

漢字文献 情報処理研究

第7号

漢字文献情報処理研究会 編
日本中国語 CAI 研究会 編集協力

好文出版

漢字文献情報処理研究 第7号

目次

論文	4	Microsoft Access を用いた明代私家書目データベースの作成	中川 諭
	12	GlyphWiki 開放型フォント開発環境の構築に向けて	上地 宏一
<hr/>			
漢情研 2006 年公開講座報告 国際化時代のデータベースとコンプライアンス	19		
	20	夏期公開講座レポート	上地 宏一
	22	東洋学情報化と法律問題——第4回 音楽／楽譜の校訂と著作権法（前編） 校訂権とその周辺（その二）	石岡 克俊
	27	中国の知的財産権関連の書籍と Web サイトの紹介	山田 崇仁
<hr/>			
特集 1：極める！ デジタルテキスト	31		
	32	私的検索のススメ	師 茂樹
	34	オンラインテキストアーカイブ概説	千田 大介
	42	データの入手と加工	二階堂善弘
	46	データの整形と整理	小島 浩之
	52	検索・置換・GREP	山田 崇仁
	60	お勧めユーティリティ——テキスト検索、一括処理ツール	小川 利康
<hr/>			
特集 2 人文学的な情報処理教育	65		
	66	なぜ理系と文系の議論はすれ違うのか	當山日出夫
	75	授業報告： 中国語テキストの構造化とデジタルドキュメント作成を同時に訓練する試み 「外国語」と「情報リテラシー」を同じ授業で	山崎 直樹
	84	情報歴史学のこれから 花園大学・情報歴史学コースの4年間をふり返って	師 茂樹
	91	BBS より：人文学的な情報教育とは	本誌編集局
<hr/>			
中国語 CAI 実践レポート	99		日本中国語 CAI 研究会
	100	ポッドキャストで中国語	清原 文代
	111	Canon wordtank G90	田邊 鉄

	114	CASIO EW-V3500L 電子辞書というデバイスの行く末	小川 利康
	119	「電子辞書座談会」	田邊 鉄
	121	W-ZERO3 で中国語	内田 慶市
<hr/>			
ソフトウェア	レビュー	125	
	126	OS・フォント	Windows Vista & Office 2007 / フォント / Windows Vista の日本語フォント環境 / Intel Mac 千田大介 / 清水哲郎 / 師茂樹
	139	アプリケーションソフト	一太郎 2006&ATOK2006 / WWW ブラウザ / EmEditor Ver.6 / Chinese Writer 8 / いきなり中国語 山田崇仁 / 上地宏一 / 金子真也 / 田邊鉄
<hr/>			
学術リソース	レビュー	153	
	154	学術サイト	中国 IT・ネット業界の動向 / Google / Google と学術情報 / 中国史 / 仏教 / Wikipedia とは何か / Wikipedia アンケートのまとめ / 二松学舎大学における日本漢文学研究の取り組み / mixi の「中国語学」関連コミュニティ 千田大介 / 師茂樹 / 小島浩之 / 山田崇仁 / 藤原敦 / 町泉寿郎・上地宏一 / 山崎直樹
	187	学術ソフト・製品	『中国基本古籍庫』—— 世界最大の漢籍データベース —— 二階堂善弘
<hr/>			
書評	190	『文字符号の歴史——欧米と日本編』/ 『これからホームページをつくる研究者のために』/ 『Chinese Culture Review —— 中国文化総覧』 Vol.1 ~ 3	
<hr/>			
コラム	30	わたしたちは子供の盗みを容認している	山崎直樹
	186	デジタル化が進む民国期刊行物 重慶図書館訪問記	小川利康
<hr/>			
お知らせ	185	連絡先変更および会費支払いのお願い	
	199	漢字文献情報処理研究会 入会のご案内	
	196	漢字文献情報処理研究会彙報 / 著者紹介	

- ◇ 本誌記事中のソフトウェア名、プログラム名、会社名などは一般に各社の商標または登録商標です。本文中では、™・®等のマークは明記していません。
- ◇ 本誌記事の記述に基づいて行われた作業の結果生じたあらゆる損害について、編著者・翻訳者および出版社は一切の責任を負いません。
- ◇ 本誌記事の内容に関するご意見・ご質問は、漢字文献情報処理研究会 Web サイト (<http://www.jaet.gr.jp/>)のフォームにて受け付けます。書面・電話・FAXによるお問い合わせには応じかねます。

Microsoft Access を用いた明代 私家書目データベースの作成

中川 諭（なかがわ さとし）

回 1. 明代私家書目とデータベース

古代中国では、前漢末の劉向『別録』・劉歆『七略』以来、歴代数多くの図書目録が編纂され、現代の我々に各時代の書籍についての状況や歴代の学術史などさまざまな情報をもたらしてくれる。時代が下って明代においても、焦竑曾『国史経籍志』・黄虞稷『千頃堂書目』などの目録が編纂され、当時の書籍の状況を伝えてくれる。

ところで明代の書目^[1]の中には、歴代の図書目録には見られない特徴を持つものがいくつか現れる。すなわち、高儒『百川書志』・晁瑛『宝文堂書目』・周弘祖『古今書刻』・徐焞『紅雨楼書目』などには、中国において近代文学が成立する以前は、少なくとも表向きは士大夫・教養人の読むべきものではないとされていた通俗小説・戯曲が多数著録されているのである。したがってこれらの目録を参照することにより、通俗小説・通俗文学に関して現存する作品からだけでは分からないさまざまな事象を知ることができる。

しかしながらこれらの書目は必ずしも一般的な四部分類によって分類されているわけではない。『百川書志』のように四部分類によって分類している書目であっても、いわゆる通俗文学を何処に分類しているかを判断することは意外と難しい。また『宝文堂書目』では、たとえば『唐詩品彙』が三個所別々に著録されているように、一つの書物が分類を異にして数個所に著録されることもある。このように索引が完備していないこれら書目

から特定の書名を検索することは容易ではない。

そこで有効な手段ではないかと考えられるのがコンピュータを用いた検索システムである。これらの書目をデジタルデータ化してデータベースを構築すれば、各書目の中において特定の書名を瞬時に検索できるだけでなく、各書目の横断的検索も容易に行えるだろう。そしてこのようなデータベースを用いることで、各書目の特徴を明らかにすることができるだけでなく、明代の通俗文学研究にとって大いなる恩恵を与えることも可能になるのではないだろうか。

本稿は、『宝文堂書目』・『古今書刻』・『百川書志』・『徐氏家蔵書目』の四書目を取り上げ^[2]、Microsoft社のリレーショナルデータベースソフト Access 2003（以下、「Access」と称する）による書目データベースの作成の手順と、このデータベースを利用した明代私家書目研究・通俗文学研究の展望について述べようとするものである。

回 2. 明代私家書目データベースの概要

まず作成するデータベースの概要を述べておこう。

データベースにもさまざまな形式が存在する。たとえば電子テキストもデータベースの一種と考えていいだろう。さまざまな中国の古典文献を電子テキスト化し、それをコンピュータ上で利用して語彙検索や語法解析が行われているのは、周知のことであろう。もちろん、図書目録も電子テ

Microsoft Access を用いた明代私家書目データベースの作成（中川）

キスト化すれば、書名検索が容易になり、検索面で紙媒体のデータよりも有効であることは間違あるまい。しかし図書目録は一般的な文献と異なり、そもそもそれ自体が一つのデータベースである。その利用方法は、単純に冒頭から順を追って読んでいくというものではなく、ある文献がどのように分類されているか、図書目録によって分類がどのように異なっているかという点がきわめて重要である。このような検索を行うためには、単純に電子テキストをワープロソフトやエディタの検索機能を用いて書名を一つの語句として検索するよりも、書名・分類項目などについて複数の書目を横断的に検索できるようなシステムを構築した方が、図書目録のデータベースとしてはより有効なものになるのではないだろうか。この点に、リレーショナルデータベースを利用するメリットがあると考えられる。リレーショナルデータベースソフトにも各社からさまざまなものが開発・販売されているが、本稿では最も一般的で広く普及しているであろう Access を用いることにする。バージョンは最新版の 2003 版を用いる。

データベースを作成するためには、当然のことながら、まず各書目のデータを電子テキストにしなければならない。これについては北京書同文デジタル技術有限公司（以下、「書同文」と称する）に依頼して行った。

そして電子テキストを用いてデータベースを作成するのであるが、各書目単独で書名や作者名が検索できるようにするだけでなく、本データベースで取り上げる書目を横断的に、同時に検索できるようにしたい。さらにその検索結果は、各書目における分類項目の構造がはっきりと分かるように表示させたい。そのために、各書目それぞれのデータを一つのテーブルとして作成し、それを基礎として横断的検索を行えるようにデータベースを設計したい。そのために、各書目のテーブルを一つにまとめるクエリを作成し、さらにそれを基にして書名検索を行うクエリを設定して、各書目を横断的に検索できるようにする。最初から複数の書目のデータを一つにまとめたテーブルを作成しても同じ結果を得られるだろう。しかし、各

書目ごとにテーブルを作りそれを結合することで、一つの巨大なテーブルを作成する必要がなくなり、データベース全体の容量を多少でも小さく抑えられると同時に、コンピュータの処理速度も速くすることができるであろう。

ただ一つ留意すべきこととして、各書目のテーブルを一つにまとめるクエリを作成する必要上、各書目テーブルのフィールド名を統一しておく必要がある。しかしながら各書目で分類の方法も著録の方法も異なっている。これを如何に処理すべきなのか。この点について事項で考察したい。

3. フィールドの設定

では、本稿で取り上げる『宝文堂書目』・『古今書刻』・『百川書志』・『徐氏家蔵書目』の著録の状況とデータベースのフィールドの設定について考えてみることにしよう。

晁瑛の私家書目である『宝文堂書目』は、晁瑛個人の蔵書を著録したものである。著録する書名の下に時折著者名・版本に関する情報などが小字で記されている。そしてその分類は「諸経総録」から始まるのであるが¹³、「易」・「詩経」・「春秋」など経書関係の分類の後は「史」・「子」・「文集」と続き、ここまではおおよそ四部分類の順序に従っている。しかしその後さらに「類書」・「子雑」・「楽府」など、一般的な四部分類であれば子部・集部の下に分類に属する項目が続いている。そしてこれらの項目の下にさらに小分類は設けられてはいない。すなわち『宝文堂書目』では分類項目のレベルは一段階しか存在しないのである。

『古今書刻』巻上は明代当時の中央および各地方の公共機関に蔵される書物の目録、巻下は碑文・法帖の目録である。巻上では書名の下に時折刊行年や著者に関する情報などが小字で記されることがある。また巻下ではほとんどの碑文・法帖名の下にその所在場所・執筆者などについて記されている。そしてその分類項目は、まず中央政府の各機関名が挙げられ、続いて「北直隸」・「南直隸」などの各地方政府名が続く。各地方政府名の下にはその地方政府の機関名が挙げられている。また

論文

書物の内容による分類項目はほとんど記されていないが、「南京国子監」や「福建」の「書坊」のように著録されている書物が比較的多い機関については、内容による分類が記されているところもある。つまり『古今書刻』では、分類項目のレベルは最大三段階存在することになる。

『百川書志』では著録される書名のほとんどに巻数と解題が記されている。そして基本的に伝統的な四部分類で分類される。経・史・子・集の四つの大項目の下に一段階の小分類が設けられている。すなわち『百川書志』では、分類項目のレベルは二段階存在することになる。

『徐氏家蔵書目』では、著録される書名のほとんどに巻数と解題が記されている。そして四部分類によって分類し、経・史・子・集各分類の下に分類項目が設定される。そして、たとえば『玉臺新詠』が「集部・総詩類・漢魏六朝」に分類されるように、さらに下の分類項目が設けられることもある。すなわち『徐氏家蔵書目』では、分類項目のレベルは最大三段階存在することになる。

以上四書目の著録の状況を整理すると、次のようになる。

- 宝文堂書目
巻数なし・説明あり・分類項目
- 古今書刻
巻数なし・解題あり・地方名・機関名・分類項目
- 百川書志
巻数あり・解題あり・四部分類・分類項目
- 徐氏家蔵書目
巻数あり・解題あり・四部分類・分類項目・分類小項目

このように四書目で著録の状況は同じではない。しかし先述のようなデータベースを構築する以上、これらの著録状況を包括するようにフィールド名を設定しなければならない。そこで、以下のようにフィールド名を設定することを考えた。

まず巻数については、一つの独立したフィールドを設定する。同じ書物であっても、巻数は書

目によって異なる可能性もあり、巻数を記していない書目もあることから、書名と巻数は分けてフィールドを設定しなければ、各書目を横断的に書名の重複検索をした際に、本来抽出されなければならないデータが、巻数の有無や相違だけのために検索されないことも起こりうるからである。「解題」について、『宝文堂書目』の「説明」は、『百川書志』の「解題」に比べて字数は大幅に少ないのであるが、これを『古今書刻』・『百川書志』の「解題」と同じフィールドに含める。また「分類項目」については、書物の内容による分類という点で四書目に共通しているので、これを一つの共通のフィールドとして設定する。『徐氏家蔵書目』の分類項目の下にある小分類については、常に現れるものではないので、「分類項目」フィールドの中に入れてしまう。もし小分類を下述の汎用フィールドに入力すれば、「分類小項目」が「分類項目」よりも前に表示され、表示の順序が乱れてしまう。『古今書刻』の「地方名」と「機関名」、『百川書志』の「四部分類」は項目の性質が全く異なり、そのままでは共通のフィールド名を設定することは不可能である。そのため、それぞれ固有のフィールド名を設定し、該当のないフィールドを空欄にする方法も考えられる。しかしそのような方針でフィールドを設定しテーブルを作成すれば、該当項目のない書目にはデータの入力されないフィールドが幾つも作られることになる上、今後さらに他の書目のデータベースを追加する際にも、書目の著録状況によってはさらに別のフィールドを追加する必要が出てくるだろう。そうすればますますデータの入力されないフィールドが増えることになり、データベースの活用にも不都合が出てくるかも知れない。そこで、たとえば「項目 1」・「項目 2」という名称をつけた、特にデータの内容を問わない汎用的なフィールドを設定してみてもどうか。こうすればデータベース内に空のフィールドを作成することを極力抑えられる上、これらのフィールドにはデータの内容に関わりなくフィールド内にデータを入力できるし、今後新たな書目のデータベースを追加する場合でも、かなりの部分で対応できるのではないだろうか。そして「分

Microsoft Access を用いた明代私家書目データベースの作成（中川）

類項目」フィールドは二つの汎用フィールドの後に置く。『古今書刻』の著録の状況から、汎用フィールドを用いる「地方名」・「機関名」の後に「分類項目」を置かなければならないからである。

以上のようなことから、本データベースでは次のようなフィールド名を設定することにした。

書目 ID	書名 ID	書名	巻数	解題
項目 1	項目 2	分類項目		

『宝文堂書目』では「項目 1」・「項目 2」の二つのフィールドが空欄になる。『古今書刻』では「項目 1」に地方名を「項目 2」に機関名を入力する。『百川書志』では「項目 1」フィールドが空欄になるが、「項目 2」フィールドに四部分類の分類を入力する。『徐氏家蔵書目』では「項目 1」フィールドが空欄になり、「項目 2」フィールドに四部分類を入力する。

回 4. 本データベースの作成手順

では、データベースの作成手順を述べることにしよう。

書同文に依頼した書目の電子テキストは Unicode テキストの形式である。これを Microsoft Word 2003（以下、「Word」と称する）に読み込むところから始める。そしてまずテキストデータと影印本を対校して、外字・異体字等を訂正する。この作業は、この研究費による組織に属する 3 名^[4]が分担して行った。そして「爲」・「為」のように、旧字体・常用漢字の違いなどによる字体の違いは、すべて旧字体に統一した。また「説」・「說」のように、Unicode のコードセパレート問題のある文字については、とりあえず JIS 規格の文字に統一した。日本語環境において検索が容易になるように考慮したためである。これらの文字は、漢字文献のデータベース構築において重要な問題であろう。異体字の問題を気にする

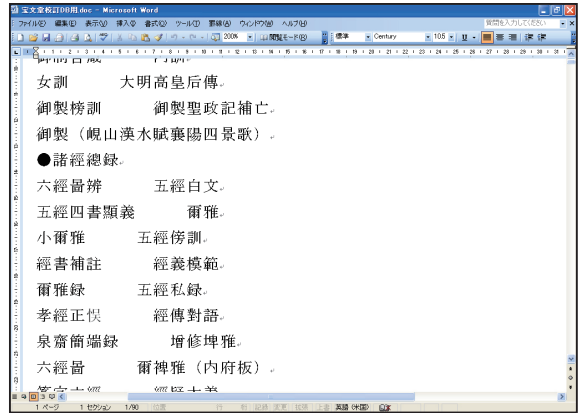
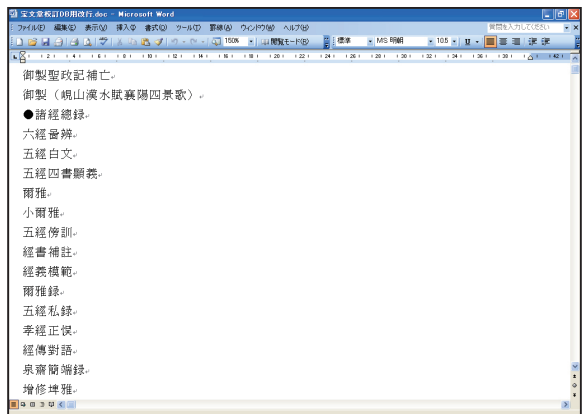


図 1 書同文作成のテキストを校訂

ことなく検索できるようなデータベース作成が理想であることは言うまでもない。本データベースにおいても、どのように異体字処理をすべきなのか、今後さらに考えていきたい。

Word 形式のテキストデータを Access に移してデータベースを作成するのであるが、ここで直接 Word から Access にデータを移すのではなく、表計算ソフト Microsoft Excel 2003（以下、「Excel」と称する）を経由する。これは、Access は Excel 形式のデータをインポートするのはきわめて容易であるだけでなく、Excel ではオートフィルの機能を利用して、連番となる書名 ID（これが Access において「主キー」となる）の入力が容易にできるからである。したがって、各書目別のデータシート（これが Access において一つの

図 2 一行に一書名となるよう成形



論文

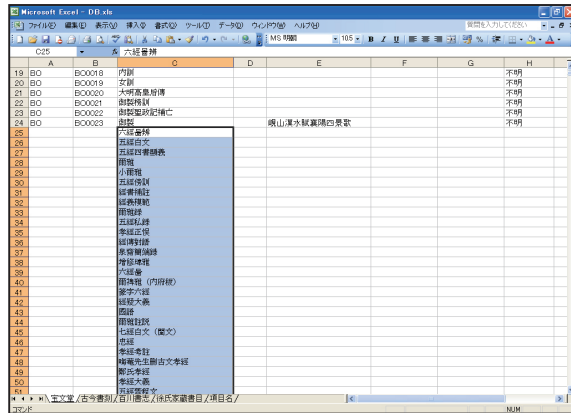


図3 Word から Excel へデータをコピー＆ペースト

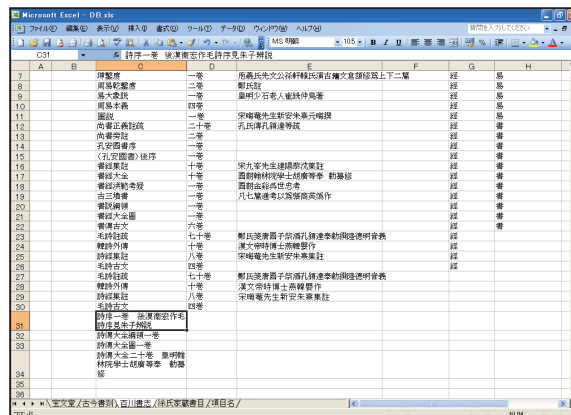
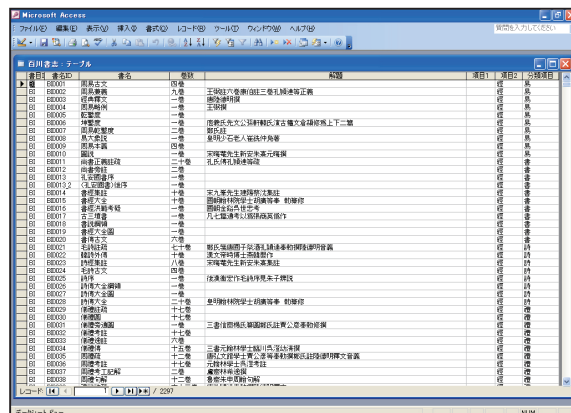


図4 Excel で「分類項目」の入力・「解題」の修正

図5 Excel から Access へインポート



「テーブル」となる)をあらかじめ Excel で作成し、それからそれらを Access へインポートする。

Word から Excel にデータを移すのは、単純にコピー＆ペーストで行えばよい。しかし、書同文作成のテキストデータは影印本の書式そのままであり、書目によっては一行の中に二、三の書名が収められている。そのためそれをそのまま Excel に貼り付けると、Excel での編集が面倒になる。そこで Word 上で「置換」機能などを利用して、一行ごとに一書名（解題などを含む）のデータのみとなるようにし、それから多少の微調整を行って、Excel へデータをペーストできるよう成形する。

Excel では先に設定したフィールド名を入力しておく。そこへ Word から Excel へ書名データをコピー＆ペーストする。Word からは解題等を含めた状態で Excel にコピーするため、Excel 上の「書名」フィールド上に解題などのデータが貼り付けられている。解題を「解題」フィールドに移すのは、一つ一つ手作業で行うしか方法がない。多少手間のかかる作業ではあるが、この作業を行う中でデータ入力ミスなどを発見することもある。この作業は筆者一人で行った。

そして「書目 ID」と「書名 ID」を入力する。「書目 ID」は、『宝文堂書目』に「BO」、『古今書刻』に「GU」、『百川書志』に「BI」、『徐氏家蔵書目』に「XU」というアルファベット二文字の ID を与えた。また「書名 ID」は「書目 ID」のアルファベット二文字に四桁の数字を付け加えた。

以上のような手順で、Excel で各書目別それぞれのデータシートを作成する。それが完成したら、続いてそれらを Access へインポートしてデータベースの構築に取りかかる。まず Access で何も設定されていない空のデータベースを作成し、インポートの機能を利用して Excel のデータを Access にインポートする。インポートする際にテーブル名を各書

Microsoft Access を用いた明代私家書目データベースの作成（中川）

目の名称にし、Accessに「宝文堂書目」テーブル・「古今書刻」テーブル・「百川書志」テーブル・「徐氏家蔵書目」テーブルを作成する。

さらに、「書目一覧」テーブルを作成する。フィールド名は「書目ID」と「書目名」のみの小さなテーブルである。このテーブルと後に作成する「全書目データ」クエリとの間にリレーションシップを設定することにより、本データベースの最終目的である書目の横断的検索の結果に書目名を表示させることができるようになる。

次に、Access上で四書目のテーブルすべてを合成するクエリを作成する。テーブルを合成するためには「ユニオンクエリ」を利用する。この時に四書目のテーブル相互にリレーションシップは設定しない。ユニオンクエリを利用するためにはSQL言語を用いなければならないが、そのSQL文は以下の通りである。

```
SELECT [書目ID],[書名ID],[書名],[巻数],[解題],[項目1],[項目2],[分類項目]
FROM [宝文堂書目]
UNION SELECT [書目ID],[書名ID],[書名],[巻数],[解題],[項目1],[項目2],[分類項目]
FROM [古今書刻]
UNION SELECT [書目ID],[書名ID],[書名],[巻数],[解題],[項目1],[項目2],[分類項目]
FROM [百川書志]
UNION SELECT [書目ID],[書名ID],[書名],[巻数],[解題],[項目1],[項目2],[分類項目]
FROM [徐氏家蔵書目];
```

このSQL文を保存しクエリを実行すると、四書目のテーブルが一つに合成される。このクエリに「全書目データ」という名称を付けておく。

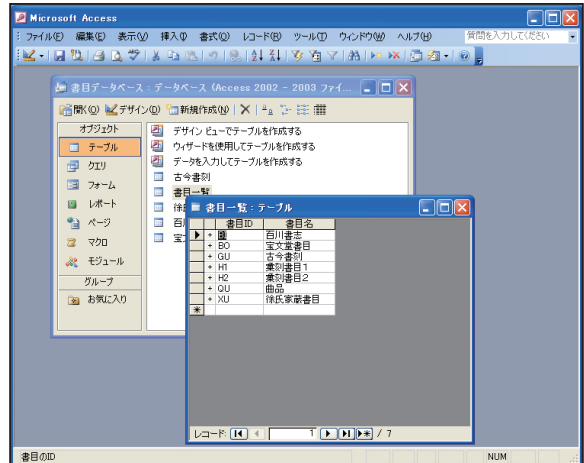


図6 「書目一覧」テーブル

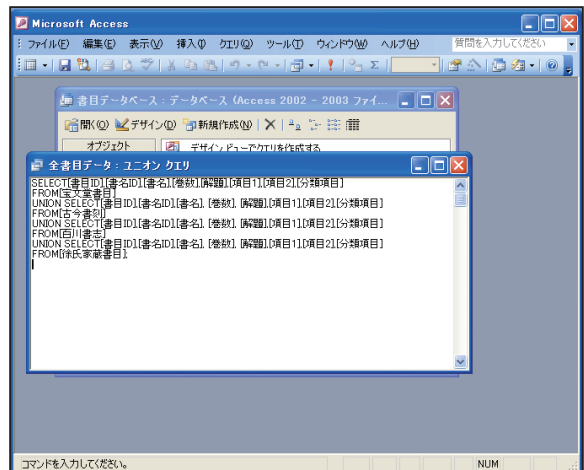


図7 ユニオンクエリのSQL文

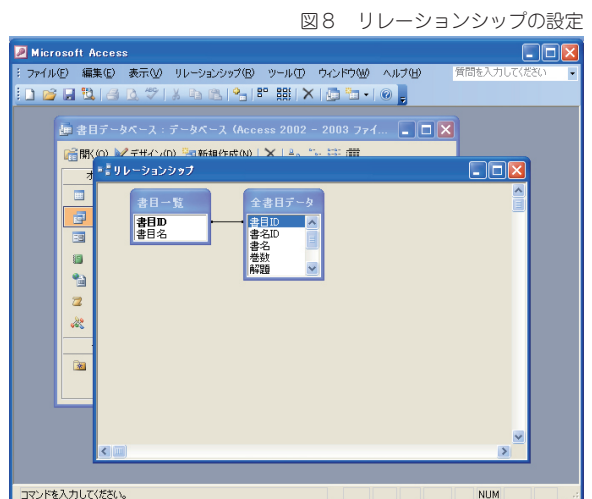


図8 リレーションシップの設定

論文

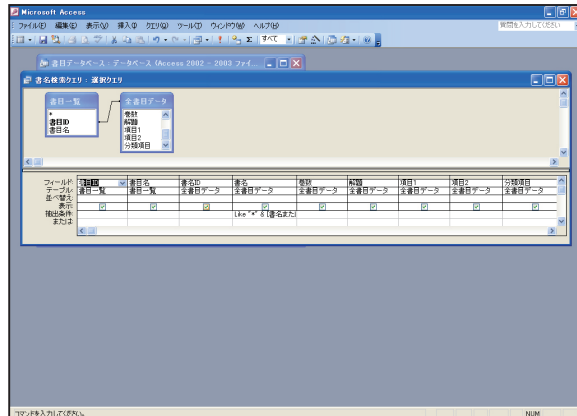


図9 「書目検索クエリ」のデザインビュー

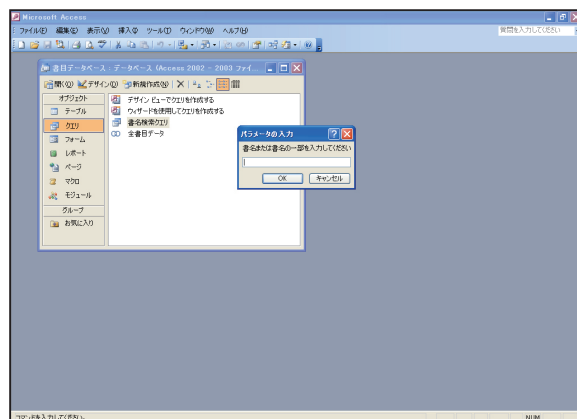
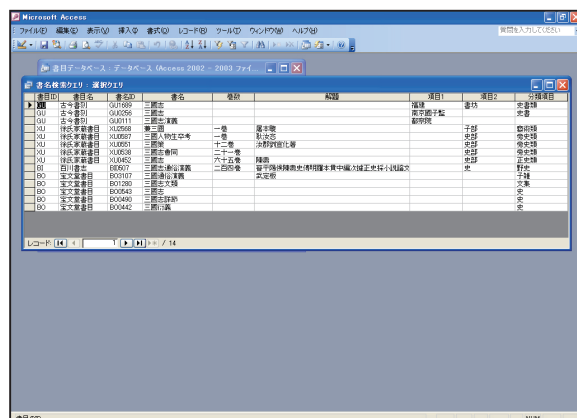


図10 書名検索クエリに検索条件を入力

図11 書名検索クエリを実行



「全書目データ」クエリを利用して、さらにもう一つクエリを作成し、四書目の横断的検索を行えるようにする。まず、先に作成した「書目一覧」テーブルと「全書目データ」クエリとの間にリレーションシップを設定する。これらに共通するフィールドは「書目ID」であるから、それぞれの「書目ID」フィールドを関連づける。

そして「全書目データ」クエリを基礎にして、そこにさらに「選択クエリ」を利用して、検索条件にパラメータを設定する。クエリを作成するためのデザインビューを表示させ、「テーブルの表示」ウインドウから「書目一覧」テーブルと「全書目データ」クエリを表示させる。新しいクエリのフィールド名には、「書目一覧」テーブルから「書目ID」を、「書目全データ」クエリから「書目名」・「書名ID」・「書名」・「巻数」・「解題」・「項目1」・「項目2」・「分類項目」を追加する。さらに「書名」フィールドの下の「抽出条件」に、

Like "*" & [書名または書名の一部を入力してください] & "*"

というパラメータを設定する。このパラメータによって、書名全体だけでなく書名の一部からも検索できるようになる。「書目検索」という名称を付けてクエリを保存し、実行する。

試みに検索条件として「三國」を入力すると、『宝文堂書目』から「三國衍義」・「三國志詳節」・「三國志」・「三國志文類」・「三國通俗演義」が、『古今書刻』から「三國志演義」・「三國志」（2データ）が、『百川書志』から「三國志通俗演義」が、『徐氏家蔵書目』から「兼三國」・「三國人物生卒考」・「三國策」・「三國志會同」・「三國志」が検索される。

5. 本データベースの意義

以上のような手順で、複数の書目を横断的

Microsoft Access を用いた明代私家書目データベースの作成（中川）

に検索することのできるデータベースを作成することができた。今後新たに別の書目を追加する場合には、やはり同様の手順で、まず Excel でその書目のデータシートを作成し、それから Access にデータシートをインポートしてテーブルを作成し、さらに「全書目データ」クエリの SQL 文に

```
UNION SELECT [ 書目 ID ], [ 書名 ID ], [ 書名 ], [ 解題 ], [ 解題 ], [ 項目 1 ], [ 項目 2 ], [ 分類項目 ]  
FROM [ ○○書目 ]
```

を付け加えればよいことになる。データ追加も比較的容易にできるであろう。

また強力な機能を持った Access を用いたデータベースであるからこそ、研究テーマによる、あるいは利用者の目的によるカスタマイズが可能である。単に検索するだけでなく、検索結果を基にさまざまな分析を行うことができる。研究者の研究テーマは人それぞれであり、したがって書目データの利用のしかたもそれぞれ異なっていて当然である。研究目的に応じて最もふさわしい形でデータベースを利用することができるので、より有効な活用のしかたが期待できるであろう。これこそが本データベースにおいて Access を用いた最大の理由である。またインターネットに接続する必要がないので、何らかの事情でインターネット接続できない環境にあっても Access さえコンピュータにセットアップされていれば、本データベースを利用することができる。これもメリットの一つであるかも知れない。

ではこのデータベースを用いて、通俗文学研究にどのように役立てればいいのかのだろう。もちろんさまざまな研究テーマが考えられ、それに応じ

た利用法があるだろう。中でも横断的検索ができるという点に着目し、ある一つの小説ないしは戯曲作品が何処に分類されているか、それが書目によってどう違うかを検討することにより、明代当時の小説など通俗文学に関する意識というものを探っていけるのではないかと。また、通俗文芸作品に限らず、その他様々な書物の分類を各書目相互に比較することを通して、その書目の分類方法の特色を考察していけるかも知れない。

本データベースが明代通俗文学研究に少しでも寄与するところがあれば幸いである。

注

- [1] 本稿で言う「書目」は、小説・戯曲など通俗文学を著録している明代の蔵書目録、あるいは本データベースにおいて取り上げる蔵書目録の意味で用いる。
- [2] 底本は『明代書目題跋叢刊』（書目文献出版社、1994年）所収の影印本による。また『宝文堂書目』については、『晁氏宝文堂書目 徐氏紅雨樓書目』（古典文学出版社、1957年）所収の活字本を参照した。
- [3] 『宝文堂書目』は現在中国国家図書館に抄本が残っているだけであるが、その冒頭部分は欠落している。したがって、「諸経総録」よりも前に著録されている書名について分類項目名は不明である。
- [4] 二階堂善弘氏を代表とし、中里見敬氏と筆者の3名で組織された研究グループである。

〔付記〕

本稿は、平成16・17年度科学研究費補助金基盤研究(C)「中国古典通俗文芸研究のための書目データベースの構築と公開」（研究代表者、二階堂善弘）による研究成果の一部である。

GlyphWiki

開放型フォント開発環境の構築に向けて

上地 宏一（かみち こういち）

回 はじめに

漢字文献情報処理の研究分野において、この十年間の研究環境の発展は目を見張るものがあったと述べても過言ではなく、これを踏まえて今後十年の学問的発展が大いに期待できるところである。

文字情報処理のベースとなる文字コードは、中国古典を中心とする数万の漢字を処理できるようになり、実際に具視化するために必要なフォントについても Windows 次期バージョンに標準で搭載されることになった^[1]。研究の対象となる大規模古典テキストについては、『四庫全書』、『四部叢刊』をはじめとして大量のデータが全文検索可能な形で入手可能であり、また現代文の文字コードについても各所で蓄積されている^[2]。これらの大量の情報を処理するために、高性能、大容量のコンピュータ環境が安価に入手できるようになり、インターネットを通じて情報の入手、研究成果の発信が容易になったのである。

しかしながら、別の側面に着目してみると、漢字文献情報処理の要の一つであるフォント環境について、次の二つの問題がおざなりになってきた感がある。

- 標準的ではない文字の処理環境の構築
- 自由な日本語漢字フォントの不足

旧来「JIS 漢字」と呼ばれてきた JIS X 0208 規

格に収録されている六千字以外の文字を処理するために、ユーザが自分でグリフを定義する「外字」（ユーザ定義文字）が活用されてきた。その後、高品質なアウトライン形式フォントの普及により、外字は「必要に応じて自分で作る」ものから「他者によって用意されたものを使う」形態へと変わってきた。これは、ボランティアによる提供や外字フォントの市場が成熟してきたと見ることもできるし、逆に外字処理環境としては衰退しているとも言えよう。

文字コードの拡張により外字の必要性が低くなったとはいえ、研究分野によっては未だに外字が必要である。また、インターネットなどによる情報交換の保証のためには、外字を利用する場合に統一的な外字処理の枠組みが必要であるにもかかわらず、これまで一つの方式が定着することはなかった。その理由としては外字処理の必要な分野が分散し、外字メーカーや組織ごとの外字の統一がなされなかったことと、OS におけるフォント処理の技術的難易度が高かったためと考えられる。

また、外字処理にも関係のあるフォントの二次利用について、フォントグリフのデザインそのものは著作権では保護されないという解釈が一般的であるが、代わりに使用許諾（ライセンス）によって、ユーザはフォントの使用形態を制限されているのが実情である。近年広まっているフリーソフトウェア運動に相反して、自由（ここで言う自由とは、単純に文字の表示や印刷にとどまらず、グ

リフの改変といった二次利用や、再配布、電子文書への埋め込みといった複製の自由も含む）に使える日本語フォントというのは、ほぼ存在しないだけでなく、フォントを自由に使うためのツールもそろっていないのが実情である。結局のところ、フォントは「自由に使えるもの」ではなく「他者から与えられたものをそのまま使う」状態にとどまっている。つまり、この十年間の日本語漢字フォント環境は停滞ぎみであったとも言えるのである。

◆「集合知」と日本語漢字フォント

近年、GNU/Linux や Mozilla に代表されるようなフリーソフトウェアの機運が高まっている。また、Wikipedia やソーシャルブックマーク、BBS やブログ（のある部分）など、インターネット上における知の結集の試みが注目され、実際に活用されている。いずれも専門家とは限らない任意の人々が参加できる開放型環境によって作り上げられた知の集合である。当然のことながら、集合知は衆愚の産物と成り果てる可能性もあり、成功するかどうかはさまざまな要素が絡んでいる。

多人数による分散ソフトウェア開発についての研究は古くから行われており、情報共有による開発工程の明確化や、部品ごとにプログラムを作成する場合のインターフェース（どのような入力に対して、どのような出力があるか、その詳細）の必要性などの要素が確立されている。さらに所属や立場の違う集団によってソフトウェアを構築するという開放型のソフトウェア開発形態は比較的新しいものであるにも関わらず、多くの成功事例が存在する。同様にフォントについても多人数による開放型開発が可能ではないだろうか。

そこで Wikipedia^[3] を比較対象としたフォントの開放型開発環境構築の検討を行った。Wikipedia の評価はさまざま^[4] ではあるが、実際に運用され多数の言語で大量の知識が蓄積されつつある。その過程で生じている問題点とその対処方法は大いに参考とすることができる。

□ 「GlyphWiki」の提案

本稿では、停滞している日本語漢字フォント環境をこの集合知によって打破できるのではないかと、という仮説を立てた。言い換えると、集合知による開放型フォントの開発および活用が可能な環境を提案するものである。この環境を「GlyphWiki」と名づけた。

従来の日本語漢字フォントは、ごく少数のデザイナーによって作られてきた^[5]。数千の字種をデザインするためには、デザインの統一性が重要であると考えられたためであり、裏を返せば、専門的でない大規模な集団が日本語漢字フォントを作ることは考えられなかった。この概念を覆し、開放型のフォント開発環境を構築して実際にフォントの開発が可能かどうかを検証する。結果として低コスト（時間）でフォントを制作できることが予想される。たとえば台湾や香港を含む中国で流通しているフォント、または日本で流通している日本向けのデザイン調整がなされた外国産フォントの一部は、筆画や部品配置による省力化デザイン方式によって生成されたものがあり、日本の伝統的なフォントメーカーによるフォントと比較すると時間的デザインコストが低い。これらのフォントは高品質とはいえないものも複数存在するが、実際には日本市場において広く受け入れられている。また ISO/IEC 10646-2 における Ext. B の全漢字集合を網羅したフォントについて、日本では未だ実装されていない現状では中国デザインの Ext. B フォントを利用せざるを得ない。このケースを見ても分かるように、必ずしもフォントデザインが最高品質である必要はなく、利用者のニーズに応じた開放型によるフォント開発が受け入れられる余地があると言える。

◆ Wiki システムによるグリフ管理

GlyphWiki は Wiki の派生環境 (Wiki クローン) で、フォント (グリフ) の開発に特化するために Wikipedia のシステムを参考に独自に実装する。Wiki 名 (Wiki の 1 ページ) は、グリフ 1 つに相

論文

馬鳥比鹿食止正齒至致片鹵車兩
舟竹足牛羊用甫甬角革身豕骨耳
目自首頁頭言詩隹誰肌殳設講石
学旦早草旬木本末末朱米来采禾
精池国門間区同広待貝火炎夕死
王玉納糸系内辻迫迢走超召開聞
考遠迫慶應環部藤漢立主氏衣合

図1 基礎データのサンプル

当し、自由な名前をつけられる。実際に蓄積されるグリフは次の4つの集合になると考えている。

- 既存の文字コードポイントに相当するグリフ
- 既存の文字コードポイントの異体字に相当するグリフ
- 私的外字集合
- 漢字を構成する部品

GlyphWiki は、用途を限定しない。つまり、自由に使えるフォントを作るために、不足している既存のコードポイントのグリフを埋める開発者もいれば、既存のグリフを活用して自分のための外字グリフ集合を作成する開発者も想定する。また、漢字一つ一つを独立してデザインするよりも、用意されている部品を活用したほうが開発が容易であり、そのための部品セットを用意すべきである。

これらを踏まえて、Wiki 名には予約フレーズ（予約 Wiki 名形式）が設定されている。たとえば「jis-x0208_1676」は JIS X 0208:1997 規格の 16 区 76 点、つまり「一」のグリフに相当する予約 Wiki 名である。このような既存の規格や標準、グリフ集合のために予約フレーズを確保する。

グリフ開発の基礎データとして、「漢字データベース計画 (Kanji Database Project)」¹⁶⁾ が 2003 年度日本学術振興会科学研究費補助金・研究成果公開促進費（データベース）（名称：より適切な漢字処理のための基礎データベース 代表者：花園大学 師茂樹）を受けて作成した JIS X 0208:1997 の漢字部分のグリフデータを登録し

ている。このデータを用いることで、多くの漢字グリフについては部品の組み合わせでデザインすることが可能である（図 1：基礎データのサンプル）。

また、GlyphWiki では開発者の ID 登録を行うことが可能である。ID を登録したユーザは私的な Wiki 名を命名（新規作成）し、編集できる。たとえば「kamichi」というユーザは「kamichi_（自由な名前）」という Wiki 名を独占して編集できるため、私的な Wiki 名を付与したグリフの集合を登録できる。ただし、命名および編集の権利が担保されるだけであり、編集されたグリフは誰でも利用が可能である。また別の開発者がそのグリフを基にしてさらに別のグリフを作ることや、グリフデータを別の Wiki 名にそのままコピーして別のグリフとすることも可能である。開発者はあるテキストデータに含まれる外字集合をまとめて登録し、GlyphWiki を通じて外字フォントを公開することが可能となる。

Wiki 形式のグリフ管理システムであるから、グリフの編集は対象となるグリフ（Wiki 名）を開き、編集し、登録する、という 3 ステップの繰り返しで実現することになる。

Wiki ページは、グリフのデータを登録するだけでなく、メタ情報を登録することができる。このメタ情報は、ユーザによるグリフ検索のために付与されるものと、グリフ同士の関係をデータ化するための 2 種類に分けられる。具体的には、検索の便として漢字の読み、字書や文字データベースなどの出典情報（認識番号）や、実際のテキストでの使用例の記録などが考えられる。グリフ同士の関係付けとしては、別のグリフへのエイリアス（同グリフとみなす）、または関連付け（異体字関係としてのリンク）を行う。たとえば別文字コード規格における同一文字へのリンクといったメタ情報のみ Wiki ページも存在しうる。さらにこれらのメタ情報を活用して、あるグリフに対する異体字関係グリフの列挙や、部首ごとのグリフ集合の表示など、フォント管理支援を実装する。

開放型開発で重要なのは、誰がどのように改変

を行ったかの記録である。そこでデータの変更履歴はすべて記録する。ID 登録者の場合は ID を、未登録者の場合は IP アドレスを代替として開発者の同意情報とする。また、過去のすべてのバージョン履歴のデータを将来的に参照できることとする。たとえばバージョンを 1 から増加していく正の整数値で表すとして、「jis-x0208_1676@12」という表現によって「jis-x0208_1676」という Wiki 名のバージョン 12 のグリフデータを指すことができる。このため開放型開発によって絶えずデータに変化が起こりうるとしても過去のデータを選択することが可能であり、ユーザは安心して利用できることになる。

◆ KAGE システムによるグリフ編集

従来のアウトライン情報の操作によるグリフ編集は非常に難易度が高く、誰にでも行えるものではない。フォントの開放型開発という新しい概念を構築するには、グリフ編集方式の改良が必要である。そこで GlyphWiki では、KAGE システム^[7]を採用し、字形の骨格情報の操作による平易なグリフ編集を可能としている。

KAGE システムの利用により、Wiki に登録されている他のグリフを部品として任意の大きさ、場所に配置して、グリフのデザインを行うことが可能である。単純に部品を組み合わせるだけではデザイン品質が低くなるため、部品を分解して個別の筆画について調整することも可能である。

また KAGE システムのデータ構造をさらに簡略化し、筆画の端形状の情報を原則「未接続」と「接続」、「跳ね」の 3 種類に統合する（従来「接続」は「箱状の左上角」、「横棒に接着」等に、「跳ね」は「左跳ね」、「右跳ね」などと細かく分類されていた）。統合された端形状はシステム側で自動判別して具体的な形状を表現することとした。さらに白紙の状態からグリフをデザインするとき、キャンバス上に書いた手書きの文字を認識してグリフデータを生成する機能も用意する。これにより飛躍的にグリフデザインコストを抑えることが可能である（図 2：上段：実際に

