

Journal of Japan Society of Information and Knowledge

情報知識学会誌

Vol.20 No.2 (May 2010)

~~~~~目 次~~~~~

## 特集 第 18 回 (2010 年度) 年次大会 (研究報告会&総会)

- オンライン書評と従来の書評との違いの分析…………… 原田 隆史, 吉村 紗和子 …… 65
- Q&A サイトにおけるベストアンサー推定の分析とその機械学習への応用  
…………… 石川 大介, 栗山 和子, 酒井 哲也, 関 洋平, 神門 典子 …… 73
- Internet Archive の Wayback Machine を使ったホームページの分析…………… 時実 象一, 杉浦 友哉 …… 86
- 環境・社会・経済のトリレンマ緩解論…………… 沢 恒雄 …… 93
- 社会変革の一般的構造…………… 安平 哲太郎 …… 103
- 人間の積極的に生きる能力を高めるために…………… 福永 征夫 …… 111
- 文芸批評の計量解析による批評行為の背景的特徴の抽出…………… 村井 源, 往住 彰文 …… 117
- 星新一ショートショート文学の物語パターン抽出…………… 佐藤 知恵, 村井 源, 往住 彰文 …… 123
- 音楽評論論文にみる作曲家の感性的特徴…………… 河瀬 彰宏, 村井 源, 往住 彰文 …… 129
- 計量分析による村上春樹文学の語彙構成と歴史的変遷…………… 工藤 彰, 村井 源, 往住 彰文 …… 135
- 業績評価に向けた正規化インパクト・ファクター,  
“IDV: Impact Deviation Value” (インパクト・ファクター偏差値) の提案 …… 根岸 正光 …… 141
- 学術の国際化による日本の産学共著関係の変化…………… 孫 媛, 根岸 正光 …… 149
- キーワード分析による環境関連研究の動向調査…………… 西澤 正己, 孫 媛 …… 155
- 資源供給情報を利用した耐熱材料の選択指針…………… 芳須 弘, 藤田 充苗, 原田 幸明 …… 163
- 学術俯瞰、特許俯瞰による商業化可能な技術要素の早期発見…………… 柴田 尚樹, 梶川 裕矢, 坂田 一郎 …… 171
- 次世代電池における国際的共同研究の構造…………… 佐々木 一, 梶川 裕矢, 坂田 一郎 …… 177
- 著者名典拠情報を拡充するための共同編集プラットフォーム…………… 川島 隆徳, 研谷 紀夫 …… 183
- 聖教書誌情報全文検索システムのスタンドアロン化の試み  
…………… 渡上 将治, 村川 猛彦, 宇都宮 啓吾, 中川 優 …… 189
- EUC/EUD を前提とした文学研究システムのモデル化に関する考察 …… 富澤 浩樹 …… 195
- カリキュラムの特徴抽出と時間割の要約生成…………… 堀 幸雄, 中山 堯, 今井 慈郎 …… 201
- 社会的ニーズを踏まえた法律情報に対する知識構造マップの開発  
…………… 川島 啓, 大竹 裕之, Adam Lobel, 山田 建智 …… 207
- MLA の記述規則に関する比較研究…………… 鈴木 良徳, 八重樫 純樹 …… 215
- 祭りデータベースモデルの研究調査…………… 山野 遼溪, 八重樫 純樹 …… 221



情報知識学会

<http://www.jsik.jp/>

# TOPPAN

## 印刷博物館。

# ここには、人類の知と創造への エネルギーがあふれています。

絵画と文字の始原を求める…。先人たちの知の遺産に触れる…。

そして、印刷とコミュニケーションの過去、現在、未来の姿を探る。

東京・文京区に開館した日本初の本格的な「印刷博物館」。

ここは人類の偉大なる知と創造へのエネルギーを感じることができるスペースです。



printing  
museum, Tokyo  
印刷博物館

〒112-8531  
東京都文京区水道1丁目3番3号  
TEL: 03-5840-2300 (代)

<http://www.printing-museum.org/>

●交通: JRおよび地下鉄有楽町線、東西線、南北線、大江戸線飯田橋駅より徒歩約13分。地下鉄有楽町線江戸川橋駅より徒歩約8分。地下鉄丸の内線、南北線後楽園駅より徒歩約10分。●開館時間: 10時～18時(入場は17時30分まで) ●休館日: 毎週月曜日(但し祝日の場合は翌日)、年末年始、展示替え期間 ●入館料: 一般(中学生以上) 300円、小学生100円、団体割引あり(税込)

# オンライン書評と従来の書評との違いの分析

## The analysis of differences between online book reviews and those in newspapers

○原田隆史<sup>1</sup>, 吉村紗和子<sup>1,2</sup>

Takashi HARADA, Sawako YOSHIMURA

1: 慶應義塾大学文学部図書館・情報学専攻

School of Library and Information Science, Keio University

〒108-8345 東京都港区三田 2-15-45

Email: ushi@slis.keio.ac.jp

2: 現在の所属は、イー・モバイル株式会社

オンライン書店のサイトをはじめとして、読者自身が図書の感想などを投稿するオンライン書評サイトが増加してきている。本研究は、このようなオンライン書評の持つ特徴を、新聞書評などと比較することで明らかにするものである。書評中の各文を、評価対象、評価の視点、評価の客観性、評価極性(肯定的か否定的)かという4つの観点から分類し、集計した。その結果、①評価対象は「作品に対する評価」がどの書評でも評価組全体の約9割を占め、書評ごとの変化は見られない、②評価の視点について、新聞書評では「作家の表現手法」などが全体の48%を占めるのに対し、オンライン書評では「ストーリー」や「場面」がほとんどである、③新聞書評では客観的な表現や肯定的な評価がほとんどであるのに対し、オンライン書評では主観的な表現や否定的な評価も多く多様な内容であることが明らかとなった。

Book review sites have been growing to which readers submit their own review. This study was aimed at investigating the characteristics of the online book reviews in comparison with those in newspapers. The sentences in the book reviews were analyzed based on the following points: target of evaluation, viewpoints, objectivity, and favorable evaluation or negative. The results were as follows: 1) 90 percent of the targets of evaluation were “the work itself.” There was no difference between online reviews and those in newspapers. 2) 48 percent of viewpoints were “the writing style” in reviews in newspapers, while “plot” and “setting” were the main viewpoints in online reviews. 3) Most of the reviews in newspapers were objective and favorable, on the other hand online reviews vary in tone including subjective and negative evaluation.

### 1. 書評の持つ特徴

#### 1. オンライン書評と従来の書評

書評とは、刊行された書物を紹介するための論評や感想の記述である。新聞、雑誌やテレビ、ラジオなど様々なメディア上に書評が存在しており、最近ではWebページを利用した書評の公開が

盛んに行われるようになってきた。Webサイト上で公開される書評の中には、新聞社が紙面に掲載した書評と同内容のものをWeb上に掲載する例もあるが、従来は公開される書評を執筆してこなかったような一般の読者が、その感想や評価を公開する例も多い。このような一般の人々がWeb

上で公開する書評を、新聞や雑誌などで公開される書評(新聞書評と略す)と区別してオンライン書評と呼ぶ。

このオンライン書評は、新聞書評とは、その形式・内容ともに異なることが予想される。本研究ではオンライン書評を、その評価している記述内容から新聞書評と比較分析する。

オンライン書評を公開しているサイトの例としては、以下があげられる。

#### 1) 個人のブログ

図書の感想や評価を掲載するブログは多い。その中には、書評が主な内容であるブログもあれば、日々の話題のひとつとして書評を取り上げるブログもある。評者としては、作家などが自身のブログで書評を掲載している場合や、一般の人が読書コンシェルジュなどと自称している場合も存在する。

#### 2) オンライン書店のレビュー

Amazon.co.jp や楽天ブックスなどのオンライン書店には、その提供するサービスのひとつとして、読者が図書に関する感想・評価文を自由に投稿でき機能を持つものがある。多いものでは1冊の図書に500件以上のレビューが投稿されている場合もある。

#### 3) 書評の投稿サイト

そのサービスの目的として書評を自由に投稿し、公開するWebサイトが多数存在している。その中には、図書の評価や感想を投稿すること限定しているものも、書評以外に映画やゲームなどの感想の投稿を募集するものも存在する。特に前者では、同じような対象に興味を持つ人が集まるコミュニティ的な要素が強い。オンライン書店と異なり本を販売すること自体は目的ではないが、オンライン書店へのリンクが各書評のページに貼られていることも多い。なお、個人のブログで書評を掲載している人が、同様の内容で投稿サイトに投稿している例もみられる。

### 1.2 書評を評価する際の評価要素

書評のような意見や評価を内容とする文章の

分析に関しては、意見抽出の分野で多くの研究が行われている。それらの研究では、文章中に表現される意見や評価に関する記述をもとに、いくつかの要素(評価の観点、表現されている項目)を設定して分析を行っている。

たとえば、小林のぞみらは各種の意見や評価表現を含む文章を収集・分析し、「評価対象」「評価の視点(属性)」「評価値」という3つの要素に注目することが有効であることを明らかにしている<sup>1)</sup>。また、これらの3つの要素を拡張して、以下の6つの要素を設定した分析も行っている<sup>2) 3)</sup>。

#### a) 評価主体

評価を行う主体となる人物、機関など

#### b) 評価対象

製品(書評の場合には図書)やサービス、人物、組織、政策など、評価分析の対象となる実体

#### c) 評価の視点(任意)

評価対象のどの構成要素あるいは属性が評価の焦点になっているか

#### d) 評価値

評価の視点に対応する評価の程度。肯定的な評価もあれば、否定的な評価もありえる

#### e) 比較対象(任意)

相対的な評価で参照される比較対象

#### f) 評価条件(任意)

評価が成立するための条件

これらの要素を用いた意見抽出や評価分析は、他にも多くの研究者が行っている。たとえば杉木健二らは、商品を対象とした消費者の評判分析を行う際に、「評価対象」「評価の視点(属性)」「評価値」という3つの要素を用いている<sup>4)</sup>。また水口弘紀らも、同様の3つの要素を用いてブログを対象にした評判情報分析システムの構築を行っている<sup>5)</sup>。

さらに倉島健らは、文書の中でも特に個人が体験した複数の商品の優劣を比べた文章では相対的な関係についての記述が重要であると考え、

「評価対象」「比較対象」「評価の視点(属性)」「評価値」という4つの要素を設定してランキングを行っている<sup>6)</sup>。



一方、立石健二らは、評価表現の分析を行い、人の感覚や感情を表す表現(「良い」、「好き」など)と、ものの性質や特徴を表す表現(「速い」「小さい」など)の2つが存在するとし、「評価値」はさらに「属性値」と「主観的評価」に分けられるとしている<sup>7)8)</sup>。青木優らも消費者の価値観に関する研究を行う中で、評価値に関する記述としては商品そのものに関する評価と、商品に対する印象や思い入れという二種類の表現があるとのべ、前者を「機能的価値」、後者を「情緒的価値」と表現している<sup>9)</sup>。

また、中山記男らは、映像や音楽、図書などの分野は、評価が二値では扱えないことを指摘し、多様な評価を扱うために、評者の感情表現を対象とした分析を行うことを提案している<sup>10)11)</sup>。中山らは感情表現を構成する要素として、「評価主体」「評価対象」「評価理由」「評価態度」を設定した分析を行っている。また、中山らは Amazon.co.jp のカスタマーレビューに対する評価を対象とした分析を行い、参考になったレビューには、参考にならないと評価されたレビューと比較して、より多くの「評価理由」が記述されていることを明らかにしている<sup>11)</sup>。

立石らの言う「属性値」は青木らの「機能的価値」と、また「主観的評価」は「情緒的価値」とはほぼ同じ内容と考えられる。さらに、中山らの「評価理由」も、評価理由という要素の設定が有効であるというだけでなく、客観的な評価基準の重要性を示しているとも考えられる。その意味で、中山らや青木らの設定しているものと同種の要素も含んでいると考えることができよう。

## 2. 書評の構造の分析実験

### 2.1 実験対象とする図書と書評

本研究では、オンライン書評と従来の書評との比較を行う。ここで、従来の書評としては、長い歴史を持ち、一般的に作家や評論家が評者となることが多い新聞書評を用いる。またオンライン書評としては、Web 上で一般の人が自身のブログに書き込んだ書評(以下、ブログ書評と呼ぶ)、およ

び、オンライン書店のレビューとして投稿された書評(投稿書評)の2種類を、その比較・分析を行う。

分析の対象とする図書は小説とし、オンライン書評と従来の書評の両方が存在する図書の中から選択する。具体的には、以下の全ての条件を満たす10冊とした。第1表に選択した図書を示す。

- 1) 2000～2007 年度に、朝日新聞・読売新聞・毎日新聞のうち、いずれか2紙以上の書評欄でとりあげられた図書
- 2) 「Yahoo!ブログ検索」で検索した際に、7つ以上のブログで取り上げられていると表示される図書
- 3) オンライン書店である Amazon.co.jp で、15件以上のカスタマーレビューが作成されている図書

第1表 評価対象とした図書

|    | 作品       | 著者        |
|----|----------|-----------|
| 1  | 模倣犯      | 宮部みゆき     |
| 2  | 東京奇譚集    | 村上春樹      |
| 3  | 悼む人      | 天童荒太      |
| 4  | ポトスライムの舟 | 津村記久子     |
| 5  | 乳と卵      | 川上未映子     |
| 6  | ひとがた流し   | 北村薫       |
| 7  | 東京島      | 桐野夏生      |
| 8  | 有頂天家族    | 森見登美彦     |
| 9  | 東京タワー    | リリー・フランキー |
| 10 | 終の住処     | 磯崎憲一郎     |

本研究で分析対象とする書評は、第1表に示す図書を対象とするもので、新聞記事について2つ、ブログ書評について3つ(文章の長い順に選択)、投稿書評について5つ(Amazon.co.jp のカスタマーレビューに対する評価で「有用性の高い順」に選択)とした。なお、投稿書評の選択にあたっては、その評者がブログ書評と同じかどうかを調べ、ブログ書評と同じ評者である場合には採用しないものとした。

## 2.2 評価対象要素の設定

本研究では、意見抽出や評価分析に関する先行研究をもとに、「評価対象」「評価の視点(属性)」「評価値・評価極性」「評価の客観性」「その他」という5つの要素を分析の際の評価指標として用い、これらの内容が含まれる文章を抽出して分析した。

### 1) 評価対象

評価対象そのものについて。具体的には「作者(作品の著者, 代名詞なども含む)」「作品(評価の中心である作品)」「評価の中心以外の作品(「著者の芥川賞受賞作」のような表現など)の3種類を設定した。

### 2) 評価の視点(属性)

評価対象の, どのような性質や特徴を評価しているかについて。評価の視点(属性)には大きく分けて「ストーリー(物語の全体像や流れに関する表現)」「登場人物(作品の登場人物に関する表現)」「場面(作品中のある一場面に関する表現)」「描写(作者の表現方法の特徴や癖など)」「その他(読後感など)」の5種類を設定した。

### 3) 評価値・評価極性

評価の視点に対する値について。ただし, 本研究では, 評価値そのものではなく, 評価が肯定的か否定的か, もしくは中立的か(評価極性)だけについて書評の種類ごとの違いを分析する。ここで, 中立は, 評価値が「良いとも悪いとも取れない」もしくは「良くも悪くも取れる」と判断される場合とする。

### 4) 評価の客観性

評価が主観的であるか客観的であるかについて。具体的には, 「面白い」「じーんとする」「好きだ」などの個人の感覚や, 「夢中になって読み進めた」など読んでいる最中の行動などが記述されている場合に主観的とした。また, 「洗練されている」「優れている」「良い」などの評価対象または属性に対する特徴を捉えている表現が記述される場合には客観的であるとした。

### 5) その他

その他, 以下のa)～c)に示す比較表現や評価

条件, 評価理由などについても, 記述されていれば抽出した。これらの項目は内容としては有効であるが, 書評の種類ごとの違いはほとんど見られなかった。

### a) 比較表現

「～と比較して」「～には見られない表現として」などのような他作品との相対評価

### b) 評価条件

- ・作品の読者あるいは書評の著者の性質や状態  
「関西弁がネイティブである人にとっては」,  
「精神的に安定している時には」など
- ・読書時の行動  
「先に他作品を読めば」など
- ・作品そのものに対する要望や条件  
「もっとストーリーが良かったら」など

### c) 評価理由

- ・「…なため」「…から」など

## 3. 書評の構造の分析結果

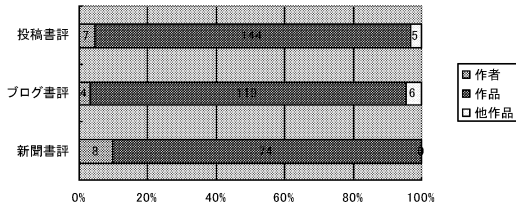
### 3.1 書評中の評価表現数

本研究で対象とする書評中の評価表現は, 新聞書評で82組, ブログ書評で129組, 投稿書評で156組存在していた。1つの書評あたりの評価組の数は, 新聞書評4.1組, ブログ書評4.3組, 投稿書評3.1組となり, 新聞書評とブログ書評がほぼ同数で投稿書評が最も少なかった。

ただし, それぞれの書評ごとの合計文字数は新聞書評22,538文字, ブログ書評27,364文字, 投稿書評16,951文字であり, 新聞書評274.9文字, ブログ書評212.2文字, 投稿書評108.6文字ごとに1つの評価表現となる。すなわち, 投稿書評は書評あたりの評価表現の数は少ないが, 評価文中に評価表現が出現する頻度は最も高い。これは新聞書評の方があらずじや内容などの説明部分が多いことが影響しといていると考えられる。

### 3.2 評価対象

評価対象(「作者」「作品」「評価の中心以外の他作品」の3つ)の, 各種類の書評における割合を第1図に示す。



第1図 書評の種類ごとの評価対象

第1図に示すように、全体の9割以上が作品に対する評価である。書評とは図書の感想や評価を記述するものであり、この結果は自然であろう。

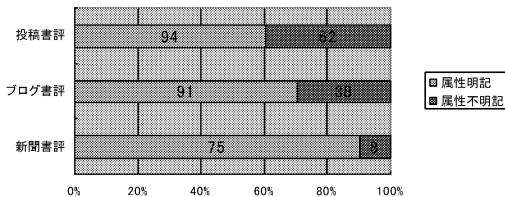
各種類の書評ごとの評価対象の割合は、いずれもそれほど大きくないが、ブログ書評と投稿書評には、新聞書評には存在しない「他作品」を対象にしたものが存在した。新聞書評はその書評だけで完結することが多く、記述内容はその対象となる作品を中心となるのに対して、ブログ書評や投稿書評では、同じ作者の違う作品や他の受賞作品も含めた広い範囲についての記述が行われることがあることを示している。

### 3.3 評価の視点(属性)

評価の視点とは、図書のどのような属性に関する評価を行うかということであり、「ストーリー」「登場人物」「場面」「描写」「その他」の5種類がある。グラフ中などでは属性とも表記する。

#### 3.3.1 評価の視点の明記・不明記

評価の視点が記述されているかどうかは、新聞書評・ブログ書評・投稿書評の間で明らかな違いが見られた。各種類の書評ごとの評価の視点が明記されている(属性明記)か、不明記(属性不明記)かの割合を第2図に示す。



第2図 書評の種類ごとの視点の明記

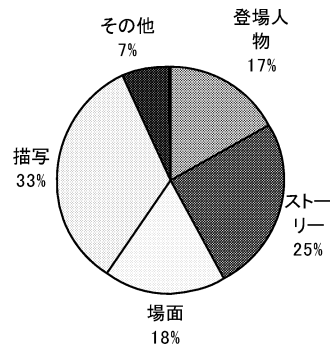
第2図に示すように、新聞書評では83件中の75件(90.4%)の評価表現において評価の視点が明記されているのに対し、ブログ書評では70.5%、投稿書評では60.3%と少なかった。

また、新聞書評では、評価値と直接的に関連づける形で評価の視点が明記されていない場合でも、前後の文脈から評価の視点が容易に類推できる場合が多かった。さらに、用いられている語句も評価の視点が明記されている場合と同じものが多い傾向が見られた。

一方ブログ書評および投稿書評においては、評価の視点が明記されていない場合には、作品全体もしくは作者自身に対する評価を示していると思われる例が多かった。たとえば、「素敵でした。」「いやあ、おもしろかった。」「毎度のことながら、惚れ惚れさせられます。」など、評価だけに言及した文章がこれにあたる。属性を記述しないことによって肯定・否定の評価をより強く書評の読者に印象付ける効果があるのではないかと考えられる。

#### 3.3.2 評価視点の種類

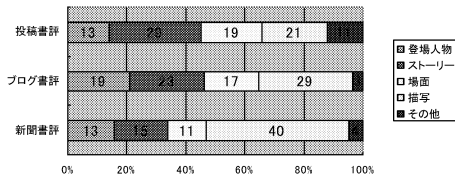
評価の視点が明記されている文書について、どのような視点について言及されているかについての出現比率を第3図に示す。



第3図 書評全体における評価視点の割合

第3図に示すように、作品中の「ストーリー」や「描写」に対する評価の割合が多く、特に「描写」は全体の3分の1、「ストーリー」は全体の

4分の1を占める。「人物」や「場面」に対する評価はほぼ同数で、いずれにも当てはまらない「その他」も7%存在する。ただし、この比率は書評の種類ごとに違いが見られた。評価視点の種類を、書評の種類ごとに第4図に示す。



第4図 書評の種類ごとの評価視点の割合

第4図に示すように、新聞書評では、「描写」に対しての評価が83件中の40件(48.2%)を占めている。また、「場面」は13.3%(11/83)と少なかった。

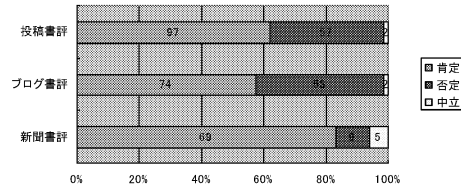
それに対して、ブログ書評や投稿書評では「描写」と同程度に「ストーリー」に対する評価表現が多く、「登場人物」や「場面」に対する評価表現も一定数存在した。特に、ブログ書評では、「登場人物」についての表現が多かった。また、投稿書評では、「その他」の割合が11.8%と他の書評に比べて多く見られた。

新聞書評の「描写」が多い理由としては、評者が専門家であることが多く、専門家として多くの書評を書いていることから、ストーリーなどの内容に加えて、作者の表現技術などにも言及する多様な文章を作成することが多いためと思われる。

さらに、投稿書評の「ストーリー」や「場面」に対する評価が多いのは、評者にとって評価をするよりも感想を書く目的の方が強い場合が多いため、より書籍の内容に沿った感想を記述しているためであると考えられる。また、投稿書評に多い「その他」の属性のひとつとして「読後感」が挙げられるが、これは、投稿書評が評価対象の書籍を読み終わってすぐに、気軽に感想を書いて投稿できることの影響があると思われる。

### 3.4 評価値・評価極性

評価極性としては、「肯定」「否定」「中立」の3種類がある。書評の種類ごとの評価極性の割合を第5図に示す。



第5図 書評の種類ごとの評価極性

第5図に示すように、いずれの種類の手書評においても全体の半分以上(投稿書評 62.2%, ブログ書評 57.4%, 新聞書評 83.1%)を肯定極性が占めていた。また、中立の評価値は極めて少なかった。書評は、図書を薦めるという目的で書かれることが多いことから、この結果は当然とも言える。ただし、書評の種類ごとには差異がみられた。

新聞書評では肯定が69件に対して否定が9件であり、全体の80%以上を肯定の評価極性が占めている。それに対してブログ書評や投稿書評では、否定の評価極性がそれぞれ4割前後存在する。また、中立の極性を持つ評価値はどの書評でも数は多くないが、新聞書評では他の種類の書評よりもやや割合が高かった。また新聞書評は、否定的な表現がされる場合でも「〜かもしれない」などと曖昧な語尾が用いられたり、評価者を読者に置き換えるなど、ブログ書評や投稿書評に比べて間接的な表現が多く見られた。

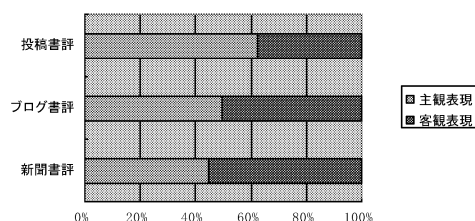
一方、ブログ書評と投稿書評では、否定的な表現も多く見られた。ただし、肯定的か否定的かについては作品によって大きな偏りがみられ、特に投稿書評ではその傾向が強かった。作品ごとの評価極性の割合を第2表に示す。第2表に示すように、投稿書評の方がより評価が極端な傾向にあるのは、他の人の評価を見ることができ環境下で書評が作成されることも多いことが影響していると考えられる。

第2表 作品ごとの評価極性の違い

|    |          | ブログ書評 |    |    | 投稿書評 |    |    |
|----|----------|-------|----|----|------|----|----|
|    |          | 肯定    | 否定 | 中立 | 肯定   | 否定 | 中立 |
| 1  | 模倣犯      | 11    | 21 | 0  | 22   | 10 | 0  |
| 2  | 東京奇譚集    | 10    | 4  | 0  | 14   | 3  | 0  |
| 3  | 悼む人      | 6     | 4  | 0  | 14   | 4  | 1  |
| 4  | ボストライムの舟 | 2     | 5  | 0  | 0    | 14 | 0  |
| 5  | 乳と卵      | 4     | 4  | 0  | 13   | 5  | 0  |
| 6  | ひとがた流し   | 10    | 5  | 0  | 11   | 2  | 0  |
| 7  | 東京島      | 4     | 2  | 2  | 3    | 12 | 0  |
| 8  | 有頂天家族    | 16    | 1  | 0  | 18   | 1  | 0  |
| 9  | 東京タワー    | 9     | 0  | 0  | 9    | 0  | 0  |
| 10 | 終の樓家     | 2     | 7  | 0  | 4    | 6  | 0  |

### 3.5 評価の主観性

評価表現における主観・客観表現の割合を第6図に示す。

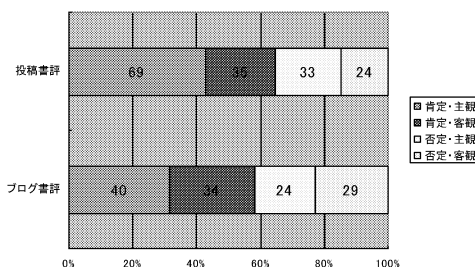


第6図 書評の種類ごとの評価の主観性

第6図に示すように、書評全体では、主観評価・客観評価の表現の割合は、主観表現がやや多いもののどちらもほぼ同数あることがわかる。ただし、新聞書評では主観表現 37(44.6%)、客観表現 46(55.4%)と全体の傾向とは異なり客観表現の方が多いのに対し、ブログ書評や投稿書評では55~60%を主観表現が占める。これは、新聞書評は他の書評に比べて文章やストーリーの特徴を述べた表現が多く、投稿書評では自身の感情を述べた表現が多いという内容の差を示している。

### 3.6 評価極性と主観・客観表現間の関係

評価極性と主観・客観表現の関係性を、作品ごとに肯定・主観、肯定・客観、否定・主観、否定・客観に分類し、第7図に示す。中立の評価極性をもつ場合は、数が少ないため省略する。また新聞書評は肯定の評価極性の割合が非常に高いため、分析する対象はブログ書評と投稿書評のみとする。



第7図 評価極性と主観・客観表現の関係

第7図に示すように、ブログ書評、投稿書評いずれも肯定・主観関係の評価組が一番多いが、ブログ書評では全体の31.5%(40/127)なのに対して投稿書評では42.9%(69/161)と全体に占める割合がより高い。また否定の評価では、投稿書評では客観表現よりも主観表現、ブログ書評では客観表現のほうが多かった。

### 4. オンライン書評と従来の書評

新聞書評では、評価の視点が明確に表現されることが多く、また表現も客観的な記述が多い。また評価以外に内容の説明も多く記述されている。これは、書評が新刊や各賞受賞作品を中心に、作品自体の紹介を行うという役割もになっていることが要因であろう。作品についての情報量が多いことから、新聞書評を読むだけである程度の内容や作風の特徴なども知ることができる。

一方、投稿書評では図書の内容などについての記述は少ない。これは、オンライン書店などに掲載された「あらすじ」などを参考にすることを前提にしていることがその背景にあると考えられる。また、多くの評者がレビューとして記述している内容は新聞書評やブログ書評に比較して短いことが多い。評価の表現としては、読後の感想を、自由な文章で記述したものが多く、その表現は、時としてわかりにくさを感じるほどに多様である。特に、新聞書評と比較して、肯定的な評価だけではなく、否定的な評価も多数掲載される点に特徴があると考えられる。

ブログ書評は、投稿書評と非常に近い内容と表



現形式を持つ。ただし、文章の長さや評価の主観性などの点で新聞書評と投稿書評の中間的な特徴を持つもっている。これは、評者が専門家ではないものの、多くの書評を他の人に読んでもらう目的で Weblog に掲載しているため評価属性等の記述は比較的明確なものが多いためと考えられる。

このように、オンライン書評は、従来の書評と比較して、内容的にも表現的にも違いが存在することが明らかとなった。近年、図書の分類やデータマイニングのための情報源として、このようなオンライン書評を利用する研究も行われるようになってきている。このような研究をより効果的に進めるためにも、オンライン書評についてのさらなる分析が望まれる。

## 5. 注・引用文献

- 1) 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. “テキストマイニングによる評価表現の収集”. 情報処理学会研究報告, NL154-012, 2003, p. 78-84
- 2) 小林のぞみ, 乾健太郎, 松本裕治. “意見情報の抽出／構造化のタスク仕様に関する考察”. 情報処理学会研究報告. Vol. 171, 2006, p. 111-118
- 3) 小林のぞみ, 乾健太郎, 松本裕治. “意見情報抽出のための評価対象・評価視点間の関係同定”. 言語処理学会第 12 回年次大会論文集, 2006, p. 65-68
- 4) 杉木健二, 松原茂樹. “消費者の意見に基づく商品検索”. 情報処理学会論文誌. Vol. 40, No. 7, 2008, p. 2598-2603
- 5) 水口弘紀, 土田正明, 久寿居大. “Weblog を対象にしたリアルタイム評判情報分析システム eHyouban”. 電子情報通信学会第 19 回データ工学ワークショップ論文集, 2008
- 6) 倉島健, 別所克人, 内山俊郎, 片岡良治. “比較評価情報の抽出とそれに基づくランキング手法の提案”. 電子情報通信学会第 18 回データ工学ワークショップ, 2007
- 7) 立石健二, 福島俊一, 小林のぞみ, 高橋哲朗, 藤田篤, 乾健太郎, 松本裕治. “Web 文書集合からの意見情報抽出と着眼点に基づく要約生成”. 情報処理学会研究報告. Vol. 93, 2004, p. 1-8
- 8) 立石健二, 福島俊一. “意見抽出を目的とした機械学習による属性－評価値対同定”. 情報処理学会研究報告. 2005-NL-165, 2005, p. 21-28
- 9) 青木優, 山本和英. “Blog を対象とした商品価値の分析実験”. 言語処理学会年次大会発表論文集, vol. 13, 2008, p. 680-683
- 10) 中山記男, 江口浩二, 神門典子. “感情表現の抽出手法に関する提案”. 自然言語処理研究会. No. 165, Vol. 3, 2004, p. 13-18
- 11) 中山記男, 神門典子. “理由に着目した感情表現の分析”. 電子情報通信学会技術研究報告. No. 105, Vol. 291, 2005, p. 51-56

研究報告

# Q&A サイトにおけるベストアンサー推定の分析とその機械学習への応用

## Analysis of Best-Answer Estimation for a Q&A Site and its Application to Machine Learning

石川 大介<sup>1\*</sup>, 栗山 和子<sup>2</sup>, 酒井哲也<sup>3</sup>, 関 洋平<sup>4</sup>, 神門 典子<sup>5</sup>

Daisuke Ishikawa<sup>1\*</sup>, Kazuko Kuriyama<sup>2</sup>, Tetsuya Sakai<sup>3</sup>, Yohei Seki<sup>4</sup>, Noriko Kando<sup>5</sup>

<sup>1</sup> 国立情報学研究所 (National Institute of Informatics)

〒 101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: dais@nii.ac.jp

<sup>2</sup> 白百合女子大学 (Shirayuri College)

〒 182-8525 東京都調布市緑ヶ丘 1-25

E-mail: kuriyama@shirayuri.ac.jp

<sup>3</sup> マイクロソフト リサーチ アジア (Microsoft Research Asia)

5F Beijing Sigma Center, 49 Zhichun Road, Haidian District, Beijing 100190, P.R.China

E-mail: tesakai@microsoft.com

<sup>4</sup> 豊橋技術科学大学 (Toyohashi University of Technology)

〒 441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1

E-mail: seki@tut.jp

<sup>5</sup> 国立情報学研究所 (National Institute of Informatics)

〒 101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: kando@nii.ac.jp

\* 連絡先著者

本研究では、Q&A サイトにおけるベストアンサーを計算機が推定可能か検証した。まず最初に、人間の判定者によるベストアンサー推定実験を行った。ベストアンサー推定実験には Yahoo!知恵袋データを利用し、「恋愛相談」「パソコン」「一般教養」「政治」の4つのカテゴリからそれぞれ無作為抽出した50問を使用した。判定者二人による推定結果の正解率(精度)は、「恋愛相談」では50%と52%(ランダム推定:34%)、「パソコン」では62%と58%(ランダム推定:38%)、「一般教養」では54%と56%(ランダム推定:37%)、「政治」では56%と60%(ランダム推定:35.8%)であった。次に、この実験結果を分析し、ベストアンサーを選ぶ要因として「詳しい」「根拠」「丁寧」を素性とする機械学習システムを構築した。判定者らと同じ50問を用いた推定実験の結果、機械学習システムの精度は、「パソコン」では判定者らの結果を上回り(67%)、「恋愛相談」では判定者らの結果を下回った(41%)。「一般教養」と「政治」では機械学習システムと判定者らの結果はほぼ同等であった。

In this research, we investigated whether a computer could estimate the best answer on a Q&A site. First, a best-answer estimation experiment was carried out with human assessors. The data of Yahoo! Chiebukuro was used for the experiment; 50 questions extracted at random from four categories, viz., "Consultation of love," "Personal computer," "General knowledge," and "Politics," were used. The accuracy rate(precision) of the estimation by two assessors was 50% and 52% (random estimation: 34%) for "Consultation of love," 62% and 58% (random estimation: 38%) for "Personal computer," 54% and 56% (random estimation: 37%) for "General knowledge," and 56% and 60% (random estimation: 35.8%) for "Politics." Next, the experimental results were analyzed, and the machine learning system with "Detailed", "Evidence", and "Polite" in the feature as a factor to choose the best answer was constructed. The precision of the machine learning system exceeded the assessors' results in the "Personal computer"(67%) category, and it fell below the assessors' results in the "Consultation of love"(41%) category. In the "General knowledge" and "Politics" categories, the precision of the machine learning system was almost equal to the assessors' results.

キーワード: Q&A サイト, ベストアンサー推定,  $\kappa$  係数, 機械学習

## 1 はじめに

近年, Q&A サイト (「Yahoo!知恵袋」<sup>[1]</sup> や「教えて!goo」<sup>[2]</sup> など) を対象としたソーシャルメディア研究が活発になってきている<sup>[3][4]</sup>.

これら Q&A サイトには, 最適な回答を選択するための枠組が備わっている. 例えば, 「Yahoo!知恵袋」では, (質問したユーザが) 最も納得・満足した回答を「ベストアンサー」として選択することができる. 「ベストアンサー」として選ばれた回答者は, 知恵コインと呼ばれる報酬を質問者から受け取ることができる. 知恵コインは質問を行う際, 懸賞金として利用することができる.

この Q&A サイトにおけるベストアンサーの中には, 質が高い回答が存在する. これら質の高い回答 (high-quality content) を識別して抽出する方法は, 現在盛んに研究されている<sup>[5][6]</sup>. 手法も様々であり, 例えば多様な素性 (textual, nontextual など) を用いて回帰分析によりベストアンサーを推測するもの<sup>[8]</sup> や, HITS アルゴリズムを適用しているもの<sup>[7]</sup> などが提案されている. これらは主に Yahoo! Answers<sup>[9]</sup> を対象とした研究である.

Yahoo!知恵袋を対象とした研究としては, 例えば西原らは<sup>[10]</sup>, 質問者と回答者の相性に着目し, 質問と回答の文末表現の相性に基づく機械学習によってベストアンサーの判別を行っている. 横山ら<sup>[11]</sup> も質問者と回答者の相性に着目し, 50 個の印象語を用いて印象評価実験を行い, その有効性を検証している. 栗山らは<sup>[12]</sup>, 回答者のカテゴリ毎のベストアンサー率などの履歴情報をベストアンサー推定へ利用可能かを検証している.

これら従来研究に対して, 本研究では第三者の人間 (判定者) がまずベストアンサー推定実験を行い, その結果からベストアンサーを選ぶための要因を分析する. そして, その要因に基づいて機械学習システムを構築し, 機械学習システムと人間 (判定者) の推定結果を比較する. これにより, 機械学習システムによるベストアンサー推定が, 人間の推定結果にどの程度接近できるかを検証することが狙いである.

以下, 本稿の構成を述べる. 本節は 1 節である. 2 節では, 実験に使用する Yahoo!知恵袋データの概要と詳細を述べる. 3 節では, 判定者によるベストアンサー推定実験の方法を述べる. 4 節ではその推定実験の結果を述べる. 5 節では, 4 節の実験結果からベストアンサーが選択される要因を分析する.

6 節では, その分析結果に基づいて機械学習システムを構築し, ベストアンサー推定実験を行う. 最後に 7 節で本稿をまとめる.

## 2 知恵袋データの概要

### 2.1 ベストアンサーとは

「ベストアンサー」の説明を, Yahoo!知恵袋のヘルプ<sup>[13]</sup> から抜粋したものを以下に示す.

1. 質問者は, もっとも納得, 満足した回答を「ベストアンサー」に選べます.
2. また, 自分でベストアンサーを選ばない場合は, ほかの利用者からの投票でベストアンサーを選ぶこともできます.

今回の実験で使用する知恵袋データ第一弾は, 1. の質問者本人によって選択されたベストアンサーのみ収録されている<sup>[15]</sup>. なお, ベストアンサーの選ばれ方の詳細については図 1 に示す. 以下, 本稿で「ベストアンサー」を略す際は **BA** と表記する.

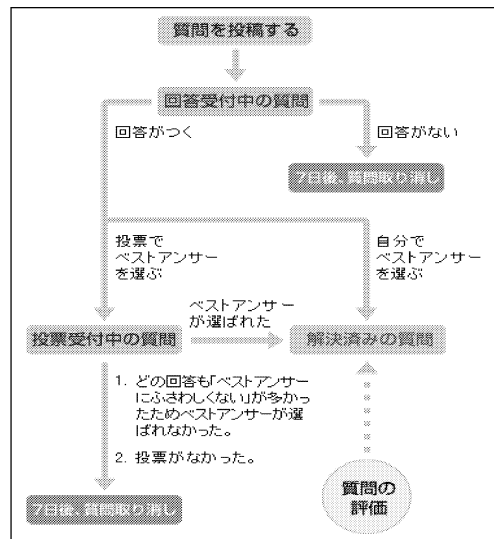


図 1: ベストアンサーの選ばれ方 (Yahoo!知恵袋のヘルプ<sup>[14]</sup> より引用)

表 1: Yahoo!知恵袋データの詳細 (仕様<sup>[15]</sup>より抜粋)

|         |                       |
|---------|-----------------------|
| 収録期間    | 2004/4/1 - 2005/10/31 |
| 解決済みの質問 | 3116009 件 (約 916 MB)  |
| ベストアンサー | 3116008 件 (約 935 MB)  |
| その他の回答  | 10361777 件 (約 2.3 GB) |

## 2.2 データの詳細

Yahoo!知恵袋データ第一弾の詳細は表 1 の通りである。以後、本稿では表 1 の「解決済みの質問」を質問、「ベストアンサー」と「その他の回答」とを合わせて回答と呼ぶこととする。また、Yahoo!知恵袋データ第一弾は、知恵袋データと呼ぶこととする。

## 2.3 カテゴリ毎の質問件数と平均回答数

知恵袋データのカテゴリ毎の質問件数と、そのカテゴリにおける平均回答数を集計した。質問件数の多い順にソートした順位、カテゴリ名 (カテゴリ ID)、質問件数 (その全体に対する割合)、平均回答数を表 2 に示す。

表 2 から、第 1 位から第 4 位までのカテゴリにおける質問件数は、それぞれ 17 万件～21 万件である。一方、第 5 位以下のカテゴリにおける質問件数は 10 万件を下回っている。よって Yahoo!知恵袋では、これら第 1 位から第 4 位のカテゴリが特に他のカテゴリに比べて活発に質問が投稿されていると考えられる。

次に、表 2 の平均回答数について述べる。まず、全質問における平均回答数は 4.32 である。これは、一つの質問に対して、(ベストアンサーを含めて) 平均約 4 件の回答が付いていることを表している。

質問件数が第 1 位の「恋愛相談、人間関係の悩み」では平均回答数は 6.74 であり、全質問における平均回答数 (4.32) よりも多い。一方、質問件数が第 4 位の「パソコン、周辺機器」では平均解答数は 2.76 であり、全質問における平均回答数 (4.32) よりも少ない。この理由の一つに、「恋愛相談、人間関係の悩み」では、誰もが自分の経験から回答することが可能であるに対して、「パソコン、周辺機器」では専らパソコンの専門知識を持つユーザのみが回答することができると考えられる。

さらに、質問件数が多い順の 100 位までの平均回答数について、平均回答数の上位 10 件と下位 10 件のカテゴリを表 3 に示す。表 3 から、「恋愛相談、人間関係の悩み」は平均回答数の上位グループに属し、「パソコン、周辺機器」は平均回答数の下位グループに属していることが分かる。

本研究では、上記「恋愛相談、人間関係の悩み」と「パソコン、周辺機器」に加えて、「一般教養」と「政治、社会問題」を対象に実験を行う。これらカテゴリを選択した理由は 3.1 節で述べる。

## 2.4 回答数に対する質問件数

表 4: 回答数の度数分布表

| 回答数 | 質問件数   | 割合     | 累積割合   |
|-----|--------|--------|--------|
| 1   | 539291 | 17.3 % | 17.3 % |
| 2   | 568982 | 18.2 % | 35.5 % |
| 3   | 502319 | 16.1 % | 51.6 % |
| 4   | 392856 | 12.6 % | 64.2 % |
| 5   | 294957 | 9.4 %  | 73.7 % |
| 6   | 216797 | 6.9 %  | 80.7 % |
| 7   | 159345 | 5.1 %  | 85.8 % |
| 8   | 115540 | 3.7 %  | 89.5 % |
| 9   | 84611  | 2.7 %  | 92.2 % |
| 10  | 62028  | 1.9 %  | 94.2 % |

回答数に対する質問件数を表 4 に示す。例えば、表 4 の最上位の回答数が 1 とは、回答が 1 件のみ付いている質問件数の合計が 539291 件 (17.3%) であることを示す。なお、回答数が 1 個だけの質問とは、ベストアンサーしか付いていない質問のことである。

次に、総質問件数に対する累積割合に着目する。回答数が 3 件の総質問件数に対する累積割合は 50% 弱である。つまり、知恵袋データの全質問は、回答数が 1 件のもの、2 件のもの、3 件のものが、全体の半数を占めていることを指す。回答数が 6 件のものまで含めると、累積割合は約 80% である。

## 3 ベストアンサー推定実験の方法

### 3.1 実験対象のカテゴリの選択

栗山と神門は、Yahoo!知恵袋に投稿される質問のタイプを社会調査型タイプと情報検索型タイプに分類している<sup>[16]</sup>。社会調査型タイプとは、広く意見

表 2: カテゴリ毎の質問件数

| 順位 | カテゴリ名 (カテゴリ ID)           | 質問件数 (割合)             | 平均回答数       |
|----|---------------------------|-----------------------|-------------|
| 1  | 恋愛相談、人間関係の悩み (2078297875) | <b>210105 (6.74%)</b> | <b>6.74</b> |
| 2  | Yahoo!知恵袋 (2078297310)    | 208793 (6.70%)        | 5.43        |
| 3  | Yahoo!オークション (2078297287) | 190432 (6.11%)        | 6.52        |
| 4  | パソコン、周辺機器 (2078297424)    | <b>171848 (5.51%)</b> | <b>2.76</b> |
| 5  | 病気、症状、ヘルスケア (2078297872)  | 84394 (2.70%)         | 3.43        |
| 6  | 政治、社会問題 (2078297787)      | <b>78777 (2.52%)</b>  | <b>5.11</b> |
| 7  | テレビ、ラジオ (2078297526)      | 75194 (2.41%)         | 3.25        |
| 8  | インターネット (2078297392)      | 70529 (2.26%)         | 2.45        |
| 9  | 言葉、語学 (2078297830)        | 68513 (2.19%)         | 3.67        |
| 10 | 動物、植物、ペット (2078297833)    | 61541 (1.97%)         | 3.91        |
| 11 | 国内 (2078297926)           | 61380 (1.96%)         | 3.24        |
| 12 | アダルト (2078297371)         | 61369 (1.96%)         | 4.53        |
| 13 | 自動車 (2078297760)          | 57627 (1.84%)         | 4.57        |
| 14 | 芸能人、タレント (2078297558)     | 55783 (1.79%)         | 4.40        |
| 15 | 野球 (2078297776)           | 53341 (1.71%)         | 4.42        |
| 16 | レシピ、調理法 (2078297961)      | 52430 (1.68%)         | 4.76        |
| 17 | 一般教養 (2078297828)         | <b>51940 (1.66%)</b>  | <b>3.54</b> |
| 18 | ファッション (2078297856)       | 48984 (1.57%)         | 3.33        |
| 19 | メンタルヘルス (2078297858)      | 47807 (1.53%)         | 4.37        |
| 20 | 健康、病気、ダイエット (2078297857)  | 41978 (1.34%)         | 3.27        |
|    | (全質問件数の合計とその平均回答数)        | 3116009 (100%)        | 4.32        |

を聞く質問タイプであり、情報検索型タイプとは、特定の回答を尋ねるタイプである

本研究では、この質問タイプに基づいて、社会調査型タイプから2カテゴリ、情報検索型タイプから2カテゴリ、合計4カテゴリを選択する。

4カテゴリを選択する方針として、まず質問件数が多い上位20位までのカテゴリ(表2)から選ぶこととした。

社会調査型タイプとしてはいくつか候補が考えられるが、特定の専門知識や背景知識を持っていないと答えられないようなカテゴリは除いた。すると、「(1位)恋愛相談、人間関係の悩み」「(6位)政治、社会問題」「(11位)国内」などが社会調査型タイプと考えられる。より質問件数の多いカテゴリ「(1位)恋愛相談、人間関係の悩み」と「(6位)政治、社会問題」を今回の実験対象に選んだ。

情報検索型タイプとしてはいくつか候補が考えられるが、特定の専門知識や背景知識を持っていないと答えられないようなカテゴリは除いた。すると、「(4位)パソコン、周辺機器」「(8位)インターネット」「(9位)言語、語学」「(17位)一般教養」などが情報検索型タイプであると考えられる。ここから、なるべく背景知識の影響が無いものとして、「(4位)パソコン、周辺機器」<sup>1</sup>と「(17位)一般教養」を今

回の実験対象に選んだ。

以下、本稿では「恋愛相談、人間関係の悩み」は「恋愛相談」、「パソコン、周辺機器」は「パソコン」、「一般教養」はそのまま「一般教養」、「政治、社会問題」は「政治」と表記する。

### 3.2 推定実験用の質問の作成方法

前節で選んだ4つのカテゴリ、「恋愛相談」「パソコン」「一般教養」「政治」を対象に、回答件数が2件から4件付随する質問に限定して、知恵袋データからランダムに50件の質問・回答セットを抽出した。

回答件数が2件から4件のものを対象とした理由として、まず一つは全体の質問件数の平均は4.32であることが挙げられる。また、2件から4件の回答を対象とすることで、判定者らには2択、3択、4択問題となり、感覚的に分かりやすい。さらに、カテゴリ間による回答件数の差を極力減らすことを考慮した。

よって、回答数が1件のみ(ベストアンサーのみ)の質問と、回答数が5件以上付随する質問は、今回の実験では使用していない。

抽出した50問の質問に付いた回答群はシャッフルした。これは、知恵袋では最後に投稿された回答がベストアンサーになりやすい<sup>[12]</sup>ため、その回答

<sup>1</sup>判定者らは共に情報学分野の研究者であるため、「パソコン、周辺機器」に関して十分背景知識があると考えた。

表 3: 平均回答数の上位 10 件と下位 10 件

| 平均回答数の上位 10 件 (降順) |         |                           |                |
|--------------------|---------|---------------------------|----------------|
| 平均回答数              | 質問件数の順位 | カテゴリ名 (カテゴリ ID)           | 質問件数 (割合)      |
| 6.74               | 1       | 恋愛相談、人間関係の悩み (2078297875) | 210105 (6.74%) |
| 6.52               | 3       | Yahoo!オークション (2078297287) | 190432 (6.11%) |
| 6.38               | 40      | マナー (2078297813)          | 23693 (0.76%)  |
| 6.22               | 81      | 家計、節約 (2078297796)        | 6659 (0.21%)   |
| 6.20               | 24      | 子育ての悩み (2078297890)       | 35153 (1.12%)  |
| 6.18               | 64      | ストレス (2078297874)         | 13274 (0.42%)  |
| 6.08               | 65      | 幼児教育、幼稚園、保育園 (2078297888) | 12944 (0.41%)  |
| 5.88               | 47      | 結婚 (2078297821)           | 20468 (0.65%)  |
| 5.51               | 32      | 小・中学校、高校 (2078297881)     | 26823 (0.86%)  |
| 5.50               | 50      | 新車 (2078297781)           | 19225 (0.61%)  |
| 平均回答数の下位 10 件 (降順) |         |                           |                |
| 平均回答数              | 質問件数の順位 | カテゴリ名 (カテゴリ ID)           | 質問件数 (割合)      |
| 2.37               | 87      | これ、探してます (2078297946)     | 6087 (0.19%)   |
| 2.45               | 8       | インターネット (2078297392)      | 70529 (2.26%)  |
| 2.54               | 54      | Yahoo!サービス (2078297295)   | 15744 (0.50%)  |
| 2.60               | 90      | 企業と経営 (2078297793)        | 5918 (0.18%)   |
| 2.62               | 25      | 携帯電話、モバイル (2078297430)    | 35001 (1.12%)  |
| 2.66               | 89      | 資格 (2078297903)           | 5944 (0.19%)   |
| 2.68               | 49      | スポーツ (2078297757)         | 19322 (0.62%)  |
| 2.69               | 71      | CM (2078297527)           | 10014 (0.32%)  |
| 2.70               | 28      | ゲーム (2078297514)          | 30085 (0.96%)  |
| 2.76               | 4       | パソコン、周辺機器 (2078297424)    | 171848 (5.51%) |

順序の情報によりベストアンサーが推測しやすくなることを防ぐためである。

その以外にも、質問番号や投稿日も同じく投稿順が判明する情報であるため、質問・回答セットからは外した。

ユーザ ID に関しても、特定のユーザが多くベストアンサーに選ばれるケースがあるため、ベストアンサーを推測する情報として利用できる可能性がある。今回は、質問文および回答文のテキストの中身から、ベストアンサーを選択することを考えているため、ユーザ ID の情報も質問・回答セットからは外した。

以上のようにして 4 カテゴリの質問・回答セットを構築した。これら 4 カテゴリの質問・回答セット (1 カテゴリにつき 50 件の質問とそれに付随する回答群のセット) を判定者らに配布した。質問・回答セットのサンプルを図 2 に示す。この例では A2 がベストアンサーに選ばれている。

### 3.3 判定者らによる実験方法

今回の実験では、判定者は二名 (第一著者と第二著者) であり、共に情報学分野の研究者である。この判定者らには、上記方法で作成した一つの質問に

対する回答 (2~4 件) の中から、質問者がベストアンサーに選んだと考えられる回答の一つ選択してもらった。

また、比較のために、判定者らに配布した質問・回答セットと同じものを、計算機を用いてランダムに推定するプログラムを作成した。このプログラムによってランダムに推定を 10000 回行い、その平均から推定結果を算出した。

## 4 ベストアンサー推定実験の結果

### 4.1 推定実験に要した時間

- 判定者らによる推定: (50 問の推定時間) 約 1 時間弱
- ランダムによる推定<sup>2</sup>: (50 問 × 10000 回の試行時間) 約 1.27sec(99%CPU)

### 4.2 ベストアンサー推定実験の結果

ベストアンサー推定結果を表 5 に示す。表 5 の結果から、以下の点が示される。

<sup>2</sup>使用した計算機環境は 6.2 節と同じ

```

<Q14>
<ID> 2110190 </ID>
<CATEGORY> パソコン、周辺機器 (2078297424) </CATEGORY>
<ANSERS> 2 </ANSERS>
<QUESTION 14>
ゴミ箱って、デスクトップから消せますか？
</QUESTION>
<A1>
WindowsXP なら出来たと思います。
</A1>
<A2>
消せますよ。下記リンク先で windows をカスタマイズするツールがダウンロードできます。これは使っている
も多いから安心です。他にもいろいろな設定が出来るよ。詳しくはリンク先で。
http://www.asahi-net.or.jp/~vr4m-ikw/
</A2>
</Q14>

```

図 2: サンプルデータ

- 判定者らの正解率は、どのカテゴリにおいてもランダムによる正解率を上回った。
- 判定者らの正解率は、「恋愛相談」では 50% と 52%, 「パソコン」では 62% と 58%, 「一般教養」では 54% と 56%, 「政治」では 56% と 60% であり、判定者間の差はほとんどなかった (正解数 1~3 件差)。
- むしろ、正解率の差は異なるカテゴリ間で生じている。例えば最も差が生じたのは、「恋愛相談」(判定者らの平均正解率が約 50%), 「パソコン」(判定者らの平均正解率は約 60%) であり、両者の間で 10% の差がある。
- 一方、ランダムによる正解率はそれぞれ 34% と 38% であり、カテゴリに対してはほとんど差が無かった。なお、「パソコン」と「一般教養」ではランダムによる正解率がやや高い (38% と 37%) が、これはこれらカテゴリでは選択肢が少ない質問が多いためである。

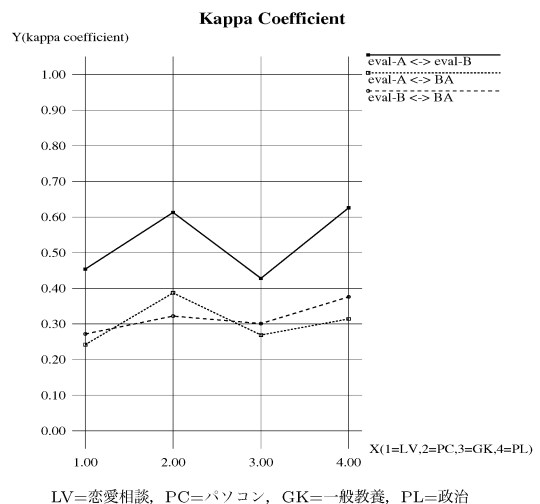


図 3:  $\kappa$  係数のグラフ

( $P_o$ : 観測された一致率,  $P_c$ : 偶然による一致率)

上記の計算式によって算出された  $\kappa$  係数の判定基準は次の通り [18].

- 0.0 ~ 0.2: slight
- 0.21 ~ 0.4: fair
- 0.41 ~ 0.6: moderate
- 0.61 ~ 0.8: substantial
- 0.81 ~ 1.0: almost perfect

上記で示した  $\kappa$  係数の計算式と判定基準を用いて、判定者 A, 判定者 B, ベストアンサーそれぞれの二者間の値を算出した。これら結果を表 6 に示す。

### 4.3 判定者間の一致度

次に、表 5 の結果を用いて、判定者間の  $\kappa$  係数を算出した。 $\kappa$  係数の算出方法は以下の通りである [17].

$$\kappa \text{ 係数} = \frac{P_o - P_c}{1 - P_c}$$



表 5: ベストアンサーの推定結果

| 判定者   | 恋愛相談<br>正解数 (正解率) | パソコン<br>正解数 (正解率) | 一般教養<br>正解数 (正解率) | 政治<br>正解数 (正解率) |
|-------|-------------------|-------------------|-------------------|-----------------|
| 判定者 A | 25/50 (50%)       | 31/50 (62%)       | 27/50 (54%)       | 28/50(56%)      |
| 判定者 B | 26/50 (52%)       | 29/50 (58%)       | 28/50 (56%)       | 30/50(60%)      |
| ランダム  | 17.0/50 (34%)     | 19.0/50 (38%)     | 18.75/50 (37%)    | 17.9/50 (35.8%) |

表 6: 判定者 A, 判定者 B, ベストアンサー間の  $\kappa$  係数

| 判定者 A と 判定者 B の一致率       |                                                |                                                   |                                                |                                                  |
|--------------------------|------------------------------------------------|---------------------------------------------------|------------------------------------------------|--------------------------------------------------|
| 項目名                      | 恋愛相談                                           | パソコン                                              | 一般教養                                           | 政治                                               |
| 判定者間の一致数                 | 32/50                                          | 38/50                                             | 32/50                                          | 38/50                                            |
| 観測された一致率 $P_o$           | 0.64                                           | 0.76                                              | 0.64                                           | 0.76                                             |
| 偶然による一致率 $P_c$           | 0.34                                           | 0.38                                              | 0.37                                           | 0.358                                            |
| $\kappa$ 係数              | $(0.64-0.34)/(1-0.34)$<br>= 0.454...(moderate) | $(0.76-0.38)/(1-0.38)$<br>= 0.613...(substantial) | $(0.64-0.37)/(1-0.37)$<br>= 0.428...(moderate) | $(0.76-0.358)/(1-0.358)$<br>= 0.626...(moderate) |
| 判定者 A とベストアンサー (BA) の一致率 |                                                |                                                   |                                                |                                                  |
| 項目名                      | 恋愛相談                                           | パソコン                                              | 一般教養                                           | 政治                                               |
| 判定者 A と BA の一致数          | 25/50                                          | 31/50                                             | 27/50                                          | 28/50                                            |
| 観測された一致率 $P_o$           | 0.5                                            | 0.62                                              | 0.54                                           | 0.56                                             |
| 偶然による一致率 $P_c$           | 0.34                                           | 0.38                                              | 0.37                                           | 0.358                                            |
| $\kappa$ 係数              | $(0.5-0.34)/(1-0.34)$<br>= 0.242...(fair)      | $(0.62-0.38)/(1-0.38)$<br>= 0.387...(fair)        | $(0.54-0.37)/(1-0.37)$<br>= 0.269...(fair)     | $(0.56-0.358)/(1-0.358)$<br>= 0.314...(fair)     |
| 判定者 B とベストアンサー (BA) の一致率 |                                                |                                                   |                                                |                                                  |
| 項目名                      | 恋愛相談                                           | パソコン                                              | 一般教養                                           | 政治                                               |
| 判定者 B と BA の一致数          | 26/50                                          | 29/50                                             | 28/50                                          | 30/50                                            |
| 観測された一致率 $P_o$           | 0.52                                           | 0.58                                              | 0.56                                           | 0.6                                              |
| 偶然による一致率 $P_c$           | 0.34                                           | 0.38                                              | 0.37                                           | 0.358                                            |
| $\kappa$ 係数              | $(0.52-0.34)/(1-0.34)$<br>= 0.272...(fair)     | $(0.58-0.38)/(1-0.38)$<br>= 0.322...(fair)        | $(0.56-0.37)/(1-0.37)$<br>= 0.301...(fair)     | $(0.6-0.358)/(1-0.358)$<br>= 0.376...(fair)      |

表 6 から, 判定者 A と判定者 B の  $\kappa$  係数は, 「恋愛相談」では 0.454(moderate, 中程度の一致), 「パソコン」では 0.613(substantial, かなりの一致), 「一般教養」では 0.428(moderate, 中程度の一致), 「政治」では 0.626(substantial, かなりの一致) を示した. どのカテゴリにおいても, 判定者間で選んだベストアンサーは, 中程度もしくはかなりの一致が認められる.

一方, 判定者 A とベストアンサーおよび判定者 B とベストアンサーの  $\kappa$  係数は, いずれも約 0.24 ~ 0.38 (fair) であり, 判定者間の値よりも著しく低い. 判定者間の一致よりも, 各判定者とベストアンサーとの一致の方が著しく低い原因については 5.4 節で議論する.

## 5 ベストアンサーを選択する要因の分析

本節では, 判定者 A, 判定者 B, ベストアンサーの三者の回答が一致した結果を中心に, ベストアンサーを選択する要因について分析する.

### 5.1 ベストアンサーを選ぶ要因

まずベストアンサーを選択する要因についてまとめる. 4 節の BA 推定実験を通じて得られた知見に基づき, 判定者間で議論してベストアンサーを選ぶための要因として考えられるものを以下にリストアップした.

1. 【分かりやすい】記述が分かりやすいもの
2. 【詳しい】詳しい説明があるもの: 詳しいけど, 説明がよくわからないものは BA としない. 場合によっては, 簡潔に書いてあるものを BA にすることもある.

3. 【根拠】 情報の根拠となる出展 (URL) があるもの：自分の体験に基づいているものなど (自分の体験>他人の体験>体験なし, 推察)
4. 【丁寧】 日本語が丁寧であるもの：日本語が丁寧な回答を BA に選ぶことが多かった。
5. 【ポジティブ】 質問者にポジティブなコメントを含むもの：例えば、「がんばってください」などの表現があるもの。逆に、質問者にネガティブなコメント (罵声など) があるものは BA に選んでいない。
6. 【的確】 的確で説得力があるもの (ただし根拠は示されていない)

## 5.2 三者間の一致・不一致

次に、三者間の一致・不一致を集計したものを表 7 に示す。表 7 の太字の項目は、判定者 A, 判定者 B, ベストアンサーが全て一致した件数である。以下本節では、これらを三者一致と呼ぶこととし、「恋愛相談」の 17 件、「パソコン」の 24 件、「一般教養」の 19 件、「政治」の 24 件を対象に分析する。

## 5.3 三者一致における要因分析

前述のベストアンサーを選ぶ要因の項目を用いて、表 7 における三者一致の結果を分析した。抽出された要因を表 8 にまとめた。さらに、表 8 の結果を要因毎に集計したものを表 9 に示す。

表 9 から、「恋愛相談」「パソコン」「一般教養」カテゴリでは、特に多い要因は、(2)【詳しい】と(4)【丁寧】である。「政治」カテゴリで特に多い要因は、(2)【詳しい】と(1)【分かりやすい】である。

次いで多い要因は、「恋愛相談」では(6)【的確】、「パソコン」と「一般教養」では(3)【根拠】、「政治」では(4)【丁寧】と(6)【的確】である。

一方、(5)【ポジティブ】は、どのカテゴリにおいてもほとんど抽出されない要因である。実際に、質問者に対して苦言であったり、論じたりする表現があるものの、質問に対しては的確に回答している場合は、それがベストアンサーに選択されるケースが見付かっている。ここから、質問者は必ずしも自分にとって都合が良い (ポジティブなだけのコメント) よりも、少々ネガティブなコメントが入っていても、質問に的確に答えて質問者が役に立つと思える回答をベストアンサーに選んでいるというのが実態であろうと推察できる。

## 5.4 一致しなかった結果

判定者間で選んだ回答は一致しているにもかかわらず、ベストアンサーとは一致しなかった結果がある。表 7 の「判定者 A と判定者 B が一致 (BA とは不一致)」の項目がそれに該当する。この項目は、全てのカテゴリにおいて三者一致の次に件数が多い (恋愛相談:15, パソコン:14, 一般教養:13, 政治:14)。

この項目に該当する質問・回答を調べてその原因を探った。ここでは、「パソコン」カテゴリを調べた結果、ベストアンサーと判定者らの選択が一致しなかった原因と考えられる理由を以下に示す。

1. ベストアンサーと判定者らが選んだ回答は、それぞれ若干の内容の差異はあるものの、ほぼ同等の回答である。どちらがベストアンサーかは甲乙つけがたく、どちらが選ばれても不思議ではない。
2. ベストアンサーと判定者らが選んだ回答は、それぞれ別の解決方法を示した回答であり、どちらも解決しうる方法である。どちらがより優れた回答かは判断できない。
3. ベストアンサーが適切な回答とは考えられないもの (Q:モデムが点滅している理由は?、A1:特に異常はありません、A2:あなたの PC はウィルスに感染しています、BA=A2)
4. 判定者らが選んだ回答の方がベストアンサーに相応しいと考えられるが、実際には別の回答がベストアンサーになっているもの。
5. どれもベストアンサーには相応しくない。
6. ベストアンサーが納得できる場合もある (再度、質問と回答を確認すると、確かにベストアンサーの方が納得のいく回答であった)

「恋愛相談」カテゴリでも一致しなかった結果を調査したが、ほぼ上記の内容 (特に 1,2,4 など) と同じ理由に基づくものが多かった。

今後、一致しなかった結果の分析をさらに進めてその原因を探る。

## 6 機械学習システムによるベストアンサー推定

前節で提示したベストアンサーに選ばれる要因に基づいて、これら要因を素性とする機械学習 (SVM) を用いてベストアンサー推定を行う。

### 6.1 選んだ素性とその計算方法

前述の分析結果から、今回は特に、2.【詳しい】、3.【根拠】、4.【丁寧】の要因を素性として機械学習に利用する。

表 7: 三者間 (判定者 A, 判定者 B, ベストアンサー) の一致・不一致の集計

| 項目名                          | 恋愛相談  | パソコン  | 一般教養  | 政治    |
|------------------------------|-------|-------|-------|-------|
| 判定者 A と判定者 B と BA が一致        | 17/50 | 24/50 | 19/50 | 24/50 |
| 判定者 A と判定者 B と BA が全て不一致     | 1/50  | 0/50  | 1/50  | 2/50  |
| 判定者 A と判定者 B が一致 (BA とは不一致)  | 15/50 | 14/50 | 13/50 | 14/50 |
| 判定者 A と BA が一致 (判定者 B とは不一致) | 8/50  | 7/50  | 8/50  | 4/50  |
| 判定者 B と BA が一致 (判定者 A とは不一致) | 9/50  | 5/50  | 9/50  | 6/50  |
| (合計)                         | 50/50 | 50/50 | 50/50 | 50/50 |

表 8: 三者一致 (判定者 A, 判定者 B, BA) の要因分析

| 恋愛相談 |              | パソコン |            | 一般教養 |            | 政治   |          |
|------|--------------|------|------------|------|------------|------|----------|
| 質問番号 | 要因番号         | 質問番号 | 要因番号       | 質問番号 | 要因番号       | 質問番号 | 要因番号     |
| Q3   | 4,6          | Q1   | 4          | Q2   | 3(URL),4   | Q1   | 2        |
| Q4   | 1            | Q2   | 2,3(自分),4  | Q7   | 2,4        | Q3   | 1        |
| Q6   | 2,6          | Q3   | 2,3(URL),4 | Q8   | 2          | Q4   | 2        |
| Q8   | 2,4,6        | Q6   | 2,4,6      | Q10  | 2,4,6      | Q8   | 1        |
| Q24  | 2,6          | Q8   | 2          | Q12  | 2,3(URL),4 | Q11  | 1        |
| Q25  | 4,6          | Q11  | 2,3(自分),4  | Q15  | 2          | Q14  | 2        |
| Q28  | 2,4,5        | Q12  | 2,4        | Q16  | 4,6        | Q15  | 4,6      |
| Q29  | 1,3(知人),4    | Q13  | 1(答えのみ)    | Q17  | 3(自分),4    | Q17  | 1        |
| Q31  | 2,4,6        | Q14  | 3(URL),4   | Q22  | 2,4        | Q20  | 4,6      |
| Q32  | 2,4,6        | Q16  | 2,4        | Q25  | 4,6        | Q22  | 4,6      |
| Q37  | 2            | Q18  | 2,4        | Q32  | 2,3(URL),4 | Q23  | 3        |
| Q39  | 4            | Q24  | 2,3(自分),4  | Q33  | 1          | Q25  | 3(URL),4 |
| Q40  | 2,3(うちの主人),4 | Q26  | 2,4        | Q37  | 3(URL),4   | Q26  | 2        |
| Q41  | 1            | Q27  | 2,4        | Q38  | 2,4        | Q27  | 3(URL)   |
| Q45  | 2,4          | Q28  | 2,4        | Q39  | 2,4        | Q28  | 2        |
| Q48  | 2,5,6        | Q29  | 2,3(URL),4 | Q43  | 2,4        | Q34  | 1        |
| Q49  | 3(自分)        | Q30  | 2,3(自分),4  | Q46  | 2,3(例),4   | Q35  | 6        |
|      |              | Q31  | 2,4        | Q48  | 2          | Q36  | 2        |
|      |              | Q32  | 2,4        | Q50  | 4,6        | Q38  | 2        |
|      |              | Q34  | 2,4        |      |            | Q41  | 2        |
|      |              | Q37  | 2,4        |      |            | Q42  | 2        |
|      |              | Q45  | 2,4,6      |      |            | Q46  | 1        |
|      |              | Q49  | 1          |      |            | Q47  | 2        |
|      |              | Q50  | 2,4        |      |            | Q49  | 1        |

まず、これら要因は次の方法で計算する。

- 【詳しい】→ 回答されたテキストの文字数をカウント。他の回答との比較のために、全回答の回答文字数の平均も併用する。
- 【根拠】→ /http/体験/経験/ のいずれかのパターンが含まれているかどうか
- 【丁寧】→ 回答テキスト中の /です/ と /ます/ の出現回数をカウント。他の回答との比較のために、全回答における /です/ と /ます/ の平均出現回数も併用する。

これらを用いて、素性 1~5 を以下のように定義する。

素性 1 回答文字数

素性 2 /です/ます/ パターンの出現回数

素性 3 /http/体験/経験/ のいずれかのパターン (有り=1, 無し=0)

素性 4 回答群の平均回答文字数

素性 5 回答群の /です/ます/ の平均出現回数

この素性に基づいて、それぞれの回答を

- 正例: ベストアンサー
- 負例: その他の回答

として上記の素性を計算して学習データを作成した。学習データのサンプルを図 6 に示す。

## 6.2 学習環境

SVM を用いた機械学習を行うシステムの詳細条件と計算機環境は以下の通り。

- 機械学習: SVM(TinySVM<sup>[19]</sup>)
- Solver Type: C-SVM (default)
- kernel: linear (default)
- 素性の数: 5 個 (前節参照)

表 9: 三者一致 (判定者 A, 判定者 B, BA) の要因分析の集計結果

| 要因          | 恋愛相談  | パソコン  | 一般教養  | 政治    |
|-------------|-------|-------|-------|-------|
| 1. 【分かりやすい】 | 3/17  | 2/24  | 1/19  | 7/24  |
| 2. 【詳しい】    | 10/17 | 20/24 | 12/19 | 10/24 |
| 3. 【根拠】     | 3/17  | 7/24  | 6/19  | 3/24  |
| 4. 【丁寧】     | 10/17 | 21/24 | 15/19 | 4/24  |
| 5. 【ポジティブ】  | 2/17  | 0/24  | 0/19  | 0/24  |
| 6. 【的確】     | 8/17  | 2/24  | 4/19  | 4/24  |

表 10: 訓練データ数とその学習結果

| 訓練データ | 精度    | 再現率   | F 値   | 処理時間 (CPU 効率)     |
|-------|-------|-------|-------|-------------------|
| 25    | 0.586 | 0.680 | 0.629 | 4.44sec(99%)      |
| 50    | 0.616 | 0.740 | 0.672 | 25.31sec(77%)     |
| 100   | 0.672 | 0.820 | 0.738 | 53.79sec(93%)     |
| 250   | 0.689 | 0.800 | 0.740 | 8min20.70sec(62%) |
| 500   | 0.672 | 0.820 | 0.738 | 18min50.1sec(99%) |

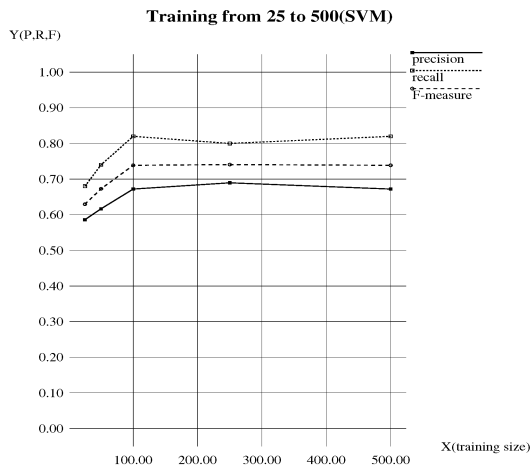
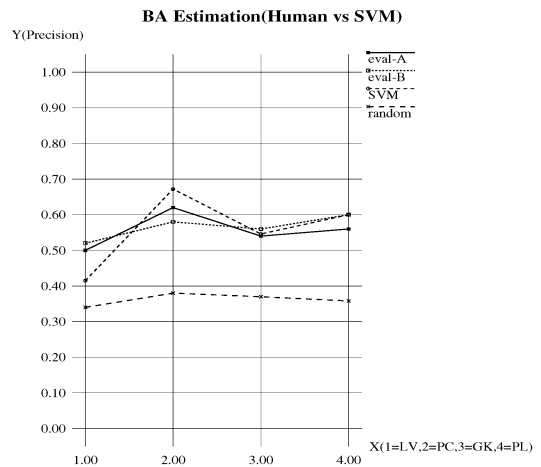


図 4: 訓練データ数と学習結果のグラフ



LV=恋愛相談, PC=パソコン, GK=一般教養, PL=政治

図 5: 判定者らと SVM による BA 推定精度

- 訓練データ: ランダムに抽出した 25,50,100,250,500 件, 回答数 2(詳細は下記参照)
- 評価データ: 判定者実験で使用した 50 問 × 4, 回答数 2~4
- 推定方法: 素性で表現された回答文に対してベストアンサーかそうでないかを 2 値分類
- 使用計算機:
  - OS: CentOS 5.3 64bit 版
  - CPU: Xeon 2.0GHz Quad Core
  - メモリ: 16GB

- ディスクアレイ: 1TBx12 (RAID 6.0, 4Gbps FC)

この環境により, 機械学習を行うシステムの性能評価を行う。なお, システムは, 各回答文に対してベストアンサー (positive) か, そうでない (negative) かを判別する。この実験では質問文は使用していない。

得られた結果の精度, 再現率, F 値はそれぞれ以

表 11: SVM による 4 つのカテゴリにおけるベストアンサー学習結果

| カテゴリ | 精度            | 再現率           | F 値   | 処理時間 (CPU 効率)      |
|------|---------------|---------------|-------|--------------------|
| 恋愛   | 0.415 (32/77) | 0.640 (32/50) | 0.503 | 44.15sec (99%)     |
| パソコン | 0.672 (41/61) | 0.820 (41/50) | 0.738 | 50.36sec (99%)     |
| 一般教養 | 0.546 (35/64) | 0.700 (35/50) | 0.614 | 58.61sec (99%)     |
| 政治   | 0.600 (36/60) | 0.720 (36/50) | 0.654 | 1min10.95sec (99%) |

```

-1 1:86 2:1 3:0 4:175 5:3 # Q1-A1 NA
1 1:264 2:5 3:0 4:175 5:3 # Q1-A2 BA
-1 1:407 2:3 3:0 4:585.5 5:4 # Q2-A1 NA
1 1:764 2:5 3:1 4:585.5 5:4 # Q2-A2 BA
-1 1:118 2:0 3:0 4:446.5 5:4 # Q3-A1 NA
1 1:775 2:8 3:0 4:446.5 5:4 # Q3-A2 BA

```

図 6: TinySVM に与えた学習データのサンプル

下の式に基づいて算出する.

