

解説 人文・芸術系のデータベース—今そしてこれから—

5. 文学データベース—急がれる総合的な環境整備—

Database of Literary Texts: Development of the Infrastructure by Mari NAGASE (Universite Charles de Gaulle Lille 3, Centre de Recherche sur la Documentation et, l'Information Science et Technique).

長瀬 真理¹

¹ シャルル・ド・ゴールリール第3大学科学・技術文献情報研究センター

1. はじめに

先端技術の発達とともに文学や古典の研究にもコンピュータが使われるようになった。これまでのような書籍の形態の資料に代わって、電子媒体に乗った電子資料を電子機器を使って研究する時代が始まろうとしている。

しかし日本ではいまだ基礎となるテキストの全文をデジタル化したテキストデータベースをはじめ、大規模データベース、電子化辞書、解析ソフトの開発は遅れている。本稿では、先行する欧米の歩みを振り返りながら、何が文学研究のための環境整備に必要な総合的な見地から検証する。ただ、英米のデータベースに関していえば、すでに古典といってもよい Corpus をはじめ British National Corpus や COBUILD など、日本でも利用者が増えている。そこで本稿では、1996年4月よりフランスに来ていることでもあり、当地における人文データベースを中心に報告する。最近の日本のデータベース事情については、安永尚志氏の解説記事が詳しいので、そちらをあわせて参照していただきたい¹⁾。

2. 西欧における文学データベース

2.1 構築の歴史

コンピュータへのテキストの蓄積は、1960年代後半、まず個々の文科系研究者達を中心に始まった。当時はまだコンピュータは高額で、しかも入力是非常に骨の折れる作業であり、利用も個人に限られていた。また研究者の数が多きテキストの場合には、重複して同じテキストの入力が行われるなど無駄も多かった。そこで研究者間の入力テキストの共有や、コンピュータの利用方法に関する情報交換の必要性が増し、いくつかの関連学

会の設立が準備された。

1970年代になると、単に個人の研究支援ということだけでなく、各国の言語文化の保存という観点から、主に大学や研究所が中心になって、テキストデータベースの作成が組織的に行われるようになった。これにはコンピュータの発達とともに、光学読み取り装置などの入力機器の発達も見逃せない。欧米では文化資産としての古典や文学に対する保存意識が高く、徐々にではあるが、多額の子算の支援も受けたり、国家プロジェクトとして開発されるようになった。

1980年代以降は大規模なデータベース開発が盛んになっている。とりわけ1990年代に入ってから、グローバルなシェア(共有)を目指した企画が増え、国際標準に則ったアーカイブやコーパス開発が盛んになっている。同時に多様化も進み、新たに文字データばかりでなく画像も含めたハイパーテキストの開発も行われるようになった。利用形態はネットワークを介してアクセスできるものから、CD-ROMやフロッピー・ディスクなどさまざまな形があり、自国以外の研究者にも安価にサービスしている。

2.2 総合的な環境整備

ここで文学研究にコンピュータを導入するためにはどのような環境整備が必要かを簡単にまとめておこう。

- a. 研究対象そのものであるテキストデータベースの作成
- b. 同一テキストのほかのバージョンや参考文献のテキストデータベース
- c. 文章解析を可能にする検索ソフトやハイパーテキストの開発
- d. 電子化辞書の開発
- e. 単一のテキストのみならず大量のテキスト

データベース

古典研究の場合、1つの校訂本だけでは足りずいくつもの写本を参照しなければならないことが多い。重要な箇所ほど、写本の間には相違が多く文献学的跡づけが欠かせない。先行する研究論文や参考文献の量も多く目を通すだけでも大変である。加えて引用や出典箇所を網羅した正確な辞書やインデックスも必要不可欠である。コーパスなど大量のテキストの蓄積が充実すれば、関連するさまざまな外在資料の参照作業や比較研究も容易になる。つまり文学研究には上記の項目のすべての整備が必要なのである。こういった事情を反映して開発も連携して進められることが多い。とくに上記項目 d. の電子化辞書作成のプロジェクトと、e. の大量のテキストデータベース開発は車の両輪のように補完関係にある。

