間閉じこめておくこと自体に無理があるのである。現代的手法を標榜して構築された概念体系も、いずれは古典世界と混じり合う宿命を持つと言っても過言ではあるまい。増大するデータに対し、その組織化を果たすという営みは、とても手間と時間がかかるというのに、何とも皮肉なことである。

そこで考えられたのが、物事を単線的で、一意に収斂する概念木に無理矢理統合しようとするをやめ、比較的明確な定義の集まりとなっている、「あるがまま」の知識セットを運用することによって、柔軟かつ精密な知識ナビゲーションを可能にしようというコンセプトである。それが"オントロジ"である。

1.5. オントロジ

こうした状況下、哲学用語「存在論」に由来する「オントロジ」(Ontology)というコンセプトが注目を浴びつつある。

「オントロジ」とは、ハイデggerらに代表される「存在論」とも訳されようが、現在は、さらにそのギリシア語の本義に回帰する方向に概念を拡張されつつもある。これは1980年代半ばの第五世代コンピュータ開発プロジェクトの頃に提唱され、近年再び脚光を浴びつつある概念である。端的に言えば、「すでに存在する知識体系を分類・集約し、その成果をコンピュータを利用することによって人間科学に反映させる」ための手法

と資源とでもまとめられ、具体的には「概念間の関係の明確な定義の集まり」として、それを実装する「トピック・マップ」とともに、情報リソースから独立した上位層^{メタレイア}に位置付けられ、情報を意味的に組織化、検索、ナビゲートするためのパラダイムとして理解されている。

この名称自体は、人文・社会学研究においては比較的早い段階から使用されていた。もっとも、その際には、「オントロジ」という用例の方が熟していたようだが、いずれも原義に則した「存在論」あるいは、事象の「底流」「(心理学的・社会的) 基盤」の意味で使用されるものである。

一方、情報学的では、オントロジは、用語モデルにおける知識構造的見地からの捉え直しが始まり、現在、用語の標準化による知識共有の基盤技術として、その重要性が再評価されつつある。その背景の一端には、知識という一般的命題を直接考察の対象にした知識工学が行き詰まり感を持ち始め、その根本資源である言語資源を再構築する必要に迫られたことがあった。

オントロジ的とらえ方では知識を構成する基本概念に立ち戻っての考察が行われることに特徴がある。すなわち、知識の階層性、コンテキスト依存性、知識の分解可能性などを注意深く考察することによって、物事や対象の成り立ちを基本から検討し、知識ベースが持つ様々な問題を回避することの糸口を与えることが期待されるのである。

そうしたコンセプトは、具体的には、情報検索に際して、それぞれの知識構造に沿ったシソーラス辞書群を基に、ユーザ個別の要求、すなわち"個"に特化した、最も目的に適った検索や知識の発見(データマイニング)を求めるという考え方への転換となって反映される。

オントロジモデルでは、さまざまなレベルの概念カテゴリー間の相互関係を示したものが使われる。よく使われるのは、「すべての概念の集合」を木の根に見立て、そこから特定の性質を持った概念の集合を代々枝分れさせていくもので、「概念木」とも「概念樹」ともよばれる。

これは、知識構造をモデル化したものともいえる。知識構造はカテゴリ単位に分類され、カテゴリには分野別に分類された概念階層の木(ツリー)が複数存在する。シソーラスにおける概念階層木は、ある概念を表す用語の上位・下位概念を表す用語へのリンクから構成され、各概念は概念名、同義語、意味を定義する述語論理によって定義される構造を基本としている。

こうした知識構造モデルの構築は、「人間」というものの認知構造の仕組みを「一般性」普遍的に捉えることを究極目的とするこの観点から、これまで「汎用性」ということを重視して進められてきた。

もっとも、「汎用性」「一般性」という定義自体、極めて相対的なもので、そもそも「日本語」という括り自体も我々からすれば大きな括りに見えるが、国際的には極めて局所的な存在とならざるを得ないものである。

しかし、我々が歴史的、文化的に培ってきた概念の豊かさは、そうした局所的な集合の中にも膨大に存在する。しかも、限られた分野における概念体系にこそ、情報爆発の被害を免れた、確度の高い情報が保全されているものである。その意味で、量に圧倒された傷を受けることなく構築が果たされた、分類体系という情報資源はきわめて貴重である。

1.6. 分類資源としてのオントロジ

情報学におけるオントロジは、すでに存在する知識体系の分類や集約を進めることにより、知識モデルの多様性を認める内容指向性研究をめざすことを特徴とする。

これを分類や語彙の実装法としてとらえた場合、関連する術語に、「分類」という行為に不可欠な語彙体系を意味するシソーラス(thesaurus)やあるいはそれに類する語句を戴く辞書群もあるが、オントロジ辞書と呼称される場合には、適切なサイズの領域内に蓄積された知識が積み上げられたものという側面と、知識を構成する基本概念への回帰ということが強調される。すなわち、限定分野の概念木を積み上げて構築することにより、形成情報爆発の被害を免れた、確度の高い情報が保証されることになるのである。そのため

か、[図 1]で示すように、著名な辞書に使用されるタームの命名由来を比較してみると、他のものが集積・言葉・日常など意味に込めているのに対して、「オントロジ（哲学用語「存在論」）」が、いかに特異な生い立ちをしているかがわかるだろう。

オントロジと関連する術語に「分類」という行為に不可欠な語彙体系を意味する、シソーラス(thesaurus)をはじめとして、辞書に使用された著名な語源や関連語として以下のような言葉がこれまでに使われており、中には情報学で頻用されているものもある。

「情報爆発」的現状を打破するために、巨大なシソーラスを整備しようとする試みは、歴史上幾度も繰り返されてきた。しかし、その成果は、取り込まれる情報量が多くなればなるほど、その分量に比例して、多くの矛盾・誤謬を内包する。シソーラス自体が巨大な情報量に耐えきれなく「情報爆発」を引き起こしてしまうのである。

そこで、これまで長い間、情報学の世界における潮流であった。グローバルな意味での”標準化”、さらには、“一意の世界”へと収斂しようとする”統合化”を指向する世界とは対極的に位置するパラダイムが生み出されるに至ったわけである。