$$\text{精度 } p = \frac{\text{システムが選んだ } BA \text{ の正解数}}{\text{システムが } BA \text{ として選んだ数}}$$

$$\text{再現率 } r = \frac{\text{システムが選んだ } BA \text{ の正解数}}{\text{全ての } BA \text{ の数}}$$

$$F \text{ 値} = \frac{2 \times p \times r}{p + r}$$

### 6.3 訓練データの件数

まず, 今回の実験にはどのぐらいの訓練データが必要かを「パソコン」カテゴリを対象に調べた. 「パソコン」カテゴリを選んだ理由は, 判定者らのベストアンサー推定の正解率が一番高く, また実験を通してベストアンサーが一番分かりやすかったため, 最初の学習例として適していると考えた.

評価データは判定者らが実験した「パソコン」カテゴリの 50 問を使用する.

訓練データは, 同じく「パソコン」カテゴリから, 新たにランダムで 25 件, 50 件, 100 件, 250 件, 500 件の質問・回答セットを抽出して作成した. なお, ベストアンサー (正例) とその他の回答 (負例) との件数を合わせるために, 回答数が 2 件のものだけを対象に抽出した.

以上の条件で学習させた結果を表 10 に示す. また, 表 10 の精度, 再現率, F 値をグラフにした結果を図 4 に示す.

この図 4 のグラフから, 訓練データ 100 件以上の学習結果はほぼ横ばいである. よって, 今回の実験

では訓練データ数は 100 件で十分と考え, 以後の実験を行う.

### 6.4 判定者らの推定結果との比較

次に, 「パソコン」以外のカテゴリ「恋愛相談」, 「一般教養」, 「政治」について学習を行う.

「恋愛相談」, 「一般教養」, 「政治」カテゴリから, 新たにランダムに 100 件の質問・回答セット (回答数は 2 件) を抽出し, 訓練データを作成した. それぞれのカテゴリの訓練データによって学習させたモデルを, 同カテゴリの評価データ (判定者実験で利用したもの) を用いて評価した.

これらの結果と, 「パソコン」カテゴリにおける結果を合わせて表 11 に示す.

表 11 の SVM による BA 推定の精度と, 判定者らの BA 推定結果 (表 5) と, ランダムによる BA 推定結果 (表 5) とを合わせてグラフにした. これを図 5 に示す.

SVM による精度と判定者らによる精度は共にランダムによる精度を上回っている.

SVM と判定者らの結果を比較すると, SVM の精度は, 「パソコン」においては判定者らの結果を上回り (67%), 「恋愛相談」においては判定者らの結果を下回った (41%). 「一般教養」と「政治」では SVM と判定者らの結果はほぼ同等の結果が得られた.

「恋愛相談」では SVM による結果は判定者らの結果に比べ, 約 10% も差があることから, 今後は「恋愛相談」に適した素性などを適切に選択・計算することによって精度が改善する可能性があると考えられる.

### 6.5 素性毎の学習結果

素性毎の学習効果を調べるため, 「パソコン」カテゴリを対象とし, 各素性毎の訓練データ (100 件)

を用意して機械学習を行った。この結果を表 12 に示す。

表 12: 素性毎のベストアンサー学習結果

| 素性        | 精度    | 再現率   | F 値   |
|-----------|-------|-------|-------|
| 1         | 0.615 | 0.640 | 0.627 |
| 1+4       | 0.656 | 0.840 | 0.736 |
| 2         | 0.545 | 0.600 | 0.571 |
| 2+5       | 0.600 | 0.600 | 0.600 |
| 3         | 0.625 | 0.200 | 0.303 |
| 1+3+4     | 0.645 | 0.800 | 0.714 |
| 2+3+5     | 0.592 | 0.640 | 0.615 |
| 1+2+4+5   | 0.677 | 0.840 | 0.750 |
| 1+2+3+4+5 | 0.672 | 0.820 | 0.738 |

表 12 から、素性 1 単体より、素性 1+4 の方が性能が高く、また素性 2 単体より、素性 2+5 の方が性能が高いことが示される。

素性 3 は単体では意味があり、素性 2+5 に対して素性 2+3+5 の方が性能が上回っている。しかし、それ以外の他の素性との組み合わせでは、性能向上に寄与していない。素性 3(根拠)の計算方法については今後改善の余地があると考えられる。

## 7 おわりに

本研究では、Yahoo!知恵袋から「恋愛相談」「パソコン」「一般教養」「政治」のカテゴリを対象に、判定者(二名)によるベストアンサー推定実験を行った。その結果から、ベストアンサーを選ぶための要因を分析した。また、その要因の中から特に【詳しい】【根拠】【丁寧】を用いて機械学習システムを構築し、判定者らの推定結果と比較した。その結果、機械学習システムの精度は、「パソコン」では判定者らの結果を上回り、「一般教養」と「政治」では判定者らとほぼ同等の結果が得られた。しかし、「恋愛相談」では判定者らの推定結果を下回った。今後、「恋愛相談」で機械学習システムの精度が低い原因を調査し、より精度の高い機械学習システムを構築する方法を検討する。

## 謝辞

本研究の実施にあたって、ヤフー株式会社が国立情報学研究所に提供した「Yahoo! 知恵袋データ(第 1 弾)」を利用いたしました。

## 参考文献

- [1] Yahoo!JAPAN: Yahoo!知恵袋,  
<http://chiebukuro.yahoo.co.jp/>
- [2] OKWave: 教えて!goo,  
<http://oshiete.goo.ne.jp/>
- [3] 第 1 回 知識共有コミュニティワークショップ,  
[http://www.infosocio.org/cfp\\_workshop\\_a2008.html](http://www.infosocio.org/cfp_workshop_a2008.html)
- [4] 第 2 回 知識共有コミュニティワークショップ,  
[http://infosocio.org/cfp\\_workshop\\_a2009.html](http://infosocio.org/cfp_workshop_a2009.html)
- [5] Agichtein, E. and Castillo, C. and Donato, D. and Gionis, A. and Mishne, G.: Finding high-quality content in social media, Proceedings of the international conference on Web search and web data mining, pp.183–194, 2008.
- [6] Wang, X.J. and Tu, X. and Feng, D. and Zhang, L.: Ranking community answers by modeling question-answer relationships via analogical reasoning, Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, pp.179–186, 2009.
- [7] Jurczyk, P. and Agichtein, E.: Discovering authorities in question answer communities by using link analysis, Proceedings of the sixteenth ACM conference on Conference on information and knowledge management(CIKM), pp.919–922, 2007.
- [8] Bloom, M.J. and Chua, A.Y.K. and Goh, D.H.L.: A predictive framework for retrieving the best answer, Proceedings of the 2008 ACM symposium on Applied computing, pp.1107–1111, 2008.
- [9] Yahoo!Answers:  
<http://answers.yahoo.com/>
- [10] 西原陽子, 松村真宏, 谷内田正彦: QA サイトにおける質問に適した回答の判定, NLP 若手の会 第 2 回シンポジウム, 2007.
- [11] 横山友也, 宝珍輝尚, 野宮浩揮, 佐藤哲司: 質問回答サイトの質問文と回答文の印象評価, 第

- 2 回データ工学と情報マネジメントに関する  
フォーラム (DEIM Forum 2010), C4-2, 2010.
- [12] 栗山 和子, 神門 典子: Q&A サイトにおける質  
問と回答の分析 (3) -質問・回答履歴を用いた  
ベストアンサー推定-, 情報処理学会 第 97 回情  
報学基礎研究会, 2009-FI-97, 2009.
- [13] Yahoo!知恵袋ヘルプ - ベストアンサーとは,  
[http://help.yahoo.co.jp/help/jp/chiebukuro/qa/qa-  
22.html](http://help.yahoo.co.jp/help/jp/chiebukuro/qa/qa-22.html)
- [14] Yahoo!知恵袋ヘルプ - ベストアンサーを選ぶ,  
[http://help.yahoo.co.jp/help/jp/chiebukuro/qa/qa-  
23.html](http://help.yahoo.co.jp/help/jp/chiebukuro/qa/qa-23.html)
- [15] Yahoo!JAPAN 「Yahoo!知恵袋」 データの提  
供について,  
<http://research.nii.ac.jp/tdc/chiebukuro.html>  
(研究機関用データは国立情報学研究所 IDR  
事務局から配布されている)
- [16] 栗山 和子, 神門 典子: Q&A サイトにおける質  
問と回答の分析, 情報処理学会 第 95 回情報学  
基礎研究会, 2009-FI-95, 2009.
- [17] Sim, J. and Wright, C.C., The kappa statistic  
in reliability studies: use, interpretation, and  
sample size requirements, Physical Therapy,  
Vol.85, No.3, pp.257-268, 2005.
- [18] Teufel, S. and Moens, M., Summarizing scien-  
tific articles: experiments with relevance and  
rhetorical status, Computational Linguistics,  
Vol.28, No.4, pp.409-445, MIT Press, 2002.
- [19] TinySVM:  
<http://chasen.org/taku/software/TinySVM/>

## Internet Archive Wayback Machine を用いたホームページの分析

### **Web Site Analysis Using the Wayback Machine of Internet Archive**

時実象一<sup>1\*</sup>, 杉浦友哉<sup>1</sup>

Soichi TOKIZANE<sup>1\*</sup>, Tomoya SUGIURA<sup>1</sup>

1 愛知大学

Aichi University

〒441-8522 愛知県豊橋市町畑町 1-1

E-mail: tokizane@aichi-u.ac.jp

\*連絡先著者 Corresponding Author

Web 上のホームページを収集して保存・公開する Web アーカイブは最近注目を集めている。米国の Internet Archive の Wayback Machine はこの分野のパイオニアで、1996 年からわが国を含め世界の Web ページを収集して公開している。ここに収集されているわが国の数機関のホームページを経年的に比較調査し、それら機関の情報発信の状況を分析した。

Web archiving is to collect internet web pages, archive them, and make them available to the public. Wayback Machine of Internet Archive is a pioneer in this area, collecting web pages since 1996. Authors studied yearly changes of web pages of several Japanese organizations in Wayback Machine, and analyzed their public relation activities.

Web アーカイブ, ホームページ, Internet Archive, Wayback Machine, PR

Web archiveing, homepage, Internet Archive, Wayback Machine, public relations



## 1 はじめに

WebアーカイブとはWeb上のホームページを収集して保存・公開する取り組みであり、変化が激しくすぐに改廃されるWebページを文化遺産として記録して将来に残そうというものである。たとえば Google で "a" を検索すると 117 億件のヒットがある (2010/4/2) ことから見て、英語圏に限ってみても、この数字を超える Web ページがあると推察される。こうした莫大な情報が急速に改廃され消滅していくことは人類の知識と遺産が失われることといえる。

最新のページだけを利用者に見せる検索エンジンとは異なり、Webアーカイブはある時点の、または数世代にわたる過去のWebページを保存する点に特徴がある。なお Web アーカイブの動向については「情報の科学と技術」誌の特集が参考になる [1-6]。

## 2 Web アーカイブ活動の現状

### (1) IIPC (International Internet Preservation Consortium)

IIPC (International Internet Preservation Consortium) は2003年 7月に発足した組織で、後述する Internet Archive の呼びかけに応じた欧米の国立図書館を中心とする12機関で構成されている。インターネット上の情報資源の保存のための協力や標準策定を目的としている。わが国からは国立国会図書館が参加している。

IIPC の主な成果には Web ページの保存用フォーマットの規格 WARC、Web ページ収集用のクローラ Heritrix などがある。

### (2) Internet Archive

アメリカの非営利団体であるInternet Archive は1996年Brewster Kahleによって設立された。現在年間運営費は1000 万ドルであり、政府や

財団の補助や寄付で運営している [8]。設立時より、全世界の Web ページを対象に包括的に収集をおこなっており、収集したページは Wayback Machine で閲覧できる。2009年5月現在で約1500億ページ(5000万サイトに対応する)もの Web ページが収集されており、事実上世界最大の Web アーカイブである。

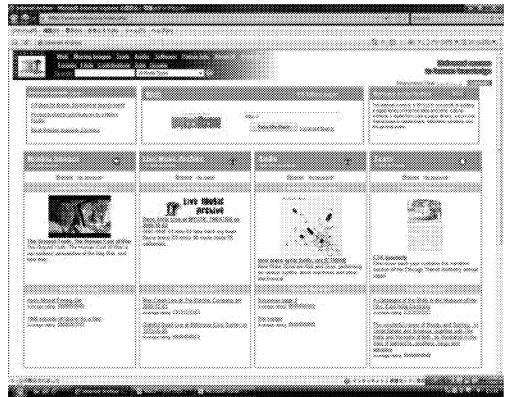


図 1 Internet Archive のトップ画面

### (3) 国立国会図書館 WARP

WARP は国立国会図書館が 2006 年から運営している Web アーカイブであり、国、地方公共団体、独立行政法人などのサイトを対象に収集している [5, 7]。2009 年に国立国会図書館法の一部改正がおこなわれ、この収集が業務として制度化されるとともに、著作権法の第 42 条の三が設けられ、このような収集において



図 2 国立国会図書館 WARP のトップ画面

は許諾が不要となった [9]。これにより今後は大幅に作業が効率化されると思われる。ただし公開には許諾が必要である。

#### (4) その他の活動

次のようなものがある。

| 名称            | 収集者          |
|---------------|--------------|
| MINERVA       | 米国議会図書館      |
| PANDORA       | オーストラリア国立図書館 |
| netarchive.dk | デンマーク王立図書館   |

### 3 Internet Archive を用いた Web ページ変遷の分析

Internet Archive の Wayback Machine では、各企業・組織のホームページの URL を指定すると、図 3 のように一覧が表示される。各年代ごとに収録されているアーカイブを閲覧しその変化等を見た。対象サイトとその URL は次のとおりである。

- 本田技研工業株式会社 <http://www.honda.co.jp/>
- 愛知県警察本部 <http://www.pref.aichi.jp/police/>
- 西日本旅客鉄道株式会社 (JR西日本)

図 3 Wayback Machine の検索結果例

<http://www.westjr.co.jp/>

- 厚生労働省 <http://www.mhlw.go.jp/>



図 4 本田技研のトップページ (1997, 2000, 2010)

(1) 本田技研工業株式会社

- (a) 表示されている製品は「四輪」、「二輪」、「汎用製品」の 3 種類にわかれており、これはほぼ調査全期間にわたって変わらない。
- (b) 「お客様窓口」が 2005 年から設けられている。また「コミュニティ」が 2002 年から設けられている。
- (c) 2009 年に「リコール修理」の窓口が設けられた。これは米国において、ホンダ インスパイア / セイバーのエアバッグ破損による死亡事故が発生したこと（朝日 2009/7/30 朝刊）を受けたものであるが、リコールに至らない「改善対策」も記載されている。
- (d) 会社情報については 2005 年に「会社案内」、「投資家情報」、「グループ」、「テクノロジー」というスタイルができてから、今日まで変わっていない。
- (e) メセナなど企業活動についてもホームページで詳しく記載されている。



(2) JR 西日本

- (a) 2005年の福知山線の脱線事故の際、トップページ全体が謝罪文となった。
- (b) 以後トップページの中央が「おわびの言葉」となって現在に至っている。
- (c) 2003 年に基本的なスタイルができてから、デザイン・色とも大きな変化はない。
- (d) 企業情報に関しては、2003年に「企業案内」「グループ会社案内」「支店情報」「企業IR情報」「法人お取引窓口」が設けられ、2008年までスタイルが変化しなかった。
- (e) 2008 年には「安全への取り組み」が設置された。



図 5 JR 西日本のトップページ (1998, 2003, 2005, 2010)

(3) 愛知県警察本部

- (a) 1997 年から 1999 年までは何のためにホームページを作成しているかのコンセプトが不

明に必要な情報を得られにくい状態であったが、1999年に「県警察案内」というスタイルができ、必

要な情報への手がかりが記載されるようになった。基本的なコンセプトはそれから今日まで変化していない。



図 6 愛知県警のトップページ (1997, 2004, 2010)

(b) デザイン的には 2004 年確立したデザインが現在も使われている。

(c) 2000 年では「運転免許」「交通安全」「警察署」など今日まで続いているスタイルが多く確立している。

(d)「捜査にご協力を」が2003年に設けられた。

(e) 2000 年から「犯罪対策室」などが設けられ、犯罪の詳細や対策の情報を見ることができる。このとき「ハイテク犯罪」が初めて取り上げられた。その後 2003 年には「地域安全」、2004 年には「街頭犯罪」の項が追加されている。

(f) 2005 年には「統計」「110番」「プレゼントコーナー」が設けられた。

(g) 2009 年には「県警ニュース」が新しく設置され、トップページで最新の更新ニュースが確認できる。

#### (4) 厚生労働省

(a) 役所の性格を反映し、「緊急情報」の項が一貫して重視されている。

(b) 旧厚生省系の情報が多く目に付く

(c) 報道発表資料と統計情報は初めから一貫して掲載されている。

(d) パブリック・コメントは旧労働省が先に掲載した。

(e) 担当業務が国民生活の他分野にわたる事から、「行政分野ごとの情報」が使いやすい。

(f) 2005 年頃から情報公開や政策評価を反映した項目が追加された。

### 5 デザインの変遷

消費者向け企業である本田技研は早くから垢抜けたデザインで、イメージも一貫していた(2008 年まで)。個々の情報を提供するというより、企業と製品のイメージを発信することに重点が置かれていたと考えられる。



図 7 厚生労働省のトップページ (1998, 2003, 2005, 2010)

JR 西日本は 2002 年までは試行錯誤的にデザインが頻繁に変わったが、2003 年よりコンセプトが確立され、現在まで使われているすっきりしたデザインとなった。しかし尼崎事故で「おわびの言葉」がページのトップに置かれるようになり、実用性が低下している。

愛知県警は 2003 年まではデザインが頻繁に変わったが、2004 年よりコンセプト確立し、現在までまったく変化がない。

厚生労働省は 2001 年創設当時は単純なデザインであったが、2004 年に実用的なデザインとなり、2008 年までは類似のデザインが続いた。しかし親しみやすさに欠けると考えたのか、現在は色数の多いデザインに変えているが、利便性は低下している。

全体として、2003, 2004 年ころから色数の少ない簡素なデザインが主流となっている。また多様な利用者に適切な情報を提供するために工夫しているといえる。

## 6 Web アーカイブの利用可能性

Web アーカイブにはすでにホームページから消去された情報も収録されている場合があり、ある企業や組織の活動の歴史を調査する上で貴重な資料となる。さらに、本研究のように経年的に比較することにより、該当機関の情報発信戦略の変遷を研究することができる。さらには複数のサイトの比較により、ホームページ・デザインの変遷の研究にも役立つと思われる。今後の研究が期待される。

## 参考文献

- [1] 喜連川優, 豊田正史, 田村孝之, 鍛冶伸裕. ボーンデジタル時代におけるウェブアーカイブとその活用基盤としての Socio-Sense. 情報の科学と技術. 2008, 58(8), 372-375

- [2] 新保史生. ウェブ・アーカイビングと法. 情報の科学と技術. 2008, 58(8), 376-382
- [3] 原田隆史. Webアーカイブの仕組みと技術的な特徴. 情報の科学と技術. 2008, 58(8), 383-388.
- [4] 柊和祐, 坂口哲男. 世界のWebアーカイブーIIPC (International Internet Preservation Consortium)を中心にして. 情報の科学と技術. 2008, 58(8), 389-393.
- [5] 武田和也. 海外動向との対比からみた日本のWebアーカイビングの課題と展望ー国立国会図書館の取り組みを通じてー. 情報の科学と技術. 2008, 58(8), 394-400.
- [6] 崔錫斗, 田窪直規. 韓国における国家知識ポータルとオンライン・デジタル資料の納本制度によるWebアーカイビング. 情報の科学と技術. 2008, 58(8), 401-407.
- [7] 廣瀬信己. Web情報のデジタル・アーカイビング:WARPを中心に. 情報管理. 2005, 47(11), 721-732.
- [8] 時実象一. 世界の知識の図書館を目指す Internet Archive: 創設者 Brewster Kahle へのインタビュー. 情報管理. 2009, 52(9), 534-542.
- [9] 文化庁. 「国立国会図書館法の一部を改正する法律」に伴う著作権法改正について. [http://www.bunka.go.jp/chosakuken/kokkai\\_toshokan.html](http://www.bunka.go.jp/chosakuken/kokkai_toshokan.html) (閲覧 2010/2/7)..

## **GMA I Sによる環境・社会・経済トリレンマ緩解論**

### **A study of Global Model Architecture Information System**

**沢 恒 雄**

**T s u n e o S A W A**

遊工学研究所 Y u k o u g a k u I n s t i t u t e

〒215-0011 神奈川県川崎市麻生区百合丘1-17-6-502

E-mail : [yrc00736@nifty.ne.jp](mailto:yrc00736@nifty.ne.jp)

連絡先著者 沢 恒雄

GMA (Global Model Architecture)概念による情報システムの開発と応用研究を10数年来にわたり研究してきた。基本的な機能は、思考支援環境、集団意思決定支援環境および合意形成支援環境の提供に有る。環境・社会・経済のトリレンマ(ESETと略す)の緩解の概念モデルの構築をし、2大モデルを提唱した。1つは、人類の種だけが保有している言語文化温存モデル、2つめは、人類生物温存モデルである。相互に密接に関連を持ち21世紀の人類の最大の課題である。2大モデルの実現の手段は、人口増加速度と工業化進度の抑制しかないと確信している。そのための最終目的がESETの緩解である。日本文化の良質性を確認して、日本のみが出来る2大モデルの実現とそのCOEとして情報発信を行うために、先人と最近の叢智を紹介して、ESETの緩解のスパイラルにより21世紀最大の課題解決の方法論を考察した。

We studied development of the information system by the GMA (Global Model Architecture) concept and applied study 10 for several years. A basic function is an environmental offer to support thought supporting circumstances, group decision aiding environment and the agreement formation. WE built the concept model of the remission of the triple handicap (ESET) of environment / society / the economy. And I proposed two major models. The first is a model keeping a language and the culture that only the human holds. The second is the human and a model keeping biology. Both have connection closely mutually, and it is the greatest problem of the human of the 21st century. The method to be realized of two major models is speed of the increase of the population and restraint of the degree of progress of the industrialization. The last purpose is remission of ESET. Two models have only Japan from the reason of the high quality nature of the Japanese culture. In this article, I assume it realization of two major models and COE, and information sends result. Therefore I introduced the wisdom of ancient people and these days. The two major models become minute by steady spiral activity. The effect will solve the greatest problem in the 21st century.

GMA I S (Global Model Architecture Information System)、人類・生物温存モデル、言語・文化温存モデル、ESET : 環境・経済・社会のトリレンマ、文化経済立国

GMA I S (Global Model Architecture Information System)、人類・生物温存モデル、言語・文化温存モデル、ESET : 環境・経済・社会のトリレンマ、文化経済立国

A human creature preservation model, a language culture preservation model, environment, economy and social mutual contradiction, a culture large country and a major economic power

## 1. 人類存亡の危機

地球に生物が存在することは奇跡的で、人類はさらに進化し2足歩行と脳の進化を遂げ言語を獲得した。言語獲得により現在の科学技術の進歩をした社会システムを構成するに至った。しかし、その人類が自らの進化の過程で存亡に関わる大きな変極点にある。産業革命後の工業化社会は、過去に比して飛躍的な成果を得たが同時に膨大な人口増加と人類生存に直接関わる環境破壊を招いた。今後更にこの傾向は続く。20世紀の工業化社会の仕組みは、大量生産、大量物流、大量消費と大量廃棄のオープン・システムであり、人類が息する自然環境はその復元力と回復力を喪失しつつある。専門領域の学者・科学者・技術者たちは、このままでは人類消滅の危機という警告を発している。今後の世界は、環境破壊による人類への悪影響、エネルギー消費による資源枯渇と人口増加による食糧難等の、ESETに満ちた世紀となる。

人類が存続するためには、自然環境の劣化具合を観測し、その修復と劣化傾向を抑制しうる環境経営が急務である。それに対応した包括的な世界システムとあらゆる組織で具体的な組織経営モデルの構築も急務である。工業化社会の負の遺産を軽減しつつ、知識社会として新たな価値観を導出し、富の配分を行い政治的、経済的な安定性を実現することである。さらに、工業化の進化率と人口増加率を緩解させる意外に方法がない。そのために、あらゆる組織が環境経営を核とした経営形態をとること、大局的な視点からは、地球環境を監視して環境を劣化させる各種の要因を制御しつつ、「自然環境の修復と回復をして復元可能な範囲」に抑制することである。

人類の消滅を回避するための必須要件は、ESET緩解である戦略的環境マネジメントシステムの概念構築と実現である。日本は、本質的にそれらの実現に近い伝統と文化を有する文化力と経済力、即ち最強の文化経済国といえる。日本の義務と役割は、文化経済構想による情報バンクを構築し、COE (Center Of Excellence) として人類存続のための情報発信により、地球経営のあり方を世界に示し、ESET緩解の重要性を諸国に認識させるために先導的な役割を担うべきである。この主張は、沢恒雄(2006)の提言である。これらを先行研究として総括的な仮説の提言に対する終章としてESET緩解論として提言する。

## 2. GMA概念と人間社会のトリレンマ

GMA概念のシステムであるGMAISによるESET緩解の具体的な方法論は、コミュニティ政策学部紀要沢(1999)から沢(2006)まで継続的に研究してきた。人類の歴史は、戦争の歴史でもあった。これは、主の存続の方法論でもある。戦争論が防衛大学でしか教えられていないことが現在の日本の平和ボケ現象である。戦争システムは、政治システムのサブセットであること程度のことは、教養課程の政治学の範疇に入れて講じられるべきである。市民と権利ばかりが強調され、国と義務をバランスよく義務教育から高等教育まで織り込まれない限り文化経済力の保持は、困難である。極貧国に対しこの保存のための施しは最低限には必要であるが、個の温存が保たれる未開発国や低開発国に対する施しは、人口増加率と工業化進捗率の抑制の認識には繋がらない。

日本は、ESET緩解の政策策定と実現に向けて最適の国である。最良の伝統・文化力と経済力を具備する文化経済大国の責務であろう。

GMAISによるESET緩解の概念総括を第2図に示した。以下、GMAISについての記述する。

### 2.1 GMAISの使用形態 沢(2006)

GMAISの使用を個人と集団に分けて、思考と討議の性格の異なる知的活動の支援環境を提供する。個人使用者は、思考支援環境を、集団使用者には、集団意思決定支援や合意形成支援環境を提供する。モデル化対象領域の制限は無いので使用の機会も森羅万象である。

GMAISの基本機能は、適用業務領域に依存しない方法を採用した。即ち、個人の使用は、思考支援機能とし、集団での使用は、使用者が討議に専念できるようにオペレーションフリーとした。情報システムの操作と討議支援環境の提供をパワー・ユーザーに委ねる方式とした。集団討議、意思決定課程に必要な情報を使用形態モードや表現形態モードのマルチモードとマルチスクリーンにより、支援環境を提供して知的活動の効率化と効果化を増大させる。GMAISの使用形態の展開を下記に示す。

1) メンタルモデルとメンタルスペースのヒューマンインターフェースは、思考支援、集団意思決定や合意形成の支援環境を提供するための情報システムGMAISとの思考空間といえる。単なる情報システムとのインタフェースではなく、メンタルスペースを使用



者が共有することができる思考空間である。サイバー・ブレインとでもいえる。ユビキタス技術でほとんどの使用者は、テレビを扱うように情報バンクにアクセスするようになる。しかし、GMAISのモデリング&シミュレーションは、地球経営の専門家集団で管理運営される。合意形成や意思決定に係わるユーザーは、政策決定や戦略策定に係わる集団である。GMAISと対話するのは、地球経営の専門知識をもつ政策立案集団とGMAISを操作可能なパワー・ユーザーである。

2) 集団意思決定支援環境は、組織活動の組織成熟度とIT成熟度による管理と環境経営への応用が組織統治や国の統治の標準となるまで人類が進化しなければならない。

GMAISは、集団討議、意思決定過程に必要な情報を多彩な使用形態モードや表現形態モード、即ちマルチモードやマルチスクリーンに可視化した情報として表示する。組織活動に必要な知的資源と知的資産の蓄積や編集機能により「個」と「集」の創造・思考支援環境を提供し知的活動の効率化・効果化を増幅する。

## 2. 2 GMAISの知的資源・資産管理 沢(2006)

知識・知恵・知謀社会における知財の運用管理の知恵を習得するには、新社会システムでの新たな活動形態や旧来のシステム再構築が前提となる。下記に示すように新社会システムに向けてGMAISの活用でそれらが可能となる。その手順を下記に示す。

- 1) 知識・知恵・知謀社会の「個と種」のモデル化と諸領域単位での統合的な日本語辞書体系を整備する。
- 2) 知識・知恵・知謀社会の知的活動で「個と組織」に有益な知的資源・資産の管理・運用法を確立する。
- 3) 知識・知恵・知謀社会における人類とESET緩解の相互関連の概念化・体系化を謀り、構築・運用する。

文化経済立国の知的資産として2大モデルの「言語文化温存モデル」と「人類生物温存モデル」の位置付けは、地球環境経営としての新しい社会システムを構築するためである。

異文化を相互理解するためには、現存する言語、即ち文化を現存する数より減少させてはならない。地球環境経営システムを具現化するには、20世紀の民主主義と市場主義を前提とした経済優先の社会システムでは人類の存続は適さない。21世紀の知識社会は、上記の2大モデルを前提にした社会システムを構築し、人類の知的資源・資産で価値あるものとして蓄積され

たモデルやシナリオ等を世界に発信し流通させ、COEとしての機能を果さねばならない。

**0 統合化辞書**：内部のDD/Sに加え外部の業務に関するメタ情報

**1 データ**：数値データやプロットに係わるログなどを含む：宇宙区間を飛翔する1水素原子から全宇宙現象

**2 ルール**：述語論理の形式や短文で表現されるセンテンス：意味を持つ有効な知識

**3 ケース**：事例を記述する情報の塊；状況に応じて問題を解決しうる解答の例と仮説・立証された例

**4 モデル**：活動や時系列的な遷移をする現象を処理の単位とした情報の塊；環境と状況に対応した問題・解決の候補と過去の事例

**5 ロジック**：論理を主体にしたソフト的な機能に業務実体などを組み込んだ情報の塊；外山氏が提唱

**6 ステディステイト**：統計処理機能に業務実体などを組み込んだ情報塊；統計学の範疇で提示された事例

**7 シナリオ**：戦略指向の計画等を記述した情報塊；孫子やCLAUSEWITZ等の戦略論をベースにしたシナリオ  
これらの1から7への順序は、下位から上位の階層をなしている。相互に包含関係にある。

## 2. 3 知識・知恵・知謀社会における

### 日本の役割はCEO

日本は20世紀末に瞬間的に工業化社会で世界一の経済大国になった。その直後に土地と株のバブル経済がはじけ算出不能な不良債権の処理を誤り、世界的な投機経済に翻弄され、その後遺症から抜けきれていない。しかし、まだ良質な文化的特性や国全体としての強大な経済力を保有している。その国力で地球経営に貢献しなければならない。

経済力、知的能力に加え、第4章で示した日本と日本人が持っている本質的特質を活かしたESET緩解は日本のみができることである。さらに、保有している知的資産と国としての経済力を利用し実態経済から乖離した博徒経済から脱却した文化経済に価値観をおく生活に変えるよう主張すべきである。

地球環境経営に関する要諦は、異文化を相互に理解し、生きることの価値観を相互に容認する。この価値観が物欲と金銭欲のみの間はこの目論見は成就しない。日本の特質を再確認して地域と地球環境経営にその拠り所を求めるべきである。新たな地球環境経営の組織化と知識社会の特質を活かしたCOE(Center Of

Excellence)としての日本のID化を謀るべきである。そのため、日本人の最も弱い戦略発想と戦略行動、さらに危機管理のマネジメントの長けた人材育成も急務である。その前提として日本と日本人のIDを取り戻し、日本の先人が残してくれた特質をDNAから顕在化させ行動化出来るようにならねばならない。

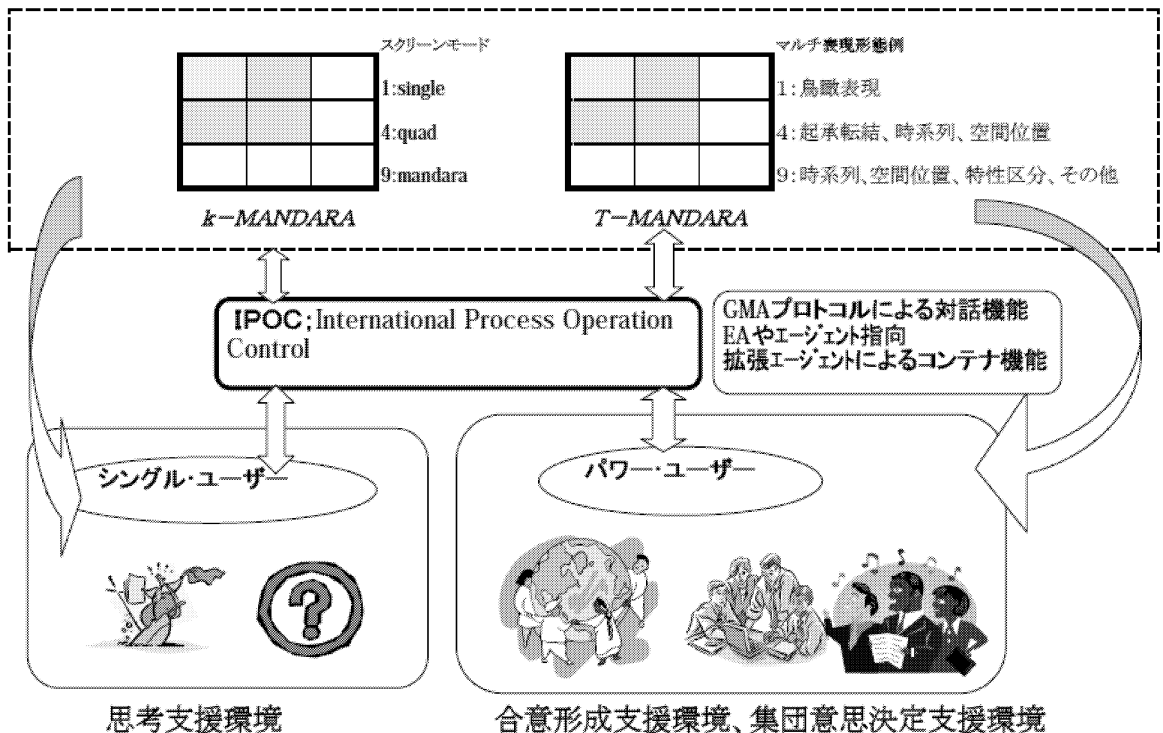
異文化の相互理解は非常に難事業であるが、今後は

その超克として相互理解が前提となってくるのである。文化の認識として日本の文化を説明でき相対化して見る必要がある。その重要性の認識と、自らの文化価値を「再発見」することを政策とし自らの文化を維持して、「武士道」精神などを発展させていくことが重要である。

第1図表 Global Model Architecture Information System :GMAISの全体図

| システム機能緒元 | 機能概要                                            |
|----------|-------------------------------------------------|
| 1. 概念    | 文化経済立国論を背景にした統合的知的資産の創製、管理・運用する情報システム           |
| 2. 機能    | 思考支援環境、合意形成支援環境および集団意思決定支援環境                    |
| 3. 特徴    | 思考過程のメンタルモデルをメンタルスペース空間(インターフェース)に展開して思考支援      |
| 4. 使用形態  | 個人の思考支援機能、集団の合意形成や意思決定支援の環境提供⇒トリレンマ緩解政策モデル      |
| 5. 知的資源種 | 情報バンクの体系的構造は、データ、ルール、ケース、モデル、ロジック、スタティステカル、シナリオ |
| 6. 統合化辞書 | 内部辞書は、情報区分の為の構造種やディレクトリー、外部辞書は、適用業務の用語辞書など      |
| 7. 多次元情報 | 表現構造はマルチモード、使用局面はマルチリテラシー&表現形態はマルチメディアとマルチリンガル  |

多様多彩マルチ 1. モード、2. スクリーン、3. リテラシー、4. リンガル、5. メディアなどの多層、多様一元的表現方式  
複雑性内包



### 3. 文化経済立国論としての自律した

#### 国の再構築 【参照：藤原正彦(2007)】

この十年来、日本は、中途半端な改革につぐ改革であった。日本改造へのアメリカの強い意志と圧力は、ビッグバン、BIS 規制から郵政民営化、三角合併解禁に至るまで、いわゆる構造改革のほとんどは、アメリカ政府が「年次改革要望書」や「日米投資イニシアティブ報告書」として日本政府につきつけたものの実施によるものである。WHO が世界でもっともすばらしいと評した、国民皆保険を軸とする日本の医療システムまでが、いま劣悪なものに改革されようとしている。

「国家の品格」を著した藤原雅彦は、経済立国に加え教養立国でなければならないと提言している。それを要約して、文化経済立国と等価な提言として紹介する。小泉・竹中が日本を壊した構造改革を、文化経済立国として再構築するための条件を考察する。

世界一の基礎学力を持った日本国民がなぜ、伝統や国柄を木端微塵に打ち砕く構造改革を支持したのか。アメリカ政府の日本に対する「年次改革要望書」などが、なぜかマスコミその他で取り上げられず、構造改革が日本のためというよりアメリカのためのものであることが陰蔽されていたことと共にマインドコントロールを狙ったアメリカの情報戦略にやられたことに原因がある。

#### 3. 1 日本人自身の2つの大きな問題

第一が経済至上主義である。戦後、焼け野原から立ち上がったわが国は、遮二無二、国土再建へと進んだ。国民の圧倒的基礎学力を背景に、民族的特性とも言える勤勉、誠実、忍耐、責任感、忠誠心などを十分に発揮した。戦時中、理工系学生を徴兵せず温存したから技術立国のための人材は豊富に残っていた。その結果、奇跡の復興をとげ、たった30年ほどで世界第二の経済大国となった。焼野原となった都市、山ばかりの狭い国土、乏しい天然資源という三重苦の中での著しい復興は、世界中の国々を驚嘆させた。1980年代に一気に世界一になり、驚嘆から感嘆、そして羨望、嫉妬、警戒、敵意と少しずつ変化した。

第二の問題は、日本人が祖国に対する誇りを失っていたことである。終戦後、大東亜戦争の敗戦で自信を粉々にされた日本人に、GHQ の魔の手が襲いかかった。厳しい言論統制の下で洗脳を敢行したのである。戦争

は一方的に日本の責任であること、戦前日本のすべては恥ずべきものであること、などがマスコミを通して垂れ流された。これは、大嘘である。

そこで誤りを是正するために「教養主義」の復活が妙手と考える。「**経済軸**」一本でやってきた戦後日本に、「**教養軸**」を加え二本柱にすべきである。教養とは文化、芸術、学問などである。

#### 3. 2 経済軸に教養軸の並立

教養がなぜ人間にとって、経済と並立するほど大切かを考察する。元来、教養は、日本の特質なのだ。

①**大局観**；日常の細々とした判断は、論理的に考えたり経験に即して考えたりするだけで足りる。しかし大局観とか長期的視野といったものはそうはいかない。例えば現代政治を表面的なものに流されずに考えようとする歴史の教養に支えられた大局観がいる。現代は過去の続きだからである。より具体的には、日米関係や日中関係を考える際、少なくとも百年前からさか上って日米中に関わる流れというものを考えざるを得ない。さらにアメリカ人や中国人の特質を知らなければならない。彼等の多岐にわたる文化や伝統、地理や気候、長所と短所なども知らねばならない。読書や体験などで得た教養が必要なのである。また、農家が潰れたら、日本の美しい田園は荒れ果てる。美しい自然こそは世界に冠たる日本文学を生み出した「もののあわれ」など美しい情緒の源泉である。これを失ったら日本が日本でなくなる。経済成長を多少鈍らせても農業を振興し自給率を上げるべきである。飽食を戒め残飯を出さない事だけでも自給率は向上される。

②**教養は人間的魅力を高める**；ギリス外交官の多くは、オックスフォードやケンブリッジで古典語や歴史を学んだ人々である。元駐タイ大使の岡崎久彦氏によると、「外交官とは、外交、政治、経済の知識はもちろんだが、哲学、歴史、文学、芸術により人間性を高め、その人間的魅力で相手と渡り合うもの」とのことである。イギリスは歴史的に見て世界でもっとも外交の巧みな国であることを忘れてはならない。明治の目覚ましい成功は、明治人の教養を抜きに語ることにはできない。日英同盟の解消は、大失敗だった。

③**教養は日本の国柄**；江戸時代の識字率は日本全体で50パーセントと推定されている。これは恐らく世界でもスウェーデンやドイツの一部地方とともにトップであろう。幕末に来日した「エルギン卿遣日使節録」の著者ローレンス・オリファントは「子供達は男

女を問わず、貧富を問わず、学校で読み書きを学んでいる。もっとも貧しい農夫でも学んでいる。彼等が我々より進歩していることは明らかと思われる」という趣旨のことを述べている。

**④教養は愉しみ**；教養を獲得するための主たる手段である読書は愉しみでもある。神社・仏閣はやがては崩れて廃墟と化し、絵画や彫刻は破損してしまうが、書物はそのままの形で生き残る。偉大な思想は時の流れと関係なく、はるか昔にはじめて作家の心に芽生えた時のみずみずしさを今も保っている。時間が演じたただ一つの役割は、悪書を追放したことである。時空を越える愉しみである。知識を得る、感動を得る愉しみである。人間は知識を得たく、感動したい生物であり、脳はそうにできている。皇紀2670年の文化に触れることで教養立国ニッポンは、再現できる。

**⑤誇り**；最近の日本の荒廃は、国民が日本人としての誇りを失ったことに主たる原因がある。これを取り戻すのは容易ではない。親や先生が「誇りを取り戻せ」とか「国を愛せ」を連呼しても効果は期待薄である。豊かな経済を得たとしても誇りにはつながらない。

日本が世界一の経済繁栄だけでは、世界が日本を尊敬することはない。世界が尊敬するものは、国民の道徳とか教養、そしてその国の産んできた文化的遺産などである。文化的遺産を産むには、広汎な教養の行き渡っていることが不可欠である。国民の間に学問、文学、芸術などへの理解、少なくとも憧憬なくして文化的遺産の創造や継承は難しいからである。道徳を保つには、国民一人一人の誇りがぜひとも必要である。誇りをなくした人間には倫理も道徳も礼節もない。自らへの誇り、国家への誇りなどがどうしても必要である。

経済は豊かな社会を実現するためであり、教養は自らを豊かにするためのものである。経済繁栄とは、これにより衣食住の向上安定や労働の軽減を達成しようとするものである。そしてこれは、そうすることで得られた自由や余暇を、文化、芸術、読書、学問といった教養の充実に向けるためのものである。

## 4. GMA I Sによる2大モデル

### 4. 1 実現のヒント

文化経済立国として日本が自律することを提案している。第3章に加えて、文化経済立国として、世界で尊敬される自律した国として新社会システム構築のためのヒントを挙げてみる。

先人が残してくれた良質の文化遺産や技術遺産を継承する社会システムが危うくなっているという現実がある。そこで日本が、振り子のように軟弱に振れすぎた振り子を元へ戻そうとする方法論的な提言である。

日本の将来は、これらの現代の良質の思想などを資源としてより自律した国として存在して、2大モデルの実現により言語、文化、人類及び生物の種の滅亡の抑止に貢献できる国となるべきである。

第2次世界大戦の大敗で大きな負の遺産を背負った国が、わずか50年で桁違いの国のアメリカに製造分野から経済力までもが、瞬間ではあるが追い抜くとは、マッカーサーも予測できなかっただろう。彼の日本統治は、旧来の日本システムの1つを残して全てを破壊した。明治憲法、教育制度、財閥、歴史までも、しかし、官僚制度だけを残した。日本統治のために必要とする膨大な資源の軽減化を謀ったと解釈している。

日本が自律・自律した国として、文化経済立国として、2つのモデルを実現していくためには新たな社会システムを再構築しなければならない。

政治家は、政官財癒着による無駄遣いを規制する制度を創ろうとせず、既得権益の擁護しか頭にない。地方の疲弊の解消が進まないのは、これらが原因である。

新社会システムの構築は、地方分権化が前提条件となる。その方法は、霞ヶ関の官僚の上から3割を地方へ返して、地方の政策立案に従事させる。

次に中央官庁の残る官僚の人事制度を「同一の省に10年以上在籍させない」制度を採用すればよい。

会計制度は、発生主義をやめ複式簿記を採用して、国民に対して年度初に前年度の実績報告と当該年度の実行計画案を説明させる制度にすれば、アングロアメリカンの走狗ではなく、自律した国として、過去の良い遺産を継承して、2大モデルの推進役として尊敬される国となりうる。そのために手段だけではなく、日本の哲学、歴史から将来を見据えることができる識者の提案や提言をジックリと習得して熟慮することが前提となる。2大モデルの早期の実現の応用できるようにそれらの候補の幾つかの主張を示した。

実体経済と賭博経済の乖離は、サブプライムローンの成果的経済の混乱をグリーンスパンに「打つ手無」と言わしめている。一方で生物学の領域では、「たんぱく質」1つ創られていない現状では、言語や脳の本質的な解明は程遠い課題である。新しいパラダムを構築して新社会システムを構築するヒントとして開かれた多重フィードバックを解明することがその重要な前提であるとしている。

## 4. 2 聖徳太子に学ぶ外交；豊田有恒(2007)

日本の官僚は、無条件で優秀であると言える。しかし、好き勝手に遣りすぎである。聖徳太子は、現在の日本を創った始祖と言えよう。1 神教の世界は、エゴ剥き出しの抗争を過去から将来まで続けるだろう。聖徳太子のすごさは、現在の日本の唯一欠点である官僚の驕りを抑える規制化と、神道と仏教を相互依存させる戦略を立て実行してみさせたことにある。何とすごい人間であったかといえる。

これで、日本の礎、活き方、考え方だけではなく近隣の大国と伍して、自律した国としての日本の仕組みとシステムを構築した英雄、豪傑と戦略家、宗教家など多彩な分野で大変革・大偉業を為した空海、織田信長等と共に数百年に1人しか出ない筆舌ににくい偉人であり、日本人としての誇りである。

ここで「17条憲法」を下記にその概要を示す。政官財癒着の糜爛した社会システムが原因で低迷している日本のあるべき姿を再構築する重大な指針となる。官僚諸君に見せたい熱望に駆られる。

(豊田有常：2007 世界に誇りうる「17条憲法」参照)

**第一条**；大いに議論すること：和を貴ばねばならない。

反逆してはならない。上のものが和やかであり、下の者たちが睦み合い、そのうえで議論を行なえば、ひとりでに理が通じ合うようになるもの。

**第二条**；仏教は民族統合の手段：篤く三宝を教わねばならない。三宝とは、仏、法、僧のことである。すべての生き物が生命の終わりにあたって帰するものであり、また、諸外国でも信仰されている教えである。

**第三条**；時にはトップダウンの決定も必要：君命を受けた場合は、かならず謹んで実行しなければならない。君主を天とすれば、臣下は地である。天は覆い尽くし、地はすべての生命を載せている。地が天を覆そうとすれば、破滅が起こる。

**第四条**；礼によって国は治まる：大臣や官僚は、礼を基本としなければならない。国民を治める要は礼にある。上に礼がないときは、巧くいかない。政府に礼があれば、後継者争いなども起こらない。国民に礼があれば、国家はひとりでに治まるものである。

**第五条**；官僚の不正を戒める：美食を求めることをやめ、物欲を捨てるようにして、問題が起こったら、訴訟によって決着をつけねばならない。資産家の訴えは、水に石を投げたように反響を呼ぶが、貧乏人の訴えは、石に水をかけたようになって貧乏人は、頼るべき相手もいなくなる。

**第六条**；勸善懲惡を徹底せよ：勸善懲惡は、古来からの規範である。善行は表彰し、悪行は懲罰すべきである。口先だけで媚びへつらう人間は、悪口、誹諷の多い人間は、上司への忠誠心に欠け、部下への配慮も持たない。大きな混乱のもとである。

**第七条**；人のために官を求めず：人それぞれに向いた仕事があるから、その職掌を守らねばならない。賢明な人材を登用すれば、称賛の声が起こるが、不適任な人間を任官させれば、災厄が繰り返され訪れる。仕事には、大きいも小さいもない。それぞれに向いた人を得て、はじめて巧くいくのである。適材適所で国家は永久に安泰になる。

**第八条**；公務員は朝から晩まで働くこと：大臣や官僚は、早く出勤し、遅く退庁しなければならない。公務は数限りなくあり、一日ではこなしきれない。

**第九条**；信念があるかどうかがすべて：信念は、道義の基本となる。どんなことでも、信念を持って当たるように。ことの良し悪しや成否は、信念の有無にかかっている。大臣や官僚が、みな信念に基づいているときは、どんな計画も成就する。

**第十条**；客観的なものの見方；内面の怒りを断ち、外面の怒りを捨て、他人の過ちに怒ってはならない。賢さと愚かさは、車の両輪のように同じものである。

**第十一条**；過失を罰しないと、どうなるか：功績と過失は、はっきりせねばならない。賞罰をかならず行なう。担当部署の責任者は、かならず罪と罰を明らかにしなければならない。

**第十二条**；公務員は懐を肥やすな：地方官吏は、住民を収奪してはならない。任命される官吏は、すべて王の僕である。地方官は、公の税といっしょに、私的に住民を搾取してはならない。

**第十三条**；公務員は手を抜くな：官吏に任官したら、自分の職掌を、きちんと把握せねばならない。病気や出張などを口実にして、なすべきことを怠るな。職掌に関わることを知ったら、会ってたしかめてから検討するようにせよ。

**第十四条**；公務員は他人に嫉妬するな 大臣や官僚は、羨んだり、妬んだりしてはならない。自分が他人を妬めば、他人も自分を妬むようになる。羨んだり、妬んだりすることの弊害は、際限がない。ほんとうの賢人は、五百年に一人しか、ほんとうの聖人は千年に一人しか現われぬ。賢人や聖人に頼れない以上どうしたら国を治められるか、よく考えるべきである。

**第十五条**;公務員は私情に流されるな：私情を捨てて、公共につくのは、公務員の道である。誰しも私情に流されれば、私怨が出る。私怨があれば、職場の一体感が損なわれる。一体感が損なわれれば、私情によって公益が害されることになり、制度や法が破られたりする。第一条にも述べたのは、このことである。

**第十六条**;国民をみだりに使役するな：国民を労役に使うときは、時を選ばねばならないと国民を使役すべきではない。農業を行なわなければ、なにを食えというのか、また、もし養蚕を行なわなければ、なにを着るというのか。

**第十七条**;独断せずに議論にはかること：大きなことは、独断すべきではない。かならず大勢でいっしょに議論すべきである。そのときは、なにが誤った論理展開がないかどうか疑ってかかり、大勢でたがいに弁論をかわせば、かならず 結論が出るものである。

#### 4. 3 その他の現在の有益な智恵

①**本質を見抜く「考え方」**；中西輝政(2007)；考え始める技術、考えを深める技術、間違いを減らす技術、世の中を考える技術、疑問を抱く技術及び情報を考える技術に分けて53の「考え方」を提唱している。

②**逆システム学**；金子勝、児玉龍彦(2007)；現在の科学技術の進歩は、西欧の要素還元論が核となっている。しかし、市場と生命のしくみを解き明かすことヒントとして多重フィードバック系を提唱している。

③**文化力 日本人の底力**；川勝平太(2006)；明治維新と日米戦争で零落になった日本が60年単位で世界1に近づくことが出来たのは、先人が残してくれた良質のソフト的な遺産である。ハード遺産が崩壊しても再起を計れる日本の底力として、過去の歴史的展望から現在の潜在的に持っている日本の文明・文化論が筆者と現在の各界の重鎮との対話で構成されている。

「美しい地球を支える美しい日本」は、終章の伊藤俊太郎との対談は、感動ものである。

④**日本のもの造り哲学**；藤本隆弘(2004)；織田信長や武田信玄は、「種子島」、鉄砲を戦闘に組み込んだ戦略家である。武田の騎馬軍団が織田軍の鉄砲隊に負けたというのは、作戦と運が勝敗を分けた歴史の1面であり、「種子島」の日本独自の開発法は、日本のもの造りのこだわりであり最高に良質な特質である。豊田を例にして、もの造りの組織能力を核にして、アーキテクチャの相性を論じている。

GMA I Sの知識・知恵・知謀の知の階層レイヤーと類似した概念として、「インテグラル」「モジュラー」「オープン」「クローズ」をアーキテクチャの基本タイプとして、産業地政学まで展開し、あるべき姿としての「もの造りの日本の進路」を提言している。

⑤**一度も植民地になったことがない日本**；デュラン・れい子(2007)；日本文化力の強度があったからオランダ、スペイン、イギリス、フランスやアメリカの植民地にならずに済んだというのが結論である。又、世界でフランスに次いで2番目に義務教育を実施した点は、教育の重要性をいくら強調してもしすぎではないことを物語る。

言語文化温存モデルの目指す初期の成果は、識字率の向上により自国の歴史と文化などを母国語から学び、それを残そうとする努力につながる。その次に、人類の置かれている現状を認識する過程へと「思考」が成長し、人類が地球上から消え去る可能性のある現実を知り、人口増加率の抑止力にはなるのである。

⑥**欲望する脳**；茂木健一郎(2007)；茂木健一郎は、テレビに出すぎの感があるがクオリアの研究では第1人者である。人間の本質的に保有する“欲望”について際限のない欲望、欲望と社会について考察して生命哲学として孔子の「七十従心」は「無為自然」と等価であるとして、従来の人間や自然を機械論的な世界観を超えた宇宙の不可思議な場所の複雑な豊穡をより深く理解するためのヒントであるとしている。脳のメカニズム研究の困難さが垣間見える結論である。

必読すべき書を参考文献に記したが、これらはまさに日本と日本人の遺産としての良質な文化・文明の特質を温故知新たるべくすべきである。

## 5. 結 言

言語文化温存モデルは、沢(2006)で、適正人口の定着化を狙った識字率向上の教育浸透と環境を考慮した「生き・活き・逝き」モデルを提案した。

知識社会で環境問題を解決する前提となる意識革命のためのマルチメディア教育を採用した。また、多様な文化を知的資産に変換するためにオントロジカル概念による統合化辞書の考え方を提案した。モデル各層の辞書体系を整備して、各層間の関係を辞書の連携により意味付けすることでシステム構築ができる。文化と言語を温存し、廉価で情報発信をすれば、識字率の向上につながる。

結果として異文化の相互理解と人口抑制と、工業化進度の制御の必要性、さらに富の配分の重要性が理解できるようになる。

環境問題を意識した日常の生活が環境問題解決へむけて変化する。敗戦後60年間のアメリカ研究から脱してこの間に忘れてしまった日本研究をして、連帯を地球と地域にもとめ、人類が「身丈けの生活」に気づき、異種文化を温存、理解、統合化を指向する。そのために日本は、2大モデルを、世界にCEOとして情報発信をする。

その結果として異文化の相互理解と人口抑制、工業化進度の制御の必要性を認識して、個々の生きる上での底なしの沼のような欲望を抑制し行動する意識改革ができる。この時点で最終目標とする2つのモデルが実現してESET緩解が円滑に初めて可能となる。

人類に課せられた最終的な研究課題は、言語、頭脳とDNAの全解明と核融合発電である。人類にとって至難の課題の解明の前になすべきことは、2大モデルの実現がその前提となるであろうことを確信する。

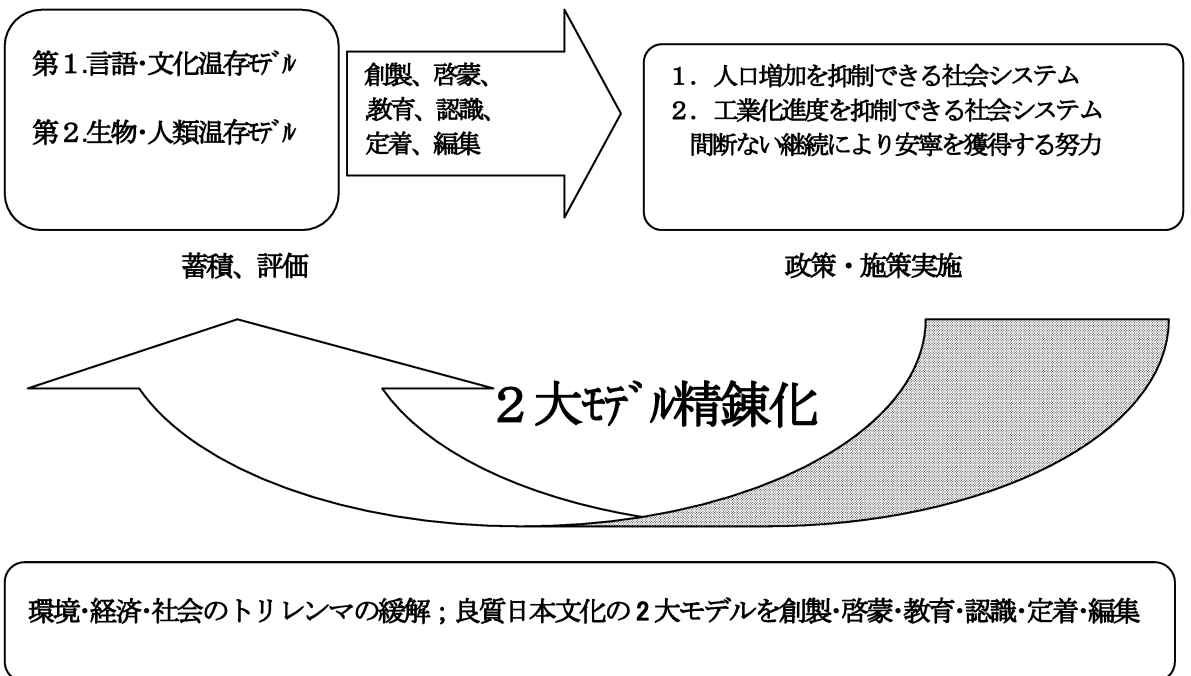
(第2図)

以上

## 第2図 GMAISによるESET緩解の方法論；ICT社会の知的資源・資産の蓄積、流通と啓蒙

スパイラル的に2大モデルの実現への概念図（日本のCOEが世界平和に貢献し尊敬される国になる）

|                                                       |
|-------------------------------------------------------|
| 日本良質文化の定着；先人の偉大な遺産である良質日本文化は哲学・思想・統治の諸相で「和」と「平和」を根底   |
| ①核兵器廃絶；世界で唯一の被爆国として廃絶を主張できる唯一の国である。何故オバマが主張できるのか疑問    |
| ②銃火器廃絶；秀吉の刀狩の制度と儒教・朱子学や武家諸法度(武士道と匠)に育まれた日本人のDにアイシユル驚愕 |
| ③地雷廃絶；ODA等で廉価な探索技術を利用した先進的ボットを開発し100年かかる期間を30年に短縮化する  |
| ④麻薬廃絶；制度と慣習(社会倫理)面で水際防御を実現済である。違反した者の加熱報道がその証拠である。    |
| ⑤原理主義・多神教の相互容認；日本では、既に聖徳太子・空海・信長により哲学・思想・統治面で達成済み社会   |



## 引用文献

- 沢恒雄(1997) “知識時代の経営情報システム論”  
白桃書坊 東京都, 59pp.
- 沢恒雄(1999a) “知識社会における知的資産創製と  
管理の研究” 愛知学泉大学 紀要 第1号 67-95.
- 沢恒雄(1999a) “異分野統合型学部の情報教育・マテリアル  
-教育” 愛知学泉大学 紀要 第2号 37-51.
- 沢恒雄(2000) “GMA概念モデルによるマテリアル-教育の  
研究” 愛知学泉大学 紀要 第3号 91-110.
- 沢恒雄(2001a) “知識・知恵・知謀社会における新組織  
愛知学泉大学 コミュニティ政策研究 第3号 95-111.
- 沢恒雄(2001b) “文化経済立国論(構想編)”  
愛知学泉大学 紀要 第4号 45-67.
- 沢恒雄(2002) “文化経済立国論(環境経営システム編)  
“ 愛知学泉大学 紀要 第5号 53-86.
- 沢恒雄(2003a) “戦略的地球環境経営システムの研究 “  
英国ウエールズ大学 修士論文
- 沢恒雄(2003b) “戦略的地球環境経営システムの研究 “  
愛知学泉大学 紀要 第6号 27-61
- 沢恒雄(2004a) “第4セクター方式組織の形態論考 “  
愛知学泉大学 政策研究 第6号
- 沢恒雄(2004b) “GMAISによる文化言語温存モデル構築法の  
研究 “ 愛知学泉大学 紀要 第7号 45-66
- 沢恒雄(2005) “GMAISによる統合化CSR管理システム(TCMS)”  
愛知学泉大学 紀要 第8号 43-57
- 沢恒雄(2006) “GMAISにおけるモデル・シナリオの研究 “  
愛知学泉大学 紀要 第9号 73-86
- 藤原正彦(2007) 教養立国ニッポン 文芸春秋  
平成19年12月号 94-108
- 中西輝政(2007) 本質を見抜く「考え方」  
サンマーク出版
- 金子勝、児玉龍彦(2007) 逆システム学 岩波新書
- 豊田有恒(2007) 聖徳太子に学ぶ外交； 祥伝社新書
- 川勝平太(2006) 文化力 日本人の底力 ウエッジ
- 藤本隆弘(2004) 日本のもの造り哲学  
日本経済新聞社
- デュラン・れい子(2007) 一度も植民地になったことが  
ない日本 講談社新書
- 茂木健一郎(2007) 欲望する脳 集英社新書

\*\*\*\*\*

## 参考文献

- 野村(2007) 「千年、働いてきました」、角川書店
- 出井伸之(2007) 「日本進化論 2020年に向けて」  
幻冬舎新書
- ヨアヒム・ヒルシュ(2007) 「国家・グローバル化・  
帝国主義」
- 永井潜(1913) 生命論 洛陽堂
- 新渡戸稲造(1899) 武士道
- 高坂正堯(1981) 文明が衰亡する時 新潮選書
- 奈良本達也・梅原猛・陳瞬臣・渡部昇一(1980)  
歴史から今なにを学ぶか 講談社
- 石原莞爾(1940) 世界最終戦論 立命館出版部
- 菊池 寛(1940) 二千六百年史抄 同盟通信社
- 草柳大蔵(1979) 実録満鉄調査部 朝日新聞社
- 秋永芳郎(1997) 満州国 虚構の国の彷徨 光人社
- 太田尚樹(2005) 満州裏話 講談社
- 小室直樹・色摩力夫(1993)  
国民のための戦争と平和の法 総合法令
- 永(2009) 日本人でいる奇跡 文芸社

追記；大東亜共栄圏や満州国の桃源郷は、戦略レベルでも最高の構想だった。当時の大戦略家の石原莞爾、北一輝、甘粕正彦、岸信介などの**日本の人的至宝**が当時の時代背景の中で十分に地位と権限を与えられなかったのが日本の悲劇であった。満州国建設の功績は、毛沢東さえ、「中国は要らない、但し東北さえあればよい」と言わしめた。(東北とは満州国をさす)。満州国の建設の実績は、敗戦後の昭和20年以降の日本の経済的な発展のベースになったのである。例えば、新幹線建設を推進した十河信二、日米安保の堅持に対する岸信介の眼力と腹の据わり方には敬服する。また、文化面での映画産業は、殆ど甘粕正彦が起こした満映で育まれた人材によるものであった。

「匠」の国、日本が製作した大傑作「戦艦大和とゼロ戦」がなかったら日本は、アメリカと戦争をしなかった。また、ミッドウエー海戦で敗退時が戦争終結の時期であったが、それを出来る人材が不在であった事を熟慮すべき。歴史には、「IF」があり得ないのだが・・・



## 「社会変革の一般的構造」

### Generalized structure of social revolutions

安平哲太郎<sup>1\*</sup>

Tetsutarou YASUHIRA

1 産業技術総合研究所

Advanced Institute of Science and Technology

〒305-8568 茨城県つくば市梅園 1-1-1 つくば中央第2

E-mail : tyasuhiro@zf7.so-net.ne.jp

\*連絡先著者

「本来同じであるべき事」を同じと認識し、「本来異なると認識すべき事」を異なると認識していない時矛盾が生じ、異なると認識した時矛盾が解消する事を社会の変革に適用することによって以下の事を明らかにした。すなわち、人間社会には一人の人間がとり得る幾つかの立場がある事。そして、通常は必然的にそのどれかの立場をとって共存しているが、やがて何らかの働きによって本来同じであるべき事を同じと認識することで矛盾が始まる事。さらにその認識が共有され、矛盾の原因が差別や不公平をもたらす立場の違いやそれを支えている制度にある事に気づく事で認識上の矛盾は解決する事。やがて同じという認識が必要な事態があつて、かつその認識が通用しない自他の立場故に事態を解決できない事に気づいた時、その認識が通用する立場へと社会の変革を非可逆的に始める事。しかし、同じという認識とそれぞれの立場の距離の違いから変革に際しては変革の抵抗勢力と変革の推進勢力とが出現し混乱を招きやすい事がわかった。

社会の変革をこのように見る意義は、これから迎える様々な変革に対して社会がどの段階にあるかを理解し備える事が出来るという点である。

We could make the following thing clear by applying the conditions causing and solving contradictory to some social revolutions:

1. There are some positions a human can put himself in and normally he put himself in any position and coexist with other those who put themselves in the same and other positions.
2. Presently they begin to feel the contradictory between the new recognition and the present situation by recognizing as the same thing what he didn't recognize as the same thing.
3. Further they can solve the contradictory by understanding that the cause of contradictory is the differences of old positions bringing discrimination or unfairness because of what new recognition cannot apply to and the social institution supporting them .

4. Shortly the situation which needs the new recognition appears and when they understand that they can not resolve it because of their old positions, they begin to revolutionize their society irreversibly to the new position which their new recognition can apply to.
5. But the difference of their distances between the new recognition and their old positions cause prompting power for revolution and resisting power against revolution and it can become the origin of the social confusion.

The significance which understand the revolution of our society in this way is to be able to know what stage our society is in when we encounter various revolutions and to prepare for.