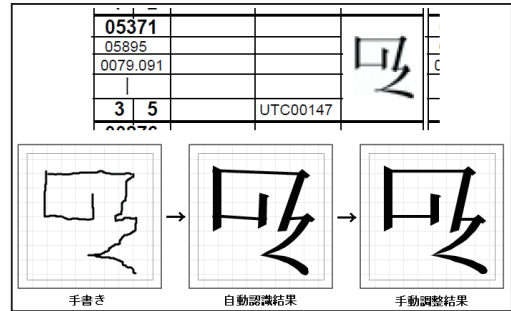
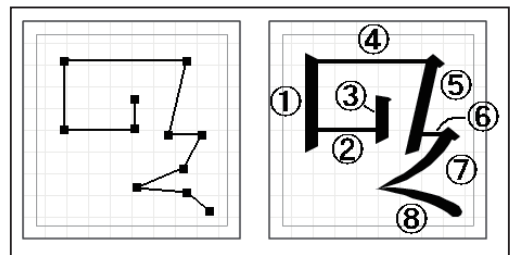


図 2 上段：実際に ISO/IEC 10646 に提案された文字
下段：手書き認識機能を利用して上段の文字をデザインしたもの

ISO/IEC 10646 に提案された文字。下段：手書き認識機能を利用して上段の文字をデザインしたもの。

筆画線種は KAGE システムの 5 種類（直線、曲線、複雑な曲線、縦はらい、直線折れ）をそのまま利用する。このため、線種を元に 2,3 個の座標と端形状を指定すると、各座標の位置関係から自動的に筆画種別が決定されキャンバス上への描画が行われる。それを微調整することでグリフデザインが完了する。

図 3・表：図 2 のグリフデータ



接…他ストロークに接続している 未…未接続

ID	線種	端形状		座標 1		座標 2		座標 3	
		始	終	X	Y	X	Y	X	Y
1	直線	接	接	39	37	39	100		
2	直線	接	接	39	100	104	100		
3	直線	未	接	104	69	104	100		
4	直線	接	接	39	37	151	37		
5	直線	接	接	151	37	136	104		
6	直線	接	接	136	104	165	104		
7	曲線	接	未	165	104	148	139	100	154
8	曲線	未	未	98	154	147	158	175	177

論文

骨格情報によって表現されたグリフは、明朝体およびゴシック体の特徴を持つ肉付けエンジンによって字形が具視化されることとなる（図3・表：図2のグリフデータ）。

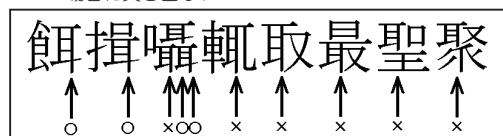
※ 他形式データの登録

PNG, SVG 形式などの一般的な画像形式によって保存したグリフ画像や、自由利用可能な他のフォントに含まれるグリフデータを GlyphWiki に登録する機構も考えている。これらのデータについては、KAGE システムを利用したのグリフ編集ができないため、図形情報そのものを登録することとなる。たとえば金文、甲骨文、西夏文字などのデータや、ロゴ・アイコンの登録が考えられる。また、現在の KAGE システムでは表現できないグリフ（例：U+3403, U+3514, U+20137, U+219B9, U+26B99）にも活用することになる。

※ フォントとしての利用

登録したグリフは、直ちに PNG 形式や SVG 形式の画像ファイルとして利用することができる。これは Web ページへの埋め込み外字画像として GlyphWiki を活用する利用形態を想定したものである。また、任意のグリフ集合に対して、Unicode のコードポイントの対応関係を指定することによって TrueType 形式のフォントファイルを生産することも可能である。これは外字ファイルの作成を想定したものである。このほかに、予約 Wiki 名として登録されている既存の文字コード等に対応するフォントファイルを定期的に生成して入手可能な状態とすること（いわゆるナイトリービルド）も考えている。このほかに Adobe 社の提供する SING^[8] とよばれる外字アーキテクチャへの対応も予定している。具体的には

図4 フォント「MS 明朝」における部品「耳」の扱い
MS 明朝では「耳」部品の右に別の筆画が存在する場合に突き出ない



GlyphWiki の各 Wiki グリフを SING のグリフレットとしてファイルに出力する。このように複数の利用形態により実用的なフォント環境がもたらされる。

回 統一的なフォントデザイン

多人数によるフォント開発で一番の問題となるデザインの統一性の確保については、次のようにその種類に応じた対策を考える。ただし、統一性を求めるのは、既存の文字コード規格等に対する予約 Wiki 名のグリフのみであり、開発者が独自に作成する漢字集合については、いかなるデザインも許容することになる。

※ 字体の統一性

たとえば「耳」の5画目が縦棒から突き抜けるかどうか、「文」の筆押さえの有無など、グリフの骨格部分に相当する字体デザインの統一性について、既存の文字コード規格に対するグリフの場合は原則的に文字コード規格の例示字形に準拠するものとする（図4：フォント「MS 明朝」における部品「耳」の扱い。MS 明朝では「耳」部品の右に別の筆画が存在する場合に突き出ない）。

また、同じ部品を利用する他のグリフと一覧形式による比較表示ができるようにすることで開発者がその場でチェックすることができる。

※ ベースラインやボディサイズの調整

一つの文字をデザインしたときに、その文字を文書中に埋め込んだ場合の見目の調整が必要である。このため、デザイン対象の文字を任意の文書に当てはめた表示チェックや、ランダムに選ばれたグリフ集合との一括表示によって、ベースラインやボディサイズの調整を行うこととする。

※ 議論ノートの活用

あるデータの改変において、それをどのような意図で行ったのか、また他バージョンに対するコメントの付与などを、Wikipedia と同様にノート機能を実装することで実現する。これにより、複

数の開発者が同じグリフを対象とした編集行為を繰り返したときに、お互いの意図を確認し、事態の收拾を図る手段となることが期待される。

※ 既存グリフの分析による推定値の導出

大量に蓄積されているグリフデータを分析し、あるグリフの部品の配置位置や大きさ、ベースラインなどの値を推定することが可能となる。その推定値およびランダムな要素を加味してデザイン調整を行ったグリフを複数提示し、開発者がその中から最も適切と判断したデザインを新たなデータとする調整方法が考えられる。これを応用すれば、IDS^[9]で表現された漢字集合のグリフデータを機械的に仮生成することが可能となり、さらにデザインコストを抑えることが可能となる。今後審議される ISO/IEC 10646 の Ext.C2 集合からは、収録文字候補の提出時に IDS データの提出を行うことが提案されている^[10] ため、実現すればフォントセットを部分的ではあるが自動機械生成することができる。

※ グリフデータの権利管理

開放型開発によって自由な日本語フォント環境を実現するためには、グリフデータの権利管理に注意する必要がある。現在の日本ではタイプフェイスに著作権が発生しないと解釈するのが一般的であるが、GlyphWiki に登録されるグリフデータについては、Wikipedia と同様に GNU GFDL もしくは GNU GPL によって投稿されるものとみなす。先に述べたように、開発者は ID を登録し、ログインしている場合は、投稿者を ID によって同定する。ログインしていない場合や ID を登録していない開発者は IP アドレスによって記録されるが、IP アドレスは必ずしも開発者ごとに固定されるわけではないため、同定情報としては便宜的なものとなる。

GlyphWiki に登録されたデータが、自由利用の認められないデータであると判明した場合には、登録者との協議により該当データの削除といった対応をとることになる。また、このようなトラブルに対応する管理業務を行うメンバの設置が必要

であると考えている。

回 さらに期待される効果

主目的である自由な日本語漢字フォント環境の実現と、多人数によるフォント開発の可能性の検証だけでなく、以下のようないくつかの知見が得られるものと期待できる。

※ 外字データベース

GlyphWiki は同時にネットワーク共有型の外字データベースでもあるため、さまざまな分野、用途における外字データベースの集積が可能であり、積極的な活用が期待される。そのためには既存の文字データベースのインポートも視野に入れている。

※ デザイン情報

あるグリフに対して施されたデザイン改良のデータ値を抽象化し分析することで「美しい文字」をモデル化することが可能となる。また、編集履歴や議論ノートの集積を分析することで、大規模フォント開発のための字体・デザイン統一のための編集ガイドラインが提案できるのではないかと考えている。

回 おわりに

GlyphWiki は現在 Wiki システムの構築を行っている段階であり、2006 年度中に運用とその評価を行う予定である。また、IDS データによるグリフ生成の自動化についても、CHISE プロジェクト^[11]において蓄積されている多数の IDS 情報をもとに生成実験を行う予定である。活発なグリフ開発が行われ、一通りの検証が達成できた場合は、別の書体や別のウェイトについても着手したいと考えている。

GlyphWiki の運用に際してフォントの専門家だけでなく、協同作業によるフォント開発に興味のある人や、外字処理の実務者など、多数の参加を期待し、忌憚のない意見を乞うものである。

論文

GlyphWiki : Web サイト

<http://glyphwiki.org/>

注

- [1] 本誌ソフトウェア・レビュー「Windows Vista & Office 2007」を参照。
- [2] 張玉潔、山本和英「中国語のコンピュータ処理について—コンピュータによる中国語処理の発展と課題」『漢字文献情報処理研究』第6号、pp.102-109、好文出版、2005。
- [3] <http://wikipedia.org/>
- [4] 本誌インターネットリソースレビュー「Wikipedia アンケートのまとめ」および「Wikipedia とは何か」を参照。
- [5] 「文字を作る人たち—株式会社モリサワ」『プラスデザインング』2006年夏号、pp.138-141、毎日コミュニケーションズ、2006。
- [6] <http://kanji-database.sourceforge.net/>
- [7] 上地 宏一「KAGE - An Automatic Glyph Generating Engine For Large Character Code Set」『書体・組版ワークショップ報告書』、pp.85-92、京都大学 21世紀 COE 東アジア世界の人文情報学研究教育拠点、2003。
- [8] 山田崇仁「Adobe InDesign CS2」漢字文献情報処理研究』第6号、pp.184-186、好文出版、2005。
- [9] 上地宏一「漢字フォント自動生成サーバ"影 KAGE"の構築—文字コードの枠組みを越える次世代漢字処理の提案—」『漢字文献情報処理研究』第3号、pp.4-13、好文出版、2002。国際規格としては ISO/IEC 10646-1:2000 Annex F を参照。
- [10] ISO/IEC JTC1/SC2/WG2/IRG 第25回会議資料 N1162 Suggested Procedural Principles for C2 Work。
- [11] <http://www.kanji.zinbun.kyoto-u.ac.jp/projects/chise/>

漢情研 2006 年公開講座報告

国際化時代の データベースと コンプライアンス

本公開講座もはや四年目である。毎回、講師としてお招きしている法学者の石岡克俊氏を中心に、中身の濃い議論が行われてきた。

これまでは、主に「データを作成する際に生ずる問題」を大きなテーマとして設定してきたが、本年はこれとは別な視点である「作ったデータを公開する際に生ずる問題」をテーマとして設定し、これについての法的知識の吸収や問題点の共有を議論した。

本会会員や本誌読者層の多くが、中国などの外国を主要な研究対象としている方だと思われる。そこで扱う材料は当然外国の著作物である以上、データベースなどの手段でそれらの材料をデジタル化して公開するには、相手国の著作権法のみならず、（中国がWTOに加盟したこともあって）多国間を包括する国際的な著作権管理の枠組みをも理解しておく必要がある。今回は、国際的な著作権管理の枠組みを理解し、また、外国で生ずる著作権紛争に関わる問題などについて理解の端緒を得ることができたように思う。

本年度も、本会 BBS に専用会議室を設けて議論を行い、それを踏まえた形で石岡氏によるご講演が行われ、更に会員諸氏による活発なディスカッションが行われた。

以下は、公開講座の講師による論攷、および報告である。

※漢情研 2006 年公開講座の日程・会場等の詳細については、彙報 (P.196) を参照していただきたい。

Contents

夏期公開講座レポート	上地 宏一	20
東洋学情報化と法律問題——第 4 回		
音楽／楽譜の校訂と著作権法（前編）——校訂権とその周辺（その二）	石岡 克俊	22
中国の知的財産権関連の書籍と Web サイトの紹介	山田 崇仁	27

夏期公開講座レポート

上地 宏一（かみち こういち）

◇ はじめに

2006年7月22日、慶應義塾大学日吉キャンパスで開催された漢字文献情報処理研究会2006年度夏期公開講座「国際化時代のデータベースとコンプライアンス」についての簡単な報告を行う。慶應義塾大学産業研究所の石岡氏を講師にお迎えしての本講座は、今回で4回目となり、ますます成熟した内容となったのは嬉しい限りである。

今回は、過去の講座での議論において度々問題となっていた国際的な著作権問題についての法解釈や実務的な理解を中心とした講演が行われ、その後に漢情研BBSで募集した質問、および会場からの質問を踏まえた議論が行われた。本稿は、当日の内容を簡単にまとめたものである。

◇ 日本国著作権法と国際的保護

日本国著作権法による保護の対象とは、著作物、実演、レコード、放送、有線放送が挙げられるが、著作物を例とした場合に、第6条では次の3種類について、著作権法の保護を受ける、としている。

- ①日本国民（法令に基づく、または、国内に主たる事業所を有する法人も含む）
- ②最初に国内において発行された著作物
- ③条約により日本が保護の義務を負う著作物

特に③は、後述の国際条約の義務履行主体が主権国家のみであることを受け、相互主義の原則に基づいて外国の著作物の保護を可能とする規定で

ある。

◎ 日本は国際条約の優等生

著作権関係の条約としてベルヌ条約、万国著作権条約、実演家等保護条約、レコード保護条約、WTO設立協定、WCT、WPPTの7つが紹介されたが、これらの条約について日本はすべて加盟、批准を行っている優等生である。

例えばベルヌ条約の批准年について、日本は1889年であるが、アメリカは1989年である。もちろん、単純に日本が著作権に対するの制度化が早かったというわけではなく、アメリカや中南米諸国ではベルヌ条約とは異なった法体系をとっていた^[1]ため、国内法との兼ね合いにより批准に時間がかかったというのが真相である。とはいえ、日本は殆どの条約について、条約締結と相前後して国内批准を終えていることは事実である。

◎ 国際的な著作権保護

先述の通りアメリカを中心とした、国内法の制約によりベルヌ条約を批准できない国との橋渡しの条約として万国著作権条約がある。この条約によって加盟国同士の場合に他国の著作物を自国の著作物と同様に扱うことが可能となった。

また香港や台湾は、ベルヌ条約と万国著作権条約双方に未加盟だが、このうち香港は中国の著作権関係条約が適用され、台湾についても、WTO設立協定に加盟しており、まったく著作権が及ばないわけではない。WTO加盟国は「ベルヌ・プラス・アプローチ」という、ベルヌ条約非加盟であっても、ベルヌ条約以上の著作権保護が義務化されている。これは一見矛盾しているように見え

るが、WTO 設立協定は貿易をターゲットとしたものであり、著作権人格権など一部の権利については対象外とされるため、ベルヌ条約を批准できないケースにも対応できたとのことである。また、WTO 自体の加入主体が国家に限定されておらず、地域も参加できるという特徴がある。

このように、漢字文献を扱う上で主な対象となる中国や台湾、香港についても、日本と同様の著作権保護がなされるということが理解できた。

◇ BBS や会場からの質問への回答

後半は、各種質問に対する回答とそれに対する議論が行われた。

◎ ローマの休日 DVD

タイムリーな話題として、映画「ローマの休日」DVD 販売差し止め請求判決に関する質問があり、「保護期間の計算」についての解釈が紹介された。もともと文化庁（立法側）では、該当著作物について権利期間の延長を考えていたが、ここでは民法に書かれている保護期間の計算規定を参考に考えるべきであり、文字通りに解釈すると保護期間の延長はなされない（保護期間は終了した）とすべき、と言うことであった。

◎ 中国国内での違法行為の追及

自分の著作物について中国国内にて権利侵害がなされたときにその責任を追求できるかについては、法律上は可能であるものの、差し止め・損害賠償請求は民事訴訟であり、どこかの裁判所で扱うかが問題となる（いわゆる裁判管轄）。つまり、日本の民事訴訟法は、相手が国内にいることを前提とするため、裁判を起こしても相手が出廷しなければ無意味であり、国家権力によって強制することはできない、という趣旨であった。

◎ 外注先の違法行為に対する責任追及の範囲

データベースの構築などを外国企業に外注したときに、その企業が違法行為をもって注文物件を完成、納品し、後にそれが発覚した場合に、発注

した側に責任が及ぶかどうか、については、その行為の違法性を認識しているかどうかにかかっている、ということが説明された。

◎ その他

このほかにも商標と著作権の関係についての話題、在日外国人の扱い、国交のない国の著作物の扱い、DVD 画像の引用の問題のほか、ドイツ著作権における校訂権の扱いなどについての熱い議論が行われた。

◇ おわりに——著作権講座の意義と今後の方向性

恥ずかしながら告白するが、当初「著作権講座」と聞いた時に TV 番組の「法律相談」番組を想像し、一般的な常識と少し乖離している法律上の判断を興味深く取り上げるものと想像していた。しかし実際には、単に事象の○×を判定するものではなく、その根拠となる法律の全体像の説明からはじまり、必ずしも1つの法律だけで解釈するものではない、ということを感じるようになった。

始めは、質問に対する直接の答えとはならない説明に対してすっきりしないものを感じていたことは事実であるが、回を重ねるごとに、問題に対する自分なりの解決の切り口が自然と見えてくるようになった（気がする）のは自分でも驚きであり、聴衆がその段階に達するような内容の非常に濃い講座となっている。

一方、中身の濃さに反し参加者の少なさが唯一残念なことであった。今後、より高いレベルを目指した少人数の講座にするのか、より一般的なテーマの講座に戻るべきなのか。漢情研執行部としては会員諸氏の考えを知りたいところである。

注

- [1] ベルヌ条約は無方式主義（著作行為を行った時点で自動的に権利が発生する）であるのに対し、アメリカ等は方式主義であり、われわれがよく目にする © マークがその権利の明示行為にあたる。

2006年公開講座報告

国際化時代のデータベースとコンプライアンス

東洋学情報化と法律問題——第4回

音楽／楽譜の校訂と著作権法

（前編）

校訂権とその周辺（その二）

石岡 克俊（いしおか かつとし）

◇ 1.はじめに

引き続き「校訂」の法的取扱いについて検討を加えていく。今回は議論の前提として、人文学（特に東洋学）における「校訂」の現代的な意義について触れ、これと文献学研究との深い関わりを示すと同時に、校訂が厳密な学術的研究の上に成り立っていることを確認した。その上で、このような学術研究の成果が、実務上どのように捉えられてきたかをいくつかの実例とともに挙げ、これをわが国の著作権法において取り扱う場合には、どのような枠組みで整理することができるか、その法律上の位置付けを整理した¹⁾。

前号において示唆したように、わが国の著作権法上、校訂は明確な法律上の根拠を有しておらず、このような行為自体の性質が法廷で議論されたこともない。その意味で、校訂の取扱いについての問題は、一定の事実行為の分析を踏まえた法的認識とその法規範上の評価をめぐるあくまで理論上の問題として存在している。

しかしながら、諸外国では、校訂を実定法の枠組みの中で明確に位置づけているドイツの例²⁾や、近時、いくつかの判決においてその保護を明

らかにした英国及びフランスの例がある³⁾。

本稿では、一昨年と去年に相次いで示された英国の判決を素材に、校訂——特に、判決との関係では音楽／楽譜の校訂——の法的取扱いについて検討していくこととする。

この事件は、ルイ14世・15世治世下に活躍した音楽家ミシェル＝リシャール・ド・ラランドの作品で構成される「太陽王に捧げる調べ（Music for the Sun King）」というタイトルのコンパクト・ディスクが、ハイペリオン・レコードにより製作・発売されたことに関連し、同タイトルの版（楽譜）を用意したラランド研究における世界的権威、ライオネル・ソーキンス氏が、同社を相手取り著作権侵害の差止めと損害（賠償）額についての調査（inquiry）を求め提訴した事案である。

いうまでもなく、ラランドが作曲した音楽はすでに公有に帰している。したがって、ここにおける中心的な論点は、既に作曲された楽曲の新しい演奏版（楽譜）が作成され、その楽譜に基づいて演奏・録音された音楽作品につき校訂権（editor copyright）が認められるか否かである。

2004年7月1日、わが国の地方裁判所に当たる大法官庁裁判所 [CHANCERY DIVISION] は、問題とされた4つの曲のうち、一つを除く3つにつ

き著作権侵害を認め、原告の主張を認容した^[4]。その後、敗訴したハイペリオン社は控訴したが、その請求は2005年5月19日に控訴審裁判所民事部[COURT OF APPEAL CIVIL DIVISION]において棄却され敗訴が確定した^[5]。

この結論に対しては、複雑なかつ微妙な反応が見られた。裁判所は、過去の音楽を単一の資料に基づいて演奏することは常に可能なわけではないこと、また、知力がありかつ経験豊富な編集者の校訂が、音楽家によって演奏可能な楽譜を作るためには不可欠であることを認め、結果、音楽作品の版に著作権が認められるという新たな法的先例を築いたことを評価する反面^[6]、ハイペリオン・レコードのようなレコード会社が、重いロイヤリティーに直面し、特定の作品やレパートリーの録音をためらうのではないかという懸念も示されている^[7]。

米国音楽学会のインターネット上の公開掲示板(AMS-L:the public list-serve of the American Musicological Society)における議論は、音楽学者を二分しているという。一方は、校訂者は自ら校訂した作品から適正な承認(acknowledgement) —しばしばそれは経済的報酬であるに違いない—を受け取るべきであるとする考え方であり、いま一つは、その版(楽譜)の著作権からさらに金銭を得ようと期待することは非現実的であり反倫理的であるとする考え方である^[8]。

いずれにせよ、学術的活動の成果の保護と利用は、何も音楽の領域に限ったことではなく、この種の議論はあらゆる分野において妥当する。ここで検討する事件は、偶々裁判所で争われたことから、校訂という学術的活動の詳細が法廷において明らかにされ法的評価の素材とされた。今回と次回においてじっくりと検討を加えていくことにしたい。

まず、手始めに音楽／楽譜の校訂の背景となっている学問的潮流—古楽の復活—について触れておく必要がある。

◇ 2. 古楽の復活

「古楽の復活(early music revival)」は、20世紀音楽学の重要な成果の一つであるといわれている。

古楽(early music)とは、文字どおり「古い音楽」のことであり、西洋においては特に古典派以前の音楽を指している。しかし、音楽学上の知的営為として、これらの復活を取り扱う場合には、それ以上の意義を有することになる。たとえば、ある音楽が、作曲された当時から現在まで、楽器の構造や質に変化がもたらされ、演奏慣習^[9]上の変化が生じ、作曲者以外の者により楽譜に手が加えられるといった事実が見出される場合には、その音楽がどの時代に属するものであろうと復活の対象たる「古楽」として理解される。西洋では、近代に至るまでの間、古楽を恣意的に再構成して演奏したり、主観的な意図の下、楽譜の編纂が行なわれることもしばしばであった。その上、このような経緯を有する楽譜が無批判に受け入れられ、音楽の解釈・演奏が行なわれることも珍しくはなかったという^[10]。

このように、「古楽の復活」とは、音楽伝統の断絶に起因する近代西洋世界の自己発見への道程ともいえ、その意味で、古楽は「再生の伝統が中断された」^[11]音楽であり、「残っている楽器や論文、さらにはほかの資料を基礎にして、歴史的に適切な演奏スタイルが復元されるべき音楽」^[12]である。

今日、古楽がコンサート・プログラムにおいてその大部分を占めているが、こうした状況はほんの数十年前に生じた現象に過ぎない。100年ほど前においては、ある楽曲が作曲され50年も経つと、再びそれを耳にする機会はほとんど無かったという^[13]。このように、現在、われわれが当たり前のように聞いている古楽の演奏は、19世紀初頭に始まり20世紀後半に著しく進展した「古楽の復活」の成果であり、これを支えてきたのが音楽学ないしは音楽学者による基礎研究である。つまり、「古楽の復活」とは、音楽学研究による

2006年公開講座報告

忘れ去られた作曲家と作品の復興、古楽器や演奏慣習（歴史的な演奏様式）の創造的蘇生のこのことの意味している。

「古楽の復活」において音楽学は、主に3つの点で重要な役割を担ってきた。ひとつは、古典として定着した作曲家とその作品に対する音楽的見地からの見直し、いま一つは、忘れ去られた作曲家の再評価とその作品の復元、さらに、演奏実践における歴史的演奏様式の適用、である^[14]。

現在、古楽の演奏においては、厳密な資料検証により成立した批判校訂版／校訂版（Critical Edition）と呼ばれる楽譜が使用されている。ここでいう校訂とは、主観的解釈を極力しりぞけ、諸資料の批判的な照合を通じて作品を客観的・実証的に確定することをいう。

校訂における第一の任務は、可能な限り、作曲家自身が創作した作品を忠実に再現することにある。作曲者の意図を正確に反映するために、さまざまな原資料（典型的には、自筆譜、手稿譜や印刷譜等々）を参照しなければならない。しかし、現代においては多くの原資料が失われており、仮に存在していたとしても、多くの場合、不正確で、断片的で、しかも曖昧である。したがって、校訂を経てそれらの資料を関連づけ再構築した上で、演奏および演奏家の便宜に供される必要がある。また、古い楽譜であればあるほど、記譜法等の違いから演奏家が容易に理解できないこともしばしば起こる。このような場合には、現代的な表記法に書き改めることも校訂の一環である。殊に、バロック期の音楽作品においては、当時の演奏慣習とも関連し、いくつかの装飾が原資料において省略されている。たとえば、ベースライン（bass line）ないし通奏低音（basse continue[仏]、continous bass[英]）と呼ばれる、この時期の作品を特徴付け、かつ、作品の土台ともなる装飾はその典型例であり、これらは演奏に不可欠であるものの、楽譜上の指示がほとんどなされていない。校訂者は、これらの装飾を施し、その他の演奏指示を付加する必要がある^[15]。

このように校訂は、「古楽の復活」にあって、古楽器の復元や演奏慣習の探求とともに重要な位

置を占めており、これらを行なうためには正確な知識と経験に基づく音楽学的知見が不可欠なのである^[16]。

◆ 3. ソーキンス博士 対 ハイペリオン・レコード社事件

◎ 3.1 経緯と概要

本件は、ミシェル＝リシャール・ド・ラランド^[17]の作品で構成される「太陽王に捧げる調べ（Music for the Sun King）」というタイトルのコンパクト・ディスクが、ハイペリオン・レコード（以下、ハイペリオン社）（被告・控訴人）^[18]により製作・発売されたことに関連し、同タイトルの版（楽譜）を用意した原告・ライオネル・ソーキンス博士（以下、ソーキンス氏）（原告・被控訴人）^[19]が、同社を相手取り著作権侵害の差止めと損害（賠償）額についての調査（inquiry）を求め提訴した事案である。

2001年10月、ハイペリオン社は4つのラランド作品（グラン・モテの《Te Deum laudamus（天主にまします御身をわれらたたえ）》及び《Venite, exultemus（われら来たりて、主に喜びの声を上げん）》、グラン・モテの Sacris solemnibus から《the Panis Angelicus（天使の糧）》、そして管弦楽による小品《La grande piece royale（王宮の大なる小品）》の録音を行なった。これらの作品の演奏には、フランスバロック音楽の演奏で特に名高い合唱と管弦楽のアンサンブル「エクス・カテドラ（Ex Cathedra）」が当たり、指揮はジェフリー・スキッドモア氏（以下、スキッドモア氏）が執った。

2001年にソーキンス氏とスキッドモア氏との間で、エクス・カテドラがラランド音楽祭や英国においてラランド作品を演奏することや、ソーキンス氏が用意した楽譜を使って録音することについて話し合いがもたれ、録音は同年10月に行なわれ、その後コンパクト・ディスクとしてハイペリオン社のレコードラベルで発売された。殊に録音の可能性については1999年に話し合いがもたれ、2000年6月まで、ソーキンス氏とスキッド

モア氏との間でそれらの作品のコンパクト・ディスクへの録音に関する話し合いが進められており、その間、ソーキンス氏はスキッドモア氏に *Sacris solemnii* を含むいくつかの作品のサマリーを提供した。最終的にソーキンス氏は録音のために《*Venite, exultemus*》と *Sacris solemnii* の2つについて新しい版（楽譜）を用意し、すでにソーキンス氏によって校訂済みの《*Te Deum*》の版（楽譜）もエクス・カテドラが使うことに同意している。さらに《*La grande piece royale*》については、既存の版（楽譜）にソーキンス氏が不満を持っていたため、更に検証を加えた版（楽譜）を用意することにもなった。ソーキンス氏の示した証拠によれば、2000年の中ごろから作業に着手し録音の間に合わせるかたちで完成させたが、4つの作品の各々につきおよそ300時間をかけたとされる。2001年8月に版（楽譜）が完成し、その写しの大部分はスキッドモア氏に送付された。また、ソーキンス氏は、3つの新しい版（楽譜）を、演奏（実演）権協会（Performing Right Society[PRS]）と機械的複製権保護協会（Mechanical Copyright Protection Society[MCPS]）とに登録した。その版（楽譜）の表紙には、その作品の登録日時・権利者名と併せてその著作権が明示されている。

参考文献

1. 小中村清矩『歌舞音楽略史』（岩波書店、1928年）
2. Haskel, Harry, *EARLY MUSIC REVIVAL*, Thames and Hudson Ltd, London.（邦訳、有村 祐輔監訳『古楽の復活——音楽の「真実の姿」を求めて』（東京書籍、1992年））
3. 大崎滋生『音楽演奏の社会史——よみがえる過去の音楽——』（東京書籍、1993年）
4. W. エマリ著、東川清一訳『エディションと音楽家』（アカデミア・ミュージック、改訂版、1999年）
5. 水谷彰良『消えたオペラ譜——楽譜出版にみるオペラ400年史』（音楽之友社、2003年）
6. 石岡克俊「校訂」の著作権上の位置——校訂権とその周辺（その一）——『漢字文献情報処理研究（漢字文献情報処理研究会、好文出版刊）6号（2005年）

注

- [1] 石岡 克俊「校訂」の著作権上の位置——校訂権とその周辺（その一）——『漢字文献情報処理研究（漢字文献情報処理研究会、好文出版刊）6号（2005年）参照。
- [2] ドイツ著作権法は、「著作権の保護を受けない著作物又は原文を含む版は、それが学術的な分析の成果を表し、かつ、その著作物又は原文の既に知られていた版と本質的に異なるとき」は、著作権法による保護を規定し（同法70条1項）、その権利は「その版の著者」、つまり校訂者により享受される旨が定められている（同法70条2項）。なお、校訂者の権利の存続期間は隣接権としてその版の公表後25年とされている。
- [3] 英国及びフランス両方において提起された訴訟は、いずれも同一人物による音楽ないし楽譜の校訂に関するものであった。フランスでは2005年1月19日にナンテール大審裁判所判決（the judgement of the Nanterre Tribunal de Grande Instance of 19 January 2005 *Sawkins v. Harmonia Mundi and Ors.*）が下され、下級審レベルではあるが演奏版（楽譜）に関する著作権侵害を認めている。また、英国においても地裁・控訴審の双方でフランスと同様の結論を見ている。
- [4] [2004]EWCA 1530(Ch), [2004] 4 All ER 418
- [5] [2005]EWCA CIV 565, [2005] All ER (D) 288 (May
- [6] Tess Knighton, Editorial, *Early Music* November 2004, p.495.
- [7] Tess Knighton, Editorial, *Early Music* November 2004, p.496.
- [8] Tess Knighton, Questions of copyright, *Early Music* August 2005, pp.543-544. [doi:10.1093/em/cah131].
- [9] 音楽学ないし音楽史研究における演奏慣習の位置づけについては、大崎滋生『音楽演奏の社会史——よみがえる過去の音楽——』（東京書籍、1993年）12頁以下を参照。
- [10] この点は、わが国において古来受け継がれている音楽（たとえば雅楽を想起せよ）とは状況が異なっていることに注意されたい（小中村 清矩『歌舞音楽略史』（岩波書店、1928年）参照）。なお、西洋における古楽復活の流れについては、Harry Haskel, *Early Music Revival* (Thames & Hudson Ltd, 1988)（邦訳・有村 祐

2006年公開講座報告

- 輔監訳『古楽の復活——音楽の「真実の姿」を求めて』（東京書籍、1992年）が詳しい。なお、以下本書につき引用は邦訳から行なう。
- [11] Andreas Holschneider, 'Über alte Musik', 34 *Musica(Kassel)*345 (1980).
- [12] Harry Haskel, *Early Music Revival* (Thames & Hudson Ltd, 1988) (邦訳・有村 祐輔監訳『古楽の復活——音楽の「真実の姿」を求めて』（東京書籍、1992年）11頁）。
- [13] Walter Emery, *Edition and Musician* (Novello, 1957) (邦訳・東川清一訳『エディションと音楽家』（アカデミア・ミュージック、改訂版、1999年）1頁。なお、以下本書につき引用は邦訳から行なう。
- [14] 水谷 彰良『消えたオペラ譜——楽譜出版にみるオペラ400年史』（音楽之友社、2003年）212-213頁参照。本書のこの部分は、バロック・オペラの復活について述べたものであるが、この復活は「古楽復活」の歴史においても重要な意味を有している（Harry Haskel, *Early Music Revival* (Thames & Hudson Ltd, 1988) (邦訳・有村 祐輔監訳『古楽の復活——音楽の「真実の姿」を求めて』（東京書籍、1992年）第7章）。
- [15] See [2004]EWCA 1530(Ch), [2004] 4 All ER 418 [para.13-19].
- [16] オペラにおける校訂版の成立過程についての詳細は、水谷 彰良『消えたオペラ譜——楽譜出版にみるオペラ400年史』（音楽之友社、2003年）221頁以下を参照。ロッシーニ全集を例に細かくて厳格なその作業の全体をうかがい知ることが出来る。
- [17] Michel-Richard de Lalande (1657年生 -1726年没) は、ルイ14世が任命した4名のヴェルサイユ宮王室礼拝堂作曲家の一人であり、ルイ14世・15世両治世下における指導的な宮廷作曲家である。作品の多くは、グラン・モテ (grand motet) と呼ばれ、独唱者 (ソリスト) と歌手・合唱団の小さなアンサンブルとが部分 (passages) ごとに区切られたいくつかの楽節 (sections) を構成するものである。これは、17世紀中盤以降に確立したミサの伴奏形式であると同時にフランス王室における宗教音楽の主だった形式であった。比較的最近まで、主に「王の晩餐に捧げるサンフオニ (Syphonies pour les soupers du roi)」と称するおよそ300の短い楽章 (movements) が知られ、これらは王室がヴェルサイユ宮で催す晩餐会で演奏されていた。しかし、現代ではこれらが演奏されることはほとんどなく、1970年代まで現代演奏版と見られるようなモテの総譜が刊行されたことは一度もなかった。なお、この作曲家の名前の表記についてはいくつかあり得るが、近年ではドラランド (Delalande) と綴るのが一般的なようであり (Norbert Dufourcq et al, *Larousse de la musique* (Larousse, 1957))、わが国でもそう表記されている。しかし、ここで取り上げた判決においては、ラランド (Lalande) と記されており、この表記は権威ある音楽辞典 (J. R. Anthony and L. Saukins: 'Lalande [La Land, Delalande], Michel-Richard de', Grove Music Online ed. L. Macy (Accessed [31 January 2006]), (<http://www.grovemusic.com>)). なお、邦訳として『ニューグロヴ世界音楽大事典 (19巻)』（講談社、1995年）283-286頁。ただし、版は異なる。）においても認められる。本稿では一応これに従っておくことにする。
- [18] ハイペリオン社は、忘れ去られた音楽作品のレコーディングを専門的に行なう小さなレコード会社である。
- [19] ソーキンス氏はラランド研究における世界的権威であり、1996年には仏国芸術の振興・普及への貢献によって政府からシュヴァリエ勲章を授与され、1990年と2001年にはヴェルサイユとパリで開催されたラランド音楽祭に芸術顧問 (conseiller artistique) として招聘されたこともある。その際、コンサートでの演奏作品に関するアドヴァイスや、音楽祭の記念冊子・コンサートのプログラムノートなどの執筆にかかわっている。その他、仏国のバロック音楽やラランドに関する数多くの研究業績を有している。

中国の知的財産権関連の書籍と Webサイトの紹介

山田 崇仁（やまだ たかひと）

◇ はじめに

ここでは、中国の知的財産権を知る上で役立つと思われる書籍やWebサイトを、簡単なコメント付きで紹介する。

また、本誌第四号には小島浩之氏による「著作権についての知識を深めよう 東洋学のための著作権サイト・ページ指南」が掲載されている。こちらも併せて参考にさせていただきたい。

◇ 書籍

中国の著作権を初めとする知的財産権関連の情報は、日本語で読めるものも増えてきた。中には、中国の国制や法律・裁判制度に関する簡単な説明を記している書籍もある。その部分はとりあえず一読しておくことをお奨めする。

中国のWTO加盟に伴う関係各法令の改定などがあるため、なるべく新しい情報を含んでいるものを参考にすべきである。

◎ 著作権法

●『著作権法逐条講義』／加戸守行著／2006 五訂新版

日本の著作権法に携わる者にとっては、バイブルの書物。頻用者は必ず新版をそろえておくべし。

●シリーズ 『中国著作権「法律と運用」』

ビクターエンターテインメント社より三冊シリーズとして刊行された中国の著作権関連書籍もの。刊行年代の関係上、若干内容が古くなっている部分があると思われるので、後述の書籍やWebサイトも併せて参考にされたし。

- 1.『实用著作権知識問答』劉波林・許超・孫建紅著／近藤丸人監修・訳／1997
- 2.『著作権案例百析』馬暎剛・高華苓編著／近藤丸人監修・訳／1997
- 3.『中国著作権法「理論と運用」&著作権法及び関連法規』近藤丸人著／1997

◎ 知的財産法関連

書籍は、著作権そのものを対象としたものより、それを含めた知的財産権法関連のものが入手しやすい。ここでは発行年代が新しめのものをいくつかあげておく。

●『中国知的財産権関係法令集：対訳（日本語中国語併記）』／1巻：実務法令篇／2巻：関係法令篇／中国国際貿易促進委員会特許商標事務所翻訳監修／東洋法規出版／2006

外国の法律である以上、当然外国語で法律の条文も書かれている。そのため、通常の法に関する理解とは別に、外国語に対する理解も欠かせない。

本書は、日本語を母語とする利用者向けに、中国語との対訳方式を採用しているのが大変便利。但し、翻訳である以上、日本語はあくまで参考に

2006年公開講座報告

しかすぎないので注意されたし。

- 『中日・日中法律用語辞典』畑中和夫：他編／永井美佐子：他著（国際比較法シリーズ 別冊）／晃洋書房／1997

専門的な用語を理解するためには、専門的な辞書の助けを借りる必要がある。本書は、法律用語に関する日中両引きの辞典である。

- 『必携法令難語辞典』第3版／浅野一郎・田島信威・岩崎隆二編／三省堂／2003

法律の条文の中には、一読するだけでは理解することが難しい用語が存在する。六法全書の末尾などには、難語の解説が掲載されているものがあるが、より詳しい意味・用法を知るためには、専門的な辞典を参考にする方がよい。本書は、そのような目的のための辞典である。

ちなみに、法律用語の中には、律令以来の古い語法を残したものが見られる。それらの中には唐の律令用語に基づくものもあり、古典漢文を読む際、助字の使い方などの面において参考になる場合がある。

- 『中国の知的財産法—中国ビジネスに欠かせない法制度の仕組みと実際—』創英知的財産研究所編著／東洋経済新報社／2006

中国を対象としたビジネスをする人向けに書かれた本。中国の国家制度や司法制度の概説から始め、知的財産権に関わる法律の概説、各法律の詳細を内容とする。比較的平易に書かれており、参考書として便利な本だろう。

- 『中国知的財産法』遠藤誠／商務法事／2006

ビジネスの最前線で知財関連の問題に直面している、あるいは企業の対中知財関連の業務に携わっている人向けに書かれている。こちらはやや専門・実務的に書かれている。

- 『中国の知的財産権裁判と重要判決—実際の事件と判決および裁判過程の再現—』／周林編／

劉新宇訳／経済産業調査会／2004

中国の弁護士や裁判官による、知的財産権に関わる裁判に関する書物の日本語訳。実際の裁判手続きや裁判の実施経過の紹介や、代表的な知的財産権侵害に関わる判決とその評釈が述べられている。著作権に関わる裁判についてもいくつか紹介されており、参考になる。

- 『海外における著作権侵害の現状と課題に関する調査研究—中国調査編：Sarah 共通目的事業（平成14年度）—』著作権情報センター編／著作権研究所研究叢書：no.10／2001～

ゲーム・CD・ビデオソフトなどを対象として、中国大陆における海賊版の氾濫状況を現地調査したものの報告。今日ではインターネット上における著作権侵害（WebやP2Pなどによる流通）行為も問題となってきているであろう。

- 『知的財産権侵害物品取締ガイドブック—台湾編—』日本関税協会知的財産情報センター／1992

台湾を対象とした知的財産権問題についてのガイドブック。発行年が古いですが、参考にはなる。

◎西村幸次郎氏の著作

現代中国法の専門家として著名な西村幸次郎氏の一連の著作は、知的財産権関連法以外の現代中国法そのものを知るために参考にすべきだろう。

- 『中国における法の継承性論争』編訳／早稲田大学比較法研究所／1983

- 『中国憲法概論』監訳／成文堂／1984

- 『中国における企業の国有化：政策と法』／成文堂／1984

- 『中国憲法の基本問題』／成文堂／1989

- 『中国の家族法』共訳／敬文堂／1991

- 『現代中国憲法論』共編著／法律文化社／1994

- 『現代中国の法と社会』法律文化社／1995

- 『中国民族法概論』監訳／成文堂／1998

- 『グローバル化のなかの現代中国法』編著／成文堂／2004 補正版

中国の知財関連の書籍と Web サイトの紹介（山田）

- 『現代中国法講義』編著／法律文化社／2005
第2版

◇ Web サイト

- 社団法人著作権情報センター
<http://www.cric.or.jp/index.html>
HOME > 外国著作権法のページより中華人民共和国・台湾の著作権法の日本語訳が閲覧できる。
- 知的財産権関連ホームページ集
<http://www.jcci.or.jp/chizai/kanren.htm>
著作権情報センターの知的財産権に関する Web サイトのリンク集。
- Jetro 北京センター知的財産権室
<http://www.jetro-pkip.org/falv.htm>
日本貿易振興機構（JETRO）北京センター知的財産権室の Web サイト。中国での知的財産権に関わる情報を知りたいなら、まずはここを参照。
- 文化庁 著作権 あらたな文化のパスワード 海賊版対策について
<http://www.bunka.go.jp/1tyosaku/kaizokuban.html>
日本の海賊版対策に関するページ。

- TRIPS 協定 知的所有権の貿易関連の側面に関する協定
http://www.jpo.go.jp/shiryou/s_sonota/fips/trips/ta/mokuji.htm

特許庁の下位ページ。本協定は、今後の国際的著作権トラブル解決の指標となるものであり、中国も WTO に加盟している以上、TRIPS 協定を遵守する義務があるため、今後（建前上でも）大きな影響が出てくるとされる^[1]。

◇ おわりに

以上、簡単ではあるが、中国の知的財産権を知る上でナビゲートとなるような書籍や Web サイトを紹介した。本稿の情報が何かしらみなさまのお役に立つことができれば幸いである。

注

- [1] 中国法の世界では、伝統的に「法の条文はあれども、実際の運用は時々々の政策や有力者の意向や当事者間の人間関係で決定される。」という認識があるように思える。これはある程度現在にも当てはまるのではないかと個人的には思っているが、TRIPS 協定という国際的な建前に対し、中国的法律運用の実態はどう折り合いをつけてゆくのか、今後注目したい。

Jetro 北京センター知的財産権室

Jetro-Pkip 日本貿易振興機構北京センター 知的財産権部	
ホームページ > 中国知財関連法 (2006年08月16日更新)	
特許 「中華人民共和国特許法」第三回改正意見書 集積対照案文 NEW 中華人民共和国特許法 中華人民共和国特許法実施細則 (JPO日本語訳) 最高人民法院特許紛争案件審理の法律適用問題に関する若干規定 国防專利条例 新しく改正された專利法実施細則についての説明 特許権の強制実施権に関する弁法 一覽へ…	著作権、模倣取締関係 情報ネットワーク伝達権保護条例 NEW 著作権行政苦情申立ガイダンス NEW 集積回路配置回路設計保護条例 中華人民共和国著作権法 中華人民共和国著作権法实施条例 コンピュータソフトウェア保護条例 インターネット著作権行政保護弁法 一覽へ…
商標 商標法修正案 (新旧対照表) (PDF) NEW 中華人民共和国商標法 中華人民共和国商標法实施条例 中華人民共和国商標法实施条例 商標審査基準 (2005年12月公表) 商標審判規則 著名商標の認定と保護に関する規定 一覽へ…	その他 中華人民共和国独占禁止法 (草案) NEW 知的財産権税関保護における総担保の受付について NEW 公安部 税関総署の知的財産権における法執行の協力強化に関する暫定規定 不正競争民事案件の審理における法律適用の若干の問題に関する解釈 (意見募集稿) 権利衝突に係る知的財産権紛争案件の審理の若干の問題に関する指導的意見 (試行) 中華人民共和国知的財産権海關保護条例 中華人民共和国製品品質法 一覽へ…
IPG活動 IP News IP判例集 ニセモノ写真集 知的財産権法 政府機関 代理組織 統計	
北京市朝陽区建外大街甲26号長富宮公樓7003 post@jetro-pkip.org	