統制を避け、あるがままの内容を尊重することによって、情報爆発と呼ばれる状態に対処する試みは、分類によって情報内容を強制的にカテゴリライズしようという方向とは逆に、既存の分類体系の組合せにより運用される。そのため、それぞれの知識体系の細部にわたっての精密な組織化へとベクトルが向かうことになった。「内容指向性研究」と言われるゆえんである¹⁾。

2.和漢古典学にとってのオントロジ—情報学的類書論の試み—

2.1.文学研究に「オントロジ」概念を導入する理由—「類」の観点から—

「分類」という行為、すなわちものごとを「類」という概念で整理・体系化することは、和漢の古典学世界では基本的な学問態度であった。独自の新しい学問世界を打ち出す以前に、博学多識を旨とする過去の知識の集成が行われることが重んじられ、そうした態度こそ、学問と考える意識の方が強い傾向を持っていたのである。

この点に関しては既に春秋末の孔子が、「述べて作らず、信じて古を好む」態度で以て、「故きを温ねて新しきを知る、以て師と為す可し」（『論語』為政篇）とまで言っているが、こうした態度は、古典世界のみならず、歴史学ほか現代にも通底することであろう。

- | |
|--|
| ○アーベーケーダリウム,Abecedarium (アルファベット順：初学者のためのアルファベット順英羅辞典) |
| ○アルウェアーリエ,Alvearie (ハチの巣箱) |
| ○カトリコン,Catholicon (万能薬) |
| ○オルトゥス,Ortus (庭) |
| ○メドゥッラ,Medull (髄) |
| ○グロッサリー,Grocery (用語集) |
| ○コーパス,Corpus (※文字記録の集成) |
| ○マニプルス,Manipulus (ひとつかみ) |
| ○シルウァ,Silva (森) |
| ○プロンプトリーウム,Promptorium (宝庫・倉庫) |
| ○テーサウルス (シソーラス),Thesaurus (宝庫・倉庫) |
| ○ヴォキャブラリー,Vocabulary (語彙集) |
| ○ヴァルガー,Vulgar (日常品・通俗) |
| ◆オントロジ,Ontology (哲学用語「存在論」) |

[図 1]辞書に関する関連語

しかし、自己の思索に基づく独自の創造的思考を提示するのが真の学問である、との明白な認識を表示した人物としては、宗教分野を別としては、中国では恐らくは清代考証学、日本では近世国学の台頭までを待たねばならぬかもしれない。

「類」するものを集積することによって、物事を分類・整理することは、特に中国人に好まれた手法である。日本も漢字文化の移入に伴い、その認識の枠組みを多く取り入れてきた。

その際、認識の枠組みとして、同質のものをまとめるという「類」の概念が利用され、「類」としてまとめられた概念群を他と弁別するために名称（ラベル）も付与され、その結果、分類する行為が物の体系化を促し、その対象は世界に及ぶこととなり、ついに世界は認識可能な枠組みで再構成されるに至るのである。

このことは、「オントロジ」というコンセプトが人文科学研究分野、とりわけ 2000 年近い歴史を有する典籍群をあつかう、古典研究分野との親近性がきわめて高いことを想起させる。

紙上の世界では、有史以来、日本や中国で幾度も編纂された、類概念（分類概念語彙）によってまとめられた古典的な辞書・辞典（＝類書）が、きわめて継承性の強い、良質な「オントロジ（知識概念木）」の宝庫となっていた。

というのは、古来より、情報処理技術は、類書・辞書・事典などの形で、紙上で具現化されてきた経緯があるからである。ところが、古典的な辞書・辞典（＝類書）は、古典学研究において枢要な意義をもつものとしての認識は十分にあったが、その量が膨大で、検証のための手続きも煩瑣であったため、これまで十分な評価もなされず、研究資料の提供さえも十分になされてこなかった。その意味で、本研究で扱う資料は、情報学という新たな衣をまとうことによって、はじめて、その真価を発揮し得るものといえよう。

2.2. 「類書」ということ

類書は一般に、事物や事象に関する記事を既存の書物の中から抜きだし、これをいくつかの部門や項目に分類する手順で編纂される。そのため、部門や項目立て、さらにはその配列順序、そして分類のための概念語に類書の特徴が現れ、記事の内容自体を編纂者自らが執筆することは稀である。

類書に記載される記事は、多数の書物から抜き出される。その抄出方法には、任意の記事を目的（分類項目）に沿って書き出す方法と、それぞれの書物を一書、あるいは必要な記事の集中する部分に対して、分類概念ごとに分解・再配置を行うという方法がある。後者の編纂方法は、大規模な編纂作業に見受けられる方法で、明代に成立した類書『永楽大典』（二万三千巻、残存八百八十巻）では、そうした編纂手法のおかげで、逆に亡逸した典籍の復元が可能となった（『四庫全書』永楽大典輯佚書）²。また、類書ではないが、日本では鎌倉時代の歴史書『吾妻鏡』³、『源氏物語』注釈書の『河海抄』⁴などでも、類似の編纂手法が確認できる。

2.3. 古典的「類書」観とは

「類書」は、ものごとを類概念で体系化した古典的百科事典である。中国の研究者はこれを現代的に訳して「古代的百科全書」と宛てることもあるが、現在一般に使用される「類書」の概念は、多分に近代的意味づけが付与されて使用されていることには留意しなくてはならない。

俗に三大類書と呼ばれるものがある。それに『芸文類聚』（唐・^{おうようじゅん}歐陽詢等撰）、『初学記』（唐・^{じよけん}徐堅等撰）、『太平御覧』（北宋・^{りぼう}李昉等撰）を宛てるのは、中国最大の書籍解題書である清朝の『四庫全書総目提要』（巻 135 子部類書類 1）によるが、現在でもこの三書がもっとも類書の典型的形態を得たものと考えられている。

しかし、かかる評価を得ているにもかかわらず、「類書」という名称が定着するのは存外に遅く、中国では宋代、日本ではさら下って江戸時代のものである。

最初の「類書」の称号を得たものに魏の文帝時代に劉^{りゅうしやう} 邵等が勅を奉じて撰した『皇覧』(もと1000余編)と言われるが、その評価とても『四庫全書総目提要』に依る所が大きい。

それでは、類書はどのように呼ばれていたかという、おおむね雑書・俗書・雑俗書等に扱われ、図書分類上でその位置を占めるようになったのは、『隋書』経籍志の「雑家」、『旧唐書』経籍志の「類事」を経て、『崇文総目』(歐陽脩)子部や『新唐書』芸文志の「類書類」まで待たねばならない。したがって、中国において「類書」なる名称が一般化するの、北宋の仁宗の康定～嘉祐年間(1040～1060)と想像される。

