

「近代中国語文献資料データベース」構築にむけて

氷野善寛

Towards the Development of *Database Resource of Modern Chinese Language Materials*

HINO Yoshihiro

At the Institute for Cultural Interaction Studies, Kansai University, there is an ongoing project for the creation of a database/archive for Chinese language related documents by multiple groups in which this author is involved. This essay introduces the process and methodology for the creation of digital images of documents as well as the creation of a database/archive. It also investigates recent problems with the creation of databases, the collection of documentary materials on modern China, and the need to consider other issues outside of the creation of a database.

キーワード：近代中国語、文献データベース、デジタルアーカイブ、コーパス

はじめに

関西大学文化交渉学教育研究拠点など筆者が関わる研究拠点を中心に複数のグループで中国語関連資料を中心とする文献のデータベース化及びアーカイブ化の作業を行っており、筆者はその中で「近代中国語文献データベース」のプラットフォームとなるシステムの構築、「近代中国語アーカイブズ」で公開する文献の撮影方法の確立及び公開ツールの設計を行っている。そこで本論では現在構築中のこの「近代中国語文献資料データベース」を例に、文献資料のデジタル化の手順と方法及びデータベース、アーカイブ化について紹介すると同時に、昨今のデータベース化及びアーカイブズに関する問題点、また近代中国語に関する文献資料を収集、データベース化する上で考慮しなければならない点について検討していきたい。

一、「近代中国語文献資料データベース」の概要

本論で紹介する近代中国語に関連するデータベースは、近代中国語文献資料データベース¹⁾(近代漢語文献資料データベース, Database Resource of Modern Chinese Language Materials)、近代中国語文献アーカイブズ(近代漢語文献数碼典藏, Digital Collection of Modern Chinese Language Materials)、近代中国語コーパス(近代漢語語料庫, Corpus of Resource of Modern Chinese Language Materials)の3種類から構成される。それぞれ、近代中国語に関する文献資料の目録データベースである「近代中国語文献資料データベース」を中心に、文献のデジタル画像のアーカイブズである「近代中国語文献アーカイブズ」、関連資料の中で特に重要と考えられるものをテキスト化し、全文検索が可能なコーパスとして独立運用している「近代中国語コーパス(旧称;近代漢語文献データベース)」というように役割分担がなされている。「近代中国語文献資料データベース」については、ある特定の図書館、あるいは特定の固有のコレクションといったものに限定してデータ整理を行うのではなく、主に中国の清代から民国期にかけて発展していった近代の中国語を研究するために必要となる言語資料、特に言語的立場から必要となる文献資料を網羅的に収集して、その所蔵先を明らかにし、体系化していくことを第一の目標とし、さらには、該当する書籍のデジタルデータの公開及び収集に努めることを第二の目標としている。以下に各データベースの個別の役割及び構成についてまとめる。

日本語名：近代中国語文献資料データベース

中国語名：近代漢語文献資料データベース

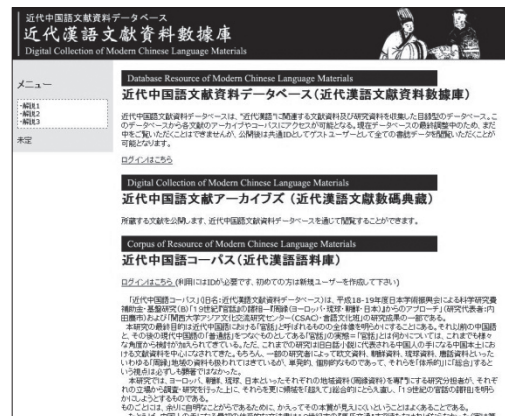
英語名：Database Resource of Modern Chinese Language Materials

中核となるデータベースで、上述の文献資料に関する情報を文献資料単位で整理するためのデータベースである。運用にはFileMakerServerを利用しており、インターネットを通じて項目の新規作成、編集、修正などの作業を共同で行うことができ、以下の5つの個別のデータベースからなるリレーショナル型のデータベースでもある。