コラム

わたしたちは 子供の盗みを容認している

山崎 直樹（やまざき なおき）

刺激の強いタイトルで恐縮だが、毎年、秋風の到来とともに、こう痛感する。

わたしの在住する市の施設では、毎年、『夏休みコドモ科学展』なるものが開かれている。内容は、市内の小学生の夏休みの宿題——自由研究とか工作とか——の展示である。ここに豚児を連れていくのが、わたしの毎年の慣わしとなっている。

会場内で、同じテーマの自由研究があちこちで目につくのは、ご愛嬌である。『これで完璧！夏休みの自由研究』という類のネタ本がたくさん出版されているせいだと思う。また、工作でも、××ハンズで売っている工作キットを組み立てただけのものもある。それもご愛嬌だと思う。これらはまあいい。

わたしが「盗み」だと痛感しているのは、以下のようなことがらである。

自由研究とやらにおいて、子供たちはいろいろなことを調べて、それを冊子にまとめたり、ポスターにまとめたりしているのだが、文面を読むと、明らかに図鑑や百科事典を引き写しただけのものもある。写真を載せているものも多いが、明らかに、図鑑や百科事典の写真をカラーコピーして切り貼したものがある。加えて、最近では、どこかのウェブサイトの文章と画像をコピー&ペーストしただけ、というお手軽なものも目につく。

なぜ、「ウェブサイト」と限定できるのかというと、今年目にした例では、貼り付けてあった画像（カブトムシの写真）をよく見ると、隅に、「購入はこちらから→」というナビゲーションリンクの文字がそのまま残っていたからである。この切り貼りを子供が自分でやったのか、大人に手伝ってもらったかわ

からないが、自分でできたのだとしたら、小学校における情報技術活用授業の成果だろう。

さて、予想はつくと思うが、これらには出処などに関する断り書きはいっさいない（断ればよいというものでもないが）。他人の著作を切り貼りしてできた作品に自分の名前だけを入れて「自由研究」と称し公表するのは、率直に言って、ドロボーの所業であると思う。そして、それに対し何の指導もしない教師や他の大人は、その所業を容認しているのだと言ってさしつかえない。小説の剽窃や論文の盗用をする大人は後を絶たないが、その習慣は小学校で涵養されていたのか……

最近、某雑誌のコラムで取りあげられていて、その存在を知ったのだが、社団法人日本教育工学振興会が、文科省の委託をうけて、『学校におけるIT活用等の推進に係る調査研究報告書』という報告書を出している（平成15年、<http://www.japet.jp/skillchk/>）。簡単に言うと、これは、小学校から高校までの教員が、コンピュータを用いて教科指導をするために身につけるべき知識とスキルがチェックリスト形式で列挙されている報告書である。この中の「小学校における各教科共通の知識とスキル」のリストに、〈知識〉として「インターネットから収集した情報について、著作権に配慮すべきであることを説明できる」が挙げられ、〈スキル〉として「児童がインターネットから収集した情報やデータについて、著作権に関する判断や指導が行える」という項目がある。気の毒に、この報告書は、数年経っても現場の教育に生かされていないわけだ。

特集 1

極める！ デジタルテキスト

伝統的文献学の研究方法の中で、「検索」という行為が重要な要素を構成してきたことは周知の通りである。研究を進める上で、いかに効率よく「(用例の) 検索」を行うかが重要視されてきたのである。

しかし、コンピュータの発達と一般化は、それまで膨大な手間と時間を費やしていた「検索」作業を一変させた。

但し、コンピュータ上での「検索」を初めとするデジタルテキストの取扱いにはちょっとしたコツがある。それを覚えることで、より効率よく研究が進むことになるだろう。

ここでは、「デジタル化されたテキストをどう効率よく利用するか」について、テキストの入手・加工・検索などを中心に説明した。

本特集は、実際に試してみながらお読みいただくことをおすすめする。その方が便利さを実感していただけるだろう。その上で、読者諸賢にその内容が少しでもお役に立てたのなら、幸いである。

Contents

私的検索のススメ	師 茂樹	32
オンラインテキストアーカイブ概説	千田 大介	34
データの入手と加工	二階堂善弘	42
データの整形と整理	小島 浩之	46
検索・置換・GREP	山田 崇仁	52
お勧めユーティリティ テキスト検索、一括処理ツール	小川 利康	60

私的検索のススメ

師 茂樹（もろ しげき）

◇ 特集にあたって

本特集では、デジタルテキスト（電子テキスト）の入手から検索や加工までの一連の流れを、初心者にもわかりやすいように配慮しながら解説している。このような特集を組んだ背景は、大きく分けて二つある。

一つは、時代の流れに追いつくためである。本研究会では、これまでも『電脳中国学』（1999年）、『電脳国文学』（2000年）、『電脳中国学Ⅱ』（2002年）を出版し、デジタルテキストの重要性と積極的な活用を訴えるとともに、より多くの人が使えようになるためのマニュアルを提供してきた。

これらの書籍で訴えてきた内容については現在もその価値を失っていないと思うが、一方でマニュアル的な部分については、日進月歩のコンピュータ業界の中ではいささか時代遅れの感も否めない（研究会の方には時折、『電脳中国学Ⅲ』はまだ出版されないのか、という問い合わせがある）。

また、ここ数年で、オンラインでの入手や検索などの利用が可能なデジタルテキストが増加する一方、以前は考えられなかったような巨大な文字数を誇るテキストデータベースも発売されるようになってきた。これらの情報についても、本誌『漢字文献情報処理研究』等を通じて逐次紹介してきたが、誌面の都合もあり詳しい使い方まで解説することはできていない。

本特集は、本研究会におけるこれまでの蓄積をまとめつつ、最新の環境にあわせたマニュアルを提供することを目的としている。誌面の都合で書

ききれなかったところも多々あるが、それについては今後、継続して提供できればと考えている。

一方、多言語環境やデジタルテキストが一昔前と比べると非常に充実してきた昨今、検索などのサービスの裏で動いているシステムは大規模化、複雑化しているはずであるが、利用者にとっては便利で使いやすいものが提供されるのが当たり前になっているためか、デジタルテキストは最近どんどんブラックボックス化しているように見受けられる。

確かに、便利なことは悪いことではないが、研究目的で使うデータがどのような技術や思想に支えられているのかをまったく理解しないというのは、研究者の態度としては危ういのではないだろうか^[1]。本特集は、このようなブラックボックス化の流れに、ささやかながらも警鐘を鳴らしたいという意図がある。

◇ なぜ「私的」なのか

さて、本稿では「私的検索のススメ」というタイトルを挙げたが、これはデジタルデータを独占して抱え込むという意味ではない。千田大介氏の記事で述べられているように、データを共有し豊かにしていくことこそが「デジタル化時代における人文科学としての中国学の発展に不可欠」であることは、いくら強調してもしすぎるということはない。」であろう。

ここで言う「私的」とはむしろ逆で、あらゆるデータを個人が自由に入手でき、検索したり加工したりすることができる（そして誰かが加工したデータを、さらに別の人が自由に入手できる）と

いう環境を理想とし、それを目指すという意味での「私的検索のススメ」である。

一般に、一部の人がデジタルテキストに対する権利を独占し、利用者に制限を加えるような場合には、お定まりの検索システムしか使えない場合がほとんどだ。

自分の目的にあわせてテキストを分類したり（小島浩之氏の記事参照）、正規表現（山田崇仁氏の記事参照）を使った曖昧検索をすることによって、検索の精度をあげたり効率化できたりするが、そのためには自由なテキストが必要である。また、検索システムを自分の目的にあわせてちょっとだけ改造するとか、文字化けを避けるためにテキストの文字を一部置換する（二階堂善弘氏の記事参照）とかいう利用をしたい場合も同様だ。

もちろん、何でもかんでも自由に利用できればよいというわけではないが、研究内容によってはお仕着せの単純な検索はまったく役に立たない場合もある。「私的」に改造できるデジタルテキスト環境というものを構築し、共有していく重要性をまず強調しておきたい

加えて、ブロードバンドが普及し、Googleをはじめとする多くの便利なサービスがインターネットの「あちら側」（自分のパソコン＝「こちら側」との対）に移行しつつある^[2]。この流れには歓迎すべき点もあるのだが、一方で小川利康氏の記事でも指摘されているように、プライバシーの流出の問題が懸念される面もある。便利な機能があるというだけで「あちら側」に依存してしまうのではなく、ある程度「こちら側」でまかなえるテクニックは身につけておいた方がよいだろう。

◆ 最後に

デジタルテキストをはじめとするデータベースの利用は、今後ますます日常化していくものと思われる。一方で、大学や大学院では工具書の扱い方について教えるのに、データベースの使い方についてはあまりつつこんで教育されることはない。現在を過渡期と言うこともできるだろうが、ごく基本的な知識ですら普及していないことに懸念を感じるのは筆者だけではないと思いたい。人文科学におけるコンピュータ教育については多くの問題があるが^[3]、それなりに使えるから教育しなくてもよいということにはならないはずだ。本特集が普及の第一歩になれば幸いである。

注

- [1] 近年、科学研究費などのネタとしてデータベースを作ることが多いようであるが、かつては「データベースを作るのってお金がかかるんでしょ？」と聞かれていたのに、最近は「データベースを作るのにそんなにお金はかからないでしょ？」と聞かれることが多くなってきたように思う。一時のIT革命＝経営合理化という幻想がしっかりと定着し、コスト感覚を欠いた発言が横行するようになってきている。企業に発注すれば数億円かかるようなプロジェクトであっても、一部の「できる」教員、院生などのボランティアをあてこんで数百万円でやっしまおうという例も耳にすることがある（これは、コンピュータ好きの人の口癖である「できますよ」が悪いという説もある）。
- [2] 本誌のレビュー「Google」「図書館とOPAC（Googleと学術情報）」参照。
- [3] 本誌の特集2「人文学的な情報処理教育とは」を参照。

オンラインテキストアーカイブ概説

千田 大介（ちだ だいすけ）

◇ はじめに

オンラインテキストアーカイブ＝オンライン（ネットワーク上の）＋テキスト（文献）＋アーカイブ（倉庫）、つまりインターネット上に作られた文献データ置き場、ということである。インターネットには、文献のテキストデータを集積・公開しているサイトが数多くある。それらのアーカイブは、学術用であったり一般の愛好者向けであったりさまざまだが、研究者にとってはテキストデータを無償で入手し検索や分析に利用できるという、大きなメリットがある。

ここでは、まずオンラインテキストアーカイブの学術的な意義と、テキストデータの電子版本鑑定について概説した上で、中国古典・仏典・中国近現代文献の代表的なオンラインテキストアーカイブとその使い方について紹介していきたい。

◎ データを共有する意味

オンラインテキストアーカイブには、インターネットにつながっていれば、だれもがアクセスできるという特色がある。そして、それらのデータはみな誰かが入力して無償で公開してくれたものである。

自分でせっかく作ったデータなのだから、他人には提供せずに自分だけのものにしておけば、研究する上で人と差を付けられて有利ではないか、と考える人がいるかもしれない。そのような立場からすれば、テキストデータを無償で公開することは愚劣な行為であるかのように思えるが、それ

は誤りである。

およそ研究というものは、他者による検証が可能でなくてはならない。資料を分析して一定の理論や法則を導き出す、その過程が妥当であるかどうかを問うのが論文なのであり、自分だけが囲い込んだ資料やデータに基づいて論文を書いたとしても、資料やデータには他人がアクセスできない以上、論文が妥当であるかどうかの検証はできないことになる。

現実には、囲い込んだデータや資料に基づいて書かれた論文は数多い。しかし、そのような態度がまかり通っていること、それ自体が他の学術分野から、人文科学ではなく人文学であると時に貶められてしまう原因なのではないだろうか。そう考えれば、オンラインテキストデータベースを使いこなすとともに豊かにしていくことは、デジタル化時代における人文科学としての中国学の発展に不可欠であることが理解されよう。

◎ 良いテキストと悪いテキスト

オンラインテキストアーカイブを利用する上では、テキストデータの質に注意する必要がある。中国学を学んだ人ならば、資料の扱いや文献学についてさまざまな訓練を積んでいて、文献資料の扱いやテキストの選定には常に細心の注意が必要であることが身に染みていることだろう。

電子テキストも研究に利用する文献資料の一種であるのだから、当然、その扱いには細心の注意が必要である。電子テキスト版本の鑑定に際しては、以下のような点に留意する必要がある。

●電子テキストの底本は何か

オンラインテキストアーカイブに収録されているテキストデータは、オンラインオリジナル小説や新聞記事・評論などを除き、大半が紙テキストを入力してデジタル化したものである。もしも悪い版本に基づいて入力されたものであれば、その電子テキストは勿論、良い電子テキストとは言えない。だから、素性のよい版本に基づいて入力されていることは、良い電子テキストの第一条件である。底本が明示されていない場合は、自分で校訂した上で利用するなどの注意が欠かせない。

●電子テキストを入力・校訂しているのは誰か

活字本の場合、信頼の置ける出版社の本であるか、また信頼のおける校訂者の手になるものであるかがテキストの善し悪しを判断する材料になる。これは、電子テキストの場合でも変わらない。

まずアーカイブが学術団体や研究者によって運営されているのか、それとも愛好者の個人サイトなのか重要な判断材料になる。

同時に、テキストデータを入力あるいは校訂・維持しているのがいかなる組織・人物であるのかにも留意しなくてはならない。活字本から電子テキストを作るのであれば、誰がやっても同じなのではないか、と思われるかもしれないが、特に古典文献の場合には、現代語では使われない漢字や語句が多く使われているので、たとえ中国人が入力していても、専門教育を受けていないとどうしても初歩的なミスが生じてしまうものだ。

●フォーマットと外字処理

以上の2点は、紙の版本鑑定の上延線上にあると言えるが、電子テキストの場合は、デジタルメディアゆえの注意点もある。

まず、文字コードと外字処理が問題になる。文字コードは、より多くの漢字が定義された国際規格に基づいている、つまり、最新の Unicode 規格に基づいているのが理想だ。古典文献の場合は、繁体字であることも必須である。

また外字については、それがいかなる字形であるのかがわかるような形式で処理されているか

が大切である。オンラインアーカイブのデータは、大半が GBK・Big5 などのローカルコードに基づいて作られており、外字は独自の Windows 外字ファイル、[口/追] (= 喰) のような合成文字、諸橋『大漢和』の漢字番号などの方法で処理されている。それらの外字には最新の Unicode 標準に収録されている文字が多いので、規則的な処理がされていれば、一括置換して手軽に検索・分析できるようにすることができるのだ。

データのフォーマットには、HTML・テキストデータ、あるいは Word 文書・PDF などさまざまである。HTML・Word 文書・PDF などは、ドキュメントをイメージ通りに再現できるメリットはあるものの、この特集で紹介するテキストの検索・分析には、テキストデータが断然便利である。もっとも、他の形式からテキスト形式に変換するのは、他の項目で解説するように、さほど難しいことではない。

以下で紹介する代表的なオンラインテキストアーカイブについては、簡単な評価を付けてあるが、それ以外のテキストアーカイブを利用する際にも、上記の点について注意していただきたい。

◇ 主要テキストアーカイブ紹介

◎ 中華文化網

<http://ef.cdpa.nsysu.edu.tw/ccw/>

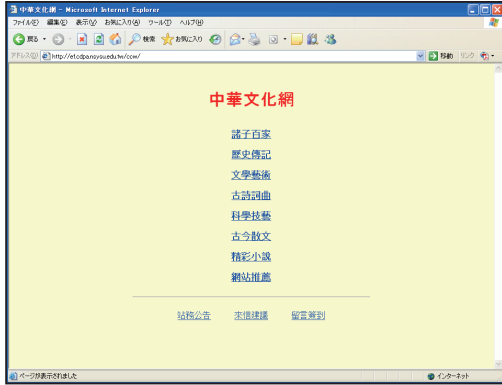
古典文献オンラインテキストアーカイブの定番。台湾中央研究院の漢籍電子文献からコピーしてまとめたデータが中核を占めている。それは違法行為ではないか、と思われるかもしれないが、いずれも著作権の消滅している古典文献であり、法律上は問題ない（学界内部の慣習・モラルの観点からは、問題なしとは言いがたい）。

ただし、「古詩詞曲」・「古今散文」の個人アンソロジーや作品はアメリカ系のオンラインデータで、信頼性がより劣る。

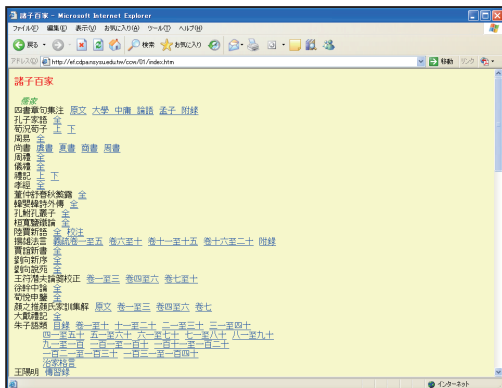
●データ閲覧手順

- ・トップページから分類項目のリンクをたど

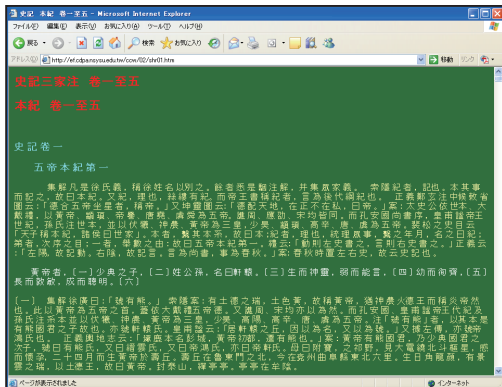
デジタルテキスト



中華文化網 トップページ



中華文化網 アーカイブ



中華文化網 データ表示

- 歴史書(二十五史・『通典』・『国語』・『戦国策』など)
- 古典籍(諸子百家・魏晋小説・『文選』・『山海経』など)

●データ形式

- 文字コード: Big5
- 外字処理: 台湾中央研究院外字を Big5 収録文字に書き換え・「●」
- データ形式: HTML (注釈は色分け表示)

●長所と短所

- 中央研究院データが元になっているので、比較的信頼性が高い。
- 南北朝以前の主要古典文献を網羅する。
- ×Big5 コードで作成されていて、中央研究院外字が一般の文字に置き換えられているため、ユーザーの校正作業が不可欠
- ×二十五史の一部など、不完全なデータが見られる。
- ×文字の間に半角スペースが入っており、また本文と注釈が HTML で色分けされているため、整形作業が不可欠。

なお、中華文化網は、ここ十年ほど URL が頻繁に変わっている。上記 URL も何時変わってしまうかわからないので、接続できない場合は繁体字中国語の Google で検索していただきたい。

◎ CBETA 中華電子佛典協會

<http://www.cbeta.org/index.htm>

国際ネットワークであらゆる漢訳仏典をデジタルテキスト化し、無償で公開している。テキストデータのみならず、仏典の検索システム・専用ブラウザ・CD-ROM イメージファイルのダウンロードなど、仏典関連のデータやツールが多数提供されている。

●データ閲覧手順

- 必要な仏典の分類がわかっている場合:
 - トップページ中央の分類検索

る。

●主要データ

南北朝以前の古典が中心。

- 儒教関連(十三経・『朱子語類』・『伝習録』など)

オンラインテキストアーカイブ解説（千田）



CBETA 仏典テキストの表示



CBETA トップページ

分類がわからない場合：

- トップページ右側の検索ボックス「經名查詢」をチェックして、検索条件を指定。台湾のサイトなので、繁体字で入力。
- リストが表示されたら必要な仏典の「經名」をクリック。
- 「目録」以下のリンクをクリック。



CBETA 検索結果

●主要データ

- 仏典（『大正新脩大藏經』・『卍統藏經』）

●データ形式

- 文字コード：Big5
- 外字処理：合成文字（[] 内に複数漢字の組み合わせで字形を説明）
- データ形式：テキスト（app 形式¹¹⁾）

●長所と短所

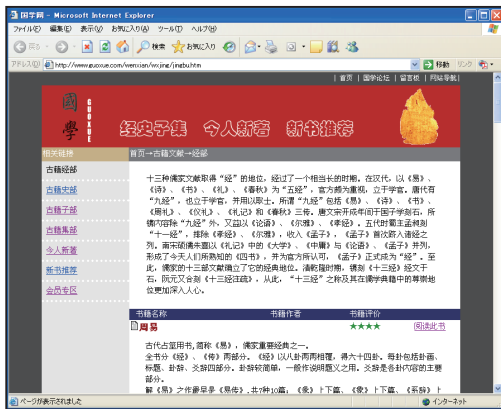
- データの信頼性が高く、収録も網羅的。
- データの形式が検索などに向いている。
- ×Big5 コードであるため、外字部分を Unicode の文字に置き換えるなどの加工が必要になる。



CBETA 閲覧メニュー



国学トップページ



経部のテキスト一覧



文献の表示

◎ 国学

<http://www.guoxue.com/>

北京の首都師範大学に本拠を置く国学ネット

は、近年、安価かつ大容量な文献 CD-ROM 製品を多く発売していることでも知られている。その Web サイトでは多くの文献テキストデータが無償で公開されており、中国古典文献テキストアーカイブとしても屈指の規模を誇っている。

● 閲覧手順

- 画面左の「経・史・子・集」から必要な文献の分類をクリック。
- 以下、分類をたどる。

文献の検索

- トップページ右の「搜国学」で、キーワードを入力して検索。

- 辞書項目の検索結果が表示される場合は、「国学书目词典—●●在内容中查」のリンクをクリック。

ただし、検索した場合は、当該テキストのキーワードの出現する巻・章などしか表示されない。

● 主要データ

古典文献（宋代以前が中心）

- 歴史書（『資治通鑑』・『続資治通鑑』など）
- 文学総集・類書（『全唐詩』・『全唐文』・『太平広記』・『太平御覧』など）

● データ形式

- 文字コード：GB コード（簡体字）
- 外字処理：「■」埋め込み・画像埋め込みなど
- データ形式：HTML（一部、空白文字埋め込みデータあり）

● 長所と短所

- データ量が豊富で、ここでしか手に入らない大部な書籍のテキストデータが多数ある。
- × 全て簡体字である。
- × 底本が明示されていない。
- × 一部文献に白い隠し文字が仕込まれている。

オンラインテキストアーカイブ解説（千田）

国学は他では入手しがたいデータや学術情報を多数公開している得難いサイトではあるが、コンテンツの構成が錯綜しており、またデータがいつでも簡体字であるというのは、中国学研究の立場からすると残念である。

もっとも、国学は研究サイトというよりも、むしろ教養サイトと言うべきものであり、その教養とは、現代中国の入試体制と文化保守主義に裏打ちされた、簡体字による大衆的中国古典教養なのであって、日本人が思い浮かべる伝統的な中国古典学教養とは異なる。つまり、簡体字であることこそが国学のアイデンティティなのである。

なお、中国には数多くのテキストアーカイブがあり、その中には簡体字と繁体字でデータを提供しているところもある。しかし、その繁体字データは簡体字データを機械的に変換しただけのものが大半で、残念ながらあまり役に立たない。

◎ 黄金書屋

<http://www.goldnets.com/>

1990年代後半に開設された、中国語オンラインテキストアーカイブの草分けである。現在は、中国のポータルサイト、マイライスの1コンテンツになっている。

● 閲覧手順

- トップページから「黄金书库」をクリック。
- 左側の一覧から必要な文献のジャンルを選び、分類をたどる。
- 必要な文献の画面が開いたら、右手の「全文阅读」をクリック。
- 表示されたファイルのリストをクリック。
- 画面下の「下頁」をクリックして次ページに移動。

● 主要データ

- 古典文学・近現代文学

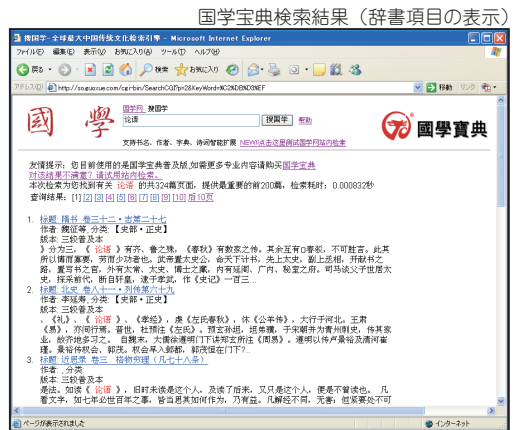
● データ形式

- 文字コード：GBコード（簡体字）
- データ形式：HTML

● 長所と短所

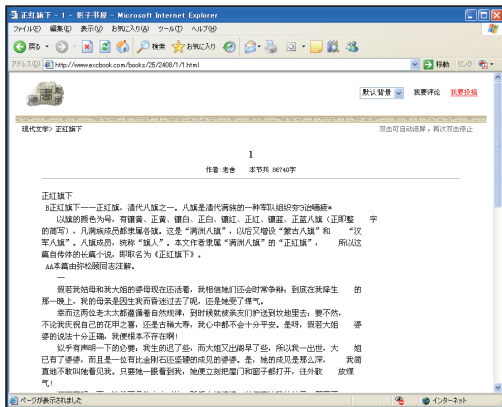
○ 近現代文学データが充実。

× 著作権法的に問題あり。





分類カテゴリの表示



文献データの表示



漢籍電子文獻トップページ

- ×データが複数ページに分かれており、コピーして一つにまとめ整形しないと使えない。
- ×底本が明示されない。

黄金書屋の近現代文学データには、著作権に関する表示のない、法的に問題があると思われるものが多数見受けられる。それでも個人がダウンロードして検索や分析に利用するだけならば、法的に許容範囲内であろう。しかし、それらは底本が不明である上に校訂も不完全であるから、學術利用に際してはしっかりした紙本に基づく校訂が必要となる。ところが、そうして校訂したデータを再配布することは、著作権法違反になってしまう。

このため、ネット上の近現代テキストデータを使った研究は必然的に、データの精度の低さに目をつぶった質の劣る研究、もしくは独自校訂の囲い込みデータによる実証不能な非科学的研究に陥ってしまう危険をはらんでいる。

この問題は、おそらく中国の作家や出版社が公式テキストアーカイブやCD-ROMを提供してくれない限り、解決困難であろう。

このほか、現代中国語研究に用いるデータとしては、インターネットに公開されている各種新聞・雑誌記事も有用である。

◇ 文献検索サイト

オンラインテキストアーカイブとは異なるが、インターネット上の文献検索サービスからデータをコピー・蓄積して検索や分析に用いることもできる。最後に、そのようなWebサイトを、ごく簡単に紹介しておこう。

◎ 漢籍電子文獻 (2.0 版)

<http://www.sinica.edu.tw/~tdbproj/handy1/>

台湾中央研究院が提供する、規模・信頼性共に最高レベルの古典文献データベース。儒教經典・二十五史のほか、古典詩歌・戯曲小説・正統道蔵・近現代史など幅広い分野をカバーしている。

文字コードはBig5コードで、独自の外字を使用している（現在、Unicode対応作業が進められている）。底本はトップページから「相關訊息」をクリックし、表示されるページからリンクをた

どると参照できる。大半が、大陸で刊行された定評のある排印本の台湾リプリント本である。

◎ 全唐詩電子検索系統

<http://162.105.161.41/tang/index.htm>

北京大学中文系が提供する、『全唐詩』および関連詩話・伝記資料などの総合データベース。表示が一首ごとであるため、大量の作品を分析するには向かないが、ある作者の作品が欲しい場合などには有効であろう。

なお、北京大学中文系では、『全唐詩』『全宋詩』のオンライン分析システムも提供しているが、こちらは全作品を収録しているわけではない。

注

[1] 「app形式」とは、ウルス・アップ氏が提唱したテキストデータのフォーマットで、以下のように各行頭にテキスト情報を埋め込んだ形式。

T08n0251_p0848c08(18) | 觀自在菩薩行深般若波羅蜜多時，照見五蘊皆空，度一切苦厄。

それぞれの数字の意味は、以下のとおり。

T08	n0251	_p0848a	01	(00)
『大正蔵』の巻数	『大正蔵』の番号	ページ番号	行番号	前の行からこの行に移された文字数

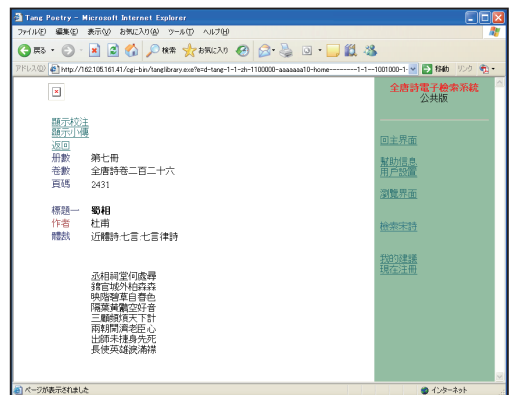
センテンスの途中で行が変わると、語句が断ち切られてしまい単語検索できなくなってしまうので、各行を必ずセンテンス単位で終わらせ、行頭に移動させた文字数を記入することで対処している。



文献データの表示



全唐詩電子検索系統トップページ



詩歌の表示

データの入手と加工

二階堂 善弘 (にかいどう よしひろ)

◇ はじめに

ここではこれまで紹介したサイトなどから、データを入手して加工するまでを扱う。早い話が「コピー」アンド「ペースト」をする作業に当たる。

それではデータを開いてコピーして貼り付ければそれで終わり、と思うかもしれない。しかし事態はそう単純でもない。

よく学生のレポートなどでも、どこかのサイトのデータをそのままワープロに貼り付けてあるのを見かける。しかし書式がメチャメチャであったり、漢字コードに無頓着でやたらと文字化けが目立ったりするものがある。こういったおサルさんレベルのやり方は、内容以前に、形式でまず減点といってよいだろう。

またテキストデータとして保存しても、検索する時に漢字コードが分かっていると、上手く検索できなかつたりする。こういったことに注意しながら、データの扱いについて簡単に紹介する。ただ、より詳細なデータの加工については、次章以降を参照していただきたい。

◇ 目的に応じて変わるデータの扱い

データについては、その使う目的によってどう保存したらよいか変わってくる。

一番単純なのが、あるデータを探して読む、それだけだろう。この場合はあまり漢字コードだの形式だのを考えずに、Webブラウザで開いたデータを、「名前を付けて保存」で、適当なフォルダに保存すればよいと思われる。むろんHTML形

式で保存されているので、あとでまたブラウザで開いてみればそれだけでかまわない。

次に、レポートなり論文なりで自分のデータとして使いたい場合である。こちらはちと工夫が必要になるかもしれない。

まず、相手のデータがどのような漢字データを元に作られているか、見極めておく必要がある。例えば前章で紹介した「国学」であれば、簡体字のGBコードで、台湾中央研究院の「漢籍電子文献」であれば、繁体字のBig5コードで作成されている。最近は、国際コードであるUnicodeでデータを作成しているサイトも増えてきている。これはデータ形式にあまり考慮しなくてよいので、一番楽かもしれない。

現代文学のデータであれば、簡体字をそのまま使ってもよいかもしれない。しかし古典の場合それではまずいと思われる。とはいえ、簡体字から繁体字から直すのは容易ではない。

サイトによってはデータをアーカイブの形で置くところもある。その場合は、テキストデータなのか、またその他ののかで対応が変わってくる。以下、幾つか実際の例を挙げながら述べてみたい。

◇ 簡体字データの場合

まず簡体字GBコードで作成された古典データをコピー&ペーストして変換するパターンを、国学(<http://www.guoxue.com/>)を例にとってやってみよう。国学サイトでは「国学宝典」の普及版を無料で公開(<http://so.guoxue.com/>)している。

データの検索は、GoogleやYahoo!などを使い

方はそう変わりはない。要するに、自分の探したいデータのキーワードを入れて、検索ボタンを押すだけである。ここはむしろ、キーワード自体は簡体字で入力する必要がある。例によって「招宝七郎」で検索してみたが、これは『水滸伝』に2例が検索できるだけである。そのひとつである『水滸伝』77回を開いてみる。するとその第77回がまるまる出ているので、必要な箇所をマウスでクリックして指定し、コピーする。

国学の『史記』などのデータの場合、白地で隠した「齋」などの文字を中に入り込ませている場合があり、注意が必要である。これは違法コピーを防止するため、単純コピーした場合にその文字が浮き出るようにわざと挿入されているのである。このことを知らないで単純にコピーすると痛い目にあう。

もっとも、ここで使用する『水滸伝』などのデータには、あまりそういった防止策は施されていない。国学の他にも、そういった様々なコピー防止の処理を行っているサイトがあるので、これも要注意だ。

さて、またコピーする際に注意しなければならないのは、貼り付ける側のソフト側の対応である。例えば、このGBのデータをそのままJIS漢字にしか対応していない普通のエディタに貼り付けると、簡体字の部分が全部「?」に化けてしまう。

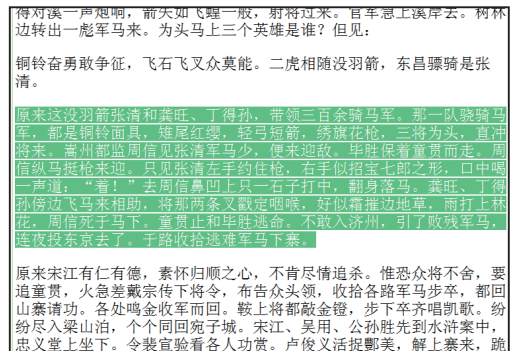
??? 猛将，正是?志、史?。???, ?口刀，却好截住?秉彝、李明?个?官廡?。李明挺?向前来斗?志，?秉彝使方天戟来?史?。??儿在山坡下一来一往，??旋旋，各逞平生武?。

これでは何がなんだか分からない。要するに、Unicodeに対応していないソフトを使うと、そうになってしまうのである。だから貼り付ける先は、Wordやワードパッド、またWindows XPのメモ帳など、Unicode対応ソフトを使わないといけない。

貼り付ける時も、そのままの何も考えずにやるとHTMLの書式が保持されたままだったりするので、例えばWordであれば「編集」から「形式

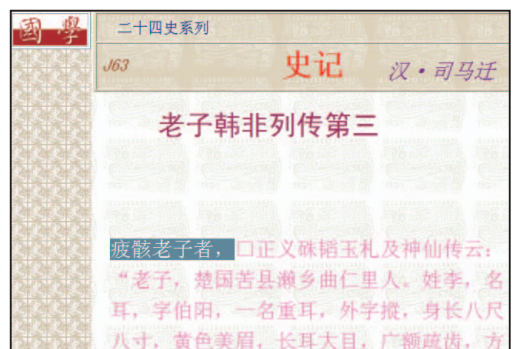
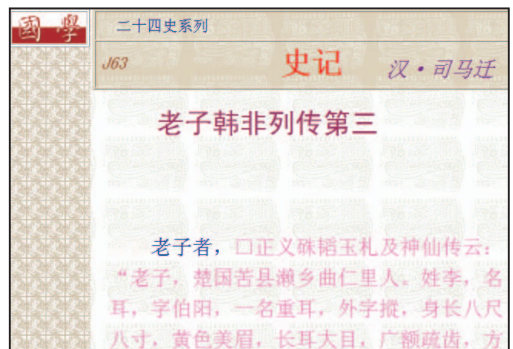


「招宝七郎」で検索する



「招宝七郎」を含む段落を選択

『史記』老子韓非列伝第三冒頭



『史記』老子韓非列伝第三 冒頭「老子者」を選択すると隠しテキストの「疲骸」が出現。

を選択して貼り付け」を選び、「Unicode テキスト」を指定して貼り付ける必要がある。上で化けたデータは、Word で貼り付ければちゃんと以下のようになる。

这两员猛将，正是杨志、史进。两骑马，两口刀，却好截住吴秉彝、李明两个军官厮杀。李明挺枪向前来斗杨志，吴秉彝使方天戟来战史进。

論文などに引用する時、このままでもかまわないとする場合もあるだろうが、しかし句読点が中国式になっているので、これは改めたほうがよい。Word やエディタの置換機能などで、「、」を「・」に変換し、それから「，」を「，」に変換する。この順序を間違えると「・」にすべき所までごっちゃになるので注意。この他、中国語の句読点は「：」や「；」なども使用されるので、日本語文書の場合より、幾つか変換を余計に行う必要がある。「！」「？」なども注意。

簡体字から繁体字への変換については、いろいろ対応ソフトがある^[1]。自分が個人的によく使うのが、Microsoft Office 用の多言語拡張ツールである「MS Proofing Tools」に付属するものである。Proofing Tools の Word の変換機能である^[2]。

使い方は、[ツール]→[その他校正ツール]→[中国語の翻訳]で、[簡体字から繁体字]を選択し、[共通の用語を翻訳する]のチェックを外して実行するだけである。すると、以下のような形式になり、古典データの引用としてもそう問題のない形式に変換される。

這兩員猛將，正是楊志・史進。兩騎馬，兩口刀，卻好截住吳秉彝、李明兩個軍官廝殺。李明挺槍向前來鬥楊志，吳秉彝使方天戟來戰史進。

この場合は外字を使ってないので問題ないが、いくつかのサイトでは GB コードに含まれない漢字を外字に置き換えている場合がある。特に文意の通じないところなどは要注意である。引用とし

て使う場合は、常に書籍や画像とのチェックを行うようオススメする。

◇ 繁体字データの場合

次に繁体字の例を、台湾中央研究院の漢籍電子文献 (<http://www.sinica.edu.tw/~tdbproj/handy1/>) をもとにやってみよう。

こちらの場合、簡体字データを使うよりはるかに楽である。中央研究院の使い方は「資料庫」を開いて、「二十五史」や「十三経」などの検索したいデータを開いてほしい。昨今は「正統道蔵」の一部データも検索できるようになっているが、外字だらけなので、かなり扱いは面倒である。

検索する時は、Big5 の漢字で引くように注意してほしい。具体的にいうと、例えば「黄巢」を「二十五史」から引きたいとする。しかし単に日本語 IME で「黄巢」と単純に入力しても、1 件もヒットしない。『旧唐書』も『新唐書』もヒットしないなんて、そんなバカなと思うだろうが、事実である。これは漢籍電子文献のバグではなく、漢字コードの違いによるものである。

普段はあまり意識しないだろうが、我々の使っている日本語の漢字も一種簡略化されたものであり、Unicode 上では、ちょっとした形の違によりそれまでの伝統的な漢字の形と区別されてしまっている。

9EC4 「黄」 9EC3 「黃」

このように「黄」と「黃」が別の漢字として認識されるために、「黄巢」は検索できない。さらに「巢」の字も「巢」と「窠」ではコード番号が違う。だからここでは「黄巢」と入力せねばならない。すると 314 件ヒットする。

黄巢，曹州冤句人，本以販鹽為事。乾符中，仍歲凶荒，人飢為盜，河南尤甚。初，里人王仙芝、尚君長聚盜，起於濮陽，攻剽城邑，陷曹、濮及鄆州。

これは『旧唐書』黄巢伝の冒頭部分である。このデータも、やはり簡体字と同様に句読点を置き換える必要がある。まず「,」を「、」に、また今度は「.」を「。」に変換する必要がある。

ところで、簡体字・繁体字双方のデータにいえることだが、保存する時は、あまりコードを気にすることのない Word 形式（～.doc）のほうがよいと思われる。

ただ、データを解析したり、他の用途に転用したりするときは Word 形式では不便である。その場合は Unicode のテキストデータとして保存する方がよいだろう。保存するときに、「ファイルの種類」をテキストファイルにするようにすれば、あとは Word がいろいろ指示してくれる。

◇ テキストデータなどの場合

上の2つの例はいずれも Web ブラウザを使って検索したデータを取り込むものであった。

Google などで検索した場合、検索されたデータは必ずしも HTML 形式とは限らない。こういったデータを扱う時は、それぞれどんな形式で保存されているのか、注意が必要になる。

PDF ファイル（～.pdf）の場合は当然だが、アドビ Acrobat を使う必要がある。この場合は、Acrobat のほうで、どういった形式か把握し、知らせてくれるのでかえって楽である。

テキストデータの場合はいささか厄介で、形式的には GB か Big5 か、JIS か Unicode か、全然区別がつかない（全部～.txt となる）。そのために開いてみたらポロポロに文字化けしているというのは、よくある話である。

Word はこのような事態によく対応しており、テキストデータを開く場合でも、Word を使って

みることをオススメする。Word はあらかじめ開く前にそのデータが GB か Big5 かなど、幾つかの可能性を提示してくれるのである。そのレビューにおいて、漢字がちゃんと表示されていれば、まずそのコードであることは間違いない。それでも文字化けする場合は、幾つかの形式を試してみるといいだろう。

なお、Unicode のデータだからといって、油断は禁物である。Unicode は当初約 2 万字の漢字しか使えなかったが、昨今はそれが拡張されて、約 9 万字にまで増加している。拡張された漢字は、拡張領域 A と B とが使えるようになっている。

これらの拡張漢字を中に折り込んで作成されたデータも見かけるようになってきた。このような場合、開いたときその文字が「・」で表示されることが多い。実はちゃんと文字が入っているのだが、フォントなどがなかったために表示されないのである^[3]。

いずれにせよ、漫然とデータをコピー&ペーストするのではなく、データの形式がどのようなものであるか、常に意識しながら扱うようにこころがける必要がある。

注

[1] テキスト形式のファイルならば、レビューで紹介している EmEditor の繁体字と簡体字変換プラグインを使うという方法がある。ただし、JIS 漢字には対応していないので要注意。

その他、テキストファイルを対象にした中国語コンバーターには、竹内信介氏による「中国語変換ツール PinConv+」<http://www.ctrans.org/> などがある。

[2] 「MS Proofing Tools」については、前号の特集 1 の拙稿「Windows で多言語・多漢字を使う」を参照のこと。

[3] フォントについては、レビューの当該記事を参照のこと。

データの整形と整理

小島 浩之（こじま ひろゆき）

◇ はじめに

本誌の読者は、これまでに、Web上にどのようなオンラインアーカイブズが存在し、それらがどのような特徴を持っているか、言わばオンラインアーカイブズ書誌学とも呼ぶべき知識を得ることができた。

またデータの簡単な加工術も身につけることができた。ただしその加工術はダウンロードしたデータをそのまま保存するためのノウハウが中心であった。

そこでここでは、この加工術に関してデータを整形・整理するレベルまで話を進めてみよう。

◇ 1. データの整形とは何か

データの整形とは、元データを加工して利用者自身のスタイルに合わせたものにするのである。

多くの保存したデータの中から、検索をして意中の字句を探索するだけであれば、手間暇掛けてデータを整形する必要はほとんどない。検索機能を利用するためには、文字列が保存されていればよいからである。

しかし、データを検索するだけで事足りるとする読者は少ないであろう。検索したいは研究のためのツール（手段）であって、最終目的ではないからである。検索して必要なデータを集め、精査し、解釈し、最終的には論文など研究発表の場で利用せねばならない。論文等への利用ということになるとどうしても多少の整形が必要になってくる。

これは、漢籍を博捜し必要な部分をカードに書き留め、それを読み込んで論文に使うという昔からの研究スタイルであっても同じだろう。白文の漢籍を読んでも、カードに採るとき、もしくは最終的に引用する際には、句読を打ち、適宜改行施すといった整形を余儀なくされる。本稿で考える整形とはこういったものだと思ってもらえればよい。

整形されたデータは、研究に利用する際の最終的なデータの姿である。したがって、整形は必要に応じて行うことが基本である。最初から整形しすぎると、他の用途に転用できなくなったり、再度の整形が面倒になったりする恐れがあるからだ。しかし中にはデータを収集した際に行った方が望ましい整形もある。

そこで以下、「最初に行っておくとよい整形」と「必要に応じて行う整形」の二つに分けて見てみることにしよう。

◇ 2. 最初に行っておくとよい整形

◎ (1) 半角スペースの置換

先に「検索をして意中の字句を探索するだけであれば、手間暇掛けてデータを整形する必要はほとんどない。」と述べたが、元データによってはそうもいかない場合がある。千田氏の「オンラインテキストアーカイブ概説」にも指摘されているが、例えば、中華文化網などのような、文字間に半角スペースが入っているものはその一例である。