一方、先に日本における「類書」の名称の定着を、江戸期と述べたものの、厳密に言えばやや曖昧とした点も残る。その理由の一つに、中国における「類書」概念の定着を見ていることと、日本的な「類書」観が現れる言説も散見されるからである。

たとえば、二西洞(前田菊叢)による元禄12年(1699)版『唐本類書考』や、向栄堂(山田三郎兵衛)の寛延四年(1751)版『唐本類書考』で扱われる書目には、『漢魏叢書』や『康熙字典』のような叢書・辞書の類の方がかえって多い。また、宝暦3年(1753)から安永9年(1780)にわたって書き継がれた山岡浚明『類聚名物考』では「類書」は、「類叢」という見出しが立てられ、そこでは、

「されども皇覧の如きは一事にもあらず。類ひをもて小部なる物を寄集めし叢書の初なるへし。され共よつて来る所を是をや初とすへからん。皇朝にては類聚国史・新撰字鏡のときを始とやいふへからんかし。」

と、類書の嚆矢とされる『皇覧』を「叢書」の初めとし、日本では『類聚国史』『新撰字鏡』をその濫觴に据えている次第である。

さらに同じ項目には、

万安方・姓氏録・蕉聚類集・藻塩草・呉竹集・春雨錦・兼名苑・名目抄・二人凡秘抄・十語五草・宗般聞書・堯孝法印日記・裁縫秘抄・庖丁譜・厨事類記

などが挙げられ、現代的観点からすれば、注釈書・日記、作文手引き書、辞書、入門書などに扱われる書名が「類叢」の書として挙がっていることが注目される。

こうした認識は、現在の考え方からすれば意外の観を以て受け止められがちだが、そもそも多量の情報資源を「聚^{あつ}めて分「類」する行為自体を重視する立場に立てば、様々な形式でまとめられる人間の思考活動の本義に適った正当な見解と言える。そして、和漢古典学におけるオントロジモデルを構築する際の対象も、こうした知識集約型の典籍を俎上に挙げてしかるべきである。

2.4.近代的「類書」観とその解体

かかる段階にまで考察を進めて来ると、「類書」という概念に代わる言葉を持ってないことに気づかされる。

そもそも、近年の研究で「類書」なるものが注目されたのは、それが一つの「工具書」として、古来より様々な方面で利用されてきたことが文学・史学の研究過程で明らかにされてきたことによる。とりわけ、通覧的機能の期待されない簡便な典籍類が、上代・平安より文学的営みと密接に関わってきたことが明らかにされたことは、文学研究における「出典」と「典拠」というものの概念を変える大きな出来事として、専門分野により10年以上の時差をとめないながらも、その後の研究方法を大幅に変えた。

しかし、そうした典拠たるものの存在自体の研究と、資料整備や解析、さらには基盤資源の提示は遅々として進まず、十年一日どころか、百年一日とも言ってよい状況にあり、それどころか、かかる状況で再認識された「類書」観が、先に述べた三大類書こそが「類書」であるといった、規範的な「類書」モデルを生み、「類聚」という観点で集められた多様な編纂物を、「分類」という観点から、同じ地平上に据えて分類する作業の障害ともなってきたのである。

したがって、我々はこうした「分類」という観点から古典学を見つめ直すためにも、多義的な「類書」という言葉、あるいは「辞書」「百科事典」などと言った、まとめあがった形式にとらわれない、人間の

知的行為や認識にのっとり、新たなパラダイムを模索しなければならないだろう。本発表で使用した「オントロジ」という概念を採用したのは、そうしたこれまでの桎梏から抜け出せる一つの可能性が孕まれていると感じるからである。そして、「オントロジ」の取材源となる、分類概念で内容が整理された典籍類を「類聚編産物」と呼称しておく。

2.5.古典的オントロジの特徴

「類書」「辞書」「字典」「百科全書」、そして用語における「オントロジ」のいずれもが、「世界の認識と記述」を志向するとは、よく言われることである。

記述の対象は、自然界・人間界の全領域（思想・制度・行為）に及び、たとえば、「類書」では、「天象」「人事」「神仏」「山地」「禽獣」「草木」……などといった配列ともに階層化が図られる。

このように限定された世界観（知識セット＝オントロジ）を設定し、それを細分化した上で事物を分化する分類・配列手法は、もっとも歴史の古いものである。この手法では、関連する事柄をそれぞれのカテゴリにまとめるため、検索者は一部分に着目するだけで必要な知識と関連する情報とを過不足なく入手できるという利点がある。しかし、欠点としては、情報が大量になるにつれて、その分類作業自体が非常に困難になってくるということがあった。

そのため、そうした欠点を克服し、大規模な情報を統御するために、配列基準を有限個に設定し、その原則で情報を統御しようとする試みも誕生した。統御規則として代表的なものには、「部首」・「韻（漢字音・五十音）」・「アルファベット」などである。あるいは「いろは」「千字文」のように、独自に設定され、親しまれてきた詩歌の詞章自体が配列規則に援用されることもあった。

「円環知⁶」などという名称も使われる、こうした配列原則に特徴的なことは、その配列規則自体が世界観・世界哲学を象徴する意味づけがなされることが少なくないという点である。

たとえば、『説文解字』は「部首法」による配列規則で統御されるが、現代の漢字・漢和辞典における『康熙字典』準拠のものとは少しく異なる。すなわち、『説文解字』の部首配列は宇宙の始まりである「一」に始まり「亥」に終わり、さらに「亥にして子を生子、復た一より起つ」（「亥」）という、あたかも時間の循環をなす十二支を連想させるかのような循環をなす世界観で統御されている。あるいは、七五調の和讃形式の歌謡の「いろは歌」は、この世の哲理である「諸行無常」を象徴する『涅槃経』の偈を背後に背負っている。

前者の配列方法に比べ、後者のように、有限個の配列規則を設定する方法の登場は、歴史的には、はるかに遅れる。それは「文字文化」というものの浸透に由来するためである。そもそも、後者の方法による検索手法を有効活用するためには、有限個の配列規則に習熟するとともに、その規則に対する理解が一般化する必要がある。しかも、立項されるべき見出し自体が、どのような名称を持つものなのか、あるいはどのように読まれるべきものなのかといった、曖昧な事柄を検索する際には、かかる配列規則は全く機能しないからである。

