

情報学の知見と手法で新たな分野を開拓した  
**研究・開発・実践の軌跡**

——研究生活を振り返って

**原正一郎 教授**

退職記念インタビュー



原正一郎教授退職記念インタビュー

# 情報学の知見と手法で 新たな分野を開拓した 研究・開発・実践の軌跡

——研究生活を振り返って

聞き手

馬場弘樹

京都大学東南アジア地域研究研究所／白眉センター特定助教

学術情報センター（現国立情報学研究所）、国文学研究資料館、京都大学地域研究統合情報センター、東南アジア地域研究研究所において、情報学的アプローチを活かして医療情報学から図書館情報学、人文情報学、地域情報学へと、境界領域における新たな学術分野を切り拓いて成果を上げてきた原正一郎先生。その研究と実践の軌跡を振り返っていただき、今後の人文学と情報学のコラボレーションの行方、さらには地域研究と情報学との協働の進展には何が求められるのかについて展望をうかがった。



——原先生が研究者になる前、高校、大学のころは、どんなことに関心があったのですか。

高校生のころ最初に関心を持ったのは生物学でした。今で言うバイオテクノロジーに類する話題を耳にすることも増えていました。遺伝子やDNAについての入門書も出版されるようになり、高校では徹夜でウニの発生実験をするなど、徐々に生物学への関心が高くなりました。また、当時の田舎高校では珍しかったと思いますが、プログラム電卓があって、これが今のデスクトップコンピュータくらいの大きさがある文字通りに電卓でしたが、自由に使わせてくれました。思い返してみると、アセンブリ言語に近いプログラム言語だったので決して使いやすいはなかったのですが、嬉々として円周率を計算したりハノイの塔の計算をしたりと、コンピュータに初めてのめり込んだ時期でした。その一

方で、時計を分解したり望遠鏡やラジオを作ったりと機械いじりも好きでした。さらに、科学雑誌などでは、近い将来に生物とコンピュータと機械が融合するような社会が到来する、といった記事も掲載されていました。そのような環境で育ったためか、生物と機械の境界分野を学んでみたいと思うようになりました。大学の進学先として東京大学の理科二類を選んだのも、進学先が生物系と工学系だったからです。

——東京大学に入学してからは、医学部保健学科に進まれていますね。

生物と工学の境界領域的な進学先としては、薬学、農芸化学、保健学などがありました。残念ながら、2年次の進学振り分けの際に第一希望で出したところは、ちょっと点数が足りずに落ちてしまったので、第二希望の保健学科に進学することにし

1 J.D. ワトソンらが DNA の二重らせん構造でノーベル賞を受賞したのが 1962 年。P. バーグラによる DNA 組み替え実験が 1972 年。W. ファイアーズらにより最初の完全なゲノム配列の解読に成功したのが 1976 年。

2 機械語のビット列に対応した文字列命令。

ました。これは結果的に良かったと思っています。保健学科では、統計学・生物学・医学・プログラミングの基礎を学び、国内限定でしたが訪問調査と称するフィールド調査もしました。また一ヶ月にも満たない期間でしたが、タイの農村（まだ電気も水道もなかった）やマレーシアを訪ねる機会にも恵まれました。これらの経験は、東南アジア地域研究所（以下、東南地域研）に来てからとても役に立っていますから、終わり良ければ全てよしですね（笑）。

——原先生が学生のころは、大学院に進学して研究者をめざす人は多くなかったと思いますが、そのなかで修士・博士課程に進もうと思われたきっかけは何でしょうか。

修士課程に進もうと思ったのは、大学で4年間を過ごした先をどうしたらいいか迷っていたからです。そこでモラトリアムを決め込んで、とりあえず大学院に進むことにしたのですが、保健学科の修士課程に進むかどうかについてはかなり悩みました。疫学や生物化学などは好きな科目で先生方とも仲がよかったのですが、保健学は社会的な色合いが強くてあまり馴染めず、そちらの先生方とは何となく折り合いが悪かった（笑）。だから、このまま保健学を専攻しても大丈夫かなと思案していた時に、当時の医学部医用電子研究施設から誘いがあり、ひとまずは研究施設のテーマで卒業論文を書くことになりました。

医用電子研究施設には、臨床医学と基礎工学と基礎医学という三つの部門がありました。臨床医学部門では人工臓器の研究をしており、1980年当時、ヤギを使った体外型完全置換型人工心臓の実験では世界最長生存記録を更新し続けていました。基礎工学部門では電気メスなどの医療機器や生体センサなどの先端装置の研究をしていました。私に声をかけてくれたのは基礎医学部門で、医療データ解析やシミュレーションさらに非観血的な生体計測法の研究などを行っていました。私の関心とも合っていたので基礎医学部門へ進むことにしました。この部門の教授は医師でしたが、研究内容は工学的でした。工学の勉強をしないとまずいと考えて、修士課程は東京大学ではなく、筑波大学の医科学研究科の医工学研究室に進みました。そこは工学系

の研究室だったので、2年の間に数学をやり直ししながら、回路や制御や通信などの基礎を学んで、それから博士課程で東京大学の医用電子研究施設に戻りました。博士課程に進んでからは「センサ回路を作れ」とか「循環系の動態モデルを作れ」などと言われたので、筑波大学での2年間はとても有益でした。

## ■人工知能、情報連携研究の端緒 ——診断エキスパート・システムの研究

——博士課程に進んでからは、具体的にどんな研究をされていたのですか。

当時の医学部では、研究テーマの選択に関しては教授の意向が強く、私には輸液自動診断の研究が割り当てられました。実は卒業論文も同じテーマでした。簡単に言えば、体内の水と電解質の量を正常に維持するために、最適な点滴の量と静脈からの投与速度をコンピュータに判断させようというものです。私が取り組んだのは、どれだけの水や電解質が不足しているかを患者の症状から推定すること、それを補正するための輸液計画を作ること、それを静脈から投与する際に体の負担にならない速度を決定すること、これら3種類のアルゴリズムを開発してプログラムを書くことでした。

この研究に没頭していた時期は1980年代半ばで、第二次人工知能ブームのまっただ中でした。ちなみに、第一次人工知能ブームは1950年代に始まりました。推論と探索のアルゴリズム研究が中心で、チェスなどのゲームや数学の定理証明などを賢く計算させることが主題でした。第二次人工知能ブームは1980年代に始まり、第一次人工知能よりも柔軟な知能システムを作るために、人間の知識をコンピュータに取り込んだ推論の研究が中心でした。専門家と同程度の推論力を目指していたので、エキスパート・システムとも呼ばれていました。第三次人工知能ブームは2000年ころから続っていますが、ビッグデータを使って機械自身が知識を獲得する機械学習や、知識を定義する要素を機械自ら習得するディープラーニングなどが登場しています。