次ページ図1は、中華文化網からエディターにコピー&ペーストしたもので、着色部分は全

て半角スペースである。もしこのまま保存すれば、検索キーワードの文字間に一々半角スペースを入れて検索しなければならない。例えば、人名の“顔師古”を検索したくても、“顔△師△古”（△は半角スペースを表す）と入力しなければヒットしないのである。保存データの中に文字間に半角スペースの有るものと、無いものが混在すると、Grepなど複数のファイルを一度に検索する際に二度手間となる。スムーズに検索するためには、あらかじめ半角スペースを取り除いてからテキストを保存するのが賢明だろう。

テキストから特定の文字列を削除する場合、エディターなどの置換を利用しよう。置換についての詳細は、次の山田氏「検索・置換・GREG」を参照されたい。山田氏が採り上げるEmEditorを利用して半角スペースを削除するには、[検索する文字列]に半角スペースを入力し、[置換後の文字列]に何も入力せず置換を実行すればよい。（以下、エディターの置換についてはEmEditorによる方法と結論のみを記すので、具体的には山田氏の論考を参照のこと。）

◎(2) 文字色の変換

HTML形式で公開されているデータの場合、フォントが色分けされていることがある。テキストファイルは、改行以外の制御情報は保存されない。このため、テキストファイル形式での保存であれば、元データのフォントの大きさ、色、種類といった情報に影響されることはない。しかしWordの場合、これらの情報も保存されてしまう。したがって中華文化網のように見出し、本文、注が色分けされているようなデータをWordに貼り付けると、文字の色まで全て再現されてしまう。文字の色によって（例えば黄色など）は、大変見づらいデータとなる。この場合、やはり最初に文字色を黒に変換した方がよいだろう。

Wordの場合は[編集]→[すべて選択]で文全体を反転させた後、ツールバーメニューからフォントを選択しプルダウンメニューで色のパレットを出し自動を選択する（図2参照）ことで、文字色を黒に統一できる。

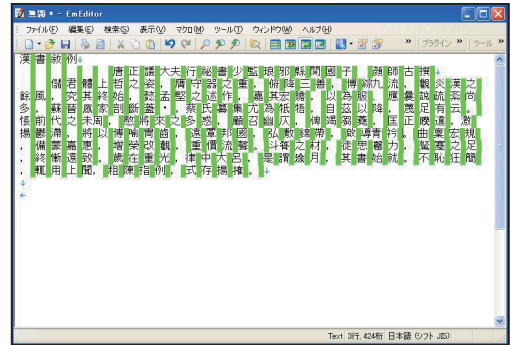


図1

また別の方法として、文全体を反転させた後、カーソルを反転した部分に置いて、右クリックメニューの[フォント]をクリック、フォントメニューのダイアログが出たら、中程左側の[フォントの色]のプルダウンメニューで自動を選択する。この状態で[OK]をクリックすれば文字色は黒一色になる。

◎(3) 句読点の置換

本特集の他の部分でも触れられているが、中国式の句読点（標点符号）も、あらかじめ日本式に置換・整形しておいた方がよいだろう。データ検



図2

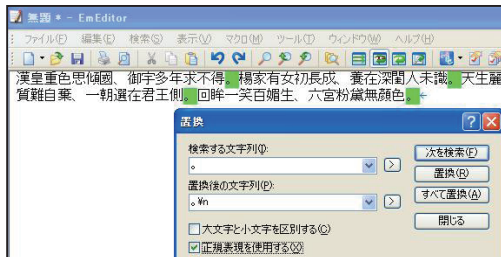


図 3

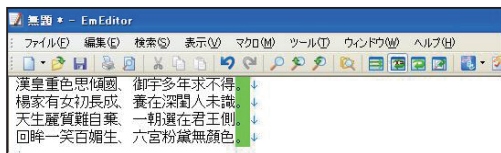


図 4

索上の問題はないけれど、日本で論文等に利用する際は、日本式の句読点を使用するのが一般的だからだ。最初の時点で一気に修正しておいた方が後々楽である。

手順としては最初に [,] (頓号) を [・] (ナカグロ) に変換する。次に [,] (カンマ) を [、] (読点) に変換する。日本語の読点は中国語では頓号すなわちなカグロの意味になる。つまり同じ記号が日中で異なる意味で使われているため、上記の変換順序を間違えると大変なことになる。このほか [?] や [!] も [。] (句点) に変換しておこう。

◇ 3. 必要に応じて行う整形

このように最初に必要な整形を加えられたデータは各自のパソコン内に蓄積される。このデータがさらなる整形を加えられるのは、検索、精査、解釈され、最終的な利用が決まった段階だろう。この場合の整形は多様であって、一概に論じられるものではない。そこで以下には、エディターと Microsoft Excel (以下エクセル) の実際に利用できそうな機能について解説する。これらを参考にして、それぞれの目的とする整形に応用していただければと思う。

◎ (1) エディターで整形

次の文章は白居易「長恨歌」の冒頭のテキスト

データである。

漢皇重色思傾國、御宇多年求不得。楊家有女初長成、養在深閨人未識。天生麗質難自棄、一朝選在君王側。回眸一笑百媚生、六宮粉黛無顏色。

これを一文一行のデータに整形する場合を考えてみる。一文一行のデータということは一文ごとに改行を入れればよい。この場合もテキストエディターの置換機能を利用することで解決する。[検索する文字列] に [。] を入力し、[置換後の文字列] に [。¥n] (¥n は改行を意味する正規表現。正規表現については後掲の山田氏の解説を参照のこと。) と入力して置換を実行すれば (図 3)、たちまち一文ごとのデータに整形される (図 4)。

特に韻文であると、句や聯、文ごとで改行して表示したい場合もあるだろう。こういった際にこの方法は有効だ。また次のデータのように、

漢皇重色思傾國、御宇多年求不得。楊家有女初長成、養在深閨人未識。▲
天生麗質難自棄、一朝選在君王側。回眸一笑百媚生、六宮粉黛無顏色。▲

既に改行 (▲は改行コードを示す) が入っているデータでこの処置を行う場合は、一旦、全ての改行コードを削除してから (削除の置換は本稿第 1 章を参照のこと)、先ほどの処置を行うこと。そうしないと元からの改行コードが残っているため、次のように空白行が生じてしまう (▼は整形により新たに加えられた改行コードを示す)。

漢皇重色思傾國、御宇多年求不得。▼
楊家有女初長成、養在深閨人未識。▼
▲
天生麗質難自棄、一朝選在君王側。▼
回眸一笑百媚生、六宮粉黛無顏色。▼
▲

◎(2) エクセルの機能を整形に利用

エクセルは表計算ソフト（英語名：spreadsheet）である。表計算ソフトとは本来、数値データの集計・分析に利用されるソフトのことをいう。ところが計算だけでなく、作表や簡易データベースとしての機能も優れているため多くの用途に利用されている。複雑な計算を行うために関数が定義され、プログラミングによって一連の処理を自動的に実行させることもできる（マクロ機能）。実は関数やマクロも使い方次第ではデータの整形に利用できる。

エクセルは作成したデータをテキストファイル形式（タブ区切り）、CSV形式（カンマ区切り）、XMLなどに変換して保存することもできる。また逆にこれらの形式で作成されたデータをエクセルに呼び出して利用することもできる。ここではエクセルの機能のうち、整形に役立つと思われる機能を紹介してみたい。なおここではエクセル2003に基づいて解説することにする。

●① Webデータの取り込み

エクセルではWeb上のデータを取り込むことができる。本来はWeb上の数値データを取り込むものなのだが、文字データを取り込んでテキストファイルとして保存するツールとしても利用できる。メニューバーの[データ]→[外部データの取り込み]→[新しいWebクエリ]をクリックすると、ダイアログが立ち上がる。ここでアドレス欄に取り込みたいデータ元のURLを入力し、移動ボタンをクリックすると、目的のサイトがダイアログ内に表示される。画面上に出ている矢印記号のうち、画面全体が反転するものを選んでクリックし（図5）、取り込みボタンをクリックする。

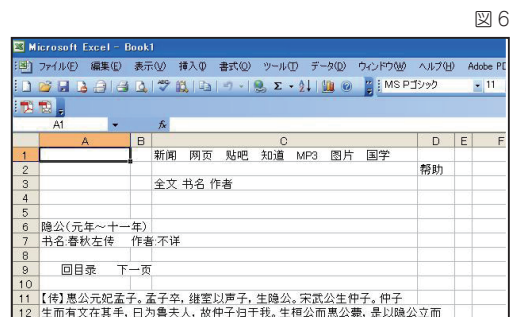
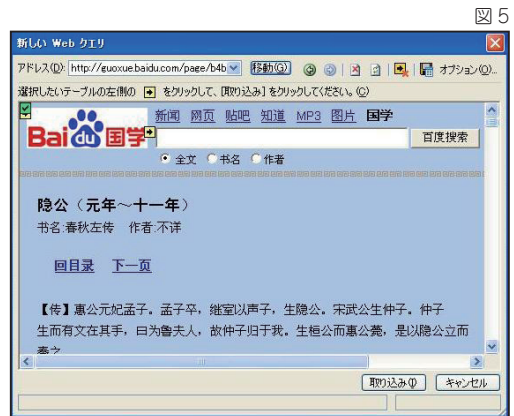
すると新たに[データのインポート]ダイアログが立ち上がるので、開始セルを選択して[OK]ボタンをクリックすれば、データが取り込まれる（図6）。

あとは[ファイル]→[名前を付けて保存]で[ファイルの種類]を[Unicodeテキスト]とし、適当な名称を付して保存すればよい。このとき注

意を促すダイアログが二つ立ち上がるが、いずれも[はい]をクリックしておく。エクセルに取り込めば、セル、行列単位で楽に整形を施して保存することができる。ただし文頭から改行までが1セルに入るので、文字列が多すぎるとエクセルの仕様に影響を受ける。エクセルの1セル当たりの最大文字数は32,767であり、これを超えない範囲での利用に制限されることに注意されたし。

●② テキストデータの読み込み

テキストファイル形式のデータをエクセルに読み込んで編集することもできる。メニューの[ファイル]→[開く]をクリックし、対象ファイルを選択する。するとテキストファイルウィザードが立ち上がる（図7）。この時データが文字化けしているようならば、文字コードのプルダウンメニューでコードを再指定すればよい。後は必要に応じて、区切り文字等を指定する。対象のデータにあらかじめタブを設定しておく、タブをセルの区切りにできる。例えば句ごとにセルを分割し



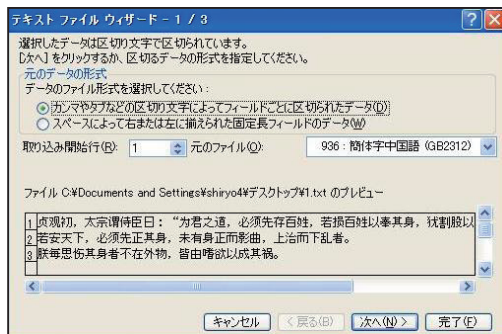


図 7

たければ、あらかじめ句読点を [句読点+タブ] に置換しておいてから読み込めばよい。

●③ 区切り位置の指定

セルに文章を打ち込んだり、貼り付けた後でも区切り位置を指定して、複数のセルに分割したりすることができる。分割したいセルを選択して、メニューの [データ] → [区切り位置] を選択すると、図 7 とほぼ同様のダイアログが立ち上がるので、区切りとする記号を指定すればよい。

●④ 置換機能

エクセルにもエディター同様に置換機能がある。詳細は当会編『電腦中国学Ⅱ』（好文出版 2001 年）286 頁^[1]を参照のこと。当該書籍はエクセル 2000 に依拠した解説であるが、エクセル 2003 においても変化はない。

置換はメニューの [編集] → [置換] から行う。ワイルドカードとしては [?] と [*] の二文字がある^[2]。

●⑤ 関数

エクセルには複雑な計算をも行えるように、様々な関数が定義されている。これらは単独で用いられるばかりでなく、複数の関数を組み合わせることによってより高度で複雑な計算が可能となる。本来は表計算ソフトとしてデータ分析に利用するものなのだが、使い方によっては文字列に対しても有効なものも多い。特に文字列操作関数と呼ばれるグループの関数は、データの整形や抽出

に利用できる。

例えば、LEFT、MID、RIGHT はそれぞれ文字列の左、指定位置、右から指定した文字数を取り出す関数である。それぞれの書式は次のようになる。

```
=LEFT ( 文字列 , 文字数 )
=MID ( 文字列 , 開始位置 , 文字数 )
=RIGHT ( 文字列 , 文字数 )
```

これらと他の文字列操作関数を組み合わせることで、データから必要とする部分だけを抽出したり、逆に不必要な部分を削除したりできる。エクセル関数の具体的な利用法については、本稿の僅かなスペースで簡単に解説できるものではない。専著も多くあり、また Web 上でも解説ページが多く開設されているので、興味のある方はそちらをご覧ください。

●マクロ (macro)

マクロとは特定の操作・処理を自動で実行するプログラムのことである。エクセルのマクロが有名だが、各種ワープロソフトやエディターでもマクロを完備しているものがある。プログラムであるから、大量のデータに対し、同一の処理を繰り返して行うのに向いている。

エクセルのマクロは VBA (Visual Basic for Applications) というプログラミング言語で書かれる。このため自分でプログラミングするには、VBA を理解する必要がある。ただし現在は Web 上で多くのマクロプログラムが公開されているので、それらを拝借できる。つまり VBA が解らなくてもマクロをデータ整形に役立たせることができる。しかし VBA が解らないということは、プログラムの検証ができないことに等しい。これでは、もし拝借したプログラムに大きな誤りがあったとしても見逃してしまうかもしれない。他人がプログラミングしたマクロはあくまで自己責任で利用されたい。マクロのプログラムを書くにはメニューの [ツール] → [マクロ] から行う。既存のマクロを拝借する際は、プログラムを自分で書く代わり

に、コピー & ペーストすればよい。

マクロに関しても多くの専著があるので、詳細はそちらに譲る。

◇ データの保存と整理

データの整形が終われば、次にはそのデータを保存・整理して活用できるようにしなければならない。

◎ (1) データの保存

データの保存にあたり、最終的な保存形式は、やはりテキストファイル形式が望ましいだろう。

テキストファイル形式であれば確実にエディターの Grep 機能が利用でき検索に便利である。また種々のデータ形式への変換が容易な点や、容量が小さい点もメリットに挙げられる。

◎ (2) データの整理

● 必ずフォルダに！

データの保存には必ず専用のフォルダを用意しよう。他の文書類との混在は出来る限り避けるべきである。もし可能であればハードディスクからして専用のものにした方がよいかもしれない。

専用フォルダの下には、分類を決め分類ごとのフォルダを作成しよう。例えば経、史、子、集、叢書、新学、その他と伝統的な四部分類を中心としたフォルダに分割するのも一つの方法である。主題分類が面倒な人は作成年月で分類するとよいだろう。

分類フォルダの下には、書籍ごとにフォルダを作成しよう。1つの書籍が1つのテキストファイルとは限らないが、たとえ1ファイルであっても Grep 対応のためにフォルダを作っておこう（Grep については後掲の山田氏の解説を参照のこと）。こうすることで、Grep の範囲を資料全体、分類ごと、書籍ごとの三段階から選べ、検索の柔

軟性と効率性を高めることができる。逆に言えば、せっかく作ったファイルをデスクトップ上に放っておいたり、他の種類のファイル類と同一フォルダに混在させたのでは、検索機能が全く生かせなかったり、検索効率を悪くしたりするのである。

また、最近はフォルダ整理用ツールがフリーソフトとして多く提供されている。これらを利用することで整理効率を上げることができるだろう。

◇ おわりに

以上、駆け足でデータの整形と整理について述べてきた。整形と整理は絶対的な指針を出すことは不可能に近い。これは、個々の研究分野によって求めるものが若干異なるからである。また同一人物であっても、いつも同じ整形・整理をすることも限らない。データの使用目的によっても整形・整理に求めるものは異なるのである。

筆者は、編集局からこの執筆を依頼された時点では楽観的に構えていた。しかしいざ執筆に取りかかってみると、大変に書きにくいテーマであった。最終的にはご覧のように整形、整理に利用できそうなソフトの機能解説に終始せざるを得なかった。従って本稿はデータの整形と整理のためのヒント程度に捉えていただければと思う。

また筆者の専門は歴史学のため、どうしても歴史学寄りの記述になってしまっている。他分野の方には読みにくい点、理解しにくい点があるかと思うが、ご海容いただきたい。

注

[1] 該当部分は山田崇仁氏の執筆。

[2] ワイルドカードとは、任意の文字を意味する特殊文字のこと。[?] は「任意の1文字」を、[*] は「任意の複数の文字列」を表す。

検索・置換・GREP

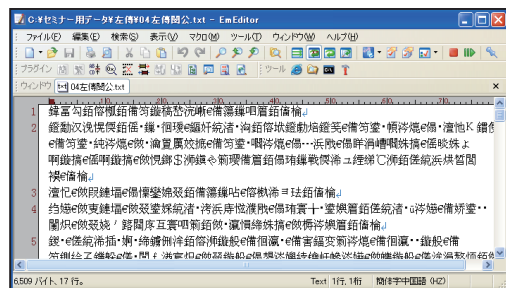
山田 崇仁 (やまだ たかひと)

◇ はじめに

これまでの課程で、自分で入力したもの、あるいは、インターネット上のテキストサイトから入

手した電子テキストを、使いやすい形式に加工し、項目別フォルダに整理してあるはずだ。次に、そこから様々な形でキーワードを指定して検索することを目標とする。

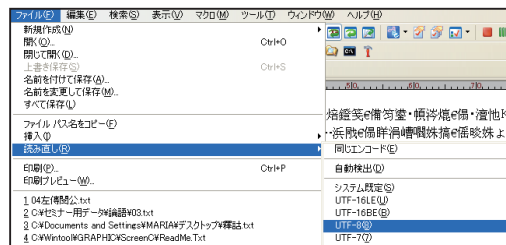
ここでは、EmEditor を対象として説明する。EmEditor を選択した理由は「Unicode の BMP 領域のみならず「CJK 統合漢字拡張 B」領域に対応しているためである^[1]。



文字コードの指定を間違えて文字化け

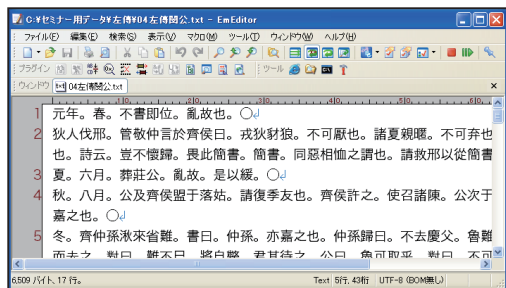
◇ ファイルを開く

検索を行う前に、EmEditor でのファイルの開き方を説明しておこう。



正しい文字コードを指定してファイルを読み直し

1. EmEditor を起動。
2. メニューの [ファイル] → 「開く」を選択し、ダイアログボックスが表示されたら、対象となるテキストファイルを選択する。
3. 開く際に、テキストファイルの文字コードは自動で解釈される。
4. もし、間違った文字コードに解釈されてしまい、画面表示が文字化けしている場合は、メニューから [ファイル] → [読み直し] → [正しい文字コード (ここでは UTF-8)] を選択すれば、文字化けが解消される。



文字コードが正しく解釈され、文字化けが解消された

◇ 単一のファイルから検索

ファイルを開き、文字化けが無いのを確認したら、いよいよ検索だ。

- 1.はじめに、[Ctrl] を押しながら [HOME] キーを押して、ファイル先頭にカーソルを移動（末尾に移動するには、[HOME] の代わりに [End] キーを押すこと）^[2]。
- 2.カーソルがファイル先頭に移動したのを確認したら、メニューから [編集] → [検索] を指定してダイアログボックスを開く。

ダイアログボックスで、キーワードの指定などを行う。ここでは、「具体的なキーワードを指定」「正規表現でキーワードを指定」について説明する。

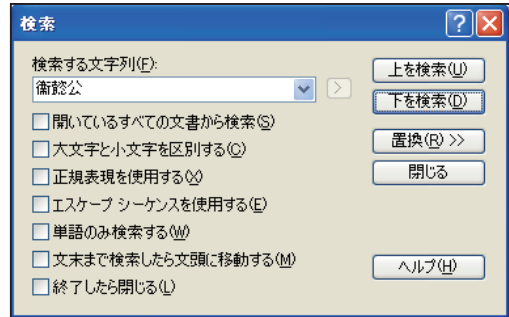
◎ 具体的なキーワードを指定

- 1.[検索する文字列] 下の文字を入力する部分（「テキストボックス」）を左クリックしてカーソルを移動。
- 2.カーソルが移動したら、テキストボックスにキーワードを入力。
- 3.入力したら [下を検索] を左クリック。
- 4.検索が実行されて、指定したキーワードに最初に一致（マッチ）した部分にジャンプする。また、ファイル中のキーワードにマッチした部分すべてが、反転表示される。
- 5.[下を検索] をクリックすると、次のマッチ部分にジャンプする。
- 6.前のマッチ部分に戻るには、[上を検索] をクリックすること。

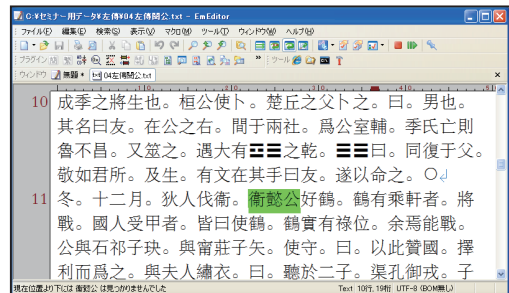
◎ キーワードを再利用するには

EmEditor では、過去に利用した検索キーワードの履歴がある程度保存されており、その履歴を、再利用することができる。

- 1.検索の際、[検索する文字列] のテキストボックス右端の [▼] を左クリックする。
- 2.クリック後、過去に使用したキーワードが一覧表示される。
- 3.その中から再利用したいキーワードをマウスで選択し、左クリックする。
- 4.テキストボックス内に、過去に使用されたキーワードが再入力される。



検索のダイアログボックス（キーワードに「衛懿公」指定）



検索結果：衛懿公が色つき表示されている

◎ 検索のオプション

EmEditor の検索機能では、単にキーワードを指定する以外に、様々なオプションを指定することで詳細な検索が可能となる。

各オプションの [ON]・[OFF] は、それぞれの項目左横の [チェックボックス] にチェックを入れる [ON] か入れないか [OFF] で切り替える。

● 開いているすべての文書から検索

[ON] にすると、EmEditor で現在開いているすべてのファイルを対象に検索を行う。

[OFF] にすると、今編集している（アクティブ）ファイル 1 つだけが対象となる。

● 大文字と小文字を区別する

[ON] にすると、検索の際に大文字と小文字を別の文字と判断する。

例：[Japan] をキーワードに検索した場合

- [ON] では [JAPAN] はマッチしないが、

[Japan] はマッチ。

- [OFF] では、[JAPAN] と [Japan] の両方ともマッチ。

● 正規表現

キーワードに「正規表現」を利用した「特殊な文字列の組み合わせを使って検索する方法」を利用する場合に [ON] にする。この機能については後で説明する。

● エスケープシーケンスを使用する

「改行文字」や「水平タブ」のような「画面に表示されない特殊な文字」を検索する場合に [ON] にする。EmEditor のエスケープシーケンスは左下表を参照（EmEditor のヘルプより抜粋）。

例：「改行文字」を検索したい場合

- キーワードに「¥n」を指定。

● 単語のみ検索する

[ON] にすると、単語だけを検索する。

例：「Japan」をキーワードに検索した場合

- [ON] では、「Japan」にはマッチするが、「Japanese」にはマッチしない。
- [OFF] では「Japan」と「Japanese」の両方ともマッチする。

エスケープシーケンスの一覧

エスケープシーケンス	内容
¥a	警告（ベル）
¥b	バックスペース
¥f	改頁
¥n	改行
¥t	水平タブ
¥v	垂直タブ
¥¥	円記号（バックスラッシュ）
¥oooooo	Unicode 文字（8進数表記）
¥xhhhh	Unicode 文字（16進数表記）

● 文末まで終了したら文頭に移動する

通常、EmEditor の検索は「カーソル位置→文末まで」を検索の対象とするが、このオプションを [ON] にすると「カーソル位置→文末まで」検索したあとに、「文頭から検索→カーソル位置の直前」の動作を行う。

● 終了したら閉じる

このオプションを [ON] にして検索を実行すると、検索後にダイアログボックスが自動で閉じる（キーワードにマッチする部分は色つき表示のまま）。

◎ 正規表現

[正規表現] とは、「一定のパターンで表記されている文字列を、抽象的な表記法に置き換えて表現する方法」を意味する。

いくつか例をあげてみた。それぞれ、「検索の要求」と「正規表現」の組み合わせになる。

- 2桁か3桁の（いわゆる半角）数字。

➡ [0-9]{2,3}

- ひらがな・カタカナのみを検索したい。

➡ [あ-ざぁ-づ]

- 「以上」「以下」以外の「以」を含む用例（例：以内）。

➡ 以[^上下]

- html 文書のタグをすべて削除。

➡ </?.+?>

※ 1つめの「?」は「0もしくは1文字」、2つめの「?」は「最短一致指定」になる。

- 新式標点本の注釈記号を検索（〔半角の四角括弧で囲まれている+漢数字と0が1～2文字使われる〕）。

➡ ¥[[一二三四五六七八九0]{1,2}¥]

上記事例の正規表現では「\{\}」などの記号が使われているが、正規表現の文法では、これらの記号のように、いくつかの文字に通常とは異なる特殊な意味を与えており、それらを [メタ文字] あるいは [メタキャラクタ] と呼ぶ。

正規表現は、「通常の文字」と「メタ文字」とを組み合わせることで、具体的なキーワードでは指定できない複雑な指定をすることができる。そのため、正規表現は検索をパワーアップする便利な機能として位置づけられており、様々なテキストエディタのみならず、Perl や Ruby 等のスクリプト言語・C 等のプログラム言語、Microsoft Word や一太郎などのワープロソフトにも使われている。

ただし、それぞれのメタ文字を用いた正規表現の表記ルールに違いがある点に注意されたし。

右に、EmEditor の正規表現の主な表記ルールをあげておく。いずれも所謂半角（1byte）で表記する点に注意されたし。他の EmEditor のメタ文字は、EmEditor のヘルプを参照のこと。

◇ 単一のファイルから置換

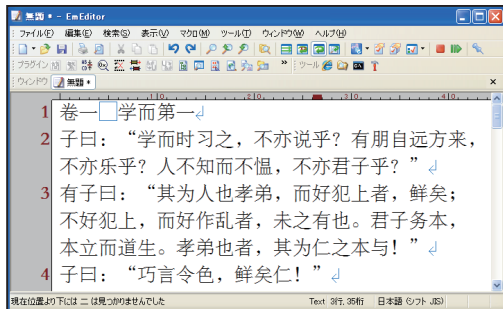
EmEditor では、指定した文字列を別な文字列に置換する機能がある。ここでは、単一ファイルを対象とした置換について説明する。

1. はじめに、置換対象とするファイルを開く（手順は上述）。
2. ここでは、中国語の句読点マークアップである「，（全角カンマ）」を「、（読点）」に置き換えてみる。
3. カーソルをテキストの先頭に移動（手順は上述）。
4. メニューから [検索] → [置換] を選択。
5. 置換のダイアログボックスが開く。
6. [検索する文字列] 横のテキストボックスを左クリックしカーソルを移動。
7. 移動したら、テキストボックスに置き換いたい文字列（ここでは [,]）を入力。
8. 同様に [置換後の文字列] 横のテキストボ

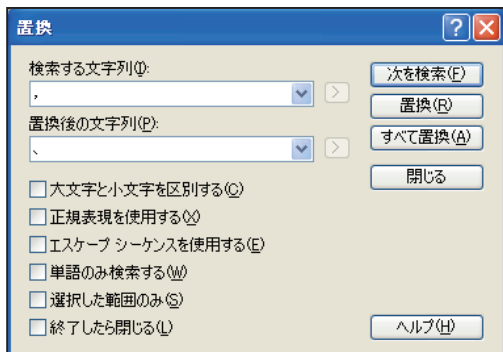
- ックスを左クリックし、置き換える文字列（ここでは [,]）を入力。
9. 入力したら [置換] を左クリック。
10. キーワードに最初にマッチする部分が置換され、次のマッチ部分にジャンプする。
11. さらに [置換] を押すと、同じ事が文末

メタ文字	内容
行の位置を指定	
^ (キャレット)	行の先頭と一致。
\$ (ドル記号)	行の末尾と一致。
文字を指定	
.	(改行以外の) 任意の 1 文字に一致。
*	「直前の文字または正規表現」と 0 回以上一致。
+	「直前の文字または正規表現」と 1 回以上一致。
?	「直前の文字または正規表現」と 0 もしくは 1 回一致。 下の最短一致と文字は同じだが、その振る舞いが異なる。
{n}	n には 0 以上の整数を指定。 「直前の文字または正規表現」と n 回一致。
{n,}	n には 0 以上の整数を指定。 「直前の文字または正規表現」と n 回以上一致。
{n,m}	m と n には 0 以上の整数を指定。 「直前の文字または正規表現」と n-m 回一致。
?	上記メタ文字の中で「.」「?」以外のメタ文字の直後に記述することで、最短一致（最短の正規表現パターン）指定になる。
文字列のパターンを指定	
(pattern)	メタ文字の pattern をグループ化。及びグループ化した文字を取得（一時的に EmEditor 内部でグループ指定した正規表現にマッチするパターンを保持。 置換の際に後方参照のメタ文字（例：¥1）を使用することで再利用可能）。
¥1, ¥2, ..., ¥9	上記メタ文字の pattern を後方参照で再利用する際に使用。 [置換後の文字列] でのみ使用可能。
x y	x もしくは y と一致。
[xyz]	角括弧内の文字いずれか 1 文字に一致。
[^xyz]	角括弧内の文字以外のいずれか 1 文字に一致。
[a-z]	指定範囲内の任意の 1 文字に一致。 範囲は文字表の配列順で指定。
[^a-z]	指定範囲外の任意の 1 文字に一致。

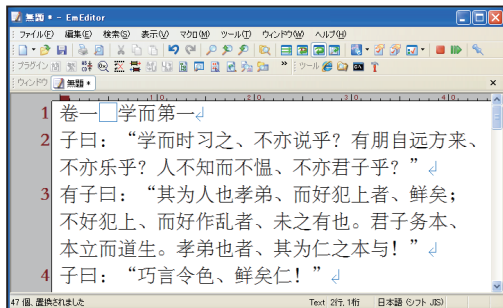
EmEditor の正規表現のメタ文字とその意味



置換実行前の文章



置換のダイアログボックス



置換実行後の文章

まで繰り返される。

- 置換を実行する際、[全置換] を左クリックすると、文書内すべてのマッチ部分が一気に置換される。

◎ 置換のオプション

置換のオプションは基本的に検索のオプションと同一である。ただし、[開いているすべての文書から検索] と [文末まで終了したら文頭に移動する] のオプションはない。

置換専用のオプションとしては、[選択した範囲のみ] がある。これは、あらかじめ選択した範囲のみを置換したい場合に使用する。

● 置換の正規表現

置換の際に使用できる正規表現は、基本的に検索と同じだが、置換の時にだけ使用可能な特殊な正規表現もある（次頁右上表を参照）。これを利用することで、「文章内容を保持したままその構造を入れ替える」といったこともできる。

以下に例を挙げてみた。右頁上図は小倉百一首の冒頭三歌だが、一行が何れも「歌番号：句：作者」の構造になっている。これを検索と置換の正規表現を利用して、「作者：歌」の構造に加工してみよう。

- 検索のキーワードとして、一行の構造を正規表現で指定する。

ここでは、「^(+):(+):(+)\$」と指定。

意味は、「先頭から一文字以上+：一文字以上+一文字以上+行末」になる。置換のキーワードで再利用するために、いくつかの正規表現にマッチした部分を丸括弧で囲んでグループ化して一時的に保持しておく（先頭から順に¥1, ¥2, ¥3の順でラベルが貼られる）。

- 置換キーワードに「¥3：¥2」を指定する。意味は、「前から三番目の丸括弧(作者：¥3)+：+前から二番目の丸括弧(句：¥2)」になる。

- キーワードを指定したら、[すべて置換] をクリックする。置換が実行され、文章の構造が変化するはずである（右下図）。

◆ ファイルから検索

EmEditor では、現在開いているファイルから、指定した文字列を高速に検索する以外に、フォルダやファイルの種類を指定して一括して検索することができる。

この機能は普通「grep」と呼ばれており、元々

UNIXのコマンドの一つだったのが、便利のため多くのソフトウェア環境で使われるようになったものである（EmEditorでは「ファイルから検索」）。

検索と同様に、EmEditorの「ファイルから検索」も、キーワードとして指定可能な文字が、Unicodeすべての文字が対象となる。また、対象となるファイルはテキストファイルであれば文字コードや形式は問わない（html・xmlファイルや、Shift-JISやUTF-8などで保存されたテキストファイルを検索することもできる）。

◎「ファイルから検索」の実行

1. EmEditorを起動。
2. メニューの「検索」→「ファイルから検索」を選択。
3. 「ファイルから検索」のダイアログボックスが表示される。
4. はじめにキーワードを指定する。「検索する文字列」の動作は通常の実行と同じなので、そちらを参照されたい。
5. つぎにファイルの種類を指定する。名前指定にはワイルドカード^[3]が使用できる。複数ファイルを指定する場合は、セミコロン「;」で区切ること。

例：すべてのテキストファイル：*.txt
 例：すべてのhtmlファイルとxmlファイル：*.htm;*.html;*.xml

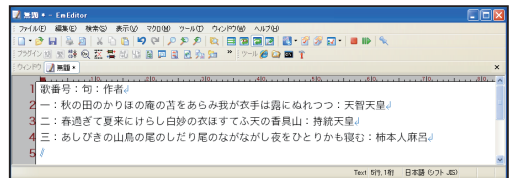
6. 検索したいファイルがあるフォルダを選択。
7. 「検索するフォルダ」横のボタンを左クリックして「フォルダの参照」ダイアログボックスを表示し、検索対象とするフォルダを指定する。
8. 検索キーワード・ファイルの種類・フォルダの場所を指定後、「検索」を左クリックすると「ファイルから検索」が実行される。

◎「ファイルから検索」の結果表示

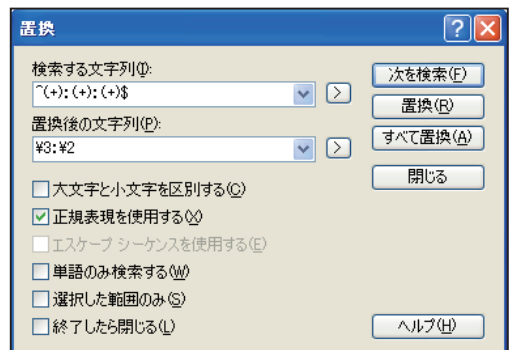
1. 「ファイルから検索」を実行すると、別なEmEditorのウィンドウが開いて、「ファイ

検索式	意味
¥0	正規表現全体への後方参照。
¥1-¥9	後方参照。
¥n	改行
¥r	「ファイルから置換」では、復帰文字として使用。
¥t	水平タブ
¥L	以後の文字を小文字に変換。
¥U	以後の文字を大文字に変換。
¥H	以後の文字を半角に変換。
¥F	以後の文字を全角に変換。
¥E	以前の¥L・¥U・¥F・¥Hによる変換を終了。

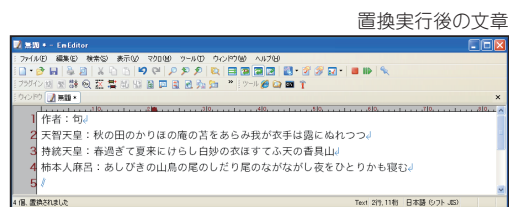
置換後の文字列で使用可能な正規表現のメタ文字とその意味



置換実行前の文章

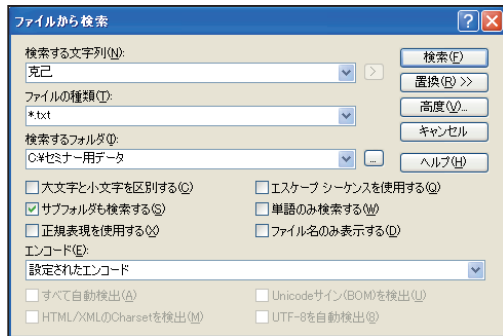


置換のダイアログボックス

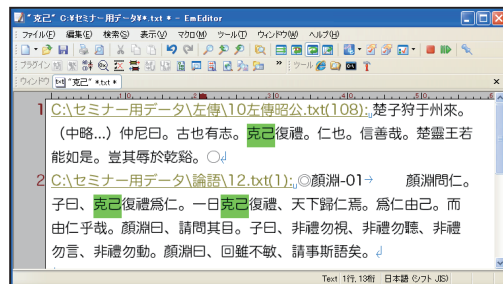


置換実行後の文章

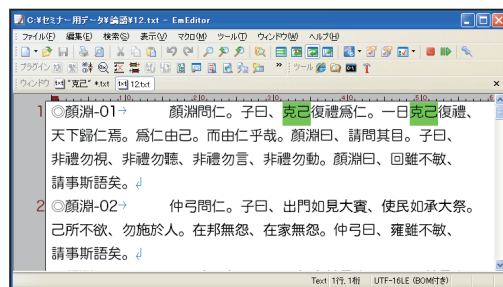
1. 「ファイルから検索」の結果が表示される。
2. 「ファイルから検索」の結果は、以下の構造で表示される。



「ファイルから検索」のダイアログボックス



「ファイルから検索」の検索結果



検索対象ファイルを開き、キーワードに一致箇所に移動

[ファイル名] + ([行数]) + [半角コロン] + [半角空白] + [キーワードがヒットした行]

3. 「ファイルから検索」の結果から、元ファイルの該当箇所へジャンプするには、その行にマウスカーソルをあわせて、マウスの左ボタンをダブルクリックすると、元ファイルが開いて、目的の行にジャンプする。

◎「ファイルから検索」のオプション

「ファイルから検索」で指定可能なオプションは以下の通り。

- 大文字と小文字を区別する
検索のオプションと同じ。

- サブフォルダも検索する
指定したフォルダの中にフォルダが含まれる場合、その中に含まれる (更にその中…) のファイルを検索するかどうか指定する。

- 正規表現を使用する
検索のオプションと同じ。

- エスケープシーケンスを使用する
検索のオプションと同じ。

- 単語のみ検索する
検索のオプションと同じ。

- ファイル名のみ表示する
上記「「ファイルから検索」の結果表示」をファイル名だけにしたい場合に使用。

- エンコードドロップダウンリストボックス
検索の際に、検索するファイルの文字コードを指定したい場合、リストから目的の文字コードを選択する。

以下のオプションも、文字コードに関する指定である。通常は、[すべて自動検出] で検索し、もし思うとおりに検索できない場合に、他のオプションを指定すればよい。

- すべて自動検出
- HTML/XML の Charset を検出
- Unicode サイン (BOM) を検出
- UTF-8 を自動検出

◆ ファイルから置換

「ファイルから置換」は、指定したファイル・フォルダから「検索する文字列」を「置換する文字列」にまとめて変換する機能である。基本的な機能は「ファイルから検索」と同じなので、説明は省略する。一度「ファイルから置換」を実行すると元の状態にもとすことができないので、実行にはくれぐれも注意されたし。

◎「ファイルから置換」のオプション

「ファイルから置換」のオプションも、基本的には「ファイルから置換」と同じだが、この機能独自のオプションもある。

●変更したファイルを開いたままにする

通常「ファイルから置換」を実行した場合、置換後のファイルを強制的に保存した状態になる。しかし、このオプションを指定すると、置換後もファイルが強制保存されないので、置換結果を見てから保存することができる。ただし、これは、現在開いているファイルのみに適応される。

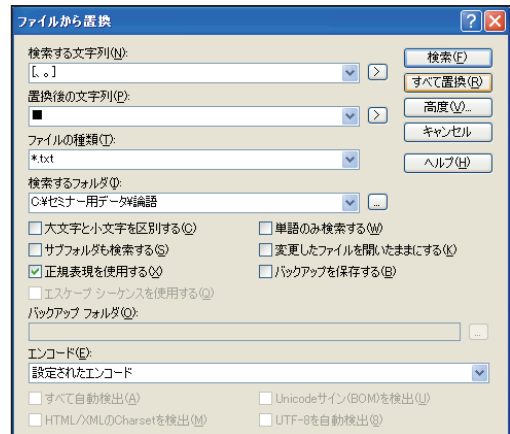
●バックアップを保存する

通常、「ファイルから置換」で置換されたテキストは強制保存され、それ以前のデータは削除される。このオプションを指定すると、「ファイルから置換」の実行の際に、強制保存前のファイルを「バックアップ」フォルダに保存してくれる。

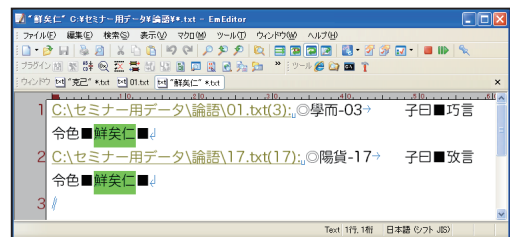
◆ おわりに

以上、EmEditor を使って検索と置換の手順を説明した。

検索と置換は、正規表現と組み合わせることで、より強力な効果を発揮する。はじめは正規表現の文法がさっぱりわからないかもしれないが、自分が普段使いそうなものから徐々におぼえておけばよい。使い慣れるうちに手放せなくなるはずだ。



「ファイルから置換」のダイアログボックス



上記「ファイルから置換」で置換された結果、フォルダ全体のファイルの当該文字列が一括して置換されている。

※「ファイルから検索」で確認。

注

- [1] EmEditor6 の最新バージョンは 6.0.2。詳細は、後述の「レビュー：EmEditor6」を参照されたし。
- [2] なぜこの手順が必要なのかといえば、EmEditor の検索が「カーソルから後に向かって検索を行い、ファイルの末尾まできたら終了（ファイルの先頭に向かって検索を行う場合には、この逆の動作をする）」という動作仕様で理由がある。そのため、カーソル位置によっては、検索結果の確認漏れが起きる可能性があり、それを防ぐために、あらかじめカーソルをファイルの先頭に持ってくるのである。
- [3] ワイルドカードとは、ファイル名の指定に使うメタ文字の一種。「*」が任意の複数の文字列。「?」が任意の一文字をそれぞれ意味する。

お勧めユーティリティ テキスト検索、一括処理ツール

小川 利康（おがわ としやす）

◇ 便利な小物

ここではテキスト検索、文字コンバートに役立つオンライン・ソフトをご紹介します。特定の目的のために使われるソフトウェアのことをユーティリティ・ソフトと呼ばれるが、そうした小物があるかないかで作業効率は格段に違ってくる。この小文が読者諸兄の研究環境改善に役立てば幸甚。

◇ テキスト検索

毎日のようにパソコンを使っていれば、メール、研究資料、論文がハードディスクに蓄積されてゆくのは当然のことである。たとえ20年前のメモであっても、まめにデータのバックアップを取っていれば紛失することはない。だが、20年前に作ったレジユメが「松 ver.5」（管理工学研究所）で作ったのか、「Mifes98」（MEGASOFT）で作ったテキストファイルだったのかなんて、覚えていないわけがない。ましてや正確なファイル名で検索なんていうのも無理な話だ。Windowsにも標準でファイル名検索の機能は付いている。だが、こんな場合にはまず役に立たない。

さりとて古いファイルをいちいち開いて確かめては日が暮れてしまう。せっかく上海図書館でコピーを取り、せっせと入力した中国語の資料なのだから、同じ事を繰り返したくない。そうとなれば日暮れどころか、深夜の2時過ぎであっても、あちこちのフォルダを開いて確かめないわ

けにはゆかなくなる。そんな苦い経験をもつ方にお勧めしたいのがテキスト検索ツールだ。

もちろんテキスト検索ツールは迷子のファイルを探すためにだけあるわけではない。特定の単語を指定して、どのような文脈で用いられているのかを調べるためにも用いられる。いわゆるKWIC（Key Word In Context）とよばれる技術で、本来言語学でコーパス研究を行う手法だったが、現在ではウェブ検索をはじめとして、かなり広い範囲で用いられている。秋山陽一郎「手軽にできる情報分析」（本誌6号）でGREP, Ngramとともに分かりやすく紹介されているように、使い方次第で色々な成果が期待できるツールである。

◎ Google デスクトップ 3

最初に紹介するGoogle デスクトップ3（Beta版、以下ではデスクトップ3）^[1]はウェブ検索のGoogleで用いられてきた技術をパソコン上のローカルデータに適用しようとするものだ。検索式も通常のGoogleとほとんど変わらず、検索結果がネットワーク上にあるのか、ローカルファイルなのかという違いに過ぎない。検索インデックスもUnicodeで作成されるので、日本語だけでなく、中国語^[2]のファイルもすべて検索できる。キーワードで関連ファイルを一网打尽に出来る快感に慣れてしまうと、フォルダ、ファイル名だけで管理する方法には戻れなくなること請け合いだ。