もっとも、後者のような有限個の配列規則により並べられた方法が採択されたにも関わらず、その欠点を多少なりとも克服するために、配列された見出し語の属性に、前者に見えるような階層的知識体系の樹形図が埋め込まれているものも少なくない。

「目次」と「索引」との関係にもたえられるであろう前者の世界観提示型（オントロジ設定型・カテゴリ分類型）と、後者の配列基準設定型による配列方法と双方の利点と欠点の葛藤は、現代のコンピュータ世界においても継続する案件となっており、省力化と効果的な検索系の実現との間で互いにせめぎ合う切実な案件ともなっているのである。

3.和漢古典学のオントロジの収集と分析

3.1.和漢古典学オントロジ資源の採取対象と方法

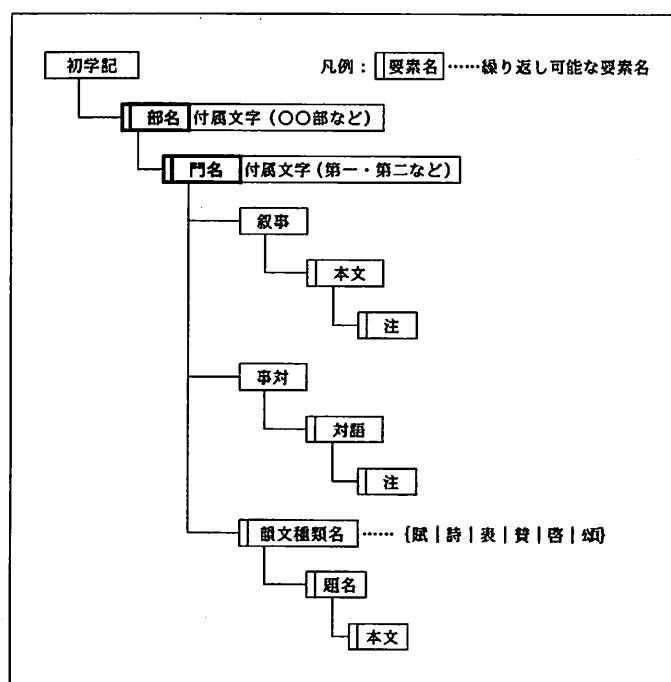
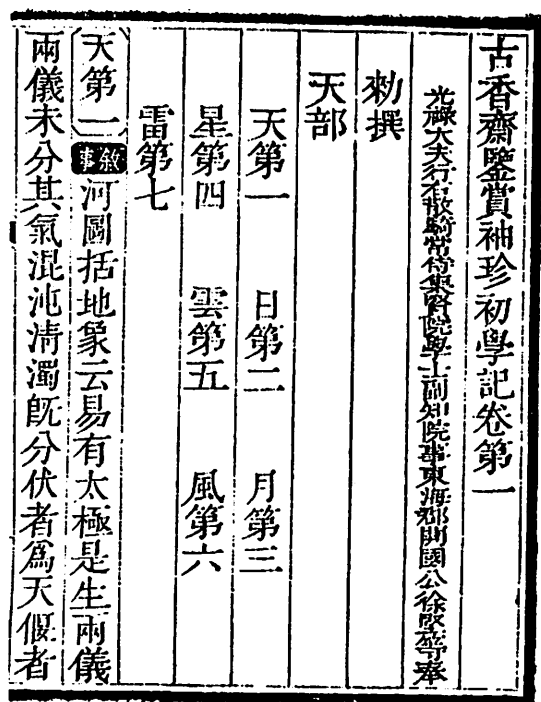
現在、稿者は、それぞれの担当者の関心の基に、類聚編纂物の分類用見出し語彙のデータを、上下の階層関係を保全したままの形で、協同で入力作業（一部収集もあり）を進めている。

現時点(2004.2)での入力済データは以下の通り。

古事類苑・北堂書鈔・事類賦・初学記・芸文類聚・李嶠百詠・万葉集・書言故事・百科全書（ディドロ、ダランベール編）・不忍文庫改正目録・四庫全書・日本館訳語・和漢書籍目録・増補書籍目録作者付大意・増補書籍目録・古今書籍目録・改正広益書籍目録・広益書籍目録・現代短歌辞典・王朝語辞典・醒睡抄・羅山文集・鷲峰文集・藻塩草・伊京集・慶長五年本節用集・元和三年板下学集・書言俗解・書言故事・俳諧類船集・和漢朗詠集・日本十進分類表（新訂第八版）・漢書芸文志（七略）（七録）・隋書經籍志・舊唐書經籍志（古今書録）
 ・新唐書芸文志・崇文總目・郡齋讀書志・直齋書録解題・文献通考經籍考・和漢朗詠集・日本分類語彙表（旧版）
 ・職原抄・百寮訓要抄・淵鑑類函・和漢朗詠集・内閣文庫和古書目録・和漢三才函会・類聚国史〔同原態推定版〕〔順不同〕

ここで採取対象とする和漢の古典籍におけるオントロジの実体は、「類書」「辞書」などの類聚編纂物中に使用される、「見出し」「部立」に使用される語句から、「〇〇部」「第一〇〇」といった付属語を取り去り、そこから「分類用概念語彙」を抽出して、「あるがまま」に集積を行う〔図2〕。

多くの典籍は、たとえば「部」「門」「目」のように、階層構造(平均2～3層)を持つものもあるが、付属語を取り去り、純粋に「分類概念語」を抽出した場合、そうした階層も含め、これらの語彙は、継承性が強いことを特色とする。