つまり、私が取り組んでいた研究は、輸液診断用のエキスパート・システムを作ることでした。研究は分厚い輸液の専門書を読んで必要な知識を取

3 出血を伴わない処置。超音波検査などが該当する。

4 溶媒中に溶解した際にイオンに電離する物質のこと。この研究ではナトリウム、カリウム、カルシウムなどが対象。

5 どんな製品つまり輸液バッグやボトルを組み合わせたら不足している水や電解質の量を実現できるかを計算すること。

り出すことから始まりました。ここからが本番で、まず取り出した知識をコンピュータが使える形に手作業で変換しました。次に、変換した知識を貯めるデータベースと、データベースに貯めた知識を使って推論を実行するプログラムを作りました。専門家の知識を貯める仕掛けを「知識ベース」、推論する仕掛けを「推論エンジン」と言って、この二つを組み合わせてエキスパート・システムとなります。ところで、第二次人工知能ブームでは論理に基づく推論が主流でした。推論とは「風が吹けば桶屋が儲かる」ことを論理的に証明することです。そのため、知識ベースは論理式の集まりで、推論エンジンは定理証明器でした。私が4年間続けた研究は、専門家と同じレベルの診断と輸液計画を作り出す定理証明器の開発であったわけです。人工知能というと煌びやかですが、実際の作業は地味なもので、専門書をひたすら読んで、知識を論理式に書き換えて、知識ベースを充実させていくという泥臭い仕事の連続でした。また推論エンジンの作成も面倒で、新しいタイプの知識が加わるたびにコードの書き換えや追加の作業が必要となって、コードやバージョンの管理が大変でした。

作りあげたエキスパート・システムが実際に使えるかどうかを確かめる必要もありました。そこで、専門病院からカルテを借りて、カルテを読んで症状を抜き出します。ただし、昔のカルテは手書きです。しかも、たいていは英語で、ドイツ語のものも混じっている。その読みづらい字を読み解いて患者データを作るというのは、まるで杉田玄白が『解体新書』の翻訳をしたときのような作業でした(笑)。そうやって作った患者データを私が作ったエキス

パート・システムに与えると、推論処理を進めて処方を出力します。それをカルテに書かれている専門医の処方と比べます。両者の処方に大きな違いがある場合、まずエキスパート・システム側の推論過程を調べます。そして「コンピュータは、このような推論の過程を経て、このような結果を出しました」と専門医に説明します。これに対する専門医の推論と診断過程、あるいはコンピュータの推論過程でおかしいと思われる部分の指摘をインタビューにより収集し、新たな知識とします。その知識を論理式に書き換えて知識ベースに追加し、あわせて推論エンジンも修正する、その作業を繰り返します。ところが、専門家の知識というのは意外にアバウトで、論理式に書き換えることが難しい知識もかなりありました。また専門家が異なると知識も異なる、つまり症状と処方は同じでも推論のパスが違うという事例も少なくありませんでした。その辻褄を合わせるために、知識ベースの修正とプログラムの調整を手作業でひたすら続けました。そうやってうまく調整できたと思った瞬間、それまで順調だった部分がおかしくなって、さらに再調整する、これが延々と続いたわけです。

このように、エキスパート・システムでは、知識を人間が記述する必要がある、知識の量が膨大になる、知識同士での矛盾が発生しやすい、曖昧な事柄に対する推論が難しい、などの問題点が明らかになってきました。同じころに開発されたエキスパート・システムとして、原子炉の自動故障診断や会社の苦情相談などがありましたが、やはり同じ問題を抱えており、そのために第二次人工知能ブームは終わりを告げることになりました。幸い、ブームが下火になる直前に、何とか博士論文を仕上げた博士課程を修了しました。

この博士課程での研究を通じて、いくつかのアイデアを思いつきました。まず、専門医から知識を聞き出して矛盾なく整理するのは大変なので、カルテの情報を使ってコンピュータに勝手に学習させようというものです。これは現在のビッグデータを使った機械学習と同じ発想です。オーバードクターの時に挑戦したのですが、カルテの電子化が進んでいなかったことで知識を自動的に取り込めなかったことと、学習プログラムが当時のコンピュータ能力ではほとんど動かず使い物にならないことがネックに

6 AならばBである、BならばCである、よってAならばCであるなど。



なって断念しました。もう一つは、体内の水と電解質の量を推論するのではなく、直接計測してしまうというものでした。これについては、計測装置を作る時間がなかったことと、モデルは単純であったものの当時の私には計算方法が分からなかったので断念しました。しかし、これをきっかけに医療情報システムに興味を持つようになり、それがデータベース研究へとつながり、関連した研究が30年ほど続いてきたということになります。地域や職域における集団健診データの集約と健診指導の支援を目指したシステムの開発は、その初期の研究成果です。

余談になりますが、当時から現在まで、人工知能という言葉はあまり使っていません。そもそも「知能」の定義あるいは概念が曖昧であったことと、マスコミが「人工知能」で騒ぎすぎていたからです。むしろ研究室では computer decision making あるいは computer-supported decision making と呼んでいました。こちらの方が研究室の研究内容に合っていました。振り返れば私も computer decision making に関する研究を続けてきたように思います。知識ベースの中身は論理式からビッグデータに、推論エンジンは定理証明器から機械学習アルゴリズムに変わりましたが、フレームワークは変わっていません。就職してから約25年は知識ベース寄りの研究、最近の数年は推論エンジンの研究に軸足が変わってきたのかもしれない。

## ■学術情報センターでの経験 ——全文データベース・システムの構築

——1990年に、現在の国立情報学研究所の前身である学術情報センターに赴任されました。ここではどんなことに取り組みされたのですか。

当時の学術情報センターのメインのミッションは、日本全国の大学図書館が所蔵する文献類の総合カタログ化、ようするに大学図書館の目録の統合でした。私より上の世代では、他大学の図書館にどんな文献があるかは、実際に図書館を訪ねて目録を見ないと分かりませんでした。その後、全国の大学図書館がネットワーク化されたので、自分の大学にない文献がどこの図書館にあるかが容易に検索できるようになりました。次にインターライブラリ・ローンが動き始めて、所蔵大学へ出向かなくと

も文献を入手できるようになりました。この取り組みを進めていたのが学術情報センターでした。

私が採用された当時の目録データベースはキーワード検索が中心で、「入力したキーワードに合致するタイトルや作者の文献」についての書誌と所在の情報が表示されるだけでした。文献の中身を見ることはできないし、語彙が適切でなければ検索は失敗します。たとえば、キーワードとして「コンピュータ」が登録されている文献は、「計算機」では検索に失敗します。もし文献のテキスト本体を検索対象にできれば、さまざまな語彙で検索できるし、文献本体を見ることもできる。これが全文データベースと言われるもので、その開発に従事しました。

ところで、少々専門的な話になりますが、文献は最初にタイトルや著者名などが書かれていて、それらに続いて幾つかの章が続きます。また章は章タイトルの後に幾つかの節が続く、といった構造があります。つまり、全文データベースを作るには、文献の文字列に加えて、テキストの構造を記述する必要があります。そのころ、テキストの構造を記述する国際規約としてSGMLが制定されたので、それを使ってテキストをマークアップすることになりました。マークアップは人海戦術で行いました。

一方、データを貯めて検索するデータベース・システムですが、当時は、今でもそうですが、関係データベースが主流でした。Microsoft社のAccessなどは関係データベースの代表例で、直感的には表のイメージです。この表において、各データは項目分けされていて、たとえば「著者」という項目には値として著者名が入っています。さらに専



集団健診支援システムの構築

7 interlibrary loan, ILL、図書館間相互貸借。

8 Standard Generalized Markup Language, ISO 8879。

9 markup、テキストに構造情報を付加する作業。タグ付け。

門的になりますが、一つの項目に書ける値に構造があってはいけません。これを第一正規形といいます。すると、テキストには先ほど述べた構造があるので、関係データベースでは扱えません。そこでマークアップされた SGML テキストの入れ物と、それを検索する仕掛けを作る必要があったわけで、それが私に与えられた仕事でした。