キーワード: 社会の変革、認識の変革期、社会制度の変革期、差別、不公平さ

Social revolution, Period of revolution for recognition, Period of revolution for social institution, Discrimination, Unfairness

## 1 はじめに

我々は情報に接し知識を獲得する事によって、今までは同じと認識していなかった対象を同じと認識するようになる事がある。ところが往々にして現状からその新しい認識が通用しないリアクションを受けて新しい認識との間で矛盾を感じる。2009 年の情報知識学会で「完全合理性と限定合理性」[1]という題で認識の矛盾の解消には「本来同じであるべき事」と「本来異なるべき事」に気づく事の重要性を指摘し、この事は、奴隷制からの解放、君主制から主権在民、明治維新等の社会変革と密接な関連がある事を指摘した。一方、筆者は 2008 年の情知学会誌に「科学の時代から科学を基礎とした神の時代へ」[2]という題で投稿し現代は時代の変革期と考えられる事を指摘した。したがって、今ほど、現在は時代の変革期の中のどの段階にあって何をしなければならないのか知る必要に迫られている時はないと言える。一方、同じであるという認識は社会や制度によって守られているうちはいいが、一旦それが緩めば経済的利益が優先してその認識を貫き

通す事が困難である事が多い。例えば、現代でも国際社会において自然災害の被災地などでは混乱に乗じて子供の奴隷化などが見られる。この事は獲得した同じであるという認識は混乱を伴いやすい変革期にこそ社会や制度によって逆行しないようにしっかりと支えられなければならないことを意味する。したがって、この認識上の矛盾の解消を奴隷制からの解放、君主制から主権在民、身分制からの解放という幾つかの社会変革に適用して、人間社会の変革の一般的な構造を明らかにし、現在進行中の環境問題を変革という観点から位置づけ直し、本格的な変革が到来する前に、人類が今までに同じと認識してきた「本来同じであるべき事」の内、基本的な事を再確認しておく事は意義があると考えられる。そこで今回はこの事について考察する。

## 2 社会変革期前後の立場の変化

社会変革の場合、社会の変革期という期間が存在し、その変革期の前後の安定した時期に立場の異なる社会が存在する。社会の変革期以前の安定した時代には、本来同じとみな

すべき事を同じとみなさない事によって可能な複数の立場が存在し、個人が必然的にそのどれかの立場をとって共存している。それに対して、変革期以降の安定した時代には、今まで同じとみなしていなかった事を同じとみなした立場を全員がとって共存する。以下、この点について具体例で考察する。

## 2.1 奴隷制度とそこからの解放

奴隷制度の場合、人間がとっている立場とは人間の所有者、被所有者（奴隷）と自由人と考えられる。すなわち、全ての人間は人間の所有者か被所有者、あるいはそのどちらでもない立場になり得る可能性があり、しかも、個人の能力、体質、負債、先祖から背負ってきている負債等によって必然的にそのどれかの立場に立ち、被所有者（奴隷）を同じ人間とみなしていなかった時代と考えられる。それに対して、社会変革期を経て奴隷制度が廃止されると「全ての人間を同じ人間」とみなし、全ての人間がそのように扱われる（すなわち、人間の所有者や被所有者にはなれない）立場をとる状態になると考えられる。

## 2.2 君主の時代から主権在民の時代へ

君主の時代における立場は、支配者（国王）と被支配者と考えられる。すなわち、全ての人間は支配者か被支配者になり得て、しかも、個人の能力、体質、負債、先祖から背負ってきている負債等によって必然的にそのどれかの立場に立ち、被支配者には主権はない（あるいは制限されている）と考えられていた時代と考えられる。それに対して、社会変革期を経て君主制が廃止されると「全ての人間に同じ主権がある」とみなし、全ての人間をそのように扱う（すなわち、支配者や被支配者にはなれない）立場をとる状態になる事

と考える事が出来る。

## 2.3 江戸時代の身分制から明治維新へ

江戸時代の身分制における立場は、士農工商の4つの身分と考えられる。すなわち、全ての人間が武士か農民、工人、商人のどれかになり得て、しかも、個人の能力、体質、負債、先祖から背負ってきている負債等によって必然的にそのどれかの立場に立ち、能力の発揮がその身分に限られていた時代と考えられる。それに対して、社会変革期を経て身分制が廃止されると、「全ての人間が同じ能力を持っている」とみなされ、全ての人間をそのように扱う（すなわち、個人の能力の発揮が身分によって制約されない）立場をとる状態になる事と考える事が出来る。

## 2.4 産業育成の時代から環境保全を考慮した産業育成の時代へ

産業育成の時代の立場は、全ての産業のありようと考えられる。すなわち、全ての人間がどれかの産業を選択でき、しかも、個人の能力、体質、負債、先祖から背負ってきている負債等によって必然的にそのどれかの立場に立ち、しかも自然環境保全を考慮していない時代と考える事が出来る。それに対して、社会変革期を経て自然環境保全を考慮するようになると、全ての人間が等しく自然環境保全を考慮した産業を行わなければならない立場をとる状態になると考えられる。

以上の考察から、社会変革というのは、同じと認識していない故に存在しえた幾つかの立場から、社会変革期を経て等しく同じと認識する立場をとる社会となる事として理解される。ただし、本来同じと認識すべき事を同じと認識する立場とは、奴隷制の場合は「全ての人を同じ人間とみなす」立場であり、

君主制の場合は「全ての人に主権があるとみなす」立場であり、身分制の場合は「全ての人が身分に制約されないで同じ能力を発揮しうる」立場であり、産業育成の場合は「全ての人が環境保全を考慮した産業に従事する」立場である。

### 3 社会変革期の構造

社会変革は、社会変革期を挟んでそれ以前の安定した時代から、それ以降の安定した時代への移行であることを説明した。ここでは、社会変革期はさらに認識の変革期と社会制度の変革期に分けられる事を説明する。

#### 3.1 社会変革のために「同じであるべき事」に気づく事の必要性

社会変革の場合、全ての人が幾つかの必然的に決まっている立場から全員がある特定の立場をとる方向に向かうためには、その方向へ向かう前に、本来同じとみなすべきなのに同じとみなしていなかった事を同じと認識する必要がある。すなわち、奴隷制度の場合は「全ての人間を同じ人間とみなす」、君主制の場合は「全ての人間を同じ主権者とみなす」、身分制の場合は「全ての人間が同じ能力を持つとみなす」、産業育成の時代には「すべての人が同じ環境保全を考慮する必要があるとみなす」という認識である。そして、この事によって、実際にはその様に扱われていない(奴隷が同じ人間として扱われていない、被支配者は同じ主権者として扱われていない、身分に制約され同じ能力を持つ者として扱われていない、種々の産業が同じ環境保全を考慮していない)現状との間で矛盾を感じるようになるのである。例えば、福沢諭吉は父親が中津藩の借財を扱う要職にあり、儒教に通じた学者でもあったにもかかわらず、身分制によって名を成すことも出来なかった事

から身分制の差別に反発を感じていたと言われるがそのような事である[3]。

#### 3.2 同じという認識が通用しない為に差別と不公平をもたらす立場の違いとそれを支える制度が矛盾の原因である事に気づく必要性

「同じであるべき事」に気づいても、その同じであるように扱われていない現状から矛盾や対立を引き起こす事になる。そこで、その矛盾や対立を解消するためには、その様な扱い(差別や不公平)を可能にする立場の違いとそれを支えている制度に気づく必要がある。例えば、奴隷制度の場合すべての人間を同じ人間とみなしても、その様に扱われない(差別を受ける)現状から矛盾や対立が生ずる事になる。差別をする立場の人間の所有者と差別を受ける被所有者という立場の違い、および、それを支える奴隷制度が矛盾の原因である事に気づいて初めて奴隷制の廃止という向かうべき方向がわかる。君主制の場合は、全ての人が主権を意識してもその様に扱われない現状から矛盾や対立が生ずる事になる。差別をし不公平を感じさせる支配者と差別され不公平を感じる被支配者という立場の違い、および、それを支える君主制が矛盾の原因である事に気づいて初めて君主制の廃止という向かうべき方向が分かる。身分制の場合は福沢諭吉と勝海舟とが咸臨丸で渡米し身分制崩壊後の世界を知る事によって、各身分の扱いの違いとその差別を支える身分制に矛盾の原因がある事を知り身分制を解消させなければならないという方向を意識したと言える。産業育成の場合は、環境保全を認識しても、環境保全を考慮せずに資源をとり廃棄物を出す産業の時代には産業活動と自然環境保全とは矛盾・対立する事になる。自然環境に公共性があると考えれば再生の困難

な資源を無償で利用する産業は不公平という事になり、そういう産業とその様な経済活動を許容する社会制度が矛盾の原因である事になり、全ての産業が環境保全を考慮しなければならない社会制度が向かうべき方向としてわかる。

以上の考察から、本来同じであるべき事を同じと認識するようになる事で認識の変革が始まり、同じと認識していなかった為に差別や不公平さをもたらしてきた立場の違いやそれらの立場を支える制度が矛盾の原因である事に気づく事で認識の変革は終了する。そして、重要な点は認識の変革が終了しただけでは社会を変革する力にはならないという事である。社会制度の変革は、同じと認識する事を必要とする事態が出現して初めて社会の変革が始まるのである。以下ではこの点を考察する。

### 3.3 社会制度の変革に着手せざるを得ない事態の出現

社会変革の場合共有した新しい認識が必要なる事態が来て、さらに、新しい認識が通用しない社会の現状故に事態の解決が出来ないことに気づいた時、新しい認識が通用する社会へと社会制度を非可逆的に変化させ始める。すなわち、リンカーンの奴隷解放でいえば、南部をアメリカ合衆国につなぎとめておく為の戦いで黒人を南部から切り離す必要があったが、南部は綿栽培の為に黒人奴隷を必要としたためリンカーンは黒人の奴隷解放に踏み切った[3]。フランス革命でいえば、国家財政の立て直しのために重税を課していた平民からの課税は無理で特権階級に課税するという平等に課税をする必要が出てきても、その努力が特権階級の反対によって失敗した時、平民が身分制の破壊に向かった場合がそうである[3]、明治維新でいえば、下関戦争、薩英戦争等で欧米列強の

力を知ったことで身分制に束縛されずに能力を発揮し欧米列強に迫っていく必要が出てきても、幕府の封建体制がその妨げとなっている事に気づいた時、倒幕派が身分制を含んだ封建体制の破壊に向かった場合がそうである[3]。また、産業の場合でいえば、全員が環境保全を意識する事が必要な事態が来て、環境保全を考慮していない産業の現状では事態を解決できない事に気づいた時、環境保全を考慮した産業育成の方向に社会制度の変革が始まる事になるが、現在はまだそういう切羽詰まった段階に達していないと考えられる。

## 4 社会変革期のその他の特徴

### 4.1 社会制度の変革において古い立場の差別や不公平さの違いがもたらす対立

同じであるべき事を同じとみなした新しい認識が古い立場では通用しない度合いを新しい認識と古い立場との距離と定義すれば、その距離はその立場によって異なり、場合によっては方向が逆でさえあり得る。例えば、奴隷制度のもとでは、「同じ人間」という認識と、奴隷、奴隷の所有者、自由人との間では、奴隷の所有者は「同じ人間」という認識を通用させるためには奴隷を失わなければならない。この事は経済的損失を招くため、「同じ人間」という認識が通用する社会への変革に最も抵抗するだろうし、奴隷に「同じ人間」という認識を通用させるためには奴隷という立場から解放しなければならない事になる。この事は経済的利益に繋がりやすいため、「同じ人間」という認識が通用する社会への変革を最も望むと考えられる。また、君主制の場合は、「同じ主権者」という認識と、支配者（君主）、および、被支配者（民）との間では、君主は自分の主権で民を動かしていた強すぎる主権の立場から

主権を弱めなければならない。この事は経済的な不利益となり、「同じ主権者」という認識が通用する社会への変革に抵抗するであろうし、民衆は命令に従わざるを得ない主権のない立場から主権を獲得する事になる。この事は経済的な利益となり、「同じ主権者」という認識が通用する社会への変革を望むと考えられる。産業育成で言えば、資源を無償で獲得し廃棄物を投棄する現在の産業の内、環境保全を考慮する事が容易な産業と困難な産業とがある。すなわち、再生が容易な資源を用いる産業は環境保全が容易と考えられ環境保全を考慮した社会への移行には小さな経済的な負担で済むだろうし、再生が困難な資源を用いる産業は環境保全が難しく環境保全を考慮した社会への移行には大きな経済的な負担を伴う事になる。このように立場によって、新しい認識との間の距離が違うのである。そして、新しい認識に対してこの距離の違いが奴隷制や身分制や君主制の場合は差別を与える方と受ける方の方向の違いとなるし、産業育成の場合は自然環境に公共性があると考えれば不公平さの違いとなるのである。そして、この差別や不公平さの違いが実際に社会変革に着手し始めた場合には変革を推進する勢力と変革に抵抗する勢力を作り出し対立が起きると考えられる。

## 4.2 認識の深化に伴う遅延時間の存在

全ての人間が本来同じと認識すべき事を同じと認識し始め、その認識が通用しない社会の現状故に矛盾を感じ、差別と不公平さをもたらす立場の違い、およびそれを支える制度に矛盾の原因を見出し、認識上の矛盾を解消するまでの過程はいわばコヒーレントに全員が足並みをそろえて進んでゆくわけではない。それぞれ遅延時間を持っているのである。

以上で、認識における矛盾の解消という観点から社会変革を考察した。

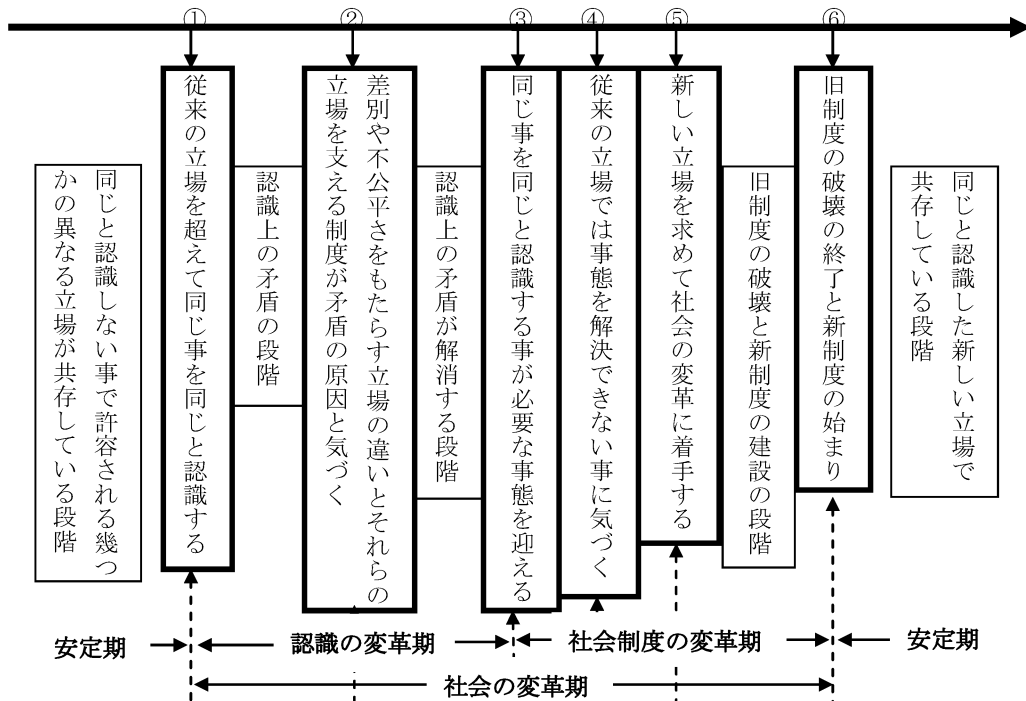
## 5 社会変革の一般的な流れ

次にこの社会変革における一般的な流れを明治維新と対比させながら図1で説明する。

図1の上段は社会変革の一般的な流れを表し、下段は明治維新の変革の流れを表す。上段の太い横線は時間の流れを表し、横線から下方に向かう矢印によってある時点で起きている事象を表す。

① 以前は、変革前の安定期であり、①から⑥が社会の変革期であり、⑥以降が変革後の安定期である。さらに、社会の変革期の中でも①から③が認識の変革期であり、③から⑥が社会制度の変革期を指す。

1) 社会変革の一般的な流れ



2) 明治維新における変革の流れ

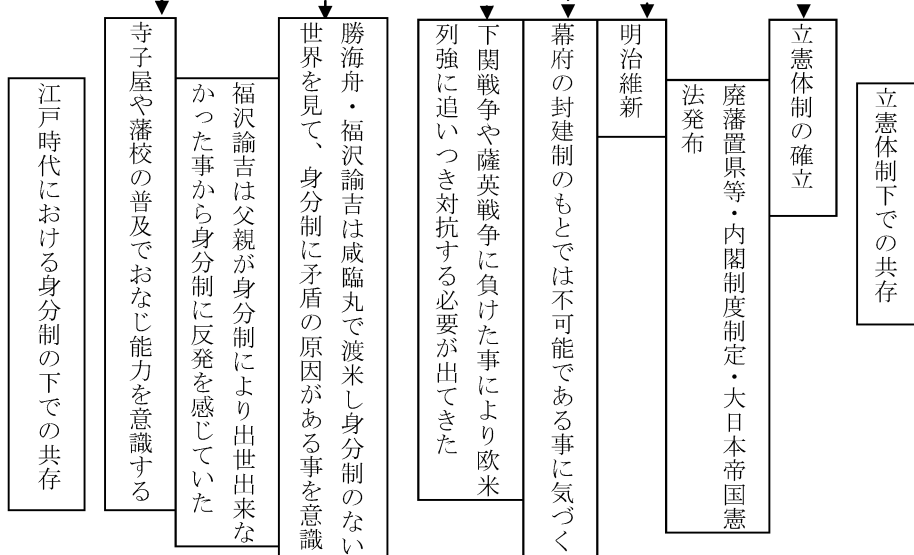


図1 1)において、認識上の矛盾の解消も含めた社会変革の一般的な流れを示す。太い横線が時間の流れる方向を示す。縦実線の矢印がその時点で起きている事象を示し、横実線はその期間の特徴を示す。2)は対応する明治維新の流れを示す。

## 6 まとめ

認識における矛盾の解消という観点から、社会変革を考察する事によって以下の事を明らかにした。社会変革とは本来同じ事を同じと認識せずしかもそれが社会制度によって支えられていた時代から、社会変革期を通して本来同じ事を同じと認識しそれが新しい社会制度で支えられる時代へと移行する事である。社会変革期にはまず、認識の変革期がありそれは今まで同じと認識していなかった事を同じと認識し矛盾を感じる事から始まり、同じと認識していなかった事による差別を社会制度が支えていた事に矛盾の原因がある事に気づく事で終わる事。そして、社会制度の変革は同じと認識する事が必要な事態を迎えて初めて同じと認識する事が通用する社会制度へと変革を始める事、旧制度の破壊が終了し、新制度の建設が一段落する事で終わる事が分かった。

社会の変革をこのように見る意義は、社会制度の変革が始まるまでに認識の深化と共有が必要であるという点を理解できるという事である。明治維新とかフランス革命とかはこの社会制度の変革からを言う。また、二番目は我々がこれから様々な変革を迎えるにあたって、我々の社会が変革のどの段階にあるかを知り、備えをする事が出来るという点である。環境問題でいえば、環境保全という概念自体がまだ公害問題に見られるように人間に対する有害物質を産業が放出しない、また、温暖化問題に見られるように温暖化物質を産業も個人も放出しないという意味であり、資源取得や廃棄物排出に対する保全ではない。しかし、現在の産業活動を始めとする人間活動の在り方が持続発展可能な開発と対立する事は認識されつつある[4]。したがって、環境保全が産業も

含めて誰もが等しく取り入れるべき事として認識され、かつ、その内容が特に資源取得や廃棄物排出に対する保全へと変わってゆく、認識の変革の途上の段階と考えられる。三番目は人類が獲得してきた基本的な「同じ」という認識は、奴隷制、君主制、身分制からわかる通り「同じ人間」、「同じ主権者」、「同じ能力」であり、これから迎える変革において社会や制度によってしっかり守られる必要があるという事が分かる点である。

## 謝辞

東邦大学名誉教授 近藤淳博士から現実と問題とが噛み合う事、そして、それまで待たなければならない事を教えて頂いた。待つ事は容易ではなかったが、環境問題に直面してこれがいかに大切であるかを考えないわけにはゆかない。この事を教えて頂いた事に心から感謝します。

## 参考文献

- [1] 安平哲太郎：「完全合理性と限定合理性」情報知識学会誌, Vol.19, No.2, 2009.
- [2] 安平哲太郎：「科学の時代から科学を基礎とした神の時代へ」, 情報知識学会誌, Vol.18, No.2, 2008.
- [3] 福沢諭吉, 南北戦争, フランス革命, 幕末, 明治維新：フリー百科事典『ウィキペディア (Wikipedia)』  
[http://ja.wikipedia.org/wiki/\(2010年4月1日参照\)](http://ja.wikipedia.org/wiki/(2010年4月1日参照))
- [4] 吉川弘之：「科学者の新しい役割」, 岩波書店, 2002.



## 人間の積極的に生きる能力を高めるために

### A study how one can enhance one's ability actively to survive changing environment positively

福 永 征 夫

Masao FUKUNAGA

アブダクション研究会

Abduction Research Institute

〒610-0357京都市京田辺市山手東1丁目28-4

E-mail: [jrfdfl17@ybb.ne.jp](mailto:jrfdfl17@ybb.ne.jp)

人間が積極的に生きる能力を高めるためには、環境の変化に対応して、生きるための新たな知識を発見し、知識を進化させ続けることが肝要である。脳の情報処理の機能を維持し向上させて、新たな知識をまとめ上げ、知識を進化させるためには、『既存の領域的な知識をベースにして、新たな領域的な知識を探索し、それらを広域的な知識に組み換える』ことがポイントになる。

To enhance one's ability surviving changing environment, one should continue to get and compose new knowledge that is necessary to adapt new situation in front, and at the instance one should try to evolve the connection between different knowledge in mind.

To achieve the above information processing behavior effectively, it is the point that one could connect the newly got special-domain-knowledge with the existent special-domain-knowledge, and could recompose them to broader-domain-knowledge.

キーワード : 記憶のネットワーク      領域的な知識      広域的な知識      <知><情><意>の情報の系

keyword : network of memory      special domain knowledge      broader domain knowledge  
information system of intellect, emotion and volition

#### 1. 人が積極的に生きる能力を高める

- ① 国内外の厳しい社会環境の中で、精神的な悩みを抱えて苦しむ人が増えている。
- ② 社会の高齢化の度合いも高まっている。
- ③ 人間の積極的に生きる能力を高めなければならない。
- ④ 明晰で判明な理解と行動の能力をいつまでも維持することが重要である。

⑤ それには、環境の変化に対応して、生きるための新たな知識を、その人なりに発見し、知識を進化させ続けることが肝要である。

⑥ 生きるための知識には、知・情・意の三位一体性を具備することが求められる。

⑦ 筆者は十数年来、自然のシステムにおける『ラティスの情報処理モデル』(Lattice Information Processing Model of Systems in Nature)

を数理的に考案して提案し、人間の知識が発展し進化する仕組みを説明している。

⑧ 脳の情報処理の機能を維持し向上させて、新たな領域的な知識をまとめ上げ、知識を進化させるためには、『既存の領域的な知識をベースにして、新たな領域的な知識を探索し、それらを広域的な知識に組み換える』ことがポイントになる。

## 2. 深刻な記憶の連続性の障害

TV で見た日本の若年性認知障害者グループの海外交流での会話。オーストラリアの人が日本のある人に『あなたはやっていることに、なお連続性を見出しているから幸せだ』と言った。今の自分と過去の自分の記憶、自分と他者の関係の記憶、場所や事物や情報の記憶、などの断絶にも繋がりがかねない記憶の連続性に関する障害を背負っている人が、中高年者の間に稀ではない頻度で発生している。これは、1970年代から今日までの40年の間に生じた現代人の生活環境の余りにもドラスティックな変化と、それへの対応のなされ方の中に、原因を探る糸口があるかも知れない。

## 3. 機械化・情報化と生身の人間の能力

1960年代以降に大きく進んだ家庭生活や職場労働の情報化・自動化・機械化は、人間の生活や労働の負荷を軽減し効率と安楽の増加をもたらした。また、近時における社会全般のデジタル・バーチャル革命は、社会に計り知れない効用をもたらした。こうした機械化・情報化技術の輝かしい光の部分の反面、前頭前野の46野の失調症というボタン制御の作業習慣やデジタルワークの過剰に関わるかも知れない、プロセスを制御する行動の能力の不調に苦しむ人が現れている。また、40～50代の人の中に、人に対する意思の一方方向的発信はできても、相手からの意思表示を上手に受信して、相手に得心を与えることの不得手な人が確実に増えているようにも思われる。更には、社会の人々の間で、

対話と言葉が貧困になって、他者と意味を共有して共感し、抽象をし、汎化をして、合意をつくるという能力が退行する傾向にあるようだ。今までの機械化・情報化技術の影の部分の影響が顕在化してきたのであろうか。

## 4. 知・情・意の情報の順序の配位

人間は過去を想起し未来を想像し予期して、現在に対処する営みにおいて、＜知＞という事実・＜情＞という価値・＜意＞という目的の情報の各々に関する、時間・空間的な順序の起・承・転・結の配位に、生存に意味のある、リスクとチャンスの、新たなストーリーを見出す。この新たな領域的な知識のアイデアが、試行と実行の後に、新たな領域的な知識にまとめ上げられて、新たな記憶として脳という認知の場に定着して行くことになる。

## 5. 知・情・意の三位一体と「物語り性」

今までの機械化・情報化技術の影の部分は、KNOW-WHAT に相当する「為すべき確定操作」という目的の系、即ち＜意＞の情報に加重がかかり過ぎて、KNOW-HOW に相当する「プロセスの制御」という事実の系、即ち＜知＞の情報およびKNOW-WHY に相当する「意味づけ」という価値の系、即ち＜情＞の情報が欠落し勝ちの傾向にあったことと関係している。機械化・情報化技術の日常的な利用に関与する人の記憶形成においては、三位一体の情報と、各々の情報がもつ時間・空間的な順序の起・承・転・結の配位によって生み出される、「物語り（ストーリー）性」が欠如しているかも知れない、という点に原因のある可能性がある。

## 6. 若年者と高年者の生きる力の基本

人間にとって、○社会環境に働きかけるための言語と説明の能力 ○物理環境に働きかける技能と技術のプロセス制御の能力 ○経験の豊かな先人や他者から、教えられたり、見聞きをしたり、先人の著作から学び取ったり

して、言語と説明の能力、および、技能と技術のプロセスの制御能力を、自らに移転させる能力、の三つの能力は、人間が環境の変化に適応して、よりよく生き続けるために求められるところの、他の能力による代替不能の基本の能力であり、若年者と高年者を問わずに、生活の中で、不断に磨かれ鍛えられなければならない能力である。

## 7. 知識は未来に向け重層化し続ける

人間の知識とは、知・情・意の情報の総体のことを意味している。人間の個体における記憶という知識の歴史は、新たに経験し学習した領域的な知識が、以前に経験し学習した知識に結合して、最新の知識のネットワークを形成し、新しい知識のネットワークが、以前の知識のネットワークの上に、次々にかぶさって、重層化する。そして、どの層も現在と未来の生存を支える歴史的な記憶情報として、活用され続ける。新たに経験し学習した知識が、以前に経験し学習した知識に結合して、最新の知識のネットワークを形成するプロセスは、新たに経験し学習した知識に基づいて、過去の経験や学習を再解釈するところの、成長のプロセスだと言える。再解釈のプロセスで、知識が、より広域的なものに組み換えられて、得られた新機軸の広域的な知識が、未来における更に新たな経験と学習の可能性を捕捉し、引き寄せて行く。

## 8. 知識を進化させるのに必要な条件

脳の情報処理の機能を維持し向上させて、新たな領域的な知識をまとめ上げ、知識を進化させるためのポイントは、『既存の領域的な知識をベースにして、新たな領域的な知識を探索し、それらを広域的な知識に組み換える』ことにある。

知識を進化させるために必要な条件と考えられるものは、次の通りである。

① 常に、異なる多様な情報の記憶のインプットが維持されていること。

② どの二つの情報の記憶の間にも、互いに意味の差異があるという、画然とした意味の非連続性（領域性、あるいは、分散性）を表象する認知の場のメカニズムが正しく働いていること。これは、「X または Y」(OR) という「分ける」「微分」のメカニズムである。

③ どの二つの情報の記憶の間にも、互いに意味の類似があるという、意味の連続性（共通性、あるいは、収束性）を表象する認知の場のメカニズムが正しく働いていること。これは、「X かつ Y」(AND) という「まとめる」「積分」のメカニズムである。

④ 主体の、現在の適応上のニーズに応じて、「分ける」「微分」のメカニズムか、あるいは、「まとめる」「積分」のメカニズムか、のいずれかが選択されるメカニズムが正しく働いていること。主体の認知の場のメカニズムや記憶の総体としては、「分ける」という「微分」の過程と、「まとめる」という「積分」の過程とが、概ね、バランスしているのが自然であろうと思われる。

⑤ 一に多を見る、「演繹」のメカニズムが正しく働いていること。つまり、一つの情報の記憶に、多様な意味の「顔」を持たせて、その記憶が、同語異義的に活用されていること。

⑥ 多に一を見る、「帰納」のメカニズムが正しく働いていること。つまり、多くの情報の記憶に、一様な意味の「顔」を持たせて、それらの記憶が、異語同義的に活用されていること。主体の認知の場のメカニズムや記憶の総体としては、「一に多を見る」という「演繹」の過程と、「多に一を見る」という「帰納」の過程とが、概ね、バランスしているのが自然であろうと思われる。

⑦ 過去から現在に至る、各々の時点において、今現在、新たに経験し学習した領域的な知識の記憶と、過去に経験し学習した既存の領域的な知識の記憶の間に、常に、アップトゥデートな情報のネットワークが形成され、より以前のネットワークの上に、新しいネットワークが、次々にかぶさって、重層的に、

積み重ねられるメカニズムが正しく働いていること。

## 9. 構造を生む自然のホメオスタシス

### 9.1 自然のシステムを規律する原理

宇宙のシステムを全体として捉えるならば、非平衡化と平衡化の狭間で、ある宇宙定数に向かって、バランスが成立し、宇宙のシステムは、エントロピーの増加速度が最も小さい経路を辿って進化し続けているものと考えられる。生命のシステムについても、同様に考えられる。部分域と部分域の相互作用は、平衡化をもたらすところの、正の共振と、非平衡化をもたらすところの、負の共振から成り立ち、正の共振は、互いを促進し、負の共振は、互いを抑制して、エントロピーの増加速度が最も小さいエネルギーのレベルに落ち着くことになる。これが、ホメオスタシス<恒常性>と呼ばれている状態である。

### 9.2 『ラティスの情報処理モデル』を構築

自然のシステムにおける『ラティスの情報処理モデル』は、自然の系のホメオスタシス<恒常性>を表わしている。

● 脳を含む自然のシステムの二つの部分域を  $P_2$  ,  $P_1$  , として、それぞれが保持するエネルギーの準位の相対的な比率を  $\varnothing P_2$  ,  $\varnothing P_1$  , とし、 $\varnothing P_2 = 1$  ,  $1 > \varnothing P_1 > 0$  , とする。そうすると、自然のシステムのホメオスタシス<恒常性>を表わす『ラティスの情報処理モデル』は、次のように表現することができる。

●  $P_2$  と  $P_1$  の間の引き込み合う関係は、 $\varnothing P_2 / \varnothing P_1 > (\varnothing P_2 + \varnothing P_1) / \varnothing P_2$  ..... [式 1] で、表わすことが出来る。

その解は、 $\varnothing P_1 < (\sqrt{5} - 1) / 2$  となる。

● 引き込み合う関係とは、 $P_2$  と  $P_1$  の間のポジティブフィードバックによって生じる、引き込み合う相互作用のことである。

●  $P_2$  と  $P_1$  の間の斥け合う関係は、 $\varnothing P_2 / \varnothing P_1 < (\varnothing P_2 + \varnothing P_1) / \varnothing P_2$  ..... [式 2] で、表わすことが出来る。

その解は、 $\varnothing P_1 > (\sqrt{5} - 1) / 2$  となる。

● 斥け合う関係とは、 $P_2$  と  $P_1$  の間のポジティブフィードバックによって生じる、斥け合う相互作用のことである。

●  $P_2$  と  $P_1$  の間のトータルとしての中立的な融合の関係である、ホメオスタシス<恒常性>の関係は、

$\varnothing P_2 / \varnothing P_1 = (\varnothing P_2 + \varnothing P_1) / \varnothing P_2$  ..... [式 3] で、表わすことが出来る。

その解は、 $\varnothing P_1 = (\sqrt{5} - 1) / 2$  となる。

● トータルとしての中立的な融合の関係である、ホメオスタシス<恒常性>の関係とは、 $P_2$  と  $P_1$  の間に生じる「引き込み合いのポジティブフィードバック」と「斥け合いのポジティブフィードバック」の間に生じるネガティブフィードバックの相互作用のことである。

●  $[(\sqrt{5} - 1) / 2 \div 0.61803398]$  を『融合準位』 FL [FUSIONAL ENERGY LEVEL] と名づける。

●  $P_2$  から  $P_1$  へ移動するエネルギーの準位を、『循環準位』 CL [CIRCULATIVE ENERGY LEVEL] と名づける。

● FL と CL の間には、 $(FL + CL) (FL + CL) = FL$  ..... [式 4] という関係が存在し、

$CL = \sqrt{\{(\sqrt{5} - 1) / 2\} - (\sqrt{5} - 1) / 2} \div 0.168117389$  となる。

### 9.3 ネットワークをつくり構造を創発する

● 上のモデルの機序によって、システムの部分域が結びつく過程を比喩的に言うと、「私」が結びついた「知人」の「知人」と、「私」が結びつくことを繰り返して、「私」が見知らぬ人と飛び石伝いに「知人」として結びつくのが、ネットワークをつくる仕組みである (式 3・式 4)。

● これで、「私」と「知人」と「知人」の「知人」から成るクラスターが、重層的なネットワークを形成して、小域的なクラスターから、中域的なクラスターへ広がり、更に、大域的なクラスターへと、限りなく発展して行く。

● そして、「知人」の中味が、「友人」(AND)

として結びつくのか (式1), または「他人」(OR)として結びつくのか (式2), が決まって行って、「友人」(AND) × 「友人」(AND) × 「友人」(AND) のクラスター関係が実現すれば、構造の創発を生む。

● 後の10. 4の記述において、脳の認知の場が、ネットワークをつくり、構造を創発するプロセスを具体的にシミュレートする。

#### 9. 4 ホメオスタシスの規律が構造を生む

自然のシステムでは、ホメオスタシス<恒常性>の規律に基づき、引き込み合うポジティブ・フィードバックと、斥け合うポジティブ・フィードバックの間にネガティブ・フィードバックが作動して、エネルギーの場の均衡からの逸脱と、均衡への回帰のバランスが逐一に調節され、このシステムの場のダイナミズムを通して、構造の自己組織化がダイナミックに達成される。人間の脳という認知のシステムにおいても、また、しかりであって、ホメオスタシス<恒常性>の規律が、認知の場に知識の構造を生み出す。

### 10. 知識の過去・現在・未来の繋がり

#### 10. 1 環境が領域的な知識を進化させる

人間の歴史の過去・現在・未来の繋がりとは、人間が、現在の新たな環境に適応するという視点から、過去に経験し学習した既存の領域的な知識を捉え直し、より抽象的で一般的な知識として、現在の適応のために転用し応用しながら、営みを効果的・効率的に変革し、現在と未来の生存を確かなものにして来た、という進化の歴史である。

人間の脳が担う情報処理の機能は、過去を想起し、未来を想像し予期して、現在に対処する営みを通じて、新たな領域的な知識をまとめ上げ、人間の知識を進化させて行く。

領域的な知識とは、特定の目的行動を達成するのに必要な、事実・価値・目的の三つの系の情報から成る、一つの総体としての知識、および、その要素知識をいう。

#### 10. 2 領域的な知識が形成・蓄積される場

人間は、過去を想起し、未来を想像し予期して、現在に対処する営みを通じて、現在から未来に向けて、新たな領域的な知識の記憶を形成する。

● それは、(1) <事物・事象>の事実の系の情報 (2) <自己の思考・自己の行動>の目的の系の情報 (3) <事物・事象>と<自己の思考・自己の行動>に対する価値の系の情報を、[時間の軸]・[事実・目的の空間の軸]・[評価(感情)の空間の軸]の三つの軸から成る、脳という認知の場に配置し、生存に役立つように秩序化する過程である。

● 各々の系の情報の時間的な配位と、空間的な配位の接合によって、各々の系に生じる情報の起・承・転・結の繋がりに、生存に意味のある、リスクとチャンスの、新たなストーリーを見出し、一つの総体としての新たな領域的な知識の記憶が形成され、蓄積される。

● 時間的な配位とは、空間の軸の同じ位置で時間的に継起して存在・生起した記憶の組み合わせを言い、空間的な配位とは、時間の軸の同じ位置で空間的に隣接して存在・生起した記憶の組み合わせを言う。

● 人間は、現在と未来のリスクの不安を減らし、チャンスの希望を増やすことを目指して、過去を想起し、未来を想像し予期して、現在に対処する営みを達成しようとする。

● その時に、○事実の系は、外部環境の状態の情報を担い、○価値の系は、生命体としての<自己>という内部環境の状態の情報を担う。○目的の系は、外部環境と内部環境の間に生じる不均衡の仲介者として、<自己の思考・自己の行動>という<自我>の情報を担う。

#### 10. 3 三系の情報が領域的な知識を紡ぐ

● 人間は、現在における内・外部の環境を認知して、過去を想起し、過去の経験や学習との類似と差異に照らして、未来を想像し予期し、外部の環境と、内部の環境である<自己>との間に生じそうな未来の不均衡を、想像

し予期する。

● そして、不均衡を解消して均衡を回復するために、過去の経験や学習との類似と差異に照らして、仲介者としての＜自己の思考・自己の行動＞の方略を構成する。

● こうして構成された、新たな領域的な知識のアイデアが、試行と実行の後に、新たな領域的な知識にまとめ上げられて、新たな記憶として脳という認知の場に定着して行くことになる。

● ＜自己の思考・自己の行動＞の方略のバリエーションがもつ、未来における可能性と現実性が想像され予期されて、選択され、現在において、実行される。

● 人間の営みにおいて、三つの系の情報の記憶は、脳という認知の場で、一対一に対応して布置され、連動し、相互に規定し合う。

#### 10.4 領域的な知識の進化の仕組みを模擬

● 先に記述した、自然のシステムのホメオスタシス＜恒常性＞を表わす『ラティスの情報処理モデル』を適用して、脳の認知の場がネットワークをつくり、広域的な知識という構造を創発するプロセスをシミュレートすることが出来る。

● このプロセスは、同時に、新たな領域的な知識が形成され、人間の知識が進化するプロセスのシミュレーションでもある。

● それは、新たな領域的な知識の、ある要素知識  $F$  に、別の要素知識が、新たに形成されて加わり、新機軸の広域的な知識の構造が創発されて、その新機軸の広域的な知識の構造が、更に、次の新たな要素知識を呼び込んで、更に、新たな新機軸の広域的な知識の構造を創発するという、相互的な作用のプロセスである。

○ 一層： $F$ に $G$ が「友人」(AND)で繋がり、 $G$ に $H$ が「他人」(OR)で繋がる。 $[F \cdot G - G \cdot H]$ 。○ 二層： $H$ に $I$ が「友人」(AND)で繋がり、 $G$ に $I$ が「他人」(OR)で繋がる。一層のネットワークに、 $H \cdot I$ と $G \cdot I$ が重層化し、 $[F \cdot G - G \cdot H - H \cdot I - G \cdot I]$ に

なる。○ 三層： $I$ に $J$ が「友人」(AND)で繋がり、 $H$ に $J$ が「友人」(AND)で繋がり、 $G$ に $J$ が「他人」(OR)で繋がる。 $J \cdot I$ に構造ができる。二層のネットワークに $I \cdot J$ と $H \cdot J$ と $G \cdot J$ が重層化し、 $[F \cdot G - G \cdot H - H \cdot I - G \cdot I - I \cdot J - H \cdot J - G \cdot J]$ になる。○ 四層： $J$ に $K$ が「友人」(AND)で繋がり、 $I$ に $K$ が「友人」(AND)で繋がり、 $H$ に $K$ が「友人」(AND)で繋がり、 $G$ に $K$ が「他人」(OR)で繋がる。 $K \cdot J$ に構造ができる。三層のネットワークに $J \cdot K$ と $I \cdot K$ と $H \cdot K$ と $G \cdot K$ が重層化し、 $[F \cdot G - G \cdot H - H \cdot I - G \cdot I - I \cdot J - H \cdot J - G \cdot J - J \cdot K - I \cdot K - H \cdot K - G \cdot K]$ になる。○  $n$ 層：・・・・・・

#### 10.5 マルチモーダルな情報のクラスター

● このプロセスで創発する広域的な知識の構造とは、各層のネットワークにおいて、新たな領域的な知識の記憶と、それ以前の二つの既存の領域的な知識の記憶から成る、三つの知識の記憶のクラスターの結合関係の全てが、「AND」の関係で自己組織化されたときに生じる、事実・価値・目的という三つの情報の系から成る、マルチモーダルな情報のクラスターをいう。広域的な知識は、無意識的な帰納の推論によって、三組の領域的な知識の記憶の意味の共通性が集約された抽象的なイメージまたは概念である。

● 新機軸の広域的な知識が、未来における更に新たな経験と学習に遭遇する可能性を生み出し、近い未来に、更に新たな領域的な知識をまとめ上げることに繋がって行く。

● 人間は、そのような一般的で抽象的なイメージまたは概念を前提とした、意識的な演繹の推論によって、事実・価値・目的の三つの系の各々の情報に関する、時間的・空間的な起・承・転・結の配位に、生存に意味のある、リスクとチャンス、新たなストーリーを見出して、新たな領域的な知識にまとめ上げ、知識を進化させながら、営みの変革を図って行くことが出来るのである。

## 文芸批評の計量解析による批評行為の背景的特徴の抽出

# Extracting Background Knowledge About Critics Through Numerical Analysis of Literary Reviews

村井源<sup>\*</sup>, 徃住彰文<sup>†</sup>

Hajime MURAI, Akifumi TOKOSUMI

<sup>\*</sup>1 東京工業大学大学院社会理工学研究科

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

〒152-8552 東京都目黒区大岡山2-12-1-W9-48

E-mail: {h\_murai, akt}@valdes.titech.ac.jp

本研究では、テキストを解釈し批評する行為を科学的にとらえるためのケーススタディとして、著名な文芸批評家個人の批評テキストを網羅的に収集し、計量的な分析を行った。テキスト中の評価的語彙の共起情報の分析からは、「新しさ」「美しさ」「深さ」などが重要であることが分かった。また、テキスト中の人名から、批評家が影響を強く受けている作家・思想家を分析し、テキストの評価軸となる思想的な背景を抽出することができた。

In order to scientifically investigate the cognitive processes of interpreting and evaluating texts, as a case study, this paper collects and numerically analyzes some literary reviews written by a famous critic. The results of co-occurrence analysis for evaluative vocabulary revealed ‘newness’, ‘beauty’ and ‘depth’ to be important. Moreover, by extracting individual names from the texts and detecting influential writers and philosophers, it was also possible to extract the background knowledge and philosophical characteristics that shape the standards of text evaluation.

キーワード: 書評, ネットワーク, 共起単語

Book review, network, co-occurrence

## 1 はじめに

テキスト読解における情報処理過程は、主に心理学や認知科学の分野において、その解明が進められてきた。文字を単語として認識し、単語を文節やセンテンス

レベルに構造化し、センテンス間の修辭関係や論理的関係を認識するという多段階の並行的なメカニズムは非常に複雑である。テキストの読解にはコンテキストの理解の問題も含まれるため、現在まで多くの研究が進められているものの未だ

解明されていない点も少なくない。

一方で、テキスト読解の後に続く評価と批評の認知プロセスもまた複雑である。テキスト批評の認知メカニズムは、何をもって批評と言いつるかという根源的な問題も含めて端緒についたばかりである[1]が批評において重要視される概念の関係性についてはジャンルごとの特徴が抽出されている[2]。

読解及び批評には種々のコンテキスト情報が必要となることが一般に知られている。しかし科学的にテキストの読解を理解するためには、具体的にどのようなコンテキスト情報が必要となるかに関してはいまだ明確にされていない。また、テキスト読解においては、哲学・思想的背景に基づく様々な批評法が用いられ、新しい思想パラダイムの出現とともに、新しい批評が誕生することが少なくない。

本研究においては、未だ未解明の点の多い批評の認知メカニズムを探るため、批評において必要となる背景的な思想に対応する価値観と、コンテキスト情報に対応する知識がどのようなものであるかを計量的に抽出することを目的とする。ただし、一般的な背景の価値観や背景のコンテキスト情報の網羅的な抽出は困難であるため、本研究においては一人芸芸批評家の批評テキストに焦点を絞って、ケーススタディ的に背景情報の抽出を行い、結果の妥当性を考察する。

## 2 対象となる批評文テキスト

本研究が対象とする批評文テキストのデータは、文芸批評家、井口時雄氏[3]の文芸批評に関連する、メディアに掲載されたテキストを収集し電子化したもの

である。年代は1983年から2009年までであり、テキスト数は396、全体のテキストサイズは 2.96MBとなっている。表1には年代と形式ごとの数を示した。

表1 対象テキスト数

|        | 1980<br>年代 | 1990<br>年代 | 2000<br>年代 | 合計  |
|--------|------------|------------|------------|-----|
| 解説     | 0          | 9          | 14         | 23  |
| 書評     | 52         | 110        | 69         | 231 |
| 作家論    | 12         | 17         | 23         | 52  |
| 文学・批評論 | 3          | 19         | 25         | 47  |
| エッセイ   | 11         | 27         | 5          | 43  |
| 合計     | 78         | 182        | 136        | 396 |

## 3 批評文の感性的側面

テキストに含まれる概念構造としては、内容的側面と感性的側面の両面が考えられるが[4]、感性的側面を表す要素としては形容詞が考えられる。そこで形容詞の頻出語彙の共起ネットワークを作成して形容詞によって示される批評文中の感性的な側面の特徴を図示する。図1には全批評テキストの形容詞共起ネットワークを示した。頻出語彙の影響を可視化するため、図でのノードのフォントサイズは、単語の出現頻度の二乗根に比例している。また、孤立ノードは割愛した。以下の図では、可視性のため語彙間共起頻度上位約50までをエッジとして描画している。

図1のネットワークにおける中心性を計量的に分析するため、UCINETを用いて計算した結果の各中心性の値が上位10位に入った語彙を表2に示す。

図1および表2の結果より、対象批評テキスト中で中心的な役割を果たす評価尺度は（「よい」が評価尺度であるのは当然であるので）、「新しい」「美しい」「深い」などであると分かる。文芸批評においては対象となるテキストが何らか



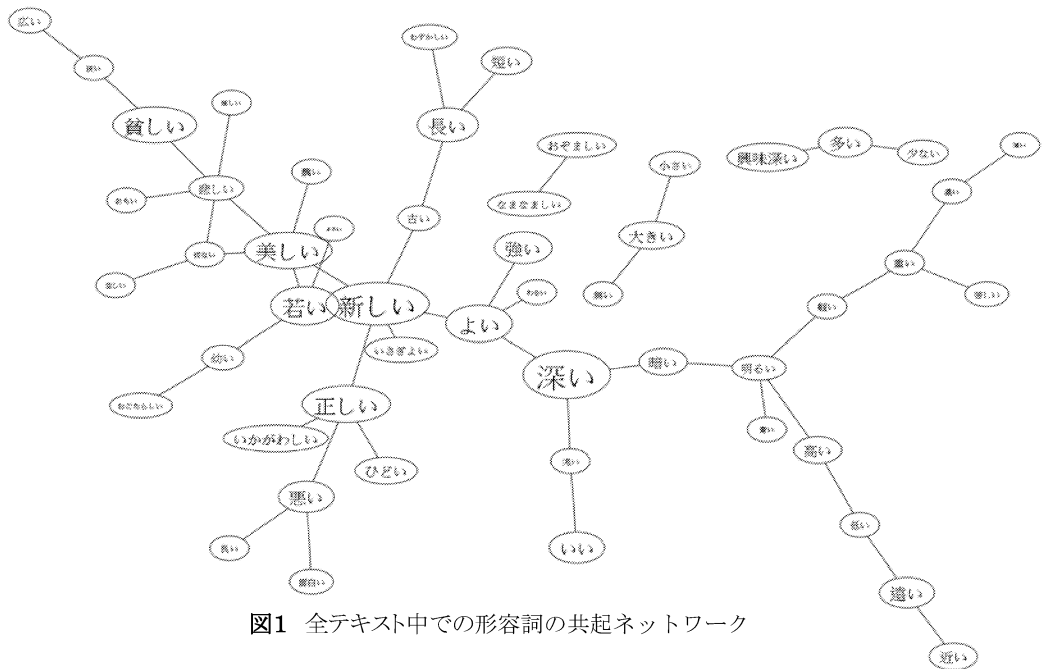


図1 全テキスト中での形容詞の共起ネットワーク

の意味で「新しい」要素を持っており、文章表現等に「美しさ」があり、対象を「深く」分析し、描けているかという点が問われているのだと考えられよう。

表2 全テキスト中での形容詞の中心性

| Closeness |      | Betweenness |       | Eigenvector |       |
|-----------|------|-------------|-------|-------------|-------|
| 新しい       | 9.11 | 新しい         | 50.04 | 美しい         | 67.06 |
| よい        | 8.98 | よい          | 37.10 | 新しい         | 66.74 |
| 美しい       | 8.79 | 深い          | 33.73 | 若い          | 50.63 |
| 深い        | 8.76 | 明るい         | 28.16 | 悲しい         | 44.81 |
| 若い        | 8.70 | 暗い          | 27.61 | 切ない         | 37.56 |
| 正しい       | 8.62 | 美しい         | 24.63 | よい          | 29.16 |
| 古い        | 8.56 | 悲しい         | 15.45 | 正しい         | 29.02 |
| 暗い        | 8.47 | 正しい         | 15.45 | 古い          | 22.94 |
| いさぎよい     | 8.47 | 軽い          | 12.24 | 醜い          | 20.42 |
| 悲しい       | 8.36 | 若い          | 9.57  | いさぎよい       | 20.33 |

このような批評における感性的な評価軸であるが、文芸批評家間で異なることは当然ながら、同じ批評家の中でも時期によって異なりうる。本研究での対象とする文芸批評家の場合、1980年代・1990

年代・2000年代で中心的な形容詞がどのように変化するかを上位5位のみそれぞれ表3から表5に示す。表3で中心的なのは「若い」という語彙だが、表4では「新しい」と「美しい」の中心性が高く、表5においては「深い」「正しい」「悪い」と時代によって大きく変化している様子が見えてくる。「若さ」「新しさ」を重視した時代から、次第に内面的な対象の分析の深さを重視するように変化した中で善悪の問題がクローズアップされてきた価値観の変化の様子が分かる。

表3 1980年代テキスト中での形容詞の中心性

| Closeness |       | Betweenness |       | Eigenvector |       |
|-----------|-------|-------------|-------|-------------|-------|
| 若い        | 57.50 | 若い          | 54.69 | 若い          | 71.16 |
| 美しい       | 46.00 | 白い          | 22.50 | 悪い          | 42.24 |
| 悪い        | 46.00 | 深い          | 19.76 | 白い          | 40.82 |
| 白い        | 46.00 | 短い          | 18.52 | 明るい         | 40.24 |
| 新しい       | 44.23 | 新しい         | 18.21 | 新しい         | 40.10 |

表4 1990年代テキスト中での形容詞の中心性

| Closeness |       | Betweenness |       | Eigenvector |       |
|-----------|-------|-------------|-------|-------------|-------|
| 新しい       | 43.04 | よい          | 23.37 | 新しい         | 58.58 |
| よい        | 39.08 | 大きい         | 21.73 | 美しい         | 50.27 |
| 美しい       | 38.64 | 新しい         | 21.13 | 若い          | 42.16 |
| 古い        | 38.20 | 美しい         | 21.00 | 正しい         | 41.41 |
| 大きい       | 38.20 | 長い          | 20.74 | 大きい         | 39.85 |

表5 2000年代テキスト中での形容詞の中心性

| Closeness |      | Betweenness |       | Eigenvector |       |
|-----------|------|-------------|-------|-------------|-------|
| 深い        | 9.71 | 深い          | 23.68 | 深い          | 52.21 |
| 正しい       | 9.62 | 正しい         | 19.69 | 軽い          | 46.21 |
| 悪い        | 9.51 | 悪い          | 18.20 | 若い          | 45.26 |
| 若い        | 9.47 | 若い          | 17.66 | 暗い          | 38.81 |
| 新しい       | 9.41 | 新しい         | 15.62 | 正しい         | 38.38 |

#### 4 批評対象となる動作

形容詞の分析と同様の手法を用いて、動詞の使用の特徴を分析する。動詞利用の特徴によって明らかとなるのは、批評で中心的な対象となる「行為」である。まずテキスト全体での共起ネットワークを図2に示す。また図2のネットワークにおける中心性の計算結果を表6に示す。

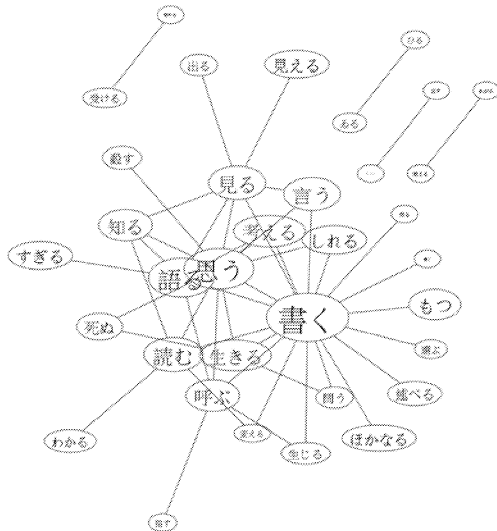


図2 全テキスト中での動詞の共起ネットワーク

表6 全テキスト中での動詞の中心性

| Closeness |       | Betweenness |       | Eigenvector |       |
|-----------|-------|-------------|-------|-------------|-------|
| 書く        | 10.83 | 書く          | 35.41 | 書く          | 69.89 |
| 思う        | 10.63 | 思う          | 12.70 | 思う          | 61.04 |
| 語る        | 10.43 | 見る          | 9.31  | 語る          | 42.66 |
| 見る        | 10.40 | 語る          | 5.78  | 見る          | 39.01 |
| 読む        | 10.30 | 読む          | 5.48  | 読む          | 34.40 |
| 呼ぶ        | 10.30 | 呼ぶ          | 4.93  | 生きる         | 31.69 |
| 生きる       | 10.27 | 生きる         | 1.55  | 呼ぶ          | 31.42 |
| 言う        | 10.24 | 知る          | 0.59  | 知る          | 28.54 |
| 考える       | 10.21 |             |       | 考える         | 27.97 |
| しれる       | 10.15 |             |       | 言う          | 27.38 |

図2および表6より、「書き」「語る」行為の結果を、「見て」「読み」、それについて「思う」ということが批評行為の中心的作業であることが考察される。他の動詞もほぼ直接的に批評行為自体に関連しているが、「生きる」だけが例外的な意味概念を持つ。対象テキスト中の「生きる」の用例としては、どのように言葉を「書き」「語る」かがどのように「生きる」かであると記述している箇所が複数あり、そのような言語行為全体に対する筆者の特徴的な姿勢が現れていると考えられる。

#### 5 作家・批評家間の関連性

文芸批評においては文学作品だけではなく、批評テキスト自体もまた文芸批評の対象となる。本研究の対象テキスト中で批評対象となった文学者・批評家を表7に上位10位まで記す。全テキスト中での単純出現回数順と、出現テキスト数での順位に分けて記述してある。また、これらの文学者・批評家のテキスト中での共起関係をネットワーク化したものを図3に示す。また批評対象となったテキストの作者のリストは表8に示す。

図3 全テキスト中での作家・批評家等の共起ネットワーク

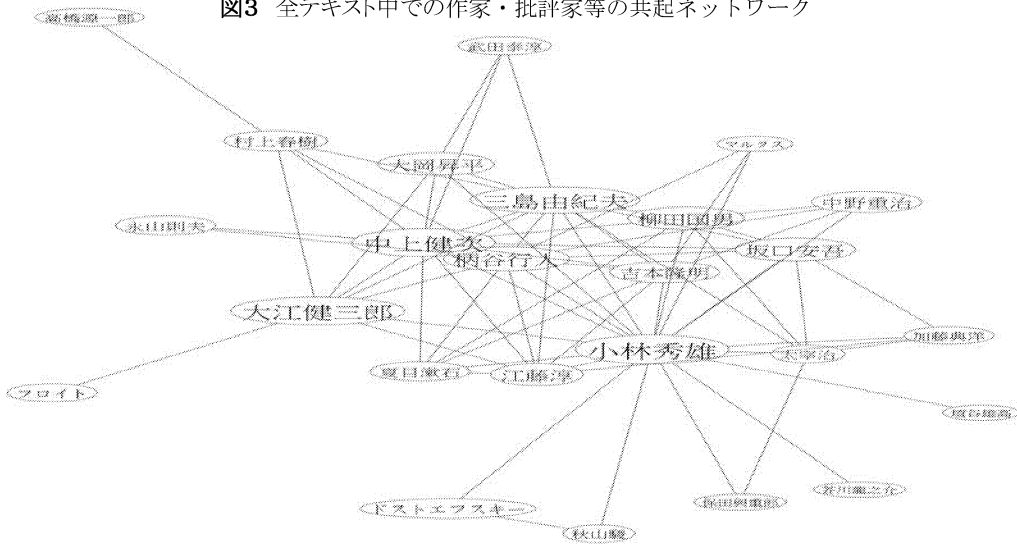


表7 全テキスト中での文筆家の出現数

| 出現回数  |     | 出現テキスト数 |    |
|-------|-----|---------|----|
| 小林秀雄  | 622 | 小林秀雄    | 63 |
| 坂口安吾  | 555 | 中上健次    | 60 |
| 中上健次  | 476 | 三島由紀夫   | 59 |
| 大江健三郎 | 448 | 大江健三郎   | 56 |
| 太宰治   | 406 | 柄谷行人    | 44 |
| 江藤淳   | 383 | 坂口安吾    | 39 |
| 柄谷行人  | 359 | 柳田国男    | 39 |
| 保田與重郎 | 350 | 大岡昇平    | 36 |
| 柳田国男  | 237 | 江藤淳     | 35 |
| 永山則夫  | 188 | 中野重治    | 32 |

表8 批評対象テキストの作者

| 作者名   | 回数 | 作者名  | 回数 |
|-------|----|------|----|
| 大江健三郎 | 30 | 江藤淳  | 8  |
| 中上健次  | 30 | 車谷長吉 | 7  |
| 柄谷行人  | 14 | 古井由吉 | 5  |
| 坂口安吾  | 11 | 中野重治 | 5  |
| 永山則夫  | 9  | 筒井康隆 | 5  |

大江・中上・坂口など批評対象となっている人物の言及が多いのは当然ながら、一度も批評対象となっていない、太宰・三島や、ドストエフスキー、マルクス、フロイトなど様々な名前がネットワーク上には現れてきている。これらが示すのは、言及回数は少なく、直接批評対象と

はしていない言外のコンテキストとして種々の古典的作品や思想系のテキストが評価メカニズムの中に深く織り込まれている可能性である。

## 6 背景的思想・作品

5節での結果に基づき、対象テキストが前提としている古典的作品と思想的な背景を抽出するため、人名に絞って抽出を行った。まず、文芸批評の前提となる作品であるが、作品登場人物の出現順位（出現テキスト数順）を表9に示す。

表9より、批評対象作品となっており、かつ連作であるため複数テキストで出現しやすい中上健二の『枯れ木灘』の登場人物を除くと、ドストエフスキーの作品の登場人物が有意に多いことが分かる。つまり、当該批評家は他テキストの読解において暗黙のうちにドストエフスキー的な価値観・世界観を前提として作品の理解を進めているということが強く示唆される。また、フロイトの心理的解釈との関連でギリシャ神話のオイディプスも

登場回数が多い。

次に、背景的思想を示すと考えられる思想家名の出現順位を表10に示す。

表9 頻出登場人物と出典

| 登場人物名    | 頻度  | テキスト | 作者       | 作品     |
|----------|-----|------|----------|--------|
| ラスコーリニコフ | 61  | 15   | ドストエフスキー | 『罪と罰』  |
| 龍造       | 172 | 10   | 中上健二     | 『枯れ木灘』 |
| オイディプス   | 46  | 9    |          | ギリシャ神話 |
| 秋幸       | 307 | 8    | 中上健二     | 『枯れ木灘』 |
| 郁男       | 53  | 8    | 中上健二     | 『枯れ木灘』 |
| ムイシュキン   | 13  | 7    | ドストエフスキー | 『白痴』   |
| イッポリート   | 14  | 6    | ドストエフスキー | 『白痴』   |

表10 頻出思想家名

| 思想家名     | 頻度 | テキスト |
|----------|----|------|
| フロイト     | 49 | 20   |
| マルクス     | 66 | 13   |
| ジャック・ラカン | 20 | 11   |
| サルトル     | 15 | 8    |
| ルソー      | 25 | 7    |
| カント      | 15 | 7    |
| 孔子       | 13 | 7    |
| ニーチェ     | 10 | 7    |

表10より、頻度はマルクスが、出現テキスト数はフロイトが最大である。表9の結果と合わせて考えるなら、貧困と暴力（ドストエフスキー・マルクス）とコンプレックス（フロイト・オイディプス）が、対象テキストにおける背景的な重要関心事項であると考察できよう。

## 7 結論と今後の課題

本研究では、文芸批評家のテキストを網羅的に収集し、計量分析することで、批評における評価軸となる感性的な特徴と、批評を行う上での背景となる前提的

な知識・コンテクストがどのような文学作品と思想家によって主に構成されているのかを明らかにした。これらに結果により、テキスト解釈と批評のメカニズムを構成する計量的なデータのいくつかを明らかにすることができた。

今後は、これらの背景的価値観や前提となる文学・思想的知識などを計量的に扱いうるためのデータ化電子化を進めるとともに、これらがどのようにテキスト自体と相互作用して解釈と評価が行われるのかを分析していく必要がある。

## 謝辞

本研究は科研費「知識共有のための価値指向型オントロジーの多分野多言語化」(20300074)の助成を受けた。

## 参考文献

- [1] 佐藤さやか, 松本斉子, 村井源, 往住彰文: “テキストの内容分析による文芸批評の認知構造の推定”, 第10回日本感性工学会大会論文集, P01-11, 2008.
- [2] 村井源, 往住彰文: “テキスト批評の計量化に向けて—書評の計量分析—”, 情報知識学会第17回年次大会, Vol. 19, No. 2, pp. 120-125, 2009.
- [3] 井口時男: 東京工業大学大学院社会理工学研究科価値システム専攻教授, <http://www.valdes.titech.ac.jp/~iguchi/> (2010年3月9日参照) .
- [4] 村井源, 松本斉子, 往住彰文: “コーパスに基づく感性と論理のネットワーク表現”, 日本認知科学会第25回大会発表論文集, 194-197, 2008.

## 星新一ショートショート文学の物語パターン抽出

### Extracting Narrative Patterns

#### from the Short Short-Stories of Hoshi Shinichi

佐藤知恵<sup>1\*</sup>, 村井源, 往住彰文

Chie SATO, Hajime MURAI, Akifumi TOKOSUMI

<sup>1\*</sup> 東京工業大学大学院社会理工学研究科

Department of Value and Decision Science, Tokyo Institute of Technology

〒152-8552 東京都目黒区大岡山 2-12-1-W9

E-mail: {chiesato, h\_murai, akt}@valdes.titech.ac.jp

日本の小説家である星新一は、日本のショートショートの第一人者であり、執筆期間の前半(1968～1973)だけで718編の作品を残している。星作品はしばしば物語の骨組みの共通性が指摘されるが、本論文は、そのような物語のパターン性がどのような点から確認できるのかを、11 の物語構造ユニットを用いて表現、分析することにより説明するものである。これらのユニットは先行研究をもとに、星作品に頻出する物語の出来事の要素として設定した。ユニットとその組み合わせを記号で表現したデータを用いて、パターンの分布、因子分析、N-gram による分析を施すことで、星作品の最も基本的な物語構造を抽出した。

Hoshi Shinichi, a Japanese writer, is known as one of the greatest short short-story writers in Japan, who wrote 718 works during his earlier years (1968-1973). Hoshi's works have often been criticized for their structural similarities. This paper investigates the narrative patterns of Hoshi's work and describes them in terms of 11 primitive units of narrative structure. Based on previous studies, these narrative units are established as the components of frequent narrative events within Hoshi's works. Employing these units and their combinations, the most basic narrative structures are extracted by conducting analyses of pattern distributions, factor analysis, and N-gram analysis.

キーワード: 星新一, ショートショート, 構造分析, 物語パターン

Keywords: Hoshi Shinichi, Short Short-Stories, Structural Analysis, Narrative Pattern

## 1 はじめに

1053 編もの作品を残し、その短さや特有の描写、パターン性などの特徴が議論されてきた[1, 2, 3, 4]。しかし、これらの特徴に

対する計量的、客観的研究は少なく[5]、解明されていない点も多い。そこで本研究では、構造分析の考えをもとに、星作品の特徴である“パターンや骨組の共通性”を抽出し、1053 編もの作品数[6]を輩出し得た、

構造上の工夫を追及する。

## 2 対象

本研究の対象作品は、星の全執筆帰還 1968–1999 中で執筆率の高い前期 1968–1973 の全 718 作品[7]である。

## 3 分析手法と結果

星作品の特徴である“パターンや骨組の共通性”を抽出するため、本研究では展開パターンによる分析を提案する。展開パターンとは、物語を構成する出来事の種類である。展開パターンを構成する展開ユニットは Propp[8]に倣って記号で表現するが、その内容は登場人物の機能ではなく、テキストから読み取れる出来事の時間的流れである。展開ユニットの設定に当たっては過去の類似の先行研究を参照した。

表 1 展開ユニットの定義

| フィールド   | パターン      | 定義                               |
|---------|-----------|----------------------------------|
| 真相      | a. 発覚     | 1. 謎の提示 2. 謎の発覚/解消               |
|         | b. 埋没     | 発覚している真相が登場人物に隠されている状態           |
| 評価      | c. 転換     | 1. 評価の提示 2. 評価の変更                |
| 立場      | d. 交替     | 1. 登場人物の立場の提示 2. 立場の入れ替わり        |
|         | e. 対立     | 人物間のゴールとプランの対立                   |
| 目的 / 問題 | f. 達成成功   | 1. 問題/目的の提示 2. 行動 3. 問題/目的の解決/達成 |
|         | g. 達成失敗   | 1. 目的の提示 2. 行動 3. 目的と異なる結果       |
|         | h. 未達成/放置 | 1. 問題/目的の提示 2. 問題の放置/目的の未達成      |
| 状態      | i. 好転     | 1. 状態の提示 2. 状態の向上                |
|         | j. 暗転     | 1. 状態の提示 2. 状態の悪化                |
|         | k. 変化なし   | 1. 状態の提示 2. 同種状態の提示              |

### 3.1 展開ユニットの設定

Propp[8]での主人公と敵の対立や、主人公が変身するという、立場の機能をそれぞれ、「対立」及び「交替」として、展開ユニットとして採用した。また、長田[5]での結びに至るまでの伏線の動きは、何らかの真相が発覚・埋没したり、それによって対象への評価が転換する箇所によく生じることから、「真相発覚」、「真相埋没」、「評価転換」を加えた。さらに、Ryan[9]と Lehnert[10]の、登場人物のゴールプランによる物事の進行に倣い、「目的達成」、「達成失敗」、「未達成/放置」と、登場人物の精神状態ではなく、状況の「好転」、「暗転」、「変化なし」を設置した。これらより展開ユニットは表 1 のように定義する。星作品の物語パターンを展開ユニットで分析した結果の一部を例として表 2 に示す。

表 2 展開ユニット分析の結果例

| 作品名        | パターン     |
|------------|----------|
| 『保証』       | iejacdb  |
| 『大黒さま』     | eaj      |
| 『あるスパイの物語』 | efacb    |
| 『住宅問題』     | faj      |
| 『信念』       | afififid |
| 『半人前』      | jjjacd   |

### 3.2 展開ユニットの頻度分布

展開ユニットによって表現された 718 作品の物語構造（展開パターン）データを用いて、展開ユニットの分布をみる（表 3）。表 3 は単純な出現頻度の比較であり、同作品中に繰り返し 3 回出現すれば 3 とする。

### 3.3 因子分析による特徴の抽出

各作品に出現する展開ユニットのデータをベクトルとして因子分析を行い、各展開

ユニットの散らばり（各展開ユニット同士の親和性及び不和合性）を調べた（表4）。本分析においては、固定値を1とし、説明された作品数は約47%であった。なお本分析では一作品に一度以上出現した場合を1として二値でベクトル化した。

表3 展開ユニットの出現頻度

|    |        |     |    |          |     |
|----|--------|-----|----|----------|-----|
| 真相 | 真相発覚 a | 644 | 目的 | 目的達成 f   | 450 |
|    | 真相埋没 b | 188 |    | 達成失敗 g   | 132 |
| 評価 | 評価転換 c | 464 | 問題 | 未達成/放置 h | 306 |
|    |        |     |    |          |     |
| 立場 | 交替 d   | 221 | 状態 | 好転 i     | 201 |
|    | 対立 e   | 182 |    | 変化なし k   | 172 |
|    |        |     |    | 暗転 j     | 335 |

表4 展開ユニットの因子分析結果

|          | 成分1    | 成分2    | 成分3    | 成分4    |
|----------|--------|--------|--------|--------|
| 真相発覚 a   | 0.514  | -0.016 | 0.036  | -0.124 |
| 真相埋没 b   | 0.489  | -0.047 | -0.145 | 0.247  |
| 評価転換 c   | -0.063 | -0.096 | 0.647  | -0.166 |
| 交替 d     | 0.107  | 0.077  | -0.586 | -0.284 |
| 対立 e     | 0.654  | -0.080 | -0.211 | -0.029 |
| 目的達成 f   | -0.270 | 0.565  | 0.401  | -0.282 |
| 達成失敗 g   | 0.073  | 0.004  | -0.030 | 0.842  |
| 未達成/放置 h | -0.592 | -0.284 | -0.451 | -0.086 |
| 好転 i     | -0.243 | 0.307  | 0.192  | 0.354  |
| 変化なし k   | -0.168 | -0.640 | 0.258  | -0.030 |
| 暗転 j     | -0.138 | 0.646  | -0.062 | 0.099  |

### 3.4 N-gram による展開パターンの抽出

N-gram モデルを用いて、N=2 及び N=3 における特徴的なユニットの連続を調べた（表5）。本分析では、同作品中の繰り返しの含め、単純な頻度として表に示した。

表5 N-gram の出現頻度

| 2-gram |     | 3-gram |    |
|--------|-----|--------|----|
| ac     | 232 | fac    | 60 |
| fa     | 127 | acj    | 37 |
| ha     | 90  | hac    | 34 |
| cj     | 70  | eac    | 30 |
| ad     | 70  | jac    | 29 |
| aj     | 66  | iac    | 24 |
| fi     | 65  | faj    | 23 |
| ea     | 63  | acb    | 22 |
| ja     | 59  | fia    | 21 |
| af     | 52  | ack    | 19 |

## 4 考察

### 4.1 頻度分布

展開ユニットの頻度は「真相発覚」が頻度 644 と最も多く、続いて、「評価転換」464、「目的達成」450、「暗転」335 となっており、星作品において、「真相発覚」が最も本質的なパターンと言うことができる。

また、フィールド別にみると「目的/問題」フィールドの出現頻度が最も高い。星の作品において、登場人物が目的や問題を持ち、それを達成、解決しようとする展開の中でストーリーが繰り返されることを示している。状態フィールドでは「暗転」が最も多く、バッドエンドを好む星の作風が、パターンの分布にも表れている。

### 4.2 因子分析

因子分析では、4つの因子が抽出された。各因子で、相関値上位3つの展開ユニットと、相関値の最も低い展開ユニットを、因子の特徴として抽出し考察する。これらの特徴と考察を表6から表9に示す。

因子1の特徴を示す表6では、「対立」、「真相発覚」、「真相埋没」が高い相関をも

ち、「未達成・放置」が強い負の相関をもつ。これは、複数の登場人物間に対立状況が発生する場合、その原因もしくはその結果として隠れていた真相が発覚あるいは埋没することにより、当事者たちは無用な争いであると知らずに対立するといった皮肉のパターンが抽出されている。

表6 成分1の特徴と考察

| 相関値    | カテゴリ   | 特徴・考察                                |
|--------|--------|--------------------------------------|
| 0.654  | 対立     | ・複数の主体の目的が対立                         |
| 0.514  | 真相発覚   | ・対立によって真実が明らかに<br>・真実が明らかになり対立が発生    |
| 0.489  | 真相埋没   | ・騙しプロット→対立者が真実隠蔽<br>・誤解プロット→無用な争い、競争 |
| -0.592 | 未達成・放置 | ・対立は目的問題フィールドと相反                     |

因子2の特徴を示す表7では「暗転」、「目的達成」、「好転」が高い相関をもち、「変化なし」が強い負の相関をもつ。この因子では、目的を達成しても、好転や暗転を繰り返し、結局元の木阿弥に戻るような状況が描かれている。

表7 成分2の特徴と考察

| 相関値   | カテゴリ | 特徴・考察                                   |
|-------|------|-----------------------------------------|
| 0.646 | 暗転   | 状況が暗転する<br>努力が水の泡になる空しさ<br>物事/結果の別側面の提示 |
| 0.565 | 目的達成 | 目的は到達したが、努力が水の泡                         |
| 0.307 | 好転   | 好転と暗転が繰り返され物語の起伏に                       |
| -0.64 | 変化なし | 目的達成すると何かが変化                            |

因子3の特徴を示す表8では、「評価転換」、「目的達成」、「変化なし」が高い相関をもち、「交替」は強い負の相関をもつ。この因子では目的を達成しても、元の木阿弥

に戻ったり、評価が転換することで達成した事実が達成しなくても同じこととなるような状況を描いている。

表8 成分3の特徴と考察

| 相関値    | カテゴリ | 特徴・考察                    |
|--------|------|--------------------------|
| 0.647  | 評価転換 | いいと思っていたことが実はそうではないことが判明 |
| 0.401  | 目的達成 | 目的は到達したが                 |
| 0.258  | 変化なし | 結局元に戻り、元の木阿弥             |
| -0.586 | 交替   | 評価転換と同時に交替は発生しにくい        |

因子4の特徴を示す表9では、「達成失敗」、「好転」、「真相埋没」が高い相関をもち、「交替」は強い負の相関をもつ。この因子では、目的達成が失敗したにも関わらず状況が好転したり、星作品にまれなハッピーエンドが起こる場合は、その真相が不確かであるという皮肉な状況が描かれている。

表9 成分4の特徴と考察

| 相関値    | カテゴリ | 特徴・考察                        |
|--------|------|------------------------------|
| 0.843  | 達成失敗 | 失敗はしたのだが…                    |
| 0.354  | 好転   | 失敗したが好転<br>物事の別側面を提示         |
| 0.247  | 真相埋没 | 達成失敗で真実はやぶの中<br>うやむやのハッピーエンド |
| -0.284 | 交替   | 失敗と同時に交替は発生しにくい              |

#### 4.3 N-gram モデルによる分析

2-gram による分析では、表5からac(「真相発覚」+「評価転換」)が有意に多く、星作品の本質的構造として、「評価を転換させるような真実の発覚」があると言える。次に頻度の高いfa(「目的達成」+「真相発覚」)は、目的達成という好ましい話の流れの後、真相が発覚する物語パターンが描かれる。真相が発覚すると、上位セットのad



(「真相発覚」+「立場の交替」), aj (「真相発覚」+「暗転」)にあるように,その後,立場が入れ替わる,状況が悪化するなどを通して,結局元の木阿弥に戻ったり,暗転したりといった流れが描かれる場合が多い。

3-gramによる分析では,表5より,最も頻度の高いパターンはfac(「目的達成」+「真相発覚」+「評価転換」)であり,目的達成はしたものの,真相が発覚することによって,目的達成した努力が水の泡となったり,目的の前提が崩壊したりする,強い皮肉の構造を見ることができる。また,上位10セットのうち,8割が2-gramによる分析結果によって抽出されたac(「真相発覚」+「評価の転換」)が含まれていることも,acを星の本質的構造とみなすことを更に強く裏付けている。

## 5 まとめ

本研究では,構造分析データを用いて,ユニット分布,因子分析,N-gramモデルによる分析を行い,展開ユニットの組み合わせや順序を考慮した,星に特徴的な物語パターンを抽出した。

ユニット分布では,最も頻度の高いユニットが「真相発覚」であることを示した。これより,「真相発覚」が星作品の展開ユニットとして最も本質的であると考えられる。

因子分析では,共通成分を反映した4つの因子を抽出し分析・考察した。星らしさを示す,騙しプロット,誤解プロット,皮肉プロット,二面性プロット,元の木阿弥プロットなどが,具体的な展開ユニットのセットとして説明された。特に,星作品に関して熱心に議論される皮肉[3]について,目的を達成したのに負の精神状態になる(またはその逆),パターンにより,物事の

二面性を強調したり,それが強い皮肉につながったりすることが示された。因子分析の結果により得られた,特徴的な高相関値のセットとその特徴考察を表10に示す。

表10 因子分析による特徴的プロット

| 特徴的な高相関値<br>セット                  | 考察                  |
|----------------------------------|---------------------|
| ・真相埋没+対立                         | ・騙しプロット<br>・誤解プロット  |
| ・目的達成+暗転<br>・達成失敗+好転<br>・好転+真相埋没 | ・皮肉プロット<br>・二面性プロット |
| ・目的達成+変化なし<br>・評価転換+変化なし         | ・元の木阿弥<br>プロット      |

N-gramによる分析では,展開パターンの並び(シリーズ)をN=2, N=3で抽出し,星らしい物語パターンとして,分析・考察した。これにより,星作品における本質的構造パターン,バッドエンドの嗜好,寓話性などがどのように表出するかを示した。特に寓話性に関しては一般的な,動物の起用[11]や道徳的教訓内容という特徴とは異なり,本分析では,寓話性を導く具体的な物語構造的性質の一つとしてha(「未達成」・「放置」+「真相発覚」)を示した。寓話性という,文章がもつ意味内容に物語パターンという構造が影響するという考えを提案した。表11にN-gramによる分析から得られた特徴的なシリーズと考察を示す。

表11 N-gramによる特徴的プロット

| 特徴的シリーズ   | 考察               |
|-----------|------------------|
| 真相発覚→評価転換 | 星作品における本質的構造パターン |
| 評価転換→暗転   | バッドエンドの嗜好        |
| 未達成→真相発覚  | 寓話要素の現出          |

## 6 今後の展望

各分析に関する具体的な改善点・課題としては、以下が挙げられる。

### 6.1 構造分析と計量分析の互換性の向上

筆者らはすでに計量的語彙分析による星新一作品の特徴抽出を行ったが[7], 本研究の構造分析で得られた結果を計量的手法やアルゴリズムの設計に用い, より科学的な物語作品の分析を実現することが望まれる。

### 6.2 展開パターン一致度算出

展開パターンの適用に関しては, 著者らが定義表に基づいてテキストを分割し分類を行ったが, 複数の分類者間の分類の一致度は算出していない。本研究ではテキストの分割と分類という二段階の過程での一致度を検証する必要があるため, このような場合に妥当な一致度の計算手法を考案し, 分析結果の正当性を検証する必要がある。

### 6.3 作品中の重要度を考慮した分析

本研究では, 11の展開パターンの全ての重要度を一律に扱い, 出現する順序通りに文字列として物語パターンとした。従って, ある展開パターンの中に内包される重要度の低い展開パターンも並びとして計算されている。そこで, 重要度の相対的高低の2極分類を適用し, より上位な物語パターンを正確に抽出することが考えられる。

### 6.4 人文学的文学研究との質的比較検証

本研究は, 作品の構造分析を取り上げたが, 今後は語りの人称や視点などによる, 他の人文学的な研究手法をも用いて, より高度なテキスト理解の認知過程を明らかにする必要がある。

## 謝辞

本研究は科研費「知識共有のための価値

指向型オントロジーの多分野多言語化」(20300074)の助成を受けた。

## 参考文献

- [1]最相葉月:「星新一 一〇〇一話をつくった人」, 新潮社, 2007.
- [2]権田萬二:「恐怖の原形質—星新一論—」, 1976 (石川喬二:「星新一の世界」, Vol. 9, No. 4, 新評社所収)
- [3]最相葉月:「あのころの未来」, 新潮社, 2005.
- [4]最相葉月:「星新一 空想工房へようこそ」, 新潮社, 2007.
- [5]長田貴誉子:“星新一の作品分類と分析 試論 一〇〇一編ショートショート群を考える”, 国文学会誌, vol. 22, pp. 76-115, 1993.
- [6]星新一:「星新一ショートショート 1001」, 新潮社, 1998.
- [7]佐藤知恵, 村井源, 徃住彰文:“文学作品群の特徴的語彙と概念カテゴリーの抽出—星新一ショートショートの計量分析—”, 情報知識学会誌, Vol. 19, No. 2, pp. 132-137, 2009.