〔図2〕唐・徐堅等：撰『初学記』（筆者架蔵）の冒頭部およびその構造

地域・時代	撰者	書名	巻数	部名	部よみ	部付国文字	部題	門名	門よみ	門序	目名	目よみ
中国・唐	徐堅等	初学記	第一巻	天	てん	部	上	天	てん	一		
中国・唐	徐堅等	初学記	第一巻	天	てん	部	上	日	ひ	二		
中国・唐	徐堅等	初学記	第一巻	天	てん	部	上	月	つき	三		
中国・唐	徐堅等	初学記	第一巻	天	てん	部	上	星	ほし	四		
中国・唐	徐堅等	初学記	第一巻	天	てん	部	上	雲	くも	五		
中国・唐	徐堅等	初学記	第一巻	天	てん	部	上	風	かぜ	六		
中国・唐	徐堅等	初学記	第一巻	天	てん	部	上	雷	かみなり	七		
中国・唐	徐堅等	初学記	第二巻	天	てん	部	下	雨	あめ	一		
中国・唐	徐堅等	初学記	第二巻	天	てん	部	下	雪	ゆき	二		
中国・唐	徐堅等	初学記	第二巻	天	てん	部	下	霧	しも	三		
中国・唐	徐堅等	初学記	第二巻	天	てん	部	下	雹	ひょう	四		
中国・唐	徐堅等	初学記	第二巻	天	てん	部	下	露	つゆ	五		
中国・唐	徐堅等	初学記	第二巻	天	てん	部	下	霧	きり	六		
中国・唐	徐堅等	初学記	第二巻	天	てん	部	下	虹蜺	こうげい	七		
中国・唐	徐堅等	初学記	第二巻	天	てん	部	下	霹靂	せいせい	八		
中国・唐	徐堅等	初学記	第三巻	歳時	さいじ	部	上	春	はる	一		
中国・唐	徐堅等	初学記	第三巻	歳時	さいじ	部	上	夏	なつ	二		
中国・唐	徐堅等	初学記	第三巻	歳時	さいじ	部	上	秋	あき	三		
中国・唐	徐堅等	初学記	第三巻	歳時	さいじ	部	上	冬	ふゆ	四		
中国・唐	徐堅等	初学記	第四巻	歳時	さいじ	部	下	元日	がんじつ	一		
中国・唐	徐堅等	初学記	第四巻	歳時	さいじ	部	下	人日	じんじつ	二		

【図3】分類の概念抽出・記録

地域・時代	撰者	件数 一致率	皇覧	取名	北堂書鈔	芸文類聚	初学記	李嶠百廿詠	事類賦	和漢胡語集	分類語彙表	書名 件数/巻数(のべ数)
中国・魏	文帝		2	0	0	0	0	0	0	0	0	
			2	0%	0%	0%	0%	0%	0%	0%	0%	/2(2)
中国・漢	劉威成	取名	0	27	3	6	5	3	7	2	15	
			0%	11.11%	22.22%	18.52%	11.11%	25.93%	7.41%	55.56%	/27(27)	
中国・唐	虞世南	北堂書鈔	0	3	847	123	85	39	26	7	170	
			0%	0.35%	14.52%	10.04%	4.60%	3.07%	0.83%	20.07%	/847(874)	
中国・唐	歐陽詢・等	芸文類聚	0	6	123	745	124	88	87	33	174	
			0%	0.81%	16.51%	16.64%	11.81%	11.68%	4.43%	23.36%	/745(773)	
中国・唐	徐堅・等	初学記	0	5	85	124	338	80	84	31	120	
			0%	1.48%	25.15%	36.69%	23.67%	24.85%	9.17%	35.50%	/338(340)	
中国・唐	李嶠	李嶠百廿詠	0	3	39	88	80	132	63	22	64	
			0%	2.27%	29.55%	66.67%	60.61%	47.73%	16.67%	48.48%	/132(132)	
中国・宋	呉叔	事類賦	0	7	26	87	84	63	111	27	55	
			0%	6.31%	23.42%	78.38%	75.68%	56.76%	24.32%	49.55%	/111(116)	
日本・平安	藤原公任	和漢胡語集	0	2	7	33	31	22	27	131	59	
			0%	1.53%	5.34%	25.19%	23.66%	16.79%	20.61%	45.04%	/131(134)	
日本・現代	国立国語研究所	分類語彙表 中の漢字語彙	0	15	170	174	120	64	55	59	17,835	/17,835(22,608)
			0%	0.08%	0.95%	0.98%	0.67%	0.36%	0.31%	0.33%		

【図4】包含関係の比較

標本数追加が今後の課題だが、このことからでもさまざまなことがわかってくる。

まず第一に、挙例の各典籍間の親疎の度合いが、語彙の一致率比較によりある程度判明することである。たとえば、『李嶠百廿詠』『芸文類聚』『初学記』『事類賦』の四種は、いずれも、平安期以前の漢詩文の表現における影響の指摘される典籍群だが、分類用概念語彙の一致率から見ると、きわめて類縁性の高い典籍群として位置づけられるということが言える。

次に、分類用概念語彙に使用される語句は、少ないものでもその四分の一が、多いものではその半数が現代日本でも使用されているということである。このことは、各古典籍中の分類語彙を『分類語彙表』⁷⁾と比較することにより求められた結果である。

そして一致する語彙は、自然景物・年中行事・人事関係の語彙に集中している……などである[図5]。

このことは、古典的語彙が人間・自然に親近的な部分でさまざまに発生し、それゆえにこそ、現代にまでその命脈を保ち得たと考えることができよう。

■出現頻度上位（出現頻度 6 以上の）の一致語■				【注記】 分類語彙表中に含まれる上記資料中の古典語彙は409語 (分類語彙中では2.29%、古典語彙全体では23.44%)。 左下側は縦覧の一致件数/縦覧の標題件数 右上側は横覧の一致件数/横覧の標題件数 「分類語彙表」の標題件数は、漢字列のみで重複のない語彙を抽出した数 (平仮名・カタカナを含むものの総計は 31,999 件) したがって、「水」(Wednesday)と「水」(Water)は一語として扱う
頻度	語句	分類コード	分類概念	
9	立春	1.1631	節・節日(.1634)	
8	落花	1.553	枝・葉・花など	
8	鈴	1.456	楽器	
7	地	1.524	地形・山野	
7	油	1.5111	鉱物	
7	喪服	1.421	衣服	
7	夏	1.1624	季節	
6	海水	1.513	水	
6	盤	1.415	輪・車・棒・管など	
6	流刑	1.3612	刑	
6	博士	1.234	人物	

【図 5】上位出現頻度語および【図 4】凡例

そこで、このような特性を持つ古典的な典籍に取材したオントロジが、集積され、分析、整理されるならば、古典と現代とを文化的・理念的に接合するだけでなく、さまざまな分野における知的体系を包摂しうる道筋が開かれることになるのではないかと予測されるのである。

こうした手法で集積された語彙の方途としては、絵画やビデオ画像の分類への応用が考えられる。特に、絵画については、その標題命名の歴史が典籍に比べて浅く、こうした古典的類書に含まれるオントロジとの親和性が高いと予想される。人物・風景画の比率の高い古典絵画に至ってはなおさらだろう⁸⁾。