てこずったのは検索です。関係データベースであれば SQL などの検索言語が発明されています。しかし、これらは全文検索には使えません。当時、SGML テキスト用の検索言語はありませんでした。そこで、同じ開発グループの先生が DQL なる検索言語を発明しました。ただし本人は仕様書を作っただけで、「実装は君がしたまえ」と言われました(笑)。何をしたかについては専門的すぎてここで述べるのは難しいですが、SGML テキスト用のパーサを作りました。これは SGML テキストを分析して、「○○というタグが付いたテキストの中身は何か」、「そのタグの属性は何か」などの情報を取り出す仕組みです。一方、DQL を使えば、「○○というタグが付いたテキストに△△という文字が含まれている文献の Author を表示せよ」といった検索命令を書くことができます。そこで、DQL の指示と SGML パーサの分析結果を照合して、指定された条件を満たす文献中の文字列を出力するプログラムを書きました。使ったコンピュータは、当時普及し始めていた Unix ワークステーションで、そうなると思える言語は C でした。データベースやパーサについては、オーバードクターの頃に一通り勉強していたので助かりましたが、ワークステーションと C 言語は使ったことがなかったものですから、いま思うとよく 2 年であんな仕事ができたと自分でも思います(笑)。そのシステムが出来上がったところで国文学研究資料館(以下、国文研)へ異動となりました。

## ■国文研での使命をめぐる労苦——漢字を扱うデータベース・システムの更新

——学術情報センターから国文学研究資料館(国文研)に移られたのは 1991 年ですね。同じようにデータベース関連の仕事に携わられたのだと思いますが、研究内容は変わりましたか。



かなり変わりました。扱う情報は医療や自然科学から古典になり、データも数値から文字になり、カウンターパートも文学者になったわけです。とくに国文研では古典を扱っているわけですが、そもそも私は昔から国語が苦手で、ましてや古典となると解釈どころか字が読めない(笑)。さらに、高校は理数科で、高校・大学・大学院では文系との付き合いはほとんどなかったのが、国文学者の気質や研究指向性についてまったく知らなかったわけです。

——データベースの構築という大枠や方向性は同じでも、入れるデータも仕事の相手もまったく違う状態になったわけですね。

そうですね。データというか字が読めない、古典文学作品には興味はあっても研究内容にはあまり興味を持ってなかったという問題は決して小さくはなかったものの、コンピュータを使ってデータベースの仕事をするという部分は同じなので、「何とか」こなしてきたという感じでした。

国文研で取り組んだ最初の仕事は、電子目録データベースの更新でした。国文研の創立は 1972 年で、大学共同利用機関としては最も古い組織でしょう。設立当初から電子目録を作っていて、当時としては世界的にもかなり先端的な人文学の研究機関だったと思います。その設立時に作られたシステムが、私が異動してきたころまで、十数年にわたって生きながらえていました。さすがにコンピュータにガタが来て、ソフトウェアも時代に合わなくなっていたので、これを作り直すのが私のミッションでした。

そのころ、大学や研究機関などのサーバはいわ

10 Microsoft Access ではセルに書き込む何か。

11 項目の繰り返し、値の連結、入れ子など。

12 ISO 9075、JIS X 3005。

13 Document Query Language。

14 parser、構文解析器。

15 例：Authors タグの中身は Shoichiro Hara。

16 例：Authors タグの中身は英語で書かれている。

17 例：Authors が Shoichiro Hara。



ゆる大型計算機と呼ばれるものでしたが、徐々に Unix サーバに置き換わる時期でもありました。ですからプラットフォームとしての大型計算機を捨てる決断に迫られました。それに加えて、当時のコンピュータ・ネットワークは、NEC とか日立とかベンダーごとに違っていました。これもインターネットへの移行期だったので、ネットワークを切り替えて、あわせて国文研の館内配線を敷設し直さなければなりません。さらに Web のホームページが普及し始めており、ユーザ・インターフェースについても全面的に再設計することになりました。

やっかいだったのは、元のデータベース・システムが、関係データベースよりも古い時代の「ネットワーク型データベース」というものだったことです。しかも、最初の仕様書はありましたが、その後の修正や更新記録はほとんど残ってなくて、データ構造の現状がどうなっているのか分かりませんでした。そこで、まずはデータベース内にあるデータをバイナリ・ファイルという数字のみで記述されたファイルとして取り出して、それを 1 バイトごとに仕様書と見比べて、仕様と違うところ、仕様書にはないがデータとおぼしき数字列がある部分などを見つける作業から始めました。仕様書と合わない部分が見つかったら、その数字を手作業で文字に変換して、古くからいる図書館職員に聞いて、そのデータが何か教えてもらう。この作業を繰り返して、新しいデータ仕様書を作りました。その仕様書に従って、正しいバイナリ・データ部分を取り出して、それを文字に変換して、さらに SGML に変換するサ

ルページ・プログラムを作りました。SGML を使った理由はデータのポータビリティを担保するためです。バイナリ・データのままで、次にプラットフォームを更新する際に、今回と同じ変換作業の繰り返しとなるので、それを避けたかったからです。せっかく SGML データにしたので、学術情報センター在職時の経験を活かして全文データベースとしました。日本の人文学の研究機関としては、かなり先進的なデータベースであったと思います。

——気の遠くなるような、たいへんな作業ですね。

そうですね、出来上がるまで数年以上はかかりました。このときもう一つ苦労したのは漢字です。当時、コンピュータで利用できる漢字コードは、JIS 第 1 水準から第 4 水準までの約 1 万文字ですが、日本の歴史や古典のテキストを扱おうとすると、5 万文字は必要であると言われていました。多くは人名や地名などで使われる漢字です。どうしても使いたい漢字が JIS コードにはない場合、その漢字は自作するしかないわけで、これを外字と言います。国文研の古いデータベースでも沢山の外字を使っていましたが、これをどうやって新しいコンピュータに移植するか、にははなはだ悩みました。悩ましかったことのひとつが JIS コードの変遷です。国文研の古いデータベースは JIS78 という日本で最初に制定された漢字コードを使っていました。これが JIS83、JIS90 と変遷する中で、コードの入れ替えや字形の変更が行われました。その結果、コンピュータは JIS78 対応なのにプリンターが JIS83 対応だと、同じコードでありながら画面とプリンターで違う字が表示・印字されるということになりました(笑)。結局これは、「JIS78 のコードは JIS80 ではこのコードに変更された」という JIS コードのすべての変遷をトレースするプログラムを作って修正しました。2000 年ごろに Unicode に統一しました。話が前後しますが、多くの外字は Unicode に含まれていたため、この問題は解消しました。Unicode へ移行する作業でやっかいだったのは、既に作られていた外字が Unicode にある似た形のどの漢字に対応するかの判断で、これは専門家に助けを求めました。