- [8]ウラジーミル・プロップ (北岡誠司, 福田美智代訳):「昔話の形態学」, 水声社, 1987.
- [9]マリー・ロール・ライアン (岩松正洋訳):「可能世界・人工知能・物語理論」, 水声社, 2006.
- [10]Wendy G, Lehnert: “Plot Units: A Narrative Summarization Strategy”, 1982 (Inderjeet Mani, Mark T. Maybury: “Advances in automatic text summarization”, 1999).
- [11]大堀壽夫:「認知言語学Ⅱ: カテゴリー化」, 東京大学出版会, 2002.

## 音楽評論論文にみる作曲家の感性的特徴

### —計量分析を用いた感性語の抽出—

## The Sensibility Characteristics of Composers Observed within

## Music Criticisms: Extracting Sensibility Words Using

## Quantitative Analysis

河瀬彰宏<sup>1\*</sup>, 村井源<sup>1</sup>, 往住彰文<sup>1</sup>

Akihiro KAWASE<sup>1\*</sup>, Hajime MURAI<sup>1</sup>, Akifumi TOKOSUMI<sup>1</sup>

<sup>1</sup> 東京工業大学大学院社会理工学研究科

Department of Value and Decision Science, Tokyo Institute of Technology

〒152-8552 東京都目黒区大岡山2-12-1 W9-29

E-mail: {kawase, h\_murai, akt}@valdes.titech.ac.jp

本研究では、音楽評論家によって語られる作曲家の概念的特徴とそれらの関係を明らかにするために、作曲家に対する感性的表現を計量的に分析した。データは音楽評論雑誌『ポリフォニー—音楽評論の開かれた場』のテキスト全13冊を対象とし、最も多く語られる7名の作曲家(出現回数が多い順に、モーツァルト、ベートーヴェン、ワーグナー、マーラー、J.S.バッハ、ドビュッシー、シェーンベルク)を形容する感性的表現、語彙レベルで用いられる頻出感性語とその用法の特徴を解析することで、作曲家概念の感性的要素を精緻に特徴づけた。

In order to highlight conceptual features and to examine the relations to composers (musicians) evaluated by music critics, we focus on sensibility words and word phrases that describe composers and analyze them using a quantitative approach. We analyze the complete thirteen-volume set of "POLYPHONE", which is a highly respected journal of musical criticism within Japan. Each article was categorized according to seven groups based on seven most frequently occurring composers within all the articles; namely, Mozart, Beethoven, Wagner, Mahler, J.S.Bach, Debussy, and Schönberg, in descending order of frequency. We characterize the elements of sensibility linked to each composer concept by analyzing sensibility expressions, frequent sensibility words employed at the vocabulary level, and the usage features characteristic of each composer.

キーワード: 音楽評論, 音楽的概念, ポリフォニー

Music criticism, Musical concepts, POLYPHONE

## 1 はじめに

ある音楽作品を初めて聴いたり、無名の作曲家の音楽作品を聴いたりした際に、「バッハのような音楽だ」「ドビュッシーのような作品だ」と代表的な作曲家の名前を挙げて感想を述べることもある。この背後には人間特有の高度な知識処理が働いており、現在の音楽体験から得た情報と、過去の音楽体験から蓄積した情報（自らが認識する作曲家像）とを比較することで、評価・価値判断を下していると考えることができる。

本研究では、人々が過去の音楽体験から構築した「バッハ像」「ドビュッシー像」といった作曲家像を汲み取るために、人間の複雑な思考や感性を記述する手段である言語データに着目するという人文工学のアプローチに基づき、作曲家の特徴的概念を捉えることを考える。人文工学のアプローチから作曲家の概念的特徴を捉えた研究に、Tokosumi and Murai [1]がある。これは一人の現代音楽作曲家・武満徹の自筆テキストを分析対象とし、いわば限定されたコミュニティにおける作曲家概念の把握と、その機械可読な形での表現を追求したものであった。これに対して本研究では、一人の作曲家に限定せず、多くの作曲家について言及された音楽評論論文を分析対象とし、作曲家一個人のテキストとは異なる音楽評論独特のコミュニティにおける作曲家概念の把握と、その機械可読な形での表現を追求していく。具体的な方針として、近年テキスト分類・検索に用いられ

始めた感性語[2]を計量的に分析することで、音楽評論論文において語られる作曲家の概念的特徴を明確にしていく。

## 2 本研究で扱う対象

### 2.1 音楽評論テキストの収集

本研究では、河瀬ら[3]の研究に引き続き1987年10月から1993年12月号までの期間、TBSブリタニカから発行された音楽評論雑誌『ポリフォーン 音楽評論の開かれた場』（以下『ポリフォーン』）全13冊を分析対象とする。

### 2.2 テキストコーパスの特徴

解析の対象を明確にするために、本研究では音楽評論テキスト全編に渡って最も言及回数の多かった7名の作曲家に焦点を当てて収集した評論記事を7名毎に分類した。言及回数の多かった順にモーツァルト(705回)、ベートーヴェン(634回)、ワーグナー(444回)、マーラー(375回)、J.S. バッハ(351回)、ドビュッシー(335回)、シェーンベルク(273回)で、これに続く作曲家は、ブーレーズ(179回)、ストラヴィンスキー(150回)、ショパン(119回)、シューベルト(116回)、シューマン(116回)であった。全体的にドイツ語圏の作曲家に関する言及が多いことが『ポリフォーン』の特徴といえる。分析対象を7名に留めたのは、上位7位と8位以下との間に著しい言及回数の差があるためである。以上の言及回数上位7名の

表2 テキストコーパスで用いられている品詞

|           | 名詞    | 動詞   | 形容詞 | 副詞  |
|-----------|-------|------|-----|-----|
| Mozart    | 16072 | 2663 | 348 | 668 |
| Beethoven | 14376 | 2447 | 322 | 610 |
| Wagner    | 11383 | 2223 | 278 | 567 |
| Mahler    | 6910  | 1344 | 197 | 386 |
| J.S.Bach  | 10026 | 1947 | 223 | 535 |
| Debussy   | 7947  | 1445 | 185 | 375 |
| Schönberg | 8209  | 1430 | 189 | 406 |

表1 使用テキストの評論数と容量

|            | Moz. | Bee. | Wag. | Mah. | Bac. | Deb. | Sch. |
|------------|------|------|------|------|------|------|------|
| 評論数<br>(編) | 63   | 53   | 42   | 20   | 36   | 20   | 22   |
| 容量<br>(MB) | 1.19 | 1.14 | 0.83 | 0.36 | 0.73 | 0.44 | 0.51 |

作曲家ごとに分類したテキストデータの概要は表1の通りである．次にテキストコーパスの特徴を語彙レベルで示す．表2は品詞毎の出現数とその作曲家毎の比率を表したものである．品詞の分割と品詞の係り受け解析にはChaSenとCaboChaを用いた．

### 3 感性語の分類

#### 3.1 抽出された感性語の概要

前節で抽出された品詞のうち、感性的表現と関連のある感性語が多く含まれる形容詞に着目する．形容詞の頻出語が作曲家によってどのように相違するものか、作曲家毎に上位30位までの結果を表3に示す．上位30位までで形容詞のおよそ8割を占めている．上位

30位以内のうち、7名全員に共通する語彙は21であった．また、7名全員に共通して出現するものの100位以内までに入らない語彙は「ぎこちない」「ものすごい」「重苦しい」の3つのみであった．また、上位100位以内のうち、7名の誰とも重複せず単独で出現する語彙は「こよない」（モーツァルト93位）、「どぎつい」（ワーグナー72位）、「正しい」（マーラー43位）、「うらやましい」（ドビュッシー74位）、「たくましい」（ドビュッシー80位）、「かたい」（シェーンベルク65位）、「めまぐるしい」（シェーンベルク72位）、「罪深い」（シェーンベルク90位）の8つであった．用いられている語彙の多少に差はあるが、全体的に似通った語彙を

表3 形容詞の使用の相違(上位 30 位まで)

|    | Mozart  |       | Beethoven |        | Wagner  |        | Mahler |        | J.S.Bach |        | Debussy |        | Schönberg |        |
|----|---------|-------|-----------|--------|---------|--------|--------|--------|----------|--------|---------|--------|-----------|--------|
| 1  | 新しい     | 8.22% | 新しい       | 11.98% | 新しい     | 11.08% | 新しい    | 11.80% | 新しい      | 15.25% | 新しい     | 13.68% | 新しい       | 24.01% |
| 2  | 多い      | 6.30% | 多い        | 5.49%  | 多い      | 5.29%  | いい     | 6.65%  | よい       | 5.81%  | よい      | 5.26%  | よい        | 5.22%  |
| 3  | いい      | 5.91% | 強い        | 5.27%  | よい      | 4.92%  | よい     | 6.24%  | 多い       | 5.81%  | 多い      | 5.03%  | 大きい       | 3.94%  |
| 4  | よい      | 5.05% | いい        | 5.10%  | いい      | 4.73%  | 強い     | 6.24%  | 強い       | 5.47%  | 強い      | 4.91%  | 強い        | 3.83%  |
| 5  | 大きい     | 4.34% | よい        | 5.01%  | 強い      | 4.54%  | 若い     | 5.29%  | 大きい      | 4.17%  | 深い      | 4.68%  | いい        | 5.10%  |
| 6  | 強い      | 3.83% | 大きい       | 4.70%  | 深い      | 4.29%  | 多い     | 5.29%  | 深い       | 4.10%  | 若い      | 4.44%  | 若い        | 3.60%  |
| 7  | 深い      | 3.60% | 深い        | 3.66%  | 大きい     | 4.17%  | 大きい    | 5.02%  | いい       | 3.56%  | 長い      | 3.63%  | 多い        | 3.48%  |
| 8  | 美しい     | 3.29% | 若い        | 3.22%  | 少ない     | 3.55%  | 深い     | 3.80%  | 少ない      | 2.87%  | いい      | 3.51%  | 深い        | 3.25%  |
| 9  | 高い      | 3.21% | 少ない       | 3.09%  | 若い      | 3.36%  | 長い     | 2.58%  | 若い       | 2.80%  | 大きい     | 3.39%  | 長い        | 2.90%  |
| 10 | 少ない     | 2.86% | 美しい       | 3.05%  | 高い      | 3.24%  | 高い     | 2.44%  | 長い       | 2.80%  | 高い      | 3.04%  | 古い        | 2.44%  |
| 11 | 長い      | 2.58% | 高い        | 2.74%  | 美しい     | 3.05%  | 少ない    | 2.31%  | 美しい      | 2.74%  | 少ない     | 2.81%  | 少ない       | 2.44%  |
| 12 | 若い      | 2.46% | 長い        | 2.66%  | 近い      | 2.30%  | 激しい    | 2.04%  | 古い       | 2.53%  | 興味深い    | 2.11%  | 興味深い      | 2.32%  |
| 13 | 広い      | 2.27% | 広い        | 2.26%  | 長い      | 2.30%  | 広い     | 1.76%  | 高い       | 2.05%  | 近い      | 1.87%  | 短い        | 2.20%  |
| 14 | 古い      | 1.84% | 興味深い      | 1.61%  | 興味深い    | 1.93%  | 古い     | 1.63%  | やすい      | 1.98%  | 激しい     | 1.87%  | 美しい       | 1.97%  |
| 15 | 楽しい     | 1.60% | 古い        | 1.57%  | 古い      | 1.74%  | うまい    | 1.49%  | 近い       | 1.78%  | 古い      | 1.64%  | 高い        | 1.86%  |
| 16 | 悪い      | 1.37% | 悪い        | 1.52%  | 広い      | 1.74%  | 暗い     | 1.49%  | 広い       | 1.78%  | 広い      | 1.52%  | 難しい       | 1.74%  |
| 17 | 暗い      | 1.37% | 近い        | 1.52%  | やすい     | 1.62%  | 近い     | 1.49%  | 難しい      | 1.71%  | 悪い      | 1.40%  | やすい       | 1.51%  |
| 18 | 近い      | 1.33% | 正しい       | 1.35%  | うまい     | 1.56%  | 美しい    | 1.49%  | うまい      | 1.50%  | 美しい     | 1.40%  | 近い        | 1.51%  |
| 19 | やすい     | 1.29% | うまい       | 1.26%  | 激しい     | 1.37%  | ひどい    | 1.22%  | 悪い       | 1.50%  | 遠い      | 1.29%  | 正しい       | 1.51%  |
| 20 | 難しい     | 1.29% | やすい       | 1.26%  | 遠い      | 1.18%  | すばらしい  | 1.09%  | 遠い       | 1.09%  | 正しい     | 1.29%  | 詳しい       | 1.39%  |
| 21 | うまい     | 1.25% | 短い        | 1.22%  | 難しい     | 1.12%  | やすい    | 1.09%  | 激しい      | 1.09%  | やすい     | 1.17%  | うまい       | 1.28%  |
| 22 | 興味深い    | 1.21% | 激しい       | 1.18%  | 悪い      | 1.06%  | ふざわしい  | 0.95%  | ふざわしい    | 1.03%  | 暗い      | 1.17%  | 広い        | 1.28%  |
| 23 | 激しい     | 1.06% | 難しい       | 0.96%  | ふざわしい   | 0.93%  | 興味深い   | 0.95%  | 正しい      | 0.96%  | 難しい     | 1.05%  | 悪い        | 1.16%  |
| 24 | 大きな     | 1.02% | ひどい       | 0.91%  | 言うまでもない | 0.81%  | 明るい    | 0.95%  | 短い       | 0.96%  | 小さな     | 0.94%  | 遠い        | 1.04%  |
| 25 | 正しい     | 1.02% | 暗い        | 0.91%  | 暗い      | 0.75%  | 悪い     | 0.81%  | 興味深い     | 0.89%  | 親しい     | 0.94%  | 難い        | 0.81%  |
| 26 | 遠い      | 0.90% | 難しい       | 0.91%  | 正しい     | 0.75%  | 楽しい    | 0.81%  | 楽しい      | 0.82%  | 明るい     | 0.94%  | 親い        | 0.70%  |
| 27 | 著しい     | 0.90% | 遠い        | 0.87%  | 短い      | 0.75%  | 詳しい    | 0.81%  | 難い       | 0.82%  | うまい     | 0.70%  | 低い        | 0.70%  |
| 28 | 明るい     | 0.90% | 暖しい       | 0.87%  | 優しい     | 0.75%  | 難い     | 0.81%  | 狭い       | 0.75%  | はかない    | 0.70%  | ふざわしい     | 0.58%  |
| 29 | ひどい     | 0.86% | ふざわしい     | 0.78%  | 難い      | 0.62%  | 遠い     | 0.68%  | 詳しい      | 0.75%  | ふざわしい   | 0.70%  | 珍しい       | 0.58%  |
| 30 | 言うまでもない | 0.86% | 言うまでもない   | 0.65%  | 親しい     | 0.56%  | 短い     | 0.68%  | すばらしい    | 0.62%  | 楽しい     | 0.70%  | 遅い        | 0.46%  |

表 4 形容詞の意味分類

|    | Moz.  | Bee. | Wag. | Mah. | Bac.  | Deb.  | Sch.  |     | Moz.  | Bee. | Wag. | Mah. | Bac.  | Deb. | Sch. |
|----|-------|------|------|------|-------|-------|-------|-----|-------|------|------|------|-------|------|------|
| 真偽 | 26    | 31   | 12   | 3    | 14    | 11    | 13    | 言語  | 8     | 10   | 7    | 1    | 1     | 0    | 1    |
| 類  | 66    | 62   | ▲ 60 | 17   | 48    | 28    | 25    | 生活  | 30    | 31   | 22   | 11   | ▽ 6   | 11   | 14   |
| 存在 | 8     | 6    | 4    | 0    | 4     | ▲ 6   | 1     | 行為  | 307   | 262  | 192  | 104  | 167   | 102  | 89   |
| 様相 | 681   | 576  | 389  | 187  | 380   | ▽ 172 | 187   | 交わり | 12    | 12   | 11   | 3    | 12    | ▲ 10 | 3    |
| 力  | 190   | 201  | 130  | ▲ 80 | 128   | 72    | ▽ 48  | 待遇  | ▲ 67  | ▲ 52 | ▽ 15 | 13   | 18    | 14   | ▽ 5  |
| 作用 | 4     | 4    | 2    | 2    | 0     | 1     | 2     | 経済  | ▲ 112 | 80   | 63   | 26   | ▽ 33  | 32   | 20   |
| 時間 | ▽ 537 | 589  | 406  | 202  | ▲ 427 | 244   | 333   | 自然  | ▲ 319 | 218  | 178  | 76   | ▽ 117 | 92   | ▽ 54 |
| 形  | 9     | 10   | 7    | 3    | 10    | 5     | ▲ 10  | 物質  | ▲ 42  | 30   | ▽ 9  | 7    | 15    | 9    | ▽ 3  |
| 量  | 948   | 823  | 585  | 243  | 494   | 295   | ▽ 251 | 天地  | ▲ 6   | 2    | 1    | 1    | 0     | 0    | 1    |
| 心  | ▲ 484 | 374  | 277  | 115  | 225   | 150   | 120   | 生命  | ▽ 188 | 209  | 137  | 90   | 142   | 85   | 71   |

用いて各作曲家を特徴付けていることがわかる。

### 3.2. 感性語全体の大分類

感性語利用の特徴と差異を解析することで各作曲家の感性的特徴を抽出することが期待される。ここでは感性語の全体的な特徴を明確にするために、テキスト中に出現した形容詞を意味分類に区分した上で解析する。区分には分類語彙表[4]を用いる。ただし、意味分類は分類語彙表における「相」の中項目に基づくものとし、テキスト中のある語彙が2つ以上の項目にまたがる場合、その語彙の出現頻度をそれらの項目の頻度に加えるものとする。分類語彙表に基づく形容詞の意味分類の結果を表4に示す。 $\chi^2$ 検定の残差分析の結果、 $p < 0.05$ の水準において有意に多い項目と有意に少ない項目について、それぞれ▲と▽の記号を表中に付してある。表4より、音楽評論論文において用いられる形容詞のうち、「量」のカテゴリーに属する語彙が最も多く、続いて「時間」「様相」「心」のカテゴリーに属する語彙の順に多い。これら上位4カテゴリーで全体のおよそ7割を占める。以下の節では、上位4カテゴリーに属する語彙をそれぞれ分析することで各作曲家の感性的特徴を抽出する。

### 3.3 「量」に関する感性語

テキストコーパスに頻出した「量」カテ

ゴリーに属する形容詞には、大小、長短、遠近を表す語彙が並ぶ。 $\chi^2$ 検定の残差分析の結果、 $p < 0.05$ の水準において有意差はみられなかった。したがって、「量」の感性的側面から作曲家の個性を捉えることができなかった。なお、「量」カテゴリーに属する具体的な形容詞の掲載は有意差がみられなかったため割愛する。

### 3.4 「時間」に関する感性語

テキストコーパスに頻出した「時間」カテゴリーに属する形容詞を表5に示す。このカテゴリーには、「新しい」「若い」「小さい」といった新しさを喚起させる語彙と、「古い」「遅い」「大きい」といった旧さを喚起させる語彙が並ぶ。 $\chi^2$ 検定の残差分析の結果、シェーンベルクとマーラーは新しい印象を喚起する語彙に、モーツァルトとバッハは古い印象を喚起する語彙に特徴がみられた。モーツァルト、バッハ、シェーンベルクに対する時代的な感性は教科書にみられる時代

表 5 「時間」に関する形容詞の出現頻度

|     | Moz.  | Bee. | Wag. | Mah. | Bac. | Deb. | Sch.  |
|-----|-------|------|------|------|------|------|-------|
| 新しい | ▽ 210 | 275  | 178  | 87   | 223  | 117  | ▲ 207 |
| 大きい | ▲ 111 | 108  | 67   | 37   | 61   | 29   | ▽ 34  |
| 若い  | 63    | 74   | 54   | ▲ 39 | ▽ 41 | 38   | 31    |
| 古い  | 47    | 36   | 28   | 12   | 37   | 14   | 21    |
| 近い  | 34    | 35   | ▲ 37 | 11   | 26   | 16   | 13    |
| 遠い  | 23    | 20   | 19   | 5    | 16   | 11   | 9     |
| 小さい | 14    | 13   | 7    | 4    | 8    | 8    | 3     |
| 遅い  | 12    | 12   | 5    | 2    | 6    | 3    | 4     |

区分と一致しており妥当である。マーラーの評論の「若い」という形容詞の係り受け関係を調べると、「音楽」「作曲家」「時代」といった名詞を受けている。現代のマーラーに対する新旧評価は語る人の世代によって異なるが、この結果から『ポリフォーン』ではマーラーは新しい作曲家に位置づけられていると考えられる。

### 3.5 「様相」に関する感性語

テキストコーパスに頻出した「様相」カテゴリーに属する形容詞を表6に示す。このカテゴリーには、「よい」「いい」「やすい」といった肯定的な評価を喚起させる語彙と、「悪い」「難しい」「難い」といった否定的な評価を喚起させる語彙が並ぶ。作曲家だけでなく、音楽や様式など作品を評価する必要性のある音楽評論において、「様相」カテゴリーに属する形容詞の頻度が高いことは自然である。 $\chi^2$ 検定の残差分析の結果、3名の作曲家に対して「よい」「いい」という肯定的な評価を喚起させる語彙について有意

差がみられた。これらの語彙について係り受け関係を調べると、「音楽」「形式」「主義」「時代」「世界」「気質」といった作曲家と作品の性質を表す名詞を受ける頻度が高い。また、バッハとシェーンベルクに対して困難さを表す形容詞「難い」が有意に多く出現しており、係り受け関係を調べると、「音楽」「印象」「想念」という名詞を受けている。この結果は、多声音楽の極致を完成させたバッハと組織的な12音技法を確立したシェーンベルクの個性を捉えたものといえる。

### 3.6 「心」に関する感性語

テキストコーパスに頻出した「心」カテゴリーに属する形容詞を表7に示す。 $\chi^2$ 検定の残差分析の結果、有意に多い語彙について係り受け関係を調べると、ワーグナー評論の「興味深い」は「内容」「テーマ」「課題」「お話」といった楽曲の題材に関係する名詞を受ける頻度が高い。これは音楽・演劇・詩の総合芸術を目指して楽劇を創始したワーグナーの一側面を捉えたものである。モーツァルト評論の「楽しい」およびドビュッシー評論の「親しい」は「音楽」「表現」といった作品に対する印象の名詞を受ける頻度が高く、明るく快い印象がある。「様相」カテゴリーの結果と併せてシェーンベルクはその反対の印象がある。なお、有意差がみられた形容詞「ひどい」は、作曲家や作品と直接関わりのある語彙と係り受け関係になかったため、必ずしも作曲家の評価を表すものではなかった。

表7 「心」に関する形容詞の出現頻度

|      | Moz. | Bee. | Wag. | Mah. | Bac. | Deb. | Sch. |
|------|------|------|------|------|------|------|------|
| 興味深い | 31   | 37   | ▲ 31 | 7    | ▽ 13 | 18   | 20   |
| 悪い   | 35   | 35   | 17   | 6    | 22   | 12   | 10   |
| 難しい  | 33   | 21   | 18   | 6    | 25   | ▽ 4  | 15   |
| 遠い   | 23   | 20   | 19   | 5    | 16   | 11   | 9    |
| 詳しい  | 11   | 14   | 7    | 6    | 11   | 6    | ▲ 12 |
| ひどい  | 22   | ▲ 21 | 6    | ▲ 9  | 5    | ▽ 1  | 3    |
| 楽しい  | ▲ 41 | 13   | 7    | 6    | 12   | 6    | ▽ 1  |
| 親しい  | 7    | 9    | 9    | 3    | 7    | ▲ 8  | 2    |
| 力強い  | 11   | 6    | 5    | 4    | 8    | 5    | 3    |
| 鋭い   | 5    | 7    | 6    | 3    | 8    | 4    | 6    |

表6 「様相」に関する形容詞の出現頻度

|     | Moz.  | Bee. | Wag. | Mah. | Bac. | Deb. | Sch. |
|-----|-------|------|------|------|------|------|------|
| よい  | 129   | 115  | 79   | 46   | 85   | ▲ 45 | 45   |
| いい  | ▲ 151 | 117  | 76   | ▲ 49 | ▽ 52 | 30   | 31   |
| 美しい | 84    | 70   | 49   | ▽ 11 | 40   | 12   | 17   |
| やすい | 33    | 29   | 26   | 8    | 28   | 10   | 13   |
| 悪い  | 35    | 35   | 17   | 6    | 22   | 12   | 10   |
| うまい | 32    | 29   | 25   | 11   | 22   | 6    | 11   |
| 難しい | 33    | 21   | 18   | 6    | 25   | 4    | 15   |
| 難い  | 19    | 22   | 10   | 2    | ▲ 12 | 9    | ▲ 7  |

## 4 クラスタ分析

分類語彙表における「相」の中項目に該当する感性語の出現頻度を指標にクラスタ分析（Ward法）を行なった。図1は分析の結果として得られた樹形図である。樹形図を闕

値60で切断することによって、ドビュッシー以前と以後の作曲家に分けて考察できるが、ここでは閾値40で切断し、 $C_1 = \{\text{モーツァルト, ベートーヴェン, ワーグナー}\}$ ,  $C_2 = \{\text{バッハ, ドビュッシー}\}$ ,  $C_3 = \{\text{マーラー}\}$ ,  $C_4 = \{\text{シェーンベルク}\}$ の4つのクラスターに分割して考察する。 $C_1$ は古典派からロマン派に活躍した作曲家群である。 $C_2$ にはバロック音楽の全様式を総合したバロック時代最大の作曲家バッハと、19世紀と20世紀の転換期として言及される作曲家ドビュッシーという時代的にも音楽様式的にも対極にある2名が属している。両者に対する感性語の使用傾向が近いことから、評論家はバッハとドビュッシーを評価軸の両端に位置付ける基準を持っていると考えられる。 $C_3$ のマーラーと $C_4$ のシェーンベルクは共に現代の作曲家である。それぞれ独特の音楽性を持つが故、他の作曲家とは異なる切り口から概念が語られるものと考えられる。

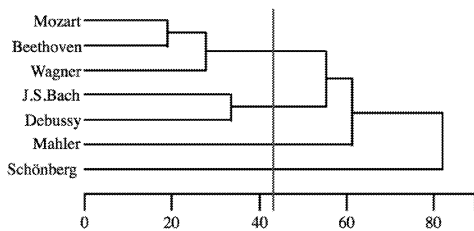


図1 意味分類に基づくクラスタリング結果

## 5 結論と今後の課題

本研究では、人文工学のアプローチに基づき、7名の作曲家に関する感性語利用の特徴と差異を解析することで、各作曲家の感性的特徴を抽出した。得られた結果は雑誌『ポリフォーン』における作曲家概念の様態、流通、共有のありさまを精緻に捉えている。分類語彙表の意味分類に基づくクラスター分析を行なった結果、『ポリフォーン』の評論家は

バッハとドビュッシーを対称的に位置付け、両者を基準に新(マーラー, シェーンベルク)旧(モーツァルト, ベートーヴェン, ワーグナー)の作曲家を捉えていると考えられる。また、モーツァルトへの言及回数が他の作曲家に比べて多く、解析結果は1991年のMozart Year(没後200年)の世相と影響を反映していた。

今後の課題として、音楽以外の芸術活動(文学, 美術, 映画)に関する評論にも同様のアプローチを適用し、抽出された概念を比較することで、創作活動全般にみられる芸術家の概念的特徴を機械可読な形で表現することを目指す。

## 謝辞

本研究は科研費「知識共有のための価値指向型オントロジーの多分野多言語化」(20300074)の助成を受けた。

## 参考文献

- [1] Tokosumi, A. and Murai, H.: "Extracting High-Order Aesthetic and Affective Components From Composer's Writings", Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Human-Computer Interaction. Novel Interaction Methods and Techniques, Vol. 5611/2009, pp. 679-682, 2009.
- [2] 原田隆史, 池内正明: 「書評中の語を用いた図書に対する感性パラメータ自動設定の最適化」, 情報知識学会, Vol. 19, No. 2, pp. 152- 157, 2009.
- [3] 河瀬彰宏, 村井源, 往住彰文: 「音楽評論論文にみる概念構造の変遷」, 情報知識学会, Vol. 19, No. 2, pp. 138- 143, 2009.
- [4] 国立国語研究所: 「分類語彙表-増補改訂版」, 大日本図書, 2004.



## 計量分析による村上春樹文学の語彙構成と歴史的変遷

# Lexical Formation and the Historical Shifts within the Works of Haruki Murakami Identified using Quantitative Analysis

工藤彰<sup>\*1</sup>, 村井源<sup>†</sup>, 往住彰文<sup>†</sup>

Akira KUDO, Hajime MURAI, Akifumi TOKOSUMI

\*1 東京工業大学大学院社会理工学研究科

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

〒152-8552 東京都目黒区大岡山2-21-1

E-mail: {akudo,h\_murai,akt}@valdes.titech.ac.jp

近年、文学や音楽、美術、映画、漫画等のさまざまな芸術を研究するにあたって、微細なデータを扱い、一貫性を有した科学的分析手法に基づいたフィクションの研究に注目が集まっている。本研究では、文学テキストの表層に見られる語彙を作家の文体的特性が宿る最小の構成要素とみなし、作風の変遷を計量的に明らかにすることを目的としている。対象としたデータは、現代の日本文学を代表する小説家村上春樹の長篇とし、そのテキスト中から抽出した語彙を品詞と意味カテゴリーに分類したのち、それぞれクラスター分析を用いて近似的な作風を持つ作品群を明示化した。その結果、品詞分類からは「初期三部作」をはじめとした通時的な区分によつてのクラスターが、また意味分類からは「初期三部作」の続篇とも呼ばれる一作品を含んだ「鼠四部作」のクラスターが形成された。

Similar to the research trends within various artistic realms, such as art, literature, music, movies, and animation, in recent years, fiction studies are increasingly employing more sophisticated techniques of computational analysis. The aims of the present study are to quantitatively analyze the surface vocabulary of a literary text into the minimum components that reflect the writer's characteristic style and to identify style transitions. The target data consists of the full-length works of Haruki Murakami, as a representative of modern Japanese literature. The works are grouped in terms of writing styles through a clustering analysis of the textual vocabulary which was initially classified according to both word class and semantic categories. While the analysis results based on word-class categories yielded cluster that reflect a diachronic division, such as an “early trilogy”, the results based on semantic categories yielded a cluster for the “Nezumi Tetrabiblos” that adds another work which is referred to as a sequel to the “early trilogy”.

キーワード: フィクション, 文学, 村上春樹, 計量書誌学, クラスター分析

Fiction, literature, Haruki Murakami, bibliometrics, cluster analysis

## 1 背景

これまで文学や音楽, 美術, 映画, 漫画をはじめとした諸芸術の意味や解釈の周囲で, さまざまな研究や批評が展開されてきた. 近年, 中でもとりわけフィクションをめぐる言説は加速の一途をたどっており, 文学研究にとどまらず記号論や表象文化論, 社会学, 美学, 言語哲学, 分析哲学あるいは精神分析等の多様な学問領域から議論がなされている. 認知科学の分野において, フィクションとは, 人間である作家の思考概念や創造力の表出と捉えることができ, 数理的ないし統計的手法の進展にともない, 微細なデータを網羅的に処理し一貫性を有した分析手法で, 対象の潜在的な特性や傾向を客観的に顕在化させる記述を可能にしはじめている. そこで信頼性の高い客観的手法が確立できれば, 今後, 諸芸術におけるフィクションの読解技術への汎用性や新たな方法論の開発, そして最終的には人工知能への搭載をも視野に入れることができる. その実現を目指すべく, まずは作家の直接的な創造力の産物であるテキストをデータとして扱うことができる文学に焦点を当てたい.

さらに, 従来の文芸批評や文学研究が言及するのが困難であった文学の周縁的な領域から, テキストの新たな意味や解釈を計算し創出するための応用可能な環境を設計するためにも, 具体的に, 一人の作家が発表した複数のテキストを, 比較検討できるサンプルとして用意する必要がある. また実際に, コンピュータによる文学の解析として, 物語構造から意味のあるまとまりを分割し抽出する研究が進んでいる[1].

以上を踏まえた上で, 本研究の目的は, 文学テキスト上に見られる語彙を, 認知科学的視点に立脚した作家の文体的特性が宿る, データとしての最小の構成要素とみなし, 統計的处理によって算出された意味概念の総和やネットワークをもとに文学を捉え, フィクションの計算処理技術を向上させる長期的展望を視野に入れた試みの一環にあるものとする.

## 2 村上春樹

本研究では, 日本の現代文学を代表する小説家村上春樹の長篇を扱う. 村上春樹という小説家は物語と文体に対して意欲的に取り組んできたことで知られる作家である. 社会や時代の流行や世相を直接的にあるいは間接的に反映させてきたそのテキスト群は, 日本の80年代以降の純文学シーンを代表する小説となっている. 国内外での受賞歴も数多く, 谷崎潤一郎賞や読売文学賞といった文学的に極めて評価の高い国内の文学賞だけでなく, フランツ・カフカ賞やフランク・オコナー短篇賞, エルサレム賞の受賞等の海外での評価や知名度も獲得している. 現存する日本の作家では例外的に世界中で流通する文学性を持ちえた作家といえる. また昨今では, 国内外問わず彼の小説に影響を受けた若手の作家も目立っている. さらに新作発表ごとに文芸批評のみならず, 一般読者の膨大な量の感想がネット上に生まれており, 村上の社会的認知や話題性は極めて高い. これらの理由から, 村上春樹という作家の小説を, 現在の日本文学を牽引する重要なテキストとみなし研究対象として採用することにした.

また作家の思考概念や創造力は当然作品

を産み出す行為の最中で流動的に変容し、それにふさわしいと決定した言葉、あるいは確率論的に導かれた言葉がテキストとして書き連ねられることになる。本研究では、作家がいかに時間的推移の中で言語体系を変化させているのか、その進行中の過程に視線を注ぐべく、通時的に比較可能とする手段として、処女作から8作目までの長篇を一作単位で扱うことにした。短篇やエッセイと比較した際に、長篇は、村上春樹自身、主戦場と位置づけており、本人の特性が最も色濃くあらわれるテキストだといえる。

本研究で8作目までの長篇を取り上げる理由としては、村上春樹という小説家の大きな作風の変化は、オウム事件や阪神大震災による90年代の転向にあるといわれるが[2]、それ以前の時期の作品群と限定し詳細に検討することによって、比較的作風の変化が緩やかに思われる前期作品群の中であっても、村上春樹のフィクションを記述する語彙が、どのようにテキスト上で変貌を遂げていったのか明白にするためである。

一般的に文芸批評で検討されることが多い主題や物語等の大きな枠組みのパラダイムの背後で、いかに微細な言葉がテキストの表象で相対的な遷移を見せるのか、それを科学的に確固たる数値として示し歴然たる所与として吟味することには大きな意義があるだろう。

### 3 データ

まずは、以下に対象とする8作品の基本的な統計の数値の詳細をあげる。形態素解析エンジンMeCabを使用し、テキストを語彙に分割後、さまざまな観点から表記上の集合として捉え処理し、新たに算出し直した。

以下、表と図から見てとれるように、小説

の一文の長さを語数と字数それぞれで調べた結果、平均して15語、25字前後で書かれていることがわかった。

また、日本語の使われ方に関していえば、ひらがなが最も多く増加傾向にあり、漢字、カタカナ、記号は減少の傾向を示した。これは平易でシンプルな文体を使って書きたいという村上春樹自身の発言[3]を裏付けるものである。

TTRは、テキスト自体の語数を踏まえたとしても、総じて年々減少しており、同じ語彙を繰り返し使用する文体に移行している。

また台詞率は、2、5作目を除きほぼ40%前後を維持している。2作目が際立って少ないことについては登場人物の内面まで描いた文体[4]が影響し、また5作目は入れ替わり立ち替わり現れる登場人物の多さがそのまま台詞の多さにつながったと考えられる。

表1 文数や単語数のデータ

|               | 文数    | 単語数    | 文字数    | 平均文長(語) | 平均文長(字) |
|---------------|-------|--------|--------|---------|---------|
| 1 風の歌を聴け      | 2348  | 34508  | 54761  | 14.7    | 23.3    |
| 2 1973年のピンボール | 3052  | 50783  | 81569  | 16.6    | 26.7    |
| 3 羊をめぐる冒険     | 8181  | 136484 | 215115 | 16.7    | 26.3    |
| 4 世界の終り       | 15105 | 258681 | 414663 | 17.1    | 27.5    |
| 5 ノルウェイの森     | 16231 | 185462 | 297464 | 16.5    | 26.5    |
| 6 ダンス・ダンス・ダンス | 18093 | 254603 | 407933 | 14.1    | 22.5    |
| 7 国境の南、太陽の西   | 6462  | 106407 | 168221 | 16.5    | 26      |
| 8 ねじまき鳥クロニクル  | 23442 | 430092 | 694773 | 18.3    | 29.6    |

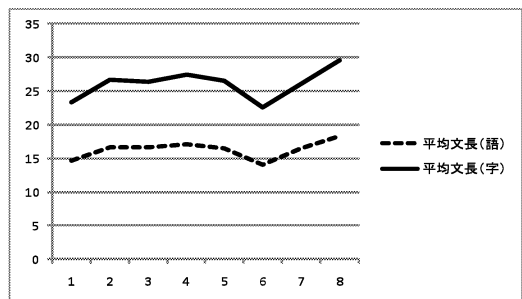


図1 平均文長（語）と平均文長（字）の推移

表2 日本語の表記比率

|               | 漢字(%) | ひらがな(%) | カタカナ(%) | 記号(%) |
|---------------|-------|---------|---------|-------|
| 1 風の歌を聴け      | 23.91 | 55.63   | 9.38    | 9.48  |
| 2 1973年のピンボール | 24.96 | 56.03   | 10.62   | 7.85  |
| 3 羊をめぐる冒険     | 24.61 | 61.48   | 6.3     | 7.19  |
| 4 世界の終り       | 22.18 | 64.74   | 6.13    | 6.5   |
| 5 ノルウェイの森     | 21.39 | 65.9    | 5.3     | 6.94  |
| 6 ダンス・ダンス・ダンス | 22.42 | 62.22   | 7.37    | 7.49  |
| 7 国境の南、太陽の西   | 22.55 | 66.84   | 3.96    | 6.46  |
| 8 ねじまき鳥クロニクル  | 22.77 | 65.89   | 4.89    | 6.15  |

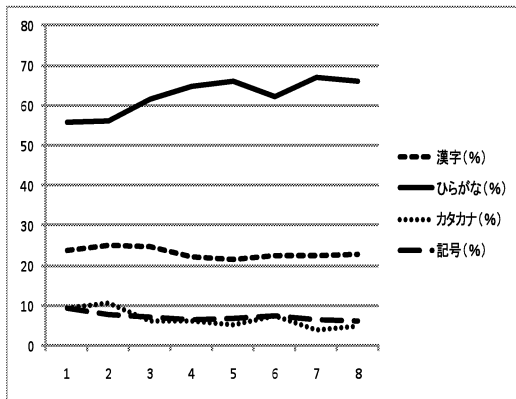


図2 日本語の表記比率推移グラフ

表3 TTRと台詞率

|               | 語彙数   | 語彙数/語数(TTR) | 台詞率(%) |
|---------------|-------|-------------|--------|
| 1 風の歌を聴け      | 3917  | 0.114       | 39.58  |
| 2 1973年のピンボール | 5297  | 0.104       | 21.52  |
| 3 羊をめぐる冒険     | 8265  | 0.061       | 37.58  |
| 4 世界の終り       | 10608 | 0.041       | 41.62  |
| 5 ノルウェイの森     | 8263  | 0.045       | 54.07  |
| 6 ダンス・ダンス・ダンス | 9967  | 0.039       | 39.77  |
| 7 国境の南、太陽の西   | 5282  | 0.05        | 35.91  |
| 8 ねじまき鳥クロニクル  | 13502 | 0.031       | 37.47  |

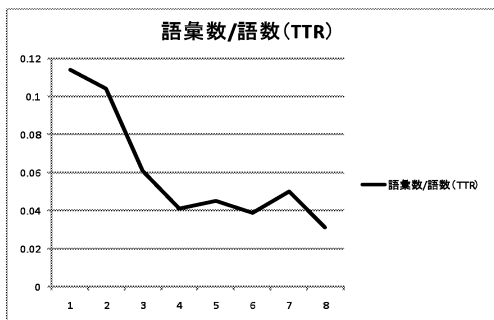


図3 TTRの推移グラフ

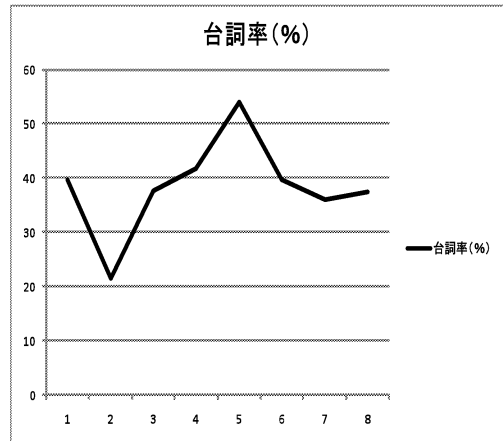


図4 台詞率の推移グラフ

## 4 クラスター分析

### 4.1 品詞分類

クラスター分析に用いるため、名詞、動詞、形容詞、副詞の4品詞を集計する。その目的は語彙が持っている最も基本的な品詞という属性をもとに通時的に比較することで、語彙の大きな変動を俯瞰的に捉えることにある。

表4 品詞の比率

|               | 名詞   | 動詞   | 形容詞  | 副詞   |
|---------------|------|------|------|------|
| 1 風の歌を聴け      | 0.61 | 0.29 | 0.04 | 0.06 |
| 2 1973年のピンボール | 0.63 | 0.28 | 0.04 | 0.05 |
| 3 羊をめぐる冒険     | 0.61 | 0.29 | 0.04 | 0.06 |
| 4 世界の終り       | 0.56 | 0.34 | 0.04 | 0.06 |
| 5 ノルウェイの森     | 0.58 | 0.31 | 0.04 | 0.07 |
| 6 ダンス・ダンス・ダンス | 0.58 | 0.31 | 0.04 | 0.07 |
| 7 国境の南、太陽の西   | 0.59 | 0.30 | 0.04 | 0.07 |
| 8 ねじまき鳥クロニクル  | 0.60 | 0.30 | 0.04 | 0.07 |

### 4.2 意味分類

品詞分類同様クラスター分析のデータとして、語彙を同義語や類義語といった、抽象度の高い意味の枠組みまで一步踏み込んでグルーピングしていくために、本研究は、テ

クスト中から最も作品の特徴やテーマが得られやすい名詞を選択し、『分類語彙表』を用いた。『分類語彙表』からは「人間活動の主体」「人間活動—精神および行為」部門に基づいて、16の分類に分類した。

表5 意味カテゴリーの比率

|               | 人間   | 家族   | 仲間   | 人物   | 成員   | 公私   | 社会   | 機関   |
|---------------|------|------|------|------|------|------|------|------|
| 1 風の歌を聴け      | 0.32 | 0.02 | 0.05 | 0.01 | 0.19 | 0.03 | 0.03 | 0.01 |
| 2 1973年のピンボール | 0.30 | 0.02 | 0.03 | 0.01 | 0.19 | 0.03 | 0.03 | 0.01 |
| 3 羊をめぐる冒険     | 0.28 | 0.02 | 0.04 | 0.02 | 0.19 | 0.04 | 0.03 | 0.01 |
| 4 世界の終り       | 0.34 | 0.02 | 0.03 | 0.01 | 0.09 | 0.08 | 0.03 | 0.01 |
| 5 ノルウェイの森     | 0.37 | 0.02 | 0.03 | 0.01 | 0.17 | 0.05 | 0.03 | 0.01 |
| 6 ダンス・ダンス・ダンス | 0.31 | 0.02 | 0.04 | 0.01 | 0.23 | 0.02 | 0.03 | 0.01 |
| 7 国境の南、太陽の西   | 0.34 | 0.03 | 0.06 | 0.00 | 0.23 | 0.02 | 0.02 | 0.00 |
| 8 ねじまき鳥クロニクル  | 0.30 | 0.03 | 0.02 | 0.01 | 0.16 | 0.06 | 0.02 | 0.01 |

|               | 心    | 言語   | 芸術   | 行為   | 交わり待遇 | 経済   | 事業   |
|---------------|------|------|------|------|-------|------|------|
| 1 風の歌を聴け      | 0.12 | 0.09 | 0.02 | 0.03 | 0.01  | 0.02 | 0.02 |
| 2 1973年のピンボール | 0.14 | 0.10 | 0.02 | 0.04 | 0.01  | 0.02 | 0.03 |
| 3 羊をめぐる冒険     | 0.13 | 0.10 | 0.02 | 0.03 | 0.01  | 0.03 | 0.03 |
| 4 世界の終り       | 0.15 | 0.07 | 0.01 | 0.03 | 0.01  | 0.09 | 0.02 |
| 5 ノルウェイの森     | 0.11 | 0.08 | 0.02 | 0.02 | 0.01  | 0.05 | 0.01 |
| 6 ダンス・ダンス・ダンス | 0.13 | 0.08 | 0.02 | 0.03 | 0.01  | 0.02 | 0.02 |
| 7 国境の南、太陽の西   | 0.11 | 0.07 | 0.02 | 0.03 | 0.01  | 0.02 | 0.01 |
| 8 ねじまき鳥クロニクル  | 0.15 | 0.09 | 0.01 | 0.03 | 0.01  | 0.05 | 0.02 |

### 4.3 品詞分類クラスター

品詞分類後のデータを用いクラスター分析を行った。可視的な分析手法であるクラスター分析を用いることによって、諸パラメータの一覧的なデータからは比較しにくい総体としての対象間の近似性を明らかにすることができる。クラスター化する方法としてはWard法を採用した。

結果、1, 2, 3の通称「初期三部作」、7, 8, また4, 5, 6と完全に年代ごとで集団傾向を示すクラスターが形成された。このことからまさに通時的な区分とともに村上春樹のテキスト文体には特徴があるということが

確認された。

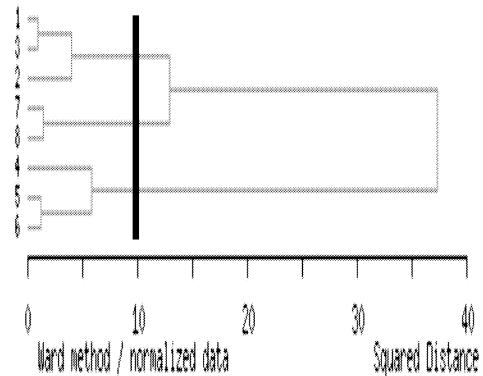


図5 品詞分類クラスター

### 4.4 意味分類クラスター

次に意味カテゴリーにも同様の分析をしたところ、1, 6, 2, 3, また4, 8, そして5, 7のクラスターが見られた。1, 2, 3, 6の4作品は「鼠」というキャラクターが繰り返し登場することから通称「鼠四部作」とも呼ばれており、この実験で扱った意味カテゴリーの語彙で見ても、高い類似性を持つことが明らかとなった。また4作目『世界の終りとハードボイルド・ワンダーランド』と8作目『ねじまき鳥クロニクル』は章によって複数の語りをテキストに組み込んだり[5], 「地下」や「井戸」といった深くて暗い場所の想像力をメタファーとして題材に扱っているといった点が共通している。5作目『ノルウェイの森』と7作目『国境の南、太陽の西』は過去の女性との決して成就されることはない恋愛をめぐる物語である。これらが示すように、一連の続篇小説であったり似通った作風の小説が、抽象的な批評の言説のみならず、文体上の語彙レベルにおいても、幾つかの意味カテゴリーに分布されるという非常に興味深い結果が出た。

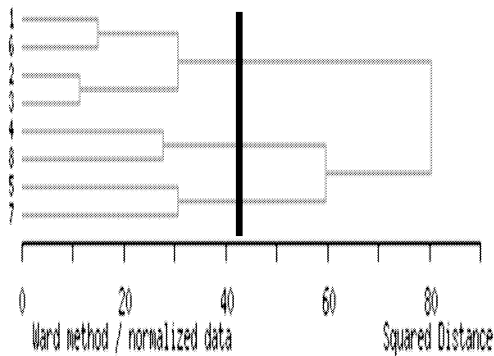


図6 意味分類クラスター

## 5 結論と課題

本研究は、フィクションの科学的解明を目指すべく、文学を対象に、村上春樹の小説テキスト8作品を実データとして、計量的に品詞と意味カテゴリーの二つの分類から、作風の近似性を明示化するためにクラスター分析をおこなった。結果、品詞分類からは「初期三部作」をはじめとした時代ごとの傾向を示すクラスターが、また意味分類からは「初期三部作」の続篇とも呼ばれる6作目の『ダンス・ダンス・ダンス』を含んだ「鼠四部作」のクラスターが形成された。これらの結果が示すのは、意識的であれ無意識的にであれ、作家の文体上の軌跡が、時代の変遷や続篇の作品群というクラスターを示しながら、確かに痕跡となって検出されたということである。

今後の展望であるが、引き続き文学を研究データとして村上春樹の長篇を扱っていく予定である。中でも本研究で触れていない後期の4作品およびそれらを含んだ全12作品の変遷を明らかにしたい。また、本研究の通時

的な観点からは漏れ落ちてしまう一作内での道筋に連動した語彙構成を詳細に考慮する必要もあり、その意味で、小説というフォーマットに適したチャプターに焦点を絞った分析は有効である。さらには、登場人物に着目しその係り受け関係とネットワークを作成することで、テキスト間同士での相関的な近似性や物語上の意味的な位置づけの変化の精査も検討している。これらを含む複数の科学的手法を接ぎ木としながら、文学のテキストにおける表層の計算が可能とするフィクションの潜在的機構の浮上を目指したい。

## 謝辞

本研究は科研費「知識共有のための価値指向型オントロジーの多分野多言語化」(20300074)の助成を受けた。

## 参考文献

- [1] 赤石美奈:「文書群に対する物語構造の動的文解・再構成フレームワーク」, 人工知能学会論文誌, Vol. 21, No. 5, pp. 428-438, 2006.
- [2] 黒古一夫:『村上春樹「喪失」の物語から「転換」の物語へ』, 勉誠出版, 294p. 2007.
- [3] 村上春樹, 柴田元幸:『翻訳夜話』, 文春新書, 245p. 1999.
- [4] 加藤典洋:『村上春樹 イエローページ①』, 幻冬舎, 249p. 2006.
- [5] Jay, Rubin (畔柳和代訳):『ハルキ・ムラカミと言葉の音楽』, 新潮社, 463p. 2006.

## 業績評価に向けた正規化インパクト・ファクター, "IDV : Impact Deviation Value" (インパクト・ファクター偏差値) の提案

### IDV or Impact Deviation Value: A Proposal of a Normalized Impact Factor for Performance Evaluations

根岸正光

Masamitsu NEGISHI

\*国立情報学研究所, 総合研究大学院大学

\* National Institute of Informatics

〒101-8430 千代田区一ツ橋 2-1-2

E-mail: negishi@nii.ac.jp

インパクト・ファクター(IF)は, 発行元トムソン・ロイターの注意喚起にもかかわらず, 研究機関でこれを年次の業績評価に使う例が多い. 一方, 個別論文の実引用数の測定は, 引用蓄積の年数を要し, 年次の評価には適用できない. この種指標で早期に利用可能なのはIFのみであり, これは評価の近似値としては有効と考える. 本稿では, 分野分類ごとのIFの平均値, 標準偏差を計算し, これを各雑誌のIFを偏差値に変換することにより, 分野を超えて比較可能は正規化指標, IDV: Impact Deviation Valueを提案する. この方式による実際の計算結果を示し, 従来のIFとの差異を分析してIDVの効果を検証する. また特にこれによる日本誌の現況についても検討する.

"Impact Factor" (IF) of Journal Citation Reports published by Thomson Reuters (or ISI) is applied to performance evaluation of researchers at research institutions, despite the warning by the publisher that IF should be used only for journal evaluations. But it takes several years to get meaningful citation counts for individual papers. The IF seems to be the only indicator instantly available for the annual evaluations, and it would be quite effective as an early approximation. The author proposes IDV: Impact Deviation Value, a normalized index comparable across fields, based on deviation values of IFs in each field. The actual IDVs are calculated and analyzed to demonstrate the effects with comparison with usual IFs. Special account for the situation of Japanese journals will also be added.

キーワード: 研究評価, インパクト・ファクター, 正規化, 偏差値

Keywords: Research Evaluation, Impact Factor, Normalization, Deviation Value

#### 1 インパクト・ファクター (IF)

インパクト・ファクター(IF)とは, ISI (現トムソン・ロイター社科学事業部) の創始者E.

Garfieldが1955に学術雑誌の重要度の評価指標として考案したもので, 前年の値を計算したものが翌年の6月頃にJournal Citation Reports (JCR)として公刊される. その算式は,

雑誌Aの2008年のIF =

$$\frac{(\text{2006年と2007年の雑誌A掲載論文が2008年にSCIの全収録誌において引用された回数})}{(\text{2006年と2007年の雑誌A掲載論文数})}$$

であり、つまり、前2年分の掲載論文の当年における平均引用回数である（SCI: Science Citation Indexとは、同社の引用索引データベースの自然科学編）[1, 2].

近年、大学等研究機関において、所属研究者の年次の業績評価のために、前年の発表論文の掲載誌のIFを単純に合計して申告させ、これを論文の品質を含めた評価指標として用いるといった方式が多くみられるようである。トムソン社ではこれをIFの「誤用」として注意喚起しているにもかかわらず、これはむしろひろがっているようである。

IFは総体的には各論文末尾の参考文献数に依存するが、研究分野間では論文スタイル（平均的参考文献数）に大差があることから、IFにも分野間で大きな格差がある。すなわちIFは雑誌評価指標としても、分野を超えた比較は無効である。それにもかかわらず、単純比較が行われがちで、このことから、IFが相対的に低い工学系分野などからはIF排斥的な論議もみられる。

個別論文を実際の引用数で「正しく」評価するには、少なくとも公刊後数年以上の観察期間を要するが、これでは前年発表の論文を即座に評価しようとする年次の業績評価には適用できない。實際上、こうした即時的評価に用いる有力な引用指標としてはIFしかないのではないか。

IFは雑誌の格付け指標であるから、その掲載論文をこのIFで評価することは一概に不合理とはいえず、近似的速報値としては一定の有効性があると考えられる。しかし、IFの分野間格差の是正はやはり必要である。

本稿では偏差値を用いてIFの分野間格差を吸収、正規化した指標としてIDV: Impact Deviation Valueを提案し、その効果を計算実例に即して検証する。なおIFの正規化に関してはE. Garfield自身も順位相関係数を用いるものなどを試みているが、計算が煩雑で実用は困難である[3]。この点、本稿にのべるIDVは計算も容易でそれなりの「実用性」があるものと考えている。

## 2 IFの概況

JCRの最新版であるJournal Citation Reports, Science Edition 2008（2009年6月刊）には、全体で6598誌が収録され、その内日本誌は174誌である。

表1 IF上位誌

（Thomson Reuters, JCR Science Edition 2008  
による。以下同様）

|    | 雑誌名                  | IF     | 発行国 | 論文数 | 原著率%  |
|----|----------------------|--------|-----|-----|-------|
| 1  | CA-CANCER J CLIN     | 74.575 | USA | 19  | 47.4  |
| 2  | NEW ENGL J MED       | 50.017 | USA | 356 | 88.8  |
| 3  | ANNU REV IMMUNOL     | 41.059 | USA | 24  | 0.0   |
| 4  | NAT REV MOL CELL BIO | 35.423 | UK  | 84  | 0.0   |
| 5  | PHYSIOL REV          | 35.000 | USA | 40  | 0.0   |
| 6  | REV MOD PHYS         | 33.985 | USA | 36  | 22.2  |
| 7  | JAMA-J AM MED ASSOC  | 31.718 | USA | 225 | 91.1  |
| 8  | NATURE               | 31.434 | UK  | 899 | 94.7  |
| 9  | CELL                 | 31.253 | USA | 348 | 88.2  |
| 10 | NAT REV CANCER       | 30.762 | UK  | 85  | 18.8  |
| 11 | NAT GENET            | 30.259 | USA | 215 | 100.0 |
| 12 | ANNU REV BIOCHEM     | 30.016 | USA | 31  | 0.0   |
| 13 | NAT REV IMMUNOL      | 30.006 | UK  | 86  | 18.6  |
| 14 | NAT REV DRUG DISCOV  | 28.690 | UK  | 62  | 29.0  |
| 15 | LANCET               | 28.409 | UK  | 289 | 94.5  |
| 16 | SCIENCE              | 28.103 | USA | 862 | 93.9  |
| 17 | NAT MED              | 27.553 | USA | 143 | 97.2  |
| 18 | ANNU REV NEUROSCI    | 26.405 | USA | 23  | 0.0   |
| 19 | NAT REV NEUROSCI     | 25.940 | UK  | 71  | 0.0   |
| 20 | ANNU REV ASTRON ASTR | 25.826 | USA | 13  | 0.0   |

収録された雑誌について、IFの高い雑誌は<表1>のようである。生物医学系のものがほとんどで、なかでもレビュー誌が多く、これらにNature, Scienceがまじるという状況である。JCRには雑誌ごとのレビュー論文、原著論文の数が収録されているので、原著論文の割合を「原著率」としてここに掲げた。一方、



論文数の多い雑誌をみると<表 2>のようであり、ここでは物理系、化学系が優位に立つ。

また日本誌のIFの高いものをみると<表 3>のようになる。

表 2 論文数上位誌

|    | 雑誌名                  | IF    | 発行国   | 論文数   | 原著率%  |
|----|----------------------|-------|-------|-------|-------|
| 1  | PHYS REV B           | 3.322 | USA   | 5,782 | 99.3  |
| 2  | APPL PHYS LETT       | 3.726 | USA   | 5,449 | 100.0 |
| 3  | J APPL PHYS          | 2.201 | USA   | 4,168 | 99.6  |
| 4  | PHYS REV LETT        | 7.180 | USA   | 3,905 | 100.0 |
| 5  | J BIOL CHEM          | 5.520 | USA   | 3,761 | 98.4  |
| 6  | ACTA CRYSTALLOGR E   | 0.367 | DENMK | 3,533 | 100.0 |
| 7  | P NATL ACAD SCI USA  | 9.380 | USA   | 3,508 | 99.7  |
| 8  | J AM CHEM SOC        | 8.091 | USA   | 3,242 | 96.6  |
| 9  | J PHYS CHEM C        | 3.396 | USA   | 2,888 | 98.4  |
| 10 | PHYS REV D           | 5.050 | USA   | 2,863 | 96.1  |
| 11 | J GEOPHYS RES        | 3.147 | USA   | 2,860 | 97.4  |
| 12 | J CHEM PHYS          | 3.149 | USA   | 2,763 | 97.3  |
| 13 | PHYS REV A           | 2.908 | USA   | 2,502 | 99.9  |
| 14 | PHYS REV E           | 2.508 | USA   | 2,361 | 99.6  |
| 15 | OPT EXPRESS          | 3.880 | USA   | 2,322 | 99.7  |
| 16 | ASTROPHYS J          | 6.331 | USA   | 2,128 | 92.3  |
| 17 | J PHYS CHEM B        | 4.189 | USA   | 2,103 | 97.5  |
| 18 | BIOCHEM BIOPH RES CO | 2.648 | USA   | 2,095 | 98.4  |
| 19 | J APPL POLYM SCI     | 1.187 | USA   | 2,050 | 100.0 |
| 20 | LANGMUIR             | 4.097 | USA   | 2,026 | 99.5  |
| 21 | JPN J APPL PHYS      | 1.309 | JAPAN | 1,963 | 99.7  |

表 3 日本誌のIF上位誌

|    | 全体順  | 雑誌名                  | IF    | 論文数 | 原著率%  |
|----|------|----------------------|-------|-----|-------|
| 1  | 348  | J PHOTOCH PHOTOBIO C | 5.360 | 9   | 0.0   |
| 2  | 521  | PUBL ASTRON SOC JPN  | 4.429 | 123 | 100.0 |
| 3  | 822  | PLANT CELL PHYSIOL   | 3.542 | 180 | 98.3  |
| 4  | 864  | CANCER SCI           | 3.471 | 343 | 87.5  |
| 5  | 1045 | HYPERTENS RES        | 3.146 | 237 | 98.3  |
| 6  | 1068 | J GASTROENTEROL      | 3.117 | 128 | 85.2  |
| 7  | 1302 | INT J INNOV COMPUT I | 2.791 | 296 | 100.0 |
| 8  | 1420 | DRUG METAB PHARMACOK | 2.641 | 58  | 84.5  |
| 9  | 1440 | J ATHEROSCLER THROMB | 2.625 | 39  | 92.3  |
| 10 | 1456 | J PHARMACOL SCI      | 2.599 | 218 | 88.1  |
| 11 | 1610 | J HUM GENET          | 2.431 | 129 | 99.2  |
| 12 | 1662 | CIRC J               | 2.387 | 359 | 99.4  |
| 13 | 1761 | DEV GROWTH DIFFER    | 2.317 | 100 | 44.0  |
| 14 | 1970 | PARASITOL INT        | 2.152 | 82  | 93.9  |
| 15 | 1972 | CELL STRUCT FUNCT    | 2.152 | 21  | 100.0 |
| 16 | 2029 | J BONE MINER METAB   | 2.100 | 90  | 90.0  |
| 17 | 2094 | J PHYS SOC JPN       | 2.058 | 542 | 99.1  |
| 18 | 2297 | J HEPATO-BILIARY-PAN | 1.914 | 105 | 91.4  |
| 19 | 2328 | POPUL ECOL           | 1.895 | 40  | 97.5  |
| 20 | 2361 | J BIOCHEM            | 1.878 | 179 | 92.2  |

次に分野別の収録状況をまとめると<表 4>のようである。雑誌数、論文数の1位は臨床医学で全体の4分の1程度を占めるが、こ

れに続いて工学、化学、物理という、後述のとおりIFが高くなく、IFに対して批判的な分野が来るのが興味深い。

同表右側に日本誌だけの統計を掲げた。最右欄は論文数における日本誌のシェアであり、全体的には2%程度であるが、農学、薬学が5.2%で相対的に大きな部分を占めるのが注目される。これは、この分野のわが国の雑誌の国際的地位が相対的に高いことを表している。

レビュー論文はその性格上引用されることが多く、従ってレビュー論文の多い雑誌のIFは高くなる。この状況を実際に統計をとってみたのが<表 5>である。なお、ここでは前述の原著率50%未満の雑誌をレビュー誌として統計した。さすがにレビュー誌のIFは高く、全体的には非レビュー誌の3.8倍に達することが分かる。最右欄は、両IFの比をとったものである。レビュー誌が少ない分野ではこれが大きな値になる傾向がみられ、こうしたレビュー誌への引用の集中が看取されるが、これらは原著率が0に近い純粋レビュー誌なのであろう。

ともあれ、レビュー誌はこの際攪乱要因であるので、以下の分析では、原著率50%以上の非レビュー誌6047誌に限って統計することにした。

### 3 IFの分野別平均、標準偏差

前掲<表 5>にみるとおり、分野別の平均IFは非レビュー誌の場合でも分子生物学・遺伝学の3.489から数学の0.934に至り、分野間で3.8倍ほどの格差がある。すなわち、分野を超えてIFを比較することが無意味であることがわかる。

表 4 分野別雑誌数・論文数と日本誌のシェア

| 分野分類                         | 全体   |       |           |       | 日本誌 |       |        |       |         |
|------------------------------|------|-------|-----------|-------|-----|-------|--------|-------|---------|
|                              | 雑誌数  | 構成比   | 論文数       | 構成比%  | 雑誌数 | 構成比   | 論文数    | 構成比%  | 論文数シェア% |
| Clinical Medicine            | 1776 | 26.9  | 235,955   | 25.3  | 38  | 21.8  | 4,271  | 19.7  | 1.8     |
| Engineering                  | 1157 | 17.5  | 146,853   | 15.7  | 33  | 19.0  | 4,116  | 19.0  | 2.8     |
| Chemistry                    | 544  | 8.2   | 145,444   | 15.6  | 17  | 9.8   | 3,509  | 16.2  | 2.4     |
| Physics                      | 422  | 6.4   | 130,497   | 14.0  | 7   | 4.0   | 3,048  | 14.1  | 2.3     |
| Biology & Biochemistry       | 670  | 10.2  | 101,494   | 10.9  | 12  | 6.9   | 1,358  | 6.3   | 1.3     |
| Materials Science            | 342  | 5.2   | 63,628    | 6.8   | 9   | 5.2   | 1,163  | 5.4   | 1.8     |
| Plant & Animal Sciences      | 521  | 7.9   | 45,460    | 4.9   | 18  | 10.3  | 1,204  | 5.6   | 2.6     |
| Molecular Biology & Genetics | 334  | 5.1   | 41,327    | 4.4   | 8   | 4.6   | 518    | 2.4   | 1.3     |
| Geosciences                  | 425  | 6.4   | 41,167    | 4.4   | 8   | 4.6   | 548    | 2.5   | 1.3     |
| Agricultural Sciences        | 353  | 5.4   | 38,286    | 4.1   | 19  | 10.9  | 2,008  | 9.3   | 5.2     |
| Mathematics                  | 477  | 7.2   | 38,019    | 4.1   | 12  | 6.9   | 477    | 2.2   | 1.3     |
| Microbiology                 | 259  | 3.9   | 37,844    | 4.1   | 7   | 4.0   | 1,047  | 4.8   | 2.8     |
| Ecology / Environment        | 273  | 4.1   | 34,570    | 3.7   | 4   | 2.3   | 278    | 1.3   | 0.8     |
| Pharmacology                 | 274  | 4.2   | 34,239    | 3.7   | 9   | 5.2   | 1,787  | 8.2   | 5.2     |
| Neurosciences                | 219  | 3.3   | 29,068    | 3.1   | 2   | 1.1   | 219    | 1.0   | 0.8     |
| Computer Sciences            | 394  | 6.0   | 28,029    | 3.0   | 4   | 2.3   | 797    | 3.7   | 2.8     |
| Immunology                   | 126  | 1.9   | 18,920    | 2.0   | 2   | 1.1   | 229    | 1.1   | 1.2     |
| Space Sciences               | 48   | 0.7   | 13,974    | 1.5   | 1   | 0.6   | 198    | 0.9   | 1.4     |
| Psychology/Psychiatry        | 146  | 2.2   | 13,708    | 1.5   | 1   | 0.6   | 105    | 0.5   | 0.8     |
| Multidisciplinary            | 42   | 0.6   | 10,435    | 1.1   | 1   | 0.6   | 28     | 0.1   | 0.3     |
| Social Sciences, General     | 75   | 1.1   | 6,872     | 0.7   | 1   | 0.6   | 39     | 0.2   | 0.6     |
| Education                    | 23   | 0.3   | 2,087     | 0.2   |     |       |        |       | 0.0     |
| Arts & Humanities            | 41   | 0.6   | 1,126     | 0.1   |     |       |        |       | 0.0     |
| のべ計                          | 8941 | 135.5 | 1,259,002 | 135.0 | 213 | 122.4 | 26,947 | 124.3 | 2.1     |
| 実数                           | 6598 | 100.0 | 932,795   | 100.0 | 174 | 100.0 | 21,681 | 100.0 | 2.3     |

表 5 レビュー誌と非レビュー誌の差異

| 分野分類                         | 雑誌数  |        |       | IFの平均値 |        |        |         |
|------------------------------|------|--------|-------|--------|--------|--------|---------|
|                              | 全体   | 非レビュー誌 | レビュー誌 | 全体     | 非レビュー誌 | レビュー誌  | RV / NR |
| Molecular Biology & Genetics | 334  | 297    | 25    | 3.898  | 3.489  | 9.321  | 2.7     |
| Immunology                   | 126  | 103    | 19    | 4.011  | 3.230  | 8.288  | 2.6     |
| Neurosciences                | 219  | 189    | 19    | 3.434  | 2.933  | 8.735  | 3.0     |
| Multidisciplinary            | 42   | 39     | 2     | 2.705  | 2.666  | 4.788  | 1.8     |
| Psychology/Psychiatry        | 146  | 139    | 5     | 2.828  | 2.661  | 7.488  | 2.8     |
| Biology & Biochemistry       | 670  | 593    | 52    | 3.013  | 2.645  | 7.245  | 2.7     |
| Space Sciences               | 48   | 44     | 4     | 3.065  | 2.388  | 10.516 | 4.4     |
| Microbiology                 | 259  | 229    | 23    | 2.827  | 2.362  | 7.611  | 3.2     |
| Clinical Medicine            | 1776 | 1634   | 106   | 2.422  | 2.262  | 5.074  | 2.2     |
| Pharmacology                 | 274  | 224    | 43    | 2.732  | 2.134  | 5.801  | 2.7     |
| Social Sciences, General     | 75   | 70     | 4     | 2.343  | 1.984  | 8.526  | 4.3     |
| Ecology / Environment        | 273  | 252    | 11    | 1.977  | 1.833  | 6.276  | 3.4     |
| Chemistry                    | 544  | 501    | 37    | 2.101  | 1.764  | 6.596  | 3.7     |
| Physics                      | 422  | 385    | 22    | 1.953  | 1.648  | 7.801  | 4.7     |
| Materials Science            | 342  | 323    | 8     | 1.623  | 1.493  | 7.595  | 5.1     |
| Geosciences                  | 425  | 404    | 8     | 1.473  | 1.440  | 4.174  | 2.9     |
| Computer Sciences            | 394  | 383    | 4     | 1.392  | 1.377  | 4.059  | 2.9     |
| Plant & Animal Sciences      | 521  | 490    | 9     | 1.367  | 1.214  | 10.488 | 8.6     |
| Engineering                  | 1157 | 1097   | 17    | 1.177  | 1.140  | 4.717  | 4.1     |
| Agricultural Sciences        | 353  | 333    | 6     | 1.149  | 1.133  | 2.512  | 2.2     |
| Mathematics                  | 477  | 465    | 1     | 0.930  | 0.934  | 2.227  | 2.4     |
| Education                    | 23   | 21     | 0     | 0.936  | 0.885  |        |         |
| Arts & Humanities            | 41   | 38     | 0     | 0.602  | 0.623  |        |         |
| のべ計                          | 8941 | 8253   | 425   |        |        |        |         |
| 実数                           | 6598 | 6074   | 345   | 2.048  | 1.798  | 6.766  | 3.8     |

表 6 IF(インパクト・ファクター)の分野別平均, 標準偏差(非レビュー誌)

| 分野分類                         | 雑誌数  | IF<br>平均 | IF<br>標準偏差 | 分野別最大IF (IF第1位誌) |                      |      |
|------------------------------|------|----------|------------|------------------|----------------------|------|
|                              |      |          |            | 最大IF             | 雑誌名                  | 総合順位 |
| Molecular Biology & Genetics | 297  | 3.489    | 3.736      | 31.253           | CELL                 | 4    |
| Immunology                   | 103  | 3.230    | 3.448      | 25.113           | NAT IMMUNOL          | 12   |
| Neurosciences                | 189  | 2.933    | 2.203      | 14.170           | NEURON               | 35   |
| Multidisciplinary            | 39   | 2.666    | 6.515      | 31.434           | NATURE               | 3    |
| Psychology/Psychiatry        | 139  | 2.661    | 2.225      | 14.273           | ARCH GEN PSYCHIAT    | 34   |
| Biology & Biochemistry       | 593  | 2.645    | 2.576      | 31.253           | CELL                 | 4    |
| Space Sciences               | 44   | 2.388    | 2.393      | 13.990           | ASTROPHYS J SUPPL S  | 37   |
| Microbiology                 | 229  | 2.362    | 2.095      | 22.297           | NAT BIOTECHNOL       | 18   |
| Clinical Medicine            | 1634 | 2.262    | 2.534      | 50.017           | NEW ENGL J MED       | 1    |
| Pharmacology                 | 224  | 2.134    | 1.287      | 7.586            | CLIN PHARMACOL THER  | 113  |
| Social Sciences, General     | 70   | 1.984    | 1.000      | 4.074            | NEUROPSYCHOLOGIA     | 599  |
| Ecology / Environment        | 252  | 1.833    | 1.322      | 9.392            | ECOL LETT            | 77   |
| Chemistry                    | 501  | 1.764    | 1.747      | 23.132           | NAT MATER            | 15   |
| Physics                      | 385  | 1.648    | 2.149      | 24.982           | NAT PHOTONICS        | 13   |
| Materials Science            | 323  | 1.493    | 2.202      | 23.132           | NAT MATER            | 15   |
| Geosciences                  | 404  | 1.440    | 0.993      | 5.304            | GLOBAL ECOL BIOGEOG  | 285  |
| Computer Sciences            | 383  | 1.377    | 0.983      | 6.800            | Vldb J               | 150  |
| Plant & Animal Sciences      | 490  | 1.214    | 1.048      | 9.296            | PLANT CELL           | 79   |
| Engineering                  | 1097 | 1.140    | 0.942      | 8.479            | INT J NONLIN SCI NUM | 92   |
| Agricultural Sciences        | 333  | 1.133    | 0.776      | 4.453            | BIORESOURCE TECHNOL  | 464  |
| Mathematics                  | 465  | 0.934    | 0.736      | 8.479            | INT J NONLIN SCI NUM | 92   |
| Education                    | 21   | 0.885    | 0.545      | 2.181            | MED EDUC             | 2315 |
| Arts & Humanities            | 38   | 0.623    | 0.690      | 4.378            | AM J BIOETHICS       | 485  |
| ALL                          | 6074 | 1.798    | 2.150      | 50.017           | NEW ENGL J MED       | 1    |
| のべ計                          | 8253 |          |            |                  |                      |      |

筆者は朝日新聞社刊「大学ランキング」各年版に大学別の論文引用度に基づくランキングを2001年以来寄稿している。その際、分野間の引用度の格差を是正するために、偏差値を応用した引用度指標を考案して用い、これによってほぼ妥当な順位付け結果を得ている[4, 5]。

そこで、本稿の目的であるIFの分野間正規化においても偏差値を応用できるのではないかと考え、後述のようなIDVの計算方式を試みた。

ここでまず、計算の基礎データであるIFの分野別平均と標準偏差、および最大値とそれに対応する雑誌名をまとめると<表 6>のようである。

#### 4 IDV: Impact Deviation Valueの提案

<表 6>のIFの分野別平均と標準偏差により、例えばNew England Journal of Medicine(臨床医学分野)の偏差値を計算すると次のようで、238.4となる。

#### 《NEW ENGL J MED(Clinical Medicine)の計算例》

$$\begin{aligned} & (50.017 - 2.262) \div 2.534 \times 10 + 50 = 238.4 \\ & \text{(NEJMのIF - Clin Medの平均IF)} \\ & \div \text{Clin MedのIFの標準偏差} \times 10 + 50 \\ & = 238.4 \quad \leftarrow \text{IDV} \end{aligned}$$

JCRでの雑誌の分野分類では、複数の分野に重複分類されている雑誌も多い。これらについては、各分野別に偏差値を計算してその平均値をとることとする。例えばNature Materialsは化学、材料科学、物理の3分野に重複分類されているので、次記のような計算になり、156.9の値を得る。

#### 《NATURE MATERIALS (Chemistry, Materials Science, Physics)の計算例》

$$\begin{aligned} & \text{Chemistry:} \\ & (23.132 - 1.764) \div 1.747 \times 10 + 50 = 172.3 \\ & \text{Materials Science:} \\ & (23.132 - 1.493) \div 2.202 \times 10 + 50 = 148.3 \\ & \text{Physics:} \\ & (23.132 - 1.648) \div 2.149 \times 10 + 50 = 150.0 \\ & \text{Average:} \\ & (172.3 + 148.3 + 150.0) \div 3 = 156.9 \quad \leftarrow \text{IDV} \end{aligned}$$

このようにして計算するIFの分野間平均偏差値を、ここでは"IDV: Impact Deviation Value" (インパクト・ファクター偏差値) と命名する。

## 5 IDV: Impact Deviation Valueの効果

IDV: Impact Deviation Valueを非レビュー誌について計算した結果として、その上位を

占める雑誌を<表 7>に示す。ここではIDVと従来のIFにおける順位の変動幅も算出し、これに注目してIDVの「効果」を検討する。

まずNew Engle J Medはじめ、最上位部には不動の上位誌が列ぶが、Int J Nonlin Sci Num (数学) が60位上昇移動して11位になるなど、平均IFの低い分野の雑誌も現れてきて、IDVによる分野間調整効果が確認できる。

表 7 IDV上位誌

| IDV Rank | Title                | IDV   | IF     | IF Rank | Rank Shift | IDV Rank | Title                | IDV  | IF     | IF Rank | Rank Shift |
|----------|----------------------|-------|--------|---------|------------|----------|----------------------|------|--------|---------|------------|
| 1        | NEW ENGL J MED       | 238.4 | 50.017 | 1       | 0          | 31       | PLANT PHYSIOL        | 96.7 | 6.110  | 140     | +109       |
| 2        | JAMA-J AM MED ASSOC  | 166.2 | 31.718 | 2       | 0          | 32       | NAT CHEM BIOL        | 96.5 | 14.612 | 25      | -7         |
| 3        | NAT PHOTONICS        | 158.6 | 24.982 | 10      | +7         | 33       | STRUCT EQU MODELING  | 96.4 | 4.351  | 360     | +327       |
| 4        | NAT MATER            | 156.9 | 23.132 | 12      | +8         | 34       | IEEE T IND ELECTRON  | 95.9 | 5.468  | 186     | +152       |
| 5        | LANCET               | 153.2 | 28.409 | 6       | +1         | 35       | CMES-COMP MODEL ENG  | 95.5 | 4.785  | 280     | +245       |
| 6        | NAT BIOTECHNOL       | 145.1 | 22.297 | 13      | +7         | 36       | NANO LETT            | 94.8 | 10.371 | 52      | +16        |
| 7        | CELL                 | 142.7 | 31.253 | 4       | -3         | 37       | NATURE               | 94.2 | 31.434 | 3       | -34        |
| 8        | CANCER CELL          | 139.6 | 24.962 | 11      | +3         | 38       | LANCET ONCOL         | 93.5 | 13.283 | 33      | -5         |
| 9        | NAT MED              | 137.0 | 27.553 | 8       | -1         | 39       | CELL METAB           | 93.0 | 16.107 | 22      | -17        |
| 10       | NAT NANOTECHNOL      | 136.7 | 20.571 | 15      | +5         | 40       | J EXP MED            | 93.0 | 15.219 | 23      | -17        |
| 11       | NAT GENET            | 121.7 | 30.259 | 5       | -6         | 41       | NAT METHODS          | 92.7 | 13.651 | 31      | -10        |
| 12       | INT J NONLIN SCI NUM | 120.7 | 8.479  | 72      | +60        | 42       | CLIN PHARMACOL THER  | 92.4 | 7.586  | 84      | +42        |
| 13       | NAT PHYS             | 120.6 | 16.821 | 20      | +7         | 43       | MOL PSYCHIATR        | 92.1 | 12.537 | 41      | -2         |
| 14       | NAT IMMUNOL          | 113.5 | 25.113 | 9       | -5         | 44       | BRIT MED J           | 91.7 | 12.827 | 37      | -7         |
| 15       | ANN INTERN MED       | 110.0 | 17.457 | 17      | +2         | 45       | PSYCHOL REV          | 90.9 | 11.765 | 44      | -1         |
| 16       | J CLIN ONCOL         | 108.8 | 17.157 | 18      | +2         | 46       | BIOMATERIALS         | 90.9 | 6.646  | 115     | +69        |
| 17       | ECOL LETT            | 107.2 | 9.392  | 61      | +44        | 47       | GASTROENTEROLOGY     | 90.8 | 12.591 | 39      | -8         |
| 18       | J CLIN INVEST        | 106.4 | 16.559 | 21      | +3         | 48       | INT J COMPUT VISION  | 90.5 | 5.358  | 202     | +154       |
| 19       | Vldb J               | 105.2 | 6.800  | 110     | +91        | 49       | ECONOMETRICA         | 89.8 | 3.865  | 490     | +441       |
| 20       | ARCH GEN PSYCHIAT    | 102.2 | 14.273 | 27      | +7         | 50       | PLANT CELL           | 89.5 | 9.296  | 63      | +13        |
| 21       | ANGEW CHEM INT EDIT  | 102.2 | 10.879 | 48      | +27        | 51       | PLOS MED             | 89.2 | 12.185 | 43      | -8         |
| 22       | MATER TODAY          | 101.9 | 12.929 | 34      | +12        | 52       | SCIENCE              | 89.0 | 28.103 | 7       | -45        |
| 23       | NEURON               | 101.0 | 14.170 | 28      | +5         | 53       | COMMUN PUR APPL MATH | 89.0 | 3.806  | 503     | +450       |
| 24       | NAT NEUROSCI         | 101.0 | 14.164 | 29      | +5         | 54       | PLOS BIOL            | 89.0 | 12.683 | 38      | -16        |
| 25       | PLANT J              | 100.4 | 6.493  | 124     | +99        | 55       | MIS QUART            | 88.7 | 5.183  | 225     | +170       |
| 26       | IMMUNITY             | 100.3 | 20.579 | 14      | -12        | 56       | B AM METEOROL SOC    | 88.4 | 5.252  | 214     | +56        |
| 27       | J NATL CANCER I      | 100.0 | 14.933 | 24      | -3         | 57       | NAT CELL BIOL        | 88.2 | 17.774 | 16      | -41        |
| 28       | IEEE T PATTERN ANAL  | 98.9  | 5.960  | 150     | +122       | 58       | NEW PHYTOT           | 87.8 | 5.178  | 226     | +168       |
| 29       | CIRCULATION          | 98.7  | 14.595 | 26      | -3         | 59       | GLOBAL CHANGE BIOL   | 87.5 | 5.876  | 158     | +99        |
| 30       | ASTROPHYS J SUPPL S  | 98.5  | 13.990 | 30      | 0          | 60       | MOL SYST BIOL        | 87.3 | 12.243 | 42      | -18        |

その一方でIDVによって順位が大きく低落了た雑誌を抽出すると<表 8>のようで、P NATL ACAD SCI USA (PNAS)の512位という大幅低落が目立つが、これは学際分野で、後述のとおり別途の検討を要する。

ともあれ、全体的には、IFで30~100位辺りあった生物医学系雑誌はIDVでは大きく順位を落とし、これらに代わって工学、理学等の有力誌がIDVの上位に進出するという状況で、IDVは相当の調整機能を果たしうるものと考えられる。

表 8 凋落誌

| IF Rank | IDV Rank | Title                | IDV  | IF     | Rank Shift |
|---------|----------|----------------------|------|--------|------------|
| 32      | 99       | GENE DEV             | 77.1 | 13.623 | -67        |
| 36      | 119      | DEV CELL             | 75.1 | 12.882 | -83        |
| 40      | 133      | CURR OPIN CELL BIOL  | 74.2 | 12.543 | -93        |
| 47      | 109      | NAT STRUCT MOL BIOL  | 76.2 | 10.987 | -62        |
| 54      | 231      | AM J HUM GENET       | 67.8 | 10.153 | -177       |
| 58      | 207      | J ALLERGY CLIN IMMUN | 69.0 | 9.773  | -149       |
| 62      | 574      | P NATL ACAD SCI USA  | 60.3 | 9.380  | -512       |
| 65      | 315      | J CELL BIOL          | 65.1 | 9.120  | -250       |
| 68      | 335      | PLOS GENET           | 64.4 | 8.883  | -267       |
| 69      | 137      | MOL CELL PROTEOMICS  | 74.0 | 8.834  | -68        |
| 73      | 127      | ANN SURG             | 74.5 | 8.460  | -54        |
| 74      | 153      | DIABETES             | 72.3 | 8.398  | -79        |
| 75      | 244      | EMBO J               | 67.4 | 8.295  | -169       |
| 76      | 154      | CLIN INFECT DIS      | 72.2 | 8.266  | -78        |
| 78      | 141      | CURR OPIN NEUROBIOL  | 73.5 | 8.102  | -63        |
| 80      | 388      | J AUTOIMMUN          | 63.5 | 7.881  | -308       |
| 81      | 479      | SYST BIOL            | 61.6 | 7.833  | -398       |
| 82      | 267      | AGING CELL           | 66.7 | 7.791  | -185       |
| 83      | 196      | STEM CELLS           | 69.6 | 7.741  | -113       |
| 85      | 317      | CELL DEATH DIFFER    | 64.9 | 7.548  | -232       |

なお、Nature、Scienceという著名誌は「学際」に分類されているが、この分類は対象誌が少ない上に、両誌のIFのみが突出した状況にあり（これに続いてPNAS）、これは標準偏差の大きさにも表れている。この点を踏まえ、両誌やPNASについては今後何らかの調整を検討するべきものと考えている。

## 6 日本誌の状況

次に日本誌だけをとりだしてまとめたのが<表 9>である。順位移動では不動の上位誌がある一方で、Primates、J Plant Resという動植物学系の上位進出が注目される。

しかし、ここでまず注意すべきは、IDVは

偏差値であり、従って、その 50 が各分野における国際平均水準であるということである。そして、IDV:50 以上の日本誌は 19 誌に止まるという点であろう。日本誌の振興を図るには、まずこのIDV:50 の水準を目指して、各誌において各般の努力がなされるべきものと思われるのである。

それには、まず当該分野においてIDVが少し高い雑誌（IFが少し高い雑誌でもある）の状況を分析的に観察し、自誌との比較を試みるといったことが手始めではなかろうか。しかし従来、こうしたことはあまり行われてこなかったようで、この際、あえて提案しておきたい。

表 9 日本誌のIDV

| 日本誌内<br>IDV順 | 雑誌名                  | IDV  | IF    | 日本誌内 |       | 非REV全誌 |      |       | 論文数  |
|--------------|----------------------|------|-------|------|-------|--------|------|-------|------|
|              |                      |      |       | IF順位 | 順位シフト | IDV順位  | IF順位 | 順位シフト |      |
| 1            | INT J INNOV COMPUT I | 66.0 | 2.791 | 6    | +5    | 293    | 1036 | +743  | 296  |
| 2            | PLANT CELL PHYSIOL   | 61.2 | 3.542 | 2    | 0     | 516    | 609  | +93   | 180  |
| 3            | PUBL ASTRON SOC JPN  | 58.5 | 4.429 | 1    | -2    | 707    | 343  | -364  | 123  |
| 4            | CANCER SCI           | 54.8 | 3.471 | 3    | -1    | 1128   | 645  | -483  | 343  |
| 5            | PRIMATES             | 54.3 | 1.670 | 25   | +20   | 1201   | 2348 | +1147 | 47   |
| 6            | DRUG METAB PHARMACOK | 53.9 | 2.641 | 7    | +1    | 1265   | 1146 | -119  | 58   |
| 7            | J PHARMACOL SCI      | 53.6 | 2.599 | 9    | +2    | 1330   | 1180 | -150  | 218  |
| 8            | J PLANT RES          | 53.6 | 1.590 | 31   | +23   | 1332   | 2463 | +1131 | 66   |
| 9            | HYPERTENS RES        | 53.5 | 3.146 | 4    | -5    | 1350   | 808  | -542  | 237  |
| 10           | J GASTROENTEROL      | 53.4 | 3.117 | 5    | -5    | 1376   | 830  | -546  | 128  |
| 11           | J BIOSCI BIOENG      | 52.1 | 1.702 | 23   | +12   | 1613   | 2307 | +694  | 217  |
| 12           | J PHYS SOC JPN       | 51.9 | 2.058 | 15   | +3    | 1664   | 1785 | +121  | 542  |
| 13           | J ATHEROSCLER THROMB | 51.4 | 2.625 | 8    | -5    | 1786   | 1165 | -621  | 39   |
| 14           | J REPROD DEVELOP     | 51.1 | 1.609 | 29   | +15   | 1895   | 2439 | +544  | 85   |
| 15           | CIRC J               | 50.5 | 2.387 | 11   | -4    | 2048   | 1376 | -672  | 359  |
| 16           | POPUL ECOL           | 50.5 | 1.895 | 17   | +1    | 2057   | 2009 | -48   | 40   |
| 17           | SOIL SCI PLANT NUTR  | 50.2 | 1.152 | 54   | +37   | 2128   | 3328 | +1200 | 102  |
| 18           | PROG THEOR PHYS      | 50.1 | 1.661 | 26   | +8    | 2172   | 2360 | +188  | 111  |
| 19           | J ELECTRON MICROSC   | 50.0 | 1.139 | 56   | +37   | 2200   | 3350 | +1150 | 30   |
| 20           | ANAL SCI             | 49.8 | 1.735 | 22   | +2    | 2254   | 2258 | +4    | 277  |
| 21           | PARASITOL INT        | 49.6 | 2.152 | 12   | -9    | 2336   | 1667 | -669  | 82   |
| 22           | B CHEM SOC JPN       | 49.5 | 1.677 | 24   | +2    | 2351   | 2337 | -14   | 199  |
| 23           | ZOOL SCI             | 48.9 | 1.100 | 57   | +34   | 2535   | 3454 | +919  | 152  |
| 24           | J HEPATO-BILIARY-PAN | 48.6 | 1.914 | 16   | -8    | 2639   | 1981 | -658  | 105  |
| 25           | J BONE MINER METAB   | 48.6 | 2.100 | 14   | -11   | 2644   | 1723 | -921  | 90   |
| 26           | JPN J APPL PHYS      | 48.4 | 1.309 | 43   | +17   | 2714   | 2988 | +274  | 1963 |
| 27           | CHEM LETT            | 48.4 | 1.478 | 33   | +6    | 2738   | 2653 | -85   | 611  |
| 28           | ODONTOLOGY           | 48.3 | 1.833 | 19   | -9    | 2760   | 2120 | -640  | 11   |
| 29           | POLYM J              | 48.2 | 1.456 | 35   | +6    | 2784   | 2693 | -91   | 176  |
| 30           | BREEDING SCI         | 48.0 | 0.989 | 65   | +35   | 2869   | 3731 | +862  | 50   |
| 31           | BIOSCI BIOTECH BIOCH | 47.9 | 1.390 | 40   | +9    | 2903   | 2834 | -69   | 500  |
| 32           | J CERAM SOC JPN      | 47.9 | 1.023 | 60   | +28   | 2931   | 3643 | +712  | 261  |
| 33           | CHEM PHARM BULL      | 47.6 | 1.623 | 28   | -5    | 3032   | 2416 | -616  | 357  |
| 34           | J EPIDEMIOL          | 47.6 | 1.642 | 27   | -7    | 3047   | 2386 | -661  | 36   |
| 35           | J OCEANOGR           | 47.5 | 1.189 | 52   | +17   | 3076   | 3253 | +177  | 79   |
| 36           | J HUM GENET          | 47.2 | 2.431 | 10   | -26   | 3185   | 1324 | -1861 | 129  |
| 37           | MED MOL MORPHOL      | 47.1 | 1.536 | 32   | -5    | 3203   | 2554 | -649  | 38   |
| 38           | BIOL PHARM BULL      | 47.1 | 1.765 | 21   | -17   | 3206   | 2210 | -996  | 438  |
| 39           | J BIOCHEM            | 47.0 | 1.878 | 18   | -21   | 3250   | 2045 | -1205 | 179  |
| 40           | EXTREMOPHILES        | 46.9 | 1.782 | 20   | -20   | 3284   | 2184 | -1100 | 84   |

## 7 まとめ

本稿では、IFが分野間格差の調整もなく、そのまま年次的業績評価に誤用されるという現実に鑑み、その分野間の調整、正規化の試みとして、IDV: Impact Deviation Valueを提案した。

**業績評価** — IDVは<表 6>の太枠内に示した分野別のIFの平均値と標準偏差があればすぐに計算でき、これによって分野間におけるIFの格差は大幅に是正される。工学系等、IFの低い分野からのIFに対する批判、不満も、このIDVへの換算によってよほど軽減されるのではないかと考えるところである。

**日本誌振興** — また雑誌編集の立場から、IDVは国際平均水準：50、標準偏差：10という正規化された値であるので、自誌の国際的位置付けが容易に確認できる点で、有用と思われる。これを参照することにより、日本の学会誌が引用度の向上に向けて、有効な対策を検討されるよう期待したい。

### <参考>

- [1] 根岸正光: 研究評価・雑誌評価のためのビブリオメトリックス指標: 現状と課題, (第6回SPARC Japanセミナー2008, 「IFを超えて: さらなる研究評価の在り方を考える」, 2008.11), 11p. (2009.5.27 掲載)  
[http://www.nii.ac.jp/sparc/event/2008/pdf/112508/document/Prof.Negishi\\_document\\_final.pdf](http://www.nii.ac.jp/sparc/event/2008/pdf/112508/document/Prof.Negishi_document_final.pdf)
- [2] トムソン・ロイター: Journal Citation Reports.  