3.2. データベース構築にむけて

このように、古典的類書に使用される分類概念語彙と現代語との親和性のある特性を持つ語彙を含む古典的な典籍に取材したオントロジが、集積され、分析、整理されるならば、古典と現代とを文化的・理念的に接合するだけでなく、さまざまな分野における知的体系を包摂しうる道筋が開かれることになるのではないかという予測も生まれてくる。

そこで、オントロジを整理・分析するための「分析型データベースシステム」を構想することにより、古典と現代とを文化的・理念的に接合することに取り組むこととし、次の5点の目標を設定した。

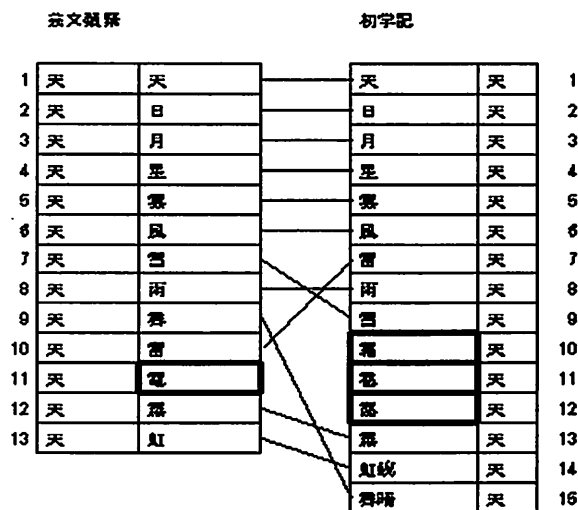
- ① 2000 年にも及ぶ叡智を注いで編まれ続けてきた古典的な類聚編纂物中の「部立」^{ふたて}に使用される「分類用語彙」の、語句の共通性だけでなく各典籍内における語句配列の一致を指標として、各典籍間の継承関係を考察することを支援する「分析型データベースシステム」を構築する。
- ② オントロジの観点から、日本・中国の古典的な辞書・辞典 (= 類書) の継承関係を分析し、各々の典籍の特性と意義について考察する。
- ③ 蓄積されたオントロジを利用して、和歌集、漢詩文集などの古典的類聚編纂物の編纂原理を分析するとともに、画像・映像などのコンテンツ分類への応用をはかる。
- ④ オントロジ概念を組み入れた、新しい文学研究 (古典学研究) を構築する。
- ⑤ 和漢古典学の発想にもとづく「オントロジナビゲーション」の提示をはかる。

3.3. 分析結果のモデル化のために

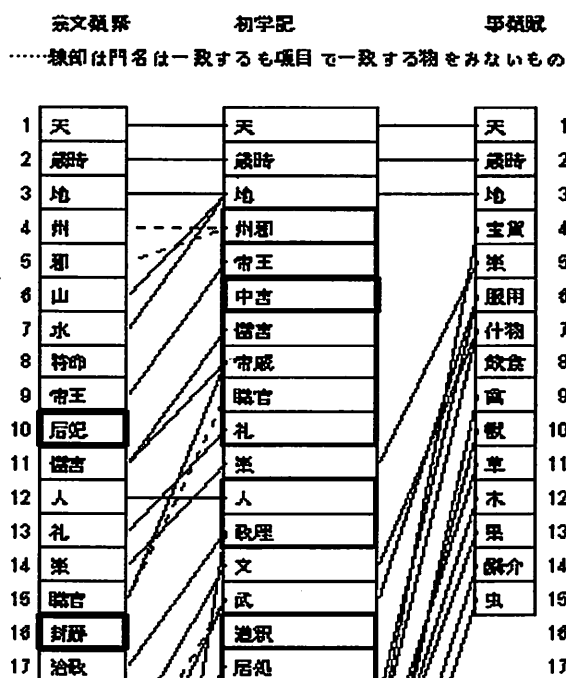
先の項目で挙げた5点の内、①の配列の分析と比較を行うことによって、それぞれの資料の特質を把握するという点については、これまで実証的な研究がなかなか進捗まなかつたものである。

その最大の要因は、視覚化と数値化というプレゼンテーションの方法に帰する。データ量があまりに膨大なために、明確な説明を行うことが難しかったのである。

たとえば、【図 6】は唐代類書の『芸文類聚』と『初学記』の分類概念語彙を比較したものである。一致する概念は——線で、類縁概念については、……線でそれぞれをつないで視覚化を試みた。