18 platform、オペレーティングシステムやハードウェアといったコンピューティングの基礎部分。

19 portability、特定のプラットフォームやアプリケーションに蓄積してきたデータを他のプラットフォームやアプリケーション

ションへ容易に移動・移行できること。

20 JIS C 6226-1978。

21 JIS C 6226-1983。

22 JIS X 0208-1990。

## ■資源共有化システムと古典文学大系テキスト・データベース

——国文研では約15年にわたって所属して研究活動をされていますが、他にはどのようなことに取り組まれたのでしょうか。

国文研で取り組んだ別の大きな仕事は「資源共有化」です。ここでは、ネットワーク上に分散しているデータベースの統合検索を目指しました。データベースの情報を共有するには、データ項目や、項目の名称や、データの書き方などの標準化が必要です。しかし、そもそも「新しいことをしなさい、これまでになかったことをしなさい」と言われて育った研究者に、標準化はそぐわないんですね。たとえば、書誌目録を作るにしても、研究で作る目録には何らかの特別な目的があるわけで、研究者や研究課題ごとに構造は違ってきます。ですから、ある目録では抜けている部分が、別の目録ではとても詳しく記述されている、ということがあるわけです。極端に言えば、研究者の数だけ異なったデータベースが存在します。この状況は、利用者にとってデータベースを一つずつ検索するには面倒だし、データ項目やデータの書き方がバラバラなので使い勝手も悪いから、データベースの統合にはメリットがあるでしょう。一方、統合しようとする、データベース作成者にはデータベース作成の目的の違いや思い入れがあるので、当然のこととして反発がくる。そこで「資源共有化」では、各々のデータベースとは独立した「仮想」のデータベースを作りました。そして、各々のデータベースのデータ項目を仮想データベースのデータ項目に紐付けるようにしました。利用者は仮想データベースの検索を経由して、個別データベースを間接的に検索する。こうすれば、データベース作成者はデータベースを作り直す必要がなく、利用者は各々のデータベースの所在やデータ構造が分からなくとも検索できる。1999年ごろのアイデアです。

まずは国文研内のデータベースの共有を試しました。また当時は大阪市立大学におられた柴山守先生が興味を持っておられたので、次に大阪市立

大学との共有を試みました。同じころ、国立大学や研究所の法人への転換が進み、国文研は人間文化研究機構という組織に組み込まれ、この時に「資源共有化」が機構のプロジェクトになりました。国文研の「資源共有化」で作った仕様書やプログラムが、この機構のプロジェクト、そのあと京都大学で展開した「資源共有化」の骨格になっています。

国文研でのもう一つの仕事が、岩波書店の旧版「日本古典文学大系」という、約100巻、566作品の全文データベース化です。このプロジェクトは、私が国文研に異動する前から始まっていて、国文研に異動したころデータは出来上がりつつありました。問題は、このデータを貯めるデータベースも検索ツールもなかったことで、その構築を任せられました。テキストデータなので、ここでもSGMLを使うことにしました。しかし、このプロジェクトが開始されたのはSGMLが普及する前で、またTEIという人文テキスト用の標準マークアップも制定途上であったため、独自のマークアップを使っていました。私が異動した少し後ごろにTEI-P3が制定されましたが、日本の古典を扱うには少々難があったので、SGMLテキスト化に際しては、TEIを使わずにKOKINルールの独自タグをそのままSGMLテキストに反映させることにしました。

ところが、このKOKINルールの定義が曖昧なもので……(笑)。専門用語で言うと、文脈自由文法のつもりであったのに、実際は文脈依存文法のレベルでした。普通のプログラム言語やSGMLなどは文脈自由文法なので文法レベルが違うわけですから、通常のパーサを使って処理することができない。さらにデータは手作業で作られていましたから誤りも多かった。簡単にSGMLテキストに変換できるシロモノではなかったということです。結局、まずはエラー検出プログラムを作って仕様と異なる記述上のエラーを探し出して、これらを人手で直しました。次に、文脈依存の部分にタグを追加して文脈自由に変換するプログラムを書いて正しい文脈自由文法レベルのテキストを作り、最後にKOKINルール用のパーサを作ってSGMLテキス

23 文献の主題名と副題名を「title」の中に押し込んで書くものや、「title」と「subtitle」に分けて書くものなど構造が異なる。

24 例：題名が「title」や「書名」のように異なる。

25 例：日付が30th January 2023や2023-01-30のように異なる。

26 Text Encoding Initiative。

27 KOKINルールと呼んでいました。↗

<https://www.dlib.org/dlib/july97/japan/07hara.html>

28 形式的に書くと  $V \rightarrow w$  となる。つまり文脈自由とは、語彙  $V$  の前後関係に依存せずに語彙  $V$  を別の語彙  $w$  に置換できる。

29 形式的に書くと  $a V \beta \rightarrow a \gamma \beta$  となる。 $a V \beta$  は語彙  $V$  の前後の語彙  $a$  と語彙  $\beta$  を意味している。つまり文脈依存とは、 $V$  の前後の文脈によって語彙  $V$  を語彙  $\gamma$  に置換できるかどうかを判断する。



トへ変換しました。この SGML テキストから、さらに Web 表示用の HTML ファイルと、印刷用の PDF ファイルを作成しました。詳細は述べませんが、この部分の作業もけっこう時間と手間がかかりました。作ったデータは、国文研のホームページからアクセスできます。

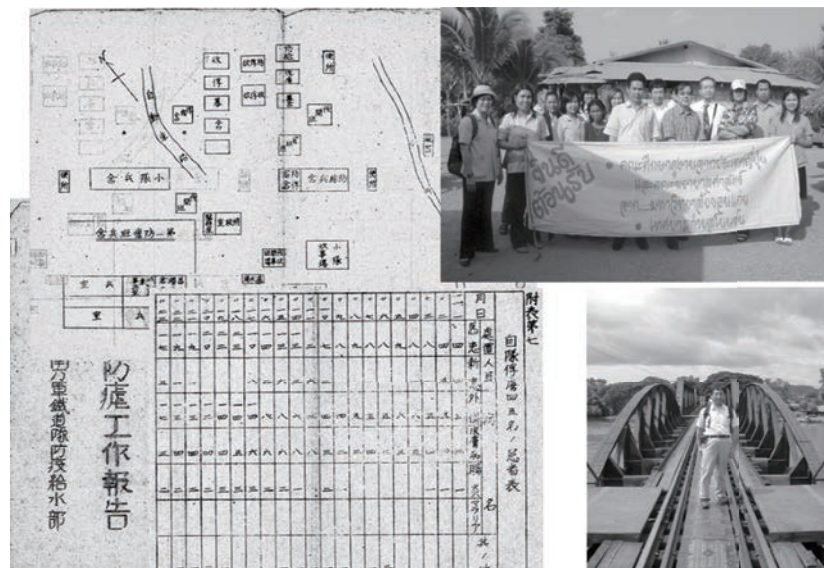
## ■地域研究統合情報センターに着任して ——「地域研究」をめぐる戸惑い

——その後 2006 年に、新設された京都大学地域研究統合情報センター（京大地域研）に異動になって東京から京都に移られたわけですが、このときはどんな変化がありましたか。

いろいろありますが、研究以外では環境の変化が大きかったですね。まずは気候です。京都に移って 15 年が過ぎましたが、いまだに冬の空には馴染めません。東京の冬だとほとんど雲がない晴れの日が続きますが京都はそうではない。晴れてもすぐに曇ったり雪が降ったりすることが多く、冬の景色は圧迫感があります。また、地名が難しいことにも参りました。たとえば「帷子ノ辻」なんてどう読んだらいいのか、間違っって読むと周囲が怖いし（笑）。よくジョークになるけど「一口」と書いて「いもあらい」とか、京都は本当に難字地名の宝庫ですね。加えて土地勘がないことにも悩まされました。今では、比叡山を見ればどちら側だか分かりますが、田の字の直線道路があだになって、どちらを向いても同じに見える。引っ越してしばらくは、地下鉄から地上に出た時に「いったいどこを向いているんだろう」という状態になることもありました（笑）。