<http://www.thomsonscientific.jp/products/jcr/>
- [3] Alexander I. Pudovkin, Eugene Garfield: Rank-Normalized Impact Factor: A Way To Compare Journal Performance Across Subject Categories, Proceedings of the 67th Annual Meeting of ASIST, vol.41, p.507-515, 2004.  
<http://www.garfield.library.upenn.edu/papers/ranknormalizationasist2004published.pdf>
- [4] 根岸正光: 論文数・引用数からみたわが国の大学間格差の動向分析, 情報知識学会誌, Vol.19, No.2, p.158-169, 2009.
- [5] 根岸正光: ISI 論文引用度指数ランキング: 納税者の視点で資金効率を評価, 大学ランキング 2010 年版, 朝日新聞社, p.222-231, 2009. ISBN: 978-4-02-274532-3

# 学術の国際化による日本の産学共著関係の変化

## Change in university-industry co-authorship relations in Japan with globalization of academic research

孫 媛\*, 根岸 正光  
Yuan SUN, Masamitsu NEGISHI

国立情報学研究所 情報社会相関研究系

Information and Society Research Division, National Institute of Informatics

〒101-8430 千代田区一ツ橋2-1-2

Email: yuan @nii.ac.jp

筆者らは、日本の産学連携について学術論文の共著関係に基づく分析を行い、大学と企業の連関が弱まる一方、海外との連関が年々強まっていることを見出した。このことは、国内知識生産システムを説明するのに海外との連携を考慮する必要性を示唆する。そこで、本研究では、海外を南北アメリカ、アジア、欧州の三地域に分け、日本の学術の国際研究ネットワークの形成および産学連携の時系列変化に関する実態をより詳細に分析する。

Our results based on analyses of co-authorship relations of academic publications between university and industry collaborations have shown that relations between university and industry are getting weaker, while collaborations with overseas are becoming stronger in Japan. This suggests that without taking account of collaborations with overseas it is no even more possible to explain national knowledge producing systems. In this study, we break down foreign sector into three regions of Americans, Asian areas and EU, to make further investigations on the formation of international research networks and time series changes of university-industry collaborations in Japan.

キーワード: 産学連携, 共著関係, 国際研究ネットワーク

Keyword: university industry collaboration, co-authorship relations, international research network

### 1. はじめに

これまで、筆者らは、日本の産学連携について、学術論文の共著関係に基づく多次元の相

互情報量やグラフィカルモデリングなどの手法による分析を行ってきた[1][2][3]。その結果、大学と企業の連関が弱まり、海外との連関が年々強まっていることが見出され、国内知識生産シ

システムを有効に説明するためには、海外との連携を考慮に入れる必要があることが示唆された。そこで、本研究では、海外を南北アメリカ、アジア、欧州の三地域に分け、日本における学術の国際研究ネットワークの形成および産学連携の時系列変化について、実態分析を行う。

また近年、単著論文に比べ、共著論文が増えていると言われるが、論文の共著者数の変化、共著者数と論文の引用度の関係などについて、日本においては、実際には検証されていない。本研究では、これについても統計的事実を示し、その国際化との関連に言及したい。

## 2. データと分析

ここで用いるデータベースは、National Citation Report for Japan (NCR-J)1997-2006 である。NCR-J は、米国 Thomson Reuters 社(通称 ISI)製作の Science Citation Index (SCI), Social Science Citation Index (SSCI), Arts and Humanities Citation Index (A&HCI)という3種類の引用索引データベースに収録された論文から、1997 年から 2006 年に発表された日本人著者を含む論文を抽出して、書誌データと引用数を収録したものである。分野を生物・医学系、理工系、人文社会系と学際分野の4種類に区分集計すると、生物・医学系(353,400 件)と理工系(369,885 件)の論文が、それぞれ総論文(728,284 件)の 48.5%, 50.8%と大半を占めているため、ここでは、生物医学系と理工系のみを分析対象とする。また産学共著を分析するにあたり、著者所属機関を、大学、公共、病院(大学病院以外)、企業および海外(アメリカ、欧州、アジア、その他)の計7セクタと不明のいずれかに区分している。ここで公共とは、政府、自治体の研究所等の他、財団法人等公益法人を含む。なお、複数著者の論文では、全著者の所属機関別を付与している。

## 3. 結果

### 3.1 著者数と論文数の関係

まず4つの年を取り出して、各年の論文総数に占める著者数別論文の割合を示す(図1, 図2)。年によらず、生医系では、4~6人の共著論文が多く、10人以上もかなりの割合になる。また、近年になるほど著者数の多い方へシフトしていることがわかる。実際、著者数6人以上の論文の割合は、1997年の42%から、2006年には54%へと増えている。理工系は生医系と比べると著者数が少ない方に偏っており、論文数最多は共著者数3のときである。

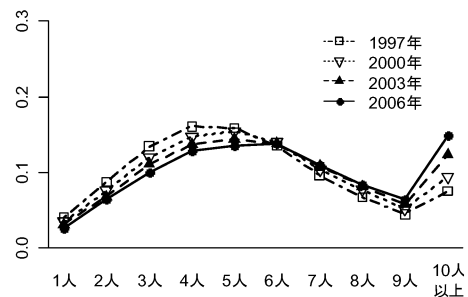


図1 著者数別の論文数割合(生医)

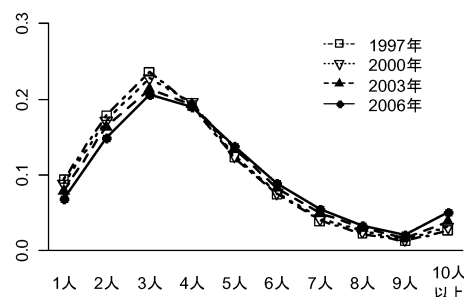


図2 著者数別の論文数割合(理工)

### 3.2 著者数と引用度の関係

図3, 図4は、1997年, 1998年に発表された著者数30人以下の論文について、著者数と引



引用度の関係を散布図で表現したものである。これらの図からは、著者が10人程度までのときには、著者数が増えるにつれて、引用度が直線的に上がる傾向を読みとることができる。とくに、生医系では、引用度の上がり方が顕著である。そこで、複数の著者の連携のあり方によって、

引用度の上がり方に違いがあるかどうかを検討するために、著者たちの所属を、「大学のみ」、「大学と企業」、「大学と海外」、「大学・企業・海外」の4群に分けて、群別に著者数(10人まで)と引用度の関係を見た(図5, 図6)。

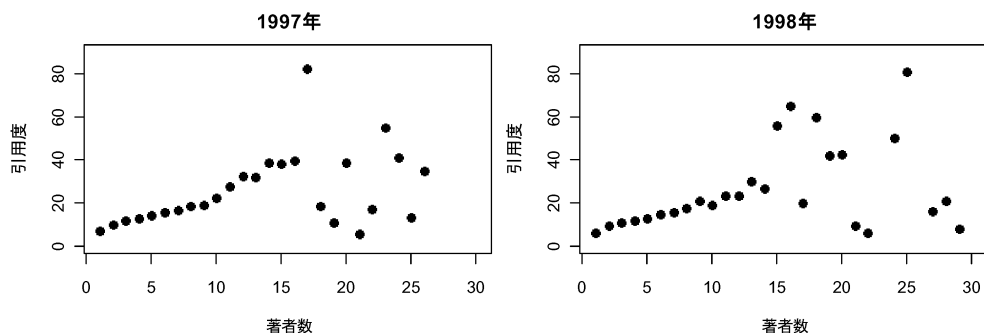


図3 著者数と引用度の関係(生医)

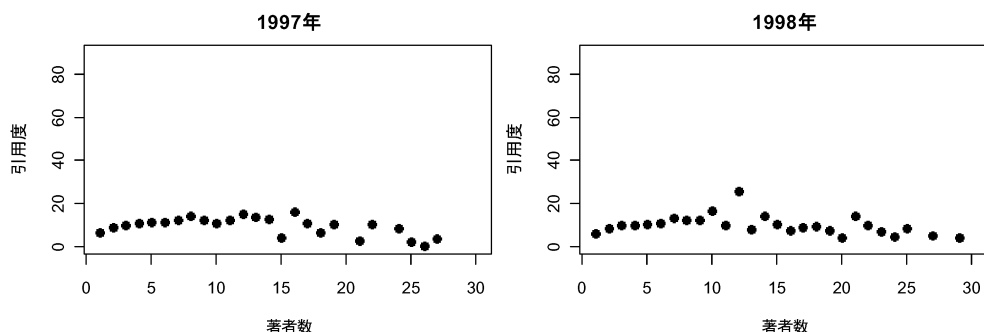


図4 著者数と引用度の関係(理工)

生医系では、すべての著者が大学に所属している場合には著者数が多くなっても引用度の上がり方はさ程大きくないが、海外著者を含む場合には、どの著者数においても一様に引用度が高く、しかも著者数が増えると引用度が大きく上がっている。これに対して、理工系では、大学・海外の著者を含む論文の引用度はやや高いが、生医系の場合と比べると、著者所属による大きな違いは見られないようである。

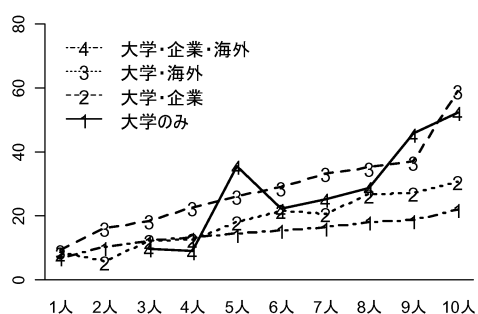


図5 機関種・著者数別の引用度 (1997年・生医)

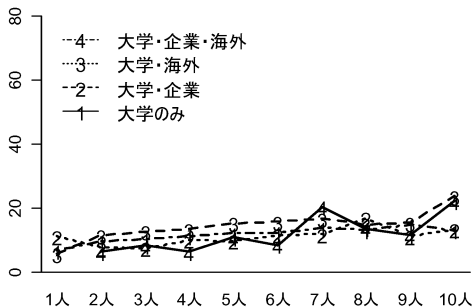


図6 機関種・著者数別の引用度  
(1997年・理工)

### 3.3 国際共著(海外との連携)と引用度

総的に海外機関に所属する著者を含む論文の引用度が高いことが示唆されたことから、著者の所属機関によって 1=国内のみ, 2=米, 3=亜, 4=欧, 5=米亜, 6=米欧, 7=亜欧, 8=米亜欧の群分けをして、群別に引用度を図示したのが図7, 図8である。これらの図では、論文発表からの経過年数を横軸にとってあるので、一番左が2006年、一番右が1997年の発表論文であることに注意されたい(海外著者を含む論文は必然的に共著論文なので、比較のために、この節以降の分析では、単著論文を除いてある)。生医系では、米, 欧, 米亜欧, 米欧のように、米と欧を共著者を含む論文の引用度は国内論文の2倍以上である。理工系では、米欧がそれほど際だつことはないが、米亜欧のすべてを著者を含む論文の引用度は非常に高い。

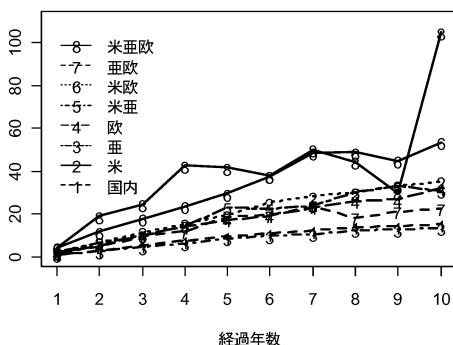


図7 海外との連携と引用度(生医)

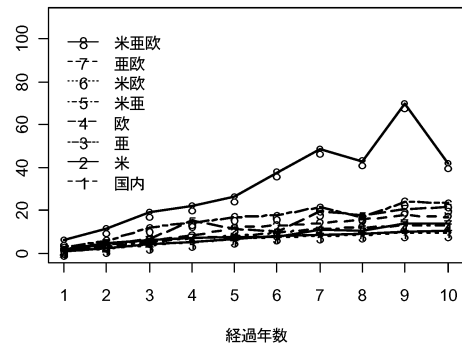


図8 海外との連携と引用度(理工)

### 3.4 国際共著と産学論文の変化

ここまでの分析により、海外とくに米欧の著者を含む論文の引用度が相対的に高いことが明らかになったが、最後に、国内機関と海外機関の共著関係、組合せの違いによる、論文数・引用度の経年変化に目を向ける。国内機関については大学、企業の著者を含むか否かに着目し、「大学(大学を含むが企業は含まない)」「企業(企業を含むが大学は含まない)」「大学・企業」「なし(大学も企業も含まない)」の4区分とし、共著海外機関については「国内(海外なし)」「米」「亜」「欧」「米亜」「米欧」「亜欧」「米亜欧」の8区分で考える。

まず、論文数について、1997年から2006年の各年の論文総数を4区分×8区分=32通りのカテゴリごとに集計した。表1(生医)、表2(理工)は、1997年と2006年について、各年の論文総数を100%として、各カテゴリに該当する論文の割合を示したものである(ここでは、2つの年だけを示したが、10年を通じて変化の方向は一貫している)。

生医系については、個別のカテゴリごとには際だった変動がないようにも見えるが、国内機関の著者のみによる論文は、85.2%から77.1%へと約8ポイント減り、米, 亜, 欧の著者を含む論文はすべて増加している。また、企業所属の著者による論文は、大学との共著の有無にかか

ならず割合を減らしている。図7で見たように、米、欧の著者を含む論文の引用度が高いが、そうした論文数の割合は、この10年で12.3%から16.8%に増えている。その一方で、亜も2.5%から6.1%へと増加している。

**表1 大学・企業と海外の共著論文の割合  
(生医, 上:1997年, 下:2006年)**

|     | 大学    | 企業   | 大学・<br>企業 | なし   | 計     |
|-----|-------|------|-----------|------|-------|
| 国内  | 68%   | 4%   | 6%        | 8%   | 82.2% |
| 米   | 5%    | 0%   | 0%        | 1%   | 6.9%  |
| 亜   | 2%    | 0%   | 0%        | 0%   | 2.5%  |
| 欧   | 3%    | 0%   | 0%        | 1%   | 3.7%  |
| 米亜  | 0%    | 0%   | 0%        | 0%   | 0.3%  |
| 米欧  | 1%    | 0%   | 0%        | 0%   | 1.1%  |
| 亜欧  | 0%    | 0%   | 0%        | 0%   | 0.2%  |
| 米亜欧 | 0%    | 0%   | 0%        | 0%   | 0.1%  |
| 計   | 79.8% | 4.4% | 6.5%      | 9.3% | 100%  |

|     | 大学    | 企業   | 大学・<br>企業 | なし   | 計     |
|-----|-------|------|-----------|------|-------|
| 国内  | 63%   | 2%   | 5%        | 7%   | 77.1% |
| 米   | 7%    | 0%   | 0%        | 1%   | 8.8%  |
| 亜   | 5%    | 0%   | 0%        | 1%   | 6.1%  |
| 欧   | 4%    | 0%   | 0%        | 1%   | 4.6%  |
| 米亜  | 1%    | 0%   | 0%        | 0%   | 1.0%  |
| 米欧  | 1%    | 0%   | 0%        | 0%   | 1.6%  |
| 亜欧  | 0%    | 0%   | 0%        | 0%   | 0.5%  |
| 米亜欧 | 0%    | 0%   | 0%        | 0%   | 0.4%  |
| 計   | 82.1% | 2.6% | 5.8%      | 9.6% | 100%  |

理工系についても、海外著者を含まない論文の割合は、82.7%から73.4%へと約9ポイント減少している。米との共著の割合はあまり変わらないが、アジアの著者を含む論文は、アジア以外の海外を含まない論文が3.7%から9.4%へとかなり増えているほか、米亜、亜欧、米亜欧の共著を含む論文がいずれも割合を増やしている。

引用度については、新しい論文ほど当然のこととして引用数が少なく引用度が低くなりがちであり、そのままでは経年比較が難しい。そこで、各年の海外著者を含まない論文全体の平均引用度を1.00として、これとの比較で各カテゴリの引用度を表現した。また、結果が煩雑になるのを避けるため、海外を細かく区分せずに、海外

著者の有無だけで分類、整理をした(表3, 表4)。

**表2 大学・企業と海外の共著論文の割合  
(理工, 上:1997年, 下:2006年)**

|     | 大学    | 企業    | 大学・<br>企業 | なし   | 計     |
|-----|-------|-------|-----------|------|-------|
| 国内  | 58%   | 12%   | 7%        | 5%   | 82.7% |
| 米   | 3%    | 1%    | 0%        | 1%   | 5.1%  |
| 亜   | 3%    | 0%    | 0%        | 0%   | 3.7%  |
| 欧   | 5%    | 0%    | 0%        | 1%   | 6.4%  |
| 米亜  | 0%    | 0%    | 0%        | 0%   | 0.3%  |
| 米欧  | 1%    | 0%    | 0%        | 0%   | 1.3%  |
| 亜欧  | 0%    | 0%    | 0%        | 0%   | 0.3%  |
| 米亜欧 | 0%    | 0%    | 0%        | 0%   | 0.3%  |
| 計   | 71.1% | 13.6% | 8.1%      | 7.3% | 100%  |

|     | 大学    | 企業   | 大学・<br>企業 | なし   | 計     |
|-----|-------|------|-----------|------|-------|
| 国内  | 55%   | 5%   | 8%        | 5%   | 73.4% |
| 米   | 4%    | 0%   | 0%        | 1%   | 5.2%  |
| 亜   | 7%    | 0%   | 0%        | 1%   | 9.4%  |
| 欧   | 6%    | 0%   | 0%        | 1%   | 7.6%  |
| 米亜  | 1%    | 0%   | 0%        | 0%   | 0.8%  |
| 米欧  | 2%    | 0%   | 0%        | 0%   | 2.0%  |
| 亜欧  | 1%    | 0%   | 0%        | 0%   | 0.8%  |
| 米亜欧 | 1%    | 0%   | 0%        | 0%   | 0.7%  |
| 計   | 75.1% | 6.6% | 9.5%      | 8.8% | 100%  |

生医系では、国内著者のみの論文と比べて、海外著者を含む論文の引用度が概ね2倍程度高いこと、大学の著者を含む論文では、企業の著者との共著であるときの引用度が高いことなどがわかる。

**表3 大学・企業と海外の共著論文の引用度  
(生医)**

|      | 海外の著者なし |      |           |      | 海外の著者あり |      |           |      |
|------|---------|------|-----------|------|---------|------|-----------|------|
|      | 大学      | 企業   | 大学・<br>企業 | なし   | 大学      | 企業   | 大学・<br>企業 | なし   |
| 1997 | 0.98    | 0.95 | 1.52      | 0.78 | 2.08    | 2.28 | 3.35      | 2.07 |
| 1998 | 0.98    | 1.12 | 1.47      | 0.81 | 1.93    | 2.94 | 2.51      | 2.29 |
| 1999 | 0.99    | 1.06 | 1.47      | 0.74 | 1.96    | 1.87 | 2.87      | 1.81 |
| 2000 | 0.99    | 0.89 | 1.44      | 0.83 | 2.02    | 2.84 | 2.45      | 1.95 |
| 2001 | 1.00    | 0.85 | 1.29      | 0.80 | 2.01    | 2.43 | 2.12      | 1.84 |
| 2002 | 1.00    | 0.93 | 1.35      | 0.75 | 1.91    | 1.77 | 2.24      | 1.98 |
| 2003 | 0.98    | 1.05 | 1.22      | 0.97 | 1.90    | 1.62 | 2.11      | 1.88 |
| 2004 | 1.00    | 0.91 | 1.20      | 0.84 | 1.89    | 1.73 | 2.09      | 1.88 |
| 2005 | 1.00    | 0.97 | 1.18      | 0.86 | 1.98    | 1.96 | 1.89      | 2.51 |
| 2006 | 1.01    | 1.06 | 1.14      | 0.78 | 1.98    | 1.88 | 1.98      | 2.11 |

理工系についても海外著者を含む論文の引用度が高いが、海外の著者を含まない論文

との違いは、生医系の場合ほどは大きくない。また、理工系では著者が企業に所属する論文の引用度が低く、とくに海外の著者を含まない場合には大学との共著でも、大学単独の論文よりも引用度が低くなっている。

**表4 大学・企業と海外の共著論文の引用度  
(理工)**

|      | 海外の著者なし |      |       |      | 海外の著者あり |      |       |      |
|------|---------|------|-------|------|---------|------|-------|------|
|      | 大学      | 企業   | 大学・企業 | なし   | 大学      | 企業   | 大学・企業 | なし   |
| 1997 | 1.04    | 0.84 | 0.95  | 0.97 | 1.45    | 1.15 | 1.09  | 1.38 |
| 1998 | 1.07    | 0.77 | 0.84  | 0.96 | 1.61    | 1.18 | 1.61  | 1.47 |
| 1999 | 1.06    | 0.76 | 0.87  | 1.02 | 1.46    | 1.35 | 1.35  | 1.38 |
| 2000 | 1.05    | 0.78 | 0.87  | 1.03 | 1.54    | 1.31 | 1.95  | 1.15 |
| 2001 | 1.03    | 0.85 | 0.83  | 1.13 | 1.45    | 1.13 | 1.06  | 1.19 |
| 2002 | 1.03    | 0.77 | 0.91  | 1.09 | 1.47    | 1.49 | 1.61  | 1.47 |
| 2003 | 1.02    | 0.79 | 0.88  | 1.17 | 1.46    | 1.54 | 1.54  | 1.28 |
| 2004 | 1.03    | 0.79 | 0.86  | 1.13 | 1.45    | 1.31 | 1.35  | 1.49 |
| 2005 | 1.03    | 0.73 | 0.89  | 1.11 | 1.44    | 1.30 | 1.07  | 1.46 |
| 2006 | 1.05    | 0.76 | 0.78  | 1.06 | 1.62    | 1.14 | 1.43  | 1.57 |

#### 4. まとめ—研究自給率の動向

わが国の科学技術・学術研究の国際化の必要性がいわれ、政策的支援も行われて久しいものがある。確かに学術研究に国境はないとはいうが、「科学技術創造立国」を科学技術政策の基本として掲げるとき、それは国際競争に打ち勝てるような研究の振興を国家的目標に据えたものであろう。してみると、本稿にみるような国際共著論文の増加をとらえて、わが国の学術研究の国際化が進展したことを一概に喜ぶわけにはいかないと考えられる。それは学術研究における自給率の低下、対外依存度の上昇を意味しているからである。

海外機関との共同研究は、その成果たる国際共著論文の露出度が高く、従って引用度も高い。そこで引用度に着目した業績評価が行われれば、研究者達は、手っ取り早く引用度を稼得する方策として、いきおい国際共同研究に傾くのはやむを得まい。しかし、これによってもたらされるであろう研究の自給率低下、対外依存度の上昇を、科学技術創造立国との関連に

おいてどのように評価するべきかは、深刻な問題である。

この際、国際共著論文の増加は、いたずらに引用度に注目した研究業績評価の「弊害」ともみなされ得るという視点も重要になってくる。このような点も踏まえ、自給率、対外依存率に注目した分析手法について、さらに検討を進めたいと考えているところである。

#### 参考文献

- [1] Yuan Sun, Masamitsu Negishi and Masaki Nisizawa “Coauthorship Linkages between Universities and Industry in Japan”, Vol.16, No.4, pp. 299-309, Research Evaluation, December. 2007
- [2] Yuan Sun, Masamitsu Negishi and Loet Leydesdorff “National and International Dimensions of the Triple Helix in Japan: University-Industry-Government and International Co-Authorship Relations”, the 11th International Conference of the International Society for Scientometrics and Informetrics (ISSI2007), pp.936-937, Madrid, Spain, 25-27 June 2007.
- [3] Yuan Sun and Masamitsu Negishi “Measuring the relationships among university, industry and the other sectors in Japan’s national innovation system”, Scientometrics, Vol.82. No.3, pp.677-685, 2010

## キーワード分析による環境関連研究の動向調査 Investigation into Environmental Studies in JAPAN by Key-Word Analysis

西澤正己\*, 孫媛  
Masaki NISHIZAWA\*, Yuan SUN

国立情報学研究所

National Institute of Informatics

〒101-8430 東京都千代田区一ツ橋2-1-2

E-mail: [nisizawa@nii.ac.jp](mailto:nisizawa@nii.ac.jp), [yuan@nii.ac.jp](mailto:yuan@nii.ac.jp)

我々はこれまでにキーワード分析により、ゲノム関連、ナノテクノロジー関連、情報関連等、科学技術基本計画の重点領域研究分野の動向とその広がりについて調査してきた。ここでは環境分野関連の研究の動向を国立環境研究所のおよそ20年にわたる研究課題と科学研究費補助金の研究課題等を比較し、キーワード分析により他の研究分野との関連や動向について調査した結果を示した。暫定結果ではあるが、用いた方法で国立環境研究所の分類分野の相互関係や、科学研究費補助金の細目分野との位置関係がよく示されることがわかった。

We have investigated genome relation, nanotechnology relation, information relation, etc. by keyword analysis about the trend and spread of the important research area of Science and Technology Basic Plan. Here, we compared the research task over about 20 years of National Institute for Environmental Studies for the trend of environmental field-related research with the research task of grants-in-aid for scientific research, and showed the result investigated about other relation and trends with an area of research by keyword analysis. Although it was an interim result, it turned out that the correlations of the classification field of National Institute for Environmental Studies and spatial relationship with the details field of grants-in-aid for scientific research are well shown by the used method.

キーワード: 科学技術基本計画, 環境研究, キーワード分析, 関連分析, 研究動向

Keyword: Science and Technology Basic Plan, environmental field-related research, keyword analysis, relationship, research trend

## 1 はじめに

政府および総合科学技術会議は科学技術基本法のもと、1996年から第1期、第2期科学技術基本計画[1]を進めてきた。第2期科学技術基本計画(平成13~17年度)では重点分野として「ライフサイエンス分野」、「情報通信分野」、「環境分野」、「ナノテクノロジー・材料分野」が特に重点を置かれ、資源配分が重点的に行われてきた。これらの分野は第3期計画(平成18~22年度)においても、引き続き重点分野とされている。

文部科学省科学研究費補助金(科研費)[2]は大学等の研究者に対しての個人申請・ピアレビュー審査による研究費配分制度である。この制度は1939年に始まり、古くから基礎研究の下支え、底上げの役目を果たしてきたと言われている。現在では、科研費以外にも多くの競争的資金が配分されているものの、研究費総額では、科研費が約半額を占めている。この科研費において、科学技術基本計画の重点分野に対してどのように研究費が配分され、どのように変化してきたかは興味を持たれるところである。

我々は、科学研究費補助金データベース(KAKEN, KAKENK)[3]からこれらの重点領域に対するキーワード(トピック関連キーワード)[4][5]を抽出したうえで、科研費のすべての採択課題の課題名からこれら重点領域(特に、ライフサイエンス分野のゲノム関連分野とナノテクノロジー・材料分野の2分野)に関連する課題を選び出し、配分額の変遷を調べた。また、科研費の分科細目表に対しても同様に特徴的なキーワードを抽出し、それらの一致度を見ることにより、重点領域との関連分野を対応分

析により示した[6]。

本論文ではこれまでとはやや異なるキーワード抽出方法や関連度の計算方法を用い、環境関連分野を特に国立環境研究所の研究成果データを用いて科学研究費補助金のデータとの比較を行った結果を述べる。

## 2 データ

### 2.1 国立環境研究所の研究成果データ

今回は環境関連の研究課題として、最も網羅的に環境関連の研究をおこなっている国立環境研究所の研究課題を基準とすることにする。

国立環境研究所(NIES)ではホームページ上に研究所の研究領域別、年度別に発表文献リストを公開している[7]。本調査ではこの文献リスト中の日本語課題名について、キーワード分析の対象とした。また、文献リストは1987年から2009年までであるが、この間に組織替え等で研究部門、プロジェクト等が変遷している。今回の調査ではNIESデータは年度別には取り扱わず、表1のように17の分析用研究分野(プロジェクト)分類に集約して調査をおこなった。

17の分類はAerosphere(大気圏)、AreaEnv(地域環境)、BioEnv(生物環境)、ChemEnv(化学環境)、EarthEnv(地球環境)、Env-risk(環境リスク)、HealthEnv(健康環境)、Infra-Info(技術・情報)、Recycle(循環型社会)、SocEnv(社会環境)、WaterClod(水土圏)、の11の研究領域とBasinZone(流域圏環境管理プロジェクト)、Bio-diversity(生物多様性プロジェクト)、dyoxin-proj(環境ホルモン・ダイオキシンプロジェクト)、ozone(成層圏オゾン変動プロジェクト)、PM-DEP(PM2.5-DEPプロ

表 1 国立環境研究所の文献データのグループ分けと調査課題数

| 分析用分類             | NIES-研究部門           | 期間        | 総論文数 | 和文論文数 |
|-------------------|---------------------|-----------|------|-------|
| Aerosphere(AS)    | 大気環境部               | 1988-1989 | 117  | 73    |
|                   | 大気圏環境部              | 1990-2000 | 400  | 168   |
|                   | 大気圏環境研究領域           | 2001-2009 | 435  | 141   |
| AreaEnv(AE)       | 地域環境研究グループ          | 1990-2000 | 1429 | 876   |
|                   | アジア自然共生研究グループ       | 2006-2009 | 251  | 100   |
| BasinZone(BZ)     | 流域圏環境管理プロジェクト       | 2001-2005 | 84   | 34    |
| Bio-diversity(BD) | 生物多様性プロジェクト         | 2001-2005 | 163  | 82    |
| BioEnv(BE)        | 生物環境部               | 1988-1989 | 93   | 39    |
|                   | 生物圏環境部              | 1990-2000 | 395  | 175   |
|                   | 生物圏環境研究領域           | 2001-2009 | 224  | 66    |
| ChemEnv(CE)       | 化学環境部               | 1990-2000 | 457  | 204   |
|                   | 化学環境研究領域            | 2001-2009 | 389  | 167   |
| EarthEnv(EE)      | 地球環境研究センター          | 1990-2009 | 715  | 461   |
|                   | 地球環境研究グループ          | 1990-2000 | 916  | 568   |
| Env-risk(ER)      | 化学物質環境リスク研究センター     | 2001-2005 | 80   | 50    |
|                   | 環境リスク研究センター         | 2006-2009 | 377  | 166   |
| HelthEnv(HE)      | 環境生理部               | 1988-1989 | 36   | 17    |
|                   | 環境保健部               | 1988-1989 | 94   | 55    |
|                   | 環境健康部               | 1990-2000 | 483  | 236   |
|                   | 環境健康研究領域            | 2001-2009 | 265  | 64    |
| Infra-Info(II)    | 環境情報部               | 1988-1989 | 45   | 33    |
|                   | 技術部                 | 1988-1989 | 56   | 50    |
|                   | 環境情報センター            | 1992-1993 | 2    | 2     |
| PM-DEP(PM)        | 環境研究基盤技術ラボラトリー      | 2001-2009 | 91   | 33    |
|                   | PM2.5-DEPプロジェクト     | 2001-2005 | 137  | 63    |
| Recycle(Rc)       | 廃棄物研究部              | 1997-2000 | 8    | 7     |
|                   | 循環型社会形成推進・廃棄物研究センター | 2001-2005 | 297  | 190   |
|                   | 循環型社会・廃棄物研究センター     | 2006-2009 | 483  | 280   |
| SocEnv(SE)        | 社会環境システム部           | 1990-2000 | 528  | 399   |
|                   | 社会環境システム研究領域        | 2001-2009 | 582  | 375   |
| WarmerClimate(Wm) | 地球温暖化プロジェクト         | 2001-2005 | 52   | 19    |
| WaterClod(WC)     | 水質土壌環境部             | 1988-1989 | 136  | 87    |
|                   | 水士圏環境部              | 1990-2000 | 397  | 298   |
|                   | 水士圏環境研究領域           | 2001-2009 | 193  | 88    |
| dyoxin-proq(Do)   | 環境ホルモン・ダイオキシンプロジェクト | 2001-2005 | 199  | 83    |
| ozone(Oz)         | 成層圏オゾン層変動プロジェクト     | 2001-2005 | 86   | 44    |

プロジェクト)、WarmerClimate (地球温暖化プロジェクト) の6つの研究プロジェクトである。プロジェクトは領域が絞り込まれているので、11の研究領域とはやや分類が異なると考えるべきである。

## 2.2 科学研究費補助金データ

国立情報学研究所では、科学研究費補助金の採択課題の速報を収録した「科学研究費補助金採択課題データベース (KAKENK)」[3]およびその成果概要を収録したデータベースである「科学研究費補助金成果概要データベース (KAKEN)」[3]が作成されている。KAKENK においては1996年度から最新の2009年度までの採択課題 (現在では遡及入力が進み、一

部の項目を除いては1965年からのデータが収められている)、KAKEN においては1985年度から2007年度までの成果概要が収録されている。KAKENK データベースには採択課題名、研究種目、分科細目 (大型研究費、奨励研究を除く)、新規・継続の別、配分額等が収録されており、KAKEN についてはKAKENK の情報に加え、成果概要、キーワード、発表文献リスト等が収録される。なお、KAKEN データベースは成果概要の報告書を収録するため、採択後少なくとも約1年半後の課題が収録されていることになる。今回の分析には1976 - 2009年度のKAKENK データベースを使用した、今回の報告では連続的な経年変化の分析が間に合っていないので、

表 2 科学研究費補助金の環境関連細目とその課題数

| 年度   | 細目コード | 細目名             | 課題数 |
|------|-------|-----------------|-----|
| 1999 | 801   | 環境動態解析          | 129 |
|      | 802   | 環境影響評価(含放射線生物学) | 127 |
|      | 803   | 環境保全            | 164 |
|      | 436   | 土木環境システム        | 112 |
|      | 442   | 建築環境・設備         | 111 |
|      | 553   | 生物環境            | 45  |
|      | 685   | 環境系薬学           | 63  |
| 2004 | 2001  | 環境動態解析          | 229 |
|      | 2002  | 環境影響評価・環境政策     | 150 |
|      | 2003  | 放射線・化学物質影響科学    | 179 |
|      | 2004  | 環境技術・環境材料       | 190 |
| 2009 | 2001  | 環境動態解析          | 259 |
|      | 2002  | 環境影響評価・環境政策     | 238 |
|      | 2003  | 放射線・化学物質影響科学    | 182 |
|      | 2004  | 環境技術・環境材料       | 195 |

分科細目表の異なる 1999 年、2004 年、2009 年の結果を掲載するに留める。

なお、2003 年には分科細目表の大改訂、2007 年には小改訂が行われており、2003 年を境に分類は大きく異なる。1999 年時点での特に環境研究を中心とした細目は 1999 年では 801 (環境動態解析)、802 (環境影響評価(含放射線生物学))、803 (環境保全)、436 (土木環境システム)、442 (建築環境・設備)、553 (生物環境)、685 (環境系薬学)などが挙げられ、2004 年、2009 年度では小変更であるので、関連分野は変更がなく細目コード、細目名は 2001 (環境動態解析)、2002 (環境影響評価・環境政策)、2003 (放射線・化学物質

影響科学)、2004 (環境技術・環境材料)が挙げられる。それぞれの課題数はまとめて表 2 に示す。

## 3 分析手法および結果

### 3.1 キーワード分析

キーワード分析の手法としては、これまで[3][4][6]とはやや異なっている。課題名から形態素解析システム茶筌[8]を用いるのは同じであるが、キーワードを抽出する場合に、以前のようにできるだけ長い複合語にするのではなく、完全一致をより多く検出するために、できるだけ短い複合語になるようにキーワード抽出をおこなった。これは、今後キーワードの組み合わせによる関連課題抽出のためにも行った変更である。

表 3 および表 4 に、国立環境研究所の課題から抽出したキーワードおよび科学研究費補助金採択課題データベースから抽出したキーワードを分野別に上位 20 件ずつ示す。紙面の都合上、表 3 は研究プロジェクトを除く 11 分野について、表 4 については表 2 の 2009 年度と 1999 年度について示した。今回は以前のように広い意味を持つキーワードの削除は行っていないので

表 3 国立環境研究所の分析用研究分野の研究課題に対する上位 20 位までのキーワード

| 順位    | キーワード | 真なり数 | キーワード | 真なり数 | キーワード    | 真なり数 | キーワード | 真なり数 | キーワード | 真なり数 | キーワード | 真なり数 | キーワード  | 真なり数 | キーワード | 真なり数 | キーワード | 真なり数 | キーワード | 真なり数 | キーワード    | 真なり数 | キーワード | 真なり数 |
|-------|-------|------|-------|------|----------|------|-------|------|-------|------|-------|------|--------|------|-------|------|-------|------|-------|------|----------|------|-------|------|
| 1     | 大気    | 111  | 環境    | 224  | 植物       | 47   | 環境    | 144  | 環境    | 228  | 物質    | 50   | 環境     | 57   | 異形    | 10   | 生態    | 112  | 環境    | 229  | 国際       | 5    | 環境    | 83   |
| 2     | 環境    | 59   | 環境    | 120  | 環境       | 40   | 物質    | 132  | 環境    | 218  | 環境    | 48   | 環境     | 54   | 環境    | 8    | 環境    | 108  | 環境    | 219  | 環境       | 4    | 水     | 52   |
| 3     | 環境    | 38   | 汚染    | 102  | 物質       | 31   | 化学    | 92   | 環境    | 221  | 化学    | 44   | 汚染     | 48   | 汚染    | 6    | 環境    | 69   | 環境    | 85   | シミュレーション | 4    | 土壌    | 42   |
| 4     | 環境    | 26   | 水     | 86   | 研究       | 28   | 汚染    | 71   | 研究    | 115  | 評価    | 35   | 環境     | 46   | 生物    | 6    | 技術    | 64   | 評価    | 73   | 役割       | 3    | 研究    | 40   |
| 5     | 汚染    | 36   | 物質    | 80   | 計画       | 25   | 研究    | 89   | 環境    | 83   | 物質    | 33   | 物質     | 42   | 環境    | 6    | 評価    | 45   | 環境    | 89   | 環境       | 3    | 特性    | 37   |
| 6     | レーザ   | 36   | 環境    | 71   | 汚染       | 21   | 分析    | 58   | 環境    | 82   | リスク   | 25   | 大気     | 39   | 水     | 5    | 物質    | 42   | 研究    | 64   | 倉庫       | 3    | 環境    | 37   |
| 7     | 地球    | 25   | 技術    | 72   | シミュレーション | 20   | 物質    | 54   | 大気    | 68   | ナノ    | 16   | 研究     | 35   | 研究    | 5    | 水     | 36   | 研究    | 56   | 交換       | 3    | 微生物   | 36   |
| 8     | 研究    | 34   | 研究    | 52   | 動物       | 19   | 環境    | 52   | 環境    | 67   | 初期    | 15   | 環境     | 24   | データ   | 5    | 物質    | 33   | 分析    | 49   | 環境       | 3    | 浄化    | 35   |
| 9     | オゾン層  | 24   | 大気    | 98   | 生物系      | 19   | 評価    | 40   | 環境    | 56   | 金属    | 15   | 評価     | 32   | 社会    | 4    | 分析    | 32   | 社会    | 44   | 吸収源      | 3    | 地下    | 31   |
| 10    | アクリル  | 21   | 環境    | 51   | 化学物質     | 17   | 化学物質  | 41   | 物質    | 55   | 粒子    | 13   | 物質     | 19   | 生物    | 1    | 環境    | 31   | 環境    | 46   | 環境       | 2    | 汚染    | 50   |
| 11    | 環境    | 25   | 環境    | 46   | 環境       | 16   | リスク   | 35   | オゾン層  | 52   | 環境    | 13   | メタボリック | 19   | 環境    | 4    | 物質    | 30   | 生物    | 46   | 分析       | 2    | 水質    | 29   |
| 12    | ライダー  | 23   | 評価    | 48   | 物質       | 15   | 分析    | 29   | 評価    | 48   | 生産量   | 13   | 物質     | 17   | 環境    | 4    | 環境    | 29   | 日本    | 36   | 物質       | 2    | 分析    | 25   |
| 13    | レーダー  | 22   | 対照    | 48   | 大気       | 19   | 生物    | 28   | 社会    | 47   | 汚染    | 13   | ディーゼル  | 17   | 影響    | 4    | 環境    | 27   | 環境    | 37   | 環境       | 2    | 物質    | 25   |
| 14    | エアロゾル | 21   | 環境    | 46   | 環境       | 19   | 水     | 27   | 環境    | 42   | 環境    | 12   | リスク    | 16   | 環境    | 3    | 物質    | 27   | 環境    | 34   | 物質       | 2    | 物質    | 23   |
| 15    | 測定    | 20   | 環境    | 49   | 環境       | 15   | 環境    | 29   | 環境    | 42   | 研究    | 11   | 環境     | 19   | 環境    | 2    | 物質    | 28   | 環境    | 39   | 環境       | 2    | 環境    | 23   |
| 16    | 化学    | 20   | リスク   | 47   | 診断       | 14   | 環境    | 24   | 汚染    | 42   | 環境    | 10   | 物質     | 15   | 環境    | 3    | 環境    | 25   | 環境    | 31   | 手法       | 2    | 環境    | 23   |
| 17    | 反応    | 19   | 環境    | 35   | 環境       | 13   | 内分    | 24   | 物質    | 42   | 環境    | 10   | アレルギー  | 15   | 環境    | 3    | 食品    | 24   | 環境    | 29   | 環境       | 2    | 汚染    | 23   |
| 18    | 環境    | 19   | 環境    | 35   | 環境       | 13   | 環境    | 24   | 環境    | 40   | 物質    | 9    | 粒子     | 13   | 生物    | 3    | 環境    | 24   | 環境    | 28   | 環境       | 2    | 物質    | 22   |
| 19    | 環境    | 18   | 環境    | 35   | 環境       | 13   | 環境    | 23   | 環境    | 39   | 環境    | 9    | 環境     | 13   | 環境    | 3    | 環境    | 23   | 環境    | 27   | 環境       | 2    | 環境    | 20   |
| 20    | 環境    | 18   | 化学    | 35   | 環境       | 10   | 環境    | 23   | 分析    | 39   | 環境    | 9    | 環境     | 13   | 環境    | 3    | 環境    | 23   | 環境    | 27   | 環境       | 2    | 環境    | 20   |
| 総真なり数 | 923   | 1847 | 1847  | 719  | 1382     | 1382 | 1829  | 601  | 1187  | 780  | 232   | 1187 | 1581   | 1581 | 1581  | 1581 | 1581  | 1581 | 1581  | 1581 | 1581     | 1581 | 1581  | 1581 |
| 総真なり数 | 2059  | 6016 | 1983  | 4141 | 8505     | 1446 | 2240  | 412  | 2688  | 5046 | 124   | 2003 |        |      |       |      |       |      |       |      |          |      |       |      |



表 4 科学研究費補助金の環境関連細目の研究課題に対する上位 20 位までのキーワード

| 研究項目  | 2009  |      |       |      |       |     |       |     |       |     | 1996  |     |       |     |       |     |       |     |       |     |
|-------|-------|------|-------|------|-------|-----|-------|-----|-------|-----|-------|-----|-------|-----|-------|-----|-------|-----|-------|-----|
|       | キーワード | 頻り数  | キーワード | 頻り数  | キーワード | 頻り数 | キーワード | 頻り数 | キーワード | 頻り数 | キーワード | 頻り数 | キーワード | 頻り数 | キーワード | 頻り数 | キーワード | 頻り数 | キーワード | 頻り数 |
| 1 環境  | 環境    | 60   | 環境    | 53   | 環境    | 44  | 環境    | 38  | 環境    | 35  | 環境    | 33  | 環境    | 31  | 環境    | 29  | 環境    | 27  | 環境    | 26  |
| 2 研究  | 研究    | 38   | 研究    | 34   | 研究    | 30  | 研究    | 25  | 研究    | 22  | 研究    | 19  | 研究    | 18  | 研究    | 16  | 研究    | 15  | 研究    | 14  |
| 3 政策  | 政策    | 32   | 政策    | 28   | 政策    | 24  | 政策    | 20  | 政策    | 18  | 政策    | 16  | 政策    | 14  | 政策    | 12  | 政策    | 11  | 政策    | 10  |
| 4 影響  | 影響    | 26   | 影響    | 22   | 影響    | 18  | 影響    | 15  | 影響    | 13  | 影響    | 11  | 影響    | 10  | 影響    | 9   | 影響    | 8   | 影響    | 7   |
| 5 評価  | 評価    | 24   | 評価    | 20   | 評価    | 16  | 評価    | 13  | 評価    | 11  | 評価    | 10  | 評価    | 9   | 評価    | 8   | 評価    | 7   | 評価    | 6   |
| 6 対策  | 対策    | 22   | 対策    | 18   | 対策    | 14  | 対策    | 11  | 対策    | 10  | 対策    | 9   | 対策    | 8   | 対策    | 7   | 対策    | 6   | 対策    | 5   |
| 7 安全  | 安全    | 20   | 安全    | 16   | 安全    | 12  | 安全    | 10  | 安全    | 9   | 安全    | 8   | 安全    | 7   | 安全    | 6   | 安全    | 5   | 安全    | 4   |
| 8 汚染  | 汚染    | 18   | 汚染    | 14   | 汚染    | 10  | 汚染    | 8   | 汚染    | 7   | 汚染    | 6   | 汚染    | 5   | 汚染    | 4   | 汚染    | 3   | 汚染    | 2   |
| 9 対策  | 対策    | 16   | 対策    | 12   | 対策    | 8   | 対策    | 6   | 対策    | 5   | 対策    | 4   | 対策    | 3   | 対策    | 2   | 対策    | 1   | 対策    | 0   |
| 10 対策 | 対策    | 14   | 対策    | 10   | 対策    | 6   | 対策    | 4   | 対策    | 3   | 対策    | 2   | 対策    | 1   | 対策    | 0   | 対策    | 0   | 対策    | 0   |
| 11 対策 | 対策    | 12   | 対策    | 8    | 対策    | 4   | 対策    | 2   | 対策    | 1   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   |
| 12 対策 | 対策    | 10   | 対策    | 6    | 対策    | 3   | 対策    | 1   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   |
| 13 対策 | 対策    | 8    | 対策    | 4    | 対策    | 2   | 対策    | 1   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   |
| 14 対策 | 対策    | 6    | 対策    | 3    | 対策    | 1   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   |
| 15 対策 | 対策    | 4    | 対策    | 2    | 対策    | 1   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   |
| 16 対策 | 対策    | 2    | 対策    | 1    | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   |
| 17 対策 | 対策    | 1    | 対策    | 0    | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   |
| 18 対策 | 対策    | 0    | 対策    | 0    | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   |
| 19 対策 | 対策    | 0    | 対策    | 0    | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   |
| 20 対策 | 対策    | 0    | 対策    | 0    | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   | 対策    | 0   |
| 総計    | 総計    | 953  | 総計    | 718  | 総計    | 543 | 総計    | 421 | 総計    | 316 | 総計    | 256 | 総計    | 203 | 総計    | 156 | 総計    | 121 | 総計    | 94  |
| 総延べ数  | 2276  | 2085 | 1697  | 1355 | 1030  | 819 | 643   | 516 | 417   | 337 | 276   | 223 | 181   | 146 | 116   | 94  | 76    | 61  | 50    | 40  |

「研究」、「開発」、「分析」等の広く使われるキーワードも含まれている。今後行う予定にはしているが、現在の状態でも分野間の比較に関してはさほど問題がないように思われる。

表 3、および表 4 の例のようにキーワードを分野別、KAKENKI に関しては分野別・年度別に抽出した後には分野間の関連度の計算をおこなう。今回は、より標準的な手法を用いて計算するため以下の余弦公式を用いた。 $W$  は今回は 1 として計算している。

$$Sim(A, B) = \frac{\sum_{i=1}^n W(A_i) \times W(B_i)}{\sqrt{\sum_{k=1}^n (W(A_k))^2} \times \sqrt{\sum_{j=1}^n (W(B_j))^2}}$$

### 3.2 国立環境研究所の分析分野間の関連

上記の余弦公式の  $A$  および  $B$  に表 1 の NIES 分析用分類分野をあてはめ、それぞれのグループのキーワードから分野間の関連度を計算した。その結果を表 5 に示す。この結果より相互に関連度の高い分野が大きい数値で示され、ChemEnv（化学環境）と Env-risk（環境リスク）や SocEnv（社会環境）と EarthEnv（地球環境）などが関連の強いことがわかる。また、WarmerClimate（地球温暖化プロジェクト）などはあまり大きな値になる分野が見られず、やや特殊な分野として考えられる。

これらの関係を図示するために、これら

表 5 キーワード分析による国立環境研究所の分析用分野間の関連度

| 分類分野          | Aerosphere | AreaEnv | BasinZone | Bio-diversity | BioEnv | ChemEnv | EarthEnv | Env-risk | HealthEnv | Infra-Info | PM-DEP | Recycle | SocEnv | WarmerClim | WaterClod | dyoxin-proj | ozone |
|---------------|------------|---------|-----------|---------------|--------|---------|----------|----------|-----------|------------|--------|---------|--------|------------|-----------|-------------|-------|
| Aerosphere    | 1.000      | 0.475   | 0.189     | 0.068         | 0.380  | 0.387   | 0.636    | 0.292    | 0.464     | 0.395      | 0.541  | 0.217   | 0.419  | 0.184      | 0.311     | 0.238       | 0.55  |
| AreaEnv       | 0.475      | 1.000   | 0.540     | 0.207         | 0.612  | 0.760   | 0.643    | 0.666    | 0.720     | 0.562      | 0.507  | 0.608   | 0.694  | 0.208      | 0.719     | 0.555       | 0.23  |
| BasinZone     | 0.189      | 0.540   | 1.000     | 0.174         | 0.326  | 0.402   | 0.395    | 0.364    | 0.265     | 0.323      | 0.160  | 0.388   | 0.505  | 0.096      | 0.525     | 0.280       | 0.08  |
| Bio-diversity | 0.068      | 0.207   | 0.174     | 1.000         | 0.179  | 0.149   | 0.155    | 0.147    | 0.103     | 0.133      | 0.075  | 0.125   | 0.180  | 0.035      | 0.196     | 0.155       | 0.03  |
| BioEnv        | 0.380      | 0.612   | 0.326     | 0.179         | 1.000  | 0.534   | 0.459    | 0.482    | 0.543     | 0.497      | 0.372  | 0.323   | 0.456  | 0.129      | 0.503     | 0.450       | 0.20  |
| ChemEnv       | 0.387      | 0.760   | 0.402     | 0.149         | 0.534  | 1.000   | 0.524    | 0.852    | 0.648     | 0.484      | 0.419  | 0.499   | 0.551  | 0.183      | 0.563     | 0.674       | 0.24  |
| EarthEnv      | 0.636      | 0.643   | 0.395     | 0.155         | 0.459  | 0.524   | 1.000    | 0.427    | 0.514     | 0.447      | 0.386  | 0.374   | 0.851  | 0.324      | 0.495     | 0.341       | 0.37  |
| Env-risk      | 0.292      | 0.666   | 0.364     | 0.147         | 0.482  | 0.852   | 0.427    | 1.000    | 0.618     | 0.359      | 0.358  | 0.411   | 0.490  | 0.161      | 0.464     | 0.825       | 0.23  |
| HealthEnv     | 0.464      | 0.720   | 0.265     | 0.103         | 0.543  | 0.648   | 0.514    | 0.618    | 1.000     | 0.430      | 0.370  | 0.319   | 0.512  | 0.164      | 0.423     | 0.504       | 0.24  |
| Infra-Info    | 0.395      | 0.562   | 0.323     | 0.133         | 0.497  | 0.484   | 0.447    | 0.359    | 0.430     | 1.000      | 0.360  | 0.399   | 0.431  | 0.146      | 0.486     | 0.336       | 0.18  |
| PM-DEP        | 0.541      | 0.507   | 0.160     | 0.075         | 0.372  | 0.419   | 0.386    | 0.356    | 0.572     | 0.360      | 1.000  | 0.223   | 0.344  | 0.157      | 0.297     | 0.276       | 0.26  |
| Recycle       | 0.217      | 0.608   | 0.388     | 0.125         | 0.323  | 0.499   | 0.374    | 0.411    | 0.319     | 0.399      | 0.223  | 1.000   | 0.475  | 0.133      | 0.495     | 0.361       | 0.11  |
| SocEnv        | 0.419      | 0.694   | 0.505     | 0.180         | 0.456  | 0.551   | 0.851    | 0.490    | 0.512     | 0.431      | 0.344  | 0.475   | 1.000  | 0.316      | 0.546     | 0.387       | 0.21  |
| WarmerClimate | 0.184      | 0.208   | 0.096     | 0.035         | 0.129  | 0.183   | 0.324    | 0.161    | 0.164     | 0.146      | 0.157  | 0.133   | 0.316  | 1.000      | 0.134     | 0.097       | 0.13  |
| WaterClod     | 0.311      | 0.719   | 0.525     | 0.196         | 0.503  | 0.563   | 0.495    | 0.464    | 0.423     | 0.486      | 0.297  | 0.495   | 0.546  | 0.134      | 1.000     | 0.373       | 0.16  |
| dyoxin-proj   | 0.238      | 0.555   | 0.280     | 0.155         | 0.450  | 0.674   | 0.341    | 0.625    | 0.504     | 0.336      | 0.276  | 0.361   | 0.387  | 0.097      | 0.373     | 1.000       | 0.18  |
| ozone         | 0.551      | 0.237   | 0.085     | 0.031         | 0.203  | 0.247   | 0.371    | 0.238    | 0.246     | 0.185      | 0.268  | 0.110   | 0.218  | 0.135      | 0.163     | 0.182       | 1.00  |

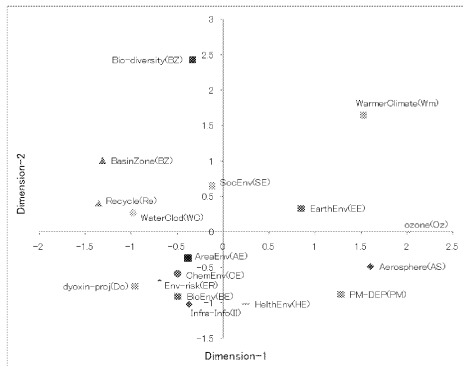


図1 国立環境研究所の分類分野の相互関係（多次元尺度法による）

の関係を関連度から距離行列に変換し多次元尺度法により図示したのが図1である。次元の解釈は難しいが、総合的な分野が中心付近に集まり、Aerosphere（大気圏）、ozone（成層圏オゾン変動プロジェクト）、PM-DEP（PM2.5-DEPプロジェクト）のグループやBio-diversity（生物多様性プロジェクト）、WarmerClimate（地球温暖化プロジェクト）などがやや離れた位置に示されている。

### 3.3 国立環境研究所と科研費細目との関連

3.2章の国立環境研究所の分類用分野間の分析と同様に余弦公式を用い、国立環境研究所の分類用分野と科学研究費補助金の年度別・細目別採択課題から抽出したキーワード群に対して関連度を計算した。

計算は国立環境研究所の17分類と1976年度から2009年度の各細目分野（約300細目）についておこなったが、時間の都合により経年変化の分析はできていないので、この中から1999年、2004年、2009年についての分析結果を紹介する。1999年については17×267細目、2004年については17×309細目、2009年については17×312細目の計算を行っている。これらの分類分野×細目分野の関連度が0.25以上になった組み合わせを選び、さらに国立環境研究所のそれぞれの分類分野に対して関連度が高い上位10細目を選んだ。この結果1999年に関しては選ばれた科研費細目の異なり数は28、2004年に関しても28、2009年に関しては31の異なり細目分野が選ばれ、年度別の3組の17×28、17×28、17×31の分類分野×細

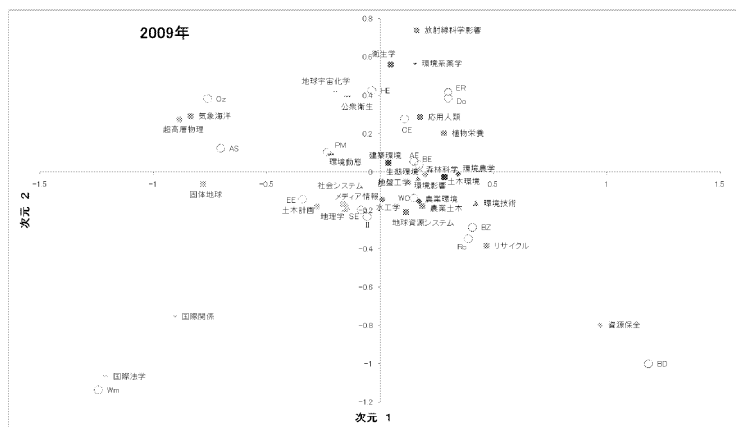


図2 国立環境研究所の分類分野と2009年の科研費細目表との関連

目分野に関する関連度行列が作成された。

これらの関係を表す場合に厳密な利用法では問題があるが、関係を図示することに関しては大きな問題が無いので、行列を変換して対応分析を行った上で結果の2成分の関係を図示する方法をとった。その結果を図2(2009年)、図3(2004年)、図4(1999年)に示す。図1の分類分野相互の関係をほぼ180度回転した形でこれらの2軸上に国立環境研究所の分類分野は示されている。図が煩雑になるので、分

類分野は表1の括弧中の2英文字表記としている。これらの結果はまだ予備的なものだが、科研費細目との関係をよく表しているように思える。

2009年の図2に関しては特にスケールの問題から煩雑となり、中心部分に集中しているが、この中でもWm(地球温暖化プロジェクト)などがやや離れた位置に示されて、関連分野に国際法学や交際関係など他とはやや違った様相を示している。この他、AS(大気圏)やOz(成層圏オゾン変動プロ

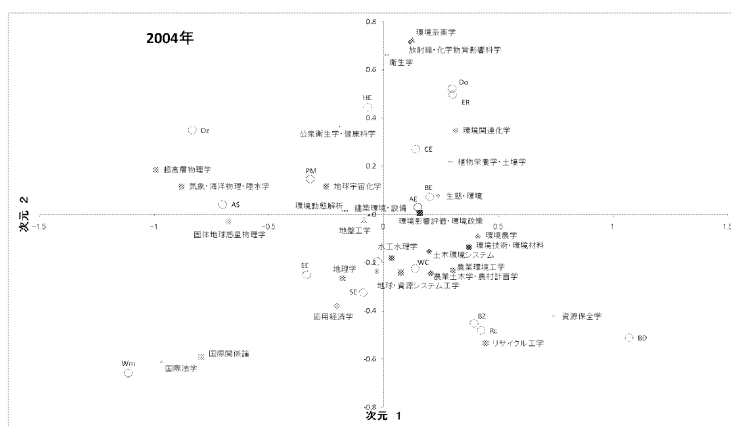


図3 国立環境研究所の分類分野と2004年の科研費細目表との関連

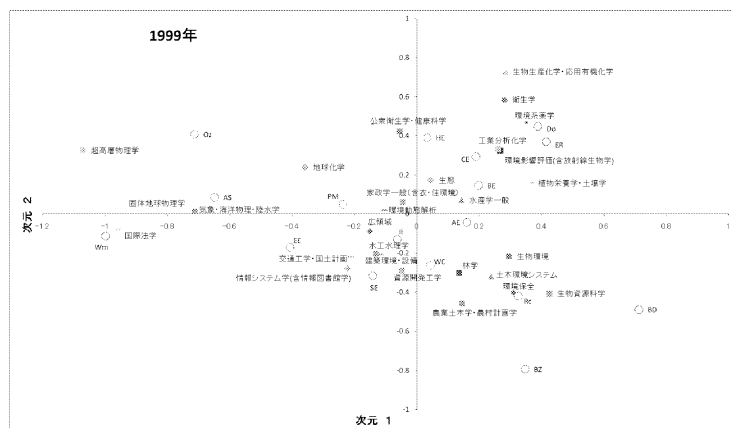


図4 国立環境研究所の分類分野と1999年の科研費細目表との関連

ジェクト)の近くには超高層物理や気象海洋などが示され、中心部分からはやや離れた様子が伺える。2004年の図3に関しては、2009年とあまり変わりはないように思われる。ただ、Wm(地球温暖化プロジェクト)やBD(生物多様性プロジェクト)が2009年ほど次元2のマイナス方向にずれていないために様子が変わって見えているのであろう。1999年の図4に関しては細目分野が大きく変わっているため様子が変わっているが、細目変更後の分野はやはり同じような位置にプロットされている。今後、年度ごとの時系列変化を見る場合にはさらに表現方法を工夫する必要があるだろう。

#### 4 まとめ

本論文では環境分野関連の研究の動向を国立環境研究所のおよそ20年にわたる研究課題と科学研究費補助金の研究課題等を比較し、キーワード分析により他の研究分野との関連や動向について調査した結果を示した。暫定結果ではあるが、キーワード群の相互の類似度と多次元尺度法や対応分析により、国立環境研究所の分類分野の相互関係や、科学研究費補助金の細目分野との位置関係がよく示されることがわかった。

今後は不要な広い意味のキーワードの除外や、類似度計算の重みの最適化、キーワードの組み合わせによる個別の関連課題の抽出方法等について考察していく予定である。また、国立環境研究所の成果における研究代表者の科研費番号の特定もほぼ終わっているため、環境関連の研究者がどの細目分野で研究を行っているかという方面からも実証したいと考えている。

#### 参考文献

- [1] 総合科学技術会議、科学技術基本計画  
<http://www8.cao.go.jp/cstp/kihonkeikaku/index3.html>
- [2] 科学研究費補助金,  
<http://www.jsps.go.jp/j-grantsinaid/index.html>
- [3] 2005年4月からKAKEN, KAKENデータベースを統合して科学研究費補助金データベース(KAKEN)としてサービスされている。  
<http://seika.nii.ac.jp/>
- [4] 西澤 正己、孫 媛、「科研費分野変更に伴う情報学分野の変遷マッピング」, 情報知識学会 第12回研究報告会講演論文集, 2004年5月, pp.9-12
- [5] Masaki Nishizawa, Yuan Sun, "Research Fields related to Information Science under a new classification in Japan", Proceedings on 4th International Conference on University Evaluation and Research Evaluation, pp67-73, Whan University, China, 27 September (2004)
- [6] 西澤 正己、孫 媛、「キーワード分析による科研費におけるゲノムおよびナノテクノロジー関連研究の動向調査」, 情報知識学会誌, Vol.17, No.2, pp.117-122, 2007
- [7] 国立環境研究所:  
<http://www.nies.go.jp/>
- [8] 形態素解析器: 茶筌システム,  
<http://chasen-legacy.sourceforge.jp/>

## 資源供給情報を利用した耐熱材料の選択指針

### Guideline on Selection in Heat Resistance Materials Using Information of Resources

芳須 弘<sup>1\*</sup>, 藤田充苗<sup>1</sup>, 原田幸明<sup>1</sup>,

Hiroshi YOSHIZU, Mitsutane FUJITA, and Kohmei HALADA

1 物質・材料研究機構

National Institute for Materials Science

〒305-0047 茨城県つくば市千現 1-2-1

E-mail: YOSHIZU.Hiroshi@nims.go.jp

材料資源の持続的供給が可能な社会を実現を目指してマテリアルリスクに関する調査を進めている。前回の報告では、耐熱材料に関する資源の持続的供給、環境変動、そして毒性の3種類のリスクについてマテリアルリスクデータベースの設計と活用として環境変動リスクについて耐熱材料の高温特性とCO<sub>2</sub>などの環境負荷ガス発生量との関係を示した。本報告では、Webで公開されている価格・生産量などの情報を収納するデータ構造を検討し、持続的供給リスクとして、各合金元素を製造するために関与する物質総量(TMR)、価格および供給リスクを考慮し、耐熱材料の最適選択や材料開発の方向を示す。

In the last annual meeting, it was introduced that a database stored three kinds of risks data for sustainable supply of the metallic resources, environmental change, and toxic of materials was necessary to consider the sustainable supply of materials. As an example of the application of the database, the guide of the materials selection of the apparatus design used at high temperature based on the optimal environmental conformity was clarified.