とはいえ個人情報が多数含まれる検索インデックスが外部へ流出する危険性を指摘する向きもあり、ネットワークの普及によって良かれ悪しかれ「あちら側」（ネット）と「こちら側」（ローカル）

の境目が無くなってきたことを示す例ともいえる [3]。利便性と安全性のバランスを考えて利用していただきたい。

●設定・インデックス作成

デスクトップ3をインストールすると、自動的に検索インデックス作成が始まる。最初のインデックス作成はデータがどれだけあるかによって異なるものの、数時間を要する。従って、すぐにファイル検索をしたいと思ってインストールしても直ちに使えるわけではない。

その後は毎回ウィンドウズが起動すると同時に常駐し、検索インデックスにデータを追加する。インデックス作成はコンピューターが30秒以上アイドル状態になっているときのみ行われるので、ユーザ側はほとんどストレスを感じないはずだ。

検索の対象とするファイル、フォルダは下の図のようにユーザが指定できる [4]。テキストファイルのほか、PDF、MS Office 系のアプリケーション用ファイル、メール、ウェブ履歴などである。MS Outlook を使用していれば予定表なども検索対象に含めることができる。圧縮ファイルでも zip ファイルなら内容を検索できる。タイムライン表示を選択すれば、検索インデックスに登録された日付順にキャッシュが表示され、書きかけの原稿の変更履歴を追いかけすることもできる。

さらに現在のバージョンでは Gmail アカウントを共有するマシン間で、以下のデータの検索インデックスも共有できるようになった。

- ①Gmail のデータ
- ②ウェブの履歴
- ③MS Word, Excel, PowerPoint のファイル
- ④マイドキュメント内の PDF ファイル, テキストファイル

Gmail はご存じの通り全てのメールを蓄積し、どこからでも自由にメールを読み書きできるシステムである [5]。デスクトップ3ではローカルファイルだけでなく、オンライン上のデータ、ネットワークでつながるマシンのファイルも検索対象と



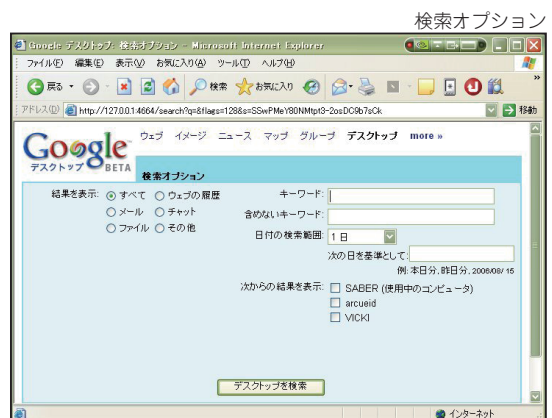
キーワード検索結果例

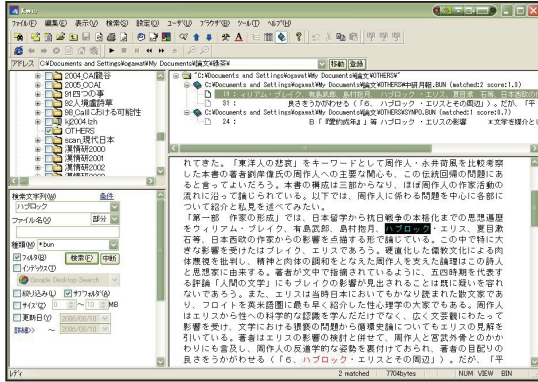
するようになった。

当然対象となるマシンすべてにデスクトップ3がインストールされているのが大前提だが、自宅のデスクトップと出先のノートパソコンとの間で検索データが共有できれば、「昨日作ったメモ、どこにやったっけ？」という時でも、キーワードで検索をかければ直ちに所在が分かり、キャッシュデータで必要な箇所を確認する事もでき、なかなか便利だ。この場合、通常の Google の検索式に加えて、対象とする端末の名前を追加してやればよい。デスクトップ3の検索オプションに端末の名前が表示される [6] のので、その中から選べば簡単だ。

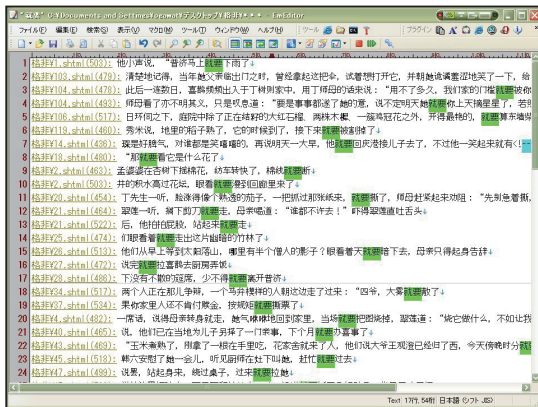
検索式としては、以下のようになる。

検索条件：「就要」を含むファイルを SABER と





KWIC Finder の検索結果



EmEditor のファイル検索結果

というマインドキュメントから探す。

就要 machine:SABER

逆に排除するのならば、「- (マイナス)」をつければよい。例えば、arcueid というマシン以外から「虽然」を含むファイルを検索するなら、

虽然 -machine:arcueid

となる。Google の検索式としては常識だろう。

ファイル探しのためのデスクトップという利用法が先になってしまったが、特定の語彙について複数のファイルを横断的に検索することもできる。検索方法は通常の Google と同様に色々な検索式で検索できるが、厳密な用例比較はできない。類似した語彙が含まれる文章をファイル単位で提示

してくれるが、厳密に該当箇所を前後の文脈のなかで把握するには KWIC による検索を利用する方がよいだろう。

◎ KWIC Finder

KWIC による検索をするソフトウェアは幾つかあるが、KWIC Finder がよく知られている^[7]。

起動したら、左上のウィンドウから検索対象となるフォルダを選択し、左下のウィンドウに検索文字列を入力する。この際、拡張子を指定して、検索対象を限定することもできる。シェアウェアライセンスを購入すれば、MS Office 系のアプリケーションだけでなく、Outlook Express など主要メーカーのフォーマット及び Adobe PDF、一太郎、Lotus123 新松、松 (Ver.5 ~) などのファイルの内容をすべて検索対象に含めることができる。

通常は検索ボタンを押してから、検索が始まるので、対象となるファイルが多いとかなり待たされる。だが、同時に Google デスクトップもインストールしてあると、その検索インデックスを利用するので、ほとんど瞬時に検索結果が表示されるようになった。ただし、この場合、検索対象になるのは Google がカバーしているファイルフォーマットだけとなる。検索結果は右側のウィンドウに表示され、該当文中のどこに検索語があるか赤字文字で表示される。事前に検索インデックスの作成なしでも使えて、主要アプリケーションのほとんどを網羅しているので、データ管理に相当な威力を発揮するだろう。

ただし、検索できるのは JIS 漢字のみだ。残念なことに中国語の解析には使えない。その場合は、EmEditor のファイル検索を利用すると良い。検索には若干正規表現の知識が必要ながら Google デスクトップより詳しい結果が得られる^[8]。

もし Google デスクトップでの検索では飽き足りないが、正規表現による検索は無理だという場合は、KWIC を利用した専用解析ソフト「百歩穿楊 II」(日中産業通信)^[9]を購入すると良い。商用ソフトだけに検索条件を比較的平易に設定できる。

◆ 一括変換処理ツール

単純作業も大量になれば、大変な手間になる。出来れば機械に任せて一括処理したいものである。ここでは一括処理するためのツールを紹介する。

◎ 文字コード変換

Unicode の普及に伴い、今まで GB で保存していたファイルを UTF-8 へ変換したり、また逆にメールに添付するため、JIS, Big5, GB へ変換し直したりする機会がむしろ増えているようだ。だが、GB, Big5 への変換までサポートしているツールは少ない^[10]。ここでは二つ紹介しておこう。

● 中国語コンバーター

針谷壯一氏^[11]による中国語の各種文字コードの変換を行うソフトとして有名。GB, Big5 だけでなく、Chinese Writer をはじめとする中国語入力ソフトメーカーの独自文字コードにも対応しており、中国語に関しては、この一本で足りる便利なもの。コード変換だけでなく、簡体字、繁体字の相互変換のためのソフトとしても活用できる。

通常なら処理するファイルが一つ二つであろうから、ダイアログ版で利用すればよいが、一括処理する場合は DOS コマンドラインから一括処理しよう。これはプログラムインストール時にスタートメニューに「中国語コンバーター用コマンドプロンプト」が登録されるので、そのメニューから簡単に利用できる。一括処理のポイントはワイルドカードの利用だ。対象ファイルが多いときはいちいちファイル名を指定していたらきりがないので、処理するファイルを一つのフォルダにまとめて置いて、全ファイルを対象として指定する。その際に、ファイル全てを指定できるのが、ワイルドカードだ。これを使えば簡単に一括変換ができる。コマンドラインから、次の順序で記述する。それぞれ項目間は半角スペースを空けること。

- ①プログラム名
- ②オプション 1

- (オリジナルファイルの文字コード)
- ③オプション 2
(変換後のファイルの文字コード)、
- ④変換したいファイル名
(ワイルドカードを使い、*.rtf と指定すれば、rtf ファイルすべてという意味)。
- ⑤変換後のファイル名
(拡張子のみ。rtf, txt, bak など)

たとえば、Chinese Writer の独自コードで書かれた RTF (リッチテキストファイル) から、GB の RTF へ変換したいなら、

```
chnconv -cCW -cGB *.rtf doc
```

という風に指定する。新しいファイルは拡張子を doc として作成してくれる。また、変換後の拡張子として bak を指定しておけば、オリジナルファイルの方を bak という拡張子で保存し、新しいファイルを rtf で保存してくれる。

一括処理で注意しなければならないのは、処理量が多いだけに全てを目視確認しながらできるわけではないので、必ず少数のファイルで試してから行うこと、後でミスに気づいたときに備えてオリジナルファイルは必ず残すことを忘れないようにしましょう。たいがいのミスはオリジナルファイルが残っていれば回復可能だが、これがないと痛い目に遭うことになる。

このほか類似したソフトとしては、北原基彦氏による KCTRANS (コマンドプロンプト専用)^[12]でも JIS, GB, Big5 相互の変換ができる。DOS での一括処理をするならばこちらでも便利だ。

● RTF コンバーター

同じ針谷壯一氏の作になるコンバーターだが、こちらは中国語互換のコンバートだけではカバーしきれないコード変換を行ってくれるソフト。たとえば日本語 JIS で作った中国語資料を Unicode へ変換したり、中国語 GB へ変換してくれる。逆ももちろんカバーされている。その際、通常の Unicode 変換ソフトでは再見→再見と変換してく

れず、再?と文字化けするが、このコンバーターは別途異体字テーブルを持っているので、かなりの部分を変換してくれる優秀なソフトだ。一括処理用のコマンドラインも用意されており、中国語コンバーターとほぼ同じ使い勝手だ。たとえば

```
rtfconv -cRTF -cU *.rtf txt
```

rtfconv というプログラムで、RTF ファイルから Unicode のテキストファイルへ出力するという意味だ。こちらを使えば日中間のみならず、韓国語も含めた RTF、テキストファイル間のコンバートが可能で、ほとんどの用途はこれで足りるだろう。このような優秀なソフトウェアを作成するには膨大な労力と時間を要するにもかかわらず、フリーウェアとして公開されている針谷壮一氏に心から感謝したい。

注

- [1] 2006年3月末、デスクトップ3が公開された。
<http://desktop.google.co.jp/ja/index.html> から入手できる。
追記：同年8月17日にデスクトップ4（日本語版）がリリースされたが、本稿はバージョン4のリリース直前に執筆したものであり、内容はデスクトップ3にもとづいたものである（公開先 URL は同じ）。ただし、本稿で紹介した機能については、デスクトップ4でも全て利用可能である。また、新たに「ガジェット」と呼ばれるミニアプリケーションをデスクトップに追加できるようになった。
- [2] Chinese Writer やオムロン cWnn などのメーカー固有の中国語フォントは当然ながら対象外となる。
- [3] Gmail はメールの内容に応じて広告が表示されることがプライバシー侵害だと問題視されたが、デスクトップ3の検索用インデックスはローカル上に保存される仕様なので個人情報が流出することはない。だが、オプションで複数のマシンで検索データを共有する設定を利用すれば、Google のサーバーを経由して検索インデックス（暗号化はされている）が交換されるので、重要な情報はインデックスに含めるべきでないだろう。
- [4] 一太郎、松などのワープロソフトのファイルも検索対象に含めたい場合は、GoogleDesktopSearch ★ Xdoc というプラグインを追加すると良い。Soft Farm.net <http://softfarm.net/ather/xdocgoogle.html> から入手できる。ただし、Unicode には対応していない。
- [5] gmail は従来までは招待状を必要とするため、利用者が原則として限定されていたが、2006年8月23日より一般公開され、招待状なしでアカウントを申請できるようになった。<https://www.google.com/accounts/ManageAccount> から申し込める。
- [6] 端末につける名前は任意だが、標準設定では「コントロールパネル」－「システム」－「コンピュータ名」で確認できる。複数の端末を使う場合、重ならないように注意しよう。
- [7] http://www31.ocn.ne.jp/~h_ishida/KWIC.html テキスト検索だけならフリーウェアだが、Word などバイナリファイルも検索したい場合は有料ライセンスが必要。
- [8] 図の結果は html ファイルをダウンロード後、句点ごとに改行コードを付加した。
- [9] 日中産業通信 <http://www.wordlab21.com/baibu2/BC.htm>
- [10] MS Office で利用できる Proofing Tools にも繁体字、簡体字の変換機能がついており、付属の辞書を利用して変換するので精度が高い。大量のファイルを一括処理するのであれば、こちらも是非勧めたい。
- [11] <http://www5b.biglobe.ne.jp/~harigaya/> から入手可能。
- [12] KCTRANS 網站 <http://www2u.biglobe.ne.jp/~kitabo/>

特集2

人文学的な 情報処理教育とは

本誌第三号で「人文学情報処理教育」についての特集を組んだが、そのときから数年を経て、再び同様の特集を組むことになった。この間パソコンやインターネット環境が急速に普及し、「とりあえず Web・メール・ワープロ」ができる学生は増えた。しかし、現状ではそこで止まってしまい、それぞれの学問分野に特化した IT 技術や方法論を学ぼうとする学生が減っているような気もする。教える側も「とりあえずできるから」という意識からか、人文学的な情報処理教育はむしろ退化しているようにも思える。

では、これらの問題を抱えた人文学的な情報処理教育はどこへ進むべきか。今回お寄せいただいた諸論叢は、「人文情報」に対する理系・文系の意識の違いへの言及（當山論文）・両者の垣根を克服する試みの記録（師論文）・「ワープロができる」の次にどう踏み出すべきか（山崎論文）をそれぞれ主題としたものであり、人文学的な情報処理教育の抱える問題と目指すべき先を示している佳品といえよう。是非御一読いただきたい。

Contents

なぜ理系と文系の議論はすれ違うのか	當山日出夫	66
授業報告：中国語テキストの構造化と デジタルドキュメント作成を同時に訓練する試み		
「外国語」と「情報リテラシー」を同じ授業で	山崎 直樹	75
情報歴史学のこれから		
花園大学・情報歴史学コースの4年間をふり返って	師 茂樹	84
BBS より：人文学的な情報教育	本誌編集局	91

なぜ理系と文系の議論はすれ違うのか

當山 日出夫（とうやま ひでお）

■ 1.はじめに

本稿においては、次の3点について、私見を述べることにする（筆者自身が「文系」にいるので、その視点からということになるが）。

なお、本稿においては、基本的に「文系」「理系」の用語で統一する。「文系」としては、主として、文学・歴史・哲学などの、オーソドックスな文学部の研究領域を想定しておく。一方、「理系」としては、物理学・化学などの、理学部・工学部に属する研究分野を考えておく。このなかには、コンピュータの研究もふくむ。おそらく、これが、現代社会における、一般的な「文系」「理系」の認識であろう。とりあえず、このような前提で、考え始めることにする。しかし、このような単純な分類意識それ自体に問題があることは、最後に指摘したい。

■ (1) 「文系」と「理系」との立場の対立

一般には、いわゆる「文系」と「理系」では、なかなか意思疎通が容易ではないとされる。特にコンピュータの利用をめぐって、文系・理系の研究者が接触する機会が増えてきている（その端的な事例が、文字コードをめぐる問題である）。しかし、そこでは往々にして話がすれちがう、あるいは、立場の対立ということが見られる。その現状について、考えてみたい。本稿では、そのため、私事にわたることになるが、筆者の、研究の履

歴・コンピュータ利用の経緯、また、大学生にコンピュータを教える経験などについて、語らせていただきたい。

■ (2) 本稿の道筋

では、なぜ、このような事態を生じるのが、その原因について考察する。特に、人文学の研究者の側から考えてみて、伝統的には、学問の教授法のなかにこそ、その学問の本質的部分がある、ということについて考えをめぐらしたい。

■ (3) 相互理解・対話の提言

だが、現在、その伝統的な教授法が崩壊の危機に瀕している。現代日本における大学の文学部は、旧来の学科や専攻の組織が再編されようとしている。その一方、コンピュータとインターネットの利用は、伝統的な学問の方法論の再考をせまるものともなっている。このような状況をふまえ、今後の課題として、単純な文理融合ではなく、お互いの専門性を認識したうえでの、相互理解・対話であることを、提言したい。

■ 2.筆者とコンピュータ

本稿執筆の時点は、すでに、「Web2.0」といわれ、また、ARGの岡本真氏による『これからホームページをつくる研究者のために』が刊行されている。このような流れのなかにあって、さらにさかのぼって、筆者自身の、コンピュータとのかか

わりについて、概略的にふりかえって述べてみたいと思う。今あることの問題の根源について、過去の歴史的経緯をさぐってみることも、決して無意味なことではないと考える。

まず、私自身の立場であるが、純然たる文系の研究者である。大学～大学院は、文学部の国文学専攻に属していた。専門とする領域は、日本文学というよりは、日本語史・訓点語学である。その後、研究のためにコンピュータをつかうようになってから、その文字の問題と直面することになり、「情報文字論」を提言して、今日にいたっている。そして、現在の研究テーマは、日本語史・訓点語学からは少しはなれて、コンピュータと文字、あるいは、文字研究にコンピュータがどのように使えるか、さらには人文学や教育におけるコンピュータ利用の諸問題などについて、考えている状態である。

では、私が大学生であったころはどのようなふうであったろうか（1975年、慶應義塾大学文学部入学）。大学の学部・大学院の時代、まだ、コンピュータといえば、大型計算機しかない時代であった。大学でのコンピュータの授業とえば、COBOLかFORTRANであり、パンチカードを使用していた時代である。三田のキャンパスで、コンピュータを利用していたのは、経済学部や商学部の一部、あるいは、文学部の図書館情報学科の一部、というのが実際であったであろう。もちろん、私自身は、このでの授業は、一切関心が無かったし、履修しようなどと思ったこともない。いやそれ以前に、文学・歴史学・哲学というような勉強にコンピュータがつかえるという認識そのものが、なかった。総合的には、文学部の教育・研究（人文学）と、コンピュータは、無縁であったと言っても過言ではない。ただし、日本語学の分野については、計量国語学会があり、また、国立国語研究所などでコンピュータをつかった研究も行われていた。しかし、そうはいつても、日本語学全体から見れば、ごく一部のことであったといえる。

大学院の学生のとき、研究課題として、『白氏文集』（唐白居易撰、卷三・四、故神田喜一郎氏旧蔵、現在は京都国立博物館蔵、院政期書写）の、

漢字索引・訓読文索引を作成することになった。太田次男先生（慶應義塾大学）の漢籍としての翻刻文と、小林芳規先生（広島大学）の訓読文の双方について、それぞれに、索引を作成する作業であった。これは、最終的には、『神田本白氏文集の研究』（太田次男・小林芳規、勉誠社、1982年）に、収録される形で結実した。このとき、筆者は、約3万枚ほどの紙の語彙カードを、手作業で作成して並べ替えて原稿用紙に書く、という経験をした。これには、ほぼ1～2年の期間を、ほぼこの仕事のためだけについやすことになった。この段階では、コンピュータを漢籍の漢字一字索引に利用しよう、というよりも、それ以前に、索引作成にコンピュータが利用可能である、という発想そのものが存在しなかった。索引は手作業のカードによるもの、ことばの用例の検索は、ひたすら本を読んで自分の目でさがすもの、というのがいわば常識であった時代である。

このことは、単に筆者個人の経験にとどまるものではなく、非常に重要な意味があると考えている。コンピュータを人文学系の研究者がつかいはじめたとき、その目的の大部分は、その検索機能（grepやsedによる）に着目したからである。つまり、索引作成の代替としての、テキスト検索である。書籍・史料があつて、本を目で読むことと、コンピュータのデータとして検索すること、この間にある乖離を、研究者の意識としてどう実感するかが、おおきな課題として残り今日にいたっている。そして、筆者自身においても、コンピュータを使い始めた最初のきっかけは、漢字索引の作成なのである。

大学院を終えて後、しばらくして、それまでの大型計算機にかわつて、個人で使用するパーソナルコンピュータが、登場した。登場した直後は、やや高価な電子工作の機材という印象であり、また、ゲーム機にすぎなかったともいえる。しかし、やがて、本格的にビジネスの業務用に使えらる機種があいついで登場するようになった。富士通やNECなどのメーカーによるものであるが、本命というべきは、やはりNECのPC-9800シリーズである。

パソコンが登場して実用的に使用できるようになったとき、筆者が、最初買ったのは、NEC製のPC-9801M2(ごく初期の9800シリーズのひとつ、CPUは8086、5インチ2HDのFDドライブが2スロット搭載されていた)であり、これをつかって、本格的に、漢字文献のコンピュータ処理を試みるようになった。先の『神田本白氏文集』の索引作成のときのような、まさに苦行というべき状況をなんとかしたいと思ったのである。この時点で目指していたのは、『白氏文集』全71巻の総合一字索引であるが、残念ながら、これは実現するにいたっていない。

だが、その準備段階として、パソコンで本当に漢字索引が作れるのか、試してみなければならない。そこで生まれたのが、『和漢朗詠集漢字索引』(當日山日出夫、勉誠社、1988年)をはじめとする、『千載佳句漢字索引』『新撰朗詠集漢字索引』など、一連の平安朝につくられた漢詩文秀句集の漢字一字索引である。これらの文献は、和漢比較文学研究の立場からも、『白氏文集』と密接な関連があり、分量的にも手頃である。

これらの仕事を、今からふりかえってみれば、特徴・課題・問題点は次のようになる。

■ (1) 特徴

見た目の体裁においては、『文選索引』にならったこと。伝統的な東洋学のスタイルを継承したことである。手作業による『神田本白氏文集』の場合も同様である。だが、実は、この私の仕事以前にも、文献の語彙索引をコンピュータで作ることが、言語研究の分野において、なかったわけではないが、しかし、それらは、KWIC形式が主流であった。

コンピュータの側の発想からすれば、索引の形式としては、KWICが通常であるのかもしれない。しかし、これでは、旧来の東洋学・訓点語学における索引のスタイルと、相容れない。索引の基本的な目的は、文字やことばの検索である。だが、それを用例つきで表示する場合、どれだけの単位を、その文字やことばにふさわしい用例として認定するか、これもまた、研究者に問われるべ

き見識なのである。単に、前後の一定数単位の文字列が表示されていけばよいというものではないのである。おそらく、このあたりの、感覚の違いが、本稿のテーマである、文系・理系の発想の違いにつながっていくのかとも、思う。

■ (2) 課題

大型計算機ではなく、まったく個人で使用するパーソナルコンピュータですべての仕事をしたこと。版下となるプリントアウトも、当時、登場したばかりのレーザープリンタ(それでも、1文字が32ドット)をつかって、自分の書齋で作成した。

つまり、大学でコンピュータにかんする専門教育をうけていないのもであっても、工夫と努力によって、かつては大型計算機でしかできなかったことが、可能になったことである。しかも、この場合、自身が人文学系研究者であるから、自分の研究目的のために、その研究領域の方式(漢字一字索引であれば文選索引方式)にあわせて、自由につかうことができる。

■ (3) 問題点

また、本稿のテーマからははずれるが、付随的に、この作業の過程で、コンピュータで使用する文字(JIS漢字)に直面することになったことがある。コンピュータで使用する文字を「JIS漢字」と名付けて、本格的に、学問的な研究対象としたのは、筆者が、その最初期のひとりであろう。(ちなみに、筆者が、最初に手にした規格票は、「JIS X 0208」ではなく、その以前の「JIS C 6226」の方である。)

この延長には、いわゆるJIS漢字の問題がある。当面の課題としては、Unicodeや次期Windows Vistaに搭載予定の「JIS X 0213:2004」をめぐる問題がある。これらの議論をみても、根底に、文系・理系の発想のずれ違いがあるとは、個人的には感じている。これが那邊に由来するのかわ、明らかにすることは難しいが、筆者の経験的な印象からいえば、「文字(文献)を読むという学術的訓練」を、どのように学んだかに大きく依存すると思う。これは、実は、文系・理系だ

けの問題だけではなく、文系内部でも、筆者のように日本語学における文字論・文字史を専門領域としている人間と、文学や歴史学を研究している研究者との間でも、微妙な違いがあるという印象をもっている。

■ 3. コンピュータ利用についての研究会など

上述した、筆者が、パソコンをつかいはじめた時期は、まさに、パソコンが、社会全体において、実用的に個人の利用につかえるものとして普及し始めたときであり、人文学の研究領域も、そのなかにあつたのである。この時期、各研究分野において、萌芽的な仕事がいくつかこなされている。また、これ以降、人文学研究へのコンピュータ利用をテーマに、幾多の研究会が組織されたりもして、今日にいたっている。例えば、次のようなものである。

- ①テキストデータベース研究会（JACH）
- ②情報処理国語国文学研究会（JALLC）
- ③統計数理研究所における研究会（村上征勝）
- ④東洋学へのコンピュータ利用研究会（星野 總、京都大学大型計算機センター）
- ⑤情報処理学会「人文科学とコンピュータ研究会」
- ⑥アート・ドキュメンテーション研究会（現在は、学会）

これら以外にも、数多くの研究会やシンポジウムなどが開催されてきている。このあたりの経緯については、雑誌『人文学と情報処理』（勉誠社）を見ることによって、かなり捕捉できるはずである。ここで実情を述べておけば、この雑誌の創刊準備から、最初の数号までは、筆者が実質的に企画・編集に関与していた。本稿の冒頭であげた、岡本真氏の『これからホームページをつくる研究者のために』が、インターネットの普及以降の状況を対象にしていることを考えれば、『人文学と情報処理』は、今となっては、パソコンの普及期

からインターネット前夜までの時期の、人文学系研究者の動きを知る貴重な資料である。

これら研究会のうち、④⑤⑥は、今も継続しているし、また、本誌の発行主体である漢字文献情報処理研究会も、これらの流れのなかで生まれた研究会のひとつと位置づけることができよう。

筆者の見るところ、これらの研究会は、次の二つの側面から生じており、異なる性格のものが同居している（あえて、誇張した立場から述べる）。

■ (1) 情報交換の必要性

コンピュータを使いはじめた文系の研究者たちにとって、お互いの情報交換の必要性である。今、コンピュータで何ができるのか、利用にはどんな課題があるのか、誰がどのようなことを試みているのか、発表し合い、意見を交換しあう場である。これは、既存の、各専門領域の学会などでは、この種のことが、なかなか出来ないという事情もある。たとえ研究業績にならないとしても、研究者同士の間での研究発表の場の確保は、是非とも必要だったのである。（この場合、人文学の方が主体であって、コンピュータはそのための道具である。）

■ (2) 理系にとっての人文学へのコンピュータ利用

理系（コンピュータの専門）の研究者にとっては、人文学へのコンピュータ利用というのが、新しい研究領域・研究対象として見えてきた、ということである。コンピュータというものが汎用性を目指す以上、そのデータ処理の対象となるものは、いわゆる理系・自然科学の研究対象に限定されることはない。例えば、画像処理技術の応用としては、美術作品や考古資料のデータベース化や復元シミュレーション、あるいは、古文書OCRの開発などがある。（この場合、コンピュータの利用可能性を探るのが中心であって、人文学は、その資料を提供してくれるもの、ということになる。）

4. 大学でコンピュータを教える

以上は、研究者としてのコンピュータとのかかわりであったが、教育の方面についても、述べておく必要がある。

たまたま各種の研究会で、理系（コンピュータ）の研究者と、同席することが多かったせいもあってか、大学教育（文学部）での情報処教育にもかかわることになった。

- ①花園大学での「国文学情報処理」
- ②立命館大学での「情報処理入門」（当初は「コンピュータ入門」）および「人文科学のための情報処理・応用」

である（非常勤）。いずれも、文学部での情報処理教育を本格化させようとした最初期から、筆者が担当することになって現在までつづいている（それぞれ10年ちかくになる）。

これらの授業を担当するようになったとき、まず感じたのは困惑である。いったい何をどのように教えればいいのか、である。

大学でも、理学部・工学部などであれば、当然ながら、授業内容に疑問が生じることはないであろう。文学部であっても、かつての筆者の学生時代とは違う。COBOLやFORTRANなど教える時代ではない。いや、そもそも、プログラミング自体が必須の知識・技能であるかさえも疑わしい。「コンピュータを使う＝プログラミング」という図式は、文学部での教育・研究には、自明のこととしては通用しないのである、と考えざるをえなかった。

花園大学の国文学情報処理の授業であれば、まずは、国文学（日本文学）の授業である、という前提で考えた。そうなると、授業で出来ることは、ほぼ限定されてしまう。

- ①文学作品のテキストを読むためにコンピュータを、どのように利用できるか。
- ②国文学（日本文学）の勉強で、レポート・

論文を書くためのコンピュータ利用として何が必要か。これは、端的にいつてしまえば、ワープロの操作法である。

現在では、かなり古典文学作品がデジタル化されており、研究領域によっては、必需品という状況にある（例えば『国歌大観』など）。しかし、授業を担当し始めた頃は、まだそうではなかった。ほとんど信頼に耐える古典作品のデータが存在しない時代であった。また、コンピュータを使うことの以前に、文学作品のテキストをきちんと読むトレーニングが出来ていなければならない。原テキストがろくに読めないのに、ただコンピュータで、ことばの検索ができて、教育的に無意味であることは、言うまでもなからう（この点では、現在でも、同様である）。

試行錯誤の結果、現在では、前期に②の基本操作の練習、後期は『百人一首』のデータベースの作成（Access）ということにしている。これについては、2005年に情報処理学会の人文科学とコンピュータ研究会で発表した（CH-66）。

立命館大学の場合どうか。ここは総合大学であるので、文学部内にも、文学・歴史・哲学などの伝統的な人文科学の分野もあれば、地理学（もはや、空間情報処理学とでもいうべきであろう）や、心理学（これは、サイエンスの一部というべきであろう）の専攻もふくんでいる。

ここも試行錯誤の結果、文書情報処理からライティング教育への橋渡し、という位置づけにしている。ワープロ（Word）をつかって、アカデミックな体裁の整った文書作成の実習としている。具体的には、タイトルや著者名の書き方、脚注のつけかた、参考文献リストの整理（Excel）と記載方法、図表の描き方（小数点揃えタブの使い方）、などである（もちろん、これは、筆者自身の意識における位置づけであって、大学のカリキュラム上の位置づけとは、異なる）。

これらのことは、情報処理教育とはいえないかもしれない。しかし、この種の基本的なことから確実にこなせるようになっておかないと、教育・研究にコンピュータを使うというレベルに到

達しないのである。今は、パソコンでかなりのことができる。いや、教育・研究の必需品であるともいえる。だが、様々な統計処理や画像データの処理ができて、インターネットから各種情報の収集ができて、その受け皿となる基礎的なアカデミックな文書作成技能を身につけていないと、学生としての勉強にならない。最終的なレポートや論文のかたちに作り上げていくことが、難しい。レポートや論文はワープロで書くのが当たり前になっていながら、コンピュータのなかにあるデータが、最終的なレポートや論文に結実せずに、それだけで空回りしてしまうのである。

例えば、学術論文の検索などは、今は、インターネットで簡単にできる。しかし、それを、実際にレポートや論文に参考文献リストとして記載するとき、どのようなルールにしたがうべきかを知らなければならぬ（これは、研究領域によっていちじるしく異なっている）。記載項目の順番や区切り記号の種類、書名を『 』に入れる入れないなど、実に様々な流儀が混在している。単に、インターネットの検索結果を、コピーして貼り付ければ（本当は、実際に自分自身で読んだうえで、書誌情報を記載すべきなのだが）、それでよいというわけではない。

さらに、文献リストの作成の発想にも、文系・理系の違いがある。一般に理系であれば、その論文を書くために直接参考にした文献を厳選してかかげる。一方、文系の場合は、文献リスト＝研究史、として、該当の研究課題についての過去の研究論文を網羅的にリストアップすることが要求される。

このようなことは、一般のパソコンやワープロの解説書からは抜け落ちていることであるし（一般には、ビジネス文書中心）、また、プログラミングを中心とした理系的な情報処理教育からも生まれてこない。最近になって、メディア・リテラシ教育というかたちで、各種のこころみがなされているが、これもまだ試行錯誤の段階であるという印象をもつ。

結局、文学部の学生にコンピュータを教えること、これひとつをとりあげてみても、そう簡単にはいかないの实情である、というのが、筆者の率

直な経験的な感想である。これは、単純に教えることの内容ではなく、学問的な知のありかたとその教え方にかかわる問題であると、筆者は考えている。パソコンで論文を書く時代になったということは、伝統的な学問の教授法のなかにパソコンが侵入してきたことを意味する。

■ 5. 文系・理系という問題

さて、本稿のテーマである「文系・理系」の問題であるが、上述の、筆者の個人的な経験から、「文系・理系」の対立が、顕著にあらわれるのは、次の二つの場面においてである。

- ①上記の研究会などでの研究発表のテーマや、質疑応答の場面。
- ②大学教育、特に、文学部での情報処理教育において、いったい何をどのように教えればよいのかを、考えるとき。

ただし、筆者個人の見解としては、必ずしも、文系・理系で議論がすれ違っている、とは、必ずしも思っているわけではない。しかし、一方で、そう簡単に、一緒になれるものでもないとは感じている。以下、コンピュータの利用を軸として、いわゆる「文系・理系」の対立のかかえる問題点について、論点を、先鋭的なかたちで整理してみたい。

■ 6. 文系では教授法にその学問の本質がある

さきに述べたように、文学部の学生にコンピュータを教えるのというのは、きわめて複雑な問題をはらんでいる。論点を整理するならば、

■(1) 基礎教養としてのコンピュータ利用技能教育

現代社会における必須の技能としてのコンピュータ操作法。これは、どのような学部であっ

ても、そう大きな違いがあるわけではない。

■ (2) 理系と文系との学問的すれ違い

各専攻分野において、コンピュータがどのように利用可能であり、また、どう使われているかという問題。また、コンピュータの利用によって、新たな研究分野が開拓されるかどうか、という問題(筆者のかかわる日本語学の分野であれば、コーパス言語学が、新たな研究領域として、新しく生まれてきている。古典籍をあつかう分野でも、画像データの利用は重要な課題となっている。方言学では、言語地図の作成・分析などにコンピュータは威力を発揮する)。

■ (3) 教授法の問題

人文学、特に、古典や歴史といった分野では、学問の教授法そのもののなかに、その学問の本質が内在している。昔の、「漢学」の世界であれば、幼少時の四書五経の素読からスタートする。筆者の経験の範囲でいえば、幸いなことに、大学から大学院にかけての専門の勉強は、師事した先生との個人教授に近いような状況であった。他の分野の人文学系研究者であっても、なんらかのかたちで、その学問分野独自の、あるいは、師事した先生個人独特の教授法の世界のなかで、研究者としての自己形成をなしてきているはずである。

このように考えたとき、文系・理系の対立の問題は、以下のレベルに整理できる。

(1)の基礎教養としてのコンピュータ利用技能教育のレベルにおいては、文系・理系の対立が、鮮明化することはない。もし、問題が生じるとすれば、必要とする技能の設定においてである。

(2)の場合、コンピュータは、文系研究者にとって、完全な「道具」である。最終的な目標は、その「道具」をつかって、どのような結果(学問的業績)を生み出せるか、である。このレベルにおいては、文系・理系の対立が生じる可能性がある。理系、特に、コンピュータの専門家にとっては、コンピュータ自体が研究の対象であり目的である。極端な場合、文系の研究テーマを借りてきて、そ

のためのプロトタイプを作成するだけでも、十分に研究者としての仕事(業績)になる。あえていえば、文系の研究課題が、ひとつの「材料」になってしまう。だが、文系の研究者にとっては、コンピュータを使った研究方法のプロトタイプをつくるだけでは、研究業績にならない。その学問分野において、検証に耐えるだけの結論を出さなければならない。

つまり、文系・理系において、つぎのような二つのすれ違いが生じる。

第一には、コンピュータを、研究の道具とみなすか、研究対象そのものとしてあつかうかの違いである。

第二には、文系の研究課題を、それ自身を目標としているか、コンピュータ利用の対象とみるかの違いである。

おそらく、一般に意識される、コンピュータ利用をめぐる文系・理系の対立は、このあたりにある。

(3)の場合、問題はさらに深刻になる。

東洋学についてみれば、前述したように、幼少時における四書五経の素読を基盤としてなりたってきた旧来のスタイルの学問と、四庫全書が、簡単にコンピュータで検索できる現代とでは、もはや同じとはいえない。同様のことは、仏教学における『大正新脩大藏経』のデジタル化、日本文学における各種古典文学作品のデジタル化、についてもいえる。

しかし、まったく別の学問になってしまったというわけではなく、筆者ぐらいの世代以上の研究者であれば、若いときの勉強は、紙の書籍をつかってであったはずであるし、その教育は、習った先生が、さらにその若いときに学んだ、昔の先生の教授法をうけついでのものであったはずである。筆者の学生時代、論文の書き方(さらには学問そのものが)学ぶものではない、師匠を見て盗むものである、と言われた(もちろん、この盗むは、剽窃ではなく、徒弟修行的な教育のありかたの謂である)。

また、人文系の研究では、過去の研究業績を尊

ぶ。そうなれば、コンピュータ以前からの研究も視野に入れなければならない。東洋学であれば、清朝考証学の世界を無視はできないし、日本文学でも、江戸時代の国学者たちの著作をひもとくことが必要になってくる。どうしても、過去の研究の方法や、学問の教授法を、ひきずらざるをえない。

このような状況のなかで、前述(2)の問題がからんでくると、事態は非常に錯綜する。過去の「コンピュータをつかわなかったがゆえに達成できた学問」の世界を、コンピュータを必須とする環境のなかで、継承していかなければならないのである。

筆者が、コンピュータを本格的に利用した仕事が『和漢朗詠集漢字索引』であることは、前述のごとくであるが、この本をつくったとき、筆者は、「あとがき」に次のようにしるした。「コンピュータをつかわなかったがゆえに達成できた学問」についてである。なかの引用は、昭和5年刊の岩波文庫版の『和漢朗詠集』（山田孝雄校訂）からである……「（筆者が）「朗詠索引」を手がけて以来、私の頭の片隅には、絶えず次の言葉があった。／『総じて、鎌倉時代を主として室町時代に及びての文芸に、これ（和漢朗詠集）の及ぼせる影響は頗る大なるものありと認めらる。（中略）この故に、鎌倉時代、室町時代の文芸を理會せむにはこの一部を殆ど暗誦せむまでに熟知せざるべからざるなり。／「朗詠索引」を作り終えたいま、まず思い浮かぶのは、「この一部を殆ど暗誦せむまでに熟知せざるべからざるなり」という山田孝雄の言葉である。／今後は、この「朗詠索引」が、山田孝雄の言葉の意味をふまえつつ、今日の学術資料のコンピュータによるデータベース化という研究環境の変化のなかで、批判的に利用されることを期待したい。」

索引を作った本人がいうのも妙であるが、現在、日本の中世文学を研究する若い研究者で、『和漢朗詠集』を暗記しようところみる人は皆無であろう。しかし、日本古典文学を本格的に研究するならば、どこかで必ず山田孝雄の仕事に接することになる。もちろん、山田孝雄は、コンピュータはおろか索引さえ無い時代に、偉大な業績を残し

た学者である。

我々は、過去の研究の業績をどう継承すべきなのであろうか。そのとき、コンピュータは、単なる道具にとどまるものであろうか。研究資料のデジタル化は、研究者の、研究方法のみならず、研究の基盤にある最も基礎的な部分（これは、まさに、教授法によって伝承されてきたものである）が、大きく変革してしまう可能性をはらんでいる。この点について、筆者は、楽観も悲観もできないでいる。よくわからないというのが、正直なところである。

だが、しかし、文系・理系の対立というテーマに即してみれば、この点が、最も深刻な論点になる。まず、人文系研究者にとっては、自らの研究者としてのアイデンティティにかかわる問題である。そのことに自覚的でなければならない。そして、そのうえで、コンピュータにかかわる理系の研究者に、このことを、きちんと伝えて理解してもらわなければならない。その努力をわれわれは、はたしてやってきたであろうか。

7. われわれは次に何をめざすべきか

文系・理系は、必ずしも対立しているというわけではないと、筆者は考えているが、少なくとも、コンピュータの利用にかんしては、対立とは言わないまでも、発想のすれ違いがあることは、上述したごとくである。このような状況のなかで、我々は、今後、どのような方向をめざすべきであろうか。ささやかな私見を述べて、本稿のくくりとしたい。

■ (1) 現実目標としての対話

現実的にめざすべきは、融合ではなく、対話である。現代のように極端に細分化した学問世界において、安易に融合をもとめるのは、中途半端な妥協の産物しか生み出さない危険性をはらんでいる。それよりも、自らの専門性に徹したうえで、さらにそれを確たるものとするための教養として、他分野の専門領域への対話の回路を確保すべくつ

とめるべきである。

■ (2) 「研究」と「教育」のあり方

学問には様々な専門領域があるが、それを、単純に、文系・理系の対立軸ではなく、他の要因にも配慮する必要がある。例えば、理学的な知と工学的な知、あるいは、科学（サイエンス）と人文学、などである。一般に、理系と言うと科学技術とむすびつけてしまうが、これを、科学（サイエンス）と技術（テクノロジー）をわけて考える視点も重要であろう。また、現代における学問のすべてが、科学であるとは言い切れないが、その考え方（学問的な思考方法）は、多くの人文学の分野でも共有し得るものであることを、理解すべきである。「文系・理系」という分類の発想それ自体、ある種の歴史的産物である。最終的には、「知」のあり方ということに帰するであろうが、当面の課題としては、総合的に、現代社会における「研究」と「教育」のあり方を考えるべきことになる。

■ (3) 今後の課題

現実的に、この問題における今後の課題は、次の二点になる。本誌の読者層の多くは、大学教育関係者であろうと推測されるので、ここに論点をしぼることにする。

第一に、研究者自身が、自らの研究の方法論について、いかに自覚的であり得るか。言い換えるならば、研究者として、謙虚でありつつも、貪欲でなければならない。

第二に、中等教育（中学・高校）から、高等教育（大学・大学院）にかけての、あまりにも硬直

した理系・文系の分類意識があるのが現実である。そのためには、大学受験（高校生）までの理系・文系意識の垣根をのりこえるものとして、大学教育（教養教育）の再検討がもとめられる。そのためには、研究者自身が自分の学問の方法論をどう考えているのか、その再検証が必須であろう。

■ 8. おわりに

ところが、現実には、伝統的な学科を単位とした大学文学部の存在そのものが崩壊しつつあることは、関係者に周知の事実である。タコツボを守ろうにも、強制的にたばねて壊されつつあるのが現状である。一方、「Web2.0」に象徴されるように、タコツボがからみあって溶解しつつあるのも、確かなことである。本稿で述べた「文系・理系」をこえて、新しい学問の未来が生み出されてくるか、あるいは、単に鶴的なのが跋扈するだけに終わるのか、今まさに、われわれは、その場面に当事者として立ち会っていることだけは確かである。

参考文献

- 『ウェブ進化論』、梅田望夫、筑摩書房、2006
- 『これからホームページをつくる研究者のために』、岡本真、築地書館、2006
- 『科学哲学の冒険』、戸田山和久、日本放送出版協会、2005
- 『コンピュータが子どもの心を変える』、ジェーン・ハーリー／西村辨作・山田詩津夫訳、大修館書店、1999（原著は、1998）
- 『系統樹思考の世界』、三中信宏、講談社、2006

授業報告：中国語テキストの 構造化とデジタルドキュメン ト作成を同時に訓練する試み

「外国語」と「情報リテラシー」を同じ授業で

山崎 直樹（やまざき なおき）

■ 1.はじめに

著者は、2006年7月から8月にかけて、関西大学外国語教育研究科で、「外国語運用実習——中国語文章表現法」なる集中講義を担当した（90分×15回、つまり、週1回1期分の授業に相当する分量の授業で、受講者は4人）。この文章は、その授業でどのような学習項目を取りあげ、どのような訓練をしたか、という報告である。