2.3 図書館の電子化

欧米では辞書学や文献学の長い伝統があり、個々の研究者との協力体制が確立している。ライブラリアンや辞書編纂者などの中には博士号をもち、資料の収集から、参考文献、古いマニスクリプトであれば原本や種々の写本について豊富な知識を蓄えた人々が多く層も厚い。すでに多くの大学図書館、研究所、また辞書の作成現場ではコンピュータが導入されデータベースの作成が行われている。さらにデ・モントフォート大学のように電子図書館の実験を始めた大学もある²⁾。テキストによっては、図書館などに足を運ばなくても自室の端末から簡単なコマンド操作で、さまざまな文献にアクセスすることも可能だ。また直接手にとることさえできなかった古文書や珍しい写本を、ネットワーク経由で画像データとして自分のコンピュータに取り込むこともできるようになった。たとえば Robinson, P.氏は7年がかりでチャウサーの「Wife of Bath's Prologue」の58種類の写本、翻訳書、研究論文を次々と画面に呼びだして比較研究できるソフト一体型のデータベースを開発し、現在 IBM の協力でネットワーク公開の実験を行っている³⁾。

3. フランスの場合

3.1 FRANTEXT

一般にフランスでは“ミニテル”が災いしてコンピュータの利用は遅れていると考えられてい

る。しかし文学データベースに関していえば、非常に進んでおり“FRANTEXT”という大規模データベースが構築されている。CNRS(国立科学研究所)の管轄である INaLF(国立フランス語研究所)が、Tresor de la Langue Francais(フランス語基金財団)という国庫補助を受け開発にあたった。16世紀から20世紀までの400年間にわたって発表された主要なテキスト3,500冊が、STELLA(version2)というソフトと共に研究者に公開されている⁴⁾。また国立カタログ・センター(Centre National de Catalogue Collectif National)では、テレテース(Teletheses)というデータベース名で、1972年以降の主要雑誌および博士論文は理科系、文科系を問わず、すべてデジタル化されている。

3.2 ミニテル

ここで簡単に“ミニテル”について説明する。1985年に国営電信電話会社のフランステレコムが電話番号の検索用に貸し出したのが始まりで、全国に650万台、だいたい3軒に1台の割りで普及している。現在は情報提供業者の数も1万あまりに増え、2万5千種類のサービスがある。その中には日本でも数多く翻訳版が出版されている「クセジュ文庫」の検索サービスも含まれている。しかしこの“ミニテル”がネットワークの普及を遅らせているのも事実である。政府計画委員会が1996年10月に発表した報告書「情報社会のネットワーク」によると、フランスは情報通信の分野で大幅に遅れている。電話回線の使用は米国の3分の1以下、ケーブルTVや衛星放送の利用はドイツの10分の1、携帯電話の普及にいたっては欧州で下から2番目である。コンピュータのある家庭は14.4%、世界的な商業パソコン通信の加入者はドイツの20万人に対し4万5千と差がついている。報告はそのほかの原因として、フランス製のコンピュータがないこと、欧州で一番高い特別回線使用料や新しい情報通信分野の法整備の遅れなどを指摘している。

ただフランステレコムは1997年より部分的に民営化され、料金も大幅に値下げした。また1998年にはEC域内で通信事業が自由化されるため、この1~2年でネットワーク利用が劇的に増えるといわれている。そのためテレコム側もこの事態に備え、すでに1995年より従来の8倍の

通信速度のモデムを内蔵し写真の送受も可能な新しい“ミニテル”を発表し、2,000フラン(約4万円)で売り出している。レンタルは月額50フラン(千円)で、一部のクレジットカードの電子決済も可能である。またネットワークに接続も可能で、米国のオラクル社が発売を始めた廉価な通信専用パソコンに対抗している。