In this report, the data structure of the database to stored necessary information provided in the internet and to calculation TMR (total materials requirement), cost and risk of the sustainable supply of each metallic elements formed heat-resistant materials is discussed. Using the database, it is indicated that the guide of the materials selection of the apparatus design using at high temperature based on the possibility of the sustainable supply in those elements.

キーワード: マテリアルリスク指標, 持続的供給リスク, データベース, 材料選択

Materials-Risk-Index, Sustainable supply risk, Database, Materials selection

## 1 はじめに

材料資源の持続的供給が可能な社会を実現するための取り組みとして、前回の報告[1]では、耐熱材料に関する資源の持続的供給、環境変動、そして毒性の3種類のリスクについて、その情報を格納するマテリアルリスクデータベースの概念設計を行った。さらに、そのデータベースの活用例を報告した。活用例として環境変動リスクについて耐熱材料の高温特性と耐熱材料製造時のCO<sub>2</sub>などの環境負荷ガス発生量との関係から耐熱材料の材料選択指針を示した。本報告は、3種類のリスクの中から持続供給リスクを取り上げ Web で公開されている持続的供給に関する価格、生産量などの情報を収集、格納するためのデータ構造を検討した。そのデータベースを用いて耐熱材料を構成する各合金元素の埋蔵量やそれらを製造するために関与する物質の総量などを考慮し、耐熱材料の最適選択や材料開発への方向を示す。

## 2 持続的供給リスクについて

前回の報告では、図1に示すような材料資源の持続的供給を目指した機器設計用材料探索の際に使用されるマテリアルリスク指標データベースの概念を示し、耐熱材料製造時の環境負荷ガスをもとにした耐熱材料の選択指針を示した。ここでは、資源（合金元素）の持続供給を考慮した材料選択指針を検討する。耐熱材料を構成する合金元素を供給する資源の持続可能性については、物質・材料研究機構の原田らにより地球環

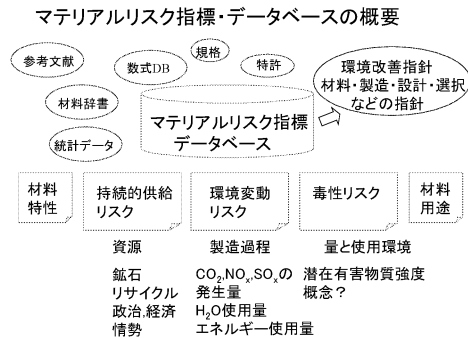


図1 マテリアルリスク指標データベースの概要

境、人類経済、そして国の経済の3つの観点から論じられている[2]。

### TMR

資源を得るための地球環境へのリスク指標のひとつとして関与物質総量（TMR: Total Materials Requirement, 以下、関与物質総量を TMR と略）[3]があげられる。TMR は、エコロジカル・リュックサック[4]とも呼ばれ、Shumidt-Bleek（独）が提唱した。物質生産時の環境影響への度合いを示す指標であり、物質1トンを得るために何トンの天然資源を改変したかという量である[2]。TMR は下記の式により算出する。

$$\begin{aligned} \text{TMR (関与物質総量)} = & \Sigma (\text{直接投入物質量}) \\ & + \Sigma (\text{間接投入物質量}) \\ & + \Sigma (\text{隠れた物質フロー量}) \end{aligned}$$

ここで、直接投入物質、間接投入物質とは、人間の経済的行為として直接もしくは間接的に投入された物質をいい、隠れた物質フローとは、直接、間接以外の行為に伴う物質の移動（採掘に伴う岩石、土砂の移動、森林の伐採など）に必要な物質をいう。

## 資源枯渇加速度

人類経済の持続性への指標は資源疲弊度（資源の耐用年数）があげられる。その算出方法は、以下の式により算出する。

資源 I の耐用年数＝

資源 I の埋蔵量／資源の消費量  
年間資源枯渇加速度＝

年間生産量×資源枯渇加速度

## 価格

国レベルの持続性への指標は、供給安定性があげられ価格の安定性と資源の偏在性が問題となる。それぞれの数値は以下の式により算出する。

価格の安定性＝

価格の平均値／価格の最大値

資源の偏在性＝

1－第一供給国のシェア

## 供給リスク

供給リスク＝

(1－価格の安定性)

×(1－資源の偏在性)

資源の持続的供給を評価する上で、上記の指数の算出のもとになる数値は、公的機関の Web サイト上にそれぞれの公的機関の所掌範囲の数値が提供されている。その数値は社会情勢によって時々刻々と変化するため、資源の持続的供給を評価するためのデータベースが必要となる。このデータベースは、図 1 の持続的供給リスクサブデータベースに相当する。

## 3 持続的供給に関する情報

資源の持続的供給リスクに関する情報は、公的機関の Web サイトや地球環境

関連の文献などから必要な情報が得られる。例えば、耐熱材料を構成する合金元素に関する資源情報、製品（耐熱材料）製造における TMR、資源の価格情報などがある。

### 3.1 情報源

持続供給リスクに関する指標を算出する上で必要となる情報源として、以下に列挙したような公的機関の Web サイトや文献などがある。

- ・財務省貿易統計[5]では、国別（品目別）の生産・出荷等。

- ・経済産業省[6]では、品目別生産額、資源の生産・出荷等。

- ・U.S. Geological Survey[7]では、各国資源の埋蔵量、生産量等。

- ・石油天然ガス・金属鉱物資源機構[8]では、金属資源マテリアルフロー、需給動向、輸出入動向、価格動向等。

- ・物質・材料研究機構[9]では、元素戦略データ、元素の特性、物質収支等データ、マテリアルフロー等。

- ・その他（環境関連の論文、資料集等）TMR、マテリアルフロー等のデータが提供されている。

以上のような国内外の公的機関の Web サイトから資源の埋蔵量、価格、あるいは持続供給の指標となる製品製造に関与する物質の総量などの情報が取得可能である。

### 3.2 データ項目と構造

Web サイト等から収集した持続的供給に関する情報を格納する持続的供給リスク（サブデータベース）のデータ構造は、図 2 のようになる。これらの情報

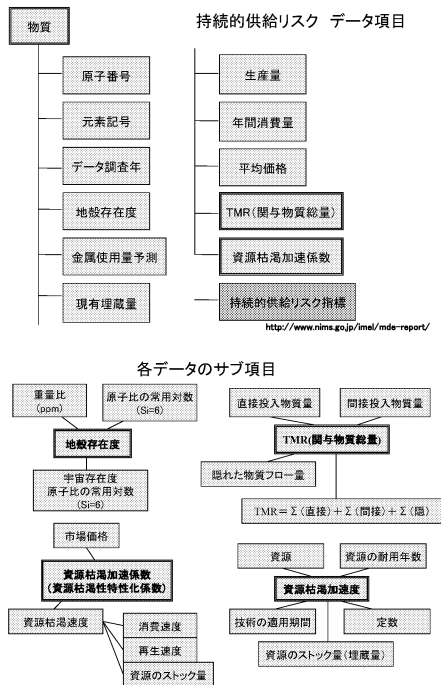


図2 持続的供給リスクのデータ項目と構造

をもとに、前述の TMR、価格、供給リスク（ここで、持続的供給リスクは総称、供給リスクは合金元素など個別レベルとする）などを各元素について計算できる。持続的供給リスクについてその主な項目を表1に示す。表1にはその算出例

として、一部元素の年間消費、現有埋蔵量、平均価格、そして元素の製造に関わる TMR などが記載されている。

### 3.3 耐熱材料の各種供給リスクの算出

耐熱材料の特性情報と持続的供給情報との関係を調べるため、表1のデータ項目から必要情報として耐熱材料の TMR、価格（その材料を構成する合金元素から計算した総価格）および供給リスクを用いた。

## 4 耐熱材料の資源供給の検討

ここでは、持続的供給リスクの算出に必要な情報の中から TMR、資源（合金元素）価格および供給リスクを、これらと関係付けるため耐熱材料の特性（負荷応力、長時間強度指定による耐熱材料の推定温度を算出）との関係について調べ、以下にその結果を述べる。

### 4.1 耐熱材料の特性

耐熱材料の特性として物質・材料研究機構クリープデータシートの 41 鋼種 [10]を用いて、負荷応力 100MPa、破断時

表1 持続的供給リスクに関するデータ項目

まであり Vol.46 No.8 (2007) p.543-548 表1より抜粋

| 原子番号 | 記号 | 地殻存在度  | 年間消費       | 現有埋蔵量       | 平均価格  | TMR(関与物質総量) | 資源枯渇加速係数 | 資源供給リスク |
|------|----|--------|------------|-------------|-------|-------------|----------|---------|
|      |    |        | t/y        | t           | \$/kg |             |          |         |
| 5    | B  | 10     | 4600000    | 410000000   | 9.31  | 140         | 9600     | 400     |
| 13   | Al | 84100  | 31601000   | 23000000000 | 1.6   | 48          | 0.4      | 2       |
| 14   | Si | 267700 |            |             | 1.45  | 34          | 0.1      | 1       |
| 22   | Ti | 5400   | 78000      | 54000000    | 9.92  | 36          | 12       | 7       |
| 23   | V  | 230    | 44000      | 13000000    | 8.28  | 1500        | 300      | 200     |
| 24   | Cr | 185    | 13500000   | 7500000000  | 6.36  | 26          | 42       | 12      |
| 25   | Mn | 1400   | 8110000    | 380000000   | 14.3  | 14          | 100      | 13      |
| 26   | Fe | 70700  | 1099000000 | 80000000000 | 0.51  | 8           | 1        | 1       |
| 27   | Co | 29     | 47600      | 7000000     | 34.27 | 600         | 1600     | 300     |
| 28   | Ni | 105    | 1246000    | 40000000    | 8.31  | 260         | 1100     | 200     |
| 29   | Cu | 75     | 16287000   | 470000000   | 2.38  | 360         | 290      | 120     |
| 30   | Zn | 80     | 10082000   | 220000000   | 1.23  | 36          | 100      | 70      |
| 40   | Zr | 100    | 864000     | 38000000    | 28.41 | 550         | 560      | 200     |
| 41   | Nb | 11     | 32800      | 4400000     | 9.34  | 640         | 2100     | 400     |
| 42   | Mo | 1      | 151000     | 8600000     | 13.19 | 750         | 10000    | 1000    |
| 73   | Ta | 1      | 1210       | 43000       | 85.16 | 6800        | 66000    | 8000    |
| 74   | W  | 1      | 59100      | 29000000    | 152.4 | 190         | 33000    | 1000    |



間 10 万時間を指定条件とした時の推定温度を算出して用いた。この推定温度は各耐熱材料の指定条件で適用できる温度を示すもので、図 3 に算出した結果を示す。図 3 は、縦軸に温度、横軸に各耐熱材料（1 材料につき複数ヒートあり）を示す。この結果では、各耐熱材料の耐食性重視あるいは原子力プラントのような高温環境下での使用などその使用目的により温度域が異なり、炭素鋼（○）は低温側にあり、以下、低合金鋼（△）、高合金鋼（□）及びステンレス鋼（◇）、そして超合金（\*）の順に高温側へ移行する。超合金のばらつきはベースとなる金属が鉄やニッケルであることと使用目的の違いによる。

#### 4.2 TMR, 価格および供給リスクの算出

図 4 に各耐熱材料の TMR, 価格および供給リスクを示す。数値算出には表 1 の各合金元素について算出されている数値を係数として引用し、各耐熱材料の主要合金元素について計算し、その数値を積算した。

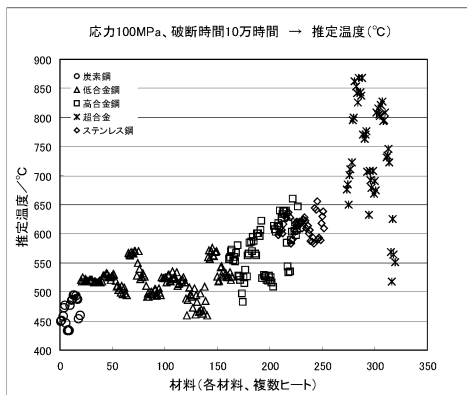


図 3 耐熱材料の特性  
(負荷応力 100MPa, 破断時間 10 万時間指定時の推定温度)

#### 耐熱材料の TMR

TMR について述べる。図 4 a に各耐熱材料 1 トン当たりについて算出した TMR を示す。縦軸に TMR, 横軸に各耐熱材料を示す。シンボルについては図 3 の特性と同じである。図 3 の結果から各耐熱材料の TMR は、鉄が多くを占める炭素鋼や低合金鋼は TMR が少ないが、Ni や Mo などの添加合金元素の種類の数やその添加量が多い超合金ほど多くなっている。超合金については特性でも述べたがベースとなる金属により TMR 量の差が大きくなってばらついている。

#### 耐熱材料の価格

各耐熱材料の 1 トン当りの価格を図 4b に示す。図 4 の結果から、安価な鉄が多くを占める炭素鋼や低合金鋼は積算した総価格が安くなっているが、超合金のように Ni や Mo などの添加合金元素の種類の数やその添加量により高価なものになっている。また、超合金の中でも特に高価な値を示している材料はベースとなる金属が高価なコバルトであることが起因する。

#### 耐熱材料の供給リスク

耐熱材料 1 トン当りの供給リスクの結果を図 4c に示す。耐熱材料の供給リスクについてもその傾向として炭素鋼から超合金へと供給リスクが高くなっている。これは炭素鋼や低合金鋼などはベースが鉄であり、鉄は埋蔵量も多く、またリサイクルなどで再利用されることも起因してリスクが低くなっている。これに対して埋蔵量が少ない合金元素を多く含有する耐熱材料は当然供給リスクも高くなる。

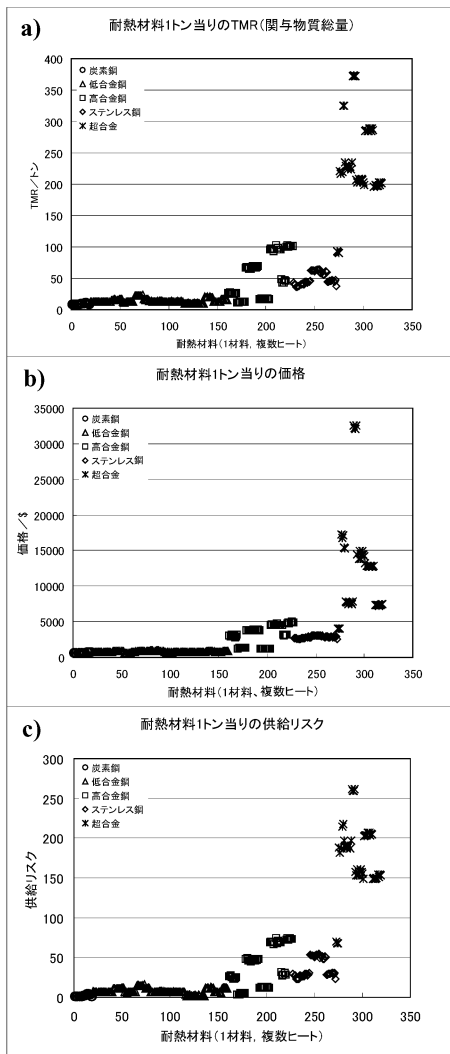


図4 耐熱材料の a)TMR, b)価格および c)供給リスク

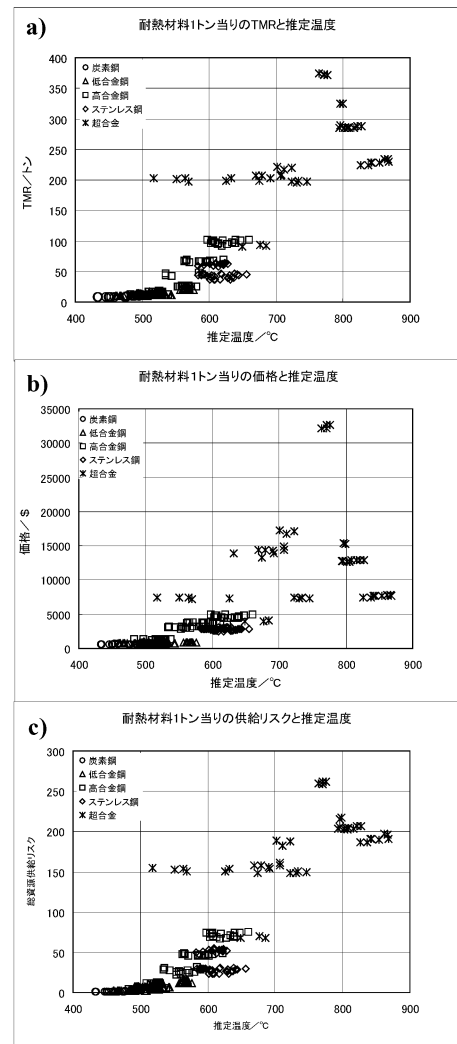


図5 耐熱材料の特性と a)TMR, b)価格および c)供給リスクの関係

### 4.3 耐熱材料の特性と各項目の関係

ここでは、耐熱材料の特性と TMR、価格および供給リスクとの関係について調べた結果 (図5) を述べる。

#### 特性と TMR

耐熱材料の特性と TMR の関係を図 5a に示す。両者の関係は、超合金の TMR の数値が特出しているが、これは採掘か

ら製品化するまでの過程において合金元素の種類の多さ、埋蔵量あるいは採掘等による物質の移動などが要因と考える。また、超合金についてはばらつきがみられるが、材料のベース金属が TMR の少ない鉄であり、対して TMR の多いニッケルが起因する。さらに、温度域についてもばらつきがあり、これは耐食性

の重視あるいは高温環境下の強度の重視などその使用目的による。

### 特性と価格

ここでは、耐熱材料を構成する合金元素の Fe, Ni などの主要合金元素についてここ数年の価格推移（計算では表 1 の平均価格を適用）をもとに耐熱材料の価格を算出し、特性との関係を調べ、その結果を図 5b に示す。両者の関係は TMR と同様な傾向を示したが、価格については安価な鉄がベースの炭素鋼や低合金鋼は低温、安価に位置し、高合金鋼やステンレス鋼は中間域に、ニッケルをベースとして合金元素の数が多い超合金は高温、高価に位置する。

### 特性と供給リスク

特性と供給リスクとの関係についても他の 2 項目と同様な傾向を示す。その結果を図 5c に示す。材料別ではやはり合金元素の数が少ない炭素鋼が供給リスクが低く、高合金鋼、超合金など合金元素の数や添加量が多くなるほどその埋蔵量や資源確保により供給リスクが高くなる。

### 4.4 各情報による材料選択

TMR, 価格, そして供給リスクの 3 項目について耐熱材料の特性との関係を調べたが、どの関係も炭素鋼から低合金鋼, 高合金鋼およびステンレス鋼, そして超合金へと各図の結果では右肩上がりの傾向を示した。これは左下に位置する材料は TMR が少なく, 安価, そして供給リスクが低いことを示し, 逆に右上に位置する材料は, TMR が多く, 高価, そして供給リスクが高いことがわかる。繰り返すがこれは合金元素の種類の数, 添加量の多少が起因する。これらの結果から, 同じ条件下（負荷応力 100MPa, 破断時間 10 万時間）で材料を選択する場合には TMR の削減, 安価, そして供給リスクの低減を考慮し, より高温域の材料（図 4, 図 5 の右下の領域）を選択することが望ましい。

## 5 おわりに

前回の報告では、耐熱材料の特性（長時間強度）と環境変動リスクとの関係を

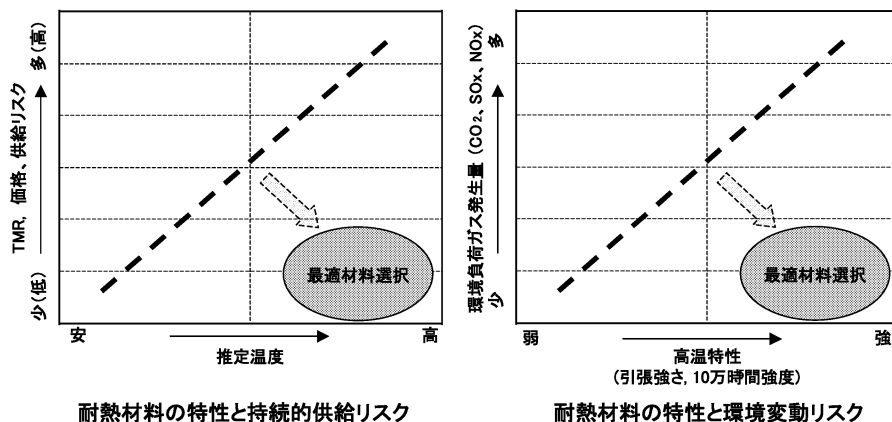


図 6 耐熱材料の特性とマテリアルリスクからの材料選択（模式図）

調べ、材料選択の方向を示した。今回は、耐熱材料の特性と耐熱材料製造時のTMR、価格、そして供給リスクの3種類の関係を調べ、各関係から資源の持続可能な材料選択への指針を図6に耐熱材料の製造時に発生する環境負荷変動リスク（環境負荷ガス）と材料特性との関係をあわせて示した。持続的供給リスクも環境変動リスクと同じように高温で使用する可能な材料ほどリスクが増加する傾向にあり、耐熱材料の供給リスクを低減する方向で材料選択すべきである。

まとめとして、

- 1) 耐熱材料の高温特性の良好な材料ほど合金元素の種類の数と添加量が多く、TMR、価格、そして供給リスクの数値がいずれも大きくなる。
- 2) 使用温度が同じであればTMR、価格、そして供給リスクを低減する方向で材料を選択すべきである。

最後に、本研究は平成 20 年度の日本学術振興会科学研究費補助金（基盤研究B：課題番号 208538）の支援を受けたことを報告する。

## 参考文献

[1] 芳須 弘他：「マテリアルリスク指標データベースの設計と活用」，情報知識学会，第 17 回年次大会，pp.112-119，2009

[2] 原田幸明：「希少資源・元素の現状」，まてりあ，vol.46, No.8, pp.543-548, 2007.

[3] 中島謙一他：「関与物質総量（TMR）の算定」，Journal of Life Cycle Assessment, Japan, Vol.2, No.2, pp.152-158, 2006.

[4] F.シュミット＝ブレーク：エコリユクサック，省エネルギーセンター，2006

[5] 財務省貿易統計 HP：

<http://www.customs.go.jp/toukei/>（2010 年 1 月参照）

[6] 経済産業省 HP：

[http://www.meti.go.jp/statistics/tyo/seidou/result/ichiran/01\\_tekko.html](http://www.meti.go.jp/statistics/tyo/seidou/result/ichiran/01_tekko.html)（2010 年 1 月参照）

[7] U.S. Geological Survey：

<http://minerals.usgs.gov/>（2010 年 2 月参照）

[8] （独）石油天然ガス・金属鉱物資源機構 HP：

<http://www.jogmec.go.jp/index.html>（2010 年 1 月参照）

[9] 物質・材料研究機構元素戦略センターHP：<http://www.nims.go.jp/ecomaterial/SMAND/database/top.html>（2010 年 2 月参照）

[10] （独）物質・材料研究機構：クリープデータシート，No.1-No.41.

## 学術俯瞰、特許俯瞰による商業化可能な技術要素の早期発見

# Early Detection of Commercialization Opportunity By Analyzing Scientific and Technological Landscapes

柴田 尚樹<sup>1\*</sup>, 梶川 裕矢<sup>1</sup>, 坂田 一郎<sup>1</sup>

Naoki SHIBATA<sup>1\*</sup>, Yuya KAJIKAWA<sup>1</sup>, Ichiro SAKATA<sup>1</sup>

<sup>1</sup>東京大学

The University of Tokyo

〒113-0033 東京都文京区本郷7-3-1

E-mail: {shibata, kaji, isakata}@ipr-ctr.t.u-tokyo.ac.jp

二次電池のケーススタディを通して、学術俯瞰、特許俯瞰を通して、商業化可能な技術要素の早期発見を行う方法論を提案する。学術論文は基礎的、特許は応用的になる傾向があった。学術論文と特許それぞれの主要5領域が扱うトピックを比較すると、リチウム二次電池と電解質に関する2組の対応関係が見られ、トピックの内容や平均出版年から学術論文と特許の差分を抽出した。その結果、商業化の可能性のある領域が発見された。

In this paper, we perform a case study of secondary batteries research and technology to develop a method of extracting commercialization gaps between science and technology. Scientific research tended to be basic, whereas patents focused on applied technology developed from the basic research. Of five science research fronts and five areas of technology, three scientific groups, Li(Mn,Co,Ni)O-, Li-Sn- and olivine-based batteries, corresponded to the largest set of patents relating to Li. These three academic outcomes are so recently published that these can be the candidates for industrial opportunities. Other than Li, the largest set of academic papers correlated semantically to the third largest set of patents. Although patents aimed physically such as covering the electrolytes more strictly, scientific researches aimed to polymerize the electrolytes used for secondary batteries, while. There also exists an opportunity for polymer electrolytes to solve the problems relating to electrolytes. Our results offer an intellectual basis for detecting possible opportunities for industry.

キーワード: 研究開発, テクノロジー・ロードマップ, 書誌分析, 引用分析, 特許分析

R&D management, Technology roadmap, Bibliometrics, Citation Analysis, Patent Analysis

## 1 序論

今日のように経済が急速にグローバル化、知識化する環境下では、競争力、成長力は、基礎的なイノベーションの源泉を発見し、速やかにアプリケーションを作り出す能力に依存する。KostoffとSchallerは科学と技術の両方を理解するだけではなく、それらを結びつける高品質なロードマップを構築することの重要性を述べている[1]。従って、R&Dマネージャーや政策意思決定者の両方にとって、科学と技術の関係を理解することは不可欠である。まだ特許にならない新しい科学知識は産業にとってチャンスの源泉となり得る。一部の科学知識は基礎的すぎて商業化が難しい場合もあるが、科学知識と技術要素の対応関係を調べることは、科学と技術の間にある差分を発見するという点で非常に重要である。

科学知識と産業技術を対応づける一つの方法は、学術論文と特許の意味的な類似度を測定することである。論文と特許のトピックスを比較する前に、本研究ではそれぞれの代表的なトピックの抽出を行った。論文群、特許群それぞれで、文書を分類しトピックを抽出する方法としては、テキストマイニングと引用マイニングの二つがある。Kostoffらは、複合語の頻度や出現近接性を分析することで、エネルギー関連研究における分類構造を抽出した[2]。後者の引用分析は知識構造を分析するために広く用いられている方法である。引用分析においては、引用している文書と引用されている文書の間には何らかの意味的な関係があるという想定がなされる。例えば、Smallは学術論文の共引用ネットワークを分析し、研究領域における新しい成長分野

を発見する方法を提案した[3]。本研究では、新しい知識を発見するのに優れていることが実証されている引用分析を用いることとする[4]。引用ネットワークをクラスタリングすることで、論文、特許といった文書をそれぞれのレイヤーで分類し、トピックを比較することで、論文と特許の間の差分を発見する。本研究の目的は、学術論文と特許の対応関係を分析することで、学術研究としては存在するが、特許にはなっていない商業化可能な技術要素を早期に発見するための方法論を構築することである。この方法論を二次電池のケースを用いて実証する。本研究によって、産業にとってのチャンスを発見するための重要な知見が得られると考えられる。

## 2 手法

学術論文と特許の差分を発見する手法は大きく二つのフェーズからなる。第一フェーズでは論文と特許のそれぞれをクラスターに分類する。この分類手法は、新知識発見に用いられるなど、既に確立された手法である[4]。第一フェーズでは、この(1)から(6)までのステップが、論文群、特許群のそれぞれに適用される。第二フェーズでは、主要な論文クラスター、特許クラスター間のトピックをエキスパートが比較する。比較的最近出版された論文からなる論文クラスターに対応する特許クラスターがなければ、そこには商業化のチャンスがある可能性がある。エキスパートはステップ(5)と(7)のみに必要であり、他のステップは計算機によって自動的に計算可能である。

第一フェーズでは、まずステップ(1)として、文書（論文、特許）を収集する。ここ

では検索クエリを用いて、論文データベース、特許データベースを検索し、対象文書群を得る。学術論文データベースとしては、Institute for Scientific Information (ISI)によって提供されている、Science Citation Index Expanded (SCI-EXPANDED)、Social Sciences Citation Index (SSCI)、Arts & Humanities Citation Index (A&HCI)を用いた。これらの3つは学術論文データベースとしては最良のデータベースであると考えられる。特許データベースとしては同じくISIによるDerwent World Patents Index (DWPI) by Thomson Innovation (TI)を用いた。DWPIは41の特許分類に対応し、1740万件以上のレコード、3720万件以上の特許を管理する最も信用できる特許データベースである。DWPIを用いることで正規化されたクリーンなデータを得ることが出来る。例えば、同じ内容の特許が異なる国に出願された場合、DWPIファミリーという概念で一つにグルーピングされる。本研究では、同じDWPIファミリーもしくは、同じInternational Patent Documentation Center (INPADOC)ファミリーを有する特許群は一つの特許ファミリーとして扱う。また、これらのデータベースからは出版年、タイトル、著者、概要といった属性データと、引用関係という関係性データの両方が取得可能である。

続いて、ステップ(2)として、論文や特許ファミリーをノード、直接引用をリンクとする引用ネットワークを構築した。ステップ(3)以降は、最大連結成分のみが用いられ、最大連結成分にリンクしていない要素は削除される。

最大連結成分が抽出された後、ステップ(4)として引用ネットワークはトポロジカ

ルクラスタリング手法によってクラスターに分割される[5]。この手法はクラスター数を入力する必要がない手法であり、密に結合した部分をクラスターとして抽出する。

ステップ(5)はデータクリーニングである。クエリ検索によって得られたデータには、本来分析対象にしたいデータ以外のノイズが含まれる。この場合、二次電池関連のデータのみを分析対象にしたかったが、実際に取得したデータには多くのバイオ系の文書が含まれていた。従って、分析対象にしたいデータ以外を除去するというのがステップ(5)である。実際には、2008年のクラスタリング結果から、エキスパートが関係するクラスターのみを抽出した。ステップ(5)の後、ステップ(6)として抽出されたデータのみが再度クラスターに分割された。

第二フェーズとして、エキスパートが主要な論文クラスター、特許クラスター間の対応関係を判定する。ステップ(7)では、各クラスター内の文書の主要文書の書誌情報(タイトル、アブストラクトなど)から、どの論文クラスターと特許クラスターの間に意味的に関係があるかをエキスパートが判定する。

### 3 ケース:二次電池分野

充電電池として知られる二次電池は二次細胞からなる。二次電池が充電可能なのは、電氣的に可逆な電気化学反応のためである。二次電池の大きさは利用される化学物質の組み合わせに依存する。広く利用されている化学物質としては、鉛酸蓄、ニッケルカドミウム、ニッケルメタルヒドライド、リチウムイオン、リチウムイオンポリ

マーなどがある。二次電池は、使い捨て電池と比べて経済的にも環境的にもメリットがある。これらの商業的なニーズに対応するために、より高エネルギー密度な電池を作るための多くの研究開発がなされている。本研究では、クエリとして

"((secondary or storage or rechargeable or reserve) and (cell\*)) or batter\*"を用いて論文、特許の特定を行った。

## 4 結果

ステップ(1)(2)の結果として、2008年までに出版された145,294論文と333,207特許が得られた。次にステップ(3)の後に、77,924論文、101,177特許がそれぞれの最大連結成分として抽出された。さらにステップ(4)(5)において、エキスパートが上位20論文クラスター(65,369 papers, 84%)、上位30特許クラスター(86,610 patents, 85%)を判定し、Table1に示されたように3論文クラスター(20,430 papers, 26%)、19特許クラスター(69,789 patents, 69%)が二次電池関連と判定された。時系列の文書数は図1に示された通りである。1990年以降論文数が急激に増加しているように見えるが、これはバイオ関係の影響によるものであると考えられる。特許は論文よりも、最大連結成分に含まれる文書の割合が小

さい。また学術論文とは別に、特許ではバイオ関連特許のノイズとしての影響が小さいということも読み取れる。

以下の4.1節、4.2節ではそれぞれ論文レイヤー、特許レイヤーでの結果を示す。それぞれのレイヤーで得られた文書群を5章にて比較し、論文と特許の対応関係を考察する。

### 4.1 学術俯瞰の結果

2008年時点での引用ネットワークがステップ(6)でクラスターに分割され、500以上の論文を含むクラスターは5つであった。クラスターS1, S2, S3, S4, S5はそれぞれ、5410, 4851, 4843, 3672, 832本の論文を含み、平均出版年は2000.0, 2003.1, 2003.1, 2000.7, 2005.5であった。5つの主要クラスターのうち、S1, S4の2つは比較的古いのに対して、S2, S3, S5の3つのクラスター内の論文は比較的最近出版されており新興学術領域であると言える。従来から存在するクラスターはS1, S4それぞれ、メタルハイブリッド、ポリマー電解質に関するテーマを扱っている論文群である。他方、新興クラスターS2, S3, S5は、それぞれLi(Mn, Co, Ni)O系、Li-Sn系、olivine系電池を扱っている。まとめると、2008年時点では主に5つの論文クラスター

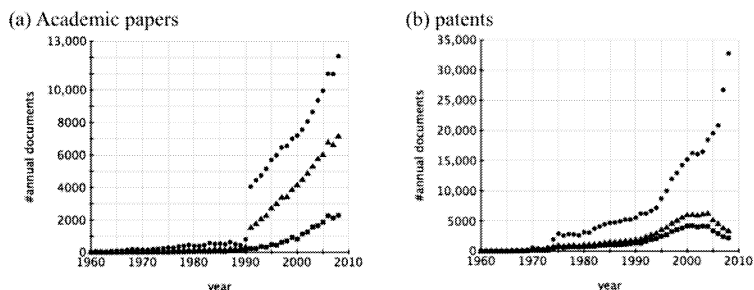


図1 クエリをタイトル、概要に含む各年ごとの文書数：(a)論文、(b)特許  
黒丸：2008年までに出版された全文書数、黒三角：2008年の最大連結成分に含まれる文書数、黒四角：データクリーニング後のデータセットに含まれる文書数



が存在し、メタルハイドロイド、ポリマー電解質に関する論文クラスターは長い研究の歴史があるのに対し、Li (Mn, Co, Ni)O系、Li-Sn系、olivine系電池を扱う3つのクラスターは新興クラスターであると言える。

## 4.2 特許俯瞰の結果

論文に適用されたのと同じ手順を特許の引用ネットワークにも適用した。その結果、2008年時点で5000本以上の特許を含むクラスターは5つであった。各クラスターT1～5はそれぞれ、リチウム系電池のセルデザイン(14892特許、平均出版年1996.1)、自動車への応用(12110特許、平均出版年1996.3)、メタルハイドロイド(10038特許、平均出版年1995.7)、光システムへの応用(7123特許、平均出版年1995.5)、センサーへの応用(6784特許、平均出版年1998.9)を扱っていた。

## 4.3 論文クラスターと特許クラスターの対応関係

最後にステップ(7)において、エキスパートは上位10の論文クラスターと上位10の特許クラスターの間の100通りの組み合わせに対して、意味的な対応関係があるかどうかを判定した。その結果を図2に示す。S1とT3はともにメタルハイドロイドを扱っており明らかに関係があると判定された。また、特許クラスターT1のトピックはリチウム系バッテリーであるが、これは3つの論文クラスターS2, S3, S5に対応する。ポリマー電解質を扱う論文クラスターは、非水性電解質を扱う特許クラスターT9(1549 特許、平均出版年1998.8、スペースの都合でFig. 5には非掲載)に対応する

と判定された。特許レイヤーにおける他の3クラスター(T2, T4, T5)は応用技術を扱っており、対応する論文クラスターは存在しなかった。

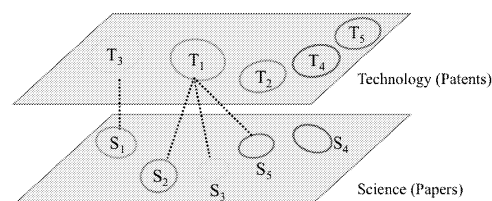


図2. 二次電池関連の学術領域と特許領域の対応関係

## 5 考察

はじめに、本提案手法による商業化可能な技術要素の早期発見の可能性に関して考察する。二次電池の場合、学術論文はより基礎研究を、特許は自動車、光システム、センサーなどより応用的な内容を扱う傾向にあった。上述のように、リチウム電池関連の特許クラスターT1は、特許レイヤーでは最大のクラスターであった。他方、論文クラスターS2, S3, S5は、それぞれLi (Mn, Co, Ni)O系、Li-Sn系、olivine系であり、平均出版年が最近であり新しい学術領域であることから、商業化のポテンシャルがある領域であると言える。特許クラスターT1の中に論文クラスターS2, S3, S5に対応する技術要素がどの程度存在するかをより詳細に調査する必要があるが、少なくともこれらの3つの論文クラスターが将来の商業化チャンスを持つ学術知の候補であると言える。

リチウム系以外では、電解質（論文クラスターS1と特許クラスターT3）に関しても言及したい。学術研究も特許も電解質が漏れることを防ぎたいという同じ目的を共

有しているが、アプローチが異なる。論文クラスターS1は、電解質をポリマー化することでこの問題を解決しようとしているのに対して、特許クラスターT3は物理的により厳密にカバーをすることで漏れを防ぐ方法を議論している。学術研究のアプローチの方が本質的な解決につながる可能性があり、ここでも電解質のポリマー化は、将来的に商業化チャンスがある領域であると言える。提案手法は、少なくとも技術経営戦略を立案し、重点分野を決めるための新たな視点を提供していると言える。

## 6 結論

本研究では、二次電池のケースを通して、学術研究と特許の差分領域を発見するための方法論を提案した。引用分析とエキスパートによる分析を組み合わせることで、学術論文と特許の対応関係を分析した。学術論文はより基礎研究を、特許はより応用的な内容を扱う傾向にあった。主要な5つの論文群のうち、Li(Mn, Co, Ni)O系、Li-Sn系、olivine系の3論文群が、リチウム系を扱う最大の特許群に対応していた。これらの3つの学術領域は平均出版年が最近であり新しい学術領域であることから、商業化のポテンシャルがある領域であると言える。リチウム系以外では、最大の論文群と3番目に大きい特許群が共に電解質を扱っていた。学術研究も特許も電解質が漏れることを防ぐという同じ目的を共有しているが、アプローチが異なる。学術研究は電解質をポリマー化することでこの問題を解決しようとしているのに対して、特許は物理的により厳密にカバーをすることで漏れを防ぐ方法を議論している。学術研究のアプローチの方が本質的な解決につな

がる可能性があり、ここでも電解質のポリマー化は、将来的に商業化チャンスがある領域であると言えた。我々の提案手法は、少なくとも技術経営戦略を立案し、重点分野を決めるための新たな視点を提供できる。

## 謝辞

本研究の一部は科研費 若手研究(スタートアップ) (21800011) の助成を受けたものである。

## 参考文献

- [1] Kostoff, R. N., Schaller, R. R.: "Science and technology roadmaps", IEEE Transactions on Engineering Management, vol. 48, pp. 132-143, 2001.
- [2] Kostoff, R. N., Tshiteya, R., Pfeil, K. M., Humenik, J. A., Karypis, G.: "Power source roadmaps using bibliometrics and database tomography", Energy, vol. 30, pp. 709-730, 2005.
- [3] Small, H.: "Tracking and predicting growth areas in science", Scientometrics, vol. 68, pp. 595-610, 2006.
- [4] Shibata, N., Kajikawa, Y., Takeda, Y., Matsushima, K.: "Detecting emerging research fronts based on topological measures in citation networks of scientific publications", Technovation, vol. 28, No. 11, pp. 758-775, 2008.
- [5] Newman, M. E. J.: "Fast algorithm for detecting community structure in networks": Physical Review E, vol. 69, 066133, 2004.

# 次世代電池における国際的共同研究の構造

## The Structure of International Collaboration in Green Technology Research

佐々木一<sup>1\*</sup>, 梶川裕矢<sup>2</sup>, 坂田一郎<sup>3</sup>

Hajime SASAKI<sup>1\*</sup>, Yuya KAJIKAWA<sup>2</sup>, Ichiro SAKATA<sup>3</sup>

1, 3 東京大学 政策ビジョン研究センター

Todai Policy Alternative Research Institute, the University of Tokyo

〒113-0033 東京都文京区本郷7-3-1 医学部1号館2F

E-mail: [sasaki@utri.co.jp](mailto:sasaki@utri.co.jp), [isakata@ipr-ctr.t.u-tokyo.ac.jp](mailto:isakata@ipr-ctr.t.u-tokyo.ac.jp)

2,3 東京大学 総合研究機構工学系研究科 イノベーション政策研究センター

Innovation Policy Research Center, the University of Tokyo

〒 東京都文京区弥生2-11-16工学部9号館1F 119室

E-mail: [kaji@ipr-ctr.t.u-tokyo.ac.jp](mailto:kaji@ipr-ctr.t.u-tokyo.ac.jp)

\*連絡先著者 Corresponding Author

本稿は、次世代電池に関わる機関間の共同研究の構造を国際的な視点から俯瞰することで、我が国の次世代電池に関わる国際提携戦略の立案等に資することを目的とする。具体的には次世代電池の代表として、燃料電池および太陽電池の2つの研究領域を取り上げる。

手法としては、学術論文データベースより抽出したデータの著者所属機関情報から、延著者数を計算することで、各機関の当該領域の研究実績を表す指標とした。次に、論文の共著関係を抽出することで、研究推進上関係の深い機関や国の特定を行った。以上2つの情報を組み合わせることで、主要機関の共同研究状況をネットワークとして可視化し、研究能力の高い機関の間における提携関係の特定を行った。

結果として、先の手法により次世代電池に関する国際的な提携状況を俯瞰することが出来ることを明らかにした。また、2つの研究領域を比較することで、提携状況に関する領域間の差異を特定した。

This paper aims at understanding a landscape of international collaboration among institutions in research fields of “Fuel cell” and “Photovoltaic” as green technology.

This paper tries to visualize the structure and identify the top institutions. The innovation policy of green technology is proposed from the results.

キーワード: 燃料電池, 太陽電池, 国際的共同研究, ネットワーク, グリーンテクノロジー

Fuel cell, Photovoltaic, International Collaboration, network, Green technology

## 1 はじめに

近年、エネルギー制約が構造的に高まりつつあり、主要国のエネルギー政策における最大の課題となっている。そうした中、クリーンなエネルギー技術に関わる研究は世界的に行われており、特に燃料電池や太陽電池はそうしたクリーンエネルギー領域を代表する次世代電池として着目されている。

我が国においても経済産業省が新国家エネルギー戦略の中で、運輸エネルギーの次世代化や新エネルギーイノベーションを掲げるなど、新経済成長戦略において次世代電池をイノベーション戦略の柱として位置づけている。

同時に、これらのエネルギー問題は世界共通の課題であることから、国際的な共同研究が世界各地で行われている。その中で、我が国が国際的潮流からはずれることは、国際競争力の観点からも好ましくない。次世代電池分野における競争優位を得るために、世界中の機関における次世代電池研究へのコミットメントを把握し、また世界的な共同研究状況を俯瞰することで、世界における我が国の位置づけを明確にする必要がある。

## 2 目的

燃料電池と太陽電池の二つの研究領域において、研究実績の上位機関の特定および、共同研究関係の強い組合せを客観的に特定する。国際研究ネットワーク内における日本の位置づけを明確にすることで、グリーンイノベーションに関わる研究開発支援策や国際提携戦略の立案に貢献することを目的とする。

## 3 手法

### 3.1 分析方針

使用するデータベースには®Thomson

Reuter社が提供する学術論文データベース Thomson isi web of knowledge®を用いる。

対象とする研究分野における、当該研究機関の研究実績としての研究実績の指標には、著者データ内に出現する延著者数を用いる。延著者数上位100機関のみを分析対象とする。

また、データベース内に収録されている著者所属研究機関情報には、それぞれの機関名のみならず、研究センター名や研究所名のような詳細な組織レベルで記載されている。ただし、これら詳細な組織名称は論文ごとに記載レベルが異なることや、略称や通称を用いられることもあり、データ間での表記ゆれが生じる確率が上がる。よって本稿ではこれらのセンター名や研究所名を除外し、大学名や機関名として呼称される最上位の組織名を分析の最小単位として考える。

なお、これら最上位の組織名を分析の最小単位として利用する場合、多国籍にまたがり存在する機関の扱いを明確にする必要がある。すなわちこれらの扱いは、当該研究機関の代表的な国と考えられる国籍に属するものとして算出する。

以上より、本稿における「機関間共同研究」の定義とは、「最上位の組織名が異なる機関に所属する著者同士の共著によって論文とされた研究」である。

### 3.2 分析手順

本分析の手法は、次の7ステップで実施した。

ステップ1:データベースより燃料電池と太陽電池に関わる学術論文を抽出する。1945年から2009年の間において、タイトルもしくはアブストラクトに”fuel cell”を含む論文を燃料電池に

関する研究論文であるとした。同様に、タイトルもしくはアブストラクトに”photo voltaic”もしくは”solar cell”を含む論文を太陽電池に関する研究論文であるとした。

ステップ2: 著者情報（筆頭著者および全共著者）の所属機関のうち、最上位の機関名のみ抽出する（すなわち、センター名等を除く）。

ステップ3: 抽出されたデータ中において、同一研究機関名の出現数を計算（筆頭著者、共著者の区別なく、また重複の区別なく、存在すれば1つとして計算）し、すべての研究機関にわたって実施する。これを延著者数とし、当該研究分野の研究実績を表す指標とする。

ステップ4: 燃料電池分野にいたっては延著者数上位101機関のみを対象、太陽電池分野では上位100機関のみを対象とし、それより下は足切りする。

ステップ5: 上位100機関同士において、延共著者数を計算（筆頭著者、共著者の区別なく、また重複著者の区別なくカウント）し、共同研究実績とする。

ステップ6: 延著者数を円（ノード）のサイズ、延共著者数を線（リンク）の太さとしたネットワーク図によって可視化し、主所属国同士でまとまるように整理する。

ステップ7: 同一機関内での共同研究（セルフープ）を削除する。

## 4 結果

### 4.1 燃料電池共同研究

1945年から2009年の燃料電池に関する研究論文を、データベースより抽出した結果、33,722件の論文を抽出することができた。（ステップ1）

上位101機関のみを抽出した結果、その主所属国は24カ国に渡り、459の機関間共同研究関

係が抽出された。（ステップ4,5）

上位国同士の機関間共同研究ネットワーク図を図1に示す。機関名のついている円（ノード）の大きさが延著者数に比例して対応しており、すなわちこれが燃料電池の研究（共同研究に限らずない）の実績を意味する。また、機関間を結ぶ線（リンク）の太さが、両端の機関間の共同研究実績を意味する。括弧内数字は、その国の所属する機関数である。（ステップ6,7）

なお、図1のうち、延著者数上位30の機関と、当該期間の共同研究相手先機関数を表1に示す。また、図1のうち、機関間共同研究組合せ上位30ペアを表2に示す。

### 4.2 太陽電池共同研究

1945年から2009年の太陽電池に関する研究論文を、データベースより抽出した結果、43,903件の論文を抽出することができた。（ステップ1）上位100機関のみを抽出した結果、その主所属国は23カ国に渡り、451の機関間共同研究関係が抽出された。（ステップ4,5）

上位国同士の機関間共同研究ネットワーク図を図2に示す。機関名のついている円（ノード）の大きさが延著者数に比例して対応しており、すなわちこれが太陽電池の研究（共同研究に限らずない）の実績を意味する。また、機関間を結ぶ線（リンク）の太さが、両端の機関間の共同研究実績を意味する。括弧内数字は、その国の所属する機関数である。（ステップ6,7）（ステップ6,7）

なお、図2のうち、延著者数上位30の機関と、当該期間の共同研究相手先機関数を表1に示す。また、図2のうち、機関間共同研究組合せ上位30ペアを表4として示した。

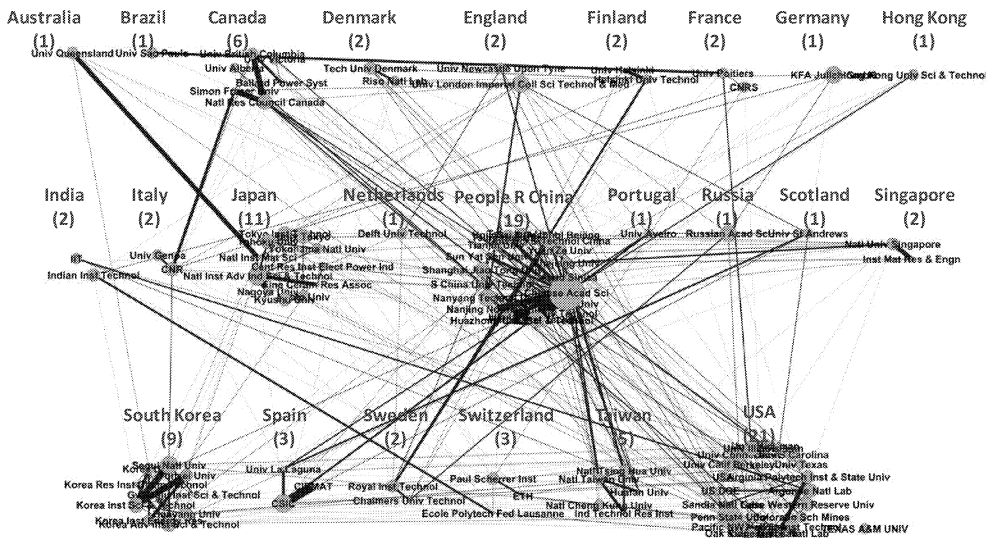


図1 燃料電池共同研究ネットワーク (延著者数上位101機関間)

表1 燃料電池研究 延著者数順上位30機関

| Rank | 機関名                                         | 所属国             | 延著者数 | 共同相手先機関数 |
|------|---------------------------------------------|-----------------|------|----------|
| 1st  | Chinese Acad Sci                            | Peoples R China | 1246 | 33       |
| 2nd  | Penn State Univ                             | USA             | 590  | 21       |
| 3rd  | Natl Inst Adv Ind Sci & Technol             | Japan           | 377  | 14       |
| 4th  | CSIC                                        | Spain           | 342  | 11       |
| 5th  | Seoul Natl Univ                             | South Korea     | 330  | 14       |
| 6th  | Harbin Inst Technol                         | Peoples R China | 313  | 20       |
| 7th  | Nanyang Technol Univ                        | Peoples R China | 307  | 12       |
| 8th  | Univ London Imperial Coll Sci Technol & Med | England         | 304  | 15       |
| 9th  | KFA Juelich GmbH                            | Germany         | 299  | 7        |
| 10th | Russian Acad Sci                            | Russia          | 272  | 9        |
| 11th | Univ Connecticut                            | USA             | 270  | 9        |
| 12th | Univ Sci & Technol China                    | Peoples R China | 263  | 11       |
| 13th | Univ Illinois                               | USA             | 257  | 14       |
| 14th | Natl Res Council Canada                     | Canada          | 251  | 16       |
| 15th | Korea Inst Sci & Technol                    | South Korea     | 251  | 9        |
| 16th | Pacific NW Natl Lab                         | USA             | 248  | 15       |
| 17th | Shanghai Jiao Tong Univ                     | Peoples R China | 241  | 10       |
| 18th | Georgia Inst Technol                        | USA             | 238  | 13       |
| 19th | Korea Adv Inst Sci & Technol                | South Korea     | 228  | 16       |
| 20th | Tohoku Univ                                 | Japan           | 221  | 12       |
| 21st | Tokyo Inst Technol                          | Japan           | 215  | 12       |
| 22nd | Argonne Natl Lab                            | USA             | 211  | 12       |
| 23rd | Univ Tokyo                                  | Japan           | 209  | 11       |
| 24th | Tsing Hua Univ                              | Peoples R China | 208  | 10       |
| 25th | Jilin Univ                                  | Peoples R China | 206  | 10       |
| 26th | Univ Calif Berkeley                         | USA             | 202  | 11       |
| 27th | CNR                                         | Canada          | 202  | 14       |
| 28th | Univ S Carolina                             | USA             | 198  | 7        |
| 29th | Kyushu Univ                                 | Japan           | 195  | 3        |
| 30th | Yuan Ze Univ                                | Peoples R China | 193  | 8        |

表2 燃料電池共同研究 機関組合せ上位30ペア

| 機関名                                 | 主所属国            | 延共同研究者数 | 機関名                      | 主所属国            |
|-------------------------------------|-----------------|---------|--------------------------|-----------------|
| Jilin Univ                          | Peoples R China | 44      | Harbin Inst Technol      | Peoples R China |
| Harbin Inst Technol                 | Peoples R China | 33      | Acad Sinica              | Peoples R China |
| CSIC                                | Spain           | 33      | CIEMAT                   | Spain           |
| Chinese Acad Sci                    | Peoples R China | 32      | Nanjing Normal Univ      | Peoples R China |
| Seoul Natl Univ                     | South Korea     | 28      | Korea Inst Sci & Technol | South Korea     |
| Korea Inst Sci & Technol            | South Korea     | 27      | Yonsei Univ              | South Korea     |
| Natl Res Council Canada             | Canada          | 27      | Univ British Columbia    | Canada          |
| Chinese Acad Sci                    | Peoples R China | 27      | Dalian Univ Technol      | Peoples R China |
| Natl Inst Adv Ind Sci & Technol     | Japan           | 23      | Nagoya Univ              | Japan           |
| Natl Res Council Canada             | Canada          | 23      | Simon Fraser Univ        | Canada          |
| Natl Inst Mat Sci                   | China           | 22      | Univ Queensland          | Australia       |
| Natl Univ Singapore                 | Singapore       | 21      | Inst Mat Res & Engn      | Singapore       |
| Virginia Polytech Inst & State Univ | USA             | 19      | Los Alamos Natl Lab      | USA             |
| Chinese Acad Sci                    | Peoples R China | 18      | Jilin Univ               | Peoples R China |
| Tsing Hua Univ                      | Peoples R China | 18      | Chinese Acad Sci         | Peoples R China |
| Helinki Univ Technol                | Finland         | 16      | Univ Helsinki            | Finland         |
| Natl Inst Adv Ind Sci & Technol     | Japan           | 15      | Eme Ceram Res Assoc      | Japan           |
| Korea Inst Sci & Technol            | South Korea     | 15      | Korea Univ               | South Korea     |
| Chinese Acad Sci                    | Peoples R China | 15      | Sun Yat Sen Univ         | Peoples R China |
| Ind Technol Res Inst                | Taiwan          | 15      | Natl Taiwan Univ         | Taiwan          |
| Korea Inst Energy Res               | South Korea     | 14      | Nysei Univ               | South Korea     |
| Simon Fraser Univ                   | Canada          | 13      | CNR                      | Canada          |
| Univ La Laguna                      | Spain           | 13      | CSIC                     | Spain           |
| Huafan Univ                         | Taiwan          | 12      | Natl Taiwan Univ         | Taiwan          |
| Univ Sao Paulo                      | BRAZIL          | 12      | Univ Poitiers            | France          |
| Ballard Power Syst                  | Canada          | 12      | Univ British Columbia    | Canada          |
| Huazhong Univ Sci & Technol         | Peoples R China | 12      | Nanyang Technol Univ     | Peoples R China |
| Natl Tsing Hua Univ                 | Taiwan          | 11      | Yuan Ze Univ             | Peoples R China |
| Tech Univ Denmark                   | Denmark         | 11      | Riso Natl Lab            | Denmark         |
| Chalmers Univ Technol               | Sweden          | 11      | Royal Inst Technol       | Sweden          |

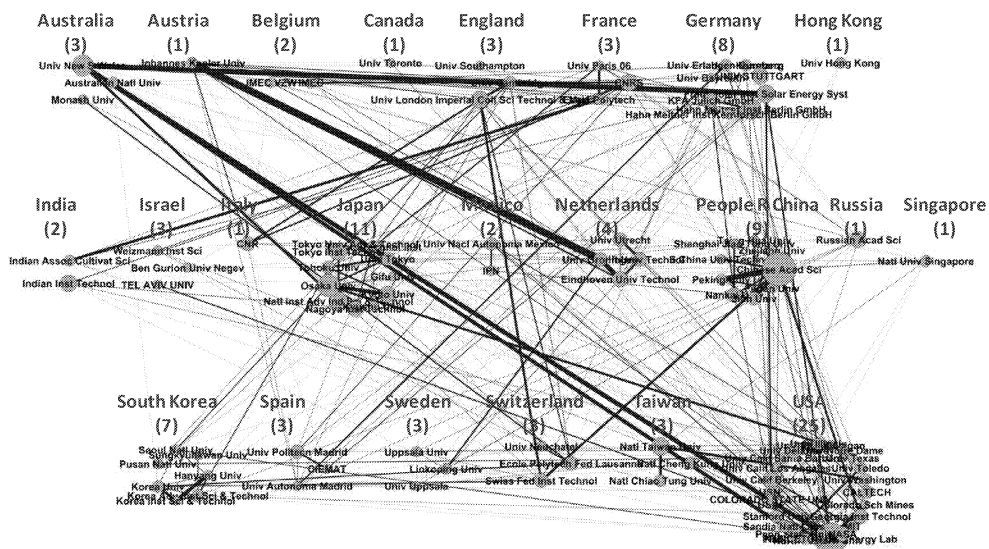


図2 太陽電池共同研究ネットワーク (延著者数上位100機関間)

表3 太陽電池研究 延著者数順上位30機関

| Rank | 機関名                                         | 所属国             | 延著者数 | 共同相手先機関数 |
|------|---------------------------------------------|-----------------|------|----------|
| 1st  | Chinese Acad Sci                            | Peoples R China | 1141 | 24       |
| 2nd  | Natl Renewable Energy Lab                   | USA             | 1046 | 34       |
| 3rd  | Osaka Univ                                  | Japan           | 531  | 13       |
| 4th  | Univ New S Wales                            | Australia       | 467  | 13       |
| 5th  | Hahn Meitner Inst Berlin GmbH               | Germany         | 445  | 15       |
| 6th  | Natl Inst Adv Ind Sci & Technol             | Japan           | 380  | 12       |
| 7th  | Natl Taiwan Univ                            | Taiwan          | 348  | 9        |
| 8th  | Fraunhofer Inst Solar Energy Syst           | Germany         | 332  | 17       |
| 9th  | Univ London Imperial Coll Sci Technol & Med | England         | 331  | 19       |
| 10th | Tokyo Inst Technol                          | Japan           | 314  | 10       |
| 11th | CNRS                                        | France          | 311  | 18       |
| 12th | Univ Calif Berkeley                         | USA             | 310  | 15       |
| 13th | Penn State Univ                             | USA             | 303  | 8        |
| 14th | UNIV STUTTGART                              | Germany         | 301  | 1        |
| 15th | Indian Inst Technol                         | India           | 299  | 7        |
| 16th | Kyoto Univ                                  | Japan           | 283  | 13       |
| 17th | Russian Acad Sci                            | Russia          | 275  | 17       |
| 18th | Univ Delaware                               | USA             | 272  | 12       |
| 19th | Eindhoven Univ Technol                      | Netherlands     | 269  | 15       |
| 20th | Naseva Inst Technol                         | Peoples R China | 264  | 4        |
| 21st | NASA Univ                                   | Peoples R China | 247  | 7        |
| 22nd | NASA                                        | USA             | 240  | 9        |
| 23rd | Georgia Inst Technol                        | USA             | 238  | 11       |
| 24th | KFA Julich GmbH                             | Germany         | 227  | 15       |
| 25th | CNR                                         | Italy           | 227  | 12       |
| 26th | Univ Utrecht                                | Netherlands     | 219  | 10       |
| 27th | Swiss Fed Inst Technol                      | Switzerland     | 218  | 15       |
| 28th | Univ Tokyo                                  | Japan           | 213  | 13       |
| 29th | Univ Groningen                              | Netherlands     | 211  | 11       |
| 30th | Univ Politecn Madrid                        | Spain           | 211  | 7        |

表4 太陽電池共同研究 機関組合せ上位30ペア

| 機関名                                         | 主所属国            | 延共同研究者数 | 機関名                                         | 主所属国            |
|---------------------------------------------|-----------------|---------|---------------------------------------------|-----------------|
| Johannes Kepler Univ                        | Austria         | 36      | Univ Groningen                              | Netherlands     |
| Univ New S Wales                            | Australia       | 26      | Fraunhofer Inst Solar Energy Syst           | Germany         |
| Peking Univ                                 | Peoples R China | 26      | Chinese Acad Sci                            | Peoples R China |
| Natl Renewable Energy Lab                   | USA             | 25      | Colorado Sch Mines                          | USA             |
| Univ New S Wales                            | Australia       | 23      | Natl Renewable Energy Lab                   | USA             |
| Eindhoven Univ Technol                      | Netherlands     | 23      | Univ Groningen                              | Netherlands     |
| Univ New S Wales                            | Australia       | 19      | Sandia Natl Labs                            | USA             |
| Osaka Univ                                  | Japan           | 19      | Kyoto Univ                                  | Japan           |
| Hanyang Univ                                | South Korea     | 18      | Korea Inst Sci & Technol                    | South Korea     |
| Chinese Acad Sci                            | Peoples R China | 17      | Jilin Univ                                  | Peoples R China |
| Natl Renewable Energy Lab                   | USA             | 16      | Sandia Natl Labs                            | USA             |
| Korea Univ                                  | South Korea     | 14      | Korea Inst Sci & Technol                    | South Korea     |
| Ecole Polytech                              | France          | 14      | Indian Assoc Cultivat Sci                   | India           |
| Univ New S Wales                            | Australia       | 13      | Natl Inst Adv Ind Sci & Technol             | Japan           |
| Natl Chiao Tung Univ                        | Taiwan          | 13      | Natl Taiwan Univ                            | Taiwan          |
| Osaka Univ                                  | Japan           | 12      | Univ Notre Dame                             | USA             |
| Chinese Acad Sci                            | Peoples R China | 12      | Swiss Fed Inst Technol                      | Switzerland     |
| Univ London Imperial Coll Sci Technol & Med | England         | 12      | Swiss Fed Inst Technol                      | Switzerland     |
| Univ Cambridge                              | England         | 12      | Univ London Imperial Coll Sci Technol & Med | England         |
| Natl Taiwan Univ                            | Taiwan          | 12      | Univ Washington                             | USA             |
| Eindhoven Univ Technol                      | Netherlands     | 11      | Univ Utrecht                                | Netherlands     |
| Univ Calif Berkeley                         | USA             | 11      | Stanford Univ                               | USA             |
| Chinese Acad Sci                            | Peoples R China | 11      | S. China Univ Technol                       | Peoples R China |
| Chinese Acad Sci                            | Peoples R China | 11      | Georgia Inst Technol                        | USA             |
| Tokyo Univ                                  | Japan           | 11      | Univ Tokyo                                  | Japan           |
| Univ Paris 06                               | France          | 11      | Ecole Polytech                              | France          |
| Jilin Univ                                  | Peoples R China | 10      | Zhejiang Univ                               | Peoples R China |
| IMEC                                        | Belgium         | 10      | CNRS                                        | France          |
| Delft Univ Technol                          | Netherlands     | 10      | Univ Utrecht                                | Netherlands     |
| Shanghai Jiao Tong Univ                     | Peoples R China | 10      | Nankai Univ                                 | Peoples R China |

## 5 考察

国籍についてみると、燃料電池研究では研究能力上位と共同研究機関ペア上位の国に大きな差異はなく、いずれも中国と米国の存在感が強い。一方で、太陽電池研究においては、研究能力上位は中国、米国が強いが、共同研究機関ペア上位には欧州勢の連携が目立つ。

それぞれの研究領域の共同研究機関ペア上位(表2と表4)を比較すると、燃料電池の共同研究においての上位30ペアは殆どが同国同士であるのに対し、太陽電池の共同研究においては、異なる国の機関間共同研究が多い。燃料電池分野は各国とも依然として同国内における共同研究の割合が強く、太陽電池分野では、国際的な連携が進んでいるものと考えられる。

地理的にみると、欧米又は中国他と欧米間の提携に比べ、アジア内での研究協力が少ないことがわかる。

日本については、いずれの研究分野でも国内機関同士で共同する傾向が強い。次世代電池分野を含むグリーンテクノロジーは、多数の国々で研究が活発化しており、また、世界共通の課題に対応する、すなわち、市場が世界的に広がっている領域であることから、提携関係が国内に閉じていることは、研究開発競争上又は市場獲得上、不利であると考えられる。中国の研究能力が高まり、また、欧米との国際提携を進みつつあることを踏まえると、将来、我が国が孤立してしまうリスクも存在する。

新技術の開発・導入による地球環境問題の早期解決と我が国の経済成長の両面から、国際的な提携を強化する必要があると考えられる。特に、アジア域内における国際研究協力の強化が大きな課題である。

## 6 まとめ

本研究では、延著者数を当該研究機関の研究実績として、延共著数を機関同士の共同研究の強さとして、それぞれネットワーク上でノードとリンクで表し、国際的な共同研究状況を俯瞰した。

燃料電池と太陽電池の双方を比較することで、研究分野によって、明らかに国際性の違いが存在することが判明した。同時に、特に我が国の次世代電池の状況として、研究実績は高いが、国際的な共同研究につながっていないことが明らかになった。アジアの中でも孤立していることが見て取れた。

今後は、著者の記載順に重みをつける、などによってより研究成果への貢献度を加味することや、過去の時点との比較を行うことによって急激に能力を高めついている機関等の特定をすることで、より具体的な構造把握が可能となると考える。

## 参考文献

- 経済産業省, “次世代自動車用電池の将来に向けた提言”, 2006.8, 新世代自動車の基礎となる次世代電池技術に関する研究会,  
<http://www.meti.go.jp/report/downloadfiles/g60824b01j.pdf>
- 経済産業省, “新国家エネルギー戦略について”, 2006.5, 資源エネルギー庁,  
<http://www.meti.go.jp/press/20060531004/senryaku-p.r-set.pdf>
- 坂田一郎/佐々木一 “サービス・イノベーション学術研究に関する世界共同研究ネットワークの可視化”, 2009.12, 内閣府サービス・イノベーション政策に関する国際共同研究



# 著者名典拠情報を拡充するための共同編集プラットフォーム Cooperative Database Editor for expanding Japan MARC/A

川島隆徳<sup>1\*</sup>, 研谷紀夫<sup>2</sup>  
Takanori KAWASHIMA<sup>1\*</sup>, Norio TOGIYA<sup>2</sup>

1 東京工業大学大学院社会理工学研究科価値システム専攻

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

〒152-8552 東京都目黒区大岡山 2-12-1

E-mail: t\_kawa@valdes.titech.ac.jp

2 東京大学大学院情報学環

Graduate School of Interfaculty Initiative in Information Studies

〒113-0033 東京都文京区本郷7-3-1

E-mail: qq26212@iii.u-tokyo.ac.jp

本研究では, 国立国会図書館の著者名典拠情報を拡充し, 有効利用するための典拠情報共同編集プラットフォームを開発した. このプラットフォームはJAPAN/MARC(A)を読み込み, ネットワークを通じて複数の識者が足りない情報を追加することが可能である. また, 必要に応じてMARCで定義されていない情報も追加することや, MADS及びMODS形式やその他の構造化記述言語で出力することも可能である. これらより, ネットワークに対応したより高度な文化情報資源の利用のために典拠情報を活用していくことが可能となる.

In this study, we developed a platform for both expanding and utilizing name authority files of the National Diet Library. This software read JAPAN/MARC(A) format and enables many experts to complement lacked elements of name authority files through the Internet. It is also possible to add element which is not defined in MARC format and output name authority data to MADS/MODS format. With this platform, name authority files can be utilized for exploitation of advanced cultural information resources.