文献データベース

著者データベース

論文データベース



1. 開発中のページ

1) <http://www.icis.kansai-u.ac.jp/ModernChinese-db/>

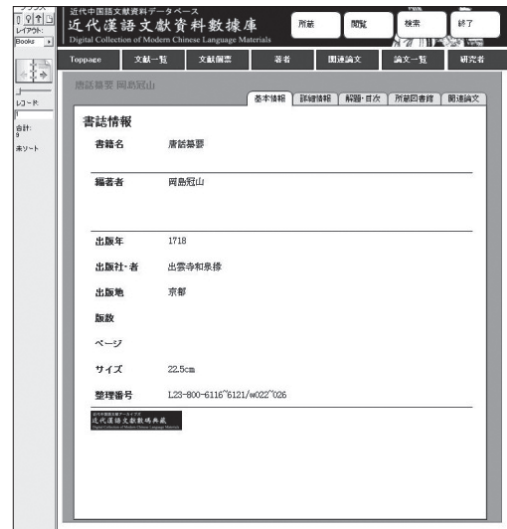
原資料データベース

所蔵データベース

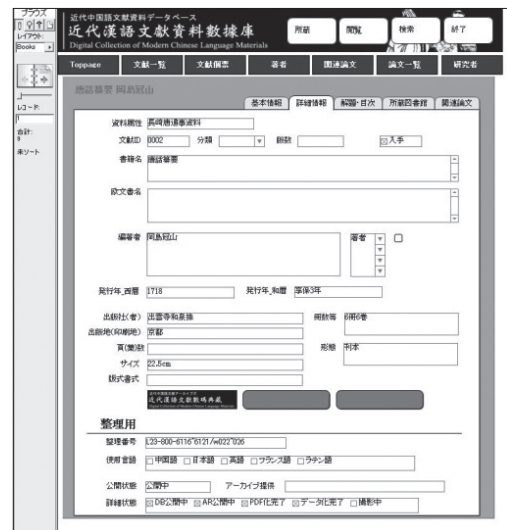
文献データベースでは、書名（和文、欧文）、編著者、資料属性、分類、版数、発行年（和暦、西暦）、出版社（者）、出版地、冊数巻数、頁数、サイズ、形態、使用言語、解題（備考）、目次、所蔵情報、関連論文などに関する情報が登録可能で、所蔵情報と著者情報と論文情報は、それぞれ所蔵データベース、著者データベース、論文データベースと情報を共有しており、たとえば各文献の所蔵先を入力しておけば、所蔵データベースの各所蔵機関の所蔵文献一覧に所蔵する文献の一覧が自動的に表示される仕組みになっている。また該当文献が本拠点でデジタル化されている場合は、「近代漢語文献数碼典藏」のロゴをクリックするとその文献画像を専用リーダーで開く仕様となっている。

文献データベースでは先述の通り、関西大学の図書館などある特定の機関の所蔵に依拠するのではなく、「近代中国語」という範疇に含まれる資料全般を収録し体系化することを目標としている。そのため、今後の課題として、この目録にどの程度の範囲の文献を収録する必要があるのか、精査する必要がある。また、収録する各文献の情報の程度についても本論の後半で検討することとする。

次に、著者データベースは文献データベースに登録された著者に関するデータを収録するデータベースで、著者名、生没年、国籍、職種、著書一覧が掲載でき、同一著者による執筆文献一覧も文献データベースからデータを取得して表示できるようになっている。論文データベースは、近代中国語に関連する研究がなされている論文データを収録するデータベースで、キーワードによって文献データベースと関連付けがされるようになっている。原資料データベースについては所蔵データとは別に文献目録などの資料など引用元となっているソース情報を表示するためのデータベースである。最後の所蔵データベースについては、各文献の所蔵データを所蔵機関別に確認できるように関連付けており、さらに各所蔵機関一覧から個別の文献データへのアクセスが可能となっている。以上のように相互参照の網を張り巡らしたのが本データベースの特徴と言える。



2. データベース（基本画面）



3. データベース（詳細画面）

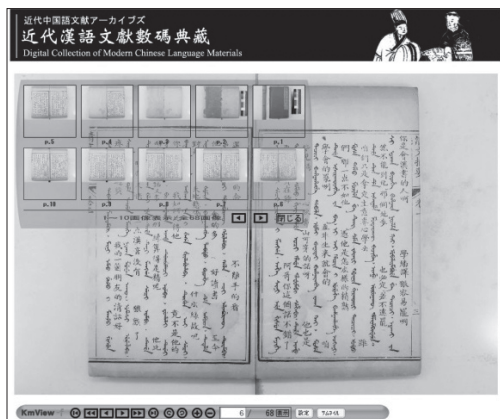
次にこの目録型のデータベースから個別の文献の画像データへのアクセスの際に利用する、近代中国語文献アーカイブズについて詳述する。

日本語名：近代中国語文献アーカイブズ

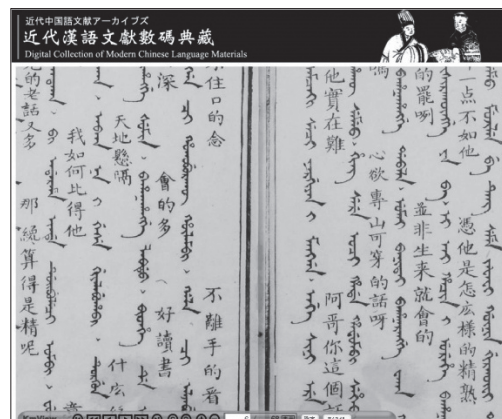
中国語名：近代漢語文献數碼典藏

英語名：Digital Collection of Modern Chinese Language Materials

「近代中国語文献アーカイブズ」は、デジタル化した文献の専用ビューワーでの表示機能を担っている。現時点ではFlash Playerを搭載したコンピュータであれば、各文献の閲覧操作が可能となっている。この専用ビューワーには国際マイクロ写真工業社が開発したKmView²⁾を採用している。2011年春を目標に公開可能なデータとしては、尾崎實氏の旧蔵書、あるいは内田慶市氏の個人蔵書の中で、著作権がすでに無く、かつデジタル化作業が終了しているものが中心となり、今後賛同が得られれば、関西大学で所蔵する個人文庫などに収録されている文献の掲載もできればと考えている。以下に掲載している画像は『新刊清文指要』をビューワーを通して表示したイメージである。