[図6] 『芸文類聚』と『初学記』の項目比較



[図7] 『芸文類聚』『初学記』『事類賦』3者比較

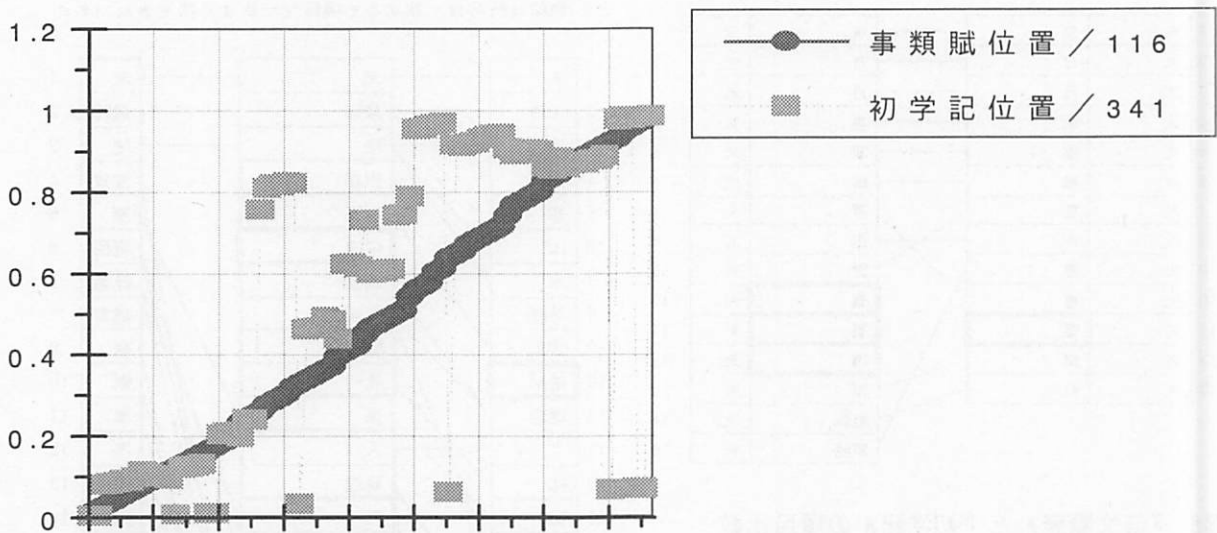
事類賦配列順	一致語	分類番号	事類賦位置比率/116	初学記配列順	初学記位置比率/341	順番変化
1	天	1	0.00862069	1	0.002932551	0%
2	天	1.01	0.017241379	1	0.002932551	0%
3	日	1.02	0.025862069	27	0.079178886	8%
4	月	1.03	0.034482759	28	0.082111437	0%
5	星	1.04	0.043103448	29	0.085043988	0%
6	風	1.05	0.051724138	31	0.090909091	1%
7	雲	1.06	0.080344828	30	0.08797654	0%
8	雨	1.07	0.088965517	33	0.096774194	1%
9	霧	1.08	0.077586207	38	0.11143695	1%
10	露	1.09	0.086206897	37	0.108504399	0%
11	霜	1.1	0.094827586	35	0.102639296	-1%
12	雪	1.11	0.103448276	34	0.099706745	0%
13	雷	1.12	0.112068966	32	0.093841642	-1%
14	歳時	2	0.120689655	2	0.005865103	-9%
15	春	2.01	0.129310345	41	0.120234604	11%
16	夏	2.02	0.137931034	42	0.123167155	0%
17	秋	2.03	0.146551724	43	0.126099707	0%
18	冬	2.04	0.155172414	44	0.129032258	0%
19	地	3	0.163793103	3	0.008797654	-12%
20	地	3.01	0.172413793	3	0.008797654	0%
21	海	3.02	0.181034483	69	0.202346041	19%

[図8]配列を数値化することによる分析の試み

この手法は、視覚的にわかりやすいという利点があるが、データ量が膨大になりすぎると、通覧性に難が生じる。また、飛び離れた位置に一致する概念語彙が発生した場合には、線が錯綜して判別が困難になるという欠点がある。

また、[図7]は『芸文類聚』『初学記』と宋代類書『事類賦』の三者で比較を試みたものであるが、実は『初学記』の観点から『芸文類聚』と『事類賦』とをそれぞれ比較してみたものにすぎず、この時点での比較においても視覚的な破綻をきたしつつある。

そこで、次に考えたのが、[図8]のように、それぞれの分類語彙が、全体語数の中の何番目に出現するかということで配列の数値化を試みた。



【図9】配列変化をグラフ化する

これは、二つのオントロジセット中の共通語彙を抽出し、それぞれが全語彙中の何番目に出現するかの割合を出し、その出現位置を比較するものである。

両者の出現順を比較して、値が「- (マイナス)」として出現するものは、当該位置において順序の変動が生じたことを示す。この形にデータが形成できると、『事類賦』から『初学記』をながめた場合（すなわち『事類賦』が『初学記』をどのように取り込んでいるか）、その配列にはどのような変化が起こるかという観点で（[初学記位置比率/341]）グラフ化が可能である。それを試みたものが【図9】である。

ただし、このような形にデータを加工して行くには、相当な手間を要する。そこで、比較的入力が行いやすい【図3】の形式から、【図9】のグラフ化の直前のデータを求めるための【図8】のデータを自動的に形成するプログラムを開発した（USC2.0以上も対応）。

4. オントロジ的発想による知識ナビゲート

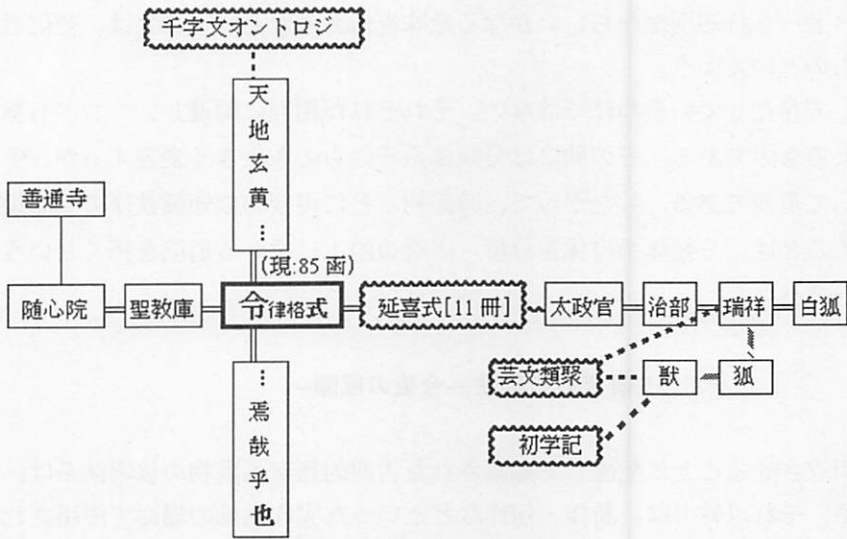
"オントロジ"という知識概念木の組合せによって立つ検索系が実現できれば、従来とは異なるパラダイムによる検索体系を構築することも可能となろう。