キーワード: 人名典拠情報, 共同構築, MADS/MODS, JAPAN/MARC(A)

Name Authority File, Collaborative, MADS/MODS, JAPAN/MARC(A)

## 1 はじめに

本論文で主な研究対象とする人名典拠情報は図書館、文書館、博物館でその概要が異なるが、主に「特定の人物に関して、その存在の同一性を確立すること」を目的としている。

典拠情報は、これまで主に図書館分野における典拠情報の構築が最も多くの実績を残している。共に英語圏にある米国議会図書館のLC Name Authority File : LCNAFと、British LibraryのBritish Library Name Authority List : BLNAに関しては早くから統合の試みが行われており、Anglo-American Authority File project : AAFAFとして統合的運用が行われている。またこれらと連携するプログラムにおいては、データの共同構築作業などが進行している[1]。

一方で、Online Computer Library Center : OCLC、米国議会図書館、ドイツ図書館の三者によってVirtual International Authority File:VIAFを作成するプロジェクトも2003年より開始されている[2]。

東アジア地域においても、日本語、中国語、韓国語の名前典拠を連携させるワークショップが2001年から2002年にかけて開催された[3]。

このように、これまで各国の特定の機関で作成されていた典拠情報は、複数の機関間の連携に基づいて構築され、さらに構造的な記述言語で出力され、データを共有していく動向が顕著になると考えられる。

本研究では、こうした動向をうけて、典拠情報の共有、共同構築、典拠情報の構造的な記述言語での出力を可能とするデータベースシステム「Shareword」を開発することとした。

既存の類似システムとしては Open

Context[4]が存在する。Open Context は、考古学における史料のための Web ベースのシステムである。このシステムは史料のデータベースを保持し、ネットワークを通じて共同で管理編集する機能を持つ。その上で、史料に単一の ID を振ることで書籍や論文などの典拠情報として利用することを目的としたものである。

Web を通じて共同でデータを編集することを目的とした汎用システムとしては、wiki やその拡張である Semantic MediaWiki[5]等が存在する。しかし、典拠情報を目的としたデータベースの場合には、そのデータの出自に厳密性が求められるため、編集者の明示やデータの出典を記述するためのデータ構造などが必須となる。Shareword はこのための機能を備えることで、典拠情報など人文科学用のデータベースの構築を容易にしている。

本研究では Shareword のケーススタディとして、JAPAN/MARC(A)フォーマットの典拠情報を登録し、データを編集して、Mads 及び Mods へと出力する実験を行った。

## 2 Shareword

### 2.1 システム構成

Sharewordはデータベースを共有・共同編集するためのサーバクライアント型アプリケーションである。システムの構成は図1のようになっている。

本システムでは、一つのデータベースをインターネットを介して複数人で共同編集する。各ユーザはGUIを通じてデータを編集し、サーバのデータベースと同期する

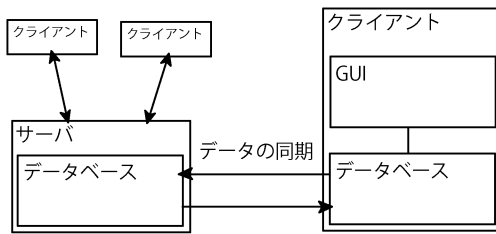


図1 Shareword のシステム構成

ことで他の編集者のデータを共有することができる。

## 2.2 データ構造

人名典拠情報においては、「同名異人」や、「異名同人」を統制する必要がある。Shareword では全てのデータに ID を割り振り、ID レベルでデータを管理することで、名前などの文字列に依存せず人物等のデータを識別することを可能にしている。

Shareword のデータ構造は、データの構造を定義するクラスと、具体的なデータを現すインスタンスを主とする。

クラスとは、例えば「人物」や「組織」のような抽象的な概念を表す。そしてインスタンスは、それぞれのクラスに所属する具体的な人物や組織を現す。

クラスには、エレメントタイプと呼ばれる属性を設定できる。これはクラスそのものに情報を付け足すものではなく、クラスに所属する各インスタンスが持つべき属性（エレメント）を規定するものである。例えば、「著者」クラスに「著作」というエレメントタイプを付与すれば、そのインスタンスである「夏目漱石」の著作として「夢十夜」というエレメントを設定することが可能になる。

エレメントには、リテラルとノンリテラルの二種類がある。リテラルエレメントとは、文字列、数値、URI、日付、真

偽値であり、具体的な値を持つ。一方でノンリテラルエレメントは、他のインスタンスへの参照を指す。これによって、エレメントにさらにエレメントを設定することが可能になる。例えば、「夢十夜」を文字列のリテラルエレメントとした場合には、それはただの「夢十夜」という文字列でしかないが、ノンリテラルエレメントとした場合には、固有の ID を持った「夢十夜」というインスタンスとして扱われ、「夢十夜」インスタンスが所属するクラスに即して「刊行年」などのエレメントを設定することが可能になる。このような仕組みを実現するために、エレメントタイプは、実際には別のクラスへの参照となっている。なお、エレメントは同じエレメントタイプに複数設定することが可能であり、例えば複数の著作を設定することができる。

同じクラスのインスタンス同士の間には、コネクションという関係を作ることができる（例：人物 A と人物 B の間に「友人」関係を設定）。コネクションも、クラスによって利用できるタイプを自由に創造することができる。

典拠情報として利用するために重要なのは、データの出自を記述することである。また、そのデータが真となる条件など、メタデータを記述したい場合もある。エレメントとコネクションに追加できるコンテキストという形式がこの役目を担っている。コンテキストの例としては、人物の略歴の出典情報や、人物同士の友好関係が続いていた期間などが挙げられる。コンテキストは実際には他のインスタンスへの参照なので、エレメントを使ってコンテキストに説明を加えることも

可能である。

## 2.3 機能

### 2.3.1 クラス構造の設計

SharewordのGUIには現在4種類のモードが存在し、その一つ、設計モードでクラスの構造を設計する。クラス構造とは、クラスと、そのクラスが持つエレメントやコネクション、そしてそれらが持つコンテキストを指す。一度クラスを作った後でもエレメントは自由に増やすことが可能であり、また他の編集者が作ったクラスに追加することもできる。ただし、設計モードが利用できるのは権限をもった編集者だけであり、他のユーザは準備されたクラス構造に従ってデータをインプットしていくことしかできない。

### 2.3.2 データの編集

編集機能には、記述、関連の2モードが存在する。

記述モード(図2)はインスタンスを作成し、そのエレメントを設定する。画面上で一つのインスタンスは一枚のカードとして表示され、エレメントを設定するためのWebフォームのようなインターフェイスが配されている。リテラルエレメントは、それぞれのデータ形式に対応したフォームが用意されている。ノンリテラルエレメントの場合、文字列を入力して新しいインスタンスを参照するエレメントを作り出す他、一覧や省入力候補から既存のエレメントを選択して編集することも可能である。また、これらのノンリテラルエレメントの表示をクリックすることで、そのエレメントの実体であるインスタンスのカードが展開され、そのエレメントが編集できる。同様の画面で、コンテキストも編集できる。

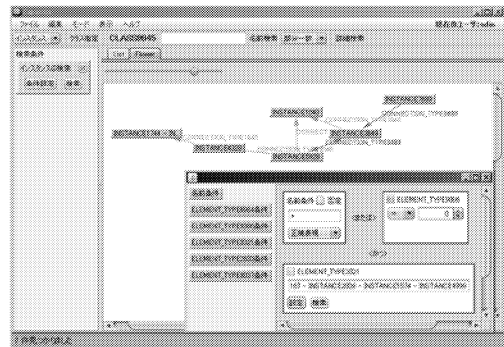


図2 記述モード

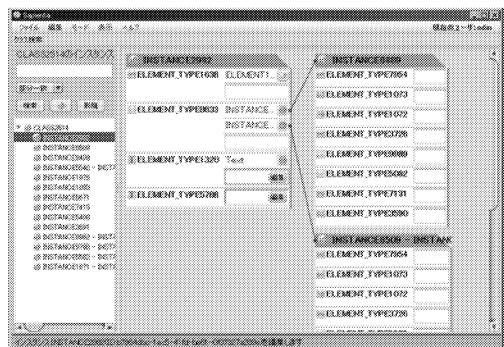


図3 検索モード

関連モードではインスタンスを検索し、画面上に表示されたインスタンス同士の間にマウスを使って線を引くことで、コネクションを作ることができる。

### 2.3.3 協調編集

Shareword のデータ編集はユーザ登録し、システムにログインしないと行えない。また、全ての編集作業はどのユーザが行ったかの情報を持つため、全てのデータはその編集責任の所在が明らかである。データの編集作業は、大きく分けて追加、更新、消去の3種類がある。追加に関しては制限がなく、例えば他のユーザが登録したインスタンスにエレメントを追加することも可能である。しかし、更新と消去は最初に登録したユーザ以外が行うことはできないため、登録したデ

表1 JAPAN/MARC(A), Shareword, MADS/MODSの対応

| JAPAN/MARC(A)            | Shareword            | MADS + MODS                                           |
|--------------------------|----------------------|-------------------------------------------------------|
| 001:レコード識別番号:            | 識別番号[リテラル・整数]        | <Mods:recordIdentifier>                               |
| 005:レコード更新情報             | 更新情報[リテラル・日付]        | <Mods:recordChangeDate>                               |
| 100:一般的処理データ:ファイルに入れた日付  | 一般処理データ[ノンリテラル]      | <Mods:recordCreationDate>                             |
| 100:一般的処理データ:目録用言語コード    | 一般処理データ[ノンリテラル]      | <Mods:languageTerm>                                   |
| 152:目録規則                 | 目録規則[ノンリテラル]         | <Mods:descriptionStandard>                            |
| 200:標目一個人名               | 名前[ノンリテラル]           | <authority><name type="personal" ID="xx">             |
| 300:名称に関する注記             | 名称に関する注記[ノンリテラル]     | <note xml:lang="xx">                                  |
| 400:「を見よ」参照指示一個人名        | 名前[ノンリテラル]           | <variant type="other"> <name type="personal" ID="xx"> |
| 801:レコード作成機関             | レコード作成機関[ノンリテラル]     | <Mods:recordInfo>                                     |
| 810:出典データ                | 出典データ[ノンリテラル]        | <note type="resource" xml:lang="xx">                  |
| 830:目録作成者一般注記            | 一般注記[リテラル・文字列]       | <note type="occupation" xml:lang="xx">                |
| 936:票目訂正に関する注記(その他の 900) | 票目訂正に関する注記[リテラル・文字列] | <Mods:extension>                                      |
| -                        | 関連 URL[リテラル・URI]     | <Mods:url>                                            |

ータを他の編集者に消されることは無い。

このように協調的な仕組みであるため、本システムではデータの真偽性に関する争いを解決するための機能としてコメント機能と投票機能を用意している。前者は、各データに掲示板のようにコメントをつけられる機能であり、後者はデータに賛成・反対の票を投じることができる機能である。

### 2.3.4 検索

検索モードではインスタンスの検索が行える。インスタンスが持つエレメントを「AND」や「OR」で組み合わせた複雑なクエリを、GUIのクエリビルダ(図3)で生成することが可能である。

検索結果は単純なリスト形式だけでなく、データ同士の関係を把握することに適したネットワーク形式で表示することができる。

### 2.3.5 インポート／エクスポート

本システムのデータ形式は独自形式であるため、現時点では汎用的なインポート機能は存在しない。本研究では、JAPAN/MARC(A)をインポートするためのプログラムを別個に作成した。

エクスポート機能に関しては、汎用の構造化形式であるRDFおよびTopic Mapへの出力に対応している。また、本研究ではMads形式への出力プログラムを別個に作成した。

## 3 JAPAN/MARC(A)のマッピング

### 3.1 MARCからSharewordへ

MARCのサンプルデータとしては、JAPAN/MARC(A)形式の国立国会図書館著者名典拠録を利用した。MARCの各データブロックのうち、データが入力されたのは表1の第一列、「JAPAN/MARC(A)」の項目である。各項目は、表1の第二列にあるように、Shareword上のデータ構造として格納された。このうち、ノンリテラルであるエレメントタイプには、その実体クラスにMARCのサブフィールドに対応するエレメントタイプを設定してある(図4)

また、MARCの各サブフィールドの指示子は、エレメントのコンテキストとして情報を格納した。これによって、MARCのデータは過不足無くSharewordに登録することができる。

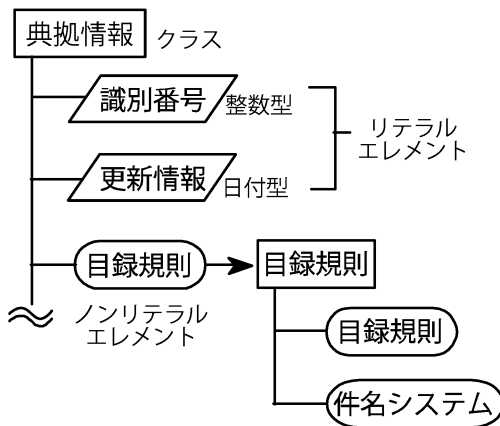


図 4 Shareword 上でのデータ構造の一部

さらに、Madsにのみ存在する要素、「URL」もエレメントタイプとして追加した。

### 3.2 SharewordからMads及びModsへ

Shareword上のデータをMadsへ変換するためには、既存のMARCからMads及びModsに変換するための手順[6]を利用し、表1の第3列のように変換した。

## 4 今後の展望

### 4.1 他種のデータへの適用

本研究ではMARCをデータとして利用したが、Sharewordは他の典拠情報や、それ以外のデータベースとしても利用可能である。他のデータに利用することで、システムの有効性を検討する必要がある。

### 4.2 システムの拡張

インポート機能、画像データの扱いなど、不十分な機能を実装する必要がある。また、登録したデータを検索するだけでなく、様々な形式で表示・処理するための機能等、構築したデータを有効活用するための仕組みも必要であると考えられる。

## 謝辞

本研究においては国立国会図書館ご協力を頂きました。また、本研究は、「科学研究費補助金 若手B ネットワーク文化情報資源で活用する人名典拠情報に関する研究」によって行われました。

## 参考文献

- [1] 三浦敬子; 松井幸子:「欧米における著者名典拠ファイルの共同作成の動向」,日本図書館情報学会誌, Vol.47, No.1, pp.29-41, 2001
- [2] Bennett, Rick. et al:「バーチャル国際典拠ファイル(VIAF: Virtual International Authority File): ドイツ図書館と米国議会図書館の典拠ファイルを利用して」, World Library and Information Congress: 72nd IFLA General Conference and Council. Seoul. [http://www.ifla.org/IV/ifla72/papers/123-Bennett\\_trans-ja.pdf](http://www.ifla.org/IV/ifla72/papers/123-Bennett_trans-ja.pdf) (参照 2010-03-30).
- [3] 国立情報学研究所:日本語, 中国語, 韓国語の名前典拠ワークショップ記録, <http://www.nii.ac.jp/publications/CJK-WS3/> (参照 2010-03-30).
- [4] Kansa, Sarah Whitcher; Kansa, Eric, C.; Schultz, Jason, M.: "An Open Context for Near Eastern Archaeology", Near Eastern Archaeology, Vol70, No.4, pp.188-194, 2007
- [5] Krötzsch, Markus; Vrandečić, Denny; Völkel, Max: "Semantic MediaWiki", Proceedings of the Fifth International Semantic Web Conference, pp 935-942, 2006.
- [6] 研谷紀夫:「著者名典拠録の情報構造化記述」, デジタル図書館, No.38, 2010

## 聖教書誌情報全文検索システムのスタンドアロン化の試み

### An approach to Full-text Search for Bibliographic Data of Ancient Documents by means of Stand-alone Operation

渡上将治<sup>1\*</sup>, 村川猛彦<sup>1</sup>, 宇都宮啓吾<sup>2</sup>, 中川優<sup>1</sup>

Yukiharu WATAGAMI<sup>1\*</sup>, Takehiko MURAKAWA<sup>1</sup>,

Keigo UTUNOMIYA<sup>2</sup>, Masaru NAKAGAWA<sup>1</sup>

1 和歌山大学システム工学部

Faculty of Systems Engineering, Wakayama University

〒640-8510 和歌山県和歌山市栄谷 930

2 大阪大谷大学文学部

Faculty of Literature, Osaka Ohtani University

〒584-8540 大阪府富田林市錦織北 3-11-1

筆者らがこれまでに構築してきた聖教書誌情報全文検索システムは、サーバに接続して使用する必要があるため、山奥などでは利用できないという問題点があった。そこで、インターネット環境に依存しない、1台のノート PC 上のみで動作するスタンドアロンのシステムの構築を行った。サーバ部の OS を Linux から Windows に変更したが、対象データと全文検索エンジンは同じものを使用し、従来のシステムに存在した機能に加えて、新たなデータの追加や、各聖教に対するコメントを残すことができる機能を実装した。本システムと従来のシステムとで検索結果が同じであり、検索時間に関しても2秒以内で結果が表示されることを確認した。

It is not possible to retrieve something that one wants deep in the mountains, using the full-text search system which we have developed for bibliographic data of documents in temple, since the system needs the access to the server. In this paper, we report a new system in stand-alone mode which works with an Ultra Mobile PC without communicating to any external computer. For constructing the system, the Windows operating system is employed instead of Linux, while the bibliographic data and the search engine software are unchanged. We implemented the interfaces for record registration and annotation as well as the features which the conventional system offered. The retrieval experiments made no difference between both search systems with regard to the retrieval results, and showed that the stand-alone system displayed results within two seconds after the user gave the search term.

キーワード: 全文検索, スタンドアロン, 聖教, デジタルアーカイブ

full-text search, stand-alone, document in temple, digital archive

## 1 はじめに

人文学の研究者は寺院や博物館に所蔵されている古典資料や図書館などにおさめられている史資料や論文などを読み解くことで新たな知見を得ている。近年、これらの史資料をデジタル化したデジタルアーカイブを利用することで、研究のさらなる発展が期待されている。実際、インターネット上に存在するデジタルアーカイブを利用することで、持ち出すことができないような史資料を電子的に参照しながら、分析・検討を行うことが可能となっている。しかし、インターネットに接続できる環境がない場所では利用することができない。この問題を解決することは、場所を選ばずにデジタルアーカイブを扱えるようになり、人文学の研究に対して大いに貢献するものと思われる。本研究で扱う聖教もデジタルアーカイブが望まれているものの一つであり、寺院などに保管されている典籍や写経、系図などがこれにあたる。

そこで本研究では、インターネットに接続することなく、1台のノートPC上のみで動作することが可能な全文検索システムの構築を行った。さらに、検索によって、目的のデータを見つけることや検索速度においても利用するのに問題がないことを確認した。

## 2 既存システム

筆者らはこれまで、全文検索エンジンであるHyper Estraier[1]を利用した聖教書誌情報全文検索システムを構築してきた[2][3][4]。これを利用することでWebブラウザを介して、聖教情報の検索を瞬時に行うことができる。システムで扱う聖教データは44,135件存在し、それぞれ「書名、巻数、出典目録、所

蔵、書写時期、装幀、訓点、奥書、備考」の情報（属性）に分割してある。多くの聖教は「奥書」と呼ばれる書誌情報を有し、そこには「いつ誰が命じて誰が書写をしたか」「経典がどこの寺院に所蔵されたか」といった情報が記載されている。そのため奥書は、聖教自体の形成・継承の問題や聖教を巡る僧侶の活動を把握するのに重要な情報である。

システムの流れを図1に示す。Hyper Estraierを利用して全文検索を行うには、あらかじめ転置インデックスを作成しておく必要がある。本研究で作成したシステムではインデックスを登録する際にドラフト文書と呼ばれるファイル形式に整形した聖教データを使うことにした。ドラフト文書とはHyper Estraierがインデックス登録の際に扱う独自のファイル形式である。自由に属性を指定することができ、これにより属性検索が可能となる。

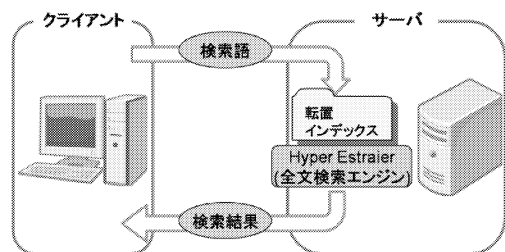


図1 既存システムを用いた検索の流れ

検索した結果は、別途データベースに登録されているキーワードがあれば、強調して表示し、クリックすることで新たな検索語として検索を行うことができるようになっている。また、聖教情報に含まれる検索語の背景色を変更することで検索語がどこに含まれているのかを瞬時に把握することができる。さらに、検索語について字体変換、例えば「弘贊法華伝」という検索語の場合、「贊」とい



う旧字体を新字体である「賛」に変換する処理や、「1234」のような3,4桁の整数値が含まれている場合、「文暦元年, 天福二年」などの該当する和暦に変換する年号変換を行うようになっている。ただし、字体変換が行われるのは検索語と全文検索用データのみであり、結果表示の際には原文のままで閲覧できる。

### 3 聖教書誌情報全文検索システムに望まれるもの

しかし、このシステムを研究者が利用するにあたっていくつかの課題があった。主要なものを以下に列挙する。(1) このシステムはサーバと通信して使用するため、インターネットに接続できない環境では利用することができない。(2) 検索に利用する聖教データに誤りがあったとしても、内容の修正を行うことができない。(3) 史資料に出現する梵字や合字などの文字を外字として適切に扱うことができない。これらの問題の中でも特に利用者の強い要望から(1)を解決することに焦点を定めた。

本システムの検索対象でもある聖教の原本は寺院や博物館などに保管され、原則的に持ち出すことができないことが多い。また、原本が所蔵されている現地にネットワーク環境が必ず整っているという保証はないためシステムを利用できない可能性が存在する。そのため、聖教の研究者が原本に関連した情報を検索するには原本が保管されているその場所で検索・調査できるような機構が必要となる。

そこで、ネットワーク環境に依存しないようなスタンドアロンでのシステムの運用に取り組むことに決めた。

### 4 スタンドアロン型聖教書誌情報全文検索システム

スタンドアロンでのシステムの運用を行うにあたり、まず何を媒体にするか検討した。条件としては持ち運びができること、素早く漏れの無い検索が利用できることが挙げられる。持ち運びができる媒体としては紙媒体やPDAなども挙げられるが、漏れの無い全文検索を行うという点から適当とは言えない。そこで、全文検索を扱え、ノートPCの中でも小型・軽量で持ち運びが容易なネットブック(ウルトラモバイルPC[5])でシステムを運用することに決めた。

続いてシステムの構築に使用した環境およびソフトウェアについて述べる。まず、OSについては利用者が扱い易いWindowsを利用することにした。また、運用環境には用意に環境を準備できるという点からWebアプリケーションの実行に必要なフリーソフトウェアをパッケージ化したXAMPPを利用することとした。これに伴い、開発言語をRubyからPHPに、DBMSをPostgreSQLからMySQLへと変更した。検索エンジンについては引き続きHyper Estraierをインストールし、使用している。なお、既存システムのサーバOSはUbuntu Linuxであり、Hyper EstraierのRubyバインディングを用いて、外部コマンドを使用することなく登録や検索を行っていたが、本システムではコマンド `estcmd.exe` をその都度実行して処理を行っている。

システムの構成を説明する。本システムはネットワークを介さずに聖教データの検索・閲覧を行うことができるWebアプリケーションであり、検索・閲覧などのシステムの動作はすべて計算機内のみで稼働するよう

になっている(図2)。本システムでは従来のシステムが持つ、検索語に対する処理や強調表示に加えて、新機能の追加や、表示方法の改善を行っている。

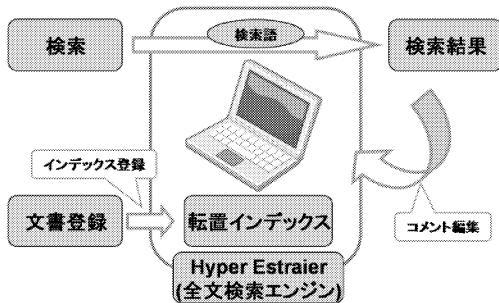


図2 スタンドアロン型システムの流れ

データベースを、2節で述べたキーワードだけでなく検索結果のテキストデータの管理にも利用している。インデックスに登録するデータは字体変換を施しているため、原文をデータベースに格納しておくことで結果表示を原文で行えるようになっている。

追加した機能のうちコメント編集機能と新規文書追加機能、表示方法の変更について説明する。「コメント編集機能」は検索により表示された各聖教データに対して、利用者がメモを残すことを可能とするものである。登録したコメントは、登録した時点から全文検索に反映されるようになっている。また、登録の際にはインデックスには字体変換を施したものを登録し、データベースには変換しないまま格納している。

「新規文書追加機能」は、本システムで新しく文書を登録するためのインタフェースである。追加するファイルはドラフト文書形式とする必要があるが、ブラウザから登録が可能になっている。登録できるのは1件ずつであり、現時点では一度に大量のファイルを登録することはできない。

最後に、表示方法の変更を行った。従来のシステムは画面解像度の比較的大きなデスクトップPCでの利用を想定したものであった(図3)。しかし、本システムはネットブック上で使用するため、従来の表示方法では非常に見にくく使いにくい。そこで、聖教データの一部を結果表示の際に隠すことにした。隠された情報はそれぞれの聖教データの行をクリックすることで見ることができる。これにより、表示がすっきりとしたものとなり、使い勝手を向上させている(図4)。

「弘誓法華伝」の検索結果

441301999ヒット

| 書名    | 巻目 | 出典目録 | 所蔵    | 書写時期 | 装幀 | 調品 | 備考                                                  | コメント                                                |
|-------|----|------|-------|------|----|----|-----------------------------------------------------|-----------------------------------------------------|
| 弘誓法華伝 | 一巻 | 法華   | 東京大学蔵 | 室町時代 | 装幀 | 紙  | 弘誓法華伝(一巻) 法華經の註釋。弘誓法華伝(一巻) 法華經の註釋。弘誓法華伝(一巻) 法華經の註釋。 | 弘誓法華伝(一巻) 法華經の註釋。弘誓法華伝(一巻) 法華經の註釋。弘誓法華伝(一巻) 法華經の註釋。 |

図3 既存システムの検索結果表示

「弘誓法華伝」の検索結果

441301999ヒット

| 書名        | 巻目 | 出典目録 | 所蔵    | 書写時期 | 装幀 | 調品 |
|-----------|----|------|-------|------|----|----|
| 弘誓法華伝     | 一巻 | 法華   | 東京大学蔵 | 室町時代 | 装幀 | 紙  |
| 弘誓法華伝(一巻) | 二巻 | 法華   | 東京大学蔵 | 室町時代 | 装幀 | 紙  |

コメント

弘誓法華伝(一巻) 法華經の註釋。弘誓法華伝(一巻) 法華經の註釋。弘誓法華伝(一巻) 法華經の註釋。

図4 本システムの検索結果表示

(上：閉じた状態、下：開いた状態)

## 5 評価と考察

本システムがスタンドアロン型でインターネット接続環境が整っていない場所でも

利用できるというだけではなく、既存のシステムと比較して検索結果が今までと変わらず信頼できるものであるかを確認するための評価実験を行った。「スタンドアロンで動作するが検索結果が誤っている」システムは信頼できず有用性を確立できない。そこで既存システムと本システムで同じ検索語を検索することで、検索結果やキーワード処理を比較し誤りが無いか調査した。また、検索時間を調べ、検索速度が利用できる範囲であるかも確認する。なお、評価するにあたり本システムを動作させる計算機はデスクトップとネットブックを用いた。環境を表1に示す。

表1 評価環境

|            | CPU                                          | メモリ |
|------------|----------------------------------------------|-----|
| ネットブック     | Intel Atom N720<br>(1.6GHz, DualCore)        | 2GB |
| デスクトップ     | Intel Pentium4<br>(3.4GHz, DualCore)         | 1GB |
| 従来システムのサーバ | AMD Athlon 64 X2 5600+<br>(2.8GHz, DualCore) | 4GB |

システムを比較するにあたり、通常の実験語10個と字体変換や年号変換の確認のための検索語3個を用いた(表2)。まず、検索ヒット数について比較してみたところ本システムの検索ヒット数は既存システムの結果といずれも一致した。字体処理を施す検索である「弘贊法華伝」に関しても全ての結果を等しく得ることができている。また、年号変換を施す検索においても全てのシステムの検索結果について同じ値となった。

表2 検索語

|        |                                              |
|--------|----------------------------------------------|
| 通常の検索語 | 東寺, 時代, 聖教, 平安<br>天保, 大和, 天保九年<br>空海, 覚厳, 硯本 |
| 字体変換あり | 弘贊法華伝                                        |
| 年号変換あり | 1000, 1234-1235                              |

次に検索速度の比較結果を記す(表3)。本システムと既存システムとの間に検索速度の差異が見られる。また、本システム間においても検索速度に差異が見られる結果となった。通常の実験語および字体変換を施す語ではデスクトップの本システムが最も時間が短く、ネットブックのものと既存システムは速度に優劣は見受けられなかった。しかし、年号変換を施す検索語においては既存システムが最も速いという結果が出た。本システムの1度目の検索時間よりも2度目の検索速度の方が総じて速くなっていることも読み取れた。既存システムでは、1度目も2度目も変わらない速度で検索が行えた。

表3 「弘贊法華伝」検索時の検索時間

|        | 検索時間(秒) |      |
|--------|---------|------|
|        | 1回目     | 2回目  |
| ネットブック | 1.80    | 1.72 |
| デスクトップ | 0.89    | 0.84 |
| 従来システム | 1.27    | 1.26 |

評価結果から考えられることを記す。本システムの検索結果はヒット数においては既存システムと比較しても同じ結果を得ることができおり、利用する際の信用性は確かであると考えられる。検索速度において本システムと既存システム間だけでなく、本システム間でも差が見られたことから単純に計算機の演算能力の違いによるものと考えられるが、ネットブックと既存システムの演算能力は1.6GHzと2.8GHzであり、この差があっても同じような検索時間を得られた事についてはシステム自体の動作速度が上昇したものと考えられる。年号変換を施した検索については本システムの検索速度が既存システムに比べて遅いが、利用に支障をきたすような遅さではないため問題はない。本シ

テムの検索速度に関して1回目よりも2回目のほうが速く、既存システムで検索速度に差が見られなかったのは、Hyper Estraierの内部構造によるものと思われる。

検索時間に関して、本システムでは通信がない一方で検索のたびに`estcmd.exe`を実行していること、またネットブックの演算能力から、懸念があったが、結果として単純な検索語に対しては1~2秒程度で結果が表示されており、十分に実用に耐えるシステムであると言える。

## 6 おわりに

本研究では、スタンドアロン型聖教全文検索システムの構築を行った。本システムを利用することで、インターネットに接続できないような環境でも、ユーザはシステムを利用することができる。しかし、本システムにはまだ改良するべき点が残されている。一つに聖教データを修正できるようにすることが挙げられる。システム開発の面から言うと、これは検索結果表示に修正へのリンクを貼っておき、修正画面へとユーザを促す機構を実装することで解決できる。しかし、修正の履歴管理や、修正後それを取り消したいという状況への対応も必要となる。筆者らは、国内古写経のための翻刻支援システムを開発しており[6]、そこではバージョン管理されたファイルに対して適切に全文検索が行える機能を試作している。これを本研究に適用することで、データ管理に関しても効率的なシステムになると考えている。

本システムを利用することにより、人文系の研究者の調査・研究が進められることを期待したい。

## 参考文献

- [1] 全文検索システム Hyper Estraier:  
<http://hyperestraier.sourceforge.net/>
- [2] 朴明哲；森本雅史；立花純児；村川猛彦；宇都宮啓吾；中川優：「人文研究を支援するデータベースシステム—聖教検索および系図表示—」，情報知識学会誌，Vol.17，No.2，pp.105-110，2007.
- [3] 村川猛彦；山中克真；宇都宮啓吾；中川優：「古典籍書誌情報におけるキーワード抽出手法」，情報知識学会誌，Vol.18，No.2，pp.87-92，2008.
- [4] 村川猛彦；宇都宮啓吾；中川優：「聖教書誌情報全文検索システム」，情報処理学会論文誌：データベース，Vol.1，No.3，pp.96-107，2008.
- [5] Intel Ultra Mobile Platform 2007,  
<http://www.intel.com/products/mid/ultramobile2007.htm> (2010年4月2日参照)
- [6] 丁敏；村川猛彦；福岡整；中川優：「Subversionを用いた仏典テキスト校訂支援システム」，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. 2008，No. 15，pp. 61-66，2008.

# EUC/EUDを前提とした文学研究システムのモデル化に関する考察

## A Modeling for the Literature Studies System assuming EUC/EUD

富澤 浩樹<sup>1\*</sup>

Hiroki TOMIZAWA<sup>1\*</sup>

1 埼玉大学大学院文化科学研究科

Graduate School of Cultural Science, Saitama University

〒338-8570 埼玉県さいたま市桜区下大久保255

情報技術の進展と普及により、利用者がEUC/EUDを実践し、情報システムに関与できるシステム環境が整ってきた。しかし現在のところ、この観点からのシステムデザインはほとんどなく、人文系データベースシステムの議論の多くもその機能面からのアプローチが中心である。本論ではこの点を人文系データベースシステムが活性化しない一因としてとらえ、文学研究者のEUC/EUDの試みに着目したシステムデザインを試みた。

Recently, as EUC/EUD has become more common, the user is ready gradually to participate in an information system. But there is almost no system design from this viewpoint in a database system for literature studies. Furthermore, a system designer almost tends to discuss only the technical aspects of a computer. In this report, we regard this point as a cause which a system does not activate. And I attempt the system design surrounding literary research activities including EUC/EUD.

キーワード: EUC/EUD, 文学研究, 情報化社会, 情報システム分析

Keywords: EUC/EUD, Literary Studies, Information Society, Information Systems Analysis

### 1 はじめに

PC(Personal Computer)や携帯電話、電子ブックリーダーなどの高性能な情報機器の普及を背景に、情報技術に長けた専門家に頼らず、利用者自らが情報システム[1]に関与する EUC(End User Computing)やEUD(End User Development)を実践することは、もはや日常となっている[2]。この社会環境の変化は、人文系データベースシステム、特に、文学研究の基礎となる手稿や原稿などの一次資料の電子的な複製を、その利用者(＝研究者)に向けて公開する取り組み(本稿ではこのデータベースシステムを研究資料デ

ータベースと呼ぶことにする)に変化を与えている。顕著なところでは、テキストデータ中心から画像データ中心へ、スタンドアロン型からクライアントサーバー型への移行がある。しかし、その利用者の多様な要求を満たすことは難しく、公開された研究資料データベースのメリットを享受していても、それぞれの視点から見た不満が常に聞こえてくる。著者の聞き取りによれば、たとえば、データの不備が訂正されない、システムマニュアルが作成されない、欲しい資料のテキストデータがないなどがよく挙げられる[3]。その一方で、そのような利用者のほとんどが、研究資料データベースの恩恵を享受

しながらも、EUC/EUD を実践することで自らの不足を補い、目的を達成している。

本稿では、そのような利用者の EUC/EUD の実践に着目する。すなわち、利用者をシステムの恩恵に享受するだけの受動的な存在ではなく、EUC/EUD を行い、積極的に貢献可能な存在としてとらえる。その上で、実際に EUC/EUD を実践した明星聖子教授(埼玉大学・ドイツ文学)の取り組みを対象にシステム分析を行う。そして、EUC/EUD の成果をどのように共有するかという観点から、文学研究システムのモデルについて考察する。

## 2 研究資料データベースと EUC/EUD

研究資料データベースの提供者は、公文書館、博物館、図書館などの公共施設や大学の付属施設、または研究者の研究グループ、そして出版社や企業などであり、成果の多くはインターネットを通して公開されている。近年では、利用者に向けてどのように提供すべきかという議論が散見されるようになってきたものの、提供者の関心の中心は、データベースの技術面・機能面にある[4]。これは、研究資料データベースの開発には、専門性と膨大な労力が必要であるためと考えられる。

しかし、たとえ利用者の要求を技術的に全て受け入れたとしても、いくつかの点で問題が生じる可能性が高い。まず、研究資料データベースのデータそのものの問題である。データベースの主要機能に検索や並べ替えがある。研究資料データベースにおいては、それら基本機能に影響を与える、転写(transcription)やタグ付け(markup)そのものが、研究対象となる場合も多い。データが直接的に基本機能に影響を与えているとなれば、様々な問題関心からシステムに接している利用者を抱えるシステムの信頼性を完全に担保することは難しい。一方で、システム開発における潜在的な問題がある。一般的に、

システム開発を巡っては、人間同士がなかなか分かりあえないという本質的な問題が関係しているため、要求定義は常にシステム開発の課題となっている[5]。つまり、提供者の取り組みだけでシステムを活性化させることは困難であると考えられる。

しかし、既存システムに不満を持っている利用者が、意識的にせよ無意識的にせよ EUC/EUD を実践し、既存システムの不足を補う役割を果たしていると考えれば、以上の問題が解消されないだろうか。利用者が主体的に情報システムに関与する EUC/EUD の意義は「データに意味を付与して情報から知恵をわき出させ行動に結びつけること[6]」ともいわれている。また、EUC/EUD においては、利用者自らが試行錯誤することで問題を克服するため、第三者が設計・開発を行うときに生じる要求定義の問題が発生しない。

そこで、次章より、EUC/EUD を前提とする文学研究システムの可能性について検討するために、明星のカフカ研究に焦点を当てる[7]。明星はフランツ・カフカの書く行為(エクリチュール)を解明する過程において EUC/EUD を実践しているためである。次章では、先ず紙メディア環境から EUC/EUD 環境に至るまでの明星のシステム環境について概観する。

## 3 対象のシステム環境

対象のシステム環境について、既に富澤と明星の先行研究[8]で示した現行システムの分析結果(図 1)に基づき、①EUC/EUD に至るまでのシステム環境、②EUC 環境、③EUD 環境に分けて詳述する。なお、この図 1[9]は、明星の文学研究をとりまくメディア環境を情報システム環境としてとらえ、その状況を示したものである。

### 3.1 第 1～第 3 世代システム環境

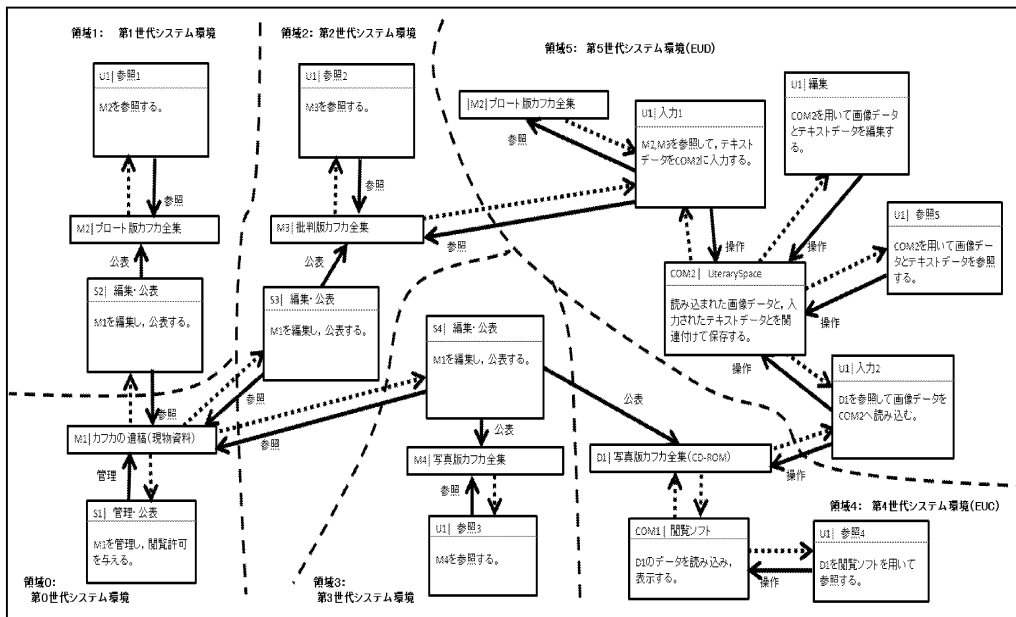


図1 現行システムの分析結果(文献[8]より)

対象のカフカ研究では、研究者が参照する紙メディア上の研究資料＝「全集」から、EUC/EUDに至るまでに3つの世代を経ている。第1世代システム環境は、端的にいえば、プロット版カフカ全集を基盤とする研究環境である。第2世代システム環境は、1980年代以降の批判版カフカ全集に基づく研究環境である。そして、第3世代は、1990年代後半以降の写真版カフカ全集の研究環境である。

これら3世代は、どれも紙の書籍という情報メディア環境であるため、ひとつの世代とまとめてしまっていていいように見える。しかし、第1世代、第2世代と第3世代とは、かたや活字かたやファクシミリという点で、媒体的な相違がある。さらに、第1世代と第2世代の間にも、前者は、一般読者を対象とした普及版、後者は、学術研究者を対象とした学術版だという点で相違がある。

なお、明星はこの点を、情報の信頼性に関わる問題として文学研究システム構想上のポイントであると、2010年3月に開かれた New Directions in Textual Scholarship 国際会議で指摘

している[10]。

### 3.2 第4世代システム環境(EUC)

第4世代以降の特徴は、PCがシステム環境に加わったことである。対象のカフカ研究においては、写真版カフカ全集に付録された、本の内容のPDFファイルを収めたCD-ROMを、読者＝利用者がPCを使って参照する。

ここで、新たな分析視点として、データの流れに着目したDFD(Data Flow Diagram)を加え、PCを含むシステム環境を分析する[11]。その結果が図2である。図2の右上にある四角5は、PCを始めとする情報機器を示しており、利用者が情報機器を介しなければデータベースにアクセスできない点で、それまでの世代と明らかに一線を画していることが明示された。

この点について、2010年3月に開かれた国際会議で明星は、第4世代システム環境を、いわゆる一般的な電子書籍の環境と繋げる議論を展開した。そして、その特徴について、コンテンツとそのコンテンツを載せる搬送体が切り離された点をあげている[12]。この明星の指摘は、対象のカフカ研究だけではなく、広く一般にも

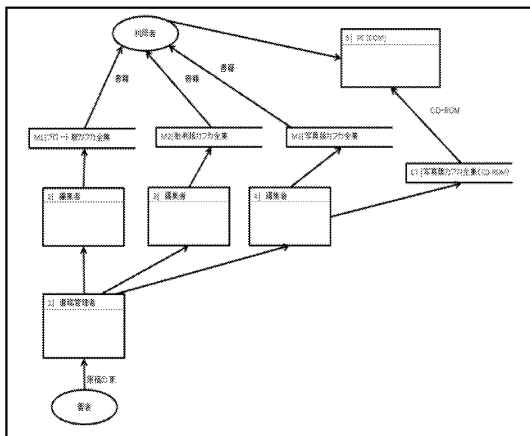


図2 第4世代/第5世代システム環境(現行物理DFD)

EUCが浸透していることを示しており、そのためのシステムデザインが必要であるとも考えられるが、本稿ではそれを示唆するに留めたい。

### 3.3 第5世代システム環境(EUD)

図2右上の四角5で、提供されたデータを指定されたツールを用いて閲覧するばかりでなく、独自に注釈や解釈データを作成し、それを系統的に整理するためのソフトウェアを利用者が独自に開発すれば、第5世代システム環境となる。明星は、第4世代システム環境として提供されたデジタル資料が紙資料特有の制約を継承していることに不満を感じ、①データのグルーピングの固定性、②データのシークエンスの固定性の打破を目指し、プログラマの菅生紳一郎の協力を得て研究ツール *LiterarySpace* を開発している[13]。

この *LiterarySpace* の開発思想については、2008年6月にフィンランドのオウルで開かれた Digital Humanities 国際会議での明星らの発表に詳しい[14]。それによると、*LiterarySpace* は、画像データとテキストデータとコントロールデータの3種類のデータを3次的に組み合わせたデータ構造からなる。画像データの1次元的な蓄積は、「経験」の層と名付けられている。この3次的なデータ構造の全体こそが、まさしく知識創出の構造のモデルであるという。そして、

その画像データひとつひとつに、1対多対応で、2次的に、transcription や interpretation や translation の結果としてのテキストデータが関連づけられる。この2次的な層は、「解釈」の層と呼ばれている。そして、その画像データとテキストデータのセットを、さらに、自由に組み合わせ、シークエンスを入れ替えて、3次的な空間に、新たなデータファイルを創出させた層が、「創作」の層といわれる。

本稿の問題意識に立ち返れば、EUC/EUD の成果といえるこのデータファイルを、いかに位置付け、どのように活用するかが、文学研究システムをデザインする上での要求項目の一つといえる。

## 4 文学研究システムのデザイン

### 4.1 現行システム環境の機能分析

ここで、図1の現行システムに現れた複数の機能について分析を行う。具体的には、同じ区分に入れることのできる機能について、より抽象的な置き換えを行った[15]。図1で現れた機能は、「管理・公表」、「編集・公表」、「参照」、「入力」、「編集」である。これら機能が何を行っているのかに注目してまとめると、表4で示されるような5つの領域として区分できる。すなわち、現物資料のアーカイブを行う機能(領域0)、:画像データとテキストデータを提供する機能(領域I)、テキストデータを

表1 現行システム(図1)の機能分析

| 現行システムの機能(図1より) | (関連) | 区分できる領域                  |
|-----------------|------|--------------------------|
| 管理・公表           | →    | 領域0: 現物資料のアーカイブを行う       |
| 編集・公表           | →    | 領域I: 画像データとテキストデータを提供    |
| 参照              | →    | 領域II: テキストデータを参照する       |
| 入力              | →    | 領域III: 画像データを参照する        |
| 編集              | →    | 領域IV: 画像データとテキストデータを活用する |



参照する機能（領域Ⅱ），画像データを参照する機能（領域Ⅲ），画像データとテキストデータを活用する機能（領域Ⅳ）を持つ5つのサブシステムの総体として，対象の文学研究が成立していると考えられるのである。

#### 4.2 EUC/EUD を前提とした文学研究システムのモデル案

第5世代システム環境では，デジタルデータを自らの PC で参照するだけでなく，それを自らのコンピュータに取り込み，自ら編集し，そして新たな解釈を付加する機能を担っていた。したがって，EUD 環境が次に求めるのは，新たな個人的エディションを作り出し，それを公表したり，既存のデータベースシステムに研究成果を反映させるためのシステム環境であるともいえる。そこで，表1で示した領域に，研究成果を反映させるための機能を追加し，EUC/EUD を前提とした文学研究システムのモデル化を試みた。その検討結果が図3である。なお，モデル化の際には，SSADM の手順に則り，誰が，どのように，予算などの現実的な実現可能性については一先ず棚上げしている。

まず，唯一無二の現物から領域0へ矢印を引き，そこから画像化およびテキスト化をした上

でデータベースに保存する領域Ⅰに繋げた。ここで，テキストと画像は構造化され，関連付けられることによって公開可能となる。このときのインターフェースを領域Ⅱ（テキスト）および領域Ⅲ（画像）として並列に繋いだ。そして，領域Ⅳ(EUC/EUD)ではそれらからデータを取得し，研究成果としてのデータセットを作成して出力する。しかし，それを受け入れるシステム領域が存在しないため，新たに「研究成果データベース」を置いた。領域Ⅰのデータベースはこの機能を経て更新される。

#### 5 問題提起(おわりに代えて)

本稿では，EUC/EUD によって作成されたデータを，研究資料データベースに反映させるための抽象的なモデルについて，対象のカフカ研究に基づいて検討した。図3を一目見て問題と指摘できるのは，たとえ各領域のシステムが分散的に存在することになったとしても，どの現物を基に画像データやテキストデータが作成されたか同定する必要があるということである。この問題について，明星は現物の名付けの問題として指摘している[16]。システムデザインの視点からみても，それが決まらなければ構想も

ままならないことが明らかである。まずは文学研究者と現物資料管理者との今後の議論に期待したい。

また一方で，2010年3月に開かれた国際会議において，デジタル環境における公表の概念の変容と文学研究活動との関連の重要性についても明星により指摘された[17]。すなわち，EUC/EUD 環境で作成される画像データに対する解釈の結果であるテキストデータについて，どの段階でどのように公表するのか，またどのような規則で関連付けるのかなどについて，文学研究者間での合意形成が必要であるという。この指摘は，文学研究システムをデザインするとき，画像データに

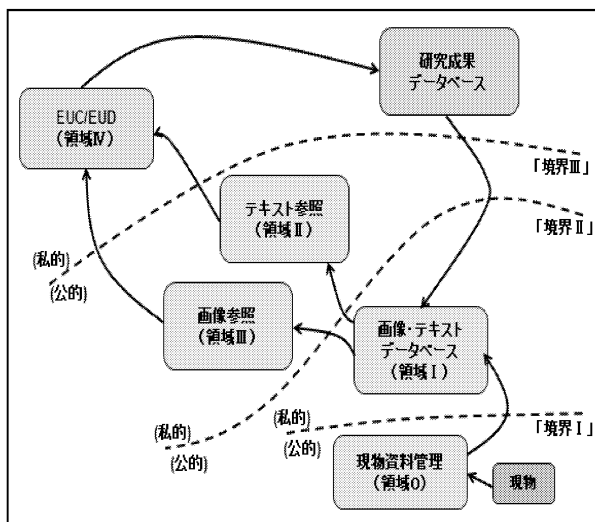


図3 対象の文学研究システムのモデル案

対するテキストデータの関連を、システム全体でどのように保つかという大きな要求項目となる。この要求に対しては、図3の各領域で稼働する機能的システムのハードな課題としても、利用者組織の合意形成などのソフトな課題としてもとらえられよう。また、公表の問題については、補助線(点線)を図3に引いて用いれば、システムデザインの視点からの議論が可能である。たとえば、明星のLiterarySpaceは、境界Ⅰの私的領域(領域ⅠからⅣ)をカバーする研究ツールといえる。また研究関心によっては、境界Ⅱを検討する可能性もあろう。この場合、テキストや画像の新たな参照の仕方そのものが研究対象となる。そして、現在稼働中の多くの文学資料データベースは、境界Ⅲの公的領域をカバーしているようにみえる。一般化をおそれずに考察すれば、利用者の声によっては、一部の機能領域をサブシステムとして私的領域に公開することで、EUC/EUDを取り込み、システム環境を活性化する可能性があるといえよう。

かつて、企業システムにおいてEUC/EUD環境を活性化するためには、情報システム部門と利用者部門との協力が不可欠といわれた[18]。文学研究システムにおいても、共通の現物を扱う研究者や現物資料管理者、そしてシステムデザイナーとの連携が欠かせない。今後も文学研究者らの協力を得ながら、システムデザインの視点から議論を深化させていきたい。本稿が文学研究システムに関する今後の議論に貢献できれば幸いである。

## 注・参考文献

- [1] 本稿では情報システムをコンピュータシステムと同義とする一般的な理解ではなく、以下の文献で示された「組織体および個人の活動に必要な情報の収集・蓄積・処理・伝達・利用にかかわる仕組み」とする広義にとらえる。浦昭二他:『情報システム学へのいざないー人間活動と情報技術の調和を求めて』培風館, 1998
- [2] 田村幸子:「エンドユーザーコンピューティングと情報リテラシー」,『九州産業大学商経論叢』第37号, pp. 95-109, 1996
- [3] ある研究者が、文献上の差異とされていたことが実は研究資料データベース上の登録ミスであったことを発見した。この研究者はこの事実を管理者に指摘したが、修正に応じてもらえなかった。現在もそのままであるという。
- [4] 以下の文献で師茂樹は、「開発された結果ばかりに目がいってしまい、モデルや方法論的な検討はあまり重視されないようである」と、研究資料データベースの研究状況について所感を述べている。高田時雄他:『人文情報学シンポジウムーキャラクター・データベース・相互行為ー報告書』, 京都大学21世紀COEプログラム「東アジア世界の人文情報学研究拠点 漢字文化の全き継承と発展のために」, 2007
- [5] Checkland, P. and Scholes, J., *Soft Systems Methodology in Action*, Wiley, 1990. (妹尾堅一郎監訳:『ソフトシステムズ方法論』, 有斐閣, 1994)
- [6] 佐藤修他:『エンドユーザコンピューティング』, 日科技連出版社, 1996
- [7] 明星の具体的な研究活動については、以下の文献を参照。明星聖子:『新しいカフカー「編集」が変えるテキスト』, 慶応義塾大学出版会, 2002
- [8] 富澤吉樹, 明星聖子:「文学研究活動に着目したEUC/EUDに関する考察」, 経営情報学会2009年秋季全国研究発表大会, pp.288-291, 2009
- [9] 四角形は、対象のシステム環境に存在する利用者(U:User), 提供者(S:Supplier), PC(COM: Computer)の役割を示している。その右横に機能を、その下には機能の説明を記した。長方形はデータ庫(=全集)である。矢印は、四角形と長方形の関係を示す。詳細は注8の文献を参照のこと。
- [10] Kiyoko Myojo: "Public Editing and/or Private Editing: Changes in the Concept of Publication in the Digital Age", International Conference "New Directions in Textual Scholarship", Saitama, 2010/3/26
- [11] Geoff Cutts: "Structured Systems Analysis and Design Methodology 2nd.ed", Alfred Waller Limited Publishers, 1991 (浦昭二監訳:『情報システムの分析と設計ーSSADMとその実践』, 培風館, 1995)
- [12] 明星は、その国際会議(注10)で、従来の紙の書籍は情報とその情報の搬送体が一体化したものだだったが、第4世代システム環境ではそれらが分離しようと指摘した。その上で、近年の電子書籍の普及をめぐる社会的な大きな進展は、人類の書物の世界が全体として大きく第4世代環境に進んでいるという旨を、英語で述べている。
- [13] 研究ツールLiterarySpaceへの要求の詳細については、注8の文献を参照のこと。
- [14] Kiyoko Myojo; Shin'ichiro Sugo: "LiterarySpace: A New Integrated Database System for Humanities Research", *Digital Humanities 2008*, pp.260-262, 2008
- [15] システムデザインでは論理化といわれる(注11参照)。
- [16] 明星聖子:「文学研究資料の未来をめぐる一考察ー編集文献学とアーカイブズ学の狭間でー」, 『アーカイブズ学研究』第11号, pp.92-111, 2009
- [17] 注10参照。
- [18] 注6参照。

## カリキュラムの特徴抽出と時間割の要約生成

### Course Schedule Summarization Base on Viewpoint of Curriculum

堀 幸雄<sup>†\*</sup>      中山 堯<sup>††</sup>      今井慈郎<sup>‡</sup>  
Yukio HORI<sup>†</sup>    Takashi Nakayama<sup>††</sup>    Yoshiro Imai<sup>‡</sup>

<sup>†</sup> 香川大学総合情報センター

Information Technology Center, Kagawa University

〒760-0016 香川県高松市幸町 2-1

E-Mail: horiyuki@itc.kagawa-u.ac.jp

<sup>††</sup> 神奈川大学理学部情報科学科

Department of Information Science, Faculty of Science, Kanagawa University

〒259-1293 神奈川県平塚市土屋 2946

E-Mail: nakayama@info.kanagawa-u.ac.jp

<sup>‡</sup> 香川大学工学部信頼性情報システム工学科

Department of Information Systems Engineering, Faculty of Engineering, Kagawa University

〒761-0396 香川県高松市林町 2217-20

E-Mail: imai@eng.kagawa-u.ac.jp

\* 連絡先著者 Corresponding Author

本研究は、学生が入力した半期分の時間割に対するカリキュラム上の観点を表す5つのキーワードを自動抽出し、各キーワードを項目としたレーダーチャートを作成する要約システムを提案する。本システムにより、カリキュラムを熟知したアドバイザーからアドバイスを受けるように、個々の学生がカリキュラム全体像を知り、潜在的に学習したいと考えているであろう分野を総合的に学ぶのにふさわしい時間割を作成することができるための機能について報告する。

This paper proposes an academic curriculum analyzing system which extracts the characteristics of them via the form of radar charts. This system treats syllabus data which constitute some curriculums, and calculates the similarity between the syllabus based on the occurrence frequency of technical terms. The radar chart view helps to find distinguishing features of the curriculums by visualizing and comparing the assignments of the syllabus to the clusters along various classification axes.

キーワード: シラバス, 文書クラスタリング, 情報可視化

Keywords: Syllabus, Document-Clustering, Information Visualization

## 1 はじめに

大学において、学生は自分の時間割を自分自身の手で決めなくてはならない。所属学科の履修条件を満たす限り、学生は自分の興味や感心などに応じて自由に時間割を決めることができる。時間割作成は、シラバスや大学から配布される修学案内を参照するなど、個々の学生の自主的な努力が期待される。時間割は単に卒業要件を満たすために、専門科目を何単位以上、教養科目を何単位以上とれば良いというわけではなく、開講科目のなから相互に関連する分

野の知識を獲得することが望ましいと考えられる。一方で大学などの教育機関は高度教育の多様化、特色ある優れた取り組みを行なうことが求められ、その結果、カリキュラムは日々進化する学問や技術の進歩、社会的なニーズに対応しながら変化し続けている。各教育機関ではシラバスの電子化が進んでいるが<sup>[7]</sup>、膨大な開講科目の中から、シラバスだけを頼りにカリキュラムの全体像を知り、卒業要件だけでなく、変化しつづける技術や分野の科目をバランス良く取るということは容易な作業ではないと考え

られる。そのため多くの大学では初めて時間割を作成する学生や時間割作成について悩んでいる学生を対象に、教員で構成される専門委員により相談を行なうアドバイザー制度を設けている。この制度は相談者に直接アドバイスをすることができるが、それには多くの時間と労力が必要である。

これまで学生の時間割作成を支援するツールとして、科目ごとのカードを整理することで各科目の理解、学習戦略を目的としたもの<sup>[1]</sup>があったが、これにはシラバスや各科目で行なわれる講義内容を理解する必要があり、このツールはまだ専門分野に知識の少ない学生にとっては効果的に利用することができないと考えられる。また電子化されたシラバスをクラスタリングし、カリキュラムの特徴把握を行なうシステムが提案されており<sup>[3]</sup>、このシステムを用いることで学生は自分が興味を持つキーワードに直接マッチする科目を知ることができるが、自分の考えた時間割がカリキュラムの中でどのような特徴を持つのか、偏った履修はしていないかといった情報を得ることができない。

我々はシラバスデータから各科目の特徴を分析して、学生の考えた時間割の特徴をレーダーチャートで表示するシステム Course Balance Support システムを開発した。本システムは所属学科、コースごとのカリキュラムを分析し、全体の特徴を把握することで、学生の 1 学期分の時間割の特徴がカリキュラム全体でどのような特徴を持つのかレーダーチャートで表示する機能を有する。また本システムは入力された時間割を元に、足りない学習内容を補う科目、同等の内容の関連科目を推薦する機能を持つ。これらの特徴により、カリキュラムを熟知したアドバイザーからアドバイスを受けるように、個々の学生がカリキュラム全体像を知り、潜在的に学習したいと考えているであろう分野を総合的に学ぶのにふさわしい時間割を作成することができると考えられる。

本論文では本システムのモデル化、実装、および評価方法について述べる。以下、2 節では本研究の問題の所在を明確にするために、シラバスを利用した先行研究、ならびに関連する研究について述べる。3 節では本システムの詳細について述べる。続く 4 節では実際のシラバスを用いて、本システムの有効性の評価する方法を提案する。5 節は結論であり、今後の課題をまとめる。

## 2 関連研究

中鉢らは学生が履修計画作成を情報収集、他人とのコラボレーション、試行錯誤という側面から支援するためのツールを提案している<sup>[1]</sup>。このツールは履修計画を立てる上での学生同士の知的生産活動を総合的に支援し、そこで創発された履修計画に関する知識を蓄積するための仕組みを提供する。このような試行錯誤の過程から人間の発想や創造的活動を支援するための試みは以前から研究がなされ、KJ 法を計算機環境上で利用するなどのシステムが提案されている<sup>[2][8]</sup>。時間割作成に含まれる科目を学生自身が分類、整理する試みはシラバスの内容理解、自身の履修戦略の確立に非常に有効であると考えられるが、シラバスに記述されている内容を十分に理解できない大学 1 年次学生等が各科目カードの整理を行なうのは困難であると考えられる。本研究はこの問題に対し、シラバスのクラスタリングを行なうことで学生の立案した時間割がどのような特徴を持っているのかをレーダーチャート形式で表示することで、まだシラバスの内容が十分に理解できない学生でもカリキュラム全体像を把握することを支援する。

野沢らはシラバスデータの分析に関して、シラバスに記述されている内容をもとにクラスタリングを行ない、各シラバスのクラスターへの帰属分布を用いてカリキュラムの特徴把握を支援するシステムを構築している<sup>[3]</sup>。このシステムは独創的なカリキュラムを設計することを目的としている。通常大学におけるカリキュラムは教員等で構成される専門委員がカリキュラム全体のバランス調整を行なうが、多数の科目の内容を記述したシラバス全てを把握することは非常に負荷の高い作業である。シラバスの内容に基づくクラスタリングは現在のカリキュラムの全体像の把握に有効な方法であると考えられる<sup>[4]</sup>。これらの研究はシラバスに記述されているテキストを情報源とする点において本研究と同じであるが、本システムは学生向けに半期ごとの時間割の特徴分析を行ない、カリキュラムからバランス良く履修を行なうことをサポートする点が異なっている。また全開講科目のクラスタリングを行なう点も同様であるが、クラスタリング結果をカリキュラムの全体像を把握するだけでなく、半期ごとの時間割におけるレーダーチャートの成分、値の決定に利用している点が異なっている。

益子らはカリキュラム全体を分析し、科目の属性に形状を与え、科目間の関係を円錐面上に可視化するシステムを提案している [5]。これにより学生はカリキュラム全体が俯瞰することができるが、1 学期分の時間割について、カリキュラム全体からみてどのような特徴があるのかを知ることができない。Zhang らは科目間の関係を 2 次元平面上にグラフで表現したシステムを提案している [6]。このシステムはあくまで科目間の関連を図示するのみであり、カリキュラム全体からみてバランスの良い学習計画を立てるための機能を持たない。本研究は 1 学期分の時間割から同様の内容の関連科目、さらにカリキュラム全体からみたときに現在の時間割に足りない科目を推薦する機能を持つものである。

我々はこれまでにシラバスデータから得られる単語ネットワークを元にした学生の興味を最大にする科目ネットワークの組み合わせを求めるシステム Active Syllabus を提案した [9]。本研究は、Active Syllabus では扱わなかった 1 学期単位での時間割をレーダーチャート形式で要約するものである。

### 3 カリキュラムバランスサポートシステム

本システムは個々の学生の 1 学期分の時間割を入力とし、その時間割をカリキュラム全体からみてどのような特徴があるのかをレーダーチャート形式で要約するものである。本節では本システムの構成、実現方法について述べる。

本システムは個々の学生の時間割の特徴をシラバスに記述された単語で表現することにした。これにより時間割の持つ特徴が抽象化され、カリキュラム全体と個々の時間割との関連の強さを算出することができる。またレーダーチャート形式で時間割を表現するために文書クラスタリングの手法を用いて、カリキュラム全体がどのような特徴を持つのかを得る。本システムの全体の処理の流れを図 1 に示す。

#### 1. データ収集

分析対象とするシラバスをシラバス Web ページから収集し、DB に格納する。各科目をシラバスに記述されているテキストに出現する単語集合に関する文書ベクトルで表す。

#### 2. カリキュラム特徴解析

得られた科目の文書ベクトルを、カリキュラ

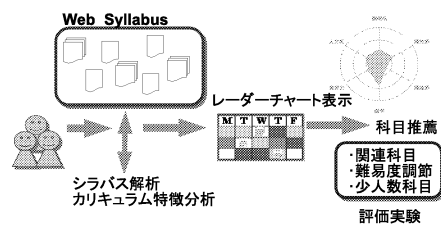


図 1: 全体の流れ

ム全体の特徴をクラスタリングし、クラスターごとの代表語を抽出する。

#### 3. チャート表示

1 学期分の時間割から、時間割中に含まれる各科目と 2. で求めたカリキュラムのクラスタ成分との類似度を算出し、結果をレーダーチャートで出力する。

#### 4. 関連科目推薦

3. で得られるレーダーチャートの形式から、足りない成分を補う科目値の高い成分を得る関連科目を表示する。

以下ではこの手順に沿って、各ステップの詳細と本システムが提供する機能を解説する。

#### 3.1 データ収集

大学提供の Web シラバス閲覧システムから全開講科目シラバスを取得する。各科目シラバスには科目名、対象学科、対象年次、区分(必修、選択、専門必修、専門選択)、担当教員名、開講学期/曜日/時間、単位数といった科目の基本情報に加え、講義概要、講義の目標、授業方法、授業内容(15 回分の講義計画)、理解目標、自己学習課題、教科書といった科目内容に関する情報が記述されている。本研究はこれら基本情報、科目内容をそれぞれテキストデータとしてシラバスより抽出する。さらに各シラバスに出現する単語を抽出し、その出現頻度に基づいて科目ごとのシラバスの文書ベクトルを作成する。科目の文書ベクトルの各単語には式 (1) の tf-idf により重み付けをしたものを特徴とする。

$$c_i = (w_{i1}, w_{i2}, \dots, w_{in})$$

$$w_{i,t_j} = \text{tf}(t_j, c_i) \cdot \text{idf}(t_j)$$

$$\text{idf}(t_j) = \log(N/\text{df}(t_j)) \quad (1)$$

ただし,  $\text{tf}(t_j, c_i)$  は科目シラバス  $c_i$  における単語  $t_j$  の出現頻度であり,  $N$  は総科目数,  $\text{df}(t_j)$  は単語  $t_j$  が 1 回以上出現する科目数である.

本システムは科目内容テキストの形態素解析に mecab<sup>1</sup> を用い, 名詞および固有名詞を抽出した. またストップワード処理として下記の単語を除外した.

- 名詞の数, 接続, 記号.
- ひらがな, カタカナ 1 文字の単語. これらは形態素解析に失敗している可能性が高いため除外した.

### 3.2 カリキュラム特徴解析

カリキュラム全体を俯瞰的に分析するには, 類似する科目を自動的にクラスタに整理するクラスターリングが有効である. 学生が計画した時間割がカリキュラム全体からみてどのような特徴を持つのかを示すレーダーチャートを表示するために, カリキュラム全体をクラスターリングした結果から各クラスターの代表語を利用する. しかし文書クラスターリングは, 自動的にシラバスを分類するために学生にとってどのような観点でシラバスを分類したのか不明であり, どのクラスタが分析に重要なのか判断することがむづかしいという問題がある. 本研究はこの問題に対し, 下記の式によりクラスターの質を定義し, その質の高さをクラスターリングの目的関数とする. これにより, カリキュラム全体からお互いに関連しあっている質の高いクラスタを得ることができると考えられる.

$$\begin{aligned} \text{Quality} &= |C| \cdot \frac{2}{(\text{density}(C) + \text{centrality}(C))} \\ \text{density}(C) &= \frac{2}{|C| \cdot (|C| - 1)} \sum \text{sim}(d_i, d_j) \\ \text{centrality}(C) &= \frac{1}{|C|} \sum |C_j| \cdot \text{sim}(c_i, c_j) \end{aligned}$$

ここで,  $|C|$  はクラスタ  $C$  の文書数,  $d_i, d_j$  は科目シラバスの文書ベクトルを表す. 上記の式はクラスターの品質をクラスターの密度  $\text{density}$  と中心性  $\text{centrality}$  の調和平均で定義したものである. 語で表現されるクラスターの重要度を測るための 2 つの指標 (密度, 中心度) は Callon らにより提案されたものであり [2], 本研究はこれをシラバスに付与さ

れた語に適用し, カリキュラム全体の特徴を把握する. カリキュラム全体をクラスターリングし, 得られたクラスタから代表語を得る.

### 3.3 チャート表示

本システムは 1 つの時間割を入力として, その時間割がカリキュラム上でどのような特徴を持つのかをレーダーチャートとして出力する. 前述のカリキュラム特徴分析で得られた各学科のカリキュラムのクラスタ群と入力された時間割との類似度を計算し, 時間割を各クラスターの代表語の成分との親密度で表現する. これにより, 学生が考えた時間割がカリキュラム全体からみたときにどのような特徴を持つのか分析しやすくなると考えられる. 図 2 はレーダーチャートの出力例である. これは我々の所属する信頼性情報システム工学科のカリキュラムを 6 つのクラスタにわけた場合の例である.

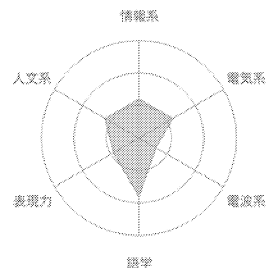


図 2: レーダーチャートの出力

### 3.4 関連科目推薦

本ステップではレーダーチャートの形式から様々な学生の学習戦略に合わせた関連科目の推薦を行なう. 具体的には下記の関連科目の推薦を行なう.

1. レーダーチャートから値の低い成分を補う科目
2. レーダーチャートから値の高い成分の関連科目

## 4 評価実験

本システムの有効性を確かめるために, 下記の 2 点についての評価実験を行なった.

1. 軸表現の評価
2. 軸スコアの評価

<sup>1</sup><http://mecab.sourceforge.net/>

まず 1. のチャートの軸表現は、本システムがカリキュラムをクラスタリングすることで得るカリキュラムの代表的な属性ラベルを、人手で付与した場合のラベルとを比較し、どの程度正しいかを評価するものである。2. の評価スコアは、ある時間割について、チャートの軸ラベルは事前に決まっているものとして、人手で各軸のスコアを与えたときと、本システムの出力するスコアとを比較することで行なう。1. および 2. の評価項目は図 3 に示す各部分のことである。

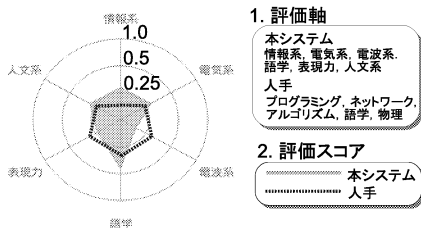


図 3: 本システムの評価箇所

#### 4.1 実験手順

軸表現の評価手順について述べる。実験対象のカリキュラムを卒業した学部 4 年生に対象のカリキュラムについて、人手で軸ラベルを付与し、これを正解事例とする。このとき各軸には 5 つのラベルを付与してもらい、軸の個数は事前に 6 個とした。そしてこの人手で作成したラベルと本システムにより作成した軸ラベルとの概念距離を求め、軸ごとの平均値を求めた。

ラベル間の類似度計算は文献 [10] による概念間の類似度をベクトルの内積で求める方法を用いた (図 4)。

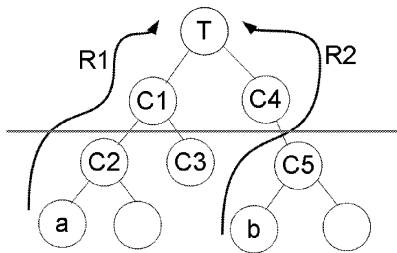


図 4: 概念間類似度計算法

表 1: 軸名の実験結果

| 人手による軸ラベル    |                                      |
|--------------|--------------------------------------|
| クラス番号        | 軸名                                   |
| 1            | 英語 リスニング プレゼンテーション<br>読解力 発表         |
| 2            | フーリエ 微分積分 通信工学<br>ネットワーク ベクトル        |
| 3            | 電気回路 電磁気学 電気 計測 電源                   |
| 4            | 統計学 確率 数学 シミュレーション<br>信頼性            |
| 5            | 特許 戦略 ビジネス システム<br>コミュニケーション         |
| 6            | プログラミング アルゴリズム<br>データベース ソフトウェア 情報工学 |
| 本システムによる軸ラベル |                                      |
| 1            | 英語 TOEIC 作文 読解 リスニング                 |
| 2            | 電気通信 フーリエ解析 情報源<br>電磁波 符号化           |
| 3            | 電気回路 電磁気学 回路設計<br>トランジスタ 電界          |
| 4            | 確率 統計 代数 微分 積分                       |
| 5            | 情報システム プレゼンテーション<br>通信システム, 特許, 科学技術 |
| 6            | プログラミング アルゴリズム<br>データ構造 ソフトウェア 探索    |
| 平均類似度        | 0.56                                 |

各軸のスコアの評価手順は、実験対象のカリキュラムを卒業した被験者により、5 名の学生の時間割 (1 年 ~ 4 年, 前期後期) にあらかじめ 6 つに分けた軸に 10 段階 (1 - 10, 1 刻み) の評価スコアを付与してもらった。評価スコアは各学期の時間割がその評価軸に関して、肯定的であるほど高いスコアを、否定的であるほど低いスコアを与えてもらうように指示した。最終的に 5 名分の時間割の評価スコアを求め、本システムにより得られるスコアとの平均絶対誤差 (MAE) を求めた

$$MAE = \frac{1}{n} \sum_{i=1}^n |cbs_i - hm_i|$$

#### 4.2 実験結果

軸ラベルの実験結果を表 1 に示す。とくに軸において低い値を示しているが、全体を通して、本システムとカリキュラムを熟知した人が作成したカリキュラムの内容は比較的近いものが得られるといえる。

軸スコアの実験結果を表 2 に示す。図 5 はある学生の 1 年次の時間割に関するレーダーチャートで

表 2: 軸スコアの実験結果

| クラス番号 | MAE   |
|-------|-------|
| 1     | 2.762 |
| 2     | 2.142 |
| 3     | 2.048 |
| 4     | 1.858 |
| 5     | 1.858 |
| 6     | 1.810 |
| 平均    | 2.078 |

比較したものであるが、低いスコアはズレが少ないが、大きい値を取るときにスコア値の差が大きいことが多かった。その理由として、被験者毎にスコア値の正規化を行っていない、システムの作成するスコア値は全学期に渡って正規化されるので人間の主観と異なることがあるといったことが挙げられる。

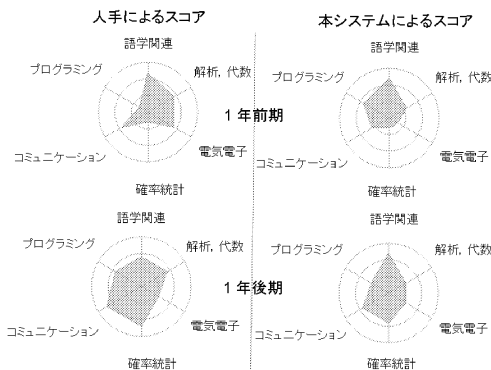


図 5: スコアの比較例

本システムが分析したカリキュラムの着眼点の軸ラベル、および軸の値がカリキュラムを熟知した教員が人手で分析したラベル、値と近いところになるように実験を通じて調整していく作業が必要となるだろう。

## 5 まとめと今後の課題

本稿では、情報抽出、クラスタリングのアプローチを用いてカリキュラムからその特徴を分類抽出し、レーダーチャートという形式で時間割の要約をする方法を提案した。今後の予定として、我々の所属する大学内での予備実験と評価値の算出、評価値向上

のためのカリキュラム分析方法の改善などが挙げられる。またレーダーチャートの形式から学生のニーズに合わせて関連科目の推薦、その効果の測定を行なう予定である。

## 参考文献

- [1] 中鉢欣秀, 石井優毅, 大岩元: 知的創発環境 en-Trace 基本コンセプトと大学環境への導入, ソフトウェア技術者協会, ソフトウェア・シンポジウム, pp.181-184 (2001)
- [2] H. Ohiwa, N. Takeda, K.Kawai, A. Shimomi: KJ editor: a card-handling tool for creative work support, Knowledge-Based Systems, vol.10, pp.43-50 (1997)
- [3] 野澤 孝之, 井田 正明, 芳鐘 冬樹, 宮崎 和光, 喜多一: シラバスの文書クラスタリングに基づくカリキュラム分析システムの構築, 情報処理学会論文誌, Vol.46, No.1, pp.289-300 (2005)
- [4] 宮崎和光, 井田正明, 芳鐘冬樹, 野澤孝之, 喜多一: 電子化されたシラバスに基づく学位授与事業のための科目分類支援システムの試作, 情報処理学会論文誌, Vol.46, No.3, pp.782-791 (2005)
- [5] 益子英俊、牧野光則: 科目間の連携を考慮したカリキュラムの可視化, 電子情報通信学会, 第 19 回データ工学ワークショップ (DEWS 2008), (2008).