4. アーカイブビューワー（基本画面）



5. アーカイブビューワー（拡大画面）

日本語名：近代中国語コーパス³⁾

中国語名：近代漢語語料庫

英語名：Corpus of Resource of Modern Chinese Language Materials

「近代中国語コーパス」は今回のデータベースとアーカイブ設置の際に、より正確にデータベースの性格付けをするために「近代漢語文献データベース」という名称から現在の「近代中国語コーパス」という名称を変更した。「コーパス」とはある特定の目的のために集められた特定言語におけるテキストデー

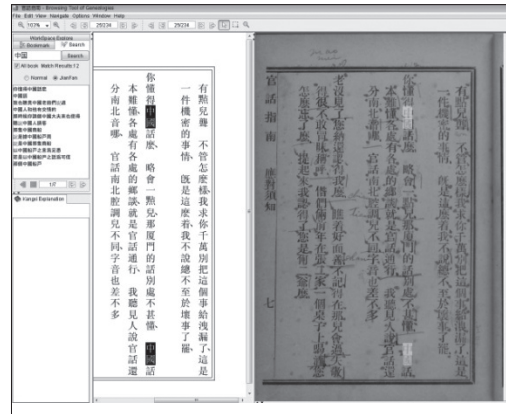
2) <http://www.kmsym.com/kmview/top.htm>

3) 「近代中国語コーパス」(旧名：近代漢語文献資料データベース)は、平成18-19年度日本学術振興会による科学研究費補助金・基盤研究(B)「19世紀『官話』の諸相——『周縁(ヨーロッパ・琉球・朝鮮・日本)』からのアプローチ」(研究代表者：内田慶市)および「関西大学アジア文化交流研究センター(CSAC)・言語文化班」の研究成果の一部である。

タや音声データなどを集めた言語データのことを指す。その名の通り、この「近代中国語コーパス」では近代中国語関連資料の中でも重要と思われる資料を順次テキストデータ化し、全文検索及び、画像データとテキストデータを対照表示できるようにしたものである。詳細については内田・氷野2007⁴⁾に詳しい。なお、このコーパスでは、これまでに欧米人の手による中国語会話書や聖書など約50種類の資料を公開しており、現在のところ収録資料が100件程度になる予定で作業を進めている⁵⁾。以下の画像は、本コーパスの全文検索の結果表示画面と、専用リーダーの対照表示画面である。



6. 近代中国語コーパス（全文検索画面）



7. 近代中国語コーパス（文献対照表示）

二、文献のデジタル化の方法と手順

デジタルアーカイブを構築する目的としては、対象資料への簡便なアクセス、資料の保全と記録などが挙げられる。特に「近代中国語文獻資料データベース」で扱う資料の中には、個人で所蔵しているも

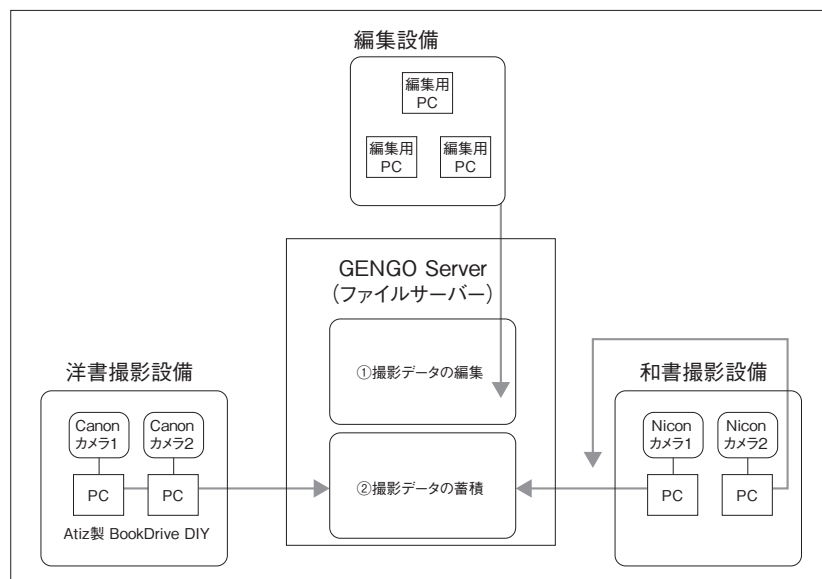
4) 内田慶市・氷野善寛2007「近代漢語文獻資料データベース」の構築『東方』318号, pp. 2-6