たとえば、【図10】は、「白狐」という事柄が、京都小野の古刹善通寺派の真言宗の寺、随心院における知識体系の中での文化的位置づけを図式化したものである。

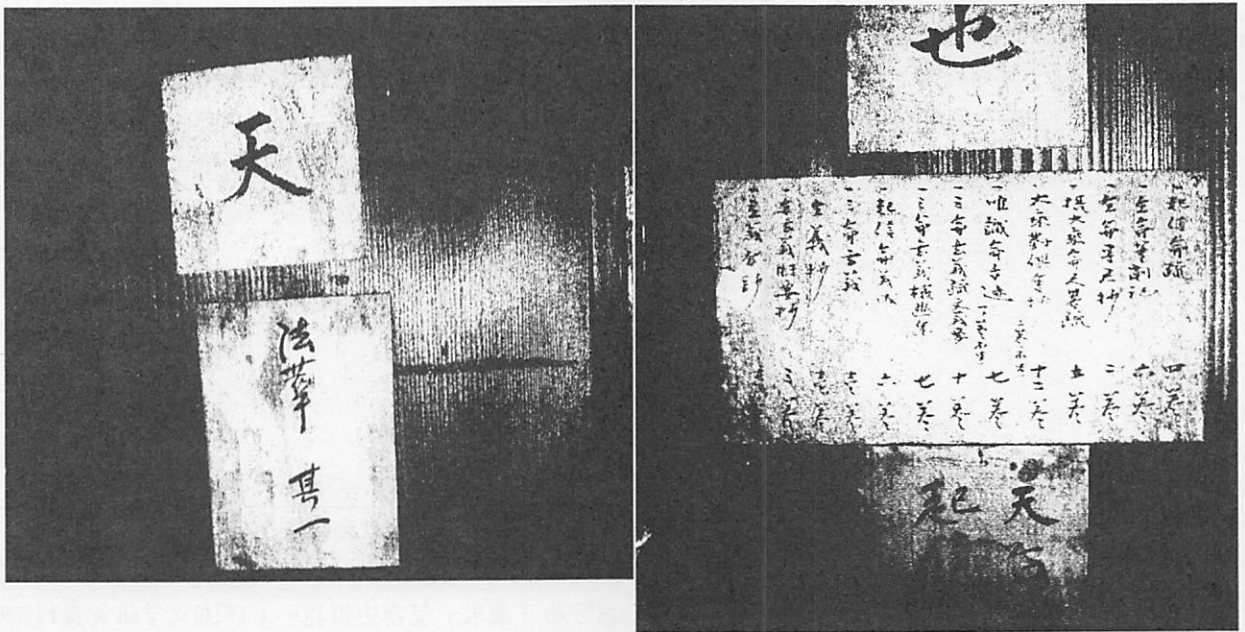
まず、随心院所蔵の『延喜式』太政官治部祥瑞の項に「白狐」の条がある。当該書は、「令律格式」と大字の「令」字に続いて「律格式」と書かれた箱に収められる。他の箱には、この「令」字は、ある時期に『千字文』に従った排架が行われていたと思われ、「天」や「也」字の書かれた箱が残っていることから、元来1,000個の箱が蔵されていた可能性が高い。

また、次に「令」に『延喜式』が納められていたことについては、後に続く「律格式」という連想が働いたと考えられ、『千字文』の各字の意味も意識した排架の体系になっていたと思われる。（『六国史』では「律令」という言い方が熟していたが、『職原抄』古本系では、「令律」という表記が使用されていた。）

そこで、それぞれのオントロジを組み立てて検索語の候補として並べたものが【図10】である。すると、



[図10] 狐をめぐるオントロジ



[図11] 『千字文』による分類ラベル

- ※白狐・瑞祥・治部・太政官・延喜式・律令格式
- ※白狐・獸 | 瑞祥・芸文類聚
- ※令・(千字文の各字)
- ※随心院・善通寺・真言宗
- ※随心院・小野・京都

などといった連想図式が想定でき、異なるオントロジの組合せ次第で、思わぬ発見を喚起するナビゲーションの可能性が暗示されるといえよう。

一方で、こうした発想は、既存資源を潤沢にを使って、いかに情報を読み解き説明するかという、データ

マイニングをサポートするための枠組みを、発信者側から意図的に情報に誘導せしめる仕組みにも有効であろう。無機質に並べ並べられる画像から、いかなる意味を伝えるかということは、発信者側のコンセプトに大きく依存するものといえよう。

情報は決して孤立して存在しているわけではない。それぞれが相互に関連し、一つの有機的な組織体として長期間の伝承されるものである。その間には分類体系そのものも大きく変容するが、その分類体系自体も、知識の記録として重要である。したがって、時系列ごとに可変的な知識秩序の原態復元を可能とする検索体系を実装することは、分類体系自体を分析・研究の俎上に乗せる道筋を拓くという点においても重要な視点といえるだろう。

5.研究の可能性—今後の展開—

通覧・検索の便を両立させることに配慮して編纂された古典的類聚編纂物の検索体系は、今なお古典研究者による出典分析や、それ以外では、詩作・句作などといった実作活動の場面で使用されている。しかしながら、それぞれの典籍の個性ゆえに、分類概念をたよりに多数の典籍を通覧するという検索手法はこれまで実現されたことはなかった。

資料操作に習熟を重ねた人は、その書の検索体系、すなわち分類体系を読み取って、初期の目的を実現しているが、その守備範囲は個別的で、限られた範囲にとどまる。

本研究の方法上の特色は、先人の残した事物の分類概念を、あるがままに集成・再構築することによる検索体系の復元・有効利用という点にある。が、その成果は、個々の古典作品の分析にも役立つ典拠資料として結実するのみならず、それをコンピュータの検索系に反映することにより、古典的分類概念が現代の事物にも適用可能かどうかについての評価を行うことにもなる。その意味で、本研究のもたらす意義は単に文学研究のみにとどまるものではなからう。

*1 溝口理一郎「オントロジー工学の試み」(1998年度人工知能学会全国大会(第12回)AIレクチャ)

*2 顧力仁『永楽大典及其輯佚書研究』(文史哲学集成 129, 文史哲出版社, 1985.7)、相田満「『永楽大典』収載の『蒙求』一群の書について—明代初期における『蒙求』受容史管見—」(『国文学研究資料館紀要』20, PP403-424, 1994.3.25)

*3 八代国治『吾妻鏡の研究』(明世堂書店、1941.11 修正再版)(藝林舎、1976 復刻版)

*4 相田満「河海抄の作られ方—『職原抄』引用から見える問題を中心に—」(国文学研究資料館紀要 26/ 国文学研究資料館/pp161-222, 2000.3.29)

*5 山岡俊明『類聚名物考』巻二百六十九書籍部第七(井上頼圀・近藤瓶城:校訂『類聚名物考』(五)、歴史図書社、1974.6.28)。句読点は稿者による。

*6 武田雅哉:『蒼頡たちの宴』,ちくま学芸文庫,夕 14--1,筑摩書房,PP.77-78,1998.5.8

*7『分類語彙表』(国立国語研究所編,秀英出版,1964.3 初版/大日本図書,1994.2.1,FDD 版)

*8 佐々木健一,タイトルの魔力—作品・人名・商品のなまえ学—,中公新書 1613,2001、内田順子,絵と詩—屏風歌以前,国語国文 69-9,200