- [6] ZHANG Jingjing, 森田 卓巳, 井上 智雄, 横山節雄, 宮寺庸造: 授業業間関係視覚化システム, 電子情報通信学会技術研究報告, ET 教育工学, 電子情報通信学会, Vol.103, No. 697, pp.185-190, (2004).
- [7] 山田信太郎, 松永吉広, 伊藤栄典, 廣川佐千男: Web シラバス情報収集エージェントの試作, 電子情報通信学会誌 D-I, Vol. J86-D-I, No.8, pp.566-574 (2003).
- [8] 大見 嘉弘, 河合 和久, 竹田 尚彦, 大岩 元: カード操作ツール KJ エディタを用いた協調作業における指示操作に関する考察, 情報処理学会論文誌 Vol.36, No. 11, pp.2720-2727 (1995).
- [9] 堀 幸雄, 滝本正志, 今井 慈郎: 活性伝播モデルを用いた履修科目計画の作成, 情報処理学会 SIG-CE, Vol.2006, No.108, pp.47-50, (2006).
- [10] Nguyen Viet Ha, 穂刈譲, 石川勉, 笠原要: 単語の意味の類似性判別のための大規模概念ベース, 情報処理学会論文誌, Vol.43, No.10, pp.3127-3136, (2002).



## 社会的ニーズを踏まえた法律情報に対する知識構造マップの開発 Legal Information “Knowledge Structure Map” Development Responsive to Society’s Needs

川島 啓<sup>1\*</sup>, アダム・ロベル<sup>1</sup>, 山田健智<sup>1</sup>, 大竹裕之<sup>1</sup>

Kei KAWASHIMA<sup>1\*</sup>, Adam LOBEL<sup>1</sup>, Taketomo YAMADA<sup>1</sup>, Hiroyuki OHTAKE

1 財団法人未来工学研究所

The Institute for Future Technology, JAPAN

〒135-8453 東京都江東区深川2-6-1

E-mail: k.kawashima@iftech.or.jp

When using text and data mining to analyze existing data bases and web contents in order to develop service solutions aimed at users’ knowledge-related needs, we must first ask in what way, and until what point, should we attempt to understand those needs? This paper summarizes results of a research project conducted within the framework of the “Needs Oriented Service Science and Engineering Research” (NEXER) program funded by the Japan Science and Technology Agency’s Research Institute of Science and Technology for Society (JST-RISTEX). This research clarified social needs for which “knowledge structure” mapping and other visualizations based on judicial precedent and other legal information, would be useful. Lastly, it explores how related knowledge products and services could best be structured to benefit society.

Keyword: Text mining, Data mining, Legal information, User, Needs, Service innovation

既存のデータベースやWebコンテンツに対して、テキストマイニングやデータマイニングを用いて、ユーザーの知識に対するニーズを満たすようなサービスソリューションの開発を考えるときに、我々はユーザーのニーズをどこまで把握しているだろうか。本稿では科学技術振興機構社会技術研究開発センター(JST-RISTEX)が昨年度実施した「問題解決型サービス科学・工学研究開発(NEXER)事業」における深堀調査の内容を紹介し、判例等法律情報を対象とした知識の構造化に対するニーズを明らかにし、社会への実装の在り方を検討した。

キーワード: キーワード1, キーワード2, キーワード3, キーワード4, キーワード5, (MSP明朝, 10pt)

テキストマイニング, データマイニング, 法律情報, ユーザー, ニーズ, サービスイノベーション

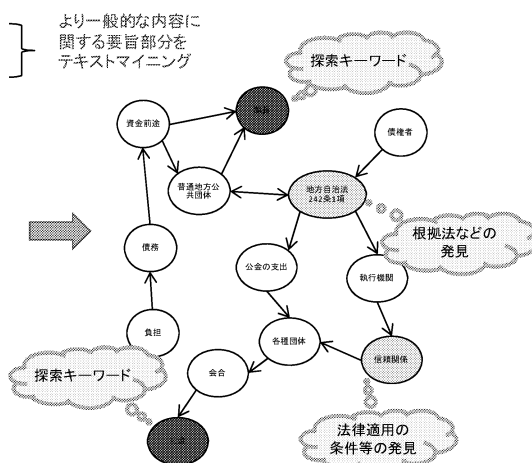
## 1 問題解決型サービス科学・工学研究開発(NEXER)事業への提案課題

2009年8月に科学技術振興機構社会技術研究開発センターはサービス科学・工学分野の研究開発課題についてアイデアをメールで公募するという試みが行われた<sup>[1]</sup>。当研究所は判例データベースにテキストマイニング技術を適用し、国民の法律リテラシーを高めるようなサービスソリューションの開発（知識構造マップ）を提案し（図1）、その後の深堀調査の採択に至った。

深堀調査は2009年11月から12月の2ヵ月間の実施期間であるため、我々はサービスソリューションに対する社会的なニーズを明らかにすることを目的に関係者に対するヒアリング調査を中心に実施した。本稿はこの深堀調査の内容を基に、法律情報サービスの高度化がもたらすサービスイノベーション、社会イノベーションの可能性について論考したい。

[illegible]

判例集子一々

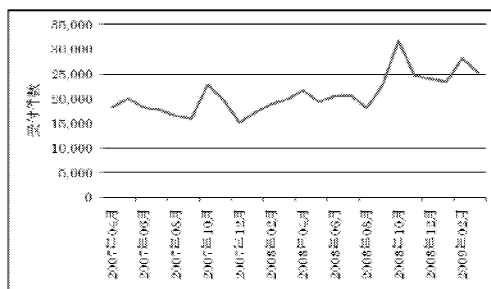


## 知識構造マップの作成

**図1** 判例データベースに対する知識構造マップの作成

## 2 法務及び法律情報に対する国民のニーズ

わが国では司法制度改革の一環として、平成16年に日本司法支援センター（通称、法テラス）が設置されている。法テラスの業務は利用者からの問い合わせに対して、法律情報や各種サービスの提供を実施している。



出所：法テラス資料より筆者作成

図2 法テラスにおける相談件数の推移

法テラスの「無料法律相談」の利用者は年々増加している（図2）。

また、企業では複雑・高度化する経営環境において企業法務やコンプライアンスなどの重要性がますます増しており、法務担当者の抱える業務は、会社法、金融商品取引法、独占禁止法、労働法、下請法、著作権法など多岐の法律分野にわたっている。

法務や法律情報サービスに関するニーズは確実に存在し、これから増加するといえる。そして重要なことは、これらのニーズは必ずしも調停や訴訟などの司法手続きを必要とするものではなく、利用者の抱える問題について判断の糸口となる正確な知識を獲得したいという欲求に基づくものであるという点である。端的に言えば、知りたいことを知るためのサービスが必要ということになる。

一方で、顧客となる利用者のニーズに合致した情報を正しく提供するための仕組みが必要となる。現状でその任を負っているのは弁護士や司法書士であるが、増加する法務ニーズに対して法曹人口の絶対数が不足しており、社会における様々な問題の解決に多くの時間と費用を要する状況にある。しかしながら、弁護士や司法書士の業務の本質は相談窓口業務ではなく、より高度な法律相談や各種司法手続き、弁護活動や交渉などである。現実には、相談窓口業務と本来の業務とが不可分になっているが、無料法律相談に訪れる人々のニーズと弁護士事務所に相談に来る人のニーズとは決定的に異なっているものと考えられる。

弁護士や司法書士の職分を侵すことなく、社会の様々な場所で法情報を正確に、かつ低コストで顧客に伝えるためのサービスの提供が求められている。

### 3 判例データベースを利用した知識構造マップの有用性

#### 3.1 対象とする知識の特性

我々は判例データベースをデータマイニング、テキストマイニングの対象として知識を構造化し、可視化することで社会に有用なシステムの開発ならびにサービスサイエンスの知見を得ることを提案した。このような研究はデータベースの対象が異なるだけで様々なものが提案・実施されている。

しかし、ある構造化されたマップが作成されたとしても、そのままでは知識を得ようとするユーザーの役に立たないことが多い。法律情報の場合にはさらに、ユーザーが知りたい内容と法律の知識との間にギャップが存在している。

例えば判例集という社会に実装されている法律知識のストックに対して、関心事項を検索して何かしらの判例や関連する文書が得られたとする。しかし、その中で何が重要であるかについての判断は法律知識のないユーザーには不可能である。その理由として、判例という文書の特殊性が挙げられる<sup>[2]</sup>。

- ・ 裁判所が個々の裁判の理由の中で示した法的判断
  - 判決は法律の条文に従って判断されるが、その結果示される判例は実務上の解釈という意味で、「判例法」としての法源性を有する。
- ・ 個々の判例は自由な解釈の余地がある（必ずしもいつも正しいとは限らない）ものの、判例は実務を実質的に支配する。
  - 「事実上の拘束力」：先例に従った判断に重きが置かれる。
- ・ 判例の価値
  - 上級審ほど価値が高い（上級審の過去の判例が下級審の上告理由となる）

- 繰り返され確立した判例ほど価値が高い。
- 新しいものほど価値が高い(古い判例に対する違反を上告理由にできない)
- ・ 判例は変更されることがある。
- 裁判官独立の原則:裁判官は条文に従って解釈ができる
- 「法的安定」:法律的判断を体系化するために最高裁判所によって解釈の統一が図られる。

すなわち、判例は社会に蓄積された知識ストックではあるが、一般性・普遍性が重視される知識体系ではなく、専門家による注釈・解釈を都度必要とする個別性の高い参照情報基盤である。一方で、個別の判例自体は口語文で記述されるようになったため、一般の人でも記述されている内容については理解が可能である。そのため、法律の知識や司法プロセスに対する理解を持たない人が判例情報に直接アクセスできたとしたら、誤った認識を持ってしまう可能性が高いことになる。

したがって、ユーザーの知識ニーズと知識ストックをつなぐための高度専門人材が必要であり、彼らの業務を支援するためのシステム開発が望ましいことになる。

### 3.2 既存の法律情報サービスとの差別化

現在、日本で稼働しているWebベースの商用法律データベース・サービスの主なものは以下の3つである。

- ・ 第一法規株式会社 D1-Law<sup>[3]</sup>
- ・ LexisNexis株式会社 LexisNexis JP<sup>[4]</sup>
- ・ ウェストロージャパン株式会社 Westlaw Japan<sup>[5]</sup>

法律情報サービスという視点でみた場合、日本全体ではおよそ800億円の市場規模があると言われている。その中で、上のような商用法律データベース・サービスは全体の10%にも満た

ない状況にあり、今後の市場の拡大が期待されている。主なユーザーは法律事務所が圧倒的に多く、近年は予防法務の観点から企業の法務担当部署の利用も増えている。

実際の実務でこれらの商用データベースをどのように活用しているかについて、関係者(弁護士、検事)のヒアリングを実施した。

テキストマイニング、データマイニング技術を利用した検索結果の可視化については、次のような意見が得られた。

- ・ 現状では検索結果の判例名をみて、事件類型を推測している状態なので、事件類型を生成するような仕組みがほしい。
- ・ 判例同士の引用関係を可視化してほしい。
- ・ テキストマイニング技術だけではなく、量刑などの属性データとの対応を取った上で、全体の分布が俯瞰できると良い。
- ・ 判例における量刑判断等の各要素が相関関係にあり、かつ複雑な案件の場合には論点がキーワードのかたちで可視化されると非常に効果的である。

総じて、実務的に有用な可視化表現に関しては、図3のように抽出結果を意味的に区別するような表現方法が求められている。

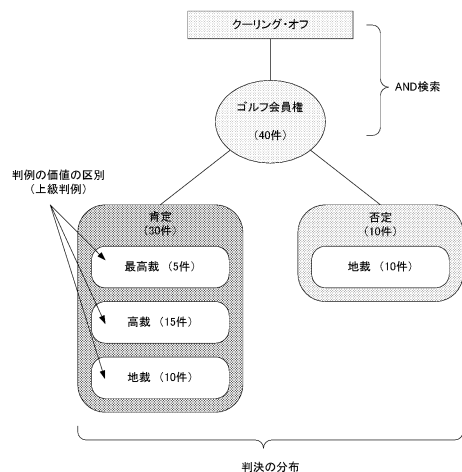


図3 判例検索結果の可視化に対するニーズ

### 3.3 想定するエンドユーザー

法律情報サービスを高度化する上で、想定すべきは法情報に精通した専門人材の業務支援に役立つインターフェースの開発や可視化表現の開発である。

社会の側に法務や法律情報への一次的接近のニーズがある一方で、法律の知識を正確に伝える仲介者の役割が必要となることはすでに述べた。高度知識社会になればなるほど、多くの情報の中から必要な知識を仲介・媒介する役割が重要になってくる(図4)。

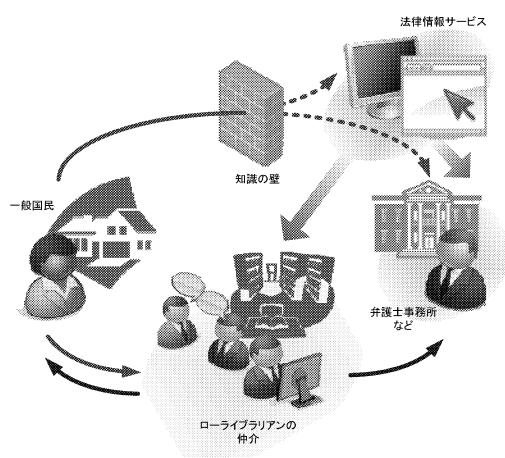


図4 判例検索結果の可視化に対するニーズ

法務サービスの先進国である米国や英国では、ローライブラリアン(Law Librarian)と呼ばれる専門サービス人材の職種が確立している。ローライブラリアンとはパラリーガル的一种であるが、基本的には図書館司書のことである。米国では、地域に設置された法律図書館に勤務するローライブラリアンが様々な情報提供サービスに従事している。ローライブラリアンは米国法律図書館協会(AALL)などの会員であり、図書館情報学の修士号(MLIS)の取得が条件となっている。また、ロースクールや大手法律事務所では、パラリーガルとしてローライブラリアン

が勤務しているが、図書館情報学や情報科学(MIS)の修士号のほかに法務博士(JD)の学位が条件となっているところも多い。

ローライブラリアンのもっとも重要な仕事は、法に関する最新知識を体系的に理解し、法曹関係者のみならず、市民やロースクールの学生、研究者、企業等の法務担当部署などに適切な法情報を提供することである。法務博士を取得しているトップクラスのローライブラリアンはバックオフィス部門ではなく、顧客を相手にしたフロントオフィス業務を担当する。この場合、顧客自らが紛争解決を行うための助言や、司法や裁判外紛争解決手続(ADR)への誘導・紹介業務を行うなど、法律情報の専門家として顧客に対してコンシェルジュの役割を果たすことになる。わが国において法情報サービスを高度化するためには、システム開発と同時に専門職としてのローライブラリアンの職種の確立と、サービスモデルが必要となる。

## 4 法律情報サービスの高度化とサービスイノベーションの可能性

### 4.1 ローライブラリアンの活躍の場

わが国では、成城大学法学部指宿教授らが中心となってローライブラリアンの必要性を訴えており、「ローライブラリアン研究会」として活動を継続されている[2]。

ローライブラリアンのような専門司書が社会の場で雇用され、国民の司法へのアクセスを身近なものとするためには、新しいサービスモデルが必要であり、そのためのニーズを把握する必要がある。法律情報サービスに関するコンシェルジュ的役割を担う人材が現実のニーズとしてどれくらいあるのかについて、関係者などからのヒアリングを下に検討した。

#### 【法科大学院図書館】

- ・ 法律の専門家(あるいはそれを志す学生)でも、得手不得手の分野があり、隣接領域や関連領域の法律情報には疎い場合がある。法教育や法研究において利用者のニーズに応じた法律情報を提供するサービスのプロとして育成することが考えられる。

#### 【法律事務所】

- ・ アソシエイトが現在担っている資料収集業務をバックオフィス機能として機能分化させ、法律情報検索・収集のプロとして育成する。弁護士はより高度な業務に専念し、法務サービスの効率化と高度化を図る。

#### 【企業の法務担当部署】

- ・ 事業や業務に関連した法律情報を整理して経営側に提供できるバックオフィス機能を高度化させ、企業法務のエキスパートを育成する。

#### 【自治体等】

- ・ 現在、地方自治体で弁護士がボランティアに実施している無料法律相談について、弁護士や司法書士の職分を侵すことなく、相談に来る人の悩みに対して適切な法律情報を提供し、弁護士や司法書士を紹介するなどのコンシェルジュ業務を行う人材として配置することが想定される。
- ・ 図書館であれば、上記無料相談業務の他に、専門司書として蔵書へのアクセス情報などを併せて提供する。
- ・ 行政手続きとして一般市民が企業等に対して裁判を起こすかわりに問題を解決できる手段(行政不服申し立て)が用意されている。しかしながら、市民が各自で不服申立書を作成しなければならず、文書作成指導のために地方自治体の窓口業務に大変な負荷

がかかっている。ローライブラリアンが関連する法律情報の知識を基に、適切な行政文書作成指導を行うことができれば、市民や行政にとっての時間費用の削減につながる。

- ・ 給付行政への不服申し立てについても、税理士と行政窓口との間に多大なやり取りが行われている。ローライブラリアンが最新の法律情報の知識を基に仲介することで、各種手続きを円滑にすることが可能になる。

#### 【法テラスなどの相談窓口】

- ・ 直接の法律相談を受けることはできないものの、ローライブラリアンとして法律情報の提供や仲介業務などに役立てることができ、市民と法務をつなげる役割がより効果的にできるようになる。

このように、社会の様々な場所でローライブラリアンが活躍できる余地があることが理解できる。ローライブラリアンの業務を支援するためのシステムを開発することができれば、システムと一緒にローライブラリアンの専門技能をサービス化することが可能となる。

## 4.2 想定される社会へのインパクト

ローライブラリアン自体はわが国ではまだ社会的に認知されているとはいえないが、システム開発と同時にシステムに精通した専門技能職の育成のあり方を検討することで、社会問題の解決にもつながる可能性がある。

#### 【新たな雇用機会の創出】

現在、わが国では司法試験制度改革によって、法科大学院終了を条件に新司法試験への受験資格が与えられるようになった。卒業者は法科大学院終了後5年以内に3回までの受験が可能となっている。法科大学院設立の決定当初、

司法試験制度改革審議会意見書において「法科大学院では、その過程を修了した者のうち相当程度（例えば約7～8割）の者が新司法試験に合格できるよう、充実した教育を行うべきである」との意見が盛り込まれていたが、2009年9月10日に法務省司法試験委員会が発表した新司法試験合格率は27.6%となっており、年々合格率は低下している状況にある。

いわゆる「三振法務博士」問題は、法科大学院修了者（法務博士）の雇用の受け皿が法曹界に依存しすぎた制度設計であったところに根源を持つといえる。新司法試験の受験資格以外に税理士試験の税法科目一部免除などの制度措置があるものの、多くの学生はロースクールを修了して司法試験に合格し、法曹界に就職することを期待しており、社会の側もまたそれを当然のことと理解している。

しかしながら、米国のローライブラリアンの例を見るまでもなく、法務博士（JD）を取得したからといって弁護士や検事の道を進むことが本人にとっての最良のキャリアパスであるとは限らない。法務サービスが多様化されていれば、法務博士の知識を活用できる場はいくらでも作ることは可能であるし、多様な社会を形成するためには法務博士の積極的な活用が不可欠である。ローライブラリアンという専門技能職の確立を目指し、社会のニーズに応じたサービスモデルを構築することで、法務博士の新たな雇用機会の創出に資することが想定される。

#### 【新たな教育サービス・ビジネスの創出】

ローライブラリアンを高度専門技能職たらしめるためには、法律に関する知識は必須であるだけでなく、図書館情報学や情報学に関する知識も重要となる。またはその情報検索に関する技能がデータベース利用技術などに立脚するものであれば、公的もしくは民間資格等のス

キルで保証される必要があるだろう。ローライブラリアンへの教育を品質保証の仕組みとしてビジネスモデル化することが想定される。

#### 【行政の窓口業務の機能分化、行政のバックオフィス機能強化】

地方自治体によっては、法務担当職員がいないケースもままある。そのような自治体では必要に応じて他の自治体に相談に行くなどの措置を採っている。窓口業務の機能分化として行政不服申し立てや給付行政不服申し立てのような行政文書作成指導だけでなく、行政職員が法律情報にアクセスする際にもローライブラリアンの存在は非常に有用である。政令や条例を制定する場合にも、関連情報検索に通じたローライブラリアンが配置されることで行政のバックオフィス機能を強化することになり、市民への行政サービスの向上にもつながる。

#### 【国民の司法へのアクセス機会の増加】

地方自治体で開催される無料法律相談は機会が限定されており、必要な時に気軽に「相談」に応じてくれる場が少ない。法テラスも十分に司法格差を解消できているとは言えず、国民の司法へのアクセス機会はまだまだ制限されている。

必要な時に必要な法情報を得られる拠点や方法がこれまで少なかったために、国民が社会生活を営む上で有用な知識に触れることなく多大なコストを支払っている場合も少なくない。特に行政の支援政策や給付政策に関わる情報への適切なアクセスは、当事者意識を持たない限りは実現し難い。可能であれば対面の場で知識を獲得できるようになれば、国民自らが様々な制度の当事者であることを認識して問題解決の道を探るようになるだろう。

システムの導入とセットで専門図書館員として

ローライブラリアンを地方の公立図書館、社会教育施設等に配置することが可能となれば、国民の司法へのアクセス機会は飛躍的に増大し、法律リテラシーの向上を図ることが可能になる。

## 5 まとめにかえて

ITを利用したサービスイノベーションを達成するために、「既存のサービスを高度化するソリューションを開発する」ことはこれまでも実際にさまざまなかたちで取り組まれている。サービスを高度化するためには、既存のサービスのプロセス分析や、コンテンツ分析、さらにはコミュニケーションの在り方をも含む総合的な取り組みが必要であり、サービスサイエンスの知見もそのような過程から生み出されていると考えられる。

一方、そうした取り組みと合わせて、より重要なことは、社会における役務の提供を専門性に応じて確立し、役務における方法論をサービスソリューションとして構築していくことである。すなわち、専門人材の新しい雇用の場を作っていくことが知識集約型サービスのみならず、社会におけるイノベーションの目標でなければならない。

## 参考

[1]JST-RISTEX:

<http://www.ristex.jp/examin/others/nexer-pj-result.html>

[2] 中野次雄編「判例とその読み方」有斐閣

[3] <https://www.d1-law.com/>

[4] <http://www.lexisnexis.jp/>

[5] <http://www.westlawjapan.com/point/>



## MLA の記述規則に関する比較研究

### A Comparative Study for the Description Rules of MLA

鈴木良徳<sup>1</sup>, 八重樫純樹<sup>2\*</sup>

Yoshinori SUZUKI<sup>1</sup>, Junki YAEGASHI<sup>2\*</sup>

1 静岡大学大学院情報学研究科

Graduate School of Informatics, Shizuoka University

〒432-8011 静岡県浜松市中区城北 3-5-1

E-mail: gs09029@s.inf.shizuoka.ac.jp

2 静岡大学情報学部

Faculty of Informatics, Shizuoka University

〒432-8011 静岡県浜松市中区城北 3-5-1

E-mail: yae@inf.shizuoka.ac.jp

博物館(M)、図書館(L)、公文書館(A)のデータベースを統合する際に、資料の記述規則が異なることは、以前から指摘されてきた問題である。近年、MLA の連携を目的とした活動が活発となり、MLA の標準化に向けてそれぞれの記述規則を見直す必要が出てきている。このような背景を踏まえ、著者らは MLA それぞれの代表的な記述規則である IGMOI、ISBD(G)、ISAD(G) の比較分析を行った。比較の結果、それぞれの機関の活動目的と記述規則が深く関わっていることがわかった。本稿では、比較の結果から、MLA 連携のために今後どのように研究を進めていくべきかを考察する。

When we integrate databases of Museums(M), Libraries(L), and Archives(A), there is a problem that the description rules of MLA are considerably different. Because the standard activity aiming for cooperation of MLA becomes active in recent years, we need to review description rules of MLA for the standardization of MLA. In this time, we compared three standards of IGMOI, ISBD(G), ISAD(G). In this paper, based on our comparison, we propose that which direction MLA cooperation should be researched in the future.

キーワード：記述規則，メタデータ，標準化，IGMOI，ISBD，ISAD(G)

#### 1 本研究の背景

博物館、図書館、公文書館（以下 MLA:

Museums, Libraries, Archives) の資料の記述規則が異なることは、以前から指摘されてきた問題である。MLA それぞれの機関におい

ては、記述対象となる資料の特性が異なるため、記述規則が異なることは当然であるが、近年、MLA の連携を目的とした活動が活発となり、MLA の標準化に向けてそれぞれの記述規則を見直す必要が出てきている。このような背景を踏まえ、著者は M (博物館) L (図書館) A (公文書館) それぞれの代表的な記述規則である IGMOI [1]、ISBD(G) [2]、ISAD(G) [3] の比較分析を行ったので以下に報告する。

## 2 研究の範囲

これまで MLA それぞれの分野に向けて作成されてきた記述規則はいろいろあるが、本論文では比較の対象を IGMOI、ISBD(G)、ISAD(G) の 3 つとする。その理由は、この 3 つの記述規則は、それぞれの分野を代表する国際機関によって作られたものであり、現在世界で用いられている多くの標準が、この 3

つの記述規則を参考としているからである。例えば、現在日本の図書館の多くで用いられている NCR は ISBD に完全に準拠しており、東京国立博物館にて作成されたミュージアム資料情報構造化モデルは IGMOI (博物館資料情報のガイドライン) に準拠した日本版である。こうした理由から、IGMOI、ISBD、ISAD(G) を MLA における記述規則の基礎と考え、比較分析を行う。

## 3 MLA の記述規則の歴史

本研究の中心は MLA それぞれの記述規則の構造と内容を比較することであるが、それらの記述規則がどのようにして今日のような形に発展したのか理解することは、研究を行う上で非常に重要である。そこで、最初はこの 3 つ記述規則が現在に至った経緯を簡潔にまとめる。表 1 は、MLA それぞれの記述規則に関する略年表である。

表 1 MLA の記述規則に関する略年表

|      | M (博物館)                   | L (図書館)                      | A (公文書館)          |
|------|---------------------------|------------------------------|-------------------|
| 1961 |                           | パリ原則                         |                   |
| 1966 |                           | MARC                         |                   |
| 1967 |                           | AACR                         |                   |
| 1968 |                           | MARC II                      |                   |
| 1974 |                           | ISBD (M)                     |                   |
| 1977 |                           | ISBD (G)、ISBD (CM)、ISBD (NB) |                   |
| 1978 |                           | AACR2                        |                   |
| 1980 |                           | ISBD (A)、ISBD (PM)           |                   |
| 1981 |                           | JAPAN/MARC                   |                   |
| 1987 |                           | NCR1987 年版                   |                   |
| 1991 | MDA データ標準                 |                              |                   |
| 1992 |                           |                              | マドリード原則           |
| 1994 | SPECTRUM、MICMO            | NCR 改訂版                      | ISAD (G) 初版       |
| 1995 | IGMOI                     |                              |                   |
| 1996 |                           |                              | ISAAR (CPF) 初版    |
| 1997 | SPECTRUM 第 2 版            | FRBR、ISBD (ER)               |                   |
| 1998 | CIDOC-CRM                 |                              | EAD1.0            |
| 1999 |                           | MARC21                       | ISAD (G) 第 2 版    |
| 2001 |                           | NCR 改訂 2 版                   |                   |
| 2002 |                           | ISBD (CR)                    | EAD2002           |
| 2004 |                           |                              | ISAAR (CPF) 第 2 版 |
| 2005 | SPECTRUM 第 3 版            |                              |                   |
| 2006 |                           | NCR 改訂 3 版                   |                   |
| 2008 | MDA が Collections Trust へ |                              |                   |

まず、MLA で最も早く記述規則の標準化に取り組み出したのは図書館であった。1961 年の国際目録原則会議でパリ原則が採択されたのをきっかけに、図書館分野ではさまざ

まな標準化活動が行われてきた。1967 年に英米にカナダを加えた 3 カ国が AACR を刊行し、1968 年には LC が MARC II の本格運用に入った。これらは図書館の目録標準化を一層加

速させる動きとなり、IFLA も資料種別ごとに ISBD を作成するに至った。博物館と公文書館は 1990 年代からやや遅れて記述規則の標準化に取り組み、博物館は SPECTRUM や IGMOI を、公文書館は ISAD(G)や ISAAR(CPF)を刊行した。やがて、図書館は FRBR のような概念モデルの開発に取り組み、それを追う形で博物館と公文書館も CIDOC-CRMや EAD のような情報統合化の基盤の開発を行い、現在に至っている。公文書館分野の記述規則につ

いては先行研究によくまとめられている[4]。

## 4 構成の比較

これら MLA の記述規則標準化の取り組みを踏まえ、記述規則の考察を行う。MLA 記述規則の代表として IGMOI、ISBD(G)、ISAD(G)を挙げているが、これらはいずれもドキュメントとしてダウンロードすることが可能である。表 2 は、それぞれ記述規則の構成を示したものである。

表 2 MLA の記述規則の構成

| IGMOI                                                                                                                                                                                                                                                                                                                                                                                         | ISBD(G)                                                                                                              | ISAD(G)2nd                                                                                                                           |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------|
| 謝辞<br>はじめに                                                                                                                                                                                                                                                                                                                                                                                    | はじめに<br>ISBD レビューグループのメンバー                                                                                           | 序文<br>はじめに                                                                                                                           |
| 情報グループ・情報カテゴリーのフォーマット<br>情報グループ<br>目的<br>情報カテゴリー<br>例<br>注記<br>情報カテゴリー<br>代替名称<br>定義<br>例<br>注記                                                                                                                                                                                                                                                                                               | 注意事項<br>範囲、目的、利用方法<br>定義<br>ISBD(G) の概要<br>句読点<br>情報資源<br>記述の言語およびスクリプト<br>要約、略語<br>キャピタリゼーション<br>例<br>ミスプリント<br>記号等 | 0. 一般規則に関連した用語集<br>1. マルチレベル記述<br>2. マルチレベル記述規則                                                                                      |
| 情報グループおよび情報カテゴリー<br>1. 取得情報グループ<br>2. 状態情報グループ<br>3. 売却・廃棄情報グループ<br>4. 記述情報グループ<br>5. 画像情報グループ<br>6. 機関情報グループ<br>7. 位置情報グループ<br>8. 記号・刻印情報グループ<br>9. 材質・技法情報<br>10. 測定情報グループ<br>11. 資料関連情報グループ<br>12. 資料収集情報グループ<br>13. 資料受入情報グループ<br>14. 資料名情報グループ<br>15. 資料番号情報グループ<br>16. 資料作成情報グループ<br>17. 資料表題情報グループ<br>18. 構成要素情報グループ<br>19. 記録者情報グループ<br>20. 参照情報グループ<br>21. 複製権情報グループ<br>22. 主題表現グループ | エレメントの仕様<br>1. タイトルと責任表示<br>2. 版表示<br>3. 資源の種類およびタイプ<br>4. 出版、頒布等<br>5. 物理的特徴<br>6. シリーズ<br>7. 注記<br>8. 標準番号および入手条件  | 3. エレメントの記述<br>3.1 識別子のエリア<br>3.2 コンテキストのエリア<br>3.3 内容と構造のエリア<br>3.4 公開および利用条件のエリア<br>3.5 関連資料のエリア<br>3.6 注記のエリア<br>3.7 記述コントロールのエリア |
| 関係組織                                                                                                                                                                                                                                                                                                                                                                                          | 附録<br>A. マルチレベル記述<br>B. 双方向記述<br>索引                                                                                  | 附録 A-1<br>附録 A-2<br>附録 B                                                                                                             |

まず、3 つに共通した事柄として挙げられるのは、最初に「はじめに」の章を設けている点と、中甸で全ての記述エレメントを列挙している点である。「はじめに」の章では、3 つとも、それぞれの記述規則が作成された理由や目的、範囲などが述べられており、

ISBD(G)については作成の歴史や謝辞なども含まれていた。また、中甸の記述エレメントの列挙について言えば、IGMOI は、ISBD や ISAD(G)と比較して記述エリアが突出して多いものの、記述エリアが重なる部分も非常に多かった。どの記述規則でも、エレメントの

列挙の中でそれぞれのエリアの記述例がよく示されていた。また、3つが異なっている点としては、ISBD(G)では、前半の「注意事項」の章で、専門用語や記号、句読点について十数ページに及んで詳しい説明がなされているのに対し、IGMOI と ISAD(G)では、前半の章では詳細な説明がなされていないかった。これは、IGMOI や ISAD(G)と比較して、ISBD の方が記述目的がはっきりしており、記述フォーマットに関して厳しい制約があることを表しているからだと考えられる。ISBD(G)では巻末に、ISAD(G)では前半に、「マルチレベル記述」に関する章があるということ

ころも異なる点であった。

## 5 記述エリアの比較

ここでは、記述エリアの比較を行う。記述規則には2種類あり、記述規則全体の構造に関する問題と、それぞれの記述エリアにどのような形で記述を行うかの記述内容の問題がある。どちらの比較も重要であるが、本論文ではMLA分野の差異を明確にするため、記述構造の問題を主として扱う。表3は、MLAそれぞれの記述エリアを比較したものである。

表3 MLAの記述エリアの比較

| IGMOI                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | ISBD(G)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | ISAD(G)2nd                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. 取得情報グループ<br>1.1 取得方法<br>1.2 取得年月日<br>1.3 取得元<br>2. 状態情報グループ<br>2.1 状態<br>2.2 状態要約<br>2.3 状態の年月日<br>3. 売却・廃棄情報グループ<br>3.1 売却年月日<br>3.2 廃棄年月日<br>3.3 廃棄方法<br>3.4 廃棄の受領者<br>4. 記述情報グループ<br>4.1 物理的特徴<br>4.2 標本の状態<br>5. 画像情報グループ<br>5.1 画像タイプ<br>5.2 画像参照番号<br>6. 機関情報グループ<br>6.1 機関名<br>6.2 部署名<br>6.3 機関住所<br>6.4 機関の所在する国<br>7. 位置情報グループ<br>7.1 現在地<br>7.2 現在地のタイプ<br>7.3 現在地の年月日<br>7.4 通常的位置<br>8. 記号・刻印情報グループ<br>8.1 記号・刻印の文字<br>8.2 記号・刻印のタイプ<br>8.3 記号・刻印の記述<br>8.4 記号・刻印の技法<br>8.5 記号・刻印の位置<br>8.6 記号・刻印の言語<br>8.7 記号・刻印の訳文<br>9. 材質・技法情報<br>9.1 材質<br>9.2 技法<br>9.3 記述の構成要素<br>10. 測定情報グループ<br>10.1 寸法<br>10.2 計測値<br>10.3 計測単位<br>10.4 計測部分<br>11. 資料関連情報グループ | 1. タイトルと責任表示<br>1.1 正式タイトル<br>1.2 一般資料指定<br>1.3 並列タイトル<br>1.4 その他のタイトル情報<br>1.5 責任表示<br>1.6 収集タイトル以外の著作<br>2. 版<br>2.1 版表示<br>2.2 並列版表示<br>2.3 版に関連した責任表示<br>2.4 追加版表示<br>2.5 追加版表示に関連した責任表示<br>3. 資源の種類およびタイプ<br>4. 出版、頒布等<br>4.1 出版地、頒布地等<br>4.2 出版者名、頒布者名等<br>4.3 頒布者の機能表示<br>4.4 出版・頒布年月日等<br>4.5 作成地<br>4.6 作成者名<br>4.7 作成年月日<br>5. 物理的特徴<br>5.1 特別資料指定および資源の範囲<br>5.2 その他の物理的特徴<br>5.3 寸法<br>5.4 付随する資料<br>6. シリーズ<br>6.1 シリーズもしくはサブシリーズの正式タイトル<br>6.2 シリーズもしくはサブシリーズの並列タイトル<br>6.3 シリーズもしくはサブシリーズに関するその他のタイトル情報<br>6.4 シリーズもしくはサブシリーズに関する責任表示<br>6.5 国際標準逐次刊行物番号<br>7. 注記<br>8. 標準番号および入手条件<br>8.1 標準番号<br>8.2 キータイトル<br>8.3 入手条件および価格<br>8.4 制限 | 3.1 識別子のエリア<br>3.1.1 参照コード<br>3.1.2 タイトル<br>3.1.3 年月日<br>3.1.4 記述のレベル<br>3.1.5 範囲および媒体の数量<br>3.2 コンテキストのエリア<br>3.2.1 作成者名<br>3.2.2 管理上の履歴<br>3.2.3 蓄積年月日<br>3.2.4 入手先および伝来<br>3.3 内容と構造のエリア<br>3.3.1 範囲と内容<br>3.3.2 評価・破壊・保存年限<br>3.3.3 追加受入<br>3.3.4 編成<br>3.4 公開および利用条件のエリア<br>3.4.1 公開条件<br>3.4.2 複写条件<br>3.4.3 史料の言語<br>3.4.4 物理的特徴および技術要件<br>3.4.5 検索手段<br>3.5 関連資料のエリア<br>3.5.1 原資料の存在および場所<br>3.5.2 複製品の存在および場所<br>3.5.3 関連資料<br>3.5.4 出版に関する注記<br>3.6 注記のエリア<br>3.6.1 注記<br>3.7 記述コントロールのエリア<br>3.7.1 アーキビストの注記<br>3.7.2 規則または協定<br>3.7.3 記述年月日 |

|                     |  |  |
|---------------------|--|--|
| 11.1 関連場所           |  |  |
| 11.2 関連年月日          |  |  |
| 11.3 関連グループ名もしくは人物名 |  |  |
| 11.4 関連タイプ          |  |  |
| 11.5 本来の機能          |  |  |
| 12. 資料収集情報グループ      |  |  |
| 12.1 収集場所           |  |  |
| 12.2 収集年月日          |  |  |
| 12.3 収集者            |  |  |
| 12.4 収集方法           |  |  |
| 13. 資料受入情報グループ      |  |  |
| 13.1 現所有者           |  |  |
| 13.2 貸出者            |  |  |
| 13.3 受入年月日          |  |  |
| 13.4 受入番号           |  |  |
| 13.5 受入理由           |  |  |
| 14. 資料名情報グループ       |  |  |
| 14.1 資料名            |  |  |
| 14.2 資料名タイプ         |  |  |
| 14.3 資料名の典拠         |  |  |
| 15. 資料番号情報グループ      |  |  |
| 15.1 資料番号           |  |  |
| 15.2 資料番号タイプ        |  |  |
| 15.3 資料番号年月日        |  |  |
| 16. 資料作成情報グループ      |  |  |
| 16.1 作成場所           |  |  |
| 16.2 作成年月日          |  |  |
| 16.3 作成グループ名もしくは人物名 |  |  |
| 16.4 作品の役割          |  |  |
| 17. 資料表題情報グループ      |  |  |
| 17.1 タイトル           |  |  |
| 17.2 タイトルのタイプ       |  |  |
| 17.3 タイトルの訳文        |  |  |
| 18. 構成要素情報グループ      |  |  |
| 18.1 構成要素の数         |  |  |
| 18.2 構成要素の記述        |  |  |
| 19. 記録者情報グループ       |  |  |
| 19.1 記録者            |  |  |
| 19.2 記録年月日          |  |  |
| 19.3 典拠             |  |  |
| 20. 参照情報グループ        |  |  |
| 20.1 参照             |  |  |
| 20.2 参照タイプ          |  |  |
| 21. 複製権情報グループ       |  |  |
| 21.1 複製権に関する注記      |  |  |
| 21.2 著作権者           |  |  |
| 22. 主題表現グループ        |  |  |
| 22.1 主題表現           |  |  |
| 22.2 主題表現の記述        |  |  |

最初に全体の数を比較すると、IGMOI の情報グループが 22、情報カテゴリーが 74 と突出して多いことがわかる。これは、IGMOI は ISBD や ISAD(G) とは異なり、記述のガイドラインという位置づけであり、必ずしも全項目を記述する必要はなく、各館が必要に応じて変更することが求められているためである。また、ISBD(G) も二番目に記述エリアの数が多いが、ISBD の場合は実際のエリアは 8 つであり、各事柄については資料から情報が得られる場合のみ記述を行うなど、記述については資料そのものからどの程度情報が得られるかに依存している。そして、ISAD(G) は、大雑把に言って IGMOI を縮小したような記

述エリアになっているが、3.7.1 アーキビストの注記、3.7.2 規則または協定など、とどこに公文書館として特徴的なエリアが見られる。3 つを比較してわかることは、IGMOI には測定情報や関連資料情報、構成要素情報など資料を深く分析して得られる情報を記述するエリアが多く、ISBD(G) は逆に、資料を見ればすぐに得られる情報を記述するエリアが多いが、一方で利用と公開の面から、タイトルや出版者、標準番号等の検索時に必須となる情報に重点を置いているということである。ISBD(G) の記述例には記号や句読点が多数用いられており、検索することを目的としているため記述フォーマットに

はやや制約がある。また、ISAD(G)はどちらかというところの中間であり、記述内容に関する制約は少ないとともに、検索に必須となる情報も、調査資料としての情報を記述するためのエリアも充実して用意されていることがわかった。

## 6 考察

近年、MLA の連携を目的とした活動が活発となり、MLA の標準化に向けてそれぞれの記述規則を見直す必要が出てきているが、MLA それぞれは活動目的が互いに異なっている。情報を統合化するためとはいえ、今回の比較分析は単純に項目比較だけで 3 つに共通した記述規則を検討したが、単純な問題でないことがわかった。図書館分野では、機械可読なメタデータとして MARC が開発されているが、MLA を連携させたメタデータの開発は、なお各分野資料の性質と意味的な検討が必要となる。その理由は、MLA の活動目的や活動内容が根本的に異なり、記述規則を統合することは MLA それぞれの活動の観点から好ましくないためである。この点についてはこれまででもよく議論されてきた[5][6]。

M(博物館)では、資料は研究や調査の目的で利用されるのが主であって、利用に際して貸出などは非常に制限される。そして、IGMOI における各エリアに記述されるエリアも調査して得られる情報がほとんどである。また逆に、L(図書館)では資料の利用に重点を置いているため、検索性能が重視されるが、その分 ISBD に関しては記述フォーマットが非常に細かく指定される。A(公文書館)はその中間と考えられ、資料は研究や調査に用いられるものであると同時に、利用も適切に行わなければならない。この辺りに各分野の隔たりがあると思われる。MLA の連携を高め、共

通した記述規則やメタデータの開発を行っていくためには、MLA の活動についてさらに研究が必要だと考えられる。

## 参考文献

- [1] CIDOC Guidelines:  
<http://cidoc.mediahost.org/pub-guidelines-1995%28en%29%28E1%29.xml>  
(2010/03/30)
- [2] Superseded ISBDs:  
<http://www.ifla.org/en/isbd-rg/superseded-isbd-s> (2010/03/30)
- [3] ISAD(G): General International Standard Archival Description, Second edition:  
<http://www.ica.org/en/node/30000>  
(2010/03/30)
- [4] 坂口貴弘:「北米におけるアーカイブズ記述規則の特性」, 国文学研究資料館紀要, 2009.
- [5] 田良島哲:「博物館の情報環境と MLA 連携」, アート・ドキュメンテーション学会創立 20 周年記念予稿集, 2009.
- [6] 田窪直規:「MLA 連携の動向とこの連携を捉える 3 つの視点」, アート・ドキュメンテーション学会創立 20 周年記念予稿集, 2009.

## 祭りデータベースモデルの研究調査

### A Study of Database Model for “MATSURI”

山野邊 溪<sup>1\*</sup>, 八重樫 純樹<sup>2</sup>

Kei YAMANOBE<sup>1\*</sup>, Junki YAEGASHI<sup>2</sup>

1 筑波大学大学院 図書館情報メディア研究科

Graduate School of Library, Information and Media Studies, University of Tsukuba

〒305-8550 茨城県つくば市春日1-2

2 静岡大学 情報学部

The Faculty of Informatics, Shizuoka University

〒432-8011 静岡県浜松市中区城北3-5-1

歴史的資料の情報化について、国や様々な機関で考えられている。これら情報化はまだ十分ではないが歴史的資料全般を一度に考えることは難しい。そこで、日本の祭りをテーマとして情報化を考えた。そもそも祭りの定義についてと他の歴史的資料の動向について勘案し、最終的に祭りデータベースを作成するためのメタデータとDublin Coreでのマッピングを図った。

専門的すぎる情報化は、のちの十分な利活用の弊害となる。基盤作成と運用段階での確固としたルールが必要であるが、まずはそれぞれの分野における基盤作成が望まれる。

Informationization of historical information is thought in a country and various organizations. It is difficult to think about the historical information at the same time. This informationization is not enough yet. Then, it thought about informationization of the festival of Japan (MATSURI). At first, I take definition of MATSURI and the trend of other historical information into consideration. Finally I attempt mapping in the Meta data and Dublin Core to make the database of MATSURI.

Too special informationization becomes bad effect of enough profit use. Ultimately it needs to make basis and rules for use. Above all it needs each to make basis for historical information.

キーワード: 情報資源化, 歴史的資料, 無形民俗文化財, 祭り, Dublin Core

Keywords: Informatization Resource, Historical Information, Intangible Folk-Cultural Property, “MATSURI”, Dublin Core

まれ受け継がれた国民の財産である。

#### 1 はじめに

文化財とは、これまでの長い歴史において生

日本における文化財保護行政は明治時代から行われ、文化財保護法が昭和25年に制定され何度かの改正を経て現在に至っている

[1].

文化財保護法では、文化財を6つに分類している[2]. 本論文では、このうちの民俗文化財の一つをテーマとして論を進めていく. そもそも文化財保護法の中で民俗文化財を『衣食住、生業、信仰、年中行事等に関する風俗慣習、民俗芸能、民俗技術及びこれらに用いられる衣服、器具、家屋その他の物件で我が国民の生活の推移の理解のため欠くことのできないもの』と定義している. よって、民俗文化財は前者を無形民俗文化財として、後者を有形民俗文化財に分けることができる.

民俗とは、人々の日常生活の中で生み出され、有形・無形という形で継承してきたものである. 長い間に培われたものを将来に残していくことを否定する理由はない. もちろん、民俗学分野においてもこれまで国や地方公共団体などにおいて調査が行われ記録が行われている. しかしながら、これら調査報告が十分に利活用されているとはいえない.

そこで、本論文では日本の祭りをテーマとして情報化を行うための方針について述べていく.

## 2 歴史資料情報化の動向

歴史資料の情報化については、全国で議論と活用が進められている. 国として行われているもの、研究機関で行われているもの、大学で行われているものなどがある. また、各博物館独自で行っている例も多々ある.

### 2.1 人間文化研究機構による取り組み

大学共同利用機関法人・人間文化研究機構は、複数の機関がそれぞれの研究分野の枠を超えた人間科学の研究を行うた

めに設立された. そこで各機関が保有している研究データベースを一元的に利用できるように研究資源共有化システムを2008年から公開している[3].

このシステムでは、100以上のデータベースを横断検索することができる. 複数のデータベースを活用するために Dublin Core をベースとして 5W1H と時空間情報から人間文化研究機構マッピングを策定している.

統合検索をすることで新たな研究について進めることができる. しかしながら、課題も多くあるとしている. その中の一つとして Dublin Core のマッピングの困難さがあげられる. Dublin Core で実資料を記述する際の定義の差異や基本 15項目で表現できないものに対しては、独自の項目追加等が考えられている. 統合検索システムの意義として各データベースを検索するためのインデックスとして考えるべきだとしている[4].

### 2.2 文化庁における取り組み

文化庁においても歴史的資料の情報化が図られている. その一つに文化遺産オンラインがある[5].

日本全国の内から登録されている博物館・美術館等から提供された情報をもとに、所蔵品の公開・検索・閲覧を行うことができる. 主な目的として全国にある博物館・美術館等に収蔵されているものを広く知らしめるために Web 上に情報を公開することである. そのために全ての収蔵品について公開されているわけではない. 主に各博物館・美術館等へと誘導を目的としている.

多くの博物館・美術館等の参加を求めるために統一したメタデータの作成を行っ



ている。

文化遺産オンラインについては、全国幾多の博物館・美術館のための総合窓口のような働きを持っている。また、館ごとの志向により情報の偏りも生じてくる。確かに普段目にする機会が少ない情報を見るきっかけとしての役割として考えるならば有益であるが、確固とした情報化とより多くの情報の公開が必要である。

### 3 祭りについて

ここでは、祭りの定義とこれまでの祭りに関する調査と先行研究について述べる。

#### 3.1 「祭り」とは

祭りとは、「神霊を招き迎え供え物や歌舞を捧げて歓待・供応し、祈願や感謝をして慰撫すること」と定義されている。祭りとは日常の生活とは異なる時期に行う行事のことであり一般的にハレの日に行う。日本における祭の多くは農業を中心とした生業における儀礼である。そのために農業における重要な時期に行われることが多い年中行事として季節感を呼び起こし、四季の節目を形作ることを目的としている。種まきから収穫までが一年で済むことが多い稲作民は一年周期での祭りを行い、長期的な作業が必要な焼畑民は数年周期で大きな祭りを開いてきた。また、非農業民は別に守護神や祖人を奉り祭りを行ってきた。祭りをを行うためにはさまざまな役割を持った人間や造り物で構成される。

[6]

これら祭りは本来、ある集団(小規模のものは家族単位から大規模のものでは部落単位)の中で行われる。特に集団の中心としてその地域の神社が使われることが多かった。

明治以降、本来の祭りの定義とは異なる祭りが外来から多く流入し定着してきている。集団における宗教的な意味合いからエンターテインメント性意味合いが強くなってきている。

また、現代は人が流動的に土地を移ることが珍しくない。このことも本来の祭りと意味合いが異なっている一つの要因である。

さまざまな文化が混同しているなかで伝統が完全に消えることはないと思われるが、本来の「まつり」の意味を考えるためには過去への理解と柔軟な定義の構築が必要になってくる可能性もあると考える。

#### 3.2 文化庁主導の民俗文化財調査

日本における初めての全国規模の民俗文化財調査は、1962-64年度に行われた民俗資料緊急調査である。これは、戦後の急激な人々の生活の変化に伴い行われた民俗調査である。広く現状を把握するために行われたため全ての網羅をしていないが、基礎的な調査要項を用いて調査が行われた。しかしながら、調査結果が広く活用されることはなかった(報告書の数が少ないなどが原因)。後にこの調査を参考に精度を上げた調査が行われているが調査結果が十分に利活用されていない。

民俗資料緊急調査などは、民俗分野全体についての調査であるが、祭りに特化したものとして平成4年度から現在まで祭り行事調査が行われている。この調査では、各県の祭りの全数把握を目的として小字・村・地域・職業集団を単位としたものを対象として実施要項を定めて3年計画で調査が行われている。

祭りの概要から祭りで使われるモノや

人間の役割や芸能についての調査項目が実施要項には用意され各県で則った調査表を作成している。

しかしながら、各県の祭りには独自性があり似たような項目が異なる表現で報告されていたりする。仮に全国調査が完了したのちの利用についてこのままでは十分な利活用がなされないと思われる。

20年近く(全県調査完了まであと10年弱かかる見込み)かけた調査が利活用されないのは至極もったいないことである。この調査が無駄にならないことが望まれる。

### 3.3 先行研究における祭りについて

祭りの情報化について、神奈川大学大学院の八重樫純樹の講義での議論がある。ここでは、祭りを「神社で何らかの伝承的な神事を行うもの」と定義した。それを基に台帳項目の基本的な骨格として「概要」「内容情報」「実行組織情報」「記入者情報」の4グループを軸として作成している[7]。

祭りの所在情報としての役割は、これで十分なものであると考えられる。実際に、先の「祭り・行事調査」の台帳と比較しても大きな違いは見られない。神社を中心としてどのような祭りをを行っているかどうかを把握するための項目である。文字情報として祭りの情報をこのモデルで蓄積を行えばWeb上においても資料検索が容易になると考えられる。

吉村案のデータベース項目では、所在情報を主としている。同様なモデル案は、先の授業の中で複数検討され別のモデル案の中には、かなり詳細な情報までを網羅するためのものも検討されていた。しかしながら、どのくらいの粒度で情報を整理するのかということを考えるとこの

程度の精度が後々の応用において柔軟性をもたらすと思われる。

## 4 モデル提案

### 4.1 統一台帳案

前章で本論文の主題である祭りとその調査について述べた。その中で、文化庁主導で行われている祭り行事調査で使用されている台帳と吉村案の台帳について見てきたが2つの台帳から有意義な情報化を行うための統一台帳をここで定義する。

先に示した台帳をそのまま統一案として利用することは難しい。なぜならば、利活用を主とする情報化においてはある程度名称を一般化する必要がある。また、その後の応用として独特すぎるものは好ましくない。管理番号や住所、名称等ではできる限り一つの項目としてまとめる必要がある。また、概要は広く文字情報として格納できるようなものにする。大きな概要という項目の中である程度のルールを定めて表現を加えれば十分役割を果たすことができる考える。

台帳案は最終的なマッピング例と共に示す。

### 4.2 マッピング

歴史的資料の進んだ情報化を行っている機関に博物館がある。博物館界では、博物館資料の標準化を古くから進めてきている。標準化を進めてきている組織の一つに国際博物館会議(ICOM<sup>1</sup>)の国際ドキュメンテーション委員会(CIDOC<sup>2</sup>)がある。CIDOCでは、1994年に「博物館資料の最小限情報分類勧告<sup>3</sup>」(MICMO<sup>4</sup>)を提案した。1995年

<sup>1</sup> The International Council of Museums

<sup>2</sup> International Committee for Documentation

<sup>3</sup> Proposal Guideline for an International Standard

にはMICMOを 詳細化した「博物館資料情報のための国際ガイドライン」(IGMOI<sup>5</sup>)を提案している。また, 1998年にはデータベース変換のための概念参照モデル (CRM<sup>6</sup>)を作成している[8]。また, 日本においても2005年に東京国立博物館が「ミュージアム資料情報構造化モデル」を提案しモデル適応に向けて様々な取り組みがされている[9]。

博物館界におけるこのような取り組みはあくまでもモノ資料を中心に考えられている。本論文で主題となる祭りは, モノではないのでこれらを使って祭り情報をマッピングすることは非効率である。

そこで, web資源として祭り情報を扱うこととしネットワーク上の情報資源の記述・発見のためであり先の研究資源共有化システムでも利用されているDublin Coreを用いてマッピングを図る。

Dublin Coreでは, 基本15エレメントとしてSimple Dublin Coreが用意され, より詳細化されたQualified Dublin Coreがのちに制定された。今回はこの2つを利用して「usage guide」を参考にマッピングを図った[9]。

具体的なマッピング例は最後に載せた[表]。国立歴史民俗博物館内においてのマッピングであげられた問題と同じようにどの項目をどのエレメントに適用するかどうかテーマが祭りでも発生した[10]。例えば, 台帳中の地区と時期についてDublin Coreでは, 広義にCoverageに分類されると思われる。Qualifiedにおいてより細かいエレメントが用意されているが, どのよう

に利用するかによって考え方が変わってくると思われる。

ある分野だけに特化したメタデータの作成は, 本来簡便なものであるメタデータという趣旨に矛盾する。ある程度のメタデータの作成の後は, 運用上のルールの必要性があると感じられる。

## 5 まとめ

本論文では, 無形民俗文化財の一つである祭り行事をテーマとして情報化のモデル提案を行った。十分な利活用のためのモデル提案を行うに際し国際標準に準拠したメタデータとして Dublin Core を利用したが, 本来祭りに対するメタデータというものは存在しないために Dublin Core の適用には無理がある部分もある。一つの方向性としては示すことができたがやはり特徴的な項目についてマッピングを行うことは難しい。今回は, 祭り・行事を例として情報化を考えた。後の応用を考えるとある程度の粒度を考える必要がある。運用段階でのルールが確固としている必要があると思われる。小さな情報基盤の作成が将来の大きな基盤の一つとなりえるわけでこれらの検討が望まれる。

## 参考文献

- [1] 文化庁文化財保護法研究会編著, 文化財保護法 改正のポイントQ&A, ぎょうせい, 1997, pp10-19 pp44-59
- [2] 文化庁 | 文化財 | 種類 | 概要  
<http://www.bunka.go.jp/bunkazai/shurui/gaiyou.html> (2010年3月31日参照)
- [3] 研究資源共有化システム  
<http://www.nihu.jp/sougou/kyoyuka/to>

<sup>4</sup> Minimum Information Categories for Museum Objects

<sup>5</sup> International Guidelines for Museum Object Information

<sup>6</sup> Concept Reference Model

ugou/index.html (2010 年 3 月 31 日参照)

[4] 山本泰則, 「データベース横断検索のための共通メタデータ - 統合検索システムにおける定義と 2 つの課題-」, 人間文化研究機構研究資源共有化シンポジウム「研究資源共有化 - その展開と可能性 -」, pp1-6, 2008. 3

[5] 文化遺産オンライン

<http://www.bunka.nii.ac.jp/index.do>  
(2010 年 3 月 31 日参照)

[6] 福田アジオ; 神田より子; 新谷尚紀; 中込睦子; 湯川洋司; 渡邊欣雄編: 「日本民俗大辞典・下」, pp577-579, 吉川弘文館, 2000. 3

[6] 吉村祐一, 「民俗資料の情報化に関する研究」, 神奈川大学大学院歴史民俗資料

学研究科修士論文, 2006 年

[7] E. Orna&Ch. Pettitt 編; 安澤秀一監修; 水嶋英治編訳: 「アート・ドキュメンテーション叢書 2 博物館情報学入門」, 勉誠出版, 2003. 6

[8] 斎藤伸雄, 「ミュージアム資料情報構造化モデル応用の検討」, 情報知識学会誌, Vol. 16, No. 2, pp. 249-252, 2006

[9] Dublin Core Metadata Initiative  
<http://dublincore.org/>

(2010 年 3 月 31 日参照)

[10] 安達文夫; 鈴木卓治: 「歴史研究データベースの Dublin Core へのマッピングとその課題」, 情報処理学会研究報告 人文科学とコンピュータ研究会報告, 2006 (112), pp. 47-54 2006. 10

表: Dublin Core でのマッピング例

| 台帳項目 | 項目       | DCMES element                                                 | 例(基礎調査から)                              |
|------|----------|---------------------------------------------------------------|----------------------------------------|
| 台帳   | 作成日      | <i>Created</i>                                                | (台帳による)                                |
|      | 更新日      | <i>Modified</i>                                               | -                                      |
|      | 作成者      | <i>Publisher</i>                                              | (台帳による)                                |
| 管理番号 | 県番号      | <i>Identifier<br/>(Bibliographic Citation)?</i>               | 22(静岡県)                                |
|      | 地区番号     |                                                               | 44(袋井市)                                |
|      | カード番号    |                                                               | (台帳による)                                |
| テーマ  | 標題テーマ    | <i>Type</i>                                                   | 17 生業                                  |
|      | 関連テーマ    |                                                               | 18 正月                                  |
| 地区   | 調査地区     | <i>Coverage→Spatial</i>                                       | 豊沢                                     |
|      | 所在地      |                                                               | 尊永寺                                    |
| 行事名称 | 正式名      | <i>Title</i>                                                  | 法多山田遊祭                                 |
|      | 別名       |                                                               | -                                      |
| 時期   | -        | <i>Date/Coverage→Temporal</i>                                 | 1 月 7 日                                |
| 概要   | 概要       | <i>Subject/<br/>Description</i>                               | (台帳による)                                |
|      | 芸能       |                                                               |                                        |
|      | 新撰・供物・料理 |                                                               |                                        |
|      | 道具立て・用具類 |                                                               |                                        |
| 状況   | 指定       | <i>Description</i>                                            | 国の記録選択指定                               |
|      | 存続       |                                                               | 順調                                     |
| 寄与者  | 伝承組織     | <i>Contributor</i>                                            | (台帳による)                                |
| 関連情報 | 地区内の祭り   | <i>Relation→</i>                                              | (台帳による)                                |
|      | 参考文献・資料  | <i>*Relation のより詳細化項目<br/>Coverage→<br/>Spatial, Temporal</i> | ex. 参考文献<br>ex. 画像・映像資料<br>ex. 報告書間の関係 |

※表中において斜体で表示されていないものは Simple Dublin Core でマッピングしたもの, 斜体で書かれているものは Qualified Dublin Core で追加されたマッピング

※(例)は静岡県の調査報告書を基に簡易に表現したものである。括弧書きは、紙面と著作権の関係上省略してある。

# 第 18 回(2010 年度)情報知学会研究大会実行委員会

|        |       |           |
|--------|-------|-----------|
| 実行委員長  | 梶川 裕矢 | 東京大学      |
| 副実行委員長 | 江草 由佳 | 国立教育政策研究所 |
|        | 高久 雅生 | 物質・材料研究機構 |
|        | 森 純一郎 | 東京大学      |

## 情報知学会誌 編集委員会

|        |       |        |
|--------|-------|--------|
| 編集委員長  | 国沢 隆  | 東京理科大学 |
| 副編集委員長 | 芦野 俊宏 | 東洋大学   |

### 編集委員

|        |              |        |                |
|--------|--------------|--------|----------------|
| 相田 満   | 国文学研究資料館     | 石井 守   | 情報通信研究機構       |
| 石塚 英弘  | 筑波大学         | 岩田 覚   | 東京大学           |
| 内田 努   | 北海道大学        | 宇陀 則彦  | 筑波大学           |
| 江草 由佳  | 国立教育政策研究所    | 大久保 公策 | 国立遺伝学研究所       |
| 岡本 由起子 | 元東京家政学院大学    | 小川 恵司  | 凸版印刷(株)        |
| 神立 孝一  | 創価大学         | 五島 敏芳  | 京都大学           |
| 阪口 哲男  | 筑波大学         | 白鳥 裕   | 大日本印刷(株)       |
| 菅原 秀明  | 国立遺伝学研究所     | 太原 育夫  | 東京理科大学         |
| 田良島 哲  | 東京国立博物館      | 時実 象一  | 愛知大学           |
| 中川 優   | 和歌山大学        | 長田 孝治  | 東京都ビジネスサービス(株) |
| 長塚 隆   | 鶴見大学         | 中山 堯   | 神奈川大学          |
| 中山 伸一  | 筑波大学         | 西川 宜孝  | みずほ情報総研(株)     |
| 西澤 正己  | 国立情報学研究所     | 西脇 二一  | 奈良大学           |
| 根岸 正光  | 国立情報学研究所     | 原 正一郎  | 京都大学           |
| 原田 隆史  | 慶應義塾大学       | 藤井 賢一  | 産業技術総合研究所      |
| 藤原 譲   | 筑波大学名誉教授     | 細野 公男  | 慶應義塾大学名誉教授     |
| 村川 猛彦  | 和歌山大学        | 安永 尚志  | 人間文化研究機構       |
| 山本 毅雄  | 国立情報学研究所名誉教授 | 山本 昭   | 愛知大学           |

### ■複写をされる方に

本誌に掲載された著作物を複写したい方は、(社)日本複写権センターと包括複写許諾契約を締結されている企業の従業員以外は、著作権者から複写権等の行使の委託を受けている次の団体から許諾を受けて下さい。

著作物の転載、翻訳のような複写以外の許諾は、直接本会へご連絡ください。

〒107-0052 東京都港区赤坂 9-6-41 乃木坂ビル 学術著作権協会

TEL: 03-3475-5618 FAX: 03-3475-5619 E-mail: naka-atsu@muji.biglobe.ne.jp

アメリカ合衆国における複写については、次に連絡してください。

Copyright Clearance Center, Inc. 222 Rosewood Drive, Danvers, MA. 01923, USA

TEL: 978-750-8400 FAX: 978-750-4744 URL: <http://www.copyright.com/>

情報知学会誌 Vol.20, No.2 2010 年 5 月 15 日発行 編集・発行 情報知学会

頒布価格 3000 円

## 情報知学会(JSIK: Japan Society of Information and Knowledge)

会長 根岸 正光

事務局

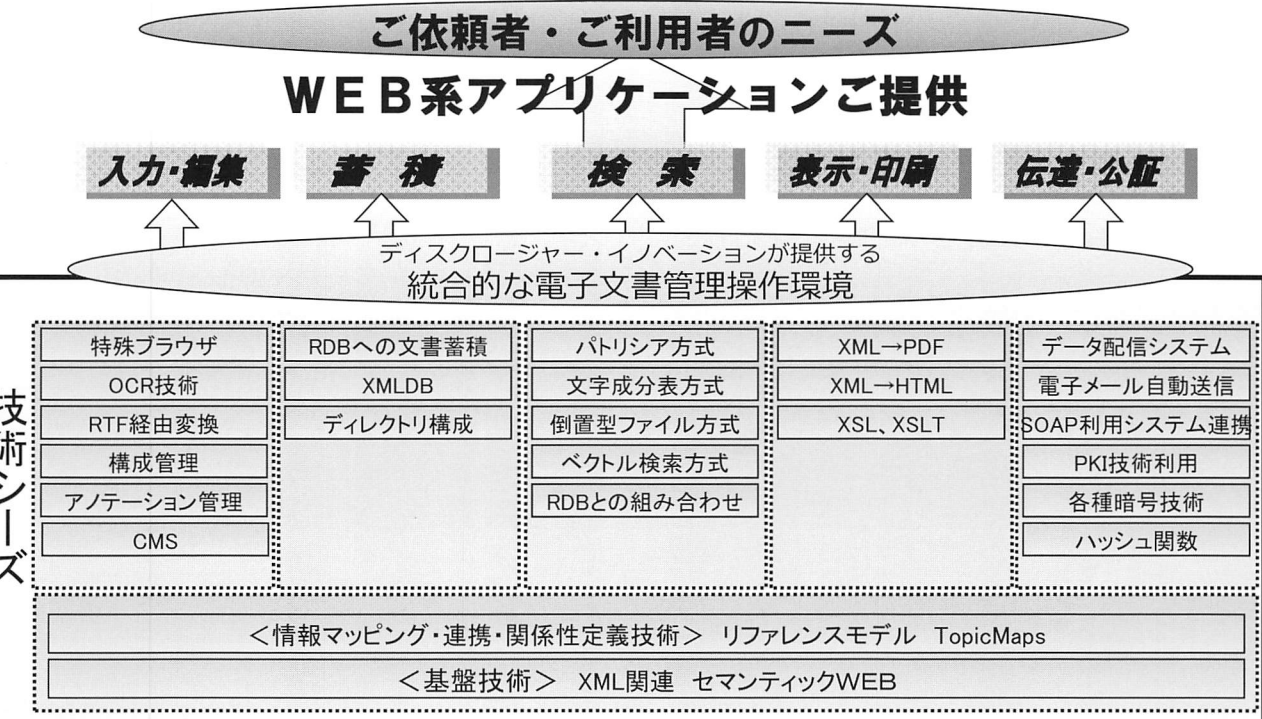
〒110-8560 東京都台東区台東 1-5-1 凸版印刷(株)内

TEL: 03(3835)5692 FAX: 03(3837)0368 E-mail: [jsik@nifty.com](mailto:jsik@nifty.com)

URL: <http://www.jsik.jp>



ディスクロージャー・イノベーション は  
電子文書 の専門家集団が運営する  
「研究開発型企业」です。



技術シース

主な事業内容

- 情報提供サービス機構の実現**

  - 大学・研究機関における情報発信・研究情報提供サービス
  - 研究情報基盤システムの構築ならびにドキュメント デリバリー機構の実現
  - 文書類のSGML・XML・HTML化関連サービス
- 電子文書を利用した社会機構の調査**

  - 地方自治体における高度情報化計画策定調査
  - Webサービス、エレクトロニックコマース、デジタルマネーの実現のための調査
  - 大学研究機関におけるバーチャルシステムズの研究
  - 地方自治体における高度情報化計画策定調査
- 電子文書関連ツール&サービス開発と販売**

  - ㈱日本電子公証機構  
電子公証サービス dPROVE、ePROVE  
電子認証サービス iPROVE
  - オントピア社 Ontopia Knowledge Suite
  - イースト社他との協業機関(Synest)による、Synest Labonote (電子研究ノート)

- ネットワーク関連(管理と開発)**

  - 研究機関におけるネットワーク構成のデザインならびに維持運用管理
  - 情報提供サービスにおけるセキュリティシステムならびに課金システムの構築
  - Webページの製作・構成ならびに維持
  - BtoB BtoCコマース機構の開発
  - コンテンツ・マネジメント構造の実現と運用
- 文書情報管理**

  - フルテキスト(全文)検索システム構築
  - 検索に関わる標準プロトコルの実装機能の設計
  - SGML・XML・HTML・PDF・TeX 等の文書形式相互変換システムの構築
  - TopicMaps・RDF・DublinCore等のメタデータを利用した情報運用のための管理システムの構築
  - 既存の紙書面による文献・資料群の電子化とデータベース化

# *Journal of Japan Society of Information and Knowledge*

## ~~~~~ Contents ~~~~~

|                                                                                                                        |                                                                                   |
|------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|
| The Analysis of Differences between Online Book Reviews and Those in Newspapers                                        | Takashi HARADA, Sawako YOSHIMURA ... 65                                           |
| Analysis of Best-Answer Estimation for a Q&A Site and its Application to Machine Learning                              | Daisuke ISHIKAWA, Kazuko KURIYAMA, Tetsuya SAKAI, Yohei SEKI, Noriko KANDO ... 73 |
| Web Site Analysis Using the Wayback Machine of Internet Archive                                                        | Soichi TOKIZANE, Tomoya SUGIURA ... 86                                            |
| A study of Global Model Architecture Information System                                                                | Tsuneo SAWA ... 93                                                                |
| Generalized Structure of Social Revolutions                                                                            | Tetsutarou YASUHIRA ... 103                                                       |
| A study How One Can Enhance One's Ability Activity to Survive Changing Environment Positively                          | Masao FUKUNAGA ... 111                                                            |
| Extracting Background Knowledge About Critics Through Numerical Analysis of Literary Reviews                           | Hajime MURAI, Akifumi TOKOSUMI ... 117                                            |
| Extracting Narrative Patterns from the Short Short-Stories of Hoshi Shinichi                                           | Chie SATO, Hajime MURAI, Akifumi TOKOSUMI ... 123                                 |
| The Sensibility Characteristics of Composers Observed within Music Criticisms                                          | Akihiro KAWASE, Hajime MURAI, Akifumi TOKOSUMI ... 129                            |
| Lexical Formation and the Historical Shifts within the Works of Haruki Murakami Identified using Quantitative Analysis | Akira KUDO, Hajime MURAI, Akifumi TOKOSUMI ... 135                                |
| IDV or Impact Deviation Value: A Proposal of a Normalized Impact Factor for Performance Evaluations                    | Masamitsu NEGISHI ... 141                                                         |
| Change in University-Industry Co-authorship Relations in Japan with Globalization of Academic Research                 | Yuan SUN, Masamitsu NEGISHI ... 149                                               |
| Investigation into Environmental Studies in JAPAN by Key-Word Analysis                                                 | Masaki NISHIZAWA, Yuan SUN ... 155                                                |
| Guideline on Selection in Heat Resistance Materials Using Information of Resources                                     | Hiroshi YOSHIZU, Mitsutane FUJITA, Kohmei HALADA ... 163                          |
| Early Detection of Commercialization Opportunity By Analyzing Scientific and Technological Landscapes                  | Naoki SHIBATA, Yuya KAJIKAWA, Ichiro SAKATA ... 171                               |
| The Structure of International Collaboration in Green Technology Research                                              | Hajime SASAKI, Yuya KAJIKAWA, Ichiro SAKATA ... 177                               |
| Cooperative Database Editor for expanding Japan MARC/A                                                                 | Takanori KAWASHIMA, Norio TOGIYA ... 183                                          |
| An approach to Full-text Search for Bibliographic Data of Ancient Documents by means of Stand-alone Operation          | Yukiharu WATAGAMI, Takehiko MURAKAWA, Keigo UTUNOMIYA, Masaru NAKAGAWA ... 189    |
| A Modeling for the Literature Studies System assuming EUC/EUD                                                          | Hiroki TOMIZAWA ... 195                                                           |
| Course Schedule Summarization Base on Viewpoint of Curriculum                                                          | Yukio HORI, Takashi NAKAYAMA, Yoshiro IMAI ... 201                                |
| Legal Information "Knowledge Structure Map" Development Responsive to Society's Needs                                  | Kei KAWASHIMA, Adam LOBEL, Taketomo YAMADA, Hiroyuki OHTAKE ... 207               |
| A Comparative Study for the Description Rules of MLA                                                                   | Yoshinori SUZUKI, Junki YAEGASHI ... 215                                          |
| A Study of Database Model for "MATSURI"                                                                                | Kei YAMANOBE, Junki YAEGASHI ... 221                                              |

~~~~~  
情報知識学会誌 第20巻2号 2010年5月15日発行

編集兼発行人 情報知識学会 〒110-8560 東京都台東区台東1-5-1 凸版印刷(株)内

TEL:03(3835)5692 FAX:03(3837)0368 E-mail:jsik@nifty.com

URL: <http://www.jsik.jp/>

(振替: 00150-8-706543)