5) これまでに収録している書目は『語言自邇集』初版、『語言自邇集』第2版、『語言自邇集』第3版、『官話類編』、『聖諭廣訓直解』、『京話指南』、『華英通語』、『新增華英通語』、『正音咀華』、『正音撮要』、『華英通語』、『英語集全』、『華英通語集全』、『漢英通用雜話』、『熊說』、『況義』、『登瀛篇』、『海國妙喻』、『意拾秘傳』、『正音彙編』、『天主正教真使實錄』、『拾級大成』、『亞細亞言語集 支那官話部』、『總譯亞細亞言語集 支那官話部』、『自邇集平仄編 四聲聯珠』、『意拾喻言』、『官話指南』2種、『改訂官話指南』、『增補華語跬步』、『英語註解』、『清文啓蒙』、『漢文華語康熙皇帝遺訓』、『格致彙編』、『時務報』、『強學報』、『中西聞見錄』、上海土本版『新約全書』、蘇州土本版『新約全書』、官話版『新約全書』、『北京官話清國學捷徑』、『聖教要理』、『增新華英貿易字典』、『字典集』、*Dialogues and Detached Sentences in the CHINESE LANGUAGE; with a free and verbal Translation in English.*, *Aesop's Fables. Compiled for the use of Chinese studying English; and English studying Chinese. In three parts., New Terms for New Ideas*である。また今後追加予定の書目としては『士民通用語録』、『英文學隅』、『日本忠臣庫』、『學官話』、『百姓話』、『新刻官話彙解』、『唐話纂要』、『漢字撮要』、『清文指要』、『清文指要漢語』、『清文要指』、『三合語録』、『初学指南』、『初學官話功課』、『語言問答』、『問答篇』、『漢譯伊蘇普譚』、『漢譯批評伊蘇普物語』、『廣訓衍』、『新刊清文指要』、『滿漢字清文啓蒙』、『官話篇』、『北京官話伊蘇普喻言』、『意拾喻言』、『新校語言自邇集 散語之部』、『官話初階』、『今古奇觀』、『無師自通官話書』、『漢語入門』、『支那文典』、*Manuel de la langue chinoise parlée, a l'usage des Français* など、近代中国語の「官話」を解明するのに必要となる欧文資料、朝鮮資料、琉球資料、唐話資料といった周縁資料を扱っている。

のや、唯一無二の貴重な文献資料も数多く存在し、図書館の中で貴重書指定がされているものも多く、容易に閲覧することができないものが多い。もちろん本物を扱うに勝ることはないが、一次的な調査等を行う上では、今後この種のアーカイブの扱いに対する比重が大きくなっていくはずである。そこで、本章ではアーカイブ化を促進することを目的として「近代中国語文献アーカイブズ」で公開する文献画像の撮影方法及びデータ化の手順について概説する。

これまで実際に本拠点を中心として、複数のグループにより文献のデジタル化が行われている。一般的に貴重書の撮影については専門の業者に外部委託を行うケースが多いが、デジタル化の作業を効率的かつ容易に進めて行くために、撮影及びデジタル化専用のスペースを設置し、貴重書の撮影から編集、公開までの工程を教員や学生の手によって行っている。

近代中国語に関連する文献としては、中国国内の資料だけでなく、欧文資料、朝鮮資料、琉球資料、唐話資料等各種の「周縁」資料を含んでおり、そのため装丁については一律同じというわけではなく、和装本もあれば洋装本もある。そのため、文献ごとに最適な撮影方法を採用するため、文献の撮影に際しては、和装本と洋装本の2つに分け、以下のような撮影システムを作り上げ、デジタル化作業を進めている。



6. システム構成

和装本の撮影に際して注意する点は、上方から垂直に撮影を試みる場合、綴じ糸より外側にある背になる部分があるために撮影がしにくくなる場所である。その問題を解決するために、背の部分の落ち込みに合わせて、写真のような専用の可動式撮影台を自作した。大きめの白い板の上に、同じ板を半分に切断したものをそれぞれ留金で固定し、左右のうち片方だけにスライダを設置し、左右に動くようにしている。そうすることで書籍の背の部分の落ち込みを挟み込むようにして配置することができるというわけである。さらに上から無反射ガラスを置き、その状態で上方から撮影する。調光装置については市販のものを利用している。