■ 2.この授業の目的

■ 2-1.文章の構造化とデジタルドキュメント

この授業の目的は次の2つである。

- (1)中国語のテキストを読み、内容を把握し、その内容を効果的に伝えるべく構造化して提示する……という一連のスキルを身につける。
- (2)デジタルドキュメント作成の初歩的な知識

と技術を身につける。

ここで言う「デジタルドキュメント」とは、ブレンテキスト、電子メール、HTML文書、ワープロ文書、などである。

要するに、この授業は、以下の2つの授業を同時に行なおうという試みである。

- (1)特定の言語（ここでは現代中国語）の運用能力を高める授業
- (2)「情報リテラシー」「情報処理技術」と呼ばれている知識やスキルを習得する授業

■ 2-2.イマージョンプログラムではない

前節で述べた2つの目的は、この授業において、密接に関連している。ある言語の運用能力を高めるのだけが目的なら、授業そのものを習得目標の言語で行えばよい。たとえば、家庭科の調理実習をすべて中国語で行う、などがその例である。このような授業はイマージョンプログラム^[1]のバリエーションとしてよく行われている。しかし、

本授業の趣旨はこれとは異なる。

■ 2-3. 知的活動のスキルとして

本授業で扱う情報処理技術は、すべて、(a)知的な構造をもったドキュメントを作成し、(b)そのドキュメントを、構造を効果的に反映したレイアウトで提示する、(c)そして、これらの作業をできるだけ効率よく行う、という目的のためのものである。

ある言語の習得目標のうちの1つに「知的活動を行う」ことを設定するのなら、上述の「知的な構造をもった……」というスキルは不可欠であろう。なお、この授業の受講生は博士前期課程の大学院生であるので、中国語を知的活動に使う必要があるのは言うまでもない。

■ 2-4. 文書の構造化に関して

「4. 授業の構成」を見ていただけるとわかるが、本授業は、HTML と CSS の基礎を学習した上で、Microsoft Word（以下、MS-Word と略）の使いかたを学ぶ順序になっている。

現実問題として、一般の大学院生が、HTML のタグをテキストエディタを使って手入力する必要があるとは考えにくい。CSS に関して同様である。しかし、この授業は、HTML や CSS そのものを学ぶことが目的なのではない。MS-Word の特定の機能——「文書を構造化し、その構造を提示する」ためのツール——を学ぶために、導入として、HTML および CSS を学んだ。

ワープロソフトにおいてはソフトウェア内部の見えないところで行われている作業を、プレーンテキスト上で、すべてが目に見える形で体験させたいと考えたわけである。MS-Word の機能の学習に先立ち、HTML および CSS を学んだ理由はそこにある。

■ 3. デジタルドキュメントに関連する受講生のリテラシー能力について

授業開始時に、まず、受講生の情報リテラシー

能力を知るために、次の質問をした。

■ 3-1. 質問項目

- ① パソコンを使ったことがありますか？
- ② 自分専用のパソコンをもちますか？
- ③ 携帯電話でなくパソコンで電子メールを読み書きしたことがありますか？
- ④ パソコンで電子メールを読み書きするとき、メール専用のソフトを使いますか、それとも、ウェブ上で行いますか？
- ⑤ (以下、使っているのが Windows だと仮定して)「メモ帳 (Notepad)」というソフトを使ったことがありますか？
- ⑥ 例えば、「学生番号は必ず半角英数字で入力すること」という指示があったばあい、すべきことが理解できますか？
- ⑦ 例えば、「文書/データは、テキストファイルでください」という依頼があったばあい、すべきことが理解できますか？
- ⑧ 「拡張子」という言葉を知っていますか？
- ⑨ 「拡張子」を表示する設定にしていますか？
- ⑩ 自分でウェブサイトを作ったことがありますか？
- ⑪ ウェブページのソースを見たことがありますか？
- ⑫ ワープロで文書を作ったことがありますか？
- ⑬ (以下、使っているワープロソフトが MS-Word だと仮定して)、「スタイルの適用」という操作をしたことがありますか？
- ⑭ メニューの「書式」から「スタイル」というサブメニューを開いたことがありますか？
- ⑮ 「見出しマップ」を表示させたことがありますか？
- ⑯ 「インデント」ということばを知っていますか？

■ 3-2. 結果

受講者は大学院生であるので、さすがに、①③⑫は問題なかった。②は、家族共有のパソコンを

使用している受講生もいた。④に関しては、メールで読み書きしていると明言する受講生もいるいっぽう、そのどちらなのかわからない、意識したことがない、という受講生もいた。⑥については、全員、問題なし。⑦は、やや頼りなげな受講生もいた。Windows を使っていないが、「メモ帳」を使ったことのない受講生もいた。⑧は、これを意図的に表示させている受講生もいたし、知らない受講生もいた。⑩⑪は、専用のソフトでウェブページを作ってみたことのある受講生もいたが、ソースを見たことのある受講生はいなかった。⑬－⑯は、すべて「否」であった。

4. 授業の構成

この授業は、以下のように構成した。

- 第1講：プレーンテキストとは
- 第2講：文字コードとは
- 第3講：電子メールは何でできているか～電子メールの作法
- 第4講：HTML の基礎
- 第5講：CSS の基礎
- 第6講：多言語混在 HTML 文書の作りかた
- 第7講：HTML をカスタマイズ
- 第8講：ワープロでの文書作成法①——レイアウトの基本
- 第9講：ワープロでの文書作成法②——テンプレートを使ったスタイルの適用

なお、この授業は、上述のとおり集中講義形式であるので、時間の区切りかたの設定は比較的的自由であった。その条件をフルに活用して柔軟に時間配分を行ったので、上記の「第0講」は、必ずしも、大学の通常授業の1回分に相当してはいない（第4, 5, 7, 9講に多く時間を割いた）。

以下、これらの授業で扱った学習項目とそれを定着させるために行った演習を、順に紹介していきたい。

■ 第1講：プレーンテキストとは

❖ 学習項目

まず、プレーンテキストとは何か、プレーンテキストでできることは何か／できないことは何かを講義した。

❖ 演習

UTF-8 でエンコーディングしたベタうちのテキストファイル（内容：『ケンタッキー漫画大賞』とでも言うべき、漫画コンテストの広告、中国語）を配付し、「メモ帳」^[2]でこれを編集させた。

与えたタスクは、「必要であれば、字句や標題、見出しなどを補い、内容がわかりやすいレイアウトにせよ」である。「メモ帳」でできること／できないこと、を通して、プレーンテキストでできること／できないことを把握してもらうのが目的である。また、この訓練は、プレーンテキストでできている電子メールの本文をわかりやすくレイアウトする訓練につながることを説明した。

■ 第2講：文字コードとは

❖ 学習項目

次の順で話を進めた。

- (1)文字コードとは何か
- (2)個別の言語に依存したローカル文字コード
- (3)プレーンテキストの中では、ローカル文字コードは共存しにくい
- (4)ユニコードという解決策

ついでに、雑談として、ユニコードでは字体の統一をしていないことを、「骨」などを例に説明した。あまり深い知識は必要ではないと判断し、実用に役立つ程度の説明にとどめた。

❖ 演習

文字コードのイメージを把握できるようなタスクを与えたかったが、著者の準備不足により、叶わなかった。よって、演習はなく、講義のみで第2講は終了した。

■ 第3講：電子メールは何でできているか～ 電子メールの作法

❖学習項目

まず、電子メールはプレーンテキストでできていることを説明した。ついでに、HTML メールがなぜ嫌がられるかについても、簡単に触れた。

次いで、メールのヘッダを見せ、どの部分が文字コードを指定しているかを説明し、主要な文字コードの名称を例示した。

さらに、内容面に踏みこんで、わかりにくい標題のメール（標題のないメールは問題外）は効率が悪いことを、筆者が実際に使用しているメーラの画面を見せて説明した。

最後に、プレーンテキストのメールマガジンを幾つかサンプルとして見せ、レイアウトなどに対していかに腐心しているかを示した。

❖演習

UTF-8 でエンコーディングしたベタうちの中国語のテキスト（内容は、ビジネスメールの一部で、納入された商品の一部が破損していたことに対するクレームとその返事）を配付し、次のタスクを与えた。「できるだけ内容をよく把握できる標題をつけよ。できれば、『結論』までを述べた標題を作成せよ」

このタスクを遂行しようとする、と、標題はどうぜん長くなる。そこで、実際に広く使われている

メーラのインターフェース画面を見せ、長い標題は末尾が表示されないことを示し、たとえ中国語の言語習慣にそぐわなくても、最重要な情報は標題の冒頭に置いたほうがよいことを説明した。

❖やらなかったこと

電子メールを作成し発信するさい、どの文字コードを選択しているのかを自覚し、主体的にこれを選択できる知識と技術は必要である。しかし、授業では、この方面の作業は省略した。

受講者が電子メールの読み書きに使用している環境がさまざまであり、統一的な説明がしにくいこと、とくに、環境によっては、“iso-2022-jp” “UTF-8” “GB2312” などという術語がユーザーインターフェースに現れないばあいもありうることを考慮したためである。

■ 第4講：HTMLの基礎

❖学習項目

まず、見本用として作った簡単なHTML文書のソースを見せ、HTML文書がテキストファイルであること、プレーンテキストにマークアップを施せばできあがることを理解させた。

次に、HTML文書を構成する基本的な要素と、それを表示するタグ（body, h1……h6, p, ol, ul, address）を導入した。

❖演習

ドキュメントタイプの宣言やヘッダなどを準備した雛形（HTMLのバージョンはxhtml1.0、コーディングの方式はUTF-8）を配付し、次に、別に用意したベタうちの中国語のプレーンテキスト（内容：日本の国民健康保険について、中国語で説明したもの）を渡し、内容を適切に階層化した上で、文書全体の標題と、中見出し、小見出しを自分で考案して付加するよう、要求した。なお、スタイルシートは、後々の作業で教材に使うことを想定したシンプルなもの、あらかじめヘッダ内に書いておいた。

この段階で、受講者が製作したHTML文書は左図のようなものである。

第4講で作成したHTML文書の一例

关于国民健康保险

在什么情况下进行国民健康保险的申报？

在进行了外国人登记的、有住处的市町村国民健康保险担当课办理加入手续。

1. 进行了外国人登记、具有一年以上在留资格并未加入其他公共医疗保险者，都必须加入国民健康保险。
2. 即使是入国时在留期间不满一年者，只要提出一年以上的在职证明书或在学证明书，也可以加入。但是，在留资格为“短期滞在”者不能加入。
3. 退出工作单位的健康保险时，必须立即加入国民健康保险。

注意事项

更换发行外国人登记证时，应办理重新制作保险证的手续。

❖やらなかったこと

画像を表示する、ハイパーリンクを設定する、などについては、まったく触れなかった。これらは、どの入門書でも触れていることだからである。同様に、ウェブサーバに場所を確保して、そこにファイルを転送して……という作業についても触れなかった。本を読めばわかることだからである。

❖感想

「段落を作る <p> と強制改行の
 のちがいがよくわからない」という質問を受講者の1人から受けた。我々が日本で受ける学校教育では、段落は、「改行して一字下げをするもの」というスタイルの問題として語られることが多く、「意味のまとまり」というブロックを作るものとして提示されることが少ない。その反映であろうと思われる。

■ 第5講：CSSの基礎

❖学習項目

第4講の演習で作成したHTML文書のソースを提示し、サンプルのスタイルシートを少し修正しては、ブラウザで読み込み直してレイアウトを確認する、という作業のデモを繰り返して行い、HTMLとCSSの関係を把握してもらった。

❖演習

第4講で制作した受講者個々人のHTML文書に対し、「階層化された構造がもっともよく表現できるレイアウトを目指す」という方針のもとに、雛形の中のスタイルシートを自分で変更させ、効果のほどを確認させた。

受講者がこの時点で製作したHTML文書は、右図のようなものである。

■ 第6講：多言語混在HTML文書の作りかた

❖学習項目

第4講で学習者に編集させたHTML文書は、文書全体の言語を、次のように指定してあった^[3]。

国民健康保険简介

[Q] 在什么情况下进行国民健康保险的申报？

[A] 在进行了外国人登记的、有住处的市町村国民健康保险担当课办理加入手续。

1. 进行了外国人登记、具有一年以上在留资格并未加入其他公共医疗保险者，都必须加入国民健康保险。
2. 即使是入国时在留期间不满一年者，只要提出一年以上的在职证明书或在学证明书，也可以加入。但是，在留资格为“短期滞在”者不能加入。
3. 退出工作单位的健康保险时，必须立即加入国民健康保险。

***注意事项**

- 更换发行外国人登记证时，应办理重新制作保险证的手续。

国民健康保险

在什么情况下进行国民健康保险的申报？

■ 在进行了外国人登记的、有住处的市町村国民健康保险担当课办理加入手续

- 1. 进行了外国人登记、具有一年以上在留资格并未加入其他公共医疗保险者，都必须加入国民健康保险。
- 2. 即使是入国时在留期间不满一年者，只要提出一年以上的在职证明书或在学证明书，也可以加入。但是，在留资格为“短期滞在”者不能加入。
- 3. 退出工作单位的健康保险时，必须立即加入国民健康保险。

【注意事项】

更换发行外国人登记证时，应办理重新制作保险证的手续。

第5講で作成したHTML文書の一例

```
<html xmlns="http://www.w3.org/1999/xhtml" lang="zh" xml:lang="zh" >
```

本講では、新たに、日本語の文章中に中国語を交える、あるいは、中国語の文章中に日本語を交える、という文書の作成のために必要な言語指定の方法を紹介した。具体的には、<*** lang="zh" xml:lang="zh">などの属性値である（***は任意のタグ名）。加えて、インライン要素を創造するためのタグ：も導入した。

❖演習

日本語の地の文の中に、中国語や韓国語が点在するテキスト文書を配付し、適切な言語指定を記入する作業を行った。同時に言語指定を行うことにより、使用されるフォントが切り替わるばあいもあることを確認した。

■ 第7講：HTML をカスタマイズ

❖学習項目

この授業では、CSS は、基本的な概念だけを導入するにとどめ、深入りする予定はなかった。しかし、後述するように、受講者に最もうけたのはCSS によるレイアウトデザインの演習で、デザインに凝りだした学習者からは、次第に、基本的な構成要素（h1～h6, p, ol, ul……）をセレクトにしたデザインのみでは自由にレイアウトができない旨の声々が挙がり始めた。

これは予期していないことであったが、その声を機会に、class 指定によるHTML のカスタマイズに話を移した。既習のタグに class 指定を加えること、ブロック要素を創造するための<div></div> およびインライン要素を創造するための にクラス指定を付加し、それらをセレクトにして、CSS のデザインをさらに細分化することを学んだ。

同時に、クラス指定を活用することにより、構造化文書の構造化文書たるゆえんがよりいっそう生かされること、そして、文書に機械可読のセマンティックスを付加しうること、などを講義した。ただ、その意義がどこまで伝わったか、はなはだ心もとないが。

受講者の知識水準によっては、ここで、XML の簡単な紹介をすることも考えていたが（構造化文書、機械可読のセマンティックス、という事項の本質を理解してもらうため）、混乱を避けるために、XML についての説明はスキップした。

❖演習

入れ子構造をなしている文章（内容は、言語起源論に関するもので、言語の起源を推定する数種類の仮説を例示する箇所が入れ子になっている）

を配付し、構造化を試みさせた。上位の階層にある段落と、下位の階層にある段落を、クラス属性によって区別させ、それぞれに異なったスタイルを指定する練習を行った。

❖やらなかったこと

“java script を用いて動的な……”という類いの話はしなかった。この授業の目的は、ウェブページの作りかたを1から指導する、というものではないし、「構造化された文書を作成するスキル」から少しく離れるような気がしたからである。

❖付記

話のついでに、HTML 文書でルビ振りを実現する規格が実現するかもしれない、という話題に触れた。この授業が行われたのは外国語教育研究科であり、受講生は、中国語教育に携わる可能性が高い——日本語話者むけに中国語の教材を作る可能性が高い——からである。

■ 第8講：ワープロでの文書作成法 (1) ——
レイアウトの基本

❖学習項目

文字フォントの変更、修飾、段落全体に対する書式の設定、とくにインデントやマージンなどの処理、その他、段落や見出しをブロック化するために必要な基本操作などを導入した。

❖演習

ベタうちの中国語のプレーンテキスト（内容は、言語の市場価値に関するもので、市場価値を決定する3要素を列挙する箇所が入れ子になっている）を配付し、まず、全体を段落に分けさせ、全体の標題や、ブロックごとの見出しをつけさせた。次に、そうして構造化された文書にふさわしいレイアウトを実現させた。

■ 第9講：ワープロでの文書作成法(2) —— テンプレートを使ったスタイルの 適用

❖学習項目

MS-Wordの「テンプレート」がどのような機能をもっているものかを説明し、これを用いることの利点を解説した。そして、テンプレートを利用して文書を作成する手順、テンプレート上のスタイル設定の変更の手順などを紹介した。

また、整った階層構造をもつワープロ文書は、「見出しマップの表示」「アウトラインの表示と編集」などの機能を有効に利用できること、また、それは、長文の文書を効率よく作成するのに便利であること、などを紹介した。

❖演習

ベタうちの中国語のプレーンテキスト（内容は、中国語の諺の言語的特徴に関するもの）を配付し、段落に分け階層構造化させ、ブロックごとにレベルの異なる見出し（大見出し、中見出し、小見出し）をつけさせた。

用意したテンプレートを配付し、その中で指定したスタイルを見出しマップで確認しながら、対応する構造に適用させ、レイアウトを整えさせた。

その後、受講者本人の好みにあったスタイルへ変更する作業をさせた¹⁴⁾。

最後に、目次を自動生成させた。右上図のような目次ができれば、階層化がうまくいっているということである。

■ 5. 検証

■ 5-1. 文章を構造化することの重要性

この授業の受講生は大学院生であるので、構造をもった文書を作成するスキルの必要性については、多くを語る必要はなかった。しかし、これが、学部生を対象にした授業であれば、この部分に、多くの時間と言葉を費やす必要があるであろう。

1. 民族的財産——谚语¹⁾
2. 中国的谚语²⁾
3. 中国谚语的语言特点³⁾
 - a. 用押韵的谚语⁴⁾
 - b. 用比喻的谚语⁵⁾
 - c. 用夸张法的谚语⁶⁾
 - d. 用文言的谚语⁷⁾
4. 在你们国家有没有相似的谚语?

第9講で自動的に作成した目次の一例

う。我々の国の学校教育は、この方面の素養を培うことに意を用いているとはいえないからである。

■ 5-2. 受講生が最も関心をもった項目

この授業では、CSSの操作は最重要学習項目ではない。HTMLに触れる都合上、基本的な事項のみを紹介するつもりであった。しかし、受講生に最もうけた——受講生がもっとも関心をもち、演習に熱心に取り組んだ——ように見えたのは、CSSによるレイアウトの試行錯誤であった。やはり、このような「デザインを変えて遊ぶ」のは、単純に「おもしろい」のであろう。

繰り返すようだが、HTMLとCSSを学習項目に入れたのは、なおかつ、MS-Wordによる文書作成の前にこれを学習したのは、(a)構造化した文書を作成すること、(b)定義した構造に対してそれにふさわしいデザインを適用すること、の基本点を理解してもらうためであった。しかし、MS-Wordによる作業よりも、CSSをいじる作業のほうが楽しそうであったのは、示唆に富む現象であると思われる。

■ 5-3. 言語の運用能力を高める作業と並行して、デジタルドキュメント作成のスキルを高める作業を行ったこと

これはあまり成功したとはいえない。

たぶん、受講生の母語である日本語の文書を素

材にして授業を進めれば、もっと多くの情報リテラシーに関する知識と技術を学習することができたと思う。あえて、中国語の文書を素材にしたのは、この授業が「外国語運用実習（中国語文章表現法）」であることも関係しているが、それ以上に、言語の運用能力には、テキストの構成を構造化して把握する能力、そしてそれを明示的に表現する能力も含まれるのだということを理解してもらいたかったからである。

しかし、これらの能力を育成するための演習のやり方があまり科学的ではなかった。HTML 文書の構築やワープロ文書におけるテンプレートの適用と異なり、ある文章を段落に分け、階層化し、見出しを付加し、という一連の作業には、「こうすれば誰もが正しい答えに到達できる」という手順がまだ完成されていない。未だに、教師の経験や嗜好に依存する部分が大きい作業である^[5]。このあたりが、この授業の重点がどっちつかずになってしまったこと、特に、運用能力を高める訓練が不十分になってしまったことの原因であろうと考えている。今後は、もっと体系的な訓練方法を考えてみたい。

■ 5-4. デジタルドキュメント作成の技術に関して

「文字コード」というものを感覚的に把握するための教材を用意し、演習を行うべきであった。講義だけでは、たぶん、十分な理解——とくに、なぜそれが文書作成に重要なのかという、実用的な見地からの理解——は不可能だと思われる。

また、HTML のタグにクラス属性を付加し、より「意味をもった」構造化をする、という作業の意義を確認するために、構造化文書のメリットを体験できる演習が必要であった。今回の授業では、単なるスタイル指定のための細分化であるような印象を与えたのではないかと心配である。

■ 6. この授業を構成するさいに参考にした文献

今回の授業を構成するために、次の文献を参考

にした。

- 『仕事のパソコン再入門——メール、ファイル、ツールを使いこなす』（館神龍彦、光文社新書 231、2005）
- 『eメールの達人になる』（村上龍、集英社新書、2001）
- 『今すぐ書ける中国語レター・Eメール表現集』（于美香／于羽、語研、2004）
- 『ユニバーサル HTML/XHTML』（神崎正英、毎日コミュニケーションズ、2000）
- 『Web 標準の教科書——XHTML と CSS でつくる“正しい” Web サイト』（益子貴寛、秀和システム、2005）
- 『Word で実践！編集レイアウトの基本と本格テクニク』（西上原裕明、技術評論社、2003）
- 『図解でわかる文字コードのすべて——難漢字からハングル・梵字まで』（清水哲郎、日本実業出版社、2001）

注

- [1] 「イマージョンプログラム Immersion Program」とは、本来、「ある言語を身につけるために、その言語が使用される社会の中に身をおいて、集中的に習得のための訓練を受ける教育」を指した。しかし、現在では、この用語は、
 - ある言語（学習者の母語ではない第二言語であることが多い）それ自体を教授の対象とするのではなく、言語を教科内容を指導する手段として使う教育、
 - 言語（第二言語であることが多い）のみを習得する教授法から、言語と一般教科を統合して教えていく教授法
 を指して使われることが多い。本稿では、後者の意味で用いている。
- [2] この授業で使用したコンピュータ演習室の端末の OS は、Windows XP Pro である。それに付属する「メモ帳」を使用し、フォントは、日本語と簡体字中国語が混在する文書であれば、SimSun に設定するのが無難であることを説明した。また、各行の「ウィンドウ右

端での折り返し」の設定のオン／オフを試させ、見かけ上の改行と、強制改行の違いを実感させた。余談だが、当該の教室の端末は、ログオフするたびに、初期設定に戻るようになっているので、「メモ帳」の設定もそのつど繰り返さねばならなかったが、これは却って、受講者にとってよい復習になったようである。

- [3] なお、この授業で扱う HTML のバージョンは、XHTML1.0 なので、言語指定は、「lang=」という書式なのだが、XHTML1.1 では、この書式は廃止され、「xml:lang=」となる。将来、「lang=」を解釈できないユーザエージェントが現れるかもしれないことを考慮し、2 種類の書式の言語指定を記述することを指定した。多言語混在文書で最も肝要なのは、文字化けをさせないことだからである。
- [4] 注 [2] でも述べたとおり、この授業で使用したコンピ

ュータは、ログオフするたびに初期設定に戻るようになっており、個々のユーザが、端末の使用中に、MS-Word のテンプレートをどのように変更しても、ログオフのたびに初期化されるので（作成した文書も消去されるが）、他者の利用環境に影響を与えることはない。このような環境は、本授業のような使用方法をする授業にはつごうがよいと言える。

- [5] 英語や日本語のような比較的良好な教授法が研究されている言語では、かなり科学的に追究された蓄積がある。その蓄積を生かした教材も工夫されている。しかし、中国語の分野では、まだまだである。もっとも、「こうすれば誰もが正しい答えに到達できる」手順が完成されてしまえば、人間並みの処理を機械が代行できてしまうことになるわけで、そう容易なことではあるまいが。

情報歴史学のこれから

花園大学・情報歴史学コースの4年間をふり返って

師 茂樹 (もろ しげき)

1. はじめに

花園大学文学部史学科の情報歴史学コースは、2006年3月、はじめての卒業生を社会に送り出すことができた。本コースは、その名の通り、コンピュータを用いた歴史研究とそのための教育活動を行うために設置された。これまで、一部の研究者や東京大学史料編纂所などの研究機関が、特定の目的についてコンピュータを利用するということはあったが、「情報歴史学」という一般的な名目で教育・研究活動を行うという例はなく、したがって担当者が手探りをしながらここまで何とか来ることができた、というのが正直なところである。

本誌にはこれまで、筆者(師)による開始直後の抱負^[1]と、佐藤・田中による中間報告^[2]を掲載していただいたが、この度4年間の総括を書く機会を与えられた。本稿では、卒業論文に向けた3、4回生のゼミでの教育実践について紹介するとともに、この4年間の意義、反省点、今後の課題について述べたいと思う。

2. 大まかな方針

花園大学の史学科は、文献史学(古代、中世、近世、近現代)だけでなく、考古学・民俗学・美術史学・禅文化史のゼミも開講し、日本史学に関

してはかなりの部分をカバーしていることが「売り」のひとつになっている。したがって、学生の興味関心も幅広く、情報歴史学コースにおいても、例えばテキスト・データベース(≒文献史学)に限定するというのは避けるべきだと考えていた。とは言え、もちろん筆者一人ではカバーしきれない部分が多いため、なるべく適切な指導が可能になるよう、情報歴史学コース担当教員を中心に史学科の他の教員にも協力を仰いでいる。

また、若い学問領域においては仕方のないことであるが、「情報歴史学」という名前を聞いても具体的にどのようなことをするのか、イメージが湧かない学生が多い。そのため、以下のような大まかな道筋を示している。

- 歴史情報サービス
- 数理的・計量的分析
- 視覚化/可視化
- シミュレーション

歴史情報サービスとは、研究文献データベースや史資料のデータベースなど、研究者や関心のある利用者が検索できるものを作り、提供することである^[3]。従来、人文科学におけるコンピュータ利用と言えば多くはこれに相当すると思われるが、これ自体は(狭い意味での)研究ではなく、研究者にとっての「道具」の提供でありサービスであるという位置づけの方がわかりやすいである

う^[4]。もちろん、このようなサービスを過小評価しているわけでは決してなく、むしろこれからの歴史学において必要不可欠のインフラであるということは、どんなに強調しても強調しすぎるといことはない（逆に、このようなサービスを研究よりも劣るものとしてしまう風潮にこそ問題があるようにも思う）。また、このようなサービスを提供するためには、歴史学の知識や方法が高いレベルで身に付いてないと不可能であることは言うまでもない^[5]。

学生にとっては、論文データベースやテキスト・データベースなど、Web上で利用可能な実例が豊富にあって、かつ情報歴史学コースの授業の中でもそれを利用する課題が多く出されることや、必修科目の中にデータベース（Microsoft Access）の授業があって技術的にある程度見通しが立つことなどから、このテーマを卒業論文のために選択する学生が多い。しかしながら、この手のデータベースは量が勝負になる面もあり、また類似事例が多いため普通に作っても新味に欠ける可能性が非常に高いため、学生の見通しほど楽な作業ではなく、また指導もそれほど簡単ではない。

次の数理的・計量的分析から下は、ある程度のデータベースを用意した上での研究になるが、中でも学生の関心が高いのは視覚化である。テレビ番組などで「コンピュータによる最新の研究」というと3DCGなどを駆使したヴィジュアル的なものが多いためか、学生側にとっても結果が想像しやすいらしい。しかしながら、統計ソフトウェアや3DCG、GISなどについては授業が開講されておらず、教員側では目的に応じたソフトウェアや環境を準備し、学生はそれらを自習しなければならない（言うまでもなく、教員がある程度、課外に指導しなければならない）、というのが現況である。

3. 具体的な実践

3.1. 論文講読

カリキュラムについては、佐藤・田中による

中間報告で示したものとその後大きな変化はない。しかし、その中間報告で述べられていたように、学生に歴史学の基礎知識や問題意識の持ち方、方法論が不足している問題を少しでも解決すべく、日本史学の方法論や史料論に関する論文の講読を、2回生向けの「情報歴史学研究」（後藤真氏担当）および3、4回生のゼミ（師担当）の冒頭15分間程度を使って行った。言わば、歴史研究の「疑似体験」である。

2回生向けには、青木和夫他編『文献史料を読む 古代から近代』（朝日新聞社、2000年）を中心に学生に読ませ、主に文献史学の方法論の基礎を学ばせた。しかしながら実際には、日本史の基礎知識を講義するような内容になってしまった。

また、3、4回生には、日本史学の従来の方と、コンピュータを利用した研究論文とを組み合わせ講読した。例えば、古代史の史料論として、加藤友康「古代史料の特質」^[6]と後藤真・柴山守「正倉院文書復原過程のXML/XSLTによる記述」^[7]とを組み合わせ、無形文化財の記録については新谷尚紀「映像民俗誌論—『芸北神楽民俗誌』とその制作の現場から—」^[8]と北川博美・磯本征雄「無形民俗文化財「古典万歳」を題材とするデジタル・アーカイブのためのWeb活用の試み」^[9]を組み合わせる、という具合である。講読においては、テーマに近い学生に概要を述べさせ、両論文の連続性、非連続性^[10]についても意見を述べさせた。

以上のような論文講読は、学生にとって負担が大きかったようで、しばしば事前に予習してこない学生が見られたのは残念であった。しかしながら、このように様々な方法論を比較し、批判的に検討させるという取り組み自体は有益であったと思われるので、読む論文をさらに厳選して今後も継続したいと考えている。

また今年度は、小中学生向けのコンテンツである文化庁の「子ども文化教室」^[11]と国税庁の「税の学習コーナー」^[12]とを比較し、対象となっている小中学生や教員にとってどちらが有益か、などのテーマでディスカッションを行った^[13]。これは日本史学というより博物館学などに近い問題設定であるが、コンピュータを用いて何らかのデ

データベースやCG作品を開発する場合、利用者の視点を欠いた開発はあり得ないし、複雑な歴史情報と様々な利用者とのインターフェースについて考えることは情報歴史学の主要テーマの一つとも言える。したがって今後は、このような課題をより多く設定する必要があると考えており、後に述べるブログなども活用しつつ、体験的にこのような問題に関心を持たせる工夫が必要であると考えている。

■ 3.2. ゼミ

■ 3.2.1. ゼミ発表

史学科を含む文学部の多くの学科において「ゼミ発表」というと、半年に一二度、教員が与えた課題（例えば文献の一部分の訓読や注釈、読解など）や自身の研究テーマについて発表をする、というのが多いのではなからうか。しかし、情報歴史学コースではそのような方法はなじまないと考え、各自の研究・開発プロジェクトの進捗を随時報告するような形式で行っているため、年間10回ほど発表するという学生も稀ではない。具体的には、だいたい以下の順序で報告させている。

- プロジェクトの構想（研究・開発対象の選択、方法・モデルの選択、問題意識の明確化）
- 先行研究・事例の確認
- 開発するデータベース等の設計、具体的なモデル
- 研究・開発の中間報告（完成するまで）
- 研究・開発の完成報告

実際には、最初の3点、すなわち日本史学界においてまだ顕在化していない問題を顕在化させ、それをコンピュータを用いて解決するための方法やモデルを確定した上で、データベースを設計する、という部分にかなりの時間を割いている。ある程度の知識や方法論を身につけている大学院生以上であればともかく、学部3回生の段階でこのような課題をこなすのはかなりしんどい。

また、設計においても、先行する類似のデータベースがある場合や、書誌データベースのように

教科書的なもの^[14]がある場合はともかく、それ以外の場合は自分で考えなければならない。先行事例については、まったく時代や地域、ジャンルが異なるものでも参考になる場合があるが、学生から見れば日本史とまったく関係がない分野のデータベースなどについて調査しなければならないことに抵抗感があるようで、教員による誘導が必要である。

■ 3.2.2. 卒業論文の題目

以下にあげるのは、昨年度の卒業生10名の卒業論文のテーマである。

- 日朝関係史における朴烈についての研究文献・史料データベース
- 織田信長と朝廷に関する論文データベースとその研究史的分析
- 新選組研究に関する書籍データベースの作成とその分析
- 自由民権運動研究史の統計分析
- 三徳円事件の叙述の変遷
- 近江六角氏研究
- 江戸時代、京都の地震災害とその現代的意義 —データベースの作成を通じて—
- 武家屋敷の3D画像作成とその意義 —明石城武家屋敷地を対象として—
- 戦後から現代にかけての朝食の変遷の研究
- 切る神事におけるふなずし切り神事の位置付け

ここにあげた題目のすべてが、情報歴史学コースにふさわしい内容のものではない。中には、コンピュータを使う必然性に乏しいもの、コンピュータでなければ解決できない問題を見出せていないものもいくつか見られた。これは、残念ながら、筆者の指導力不足が主たる原因であると考えている。

卒業論文のほかに、制作したデータベースなどに利用規程（ライセンス）を付した上でCD-ROMに収録し、付録として提出させた。

この中、「織田信長」「自由民権運動」はいずれ

も論文や研究書などの研究文献データベースであるが、単なるデータベースではなくそこから研究史や歴史叙述の歴史を分析しようとしている点が興味深い。「織田信長」においては、図1のように、研究論文の引用・被引用関係を数量化し、それによって論文の影響関係を統計的に分析している。「自由民権」も同様の方法で、研究史を分析しようとしている。

また、「新選組」では、学術的な研究が乏しく、逆に小説やムックなどが非常に多い点に注目し、ヘイドン・ホワイトのメタヒストリーの方法^[15]などを参照しながら、それらの叙述の傾向を分析しようとしている。このようなコンピュータを用いた研究史的研究は今後もっと注目されてもよいように思うし、歴史学の方法論に関する議論が情報歴史学と地続きである点は強調したい^[16]。

「武家屋敷」では、古文書や絵図、発掘調査報告書などを駆使して3DCGによる復元を試みる一方、歴史研究における復元の意義や、模型などと比較した3DCGによる復元のメリット・デメリットなどを論じている。3DCGの作成は、図2のような静止画のほか、域内を歩いてまわる簡単なムービーも作成した。

このような研究においては、歴史学的な面も当然重要であるが、これを推進するための環境が結果に大きな影響を与えることを、学生に対する指導を通じて痛感せざるを得なかった。例えば、1分のムービーをレンダリングするのに、それなりにハイスペックなPCでも10時間以上はかかってしまう。しかし、花園大学は文系の大学であるためか、理系の研究室のように夜中でも学生が実験をし続けることができるような環境ではなく、学生にとっては不自由な研究環境であつたらう。よりハイスペックなマシンを導入したりすることである程度は解決できる部分もあろうが、根本的な解決のためには全学的な合意を経た研究環境の整備が必要であろう。

なお、現在の3、4回生の研究テーマとしては、「前田利家に関する研究論文データベース」「大工道具の地域的の分布に関する確率統計的分析」「ケガレに関する研究論文データベース」「浮世絵の

引用論文ID	著者名	発行年	引用回数
「織田信長の政治的役割」	坂新	1966年	254
「將軍権力の創出 二」	朝尾直弘	1972年	254
「織田信長と朝廷」	橋本政宣	1962年	211
「織田政権の基本路線」	奥野尚広	1976年	197
「將軍権力の創出 三」	朝尾直弘	1974年	180
「將軍権力の創出 一」	朝尾直弘	1970年	179
『信長と天皇』	今谷明	1992年	173
『織田大元と天皇』	今谷明	1992年	164
『浮世絵制作成立史論 織田政権の分析Ⅱ』	藤田 隆	1977年	168
「信長の三書指図について」	立花京子	1991年	157
「織田信長と三書指図」	堀新	1997年	124
「本能寺の変拾遺—『日々記』附載天正十年夏記」	岩沢 忠彦	1968年	123
「信長への左大臣推任について」	立花京子	1993年	122
「織田政権の京都支配—村井貞勝の職掌の検討」	久野雅司	1997年	120
『新選組 幕府期の朝廷生活』	池 亨	1992年	113
「天正」	高木達太郎	1966年	102
「織田政権期における天下」	小原 康一	1987年	99
『定判義塾』	奥野尚広	1960年	80
「幕府体制における公儀と朝廷—統一権力形成」	水林 彰	1987年	74
『織田信長』	池 亨	1993年	74
『織田信長』	鈴木良一	1967年	72
「織田信長の全国制覇正当化の穴、天下静謐」	立花京子	1997年	63

図1 織田信長関連論文の引用・被引用関係の統計的分析

色彩」「キトラ古墳の3D画像による復元」「丹波国分寺の3D画像による復元」「GISを用いた姫街道の研究」などがある。

■ 3.3. Wiki、ブログの活用

情報歴史学コースでは、Wiki^[17]やブログ^[18](図3)、メーリングリストなどを使って、正規授業外での情報提供（補足説明）や課題設定、コミュニケーションなどを行っている^[19]。

ブログなどを開設するだけでは、コースの学生は自主的に見ようとはしない（逆にコース外の学生や学外の関心のある人は自主的に情報を求めにくる人が多い）ため、上に述べた論文講読の課題などをブログで告知するなどの方法で強制している。それによって、課題を自宅のパソコンで確認するついでに他のサイトも見に行ったり、関心のあるテーマのイベントに参加したりするなど、学生の潜在的な自主性を開拓する効果はあったよう

図2 明石城武家屋敷の復元3D画像

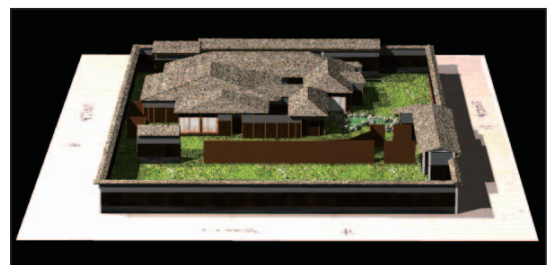




図3 情報歴史学研究室 Blog

に思われる。また、未だ一般に認知されているとは言えない情報歴史学という分野を広報する効果も期待している。

■ 3.4. 課外活動

■ 3.4.1. 課外活動の重要性

文学部史学科という枠組みの中でカリキュラムを構築しなければならない関係上、正規の授業の中だけで十全な教育プログラムを提供できているとは言えない。したがって、課外活動や課外での指導が重要になってくる。ここでは、上に述べたブログやメーリングリストでの情報提供や、3DCGソフトやGISソフトなどの習得のための個別指導以外の取り組みについて2点ほど紹介したい。

■ 3.4.2. 情報歴史学研究会

近年、大規模データベースの開発が多くの機関で進められている現状を見るまでもなく、共同研究・共同開発は情報歴史学では避けて通ることができない課題の一つである。しかしながら、正規の授業では、どうしても卒業論文を前提とした個人研究になりがちであった。加えて、情報歴史学コース以外の学生や大学院生、学外の若手研究者の中には、人文科学におけるコンピュータ利用に関心を持つ者も少なくない。そこで、情報歴史学コースの学生を中心に、学内・学外の関心のある人々によびかけて、情報歴史学研究会を発足した。

研究会では、メンバーによる事例報告や関連施設への見学会^[20]のほか、共同開発のプロジェクトを1年単位で行っている。プロジェクトでは、要件定義、設計、実装、テスト、保守などの典型的なソフトウェア開発プロセスを意識させ、また上級生に下級生を指導させるというような役割を持たせることで、ソフトウェア開発の基礎知識がある程度体験的に学習できることを狙っている。

2005年度は、QRコードと携帯電話を用いた花園大学歴史博物館の展示の閲覧を支援するシステムを共同で開発した。その際、研究会の有志5名が花園大学の2005年度学生生活動支援援助金に

図4 携帯電話による博物館閲覧支援システム



申請し、採択（20万円）され、それを元にシステムを開発、展示（学園祭の一企画としても参加）、関連する調査などを行うことができた（図4）。

今年度は、国文学科の丸山顕徳研究室と共同で、奈良県の説話に関するマルチメディア・コンテンツを開発している。

■ 3.4.3. アルバイトやインターンシップ

学生主体の実習が重要であることは言うまでもないが、この場合ある程度のものでできると、クオリティを少しでも高めようというモチベーションを持ちにくくなる。より高い目標設定をさせるためには、教員による指導もさることながら、プロフェッショナルの仕事を目の当たりにすることが大きな効果があるのではないかと考えている。

幸い、花園大学では国際禅学研究所を中心にいくつかのデジタル・アーカイブのプロジェクトが進められており、学生にはそこでアルバイトという形で参加できるようにしている。また、Webデザインの企業などをお願いして、インターンシップをする機会を作っている。

参加した学生の反応は様々であったが、自分をより高める必要性に気づく者もあり、教育効果は高いように思われる。

■ 4. 情報歴史学のこれから

以上、情報歴史学コースにおける教育実践について紹介してきた。

当初考えていた目標をすべて達成できたわけではないが、複数の異なった領域の間で顕在化していない問題を顕在化しそれを解決することができる人材の養成については、不十分ながらもある程度の訓練ができたのではないと思う。3月に卒業していった学生の就職先は必ずしも情報や歴史学に関連するものではない、ごく一般的な企業が大半であったが^[21]、このような問題発見と解決の能力はどのような分野に就職したとしても必ず役に立つのではないかと確信している。

また、コンピュータを単なる「道具」とせず、

歴史学（を含む人文科学全般）に新しい方法をもたらすものとしてその可能性を追求する点については、その萌芽を見出すことができたのではないかと思う（ただし、この萌芽も、大学院生以上の研究者の育成を通じた学問の再生産を経ない限りは、結局は小さな芽のままで終わってしまうだろう）。いずれも自己評価なので、諸賢のご指導をぜひともあおぎたい。

ただし、上に紹介した様々な実践は、思いつままま半ば場当たりに、必要だと思ったことをその時その時にどンドン試してみたため、体系化も洗練もされておらず、学生にとっても負担が大きかったかもしれない。今後これらを体系化し洗練していくことが一つの課題となるだろう。

注

- [1] 師茂樹「情報歴史学の課題 花園大学・情報歴史学コースのための主観的航海図」（本誌第3号、2002年）
- [2] 佐藤誠・田中正流「花園大学「情報歴史学コース」の現状と課題」（本誌第4号、2003年）
- [3] 歴史研究者間の情報交換や共同研究、交流を深めるためのコミュニティ・サイト（Social Networking Site）もまたサービスのひとつとして研究・開発する価値があると思われる。それについての予備的な情報提供はしているが、実際の指導には至っていない。
- [4] 柴山守「「歴史情報学」から「情報歴史学」への展開—自問自答—」（『沖縄の歴史情報』1、1998年、<http://www.tulips.tsukuba.ac.jp/liemedio/dlam/B1132580/1/mokuji/5114.pdf>）
- [5] 「道具」に込められた研究者の思想や歴史観などを無視することなく、利用する際にはそれらを意識し、また場合によっては批判的な検討の対象とする必要があろう。
- [6] 歴史学研究会・日本史研究会編『日本史講座 2 律令国家の展開』（東京大学出版会、2004年）
- [7] 『情報知識学会誌』Vol. 11, No. 4（2002年）

- [8] 『民俗学の資料論』(吉川弘文館、1999年)
- [9] 『情報処理学会研究報告』Vol. 2004, No. 110 (2004-CH-64) (2004年)
- [10] 新谷「映像民俗誌論」と北川・磯本「無形民俗文化財「古典万歳」を題材とするデジタル・アーカイブのためのWeb活用の試み」との非連続性については、師茂樹「デジタルアーカイブ」とはどのような行為なのか(『情報処理学会研究報告』Vol. 2005, No. 51 (2005-CH-66)、2005年)で述べたことがある。
- [11] <http://www.bunka.go.jp/kids/>
- [12] <http://www.nta.go.jp/category/gakusyu/kyousitu.htm>
- [13] 国税庁「税の学習コーナー」を作成した有有限会社アルファサード(<http://alfasado.net>)では、小中高生へのヒアリングを重ねた結果、文化庁のサイトで見られるようなアニメーションやゲームはかえってよくない印象を与えるという結論に達し、現在のようなデザインになったそうである。
- [14] 例えば上田修一『書誌データベース構築法』(丸善、1991年)など。
- [15] メタヒストリーについては、Hayden White. *Metahistory: The Historical Imagination in Nineteenth-Century Europe*. Johns Hopkins University Press, 1973. や富山太佳夫「物語、歴史記述、ディコンストラクション」(『思想』735、1985年)など。
- [16] 「コンピュータとは、新しい歴史研究者のタイプとか、新しい方法論とかを生み出すために必要とされるわけではなく、すべての歴史研究者向きの適切な用具なのである」(モーズリー、ムンク『コンピュータで歴史を読む』(有斐閣、1997年)という意見もあるが、これについては別稿で批判したことがある(師茂樹「データベースがもたらすもの コンピュータの中の歴史/物語」[方法論懇話会編『日本史の脱領域 多様性へのアプローチ』、森話社、2003年])。
- [17] <http://moromoro.jp/morosiki/cgi-bin/uniwiki.cgi>
- [18] <http://ihistory.blogspot.com/>
- [19] ブログの活用については、2005年12月17日の日本中国語CAI研究会大会(本誌の報告参照)における小川利康氏の発表「講読授業における検索エンジン活用」に触発された部分が多い。小川氏の講義ブログ(<http://china.ogawat.net>)も参照されたい。
- [20] 2005年には国立民族学博物館のデジタル・アーカイブの作成現場を見学させていただく機会を得た。また、小倉百人一首のデジタル・コンテンツを任天堂が開発した時雨殿も見学した。
- [21] もっとも、文学部出身の場合、専攻を直接行かせるような就職先などほとんどないのが現状であろう。一方の情報分野に関しては、よりこの道を追いたいということで情報系の専門学校に進んだ卒業生がいた。これは要するに、情報歴史学コースで身に付く能力だけではIT関連企業の技術系の職にはつけないということであり、教員としては忸怩たる思いがあるのも事実である。