3.3 電子図書館

7年がかりで構築された“フランソワ・ミッテラン国立図書館”が昨年12月17日にセヌ川岸に完成した。1千万冊を誇る研究者用施設は1998年にオープンする。最先端技術を装備した野心的な図書館で、その中には電子図書館構想も含まれている。遅れているネットワークの普及を促す上でも、またフランス語・フランス文化の発信を積極的に行うためにも、過去の優れた遺産のデジタル化が不可欠ということから、現在大量の文書のデータベース構築が熱心に進められている*

3.4 電子美術館

文学のデータベースと並んでフランスが力を入れているのは美術データの保存である。国立美術館連合(Reunion des musees nationaux)と、文化省に所属する現代美術基金(FNAC: Fonds Nationaux d'Art Contemporains)の両者が中心になって、画像データベースの作成が急ピッチで進められている。前者はルーブル、オルセー、リールなどの大型美術館が中心で、すでにCD-ROMなどの販売も手掛けている。後者は近代以降の作品が中心で、ほかにFRACという地方基金もある。そのほか美術館に所蔵されていない作品は1点6~7千フランでコピーライトを買い取り入力している。拡大・縮小をしてもぼけない特殊なソフトも開発しており、研究者も安価な使用料で使えるようになってきている。さらに現代美術に関してはポンピドー・センターのメディア・アート部門が中心に行っている。始まったばかりであるが、NASA、NTTやNECの協力でバーチャル・リアリティで作成された作品のデジタル化も始まった。3年前のルーブル美術館200年記念時に、リシュリュー館が完成したが、以来3カ月に一度研究会が開催され、“未来の美術館について”討議が続けられている。

昨年末、没20年を記念して、名誉あるパンテ

* 日本でも学術情報センターが電子図書館の実験を開始した⁹⁾。

オン入りを果たした、アンドレ・マルローは文化相として、あるいは作家として日本でも有名だが、フランス美術の保存に寄与したことで知られる。彼は芸術や文学のめづりをよく認識しフランス各地に文化会館を構築した。その結果、戦後フランスは見事に“文化大国”として再生した。“平等”を標榜するお国がら、誰もが文学や美術にアクセスできる電子図書館や電子美術館の実現はその延長線上にある。文化遺産は、その国の人々にとって、さらには人類すべての宝であり、人々の心の糧になる。実際、“芸術の伝導師”と呼ばれたマルローは、芸術があらゆる階層に浸透することを願ったといわれる⁹⁾。

最後に一言、ソフト開発に関連してフランスの特殊事情に触れておこう。同国ではネットワーク進出に遅れたために、CD-ROM利用が中心のコンピュータ購入者と、ネットワーク利用に関心をもつ層が分離している。そのため両方の利用者を結びつけるために、CD-ROMを利用しながらネットワークにアクセスして、データベースの更新が可能なソフトの開発が現在最優先課題となっている**。

4. 日本の現状

日本はコンピュータ開発では最先端に行くにもかかわらず、文科系の研究支援を明確に意図した環境整備は残念ながら非常に遅れている。しかし学術情報センターが学術文献を中心とした国産データベースの開発に力を入れるようになっており、今後国際的な協力体制も確立されると思われる⁷⁾。またネットワークの利用が進み、個人的に日本の古典テキストを公開するケースも出てきている。しかし日本語環境を想定して公開しているものも多く、海外からアクセスすると、すべて文字化けしており外国の利用者に対しては不親切なものも多い。Webサイトに公開するならば、英語の解説や、日本語表示ソフトの案内などプレゼンテーションに工夫が欲しい。

5. 問題点および将来の展望

以上簡単ながら文学データベースの現状について報告した。変化し崩れつつある日本語を保存す

** 日本でも国立西洋美術館の波多野氏を中心になって、1997年4月から電子美術館のコロキウムを開始した、との情報がフランスにも届いている。

るためにも古語辞典の電子化プロジェクトや大量の日本古典のテキストデータベースの開発を急ぐ必要がある。また個人の負担に期待するのではなく、多様な媒体で内外にサービスをする専門的な機関の設立も望まれる。