9. 和書撮影台

専用の撮影台を調光装置が設置された台の上に設置したものが次の写真である。



10. 和書撮影台全景

撮影の際に利用した機材は以下の通りである。

カメラ：Nikon D300s

撮影台：LPL コピースタンド CS-30 と自作の可動式の和書撮影台

調光：LPL コピーライト FL-2152

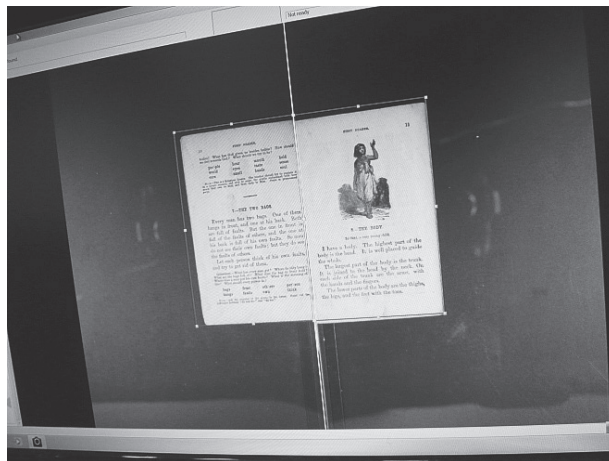
この機材を用い、コンピュータにカメラを接続し、Nikon Camera Control Pro 2 という操作のソフトをインストールしてカメラを操作、上述のネットワーク経由で直接ファイルサーバーにデータを転送する。撮影した画像データについては、数回のチェックを経て、PhotoShop等の画像処理ソフトを利用して微調整を行っている。

次に、洋装本の撮影については、和装本と異なる方法を用いている。これは洋装本では、和装本と異なり背をくぼみに落とし込んで固定する方法を使うことができず、またある一定の角度より書籍を広げた状態で撮影を行うと、この背の部分の糊付けが剥がれたり、あるいは背表紙が激しく損傷したりする可能性があるためである。そこで、上方からの垂直に撮影する方式ではなく、文献をV字型に開き、斜めから撮影することが可能な撮影台を選定した。その結果、Atiz社のBookDriveを洋装本の撮影台として採用することにした。



11. 洋書撮影

この撮影台では、文献をV字型に設置し、その上にアクリル製の専用透明ガラスが設置されており、上下移動で配置できるようになっている。この専用台に設置することで、文献の見開き面を垂直方向から設置された二台のカメラで左右のページをそれぞれ同時に撮影することができる。撮影したデータは、ネットワーク経由でホストとなるコンピュータに送信され、以下の写真のように両方のページが見開きの状態で表示される。



12. 洋書撮影（コンピュータ上での表示）

カメラ：キヤノン EOS 50D、レンズ：EF-S18-55mm F3.5-5.6 IS

撮影後は Atiz 社の BookEditor という専用画像編集ソフトを利用して画像処理を行うことになる。

以上が当グループにおける文献のデジタル化の全容である。

三、データベースの収録文献の範囲とデジタル目録について

紙の目録であれば印刷して出版した段階で完成となるが、インターネット上の目録、文献データベースには、この時点で完成という明確な線引きはなく、常に刷新することが可能である。この論文が刊行される頃には上記で説明したデータベースの内容と全く異なるものになっている可能性も否定することはできない。それが良いか悪いかという議論はひとまず置いて考えると、このことは利用者にとっても、作成者にとっても、常により良いデータベースとは何かということを考え、試行錯誤できるチャンスを与えてくれていることに他ならない。これまでの目録は、多くは個人あるいは特定のグループにより編纂作業が行われ、その人自身の文献に対する知識の集大成とも言えるものが多く、ある文献の中からどういった情報を取り出すのか、あるいはそれをどのように記述していくと誰もが分かるのかといったことを考え、文献目録が作成されている。特に紙の目録であれば、印刷の都合上、必要最低限の言葉を用いて、より正確に書籍の性格を著さなければならないという制約があり、文献目録の世界には独特な表現方法が多くある。またどのようにすれば簡潔で見やすいか、といったことは、分類、配列、各文献について記載する内容などで工夫がなされている。一方、デジタルの世界ではどうかというと、分量という制限はない。ただし制限がないからといって無制限に記載するような方法では、やはり良いとは言えず、簡潔な記載と詳細な記載を必要に応じて使い分ける柔軟さは必要である。そしてデジタルの世界の目録編集のメリットになりうるものの一つとして共同編集という概念が考えられる。ウィキペディアのような不特定多数による編集作業の場合、そのメリットと共に、デメリットが強調されるが、この種の専門性が高いデータベースにおいては、国内外の研究者による特定多数の編集作業に向いており、ある特定の分野の研究者が協力することで、それぞれの研究者が属する国家、図書館などの情報を統合していくことができるという強みを持っている。そのためにある程度以下のことについては検討しなければならない。

- ① 「項目」の分量とのレイアウト
- ② 「分類」の必要性
- ③ 「近代中国語」の範囲の定義

まず①については、データベースの構築に際して、収録するデータ項目の検討が必要となる。内田・小野2007「Nova Bibliotheca Sinicaの構築に向けて」で西洋人が残した中国語関連資料のデータベース化について提起されており、そこではフランス人中国語学者アンリ・コルディエ（Henri Cordier 1849-1925）が残したBibliotheca Sinicaを分析し、データベース構築に際してのヒントを求めている。そしてBibliotheca Sinicaでは著者名、書名などの主要な書誌情報よりも、「二次的情報」に様々な情報が詰めこめられているという特色を指摘している。参考までにこのコルディエの目録に記載されている情報について引用する。この目録では、主要な書誌情報と、二次的情報を次のように区切って扱っている。

a) 主要な書誌情報

題名／漢文の題名
著者名／漢文の著者名
出版地
出版社／印刷所
丁 合
ページ数

b) 二次的情報

書物の内容・目次
記事・書評などの二次文献の情報
著者について
著者の死亡記事に関する情報
他の書物との関連付け
手稿の移動の経緯
写本の所在と移動の経緯
現在の保管場所（目録番号）
形式的書誌学情報（紙質、印刷方法、サインなど）