研究にまつわる最初の困惑は、やはり「地域研究」ですね。何だかよく分からなかった。実は国文学もそうでしたが。国文研に異動する前に国文学の本を買って勉強しようと思ったのですが、そもそも「国文学」についての本が見当たらない。ようやく放送大学の教科書を見つけたのですが、中身は『源氏物語』とかの各論が載っているだけで、国文学が何であるかは書いてありませんでした。それと同じように地域研究も何だかよく分からなかった。実はいまだによく分かっていないのですが（笑）。

私が異動してきた京都大学地域研究統合情報センター（以下、京大地域研）は、設立直後であったた



めか、研究発表の時に「私の地域研究は……」という枕詞がよく使われていました。情報学分野では、例えば通信系、データ系などと分野は違っていても、「私の情報学は……」という人はいなかったもので、これはいささかカルチャーショックでした。様々なドメインの研究者が集まっていて、方法論も違うからそう言わなきゃいけないのかなと解釈していたので、何かの席で「地域研究はドメインじゃありませんよね」と言った瞬間、とある偉い先生に「地域研究はドメインです！」と怒鳴られて……。それ以来その用語は使わなくなりました（笑）。

結局、いまだに分かっていない部分も多いですが、地域研究は、地域を総合的に俯瞰する学問で、さまざまな学問の乗り入れる境界領域、たとえば空母みたいなもので、いろいろなファンクションが載っている「容れ物」かなと捉えています。

## ■タイ、マレーシアの研究者との協働と 学内研究者との連携

——京都に移ってからこれまでで、もっとも印象に残っている研究や活動は何でしょうか。共同研究や技術支援などもあったかと思いますが。

データベースの構築以外で、京大地域研や東南地域研内での共同研究の経験はあまりなくて、主にタイとマレーシア、それから台湾と米国において、図書館情報学や保健・看護学や人文情報学の研究者と共同研究をすることが多かったですね。それぞれの地域で同じような研究を進めている人々との共同作業というのは、とても刺激的な経験でした。たとえばタイでは、地域看護活動に関連して実施されている地域住民の健康調査結果をデータベース化するプロジェクトに参加して、データベースの設

30 Hyper Text Markup Language、World Wide Web において Web ページを記述するために用いられる。

31 日本古典文学大系本文データベース <https://base1.nijl.ac.jp/~nkbthdb/>

32 humanities computing、私は人文情報学とデジタル・ヒューマニティーズ (digital humanities) を区別せず、人文情報学を好んで使っているが、以下ではデジタル・ヒューマニティーズとする。

計作業を分担しました。そのおかげでタイ語のできない自分では行けないような町や村落を訪ねることができたのは楽しい思い出です。この調査はタンボンと呼ばれる集落ごとの悉皆調査で、都市部を除くタイ国内のほとんどの集落を網羅しているのです、学術的にもとても興味深いデータです。困ったことは、こちらの英語力が低いためか、現地の打ち合わせで「こうしよう」と互いに納得して帰国したはずなのに、後から送られてきたデータを見た瞬間、「なぜこんなデータが……」と目が点になることが何回かあったことです(笑)。

印象に残っている学内の研究活動としては、まだ生々しい部分もありますが、京都大学研究連携基盤の「学知創生ユニット」です。研究連携基盤は、研究領域を横断して新しい学術分野の創生をめざした、京都大学の附置研究所・センターによる連合体で、私が京大地域研のセンター長をしている時に設立されました。その中に学際的研究分野への試みを行うための未踏科学研究ユニットが4つ設置され、そのうちの一つの「学知創生ユニット」の運営を担当しました。このユニットでは、京都大学の学術資料を「学知」として高度利用するデータベースの開発と、それを使った研究の推進に取り組みました。

ユニットの代表としての活動は貴重な経験となりました。また、京大地域研のセンター長に就任したことで、京都大学内の他の部局の先生方と会話する機会が一気に増えたことも刺激的で印象深く、その後の研究の展開においてありがたい資産となりました。私はどちらかと言うと出不精で、部屋にこもって本を読んだりプログラムを書いたりしているほうが好きなのですが、定例の部局長会議などに否応なしに出席させられたおかげで(?)、農学研究科、医学研究科、エネルギー工学研究科などの先生方と意見を交換する機会が増え、興味深いアイデアに出会ったり、新しい繋がりを作ったりすることができました。とくに私は情報学が専門なので、学術情報メディアセンターの先生方との関係が太くなりましたが、これが、いま進めているビッグデータ研究のきっかけになっています。

## ■地域研究者自身が簡易に作り・使える「Myデータベース」

——京大地域研では、地域研究者が自らデータベースを構築して使うことができる「Myデータベース」という仕組みを開発・実装されましたね。

国文研のデータベースのところでも話しましたが、研究データベースの特徴はその多様性です。地域研究もその例外ではありません。もちろん、データベース・システムを運営する側から見れば、たとえば目録データベース・システムを一つだけ作って、そこに皆さんがデータを入れてくれればとても楽です。しかし、そうはいかないので、京大地域研のデータベース・システムでは、データの多様性に対応する必要がありました。

データベースにまつわるもう一つの問題は、「どのような条件を満たすデータならデータベースに使えるか」という、いわば情報リテラシーをデータ作成者があまり持ち合わせていないことでした。京大地域研でデータベース・システムの開発を始めたころ、皆さんが手持ちのデータの多くを提供してくれました。ですが、「Excelファイルになっているから、どうかなるでしょう」といった程度のデータの「作り」だったので、そのままデータベース化できたものは一つもありませんでした。たとえば、表の中に別の表が入っている、あるいは同じデータ項目が続くなどが典型です。これも前に述べましたが、第一正規形に反するデータ構造です。人間には便利でもデータベース・システムでは扱えません。また、項目名を2行にわたって書く、あるいは項目名を英語で書きたいために単語の間にスペースを入れるといった事例もありました。このような書き方も、多くのデータベース・システムでは許されていません。他にも、数字と文字が混在するとか、データ型がグチャグチャという場合も多々ありました。結局データファイルごとに手作業で修正したり、修正プログラムを書いたりすることになりました。問題は、一つのデータベースで問題を解決できたとしても、また別の研究者が似た問題を含んだデータを持ってくることです。そこで京大地域研のデータベース・システムでは、全ての例外事項に対応することは不可能としても、ある程度の例外には融

33 原正一郎「AIやビッグデータを活用する新しい地域情報学への試み」(たんけん動画 地域研究へようこそ [https://onlinemovie.cseas.kyoto-u.ac.jp/movie\\_hara/](https://onlinemovie.cseas.kyoto-u.ac.jp/movie_hara/))も参照。

34 例: Authorという項目を予め10人分用意する。

35 例: Title of Authorsなど。せめてTitle\_of\_Authorsとか

TitleOfAuthorsとして欲しかった。

36 例: 整数の1と文字の“1”は別物。文字列の場合は“01”と“1”は別物。

37 例: 整数の1と実数の1.0は別物。

通が利くように工夫しました。具体的には、同じ項目名が繰り返されている、項目名に空白が入っている、ID が付与されていないデータでも扱えるようになっていきます。データに対する厳しい条件を満たさなくても、ある程度の水準で作られているデータであればデータベースに登録できる仕掛けが「Myデータベース」です。説明は省きますが、「データ定義文が不要」というのも「Myデータベース」の大きな特徴です。