BBSより：人文学的な情報教育

本誌編集局

■ はじめに

本誌編集局では、特集2「人文学的な情報処理教育とは」の一貫として、人文学の分野で情報教育を行っている方々に、現場からの実情を踏まえた意見や、そこから見えてくる人文学の分野での情報教育の問題点・将来像について討論をしていただくことを計画しました。

そして、本会の掲示板（BBS）上に会議室を設置し、「人文情報学の定義」「人文情報学教育」という大きく二つのテーマを設定し、会員の皆様に活発な議論を行っていただきました。

以下に討論の内容をできるだけ元の文章を残した形でまとめてみました。なお、注意点は以下の通りになります。

- 出来るだけ元の発言順序に即して並べてありますが、編集上の都合で必ずしも発言日時順にはなっていないものがあります。
- 段落構成、表現等、編集部において文意を損なわない範囲で改めた部分があります。

■ 「人文情報学」の定義について

❖ 山田崇仁（2006/07/19 00:30:00）

ここでは、「人文情報学」そのものに対する定義をもう一度考えてみようというのがテーマです。

昨今、「人文情報学」に関連する専攻や科目を設置する大学が増えてきました。しかし、その中身は多種多様です。

「人文情報学」に対する人文系研究者の側の一般的理解は、「人文系の研究に情報系の方法論やツールを役立てる仕組みを学ぶ・作る」だと思います。これは、人文学を中核に置いた考え方と言えるでしょう。一例として、人文情報学教育を行っている研究・教育機関の定義を紹介しましょう。

- 「人間」という視点から情報や情報社会を捉える目を養います（大谷大学人文情報学科より引用。<http://www2.otani.ac.jp/hi/>）。
- 伝統的な情報科学・情報工学を中核としつつ、人間や社会を対象とする人文情報学（国立大学法人総合研究大学院大学より引用。<http://www.soken.ac.jp/rcourse/9-17/9-17-2.html>）。

両者の違いは、研究主体となる研究者のスタンスが何処に置かれているかによると解釈できますが、何かしら違う表現もあるかな？とも思うわけです。そこで、会員の皆様の考える「人文情報学」とはそもそも何かについて活発なご意見を出していただけるようにお願いします。

提議者である山田は、「人文情報学」という新しい？学問領域を開拓すると言うより、「人文学の研究に情報系の方法論やツールを役立てる」という意識です。上記分類では前者に当てはまるでしょうが、どちらかといえば「情報処理手法をも用いた人文学」だと自らの立場を定義しています。

❖ 山田崇仁（2006/07/24 16:06:15）

先の書き込みで、自らの定義（立場）について説明しました。「人文情報学」という言葉の定義そのものについては、「当該分野は存在し得ないのではないか？」になります。

それとは別に、人文系・情報系の二通りの立場からの「人文情報学」の定義を挙げました。両者は、中核とする手法や立場は異なるにせよ、最終的には「人文学の分野に貢献する」という目標設定であると定義できるでしょう。従って、「情報」を前面に出すのではなく、「人文学」の枠組みの中で論じるべきではないかと考えています。その点から言えば、（編集局注：後述の発言）千田さんの以前書かれていたスタンスに近い「人文学における情報処理手法の利用」が、私の「人文（学への）情報（的手法を応用した）学（問）」になるかと思えます。

ただ、現在の「人文情報学」という言葉の実際は、単なるキャッチコピーでしかないと考えています。大本がキャッチコピーである以上、「人文情報処理教育」をする側の実体は何でもよいわけです。例えば、データベースで調べ物をした感想を書いても、それは「人文情報学」なのかもしれません。そういう意味で「人文情報学」は、便利な言葉と言えるでしょうが、それは逆に学問領域そのものの曖昧さをも意味するかと思えます。

こちらのスレッドは、定義そのものを問題としていますが、教育の方は「人文学への情報処理の応用をどう教育するか？」的な議論で進んでいますので、その立場で論じてください。

❖ 千田大介 (2006/07/23 08:42:22)

今まで、本誌で「人文情報学」ってコトバを使ったことは無いはず。少なくとも私は、かなり意識的に「人文学情報処理」と言ってきました。

「人文情報」という学問と、「人文学」における情報処理の応用とは、本質的に全然意味が違ってしまっているので、（今回の討論は）従来の方針の踏襲で行くのか、それとも新たに人文情報学の確立・定義という方向でやるのか、お手数ですがご確認くださいませるか？まあ、各内容は、人文情報処理といった方がふさわしいような気がするし、省略すれば人文情報なんだけど（笑）。

❖ 山田崇仁 (2006/07/24 15:36:57)

過去の特集を見ますと、確かに千田さんがおっ

しゃられているように、「人文学情報処理」という定義が使われています。

「人文情報」という言葉を私が使っているのは、現在の職場の影響なんだろうが（笑）、私自身の立場や語れる事というのは、「人文学」における情報処理の応用」でしかありません。

しかし、「もしかしたら、「人文情報学」という新分野？について語れるメンバーがいたらいいなあ」という淡い期待を持って曖昧にしていました。

ですから、どの立場に立って語っていただいても構わない（というか、この定義自体、おかしいんじゃないか！・毒されすぎ！・そんな分野自体存在しない！・結局人文学の中に反映されないと意味無いでしょ！…というような突っ込みもありだと思っています。）のです。

❖ 師茂樹 (2006/08/03 22:45:35)

私は、「人文情報学」という新しい分野についてちゃんと語れた人はいないと思うけど、いずれちゃんと語りたいなあ」という人です。学として成り立たないかな？という野望です（^_^;;

ツールとしてのみコンピュータを使うというのは、私としては人文情報学とは言いたくありません。秋山さんのおっしゃっていること（編集局注：後述の秋山発言）と重なると思えます。

同志社の文化情報学部の「データサイエンスで文化を学ぶ」というコンセプトは、私の考える人文情報学ではないと思います。データサイエンスという方法は、別に対象が人文系のものじゃなくてもいいからです。対象によって「～学」と名乗る学問分野はありますけど（敦煌学とか）、基本的には方法論や問題意識によって「～学」になるのだと思います。

一方、秋山さんがおっしゃってた「計算機を離れて」というのは、気持ちはわかりつつも、コンピュータはやはり必要かと思えます。検索システムの構築をツールの作成と見なせばそれまでですが、これこそが人文学的な営みを自己言及的に記述する行為だとは言えないでしょうか。これは、言わば外部から、許慎を情報アーキテクトに見立て、それを記述するのとは違うと思えます。

❖ 山田崇仁（2006/08/07 15:15:17）

そもそも、「人文情報学」という言葉が一人歩きして、皆がとりあえず納得するような定義ができてすらいない状態なのかかもしれません。

今時の「人文情報学」は「情報学」に偏ってしまっているの、むしろ「情報人文学」の方がいいのかもしれない。また、これとは逆に「情報学的な問題を人文学の手法で解決」ってのがあっていいかと思えます。

「ツール」の定義（程度）も分かれるかと思えますが、一般的には、単なる用例検索（キーワード単位・grep）が該当するのでしょうか。私はN-gramで機械的に集めるのも「ツールの使い方」と割り切っています。そこからすると私のやっていることは単なる「人文学」なんです。N-gramの分割と頻度集計って時間さえあれば手作業でも出来るしるものです。

自分が普段何気なくやっている作業を、コンピュータに解釈させるにはどのような方法論（作業かな）が求められるか？ という仮題に日々直面している私としては、それ自体が一つのネタというか実験場ですね。これこそ「人文情報学」といえるのかもしれない。

■ 定義からの派生ネタ：「人文学的な情報処理教育」の位置づけ

❖ 山田崇仁（2006/07/19 00:34:59）

新しい学問分野の誕生というのは、何らかの差異を見つけて意識的に他者と差別化する事だと思えます。そういう意味では、上述の「新しい学問分野とっていない」という山田の意識は、単なる怠惰なのかもしれません。

しかし、現状での「人文情報学的方法論」を用いて人文学へ研究成果を出している人に対する「伝統的人文学の担い手？」からの評価は、『電脳中国学』が出た当初から「魔法使い」や「色物」的な認識しか持たれていないと思うわけです。

無論、その当時から状況は変わり、原稿のメー

ルでのやりとりも珍しくなく、漢籍電子文献を学部生が普通に使うようになりました。しかし、そこから全然進んでいないのが現状だと思います。これだけコンピュータがツール化しているにも関わらず、人文学に情報の方法論を持ち込んだ研究が色物扱いをされるのは、「伝統的人文学の担い手」の方々から根本的に理解されていない（しようとしな）のが大きな理由だと思っています。

そのため、折角おもしろい着眼点や方法論で「人文情報系」の方が研究されていても、それを「人文系本流」の方が評価できる？体制にないために、結果として全く無意味な代物が出来上がってくる場合も珍しくないわけです。

複数の異なる方法論の領域を跨ぐ研究分野の場合、両方の分野、特にメインターゲットとする分野（私の場合は人文系）で妥当と認められないと中々受け入れてもらえません。従って、「人文情報学」が自らの足場を固め、より発展するためにも、既存の分野に対する切り込み・方法論に対する理解・人文系側からの評価…をより積極的に行う必要があります。その理解の架け橋となるのが「人文情報教育」ではないかと思っています。その意味でも、既存のカリキュラムにおける「人文情報教育」の位置づけを明確にする事、これが最低限必要な行為ではないでしょうか？

❖ 秋山陽一郎（2006/07/20 02:32:58）

僕は「教育」を語るほどの経験も実績もありませんし、また「人文情報学とは何か」に対する解答も持ち合わせてはいませんが、2年間ほど「東アジア人文情報学」の拠点にいて漠然と感じたことを、折角なので書き留めておこうと思います。あくまでひとつの見方に過ぎませんし、自分の中でも成熟していない感覚で、今ひとつ自信がないのですけど。

まずは極論（暴論？）から。「そもそも「人文情報学」に計算機（コンピュータ）は必須なのでしょうか？」もちろん、「人文情報学」にとって計算機がツールとして切っても切り離せない存在であることは否定しませんが、所詮、ツールとは目的を達成するための手段であって、目的そのも

の、本質そのものではないわけです。（ある目的を達成する手段として、結果的に最適なのが計算機だということはあるでしょうけど。）自然言語処理やデータベースシステムといった手段やツールにしても、すでに限りなく似たような目的を達成するための工具書が紙媒体ベースで刊行されているケースが多々ありますが、この間の差は何でしょう？僕には媒体の差でしかないように思えるのですが、皆さんのお考えはいかがでしょうか？

というのも、最近、（特に人文系の研究者にとって）「人文情報学とは何か？」を考える時に、一度計算機を離れた方が本質にたどり着きやすいような気がしています。例えば、許慎を情報アーキテクトに見立てれば、漢字を「形」（部首）でカテゴライズした『説文解字』も人文情報学的な業績として見立てることができるのでは？とか。

上のようなことを言ってしまうと、学問領域の線引きが曖昧になってしまう気もしますが、一度こういう考え方の上に立たないと、いつまで経っても、単純な語句検索や用例抽出の枠を脱却できない気がしてなりません。もちろん、脱却している人はとっくに脱却しているわけですが、メタな議論している層と、検索システムなどの特定の手段に依存してしまっている層との間の格差や、僕自身も含めて若手が育って来ていないというのが気掛かりなところですよ。

まず身近な工具書がどういうモデルで作られているか、そのモデルがどういう利点と欠点を持っているのか、どういうモデルにすれば欠点が改善されるのか、といった、媒体差を超えたところでの議論から入るだけで、大分、敷居が低くなると思うのですが、いかがなものでしょうか？でも、2年間、朝から晩まで人文情報学の真似事をしてきて感じたことは、「紙媒体の工具書を使いこなせない者は、計算機をツールとして使いこなせない」ということと、「工具書を作れる人間は、人文情報学的なセンスがある可能性が高い」ということでした。

❖ 二階堂善弘（2006/07/20 15:28:31）

秋山さんの意図はだいたい分かるつもりですが、

敢えてこちらも、「はい、必要です。」と答えたいと思います。より暴論としては、「使えなければ人文情報学でない」と言ってもよいでしょう。

むしろパソコンなんて道具に過ぎませんし、いまやインターネットの方の重要性の方が高くなってますけどねえ。それでも、いまのところ、コンピュータをはずした形での人文情報学は存在意義が薄いと思ってます。

（秋山発言：単純な語句検索や用例抽出の枠を脱却できない気が）それは理解しますが、その語句検索が比類無い力を発揮するもの、パソコンなるツールがあつてのことです。結局はパソコンスキルを上げていくことが必要だと思います。いま「中国基本古籍庫（編集局注：本誌二階堂氏のレビューを参照）」の検索が出来ても、中身を把握してないような院生がおりますが、それでも、非常に面白い「結果」を出してきています。

ある意味、読めないのに結果が出せる、自分はいったいこの傾向も「あり」だと思ってますよ。とはいえ、読めて、工具書が使えて、さらにパソコンやネットが使えればベターなんですけど。ただ、この場合結果を出すのには、パソコンは必須なツールだったわけです。登り口はいろいろある。いまやパソコン自体もそうなんですよ。これはこれで興味深い傾向なんです。

❖ 清原文代（2006/07/31 12:42:00）

1年生の授業で、「中国語では「つかむ」も「つかまえる」も「抓」で言える。日本語を学ぶ中国語話者のために、日本語の「つかむ」と「つかまえる」の意味や用法の違いを説明しなさい。」という課題を出したら、日本語の用例を引っぱってくるのに青空文庫を使った学生が1人いました。なんでも普段は携帯電話で青空文庫を読んでいるのだそうです。勿論用例をどう解釈していくかということとはまた別の話ですが、以前なら1年生の学生が授業の課題ごときではやらなかったであろうことが比較的簡単にできてしまうわけです。これは私が大学入学した頃にはあり得なかった状況だと、ある種の感慨を憶えました。

■ 人文学的な情報教育

❖ 山田崇仁（2006/07/19 00:31:17）

本会会員の皆様は、その多くが何らかの形でいわゆる「人文情報学」や「人文学からの情報教育」なるものに携わっているかと思います。

ただし、その中身は Word, Excel 入門であったり、パソコンやインターネットで中国語を読み書きしたりする方法といった類のものである事も珍しくありません。また、それらの科目が一定のカリキュラム体系の中に確固として位置付けられているならばまだしも、他のカリキュラムとは何の関連もなく「とりあえず設置してみました」・「内容については、担当教員にお任せしています」という場合も珍しくありません。

山田は「人文学的な情報教育」について、個々の学問分野に密着したツールの部分と、情報学的方法論との橋渡しの部分と分かれてくると思います（間違っていたら指摘して下さい）。

前者の場合、専門領域から独立した科目として設置すると、どうしても抽象的・表層的な形になりがちです。また、専門領域の中で単発設置をされても、あまり突っ込んだ話が出来ません。何れにせよ、導入的な形で、後は自分で深めていってね！というスタンスが中心になるかと思います。

後者の場合、内容によっては殆どコンピュータを触らない科目も構築可能なので、前提にパソコン教室を使って実習を行う科目として定義されている場合、授業そのものが難しくなります。

上記の話は、山田の経験を踏まえていますが、皆様の経験を踏まえた「人文学的な情報教育」の現状と未来について話し合えればと考えています。

❖ 二階堂善弘（2006/07/19 23:00:42）

昔「人文情報は人文学の中に埋没すべし」とか書いた事があります…。しかし、予想に反してほとんどそういった事態は起こってこないようです。むしろ、ワープロやインターネットによる情報検索は当たり前になってきました。これ

については、むしろ変わってきたと言えるでしょう。ただ、それはあくまで「駅と駅のルートを調べる」といった使い方以上のものではないと思います。便利は便利。でも手法的には全然応用されてない。むしろ自分もそれほどの使い方が出来るわけでは無いんですけど。

関西大の大学院でも、例えば「基本古籍庫」を使ってる人も徐々に増えてきているんですけど、やはり「便利な索引」というものでしかない。そもそも、検索したデータをほとんどの人が全部印刷して、その紙をベースにいろいろ作業するのが主体になってますのでねえ。こうなりますと、本当に索引としての用でしかない。

もちろん、これは大事なことです。ただ、予想よりはなかなか「人文学との融合」という事になってきていないのが、残念といえば残念ですわ。

❖ 清水哲郎（2006/07/20 08:23:12）

最近、「今どきの高校生～大学生ならワープロぐらいできて当たり前」という「俗説」というか「先入観」について考えさせられています。

今春から1年生向けのPC実習授業を受け持っていますが、ワープロ（Word）に関しては過半がまったくの初心者です。もちろん、かな漢字変換できるだけというレベルであれば、ほとんどの者がクリアはしていますが、コピー&ペーストなり、書体（フォント）の使い分けなり、インデントなどの段落書式の設定なり、表の作成といったワープロらしい機能の活用となると、入学時点でそこそこ使っているのはせいぜい3割程度でしょうか。「高校で「情報」科授業が始まっているので、変わってくるのでは？」と思っている方もいるかも知れませんが、短期的には悲観的です。というのも、現状、高校の「情報」科教育ではほとんどワープロをやっていません（もちろん、小～中学生でも）。いま、「情報」科のサブテキストのようなものをつくるお仕事にも携わっているのですが、その監修にあっている情報系の某教授から「今どきの高校生ならワープロぐらいできて当たり前」的な発言があったのに対してわたしだけでなく、現場の「情報」科教員からも「高校の「情

報」科教員でもまともにワープロが使える先生は少数派ですし、実際に授業ではほとんど教えていません」という指摘があって、某教授も驚いて沈黙してしまうひと場面もありました。

本研究会の皆様は相当のPC経験をお持ちですので、今さらワープロなんて…とお感じになっている方が多いかと思えますし、Webや検索などをこなす大学生を見ていると信じられないかも知れませんが、「生まれてからまともにワープロソフトを触ったことがない高校生はけっして例外的ではなく、普通！」なのです。このスレッドの「人文情報学」とはまったくレベルの異なる低次元な話で恐縮ですが、たかがワープロ、されどワープロ。Wordの操作方法を教えることが目的ではなく、正しい(=標準的な)日本語文書を効率的に作成する方法を教えることを目的としたワープロ教育の重要性を痛感している昨今です。

❖ 師茂樹 (2006/07/20 12:09:25)

私の印象では、以下のような気がします。一時的なものかもしれませんが。

- 「ワープロぐらいできますよ」という妙に自信たっぷりの学生が増えた (でも実際は打てて印刷できるぐらい)
- できる学生とできない学生との格差が広がった。

(清水発言：正しい(=標準的な)日本語文書を効率的に作成する方法を教えることを目的としたワープロ教育の重要性) 確におっしゃる通り。秋山さんが別スレッドで「工具書を作れる人間は、人文情学的なセンスがある可能性が高い」と言ってますけど、「ワープロで、きちんと構造化された文書を書ける人は(以下同)」とも言えるかもしれません。いずれにせよ、コンピュータを使わなくてもよい話になってしまいますが(^_^;;

❖ 清原文代 (2006/07/20 12:42:41)

(師発言：「ワープロぐらいできますよ」という学生が増えた) はい、私もそういう学生によく遭遇します。ただべた打ちしているだけで章も節もないとか、レポートや論文に欠かせない脚注機能

を知らなかったりします。

(師発言：学生間格差が広がった) わたしもそう思います。たとえて言うのなら、鉛筆の使い方自体がわからない人と、鉛筆をどううまく使って良いものを書こうかというレベルの人が、一緒くたになっている感じがします。でもこれは一時的な問題ではないかもしれません。清水さんがおっしゃるように、結局どの教育段階でもワープロくらい使えるだろうってことで、「鉛筆の使い方」をきちんと教えていない現状がありますから。

私は情報教育に直接関わっていませんから、外野から見た単なる印象論に過ぎませんが、一時期に比べて大学で情報教育をやらなきゃという意欲が薄れてきたように思います。昔はとにかく大学でパソコンとネットを使える環境を整備してやらねば、そして環境だけでなく少なくともワープロ・メール・Webくらいは使えるように大学がしてやらねばという義務感みたいなものがあつた気がするんですが、それが今ではそのレベルのことは大学の仕事じゃないという雰囲気があるような。

❖ 二階堂善弘 (2006/07/20 15:42:55)

(師発言をふまえて) 確かにそうなんですけど、一応自分の守備範囲としては、そのくらいの学生を「ワープロが使える」と考えます。さらに次となるとなんとでもですが。とはいえ、まず多くの学生がワープロを「当たり前の道具」と思ってくれているのはありがたいかもしれません。もっとも、文字を打つのは全部「ワープロ」として考えてるので、いろいろ問題なんすけどね。「先生、ワープロで打てません。」というのは、だいたいIMEで出てこないすけどね。またMS-IMEが日本語のすべてだと思ってるし…。ATOKだってあるんだけどなあ。嘆く一方で、このカオスな状態を楽しんでもいます。我々の側も、もう少し教えるメソッドを確立すべきなのかもしれませんし。

❖ 清水哲郎 (2006/07/20 22:45:31)

4月の授業開始時に250名全員にPC経験やワープロスキルなどについてのアンケートをとったり、実際に3ヶ月間、授業を行ってみたりした

感触としても、二極化しているのは間違いないだろうと感じています。また、先週、前期の試験（実際に Word で定型的な文書を作成させる実習試験）を行い、ちょうど今日、採点を終えたところですが、その結果でもその感を強くしています。

そこで、Excel をやる後期の授業では途中で何度か、（これから行う授業に関する）実習試験を行って、できた学生さんは「以後、数回の授業は出席不要！」として、できなかった学生さんだけを対象に以後数回の授業を行うといった取り組みができないだろうかと思案しています…大学でそんなやり方が許されるのかどうかもわかりませんが（運転教習所みたいに、合格スタンプを押したら、次のステップに進むってイメージです。英語の授業などでは、学生さんのスキルに応じたクラス分けを行っている大学もあるようですが、自分の授業の中だけで同じような効果をあげたいという狙いです）。

数年前まで会社勤めをしていて、いまも何社かとビジネスを続けている経験からすると、新入社員に対して PC スキルの教育を行っている民間企業もほとんどありません。会社は会社で「ワープロぐらい、学生時代にやっているだろう」ってワケです。

（清原発言をふまえて）まったく同感です。「ワープロぐらい、Web ぐらい、メールぐらい … できるだろう」病がどこの大学にも感染している気がしています。わたしが今春からお世話になっている大学でも、ここ 2～3 年、PC 関係の実習授業数が減るなど、徐々にそういった方向に向かっていくようですが、これは「Web ができること」と「ワープロや表計算などのツールを使いこなせること」を混同した誤った判断だと考えています。そこで、1 非常勤講師の身分でありながら（だからこそ？）、学部長さんや教務委員の先生方に向かって「今こそ、大学全般の趨勢に逆らって、PC スキル教育の質と量を拡充すべし！」などと訴え始めたところですが…利益をあげればたいのことが通る民間企業とは違って、大学ってところはなかなか難しい世界の世界ですので、どこまで聞いてくれるかはわかりませんが（笑）。

● 千田大介（2006/07/31 21:55:43）

（清水発言をふまえて）これは確かにあります。というか、そもそも従来の大学の情報処理教育体制が、看板の中はどれほどのモノであったのか、すらも疑問です。情報を教える教員は大抵、理工系の数学なり情報なりの出身ですが、教える対象の人文・社会系の学生のニーズに合わせて、キチンと情報を教えられる人材をこの人数、上手く確保できている所は、非常に少ないんじゃないでしょうか。翻って、人文情報。人文情報「学」以前のリテラシーレベルの教育を、必修として課している文学部は、まだまだごく一握りでしょう。

ところが一方にあるのが、人文情報軽視です。軽視っていうのは、必要ないと思うのではなく、自分程度できていれば大丈夫、だという過信的な軽視です。そういうことを言う人に限って、メールは機種依存文字使いまくり、ワープロはスペース整形だったりするんですよ。これじゃ、人文情報リテラシーすら、普及すべくもありません。

もっとも、人文情報リテラシー、そして人文情報処理学の普及が進まない最大の要因は、もっと別のところにあると思います。

それは、「コストが見合わない」こと。

過激に極論すれば、人文系（特に文学）の論文なんてものは、体裁が整っていて、調査に汗をかいたことがアピールできれば、業績になってしまいます。また、こと中国文学分野では、手つかずの作品・作家・資料がゴロゴロしているので、ちょっと目先がきけば、幾らでもネタが見つかります。私自身、自分の研究論文への人文情報処理技法の活用は、用例検索止まりです。

そんな状況なのに、わざわざコンピュータスキルなんて余計なモノを学んで研究に役立てよう、などと考える人が居るとしたら、よほどの酔狂としか言いようがありません。

逆に言えば、人文情報処理が普及し、学問分野として確立するためには、まず、情報処理技法でなくてはなしえない、ものすごい学術的成果を上げて見せるしか無いでしょう。そのインパクトによって、コンピュータを使って人文学を研究す

るのが不可避であると認識され、更に人文学リテラシーレベルにまで普及したとき、その時初めて、人文研究の方法論としての情報処理学を、人文学から独立した学問分野として打ち立てる可能性が出てくるのではないのでしょうか。

❖ 山田崇仁 (2006/08/07 14:43:16)

(千田発言をふまえて) 確かに。ハードとソフトとネット代。どのくらい使ったかなあ…。普通の家庭では一家に一台のパソコン。我が家では一人二台以上…。さらにハードディスクはテラ単位で数えているし…。まあ、本会は好き者の集まりですからねえ。一般のニーズは「漢籍電子文献」辺りがなんとなく使えればオッケーですから。

(千田発言：情報処理技法でなくてはなしえない、ものすごい学術的成果) 問題は、そのものすごい業績をだれが評価するかですね。結局人文系の人間が最後に評価(納得か?)しないとだめなんだと思います。でも、多くの人文系の人々は評価(納得)できないから、しばらくは単なる色物扱いで進みそうです。まあ、私は研究者なんて社会の領導者じゃなくて、単なる滑稽者だと思っていますんで、それで構わないとは思いますが。

❖ 二階堂善弘 (2006/07/20 15:37:43)

(清水発言：「俗説」というか「先入観」) それは確かにありますねえ。実際、授業で教えてても、インデントやら何やらについては、全く初歩から教えないといけない。正直、人文情報学などと偉そうなタイトルの授業でも、実際には書式のやり方を教えてるわけですし。「ワープロが使える」というのは、一応ワープロを使って文章が書けるというレベルになってると思ってください。それ以上となると、やはり「ワープロを使いこなす」になるのかも…。ただ、それを踏まえて次の段階を、と望んでもいたわけですが、多くの場合、はるかに手前で終わってしまう感じです。

(清水発言：「高校で「情報」科授業が始まっているので…短期的には悲観的) これも、期待した通りにはサッパリなっていないようです。情報の授業自体が、かなり学校と学生のスキルによ

ってバラバラの内容になるようです。

いまは、「学部生は学部生なりの」「院生は院生なりの」やり方をそれぞれ覚えてくれればいいかと思ってます。検索以上を期待するのは、あくまで院生くらいのレベルです。

もっとも、それくらいに幅広い範囲を持つ「人文情報学」を、簡単に済ませようとするから難しくなるので。名称変えようかしら…。

■ むすびにかえて

以上、BBSでの議論をまとめてみました。

編集の過程で改めて議論を読み直してみると、今日における人文学からの情報教育の現状や問題点が見えてきたように思えてきます。

師氏の発言に見るように「ワープロができます」という学生は確かに増えました。しかし、「ワープロができる＝入力して印刷」程度の認識であり、逆にこれが「できるからいいや」と、それ以上の発展を阻害しているといえるでしょう。このような、なんだかあやふやな「できる」が一人歩きして、結局「何もできない」という状況が、今日的な情報教育なのではないかという気がします。

さらに、人文系の情報教育に至っては重傷です。千田発言にあるように「教える対象のニーズに合わせて人材を確保できている所が非常に少ない」事に加え、「人文情報軽視」という問題がある。これも結局は「自分程度できていれば大丈夫」という「相対的ですからない基準」意識のため、結果として「人文情報リテラシーすら普及すべくもない」状況に至ってしまっています。

極論すれば、三号の時期に比べて実は悪化しているのではないかという気もしています。どの分野でも常に関心を持つ層はいますが、それが広範囲に広がっていないのを実感するところです。

しかし、その中でもどうにかしてこれらの問題を克服すべく努力している人たちがいます。本特集に寄せていただいた諸論稿はその一例ですが、読者諸賢におかれましては、ここでまとめたBBSの白熱した議論と併せ、そこからなんらかの問題克服の方法を見いだしていただければ幸いです。



実践レポート

もはや日常生活になくなくてはならない携帯電話、急速に普及の進む携帯音楽プレイヤーや電子辞書など、私たちの身近にあるデジタルガジェットは、作り手も使い手もそれぞれに「こだわり」を持つようだ。パソコン教室を使った画一的な CALL 授業などとは一味違った、「電腦中国語」の世界をお目にかける。

日本中国語 CAI 研究会について

本会（会長：田邊鉄北海道大学助教授）は「コンピュータ援用の授業方法を中心とした中国語教授法の研究・開発・普及を推進し、同時に教員・研究者・ソフトウェア開発者の交流をはかる」ことを目的とし、1996年11月に発足した。

会員による研究発表・実践報告の場として、例会（年1～2回）、総会（秋、年1回）を開催するほか、常時メーリングリストで情報意見交換を行っている。

参加を希望される方は、中国語 CAI 研究会 Web <http://moli.cims.hokudai.ac.jp/~ccai/> を参照していただきたい。

Contents

ポッドキャストで中国語	清原 文代	100
Canon wordtank G90	田邊 鉄	111
CASIO EW-V3500L		
電子辞書というデバイスの行く末	小川 利康	114
「電子辞書座談会」	田邊 鉄	119
W-ZERO3 で中国語	内田 慶市	121

ポッドキャストで中国語

清原 文代（きよはら ふみよ）

1. ポッドキャストとは何か？

ポッドキャストはインターネット放送の一種である。ポッドキャストの「ポッド」はアップルの携帯用デジタル音楽プレーヤー iPod¹⁾ から、「キャスト」は broadcast から来ている。だが、名称の一部が iPod に由来するからと言って、iPod でなければ視聴できないわけではなく、他社製のデジタル音楽プレーヤーでもかまわないし、デジタル音楽プレーヤーがなくても、パソコンで視聴することができる。

2. ポッドキャストを受信するには？

2.1. 受信ソフトウェアのインストール

ポッドキャストを受信するためには受信ソフトウェアが必要である。代表的なものには以下のようなものがある。

- iTunes（フリーウェア、Windows2000/XP・Mac OS X 用）
<http://www.apple.com/jp/itunes/download/>
- Alligator（フリーウェア、Windows2000/XP・Mac OS X 用）
<http://www.blogcasting.jp/alligator/index.html>

2.2. ポッドキャストを探す

ポッドキャストを専門に紹介するポータルサイ

トが最近増えてきている。以下にその一部を挙げる。

● Podcasting Juice

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

パソコン通信時代から続く老舗 @nifty によるもの。

● Yahoo! ポッドキャスト

<http://podcast.yahoo.co.jp/>

ネット企業の最大手 Yahoo! Japan によるもので、現在はまだ β 版。一部の機能は Windows でないと使えない。

● Podcast navi

<http://www.podcastnavi.com/>

● podcast journal

<http://podcast-j.net/>

● pod academy

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

学習・教育用ポッドキャスト専門のポータルサイト。内容の充実はまだこれからといったところ。

● 菠萝网

<http://www.podlook.com/discover.aspx>

中国でも“播客”（ポッドキャスト）のポータルサイトが複数立ち上がっているが、これはその一つ。

その他 Web サイトではないが、2.1. で紹介した iTunes には Music Store という音楽のダウンロード販売機能があり、その中に Podcasts というポッドキャストを紹介したディレクトリがある（ポッドキャストの利用は無料）。ポッドキャストの検索・登録・受信が全て iTunes の中で行なえて便利である（図 1）。

2.3. ポッドキャストの登録

当該ポッドキャストを紹介した Web ページにある登録用のリンクやバナーをクリックする、或いは登録用バナーを iTunes 等の受信用ソフトウェアのウィンドウにドラッグするという形で行うことがほとんどである。また配信用 XML ファイルの URL がわかるのであれば、URL を受信用ソフトウェアに直接登録するという方法もある^[2]。



図 1 iTunes Music Store 内の Podcasts ディレクトリ

3. 受信者から見たポッドキャストの利点

従来のインターネット放送はストリーミング放送であり、視聴のためには毎回 Web サイトを訪ねなければならない。またそこで流れる音声や映像ファイルは原則としてハードディスクに保存することはできない。それに対してポッドキャストは以下のような特徴がある。

- 最初に 1 回登録すれば、以後は iTunes 等の受信用ソフトウェアが定期的に更新がないか自動的にチェックするため、毎回 Web サイトに行く手間がない。
- 配信される音声や映像のファイルがパソコンのハードディスクに保存されるので、いつでも視聴できる。
- iPod 等のデジタル音楽プレーヤーに転送することができ、通勤通学の合間にも視聴が可能である。
- ポッドキャストは無料である。

4. 発信者から見たポッドキャストの利点

ポッドキャストの配信の仕組みはシンプルで、ごく普通の Web サーバ上に、配信する素材（音声または映像或いは PDF）と、配信のための情報を書いた XML ファイル、そして当該ポッドキャストを紹介する Web ページを置くだけである。受信に際してはユーザー側のパソコンにイン

ストールされた受信用ソフトウェアが XML ファイルを読んで、更新分のファイルをダウンロードするという仕組みになっている。

従来のインターネット放送はストリーミング用のサーバを立てなければならぬ等、個人レベルでの制作はまず無理であったが、ポッドキャストであれば個人でも制作が十分可能である。Web サーバにしても自前でサーバを立てる必要はなく、Web サーバ上にファイルを置けるディレクトリを持っていればそれでよい。個人のできるインターネット放送、それがポッドキャストである。

5. ポッドキャストの教育利用

ポッドキャストでは音声と映像に加えて iTunes 用の拡張タグを使えば音声と共に 2000 字までのテキストを配信でき、このテキストは iPod 上でも表示される。そしてこれはあまり知られていないのであるが、PDF も配信することができる（PDF は iPod 上では表示できず、パソコン上でのみ表示可能^[3]）。教員が話して学生が聞くという伝統的なスタイルの授業で使う教材としては音声・映像・プリントがあるが、それらの全てをポッドキャストを通じて配布することが可能なのである。

ポッドキャストの教育利用はアメリカではすでに進んでおり、有名なところではスタンフォード大学が講義等を配信している^[4]。日本でも高知



図2 iTunes Music Store内のPodcasts ディレクトリ教育部門のランキング

日本の大学による中国語学習ポッドキャストとしては、関西大学中国語教材研究会メディア教材開発グループによるビデオ教材「Chinese Station」^[8]、そして私自身が制作に関わっている「大阪府立大学の初修外国語 Podcast」（後述）がある。中国語ニュースには「NHK 华语新闻」^[9]、「德国之声中文广播」^[10]があり、両者とも以前からストリーミングによるインターネット放送を実施していたが、ポッドキャストによる配信も行いうようになった。その他にも語学学校や外国語教材を出版する出版社によるもの、個人が制作しているものが多数ある。

大学が地元ラジオ局で放送している市民向け講座をポッドキャストでも配信しており^[5]、今年4月には東京大学がポッドキャストによる講義ビデオの配信を始め^[6]、マスコミにも取り上げられた。また、愛媛大学は高校生向けの大学紹介ビデオをポッドキャストで配信しており^[7]、大学広報の手段としての利用もある。その他、個人や団体が配信する教育ポッドキャストがある（図2）。

7. ポッドキャスト外国語学習番組の2つの方向性

今回後述のように自分でポッドキャストを作るという経験を通して、ポッドキャスト外国語学習番組には2つの方向性があるのではないかと考えている。

7.1. 理解するための番組

学習者の母語による解説が入る番組、例えばNHKの外国語講座のように、当該外国語を話せる日本語話者と、日本語を話せる当該外国語の母語話者のコンビによる掛け合いで進むタイプのものである。学習者にとっては、母語による解説を聞いて外国語を十分に理解し、また母語話者と日本語話者の対話からその言語が話されている国や地域の文化の一端を知ることにもできる。文化に対する興味は外国語学習への意欲を高めることになる。これは言わば理解するための教材である。子供と違って成人の学習者の場合は、学習環境や動機にもよるが、わけもわからず外国語を憶えるということは一般に非常に困難であり、母語を媒介にして外国語をきちんと理解したいという欲求は常に存在する。このタイプは従来からあるものであるが、ポッドキャストで配信することにより、放送時間に縛られない、録画録音の手間がない、どこへでも持ち歩けるという利点に加わる。ポッドキャストによる外国語学習番組においてもこの

6. ポッドキャストによる外国語学習番組

ポッドキャスト番組の中で外国語学習番組は有力なコンテンツである。ポッドキャストを専門に紹介するポータルサイトは複数存在するが、必ずと言っていいほど語学部門が設けられている。これはポッドキャストが無料であるということもさることながら、日々の継続的な学習が欠かせない外国語学習と一旦登録すれば自動的に継続配信されるポッドキャストとの相性が非常に良いこと、そしてデジタル音楽プレーヤーに転送して持ち運べ、思い立った時に学習ができるからであろう。更に学習者向けの番組以外にも、当該言語が話される国や地域のニュース番組やトーク番組等、所謂「生の外国語」に触れられる番組があることも魅力の一つである。

外国語番組のうち圧倒的多数を占めるのはやはり英語であるが、中国語の番組も増加中である。

タイプが主流を占める。

7.2. 繰り返し聞くための番組

成人の外国語学習の落とし穴は理解できたことに安心してしまい、それを文字通り身につけるための反復練習をつい怠ってしまうことである。上述のような理解するための番組はそのような反復して聞くという用途にはあまり向かないのではないかと思われる。なぜならば解説やトークを一端聞いて理解した後で果たして何度も繰り返して聞く気になるのか？それも解説入りという性質上一定の長さを持ったものをとということである。そこでその欠を補うものとして、1つ30秒程度の短いものを繰り返して聞く、何度もCMを耳にしているうちに自然とその中のフレーズや歌を憶えて口ずさんでしまうことがあるが、そのような聞き方をするポッドキャスト外国語学習番組の作り方があってはならないか？こういうタイプの市販のCD教材もあるが、ポッドキャストであれば自動配信されるので教材が徐々に増えていき聞き飽きない。また iTunes と iPod の組み合わせであれば、特に重点的に聞きたいものを手動でプレイリストに入れて聞く、或いは設定した検索条件に従って自動的にプレイリストを更新するスマートプレイリストを使って再生回数の少ないものだけ集めて聞く等のCDにはない柔軟な聞き方が可能だ（図3）。ポッドキャストと携帯用デジタル音楽プレーヤーの組み合わせというメディアの特徴を生かすことによって、学習効果が更に上がることが期待される。

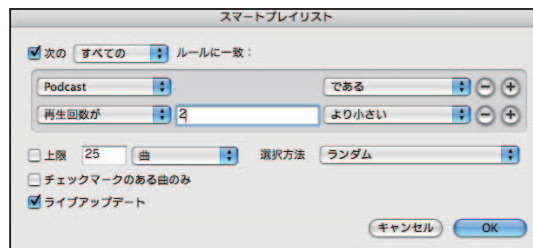


図3 iTunesのスマートプレイリストを使って、ポッドキャストであって、且つ再生回数が2回未満という条件を設定したところ

語で100個を選び、それを当該言語に翻訳、日本語と当該言語で吹き込んである。今年4月から配信を始め、週に2～3文ずつを配信している。

●「大阪府立大学初修外国語 Podcast」ホームページ（図4）

<http://www.las.osakafu-u.ac.jp/podcast-lang/index.html>

8.2. 制作の経緯

iPod 発売以後、個人的に外国語教育機器としての可能性に注目していたのだが、ポッドキャストの登場によりその可能性はますます大きくなったと感じ、まずは臆より始めよということで、2つほど自分でポッドキャストを作ってみるに及んで^[11]、それは確信に変わった。そこで職場の同僚に呼びかけて作ったのが「大阪府立大学初修外国語 Podcast」である。

「大阪府立大学初修外国語 Podcast」は前述の繰り返しして聞く番組のスタイルをとっている。こ

8. 「大阪府立大学初修外国語 Podcast」

8.1. 概略

大阪府立大学総合教育研究機構第二教室に所属する初修外国語（ドイツ語・フランス語・中国語・朝鮮韓国語）を担当する教員による音声ポッドキャストである。当該言語が話されている国や地域に旅行した際に役立つと思われる表現を日本

図4 「大阪府立大学初修外国語 Podcast」のホームページの一部



これは資金ゼロ、制作に使える時間の短さや技術的なことをする人手が足りないという物理的条件の制約があったせいもあるが、我々は普通の授業で説明はできるのだから、普通の授業ではできないもの、普通の授業を補完するものとしてポッドキャストと利用しようという意図もあった。

☞ 8.3. 制作して見えてきた問題点

■ 8.3.1. 受信側の環境をどう担保するか？

授業の補完を意図したものの、実際には授業でこういうものがありますよと紹介しただけにとどまっている。これは3月1ヶ月間で慌ただしく制作し新学期に入ることになってしまい、授業との連携を十分考慮できなかったことが原因であるが、受信する学生側の環境がこちらが思うほどには整っていなかったこともある。iPodをはじめとするデジタル音楽プレーヤーの普及率はまだCDプレーヤーほどではない。しかし、カセットテープからCDへという音声メディアの変遷の歴史を考えれば、これは早晚時間が解決する問題であろう。問題はパソコンでの視聴であった。デジタル音楽プレーヤーの普及率に比べればパソコンの普及率は高い。ポッドキャストはパソコンで聞ける、というよりはまずパソコンで受信してデジ

タル音楽プレーヤーに転送するので、デジタル音楽プレーヤーがなくても自宅のパソコンで聞けるだろうと思っていたのであるが、パソコンの基本的な操作自体に習熟していない、Windows98等古いOSを使っていてiTunesがインストールできない等、これが案外難しかったようである。学生のほぼ全員が受信できる状況でなければ授業と連携して用いることは難しく、聞ける人は聞いてくださいという扱いに終わってしまった^[12]。

しかし、ポッドキャストは授業に出ている学生のためにだけあるわけではない。インターネットに接続できる環境があれば誰でもが利用でき、潜在的聴取者は世界中に存在しうることになる。ささやかではあるが、我々が持っている教育資源を直接教えている学生だけでなく、教育機関を支えている市民や社会に還元するということではできないかと思う。

■ 8.3.2. 教材内容に関するもの

100個の文例の原案は私が提出したのであるが、以下のWebページの100個の文の一覧表を見ていただければおわかりのように、100個の文の中にはトラブル時に使用する表現が多く含まれる(図5)。

図5 「大阪府立大学初修外国語Podcast」に含まれる文の一部

	日本語	ドイツ語	フランス語	中国語	朝鮮語韓国語
16	助けて!	Hilfe!	Au secours!	救命啊!	도와 주세요!
17	警察を呼んで!	Polizei!	Appelez la police!	快叫警察!	경찰 불러 주세요!
18	医者を呼んで!	Einen Arzt!	Appelez un médecin!	快叫医生!	의사 불러 주세요!
19	泥棒!	Ein Dieb!	Au voleur!	抓小偷!	도둑이야!
20	火事だ!	Feuer!	Au feu!	着火啦!	불이야!
21	はやく逃げろ!	Weg! Schnell!	Ne restez pas là! Eloignez-vous!	快跑!	빨리 피하세요!
22	日本大使館に連絡してください。	Informieren Sie die japanische Botschaft!	Est-ce que vous pourriez avertir l'Ambassade du Japon, s'il vous plaît?	请帮我跟日本大使馆联系。	일본대사관에 연락해 주세요.
23	パスポートをなくしました。	Mein Pass ist weg!	J'ai perdu mon passeport!	我的护照丢了。	여권을 잃어 버렸어요.
24	パスポートが盗まれました。	Mein Pass wurde gestohlen.	On m'a volé mon passeport!	我的护照被偷了。	여권을 도둑 맞았어요.
25	財布をなくしました。	Mein Portemonnaie ist weg!	J'ai perdu mon portefeuille!	我的钱包丢了。	지갑을 잃어 버렸어요.
26	財布が盗まれました。	Mein Portemonnaie wurde gestohlen.	On m'a volé mon portefeuille!	我的钱包被偷了。	지갑을 도둑 맞았어요.
27	荷物が見つかりません。	Mein Gepäck ist nicht da.	Je ne trouve pas mes bagages!	我的行李找不到了。	제 짐이 안 보이는데요.
28	荷物が盗まれました。	Mein Gepäck wurde gestohlen.	On m'a volé mes bagages!	我的行李被偷了。	짐을 도둑 맞았어요.