二次的情報の中で特に注目すべき点として、ある特定の図書館に所蔵されている文献についてそこにあるという事実を述べているだけでなく、その文献がどういった経路を経て現在の場所にあるのか、あるいは著者に関する情報や、関連文献やその文献に対する記事などについても積極的に収集している点である。もちろん全ての文献に対して一律このような情報が付されているわけではなく、全くの空白のものもあるが、この点については「近代中国語文献データベース」の雛形を作成する際の参考となっている。しかしながらこれらの情報項目を基本に当初「近代中国語文献データベース」の項目整理を行ったが、レイアウト上の問題として、情報があまりに複雑、多岐化してしまったため、逆に見づらくなってしまという事態を招いた。そのため情報を分割し表示項目を整理した。参考にしたのは、内田・小野2007の結論として採用されているMARC（機械可読方式）形式を整理した11項目の情報である。この内容はハーバード大学をはじめとして多くの図書館で採用されているものである。

著者、編者
出版年月日
出版者
出版場所
版数
ページ数
所在

請求番号
目次
コメント
関連資料

結果的に「近代中国語文献資料データベース」では、コルディエの主な書誌にあたる部分を、このMARC方式とし、簡易情報として表示し、二次的情報にあたる部分を整理し、詳細情報として分割表示するという方式を利用することにした。

次に②については、データベースの整理上の問題と、いかに目録を体系化するかという問題点と関連する。たとえば上述のコルディエの目録であれば、主題別、著者別、年代別、を組み合わせて分類をしたと証言しているように、紙の目録にあつては分類、配列というものによってその目録の性格が形作られるほど大事なものである。デジタルの目録では、配置や配列については自由に入れ替えることができるため、より柔軟な対応が可能である。ひとまずは、資料整理に便宜を図る為に、文献によって「琉球資料」「欧文資料」「唐話資料」などの資料の種類による分類の他、「辞書」「会話書」「文法書」といった資料のジャンルによる分類ができるようにしている。

最後に③については、より詳細な検討が必要である。「近代中国語」関連書籍といった場合、その資料の範囲は辞書・会話書・文法書といったもの以外に、その時代の文学作品、新聞、科学関連書籍・・・など、それこそ無尽蔵に広がっていく。それらをすべて網羅するのが良いか、それとも直接的に「言語学」に関わるものを取捨選択していくのが良いか、検討の結果、「近代中国語」の形成に関わる「言語学」的に資料的価値がある文献に限るという定義付けをすることにした。その上で、基礎資料のリストアップを行う。構築中の「近代中国語文献資料データベース」のデータ作成作業において、中核として考えられているのは、「19世紀以前の欧米人による中国語学関係文献」、「琉球官話資料」、「長崎唐通事関係資料」などであるが、目録作成の基礎となる資料について先述の論文を参考にリストアップする。

- 1) *Catalogue of the Wade collection of Chinese and Manchu books in the library of the University of Cambridge*, Herbert A. Giles, 1898.
- 2) *Catalogue of Chinese printed books, manuscripts and drawings in the Li brary of the British Museum*, Robert Kennaway Douglas.1877
- 3) *Bibliothèque nationale de France*
- 4) *Soas, Western Books on China published up to 1950*, John Lusst.1987
- 5) *Catalogue of the Morrison Collection of Chinese Books*. Andrew C. West.1998
- 6) Harvard University Library, *Catalogue of the protestant missionary works in Chinese* 1980
- 7) *Catalogue of the Asiatic Library of Dr. G. E. Morrison: now a part of the Oriental Library*, 東洋文庫1924
- 8) *Memorials of protestant missionaries to the Chinese*, Alexander Wylie.1867
- 9) *Catalogue of the publications by protestant missionaries in China*. S.W.Williams.1876

- 10) 『清代民国時期漢語研究文献目録（稿）』2010
- 11) 「欧米人著作中国語関係書書目」『創大アジア研究』第14号, 六角恒広1993
- 12) 『西方人早期漢語学習史調査』張西平等2003
- 13) 『中国語関係書書目 増補版』六角恒広2001

その他、主要な図書館のカatalogも大事で、著者データベースを構築するためには国会図書館のデータベースや『近代来華外国人名辞典』などから著者に関する情報を収集するのも有用である。現在は基本データとして、関西大学で所蔵されている関連文献、内田氏などの個人蔵書の目録を整理したものを利用しているが、今後共同作業や上記の目録などの整理を経てデータベースを充実させていく。