これが「Myデータベース」の機能の半分にあたる「データを作る機能」です。残り半分は「データを使う機能」です。研究者はデータの使い方に凝るといえるのか、研究の目的に応じた使い方があります。私も情報のサービス・プロバイダの立場と研究者の立場とで態度が変わります。研究者の立場では、ユーザ・インタフェースを頻繁に変更することがあります。このユーザ・インタフェースですが、かつての情報システムでは、それなりの量のプログラムを書く必要がありました。幸いなことに、現在では Web ホームページがユーザ・インタフェースの土台になっているので、表示部分の作成はかなり簡単になりました。HTML の知識があれば誰にでも書けるようになりました。一方、データベース・システム側の操作は相変わらず複雑なまます。そこでユーザが利用できるデータベース・システムの機能を限定し、それらの機能を単純な手続きで利用するための「約束ごと」を定義しました。これを API と呼びます。「Myデータベース」にも専用の API を用意しています。こうしておく、簡単な検索命令を API の約束に従って組み上げて「Myデータベース」に送信するだけで、「Myデータベース」から書誌データや画像データを受け取ることができます。後は、受信したデータを使ってホームページを書くだけで、お望みのユーザ・インタフェースが完成です。API は、現在では当たり前の機能ですが、十数年前としてはけっこう新しい機能でしたね。

## ■情報学が人文学に浸透することで拓ける可能性

——第三次人工知能ブームが盛り上がりを見せていますが、これから情報学がどのように人文学に浸透していくか、また人文学と情報学との融合がどのように進むか、その可能性についてお聞かせください。

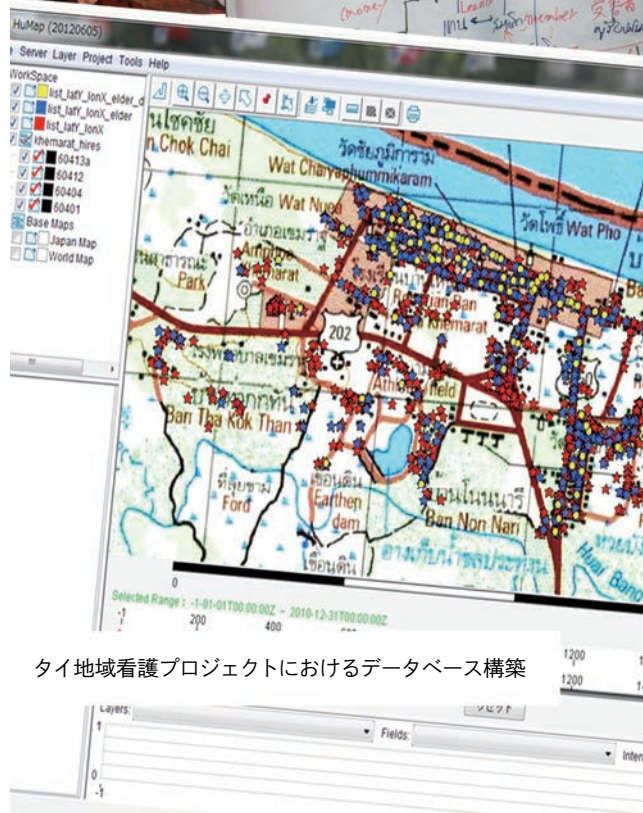
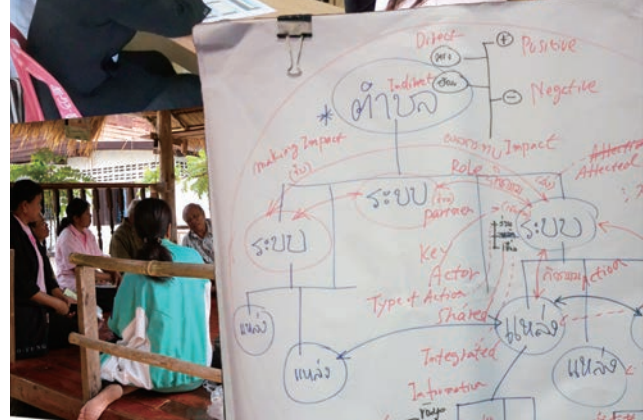
「融合」と言えるのか分かりませんが、シナジー効果により新しい研究分野が生まれることはあり得ると思います。たとえば、私が大学院を修了するころ、CT や高性能センサや第二次人工知能ブームにおける医療エキスパート・システム開発などが契機となって医療情報学会が設立され、今では医学における一研究分野となっています。人文学ではデジタル・ヒューマニティーズですが、最も初期の事例は 1946 年まで遡ることができると言われており、現在では Alliance of Digital Humanities Organizations などの国際的組織が形成され国際論誌も発行されています。このように考えると、人文学と情報学とのさらに新しい融合が進む可能性はあると思います。

人工知能あるいは機械学習の応用ではありませんが、コンピュータ

38 Application Programming Interface.

39 Computed Tomography、コンピュータ断層撮影。

40 Roberto Busa による Index Thomisticus (<https://www.corpusthomicum.org/it/index.age>) の構築。



タイ地域看護プロジェクトにおけるデータベース構築

を使うことで、それまでは困難と思われてきた分析が可能となった事例は、国文学においてもあります。たとえば、和歌には「本歌取り」といって、新しい歌を作る際に昔の歌の一部を取り込むという技法があります。昔の歌を入れることで、より複雑で豊かな表現を意図していますが、それを発見することも国文学では一つの研究対象のようです。和歌の文字列分析に生物情報学などで利用されるパターンマッチングの技法を応用して、本歌取りの部分を発見したという論文があります。また、語彙の出現パターンをコンピュータで統計的に処理することで、古文書の真贋判定を行った研究例もあります。このような事例は確実に増えると思います。

ただし、「人文学研究に本当にコンピュータは必要なのか」についてはもう少し考察が必要だと思います。文献の整理などにコンピュータを使うとしても、解釈が中心の研究であれば、コンピュータは必要ないのかもしれませんが、むしろ、こちらの方が人文学研究の本流ではないかと思うこともあります。となるとデジタル・ヒューマニティーズは表層的な人文学研究となってしまいますが、でも、コンピュータが文学作品を書くようになったら、人文学研究は激変するかもしれませんね。これはとても興味深いことです。ということで、「コンピュータの力を借りたら、これまで不可能だったことができそうだ」とか「使わないと研究が進展しない」という

人文学研究者に対しては最新の情報技術や人工知能などが浸透すると思いますが、人文学全体でどうかというと、個人的には少々疑問です。

——最近では、日本語の崩し字を解析して読めるようにするスマートフォンのアプリが開発されていますが、研究者には受容されなくても、先に一般に浸透する可能性はありそうですね。

それはあり得ます。というか、情報「学」としてではなくて情報「技術」という視点で周りを眺めてみれば、研究者には受容されていなくても一般に浸透しているツールやアプリはかなりあります。自動翻訳は典型例ではないでしょうか。私はSFが好きなのですが、原文で読むのはかなり難しい。とにかく筋を追うだけなら、誤訳が多くても日本語に自動翻訳してくれたら嬉しいと常々思っています。自動翻訳に限らず、技術の利用が回り回って情報「学」の研究に影響を与える可能性もあるでしょうね。