今後、日本でもハイパーテキストの開発が進むと思われるが、その場合、ぜひとも文学や古典の専門家、それも複数の研究者の協力を得て付加価値の高いハイパーテキストの作成を心掛けて欲しい。とくにソフト開発者に希望することは、やはり現場の研究者や利用者の意見の尊重である。技術系主導で作られたデータベースは汎用性を考慮するあまり、小回りがきかず、実際の研究では役に立たないことが多い。また既存の技術から応用発想することが多く、冒険的、実験的、個別的・特殊なテーマへの挑戦を嫌う。予算の問題もあろうが、貴重な題材がただ切り刻まれて、どこにでもある無難なできで留まってしまうのは非常に残念なことである。また、せっかくよいソフトやデータベースを作っても、開発が終わると、一件落着といわんばかりにテキストに対する興味を失ってしまう開発者が多い。古典作品の研究に終わりはないと同様、データベース開発にも終わりはない。

たとえば、評価の高いベルセウス・プロジェクトは完成までに10年かかっている。現在、米・仏・スイス・伊の4カ国協同で進められているデカメロン・プロジェクトも7年程経過しているが、完成は予想がつかないという。技術が高度になったからといって、資料整理が楽になり短時間で済むようになるわけではない。実はその逆で、より多くの資料や多様な解釈やグラフィクスを盛り込んだハイパーテキスト作成への夢が膨らんでくる。またネットワークにより国際的な協同作業がやりやすくなると同時に、複数の研究者間の意見調整も難しくなる。少しずつ公開されるたびに、参加する研究者の数も増えてくる。こうしてハイパーテキストは新しい解釈を産み出しながらネットワーク上にシェアを広げていく。こうなるとサポートや技術支援は作品の研究と同様に永遠に続くかもしれない。こういった事情を考慮して長期の協力体制を組んでいただきたい。

最後に、データベース開発には現在コピーライトの問題が大きな障害になってきている。図書館の電子化にあたっては、著作物はユーザのインタラクティブな行為によって切断、変形、再編集されることが予想され、著作人格権の保護が非常に難しくなる、という指摘がある⁸⁾。このままではせっかく電子美術館や電子図書館といったシステムとハードが完成しても、コンテンツのないものになってしまう。日本の誇る貴重な文化遺産がデータベースとして世界に発信できるよう、コピーライト問題の専門家集団の早急な対策を期待したい。

参考文献

- 1) 安永尚志：日本古典文学の本文データベース，情報処理学会誌，Vol.35，No.7，pp.642-650 (July 1994)。
- 2) Collier, M: Networking and Licensing Texts for Electronic Libraries, Interlending & Document Supply, Vol.23, No.4, pp.3-13 (1995)。
- 3) Robinson, P.: Text Collation, L&LC Vol.4, No.3, pp.174-181 (1989)。
- 4) Rebollar, P: Frantext, le Chercheur Nouveau Est Arrive, Actes du Colloque Franco-Japonais, pp.51-63 (1993)。
- 5) Adachi, J. and Hashizume, H.: NACSIS Electronic Library System: Its Design and Implementation, Proceedings of International Symposium on Digital Libraries, pp.36-41 (Aug. 1995)。
- 6) France News Digest, No.344, p.3 (Oct. 1996)。
- 7) 根岸正光：NACSISのデータベース，情報処理，Vol.33，No.10，pp.1144-1153 (Oct. 1992)。
- 8) 名和小太郎：ネットワークをめぐる法的課題，情報処理，Vol.37，No.2，pp.135-142 (Feb. 1996)。

(平成9年2月21日受付)



長瀬 真理 (正会員)

1976年東京都立大学人文科学研究科哲学博士課程修了。1980～1982年ブリティッシュ・カウンシル給費生として英国マンチェスター大学古典学科留学。1990～1991年オックスフォード大学客員研究員・講師。城西国際大学を経て現在シャルル・ド・ゴール・リール第3大学客員教授。文学データベース開発，文章解析などに従事。ALLC日本支部代表。
e-mail:nagase@univ-lille3.fr