四、デジタルアーカイブとデジタル文献目録の方向性——整理の必要性

ここまで本拠点において構築中のデータベースに焦点を当ててきたが、ここではデータベースやデジタルアーカイブズといったもの全般について着目し、その利便性の裏にある問題点についても少し考えてみたい。

「近代中国語」の諸問題について研究、あるいは今回のデータベースの設計を考える上で、お世話になった国内外のアーカイブがいくつかある。日本国内であれば、

国立国会図書館近代デジタルライブラリー
早稲田大学古典籍データベース

海外であれば

National Library of Australia
Harvard Libraries HOLLIS Classic

などである。いずれも精力的に文献資料のアーカイブ化を計っている。またこれら以外の図書館でも、所蔵する貴重書のコレクションのデジタル化が行われている例が多くある。こういった個別の図書館の動きと並行するように、Internet Archive⁶⁾ や Open Library⁷⁾ や GoogleBooks のようにデジタル化された文献の検索サイトの役割を担っているサイトもある。

まず筆者はこれらのアーカイブズを否定する立場ではなく、むしろこういったアーカイブが今後増え、各地の図書館が所蔵するデータがさらに表に出てきて、誰でもが容易にそのデータにアクセスできるような状況になることを願っている立場にあることを先に述べておく。ただし、そういった状況が何の整

6) Internet Archive <http://www.archive.org/>

7) Open Library <http://openlibrary.org/>

備も無く続けば、おそらくデジタルアーカイブズの整理がおそらく今後抱える大きな問題点となると考
える。たとえば現在「近代中国語文献資料」に関するデジタル化された文献について考えると、現在で
もそれぞれ世界中のサーバーに散らばっており、個別の図書館のサイトでひっそりと公開されているケ
ースや、GoogleBooksなどのサービスを通じて検索して表示できるケースなどが考えられる。ただ自分
が目標とする書籍の正確な名称を知らなければその文献のデータにたどり着くことは非常に難しく、現
代の情報量の多さにより埋没してしまう可能性も指摘できる。

試みとして上述のサイトでThomas France Wadeが執筆した北京語の教科書である『語言自邇集』に
ついて調べ、どれくらいのデジタルデータを得ることができるかみてみたい。

まず一般的な書誌情報を『語言自邇集』をキーワードに日本のWebcatで検索すると、

① Yü-yen tzu-erh chi = 語 言 自 邇 集 : *a progressive course designed to assist the student of colloquial Chinese, as spoken in the capital and the metropolitan department, in eight parts, with key, syllabary, and writing exercises*/by Thomas Francis Wade. - Trübner, 1867

② 語 言 自 邇 集 (Yü yen tzü êrh chi) = *a progressive course designed to assist the student of colloquial Chinese, as spoken in the capital and the metropolitan department*/prepared by Thomas Francis Wade & Walter Caine Hillier ; v.1, v.2, v.3. - 2nd ed. - Statistical Department of the Inspectorate General of Customs, 1886

③ 語 言 自 邇 集 (Yü yen tzü êrh chi) : *A progressive course designed to assist the student of colloquial Chinese as spoken in the capital and the metropolitan department*/prepared by Thomas Francis Wade & Walter Caine Hillier ; v.1. - 3d ed. - Kelly and Walsh, limited, 1903

の3冊と数冊の写本等の書誌が表示される。ここから『語言自邇集』には初版（1867年）4巻、2版（1886年）3巻、3版（1903年）2巻という3セットが中核となり、派生本が数種あることが分かる。そこで、この中国語名の『語言自邇集』、書名の当時の音声表記であるYü yen tzü êrh chi、英語名の*a progressive course designed to assist the student of colloquial Chinese*の3つをキーワードに所蔵が確認できているハーバード大学とOpen Library⁸⁾、Internet Archive、GoogleBooksの3サイトでそれぞれ検索を行った⁹⁾。

まず「Yü-yen tzu-erh chi」でInternet Archiveを検索すると、Second Edition の3冊がヒットし、それぞれ画像データを確認すると最後に「University of Toronto Library」の貸し出しカードが写り込んでいることが確認できた。このことからおそらくトロント大学の図書館が所蔵している同書をデジタル化したデータを掲載していると考えられる。さらに英語名で検索しなおすと、上記の3冊に加え、5

8) Open Libraryは全世界の書籍情報へのアクセスを目標にするオープンソースプロジェクトある。

9) 調査結果は2010年12月現在の情報による。

冊ヒットし、2冊は1903年の3版であったが、書誌情報のみで画像データはなかった。次にOpen Libraryでは、同じ条件で検索すると1件、第2版の2巻のみヒットし、オンライン上のデータ画像はInternet Archiveのデータと同じトロント大学のデータを利用していた。ただし英語名で検索すると4件ヒットし、書誌情報のみの初版の異なるデータが2つ、2版の先ほどのデータとそれとは別に書誌データのみのものが表示された。さらに、その関連項目に3版のデータが書誌のみ表示させているが、検索結果一覧では表示されないという状況であった。さらにGoogleBooksではどうかというと、GoogleBooksでは通常の検索では登録されている全ての書籍の全文検索を行うため、そのまま検索すると膨大な資料がでてくることが予想されるため、書名に限定して検索を行った。結果、1867年の書誌情報が1件ヒットし、書誌情報のみを取得することができた。