また横道に逸れますが、「研究者には受容されなくても」で思い出したことがあります。在外研究員として UC Berkeley で研究していた 1999 年のことです。大学で語学教育用の e-learning ツールを作り、語学の教員に教材データを入れるように依頼していました。その担当者が言うには、ヒンディー語やアラビア語や中国語の教員はデータを入れてくれるのに日本語（人）の教員はなかなか入

41 竹田正幸・福田智子・南里一郎・山崎真由美・玉利公一「和歌データからの類似歌発見」『統計数理』48(2): 289-310 (<https://www.ism.ac.jp/editsec/toukei/pdf/48-2-289.pdf>)。

42 村上征勝「文章の軽量分析——その歴史と現状」『計測と制御』39(3): 216-222 ([https://www.jstage.jst.go.jp/article/sicej1962/39/3/39\\_3\\_216/\\_pdf](https://www.jstage.jst.go.jp/article/sicej1962/39/3/39_3_216/_pdf))。



れてくれないとブツブツ言っていました。何で入れなかったのでしょうか？新しい情報ツールに対する日本人特有の何かが反映していたのか、e-learningは他人事なのか、新しいことへの挑戦が億劫で既存のやり方を変えたくないのか、リテラシー不足で使うのが怖いのか。今の若い人たちがスマホのアプリを縦横無尽に使いこなしている姿からは、ちょっと想像できないのですが。

## ■情報学と地域研究の協働を支える情報リテラシー

——研究の面で、これから情報学と地域研究や人文学とのコラボレーションをさらに進めるうえで、何が問題になると思いますか。

東南地域研のような人文学系の研究所に情報学研究者が参加するとき、どのように成果を出していくのかは難しい問題だと感じています。とりわけ情報学研究者の場合は、同僚の「お手伝い」になってしまうことが多い。たとえばデータベースを構築する場合、それなりに知恵を絞って時間をかけてツールを作ったり、手作業でデータを修正したり、分析したりと、かなりのエフォートを割当てています。残念ながら、多くの時間と手間をかけた割に、情報学研究としてのオリジナリティが低いので、論文になりにくいわけです。そうなる業績が少ないのでステップアップが難しくなってしまう。私の場合、デジタル・ヒューマニティーズに従事する研究者がまだ少なかったので、データベースを作ること自体がそれなりに評価されたから良かったのですが、今ではそうはいかないでしょう。データベースを作ったり情報システムを管理したりという「仕事」しながら、オリジナリティの高い情報学の「研究」も進めなければならないとしたら、これから人文学系の研究所に来る若い情報学研究者は大変だと思います。ですから、地域研究の中で情報学を「お手伝い」としてではなく、一つの研究領域として位置づける努力は情報学研究者としては不可欠だろうと思います。そのためには、地域研究者側では情報リテラシーを上げていく努力も必要です。

ところでリテラシーがないと会話が成り立たないので共同作業は難しくなりますよね。たとえば、あるデータベースを作る場合、情報の側が相手側のリテラシーを勉強してから作業に臨むケースが多

かったと思います。たとえば私が国文研に在籍していたとき、国文学者や古典籍を扱う図書館員が使う語彙を調べたり、一般図書と国文研の古典籍目録の構造の違いを分析したりしてから、打ち合わせに臨んでいました。しかし、同じ努力を相手側はしてくれないんですよね。一方で、相手側も情報技術用語を使うのですが、語彙の意味を調べず、字面から勝手に解釈しているだけのケースが多く、これが議論の際の混乱の元になったりもしました。

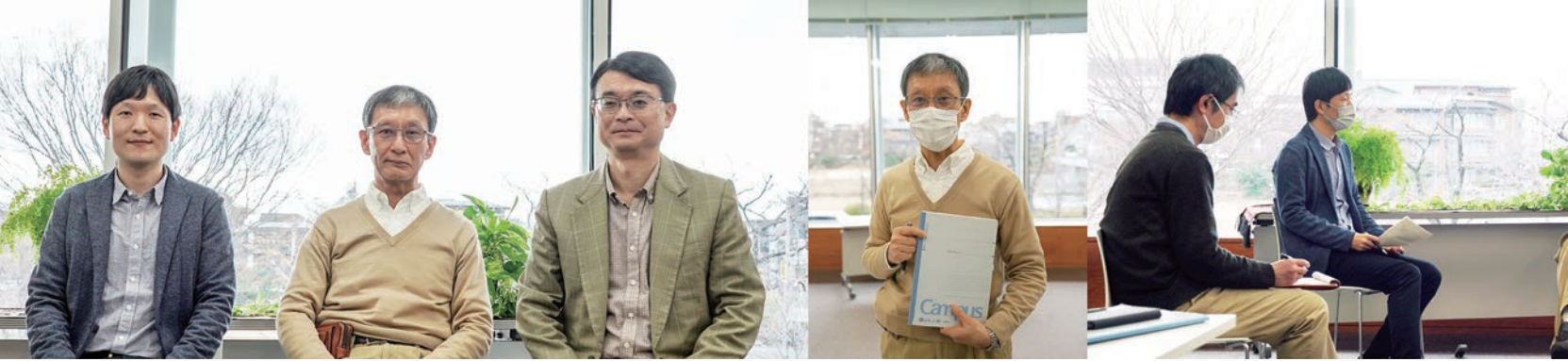
——若手になればなるほど基本的な情報リテラシーは底上げされていて、たとえばExcelはもちろん使えますし、その意味では少し改善されているかもしれません。

ことデータベースに限れば、Excelのできる人は逆にやっかいなんです。表の中に画像を入れてしまうとかね(笑)。そういうテクニックを持っていることは、何もできないよりも良いのですが、それは情報リテラシーあるいはデータリテラシーとは違います。少なくとも、データベースがどんなものかは知って欲しいですね。何度も言いましたが第一正規形を理解するとか。あるいは欠損値・外れ値・異常値の区別がつかないという例もありました。些細に思われるかもしれませんが、この違いは重要です。あるいは、統計分析をするのであれば、統計の入門レベルくらいの知識は持っておいたほうがいい。今では便利なツールを無料で利用できる環境が整っているので、データさえあれば何らかの分析結果を手に入れることは容易です。でもそのツールあるいはアルゴリズムが前提としているデータの条件を満たしていないのに、平気で使っている事例をまま見かけました。

だからといって、人文学研究者にみっちり情報学の勉強をしろというのは無理な話で、また完全なデータを作るために歴大な時間を割くことも難しいでしょう。私たちが自分の有限な研究時間の多くをお手伝いに割きたくありません。ですからデータサイエンスを支援する職種を作るべきだと思います。お金の問題はあるでしょうが、データサイエンスを専門とするURAやリサーチアシスタントのような人が一組織に一人ぐらいいるべきだと思います。

## ■後進の研究者への助言——2+1のテーマを持つ

43 外れ値とは観測値の中で真の値の推定値から大きく外れた値のこと。外れ値のうち測定ミスや記録ミスなど原因がわかっているものを異常値とよび区別する。



——情報学あるいはそれに関連した研究をしたいと考えている研究者に、これまでのご経験から、何かアドバイスをいただけませんか。

データサイエンスに関して言うと、「私はデータ分析屋だから、データ自体については知らない」と平然と言う人を時折見ますが、これはとても良くないと思います。学会のポスター発表などで、素性の良くないデータを使っているのに、平気で「こんな結果が出ました」と言う発表者がいると、ついつい突っ込んでしまいます。アルゴリズムも重要ですが、データを読む力、こちらはデータリテラシーでしょうか、を身につけておかないと、まずいだろうとは思いますがね。