最後に同文献の所蔵が確認できている、ハーバードのHollis Classicで、参考までに同様の検索をしてみると、

Wade, T. F. (Thomas Francis), 1818-1895. *Yü yen tzü êrh chi : a progressive course designed to assist the student of colloquial Chines*, 1886

Wade, T. F. (Thomas Francis), 1818-1895. *Yü-yen tzü-erh chi, a progressive course designed to assist the student of colloquial Chinese*, 1867

の2件があり、1867の版については第1巻のみGoogleBooksで閲覧できるようリンク貼られていた。なぜこれが先ほどのGoogleBooks検索にヒットしなかったかは不明である。ひとまず以上の結果から得られた情報を整理すると、以下のような結果になる。

	Internet Archive	OpenLibrary	GoogleBooks
語言自邇集初版 vol.1		書誌情報のみ	○ Harvard
語言自邇集初版 vol.2		書誌情報のみ	書誌情報のみ
語言自邇集初版 vol.3		書誌情報のみ	
語言自邇集初版 vol.4			
語言自邇集 2 版 vol.1	○ Tronto		
語言自邇集 2 版 vol.2	○ Tronto	○ Tronto	
語言自邇集 2 版 vol.3	○ Tronto		
語言自邇集 3 版 vol.1	書誌情報のみ		
語言自邇集 3 版 vol.2	書誌情報のみ		

このように、現時点においてネットを通じて『語言自邇集』全てのデータを習得することはできず、検索サイトや図書館によっては情報が部分的にしかなく、一つの版をとっても不完全な状態でしかデータが得られないことが分かる。

昨今のデジタル化作業というのは非常に進歩しており、実際に文献のデジタル化は進んでいるが、あくまで個別図書館が個別の単位でデジタル化作業を行っているのであり、ある地域、時期、種類の文献を体系化するにはまだまだ不十分な点も多い。ある文献がどこに所蔵されているのかを知っているか、あるいは検索サイトで正確な書名を入力して、運がよくその文献のデジタル化されたデータがあつてはじめて目標とするデータにたどり着くことができるのである。そこで筆者は検索以外の文献へのアプロ

ーチの方法の構築とデジタルアーカイブズにおける文献目録の総合的な整理、及びデジタルアーカイブのための文献学の必要性があると考ええる。前者については、現在の構築中の「近代中国語文献資料データベース」が一つのモデルになればと考えている。ある特定の図書館のコレクションに限定した文献データベースではなく、ある書籍のジャンルを横断的に収集したデータベースサイトを目指している点というのがそれである。このデータベースを通じて、各地にあるデジタル化された文献を体系化できればと考えている。後者については、これまでの文献目録とは違い、どのデジタル文献がどの図書館に所蔵されているどの文献を撮影したもので、同様の書籍のデジタルアーカイブが他にどういったものがあるのか、あるいはどういった条件下で撮影されたデータなのかという点に関する情報も必要である。たとえばInternet Archiveでは『語言自邇集』2版を例にとると、次のような書誌項目を記載している。

Author: Wade, T. F. (Thomas Francis), 1818-1895; Hillier, Walter Caine, 1849-

Volume: 3

Subject: Chinese language - Grammar

Publisher: Shanghai, Kelly

Possible copyright status: NOT_IN_COPYRIGHT

Language: English

Call number: AFB-9854

Digitizing sponsor: MSN

Book contributor: Robarts - University of Toronto

Collection: robarts; toronto

Selected metadata

Copyright-evidence-operator:

Copyright-region

Copyright-evidence

Copyright-evidence-date

Scanningcenter

Mediatype

Identifier

Imagecount

Ppi

Lcamid

Rcamid

Camera

Operator

Scanner

Scandate

Identifier-accesst

Identifier-ark

Bookplateleaf

Sponsordate

Filesxml

一般的な書誌項目に追加して、どういった機材を用いて、どういった所蔵場所の書籍をいつ撮影したのかといった点についても詳しく記載されているのが分かる。おそらく今後このようなデジタル文献特有の情報も必要となっていくものと考えられる。

おわりに

本論では、『近代漢語文献資料データベース』を構築するに際して行ってきたことや考えてきたことを中心に述べてきた。その上で、現在の文献のデータベース化が取り巻く諸問題について議論を展開し、結論としてデジタルアーカイブズのための目的に応じた文献データベースの整備及びデジタル文献学の必要性について指摘した。今後この『近代中国語文献資料データベース』構築を通じて、これらの問題に対する答えを見いだしていければと考える。