アドバイスというところがちょっと年寄りめいた感じがしてイヤなのですが、私の指導教授によく言われたことを一つだけ。それは「研究ではメインテーマを二つと、サブテーマを一つ、合計三つ持ったほうがいい」ということです。一つの研究テーマだけに没頭していると必ず行き詰まる時が来るが、二つ持っていればどちらかは動いている。それと、人文学分野では同じテーマをずっと続ける研究者も多いですが、理工系の場合は流行廃りが早いので、数年で一つのテーマが終わることを想定して次の手を用意する、その意味でサブのテーマを持っておく、ということですね。

指導教授の言うことに忠実に従ったわけでもないのですが、気付いたらほぼ2+1で研究を進めていました。国文研ではテキスト・データベースと「資源共有化」がメインの仕事でしたが、その脇でGISの人文学への応用研究を開始して、それが京都大学に異動するころにはメインの研究になり、時空間処理ツールを開発や歴史地名辞書の作成と公開に繋がりました。最近では機械学習がメインの研究になっていますが、これも「学知創生ユニット」でサブテーマとして始めたことがきっかけでした。

理工系の研究では、一つの研究ベースからいくつもの分岐が可能なので（私の場合は人工知能→デー

タベース→ビッグデータ）、ベースは一つで良いと思いますが、表層的(?)には2+1ぐらいのテーマを持つことは有効だと思います。

——たしかに、一つのテーマだけで研究を続けていくと、なかなか進捗がないときもありますね。

そうですね。それを言われた時には、「そんなにできるかよ」とムッと思った記憶があります(笑)。でも、いま考えると、あれはなかなかの金言でした。

## ■アイデア、ひらめきの源泉——SF、アニメ、コミュニケーション

——原先生の研究上でのひらめきは、どこからやってくるのでしょうか。SFがお好きだという話がありました。それと関わりがありますか。

何かアイデアが浮かんだ時、どう展開させていくか、どのように使えるかを考える際にはSFやアニメがヒントになることが多いですね。アニメの方が大きいかな。いまだにニコニコしてアニメを見ていて嫁さんからは響きを買っていますけど(笑)。

ひらめきについては、何か食べている時とか、散歩している時とかの話をお聞かせね。私も似たようなものですが、自分なりのちょっとしたルーティンがあります。それは問題でも何でも良いのですが、解決しなければならぬことを思いついたら「沈める」というイメージなんです。頭の中に沼のようなものがあって、そこに沈めておきます。暫くすると答えらしきものがポコッと出てくることある。それでも解決しなければ、また沼に押し戻すんです。こうなると、そのまま忘れ去って腐ってしまったものは数知れず、ですね(笑)。しかし忘れ去っていたのに再び浮かんでくることも希にはあります。たとえば「資源共有化」のアイデアもそんな感じでした。何かしなきゃいけないんだけど、どうしたらいいのかと悩んでいる時期には、頭の中で沼に押し戻すようなことをずっとしていました。

沼から出てくるタイミングとして、私の場合は、

44 Geographic Information System、地理情報システム。  
45 [https://www.nihu.jp/ja/database/source\\_map](https://www.nihu.jp/ja/database/source_map)

46 余談：ビッグデータの地域研究への応用は、アイザック・アシモフの「心理歴史学 (Psychohistory)」がヒントに……



話しているときが多いですね。誰かと話している時にふっと出てきて、「それならこうできるんじゃないか」と思いつくと、それから一人で考えるときもありますが、話せる人間が近くにいたら、相手には迷惑でしょうが一気に話します。すると相手からのフィードバックもあったりして、そうやってアイデアが固まっていくことが多いですね。

### ——人に話すこと、コミュニケーションも大事だということですね。

そうですね。H-GIS 研究会という人文学と情報学の 20 人ほどの研究者の集まりを、2 ヶ月に 1 回程度の頻度で、15 年以上にわたって開催しています。特にテーマを設定することもなく、困ったことがあったらとにかく話してみるとか、おもしろいネタを仕入れたり何かアイデアを思いついたりしたから発表するとか、時にはレジュメを作りますが、ほとんどは話っぱなしの研究会です。ですが、研究のヒントを手に入れたり、行き詰まっている研究を進めたりするきっかけを得たりすることが多いので、この研究会は私にとっては貴重なコミュニケーションの場となっています。

この研究会でさまざまなアイデアを話し合っていると、同じことを言っているつもりだったのに、人文学研究者と情報学研究者では異なったイメージを持っていたということもありました。一例ですが、ある隣土の国で事件が発生したというデータがあり、それを地図上にプロットしようとした。事件の発生した場所はほぼ同定できたので緯度と経度を割り当てた。しかし当時の国境線がどこか分からない。さて地図をどう描くか。情報学研究者は、二つの場所のあいだには情報がないから「無」であると仮定して、二つの事件の発生地点の二等分線を国境とする。後で情報が追加されれば、それに合わせて国境線を変更すれば良い。でも人文学研究者は、「いや、そうは言っても国境は川や山などの自然地形に依拠しているはずだから」と考えて、山の稜線や河川を参考に国境線を引こうとする。こうし

たもの見方の違いを認識することも楽しいことですが、それが新しい研究につながることもあります。実際、曖昧な境界線などをどう同定するかというテーマは、科研研究として採択されました。

この研究グループの活動はアクティブで、話し合いを通じて新しい研究テーマを見つけたり、それをポリッシュアップして科研申請したり、国内学会や国際学会を招致したりしています。そのつど、学内外の人を誘って研究会を続けています。

### ■退職後の活動——ビッグデータ研究の継続と OS の開発

#### ——2023 年 3 月に定年を迎えられたあとは、どのような研究活動をされる予定でしょうか。

幸いにして、科研費の基盤研究 (A)「エビデンスに基づく計量的地域研究の展開」が 2026 年まで続くので、ビッグデータの研究を引き続き進めたいと思っています。

もう一つ、本当に何もすることがなくなったら、「老人による老人のための OS」を作りたいですね。最新の OS もいいけれど、複雑すぎて使いにくいですから、もっと簡単に使えるものを作りたい。いま持っているイメージは、昔のパソコンの BASIC です。あれはプログラム言語でもあり、OS みたいなものでもありで、BASIC だけでコンピュータの制御からプログラミングまで全部できましたよね。もう一度あの単純な世界に戻れたら、余分なコマンドをたくさん覚えなくてもいいので、老人には嬉しい OS になるかもなので、取り組んでみたいと思っています。

2023 年 1 月 10 日  
京都大学稲盛財団記念館 2 階セミナー室にて

47 昔の日本の事例です。

48 operating system、コンピュータの操作・運用・運転を司るソフトウェア、Unix や Windows など。



## Center for Southeast Asian Studies, Kyoto University

原正一郎教授退職記念インタビュー

情報学の知見と手法で新たな分野を開拓した研究・開発・実践の軌跡  
—— 研究生生活を振り返って

2023年3月23日

編集協力：英明企画編集株式会社

発行：京都大学東南アジア地域研究研究所

〒606-8501 京都市左京区吉田下阿達町46

Tel: 075-753-7302 Fax: 075-753-7350