<http://www.las.osakafu-u.ac.jp/podcast-lang/4/100/4langs.html>

これが一部の教員にかなり強硬に反対されてしまった。このような文があると外国へ行けばそのような危ない目に遭うのだという印象を学生に与えてしまい良くないというのがその主旨であった。外国へ行けば必ずそのような目に遭うわけでは決してないのであるが、逆に言えば遭わないという保証もない。教科書に出てくる場面は何事もスムーズに進むというものが大半であるが、それではトラブルに遭ったときに対処しきれない。トラブルに陥った時にこそ自分を救うために言葉が必要だというのは、私自身の経験から得た実感である。教員としてはその言語や文化に興味を持って学んで欲しい、良いところやすばらしいところを知って欲しいと願うのは当然であり、反対意見も理解できる。私自身もこれまで何冊か教科書を書いているが、学生にとってはもしかしたら一生のうち中国語に触れる最初で最後の機会かもしれないと思うと、トラブル時の表現の必要性を感じながらもそれをスキットのような形で書くことを躊躇するところがある。スキットのような形になるとストーリーがあるが故に与える印象が強くなりがちだからだ。しかし、今回のようなごく短い文を細切れに流すポッドキャストであれば、さほど悪印象を植え付けずに、必要な表現だけを紹介できるのではないかと考え、教科書では出しにくいものをポッドキャストでやるのだと言って押し切ってしまったが、それが果たして学習者にとって良かったのかどうか、今のところ検証できていない。

9. 音声ポッドキャスト制作

それでは、以下にどのようにポッドキャストを制作するのかその概略を紹介する。

9.1. ポッドキャスト制作の難度

映像をポッドキャスト配信するとなると、撮影や編集のための機材もそれなりに必要になるが、音声だけのポッドキャスト及びPDFの配信であ

れば、最低限パソコン1台あれば済むため、個人レベルで十分作成できる。作業の難度について言えば、WebページをHTMLタグを打って作ったことがある方であれば、まず間違いなく作れる。Webページ作成の経験がなく、Webは見るだけ、後はワープロとメールという使い方をしている方でも、ポッドキャスト対応のブログサービス（後述）を利用すれば作成可能であろう。

9.2. 音声の録音

音源については当然のことながら教科書付属のCDや市販の音楽CDを使うことはできず、著作権上問題のないものを用意しなければならない。一番良いのは自ら録音することである。音声はパソコンのハードディスクに直接録音する。パソコンの内蔵マイクを使ってもかまわないが、より良い音質を目指すのであれば外部マイクを使用した方がよい。録音用ソフトウェアとしては例えば以下のようなものがある。

- 超録—パソコン長時間録音機フリーウェア版 (Window用)

<http://www.vector.co.jp/soft/win95/art/se297145.html>

- Audacity (フリーウェア、Window & Mac OS X用)

<http://audacity.sourceforge.net/>

- Sound it! (パッケージソフト、Window & Mac OS X用)

<http://www.ssw.co.jp/products/sit/win/sit50w/index.html>

- GarageBand3 (Mac OS X用パッケージソフト iLife'06に含まれている。現在販売中のMacにはプリインストールされている) (図6)

<http://www.apple.com/jp/ilife/garageband/>

録音したものはMP3またはAACという形式で保存する。AACは1つの音声に複数のチャプターを設定したり、iTunesやiPod上でチャプターごとに画像を表示したりすることができるが、再生

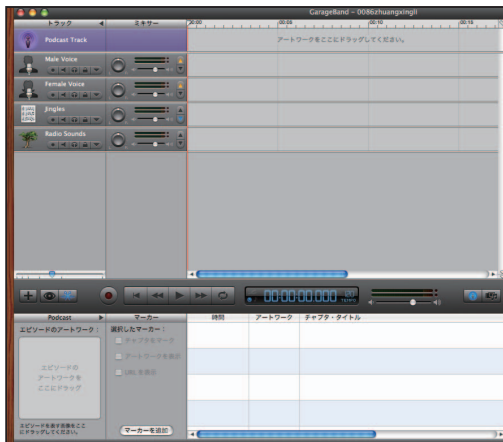


図6 GarageBand3のポッドキャスト用プリセット

環境が実質的に iTunes と iPod に限られてしまうため、汎用性という点からみれば MP3 をお勧めする。MP3 のビットレートの高さや音質、そして音声ファイルの大きさは比例するが、音楽を含まない声だけのものであれば、ビットレートはどんなに高くても 128kbps、96kbps もあれば十分である。

9.3. XML ファイルの作成——手作業で行う場合

9.3.1. XML ファイルの内容

ポッドキャストを配信するためには XML ファイルが必要である。XML ファイルと言っても、実質的には文字コード UTF-8 で保存されたテキストファイルに .xml という拡張子を付けたもので、UTF-8 を読み書きできるエディタ、例えば OS 付属のエディタ（Windows ではメモ帳、Mac OS X ではテキストエディット^[13]）で作成すればよい。

XML ファイルの中身は RSS2.0 + iTunes 用のタグで、その仕様は以下のアップル社の Web ページで公開されている。

<http://www.apple.com/jp/itunes/podcasts/techspecs.html>

但し、このページの末尾にあるポッドキャストのカテゴリ分けについては、最近行われた改訂がまだ反映されていない。最新のカテゴリにつ

いては以下の英語版の方で確認されたい。

<http://www.apple.com/itunes/store/podcaststechspecs.html>

以下にサンプルとして「大阪府立大学初修外国語 Podcast」の XML ファイルの一部を挙げておく。<!-- --> で囲まれた部分は解説であって、実際の XML ファイルには記入しない。

```
<?xml version="1.0"
encoding="UTF-8"?>
<rss xmlns:itunes="http://www.
itunes.com/DTDs/Podcast-1.0.dtd"
version="2.0">
<!-- 冒頭で必ず必要な XML 宣言。 -->

<channel>
<!-- 以下は番組全体の情報。 -->

<title> 大阪府立大学中国語講座—ぜひ憶
えておきたい 100 の表現— </title>

<link>http://www.las.osakafu-u.
ac.jp/podcast-lang/index.html</
link>
<!-- ポッドキャストの説明用 Web ページの URL。
-->

<copyright> 清原文代 </copyright>
<language>ja</language>
<!-- ポッドキャストに使用される主な言語、こ
の場合には日本語を指定している。 -->

<!-- 以下、<itunes: で始まるタグは iTunes 上で
表示される情報のためのもの。 -->
<itunes:author> 清原文代 & amp; 顧春
芳 </itunes:author>

<itunes:subtitle> 大阪府立大学中国
語講座—ぜひ憶えておきたい 100 の表現—
</itunes:subtitle>
```

```
<itunes:summary> 大阪府立大学総合教育  
研究機構第二教室に所属する初修外国語担当  
教員による中国語会話講座です。旅行やトラ  
ブルに遭遇した時に使える表現を 100 個紹  
介します。</itunes:summary>
```

```
<itunes:owner>
```

```
<itunes:name> 清原文代 </itunes:  
name>
```

```
<itunes:email>kiyohara@las.  
osakafu-u.ac.jp</itunes:email>
```

```
</itunes:owner>
```

```
<!-- メールアドレスは iTunes 上では表示されな  
い。iTunes Music Store からの連絡用。-->
```

```
<itunes:image href="http://www.  
las.osakafu-u.ac.jp/podcast-lang/  
logo.jpg" />
```

```
<!--iTunes 上で表示したいポッドキャストの口  
ゴ画像の URL。300 × 300 ピクセル以上の正方  
形の JPEG 画像推奨。-->
```

```
<itunes:category  
text="Education">
```

```
<itunes:category text="Language  
Courses"/>
```

```
</itunes:category>
```

```
<itunes:category  
text="Education">
```

```
<itunes:category text="Higher  
Education"/>
```

```
</itunes:category>
```

```
<!--iTunes Music Store の Podcasts ディレクトリ  
のためのカテゴリ情報。カテゴリは最大 3  
つまで指定できるが、一番上に書いたカテゴリ  
が優先される。-->
```

```
<item>
```

```
<!-- 以下この番組に含まれる各エピソード（音
```

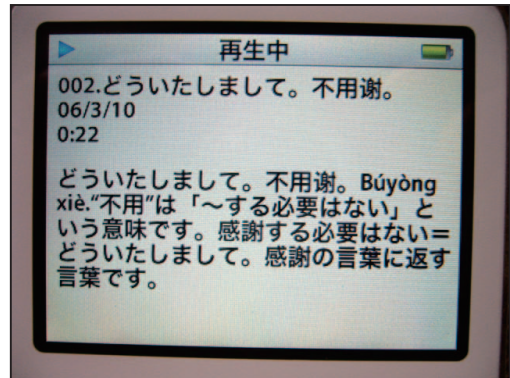


図7 iPod で summary を表示させたところ

声) の情報。この item タグで囲まれた部分を増やしていくことによって新しいエピソードを配信する。-->

```
<title>002. どういたしまして。不用谢。  
</title>
```

```
<itunes:author> 清原文代 & amp; 顧春  
芳 </itunes:author>
```

```
<itunes:subtitle> どういたしまして。  
不用谢。Búyòng xiè.</itunes:subtitle>
```

```
<itunes:summary> どういたしまして。不  
用谢。Búyòng xiè。“不用”は「～する必要は  
ない」という意味です。感謝する必要はない  
＝どういたしまして。感謝の言葉に返す言葉  
です。</itunes:summary>
```

```
<!-- この summary の部分に 2000 字までのテキ  
ストを入れることができる。summary の中で改  
行するには、&#13;を挿入する。この summary  
は iTunes 上では各エピソードの横について  
①というアイコンをクリックすることによって  
表示される。iPod 上では再生中に真ん中の決定  
ボタンを 2 回ほど押すと表示される。--> (図7)
```

```
<enclosure url="http://www.las.  
osakafu-u.ac.jp/podcast-lang/  
zh/100/zh002.mp3" length="359122"  
type="audio/mpeg" />
```

```
<!-- 配信するファイルのありかを示す URL。ファ
```

イル大きさはバイト単位。ここのファイルタイプは MP3 のもの、PDF であれば application/pdf とする -->

```
<guid>http://www.las.osakafu-u.ac.jp/podcast-lang/zh/100/zh002.mp3</guid>
```

```
<pubDate>Fri, 10 Mar 2006 12:45:00 +0900 </pubDate>
<itunes:duration>0:22</itunes:duration>
<itunes:keywords>Chinese, Mandarin, 中国語, 汉语, 大阪府立大学</itunes:keywords>
```

```
<itunes:explicit>no</itunes:explicit>
<!-- 露骨な表現が含まれているかどうかの表示、ペアレンタルコントロール[14]のため。-->
```

```
</item>
```

```
</channel>
```

```
</rss>
```

```
<!-- ファイルの末尾に必要なタグ。-->
```

■9.3.2. XML ファイル作成用ユーティリティソフト

このような XML ファイルの作成を支援するユーティリティソフトもある。

●Digion PodCreator（シェアウェア、Window2000/XP 用）

<http://www.digion.com/pc/pro/dpc/index.htm>

●Podcast Tool（シェアウェア、Mac OS X 用）

<http://www.sonoran.co.jp/podcast/>

■9.3.3. その後の作業

音声ファイルや PDF を XML ファイルに書いたディレクトリにアップロードし、XML ファイルもアップロードする。

ポッドキャストを紹介する Web ページを作り、そこに XML ファイルの URL を表記する等、登録方法を紹介します。

☞9.4. XML ファイルの作成——ポッドキャスト対応のブログを利用する場合

ポッドキャストに対応しているブログサービスを利用すれば、XML ファイルを自分で書かずにすむ上に、ポッドキャストを紹介する Web ページも同時に作れ、ポッドキャスト登録用のバナーも用意される。但し、PDF の配信には対応していないので、音声に加えて PDF を配信したい場合は、XML ファイルを手書きしなければならない。

ポッドキャストに対応しているブログには以下のようなものがある。

●ケロログ < <http://www.voiceblog.jp/>>

●Seesaa < <http://blog.seesaa.jp/>>

●ココログ < <http://www.cocolog-nifty.com/>>

ケロログと Seesaa は日本語の文字コードにし対応していないので、Web ページ上で中国語を書きたい場合は、ココログを使う。ココログの文字コードは UTF-8 であり、日本語と中国語を混在させた Web ページを作れる（図 8）^[15]。

■9.4.1. ココログでの大まかな手順

- (1) ココログにユーザー登録する。広告が強制挿入される無料のココログフリーと広告が入らない有料版がある。
- (2) ブログを作る。
- (3) 初期設定ではポッドキャスト配信機能が OFF になっているので、ON にする。「ブログ」タブをクリック→「デザイン編集」をクリック→「表示項目を変更」をクリック→「ポッドキャスト」にチェックを付け、変更を保存する。
- (4) 音声ファイルを記事の 1 行目にアップロードする。

- (5) 記事に音声ファイルに関する説明を書く。
- (6) ブラウザによっては中国語の部分が文字化けすることがある。文字化けを避けるには中国語の部分に HTML タグを付ける。記事の作成画面で「HTML の編集」タブをクリックして HTML 編集画面に入り（図 9）、中国語の部分を言語指定タグで囲む（例 1）。

```
(例1) <span lang="zh"> 请再说一遍。
</span>
```

- (7) 拼音字母を入力した場合は、これもブラウザによる表示の違いを減らすために、フォントのスタイルを指定しておいた方がよい。
- 記事の作成画面で「HTML の編集」タブをクリックして HTML 編集画面に入り、拼音字母の部分について Span 要素を設定し、style 属性を用いて、フォントのスタイルを指定しておく（例 2）。

```
(例2) <span style="font-family:
'Times New Roman'; font-style:
normal; font-variant: normal;
font-weight: normal; font-size:
x-large;">Qǐng zài shuō yí biàn.</span>
```

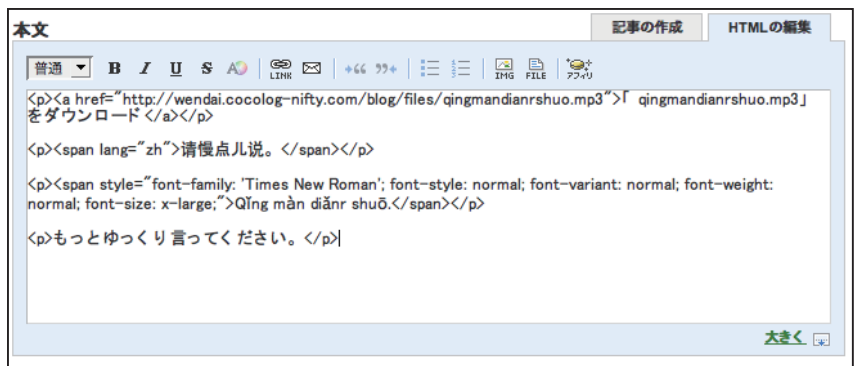


図 8 ココログフリー版の画面。右下隅の「RSS Podcasting」というパナーを iTunes のウィンドウにドラッグしてポッドキャストを登録する。

10. 小結

ポッドキャスト（特に音声ポッドキャスト）は、会社や団体でなければ負担できないような多額の資金や専門家を雇わなければならないような高い技術水準を要求しない。言わば個人が初めて手に入れた放送メディアであり、その視聴者はインターネットにつながっている限り国境を越えて全世界に広がる。また、一旦登録すれば自動的に継続配信されること、及びデジタル音楽プレーヤーに転送して常に携帯して手軽に視聴できるという利便性があり、これは日々の継続的な学習がとりわけ重要な外国語学習にとって非常にありがたい。

図 9 ココログフリー版の HTML 編集画面



無論、学生の受信環境をどう担保するか等の問題があり、ポッドキャストは万能のメディアではないが、このような特長を備えたポッドキャストの教育利用は今後益々拡大していくと考えられる。本稿を読んで、中国語学習にポッドキャストを活用してみよう、或いは中国語学習ポッドキャストを自分で制作してみようという方が増えれば幸いである。

注

- [1] <http://www.apple.com/jp/ipod/ipod.html>
- [2] iTunes では「詳細設定」メニューから「Podcast を登録」を選び、XML ファイルの URL を入力する。
- [3] iTunes には PDF を直接表示する機能はなく、Adobe Reader (Acrobat Reader) 等の外部アプリケーションを呼び出して表示する。
PDF はテキスト・画像・図表だけではなく、音声やムービー、Java Script を内包することができるため、PDF を配信することができることは、ポッドキャストの教育利用の可能性を大きく広げる。
なお、Mac OS X は初期設定では OS 付属のプレビューというソフトで PDF を表示するが、プレビューは PDF に添付された音声やムービー、Java Script を取り扱うことができないので、Adobe Reader をインストールし、Adobe Reader で PDF を表示するように設定しておく必要がある。
- [4] <http://itunes.stanford.edu/>
- [5] <http://www.kochi-u.ac.jp/~wwwlife/radio/index.html>
- [6] <http://ocw.u-tokyo.ac.jp/podcasts/>
- [7] <http://www.ehime-u.ac.jp/shokai/dvd/index.html>
- [8] <http://podchina.seesaa.net/>
- [9] <http://www.nhk.or.jp/nhkworld/chinese/index.html>
- [10] http://www.dw-world.de/dw/0,,293,00.html?maca=ch-podcast_radio_china-1169-xml-mrssl
- [11] 「清原の中国語教室」
<http://homepage.mac.com/qingyuan/Podcast/index.html>
PDF 配信の実験のために作ったもの。現在は更新していない。
《汉语动词三字经》
<http://web.mac.com/qingyuan/iWeb/>
現在もほぼ 3 日に 1 度の割合で更新している。
- [12] 学生に最も普及している端末は携帯電話である。教材の配信を考える上において携帯電話への展開は今後欠かすことができないであろう。
ポッドキャストを携帯電話に配信する「キャスピー」 というサービスが立ち上がっている。
<http://caspee.jp/>
但し、対応機種に制限がある。試験的に 1 つ作ってはみたものの、私の携帯電話ではアクセスできないため、検証できていない。
<http://caspee.jp/channels/qingyuan/>
- [13] テキストエディットのデフォルトのファイル形式は RTF なので、フォーマットメニューから「標準テキストにする」を選んでプレーンテキストファイルにする。
- [14] iTunes には iTunes Music Store 内の露骨な表現を含むコンテンツに子供がアクセスできないように親が設定できる機能がある。iTunes Music Store 内の Podcasts ディレクトリにポッドキャストを登録するには、このタグは必須である。
- [15] ココログのフリー版に筆者が試験的に作ったものがある。
<http://wendai.cocolog-nifty.com/>

Canon wordtank G90

田邊 鉄（たなべ てつ）

📎 収録辞書数 No.1

今年4月に発売されたCanon wordtank G90(以下G90)は発売直後のリコール騒ぎ^[1]で、いきなりミソをつけた格好だが、少なくとも収録される中国語辞書の数では、日本国内向けに販売されている電子辞書の中で随一の充実ぶりを誇る。中国語関係だけで以下の10コンテンツ^[2]を収録している。

- 中日大辞典・増訂第2版（愛知大学/大修館）
- 中日辞典（講談社）
- 日中辞典（講談社）
- 現代漢語詞典第4版（中国社会科学院）
- 中国語新語ビジネス語辞典（大修館）
- 現代中国語新語辞典（書籍なし）
- 日中 PC 用語辞典（日経 BP）
- 中国語会話とっさのひとこと（DHC）
- 英中辞典（Oxford）
- 中英辞典（Oxford）

辞書の数や種類からみて、比較の対象となるのはG90発売の少し前、本年2月に発売されたカシオのXD-GT7350あたりだろうか。音声機能がついているので、V90と比較するのが正しいかもしれないが、こちらは以下の8コンテンツである。

- 中日大辞典・増訂第2版（愛知大学/大修館）
- 中日辞典（小学館）
- 日中辞典（小学館）
- 中国語新語ビジネス語辞典（大修館）

- 日中 PC 用語辞典（日経 BP）
- ひとり歩きの中国語自遊自在（JTB）
- 日中英・電子技術用語対訳集（日経 BP）
- 日中英固有名詞辞典

『中日大辞典』は高電社のPC用中国語入力ソフト「Chinese Writer V8」付属の辞書として電子化されており、XD-GT7350から電子辞書にも収録されるようになった。「ハイエンド機には『大辞典』を」という流れができつつあるのかもしれない。この辞書は紙版が特に大きく重いので、頻繁に用いるなら電子辞書に収録されるのはありがたい。

メインの中日・日中辞典に、小学館ではなく講談社を採用しているのは、今のところG90だけだ。外国語電子辞書では「日本語↔○○語」が双方向で揃っていないと、営業的に厳しいと言われる。そのためこれまでは小学館一辺倒だったのだが、講談社『日中辞典』の刊行は、電子辞書のバリエーションを増やし、選択の幅を広げてくれたようだ。講談社の『日中辞典』はG90が発売される少し前、今年3月に、全データのCD-ROM付きで発売されている。してみると、講談社の『日中』は、電子辞書に収録するための辞書、「はじめに“電子化”ありき」という戦略商品として開発されたのではないか——という見方もできる。

XD-GT7350には収録されていない日中辞典として、G90は『現代漢語詞典』を収録。第5版はさすがに無理だったようだが、業界初である。

日中辞典の「マルチ構成」や日中辞典の収録は、英語以外の電子辞書としては異例の充実度である。ビジネス・新語の『小辞典』は、一見似たよう

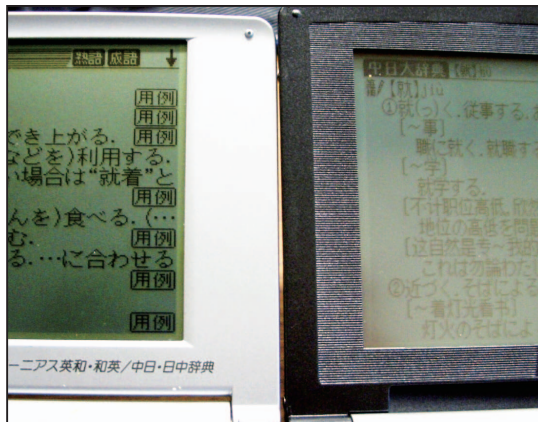


図1 G90（右）は明らかに見にくい

なラインナップだが、G90はPC+新語+ビジネス用語、XD-GT7350はこれらに加え、電子技術（半導体技術など）の用語集と固有名詞辞書を備える。

こうしてみるとXD-GT7350は「中国語を必要とするプロ（技術者やビジネスマン等）」、G90は「中国語のプロ（研究者や翻訳者）」を主なターゲットにしているように見える。ただ、そうなる（好きモノ？）研究者には、G90の収録辞書は物足りなく感じられる。『英中』『中英』を削って中中辞典を充実させるとか、専門辞書を載せるとかといったことも考えられたのではないかな。

実際のところ、G90は「（日本に留学している）中国人向けの機種」なのではないかと思う。それならば中中辞典の収録も『英中』『中英』の必要性も理解できる。

相互参照と串刺し検索のできる電子辞書は、「○日辞典」と「日○辞典」との距離を限りなく近づけた。今後、「日本製」「中国製」「日本向け」「中国向け」といった区別も溶けていくのか。興味深いところではある。

④ 外観

148×104.5×19.7、288gの筐体は、XD-GT7350とほぼ同じであるが、厚みがあるので、ずっしりとした印象を与える。

二つ折りの液晶側とキーボード側の端が、互い

違いに波形をしており、そこに指をかけて引き開けるのだが、開け方を理解するまでに時間がかかった。

キーボードはごく普通のQWERTY式。丸いキーは好みの分かれるところだろうが、十分なプッシュ感がある。キー間隔は十分とっており、押し間違いの心配は少ない。

液晶の周囲が黒いのは、おそらく液晶画面の問題（後述）を多少なりとも軽減しようとしたのだろうが、高級感のある「深い黒」ではなく、灰色っぽいため、安っぽく見える。キーボード面と同じシルバーにするか、黒に統一するかした方が、デザイン的にはいいと思う。

④ 液晶画面

5.2インチ高精細液晶は、480×240ドットのワイド画面。カシオなどが採用する480×320ドットには及ばないが、画面に表示される情報量や細かさは十分だと思う。

問題は、あちこちで指摘されているとおり、液晶画面が「暗い」ことである。図1を見て明らかなように、G90はコントラストが低く、全体に暗く、写り込みが激しい。椅子に座り、机の上に本と並べて置いて参照する、という普通の使い方が難しいほどだ。手に持って使えばそれほどでもないが、薄暗い室内・電車内などでは視認性が極端に落ちる。もともとCanonの電子辞書は手書き対応のためにタッチスクリーンを搭載しており、普通の液晶よりも若干視認性が落ちる。けれども、同様にタッチスクリーンを搭載したV80、V70と比較しても、G90の画面は特に見にくく感じる。価格に直結するだけに難しいだろうが、もう少し見やすい液晶を使ってほしい。

④ 起動・終了画面

起動時と終了時にはそれぞれ「Welcome！」「Bye Bye！」とスプラッシュ・スクリーンが表示される。起動時は「待たされる」というほどではないが、若干もたつく感じがする。気になるの

は終了画面で、電源ボタンを押してから電源が切れるまで、3秒近く時間がかかる。

使い方にもよるだろうが、電子辞書の身上は「さっと取り出し、ぱっと検索、終わったらすぐにしまって忘れてしまえる」スマートさにある。これはすなわち、電源の入/切を頻繁に行うことを意味する。ど忘れした単語を1個調べる、というだけなら、十分に実用範囲内だとは思いますが、辞書を引き引き読書、というような使い方をしていくときには、結構ストレスがたまりそうだ。

📎 検索機能

中国語の検索は、ピンイン・漢字・音訓・部首・総画数が一通り揃う。「ピンイン」と「漢字」は、どちらもピンインからの検索だが、「漢字」の方は漢字を一文字ずつ確定して検索する。どちらかひとつでよさそうなものだが、「手書き認識」と「ワイルドカード」を活用するためにあえて分けてあるのだろう。

ワイルドカードは、「?」（1文字）と「*」（0文字以上）の二つが使える。このワイルドカードはピンイン・漢字の両方で使うことができるので、ピンイン綴りの曖昧検索もできるし、「?肉?食」のような漢字単位での曖昧検索もできる。これは、ピンイン単位でのワイルドカードしか使えない機種に対しては大きなアドバンテージになる。

さて、G90のウリの一つに「例文検索」がある。語・漢字を含む例文を、列挙してくれる。複数辞書の検索にも対応しているので、中日・日中・中英・英中の各辞書から例文を探し出すことができる。中国語教員には待望久しいこの機能のデキはどうか。操作性はあまりよくない。例文検索は、上述の「漢字」検索と同じく、単漢字入力→確定→次の漢字・・・と入力するので、2文字以上の語は入力が面倒である。また、ワイルドカードが使えないので、「一*就」と言う具合に「一~就」の例文を検索できない。「一&就」として、両方を含む例文を検索できるが、語順は考慮されないので「就~一」を含むものもヒットしてしまう。高電社のPC用ソフト「Chinese Writer」に付属す

る辞書では対応しているので、実現して欲しかった。

便利なのは「一括入力」検索で、5つまでの複数の単語を入力すると、まとめて辞書引きしてくれる。各検索結果は左右のカーソルキーを使うか、画面下部のタブをペンでタップして切り替えながら参照する。この機能は例文検索でも使えるので、テストに使う例文を一発で検索終了、ということもできる（かもしれない）。

📎 その他の操作性

液晶の濃淡調節ダイヤルが付いていること、文字サイズが独立したボタンで変更できることは評価したい。パソコンの影響で、何でもメニュー画面で設定させる機種が増えているが、「専用デバイス」は、それなりのユーザ・インタフェースを持つべきだろう。もっともG90の場合、液晶画面に前述の問題があるため、画面の見え方を頻繁に調整するための機能は必須だとも言える。

「辞書本文→別の辞書で同じ語を引く（辞書リンク）」、「辞書本文→関連語句」、「辞書本文→用例」のいずれもが本当にワン・タッチなのは便利だ。この辺は画面にタッチして操作することもできるが、主な入力方法がキーボードであれば、操作も全てキーボードで行った方がラクだ。

📎 「買い」か「待ち」か

さて、ではこのG90は「買い」だろうか。

今のところ、中国語コンテンツの充実ぶりでは、G90を超える機種は出ていない。講談社の辞書を使っているとか、現代漢語詞典がぜひ必要とかいう具合に、よく使う辞書が含まれているならば、「買い」だろう。

ただ、逆に言えば、中国語辞書の数以外には、見るべき点の少ない機種だということでもある。特定の辞書を独占的に利用する契約が結ばれているとは考えにくいので、いったん電子化された辞書は、遅かれ早かれ他社製品にも収録されることが期待できる。既に他社の電子辞書を使っていて、

収録辞書の種類だけを理由に G90 に買い換えようと思っているなら、「使い慣れた機種の後継機に、欲しい辞書が収録されるのを待つ」のも一つの手だろう。

「中国語のプロ」ならば、日本向け製品ではないが、カシオの EW-V3500L（本誌別記事参照）の『現代漢語大辞典』と『漢英大詞典』の組み合わせの方が、より魅力的であろう。

注

- [1] ピンインの「yi」が検索されない、というトラブルで回収・交換が行われた。
- [2] 取り扱い説明書には「9 コンテンツ」とある。英中と中英が書籍としては 1 冊だからかと思われるが、詳細は不明。

CASIO EW-V3500L

電子辞書というデバイスの行く末

小川 利康（おがわ としやす）

品定め

一昨年は「中英日空中学習辞典」という台湾製電子辞書を買った。Oxford 英漢・漢英辞典に加え、日漢辞典も入っている優れものだ。そのうえパラレルケーブル経由で CF カードに繁体字テキストを転送して読めて、FM ラジオまで聴ける。そんな妖しさに惹かれ、一時期は随分使っていたのだが、いつの間にか使わなくなってしまった^[1]。

結局のところ、日常的に使うには液晶が旧式で見づらかったり（老眼なので）、広辞苑が入っていなかったりで、このところ CASIO XD-R7300 を使い続けてきた。中国語辞書搭載のはしりとも言うべき機種だが、それでも台湾製より遙かに液晶も見やすく、辞書も充実していた。

2006 年になって、電子辞書にも新しい動きが出てきた。特筆すべきは CANON の Wordtank G90 の発売だろう（詳細は別途レビュー記事参

照）。これまで収録された辞書はおしなべて小学館版『日中辞典・中日辞典（初版）』だったが、この機種は小学館版でなく大修館書店の『中日大辞典（第二版）』、商務印書館の『現代漢語詞典』を収録しており、目下最強と評判の辞書だ。ようやく中国語電子辞書もインターフェイスだけでなく、搭載辞書で選べる時代になった。

ここですっかり CANON を購入するつもりになったのだが、中国 CASIO から中日英電子辞書が出ていることを知った。それが CASIO EW-V3500L である。中国市場向けで、日本の小売店では買えない。だが、こちら捨てがたい魅力がある。中中『現代漢語大詞典』は 11 万語、中英『漢英大辞典』は 23 万語で、とくに中英は定評ある辞書だ。これで『中日大辞典』が入っていれば申し分ないのだが、代わりに『クラウン中日辞典』が入っている。もし中中辞典を重視するならば中国版の方が「買い」だが、中日辞典が充実しているのは間違いなく Wordtank G90 の方だ。なかなか

か甲乙つけがたい。だが、最後に液晶で比べると、EW-V3500Lの方が明らかに見やすい。やはり前回の教訓は生かさねばならない。散々迷った末、実売価格が2000円を切った所でEW-V3500Lを購入した。

基本データ

まず基本仕様から確認しよう。

●液晶画面サイズ

4.5インチ液晶（解像度 320 × 240）

日本CASIOのXD-L7350（2004年）とハードウェアとしては同等スペックのようだ。2005年に降出ている5.2インチ（解像度 480 × 320）高精細液晶と比べると、やや情報量が少ないのが気になるが、極端に狭いという感じはしない。慣れの問題だろうか。

●サイズ・重量（電池込み）

奥行 97.5 × 幅 144.5 × 高さ 11.8mm
重量 250g

見た目のコンパクトさに反して、意外と持ち重りする印象。CASIOの最近の音声対応電子辞書は軒並み250gでことさら重いわけではなく、タッチパネルで5.2インチ液晶のCANON Wordtank G90（104.5 × 148 × 19.7mm, 288g）とほぼ同じだ。

語彙数が辞書の優秀さを示すわけではないが、充実度をはかるうえで一番分かりやすい指標である。語彙数から見ても、中国人からすると、『広辞苑』が最も魅力的なコンテンツであることは言うまでもない。ただ、日本人からすると『漢英大辞典』、『日漢大辞典』が最も魅力的なコンテンツとなるだろう。次いで『現代漢語大詞典』、『クラウン中日辞典』となる。

音声については、Longmanの『英漢双解活用詞典』に含まれる英単語6600・『日漢大辞典』に含まれる日本語単語8000を収録するのみの

目録	
岩波書店 広辞苑(日大辞典)	現代汉语大词典
日漢大辞典	クラウン中日辞典
朗文 英漢双解活用詞典	汉英大辞典(第二版)
大学英语词汇手册	多词典查询

按▲/▼/◀/▶键选择辞典，按[输入]键确认

EW-V3500Lメニュー

広辞苑(岩波書店)	約23万語
漢英大辞典(上海交通大学)	約23万語
日漢大辞典(上海譯文)	約18万語
現代漢語大詞典(漢語大詞典)	約11万語
クラウン中日辞典(三省堂)	約6.5万語
英漢双解活用詞典(Longman)	約4.5万語
大学英语(上海交通大学)	約6600語

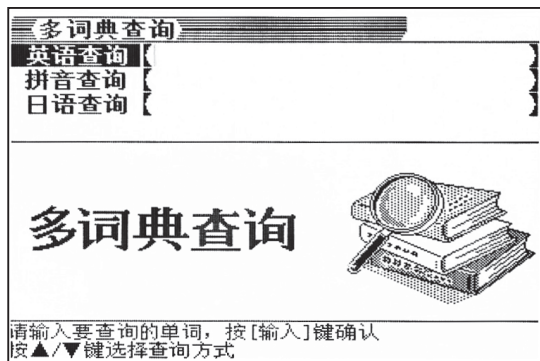
収録辞書と語彙数（カッコ内は出版社）

で、やや魅力に欠ける。中国語は当たり前だが入っていない。音声無しで軽い製品があったら、そちらを買っただろう。だが、現在中国で販売されている製品には英中辞書だけの製品か、英中辞書に法律用語辞典を追加した製品しか売られていない。

次に個々の辞書の特徴を操作性、内容から見てゆく。このEW-V3500Lも最近の電子辞書らしく複数の辞書を横断的に検索できる（中国語では「多词典查询」）。単語一つ調べるにしても、複数の辞書を同時に引けた方が効率が良い。惜しむらくは拼音検索でワイルドカードが使えないことだが²⁾、以下ではこの機能を使って検索することを前提に、中国語、日本語それぞれについて見てみよう。

中国語から検索

中国語から検索する場合は、次ページの図のように拼音を入力して検索をかける。検索対象となる辞書は三種類ある。



複数辞書検索メニュー

- (1) 汉英 漢英大辞典
- (2) 汉语 現代漢語大詞典
- (3) 汉日 クラウン中日辞典

検索の結果が辞書ごとに示され、下にはプレビュー用に辞書の語彙が提示されるので、読み比べながら適切な項目を選べばよい。実際に使ってみればすぐ分かることだが、本当は辞書ごとではなく、語彙ごとにソートされた方が便利だ。ソートし直すオプションは用意されていないので、ユーザが自分で行きつ戻りつして参照しなければならない。不便なら、声調も指定して、候補をさらに絞り込めばよいのだが、その場合は軽声も含めてすべて入力し、完全に一致するものだけが表示される。2音節以降を省略してしまうと、候補なしと表示されるので、これもまた不便だ。

個別に検索するぶんには、クラウン中日辞典、現代漢語大詞典のほか、英漢・漢英でワイルドカー

dian zi を検索した候補画面



ドが使えるので、ボタンで辞書を切り替えながら検索するか、「単語跳査」(スーパージャンプ)を使った方がよいかもしれない。

個別の語彙については、細かく見てゆく余裕がないが、『漢英大辞典』、『現代漢語大詞典』の個性がはっきり出ている、相補い合う関係にあって面白い³⁾。中検3級では厳しいだろうが、留学経験者ならずとも、中検2級を取れた学生なら十分おもしろさを分かってくれるだろう。

『漢英大辞典』は中国語から引ける小百科事典という性格を持ち、科学技術用語も多数収めている。加えて中国人学生達の英作文のために口語的表現も用意しているせいか、用例も英中併記でかなり豊富だ。例えば生活用語として、「方便」を引いてみると、下図のように出てくる。「方便」の語彙としては1) 便利である、2) 用足しに行く、3) 経済的余裕がある、という必要最低限の語彙は尽くされ、中中辞典よりも平明だ。両者の例文を並べて読めば分かるように、漢英の方は出典を厳格に明記しない代わりに口語的な表現を多数盛り込んでおり、学習者にも使いやすい例文が多い。

●漢英大辞典

③[婉] (有富余的钱) have money to spare or lend 手头不方便。have little money to spare

●現代漢語大詞典

⑤指钱财宽裕。例句：曹禺《雷雨》：“现在你手下方便，随便匀给我七块八块的好么？”

言うまでもなく、『漢英大辞典』は中国人が英語で作文するために編まれた辞書である。その意味では中国語の語義用法そのものを十分に分析していない部分もあろうが、比較参照することで得られるものは少なくない。個人的には日本語訳より例文を読むために引くことが多いので、これだけでも買って良かったと思う。

こうしてみると、語彙数の少ない『クラウン中日辞典』の存在感の希薄さは否めないが、副詞・介詞など虚詞関連の説明に関しては圧倒的な強さを発揮する。例えば「很」について引けば、『漢

英大辞典』が示す語義はわずかに一つ、『現代漢語大詞典』も二つだけであるが、『クラウン中日辞典』は十四項目に分けて詳説している。これほど違うのは例外的であるが、総じて言えることは、中中、中英には用例はあっても、用法説明についての分析は少ないのである。『クラウン中日辞典』はその欠点を補う貴重な存在といえよう^[4]。

📖 日本語から検索

日本語から検索する場合、ローマ字でひらがなを入力して（五十音入力も選択可能）検索する。中国人向けらしく、拼音でも検索できるところが新鮮だ。検索対象となる辞書は二種類。

- (1) 日漢 日漢大辞典
- (2) 日目 広辞苑

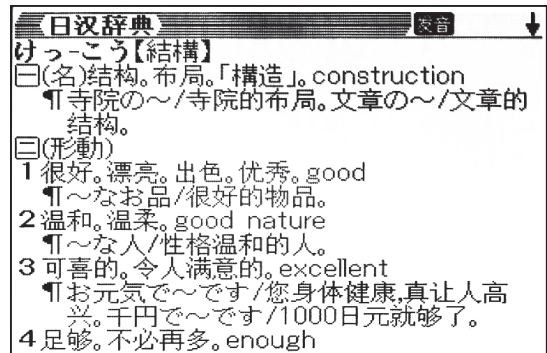
中国語の対訳を提示してくれるのは『日漢大辞典』だけだから、日本語からの検索機能はやや貧弱ともいえるが、日漢大辞典は講談社版『日本語大辞典』の中国語版で、百科事典としても利用できるように編集されている。紹介によれば、語彙数 20 万語 + 英訳 12 万語（単語）とあり、若干政治的な語彙が削除されてはいるものの^[5]、語彙量は十分だろう。

上図は「結構」を検索した画面だが、日本語に中国語対訳が付され、日本の電子辞書搭載の中日辞典と比べて大きな遜色はない。中日辞典と同様、語法説明が不足している嫌いがあるにせよ、むしろ例えば「フェリーニ」や「高倉健」についての説明が中国語で読めることが、この辞書の貴重なところであると考えべきだ。

もう一つ面白いオマケ的な機能として、「多詞典查詢」から英語で検索すると、英漢辞典の項目だけでなく、広辞苑、日漢大辞典に含まれる英語も検索できる。lisp を引くと、英漢辞典にはなかったが、日漢大辞典、広辞苑から説明が出てくる。

● リスプ [LISP]

<list processing language 的略語> 処理语言。



日漢大辞典検索画面

由美国麻省理工学院开发的函数型程序语言。广泛应用于定理的证明及符号处理，语意论，信息检索，人工智能的研究等。

linux は出てこなかったが、unix ならある。外来語の訳に困った場合など意外に役立つ機能なのではないだろうか。

📖 自由度の高いデバイスを

この電子辞書を買うとき、ネット通販サイトをあちこち見て値段を調べたので、中国の電子辞書の価格帯がだいたい分かってきた。最も手軽な普及機種だと 300 元前後で、入っているのは英語の辞書が一つとオマケの単語練習用のクイズソフト、ソリティアのようなゲームといったところ。少し高級機種になると、音声付き辞書が入り、気休めのような八か国語会話集がオマケに入る。1000 元近くの高級機種になると、メモ리카ードを追加でき、録音再生ができて、MP3 が聞ける機能がつく。英語学習を大義名分にしつつ、実は音楽プレーヤーとしても使える。高い買物だから多機能化は必然だ。辞書が沢山入っているだけでは売れない。1000 元以上となると、機種は数えるほどで、ハードウェアとしてのスペックが違うだけだ。ところが、CASIO EW-V3500L は相当安くても 1500 元前後だ（2006 年 7 月現在）。液晶も高品質で収録している辞書も中国国内メーカーのものに比べれば遙かに良い。しかし、これだけ価格帯がずれては恐らく余り売れないの

ではないかと思う。欠点もあるけれど、良い辞書だけに残念である。

良いものだけれど今は高価で売れないのだから仕方ないのかもしれないが、中国で売れている電子辞書を店頭で手にとって見ると、明らかに方向性が異なる。今後ハードウェアが良くなっていけば、電子辞書というよりハンドヘルドPCに近くなるだろう。かつてロースペックのハードウェアで大容量辞書データが手軽に素早く引けるのは魅力だったが、今後もそうといえるのか。少なくとも中国に関しては疑問だと思う^[6]。

日本市場は今後もテキスト中心の辞書が主流を占めるだろうが、今の電子辞書には、ユーザの創意工夫を生かせる余地がもっと欲しい。例えば、データ交換機能として、ユーザ間で検索履歴を共有できる機能、参照用例をパソコンに転送して印刷できる機能、あるいはメモリカード上の中国語（UnicodeあるいはGBテキスト）を表示させながら、辞書引きできる機能など、現在の技術からすれば十分実現可能なはずだ^[7]。

デバイスとして自由度が高くなれば、色々な使い方が教室で提案できる。そんな電子辞書の登場を期待したい。

注

- [1] 本誌第5号に「快譯通試用記」に書いたことがある。ご参照いただければ幸甚。
[2] 清原文代氏の中国語CAI研究会メーリングリストでの発言による。ほかにも中国語電子辞

書全般に関して、「中国語の電子辞書」(<http://www.las.osakafu-u.ac.jp/~kiyohara/chn-dic.html>)、「中国語電子辞書機能比較」(『辞書のチカラ』好文出版)で詳説されている。以下、注記3,4,6,7については同氏からのコメントを反映させていただいた。ここに記して感謝の意を示したい。

- [3] 『現代漢語大詞典』は『漢語大詞典』のサブセットという性格があるため、用例には必ず出典が記される。また、現代中国語辞典としては珍しく、部首配列であるが、電子辞書では拼音配列に改められている。
[4] クラウン中日辞典には親字に四角号碼が記載されているが、これはほかの辞書にはない。四角号碼検索には対応していないのが惜まれる。
[5] ただし、中国関連固有名詞については多数削除されている。「林彪」「江青」など項目だけ残して説明はすべて削除されているもの、「天安門事件」のように項目そのものが削除されているものがある。
[6] 例えば、カラー液晶を採用したSharpのPW-N8000のようなマルチメディア志向が今後の方向性として、一つの解になるのではないか。さらにはNintendo DS対応の辞書もすでに出ている。
[7] 日本語の「青空文庫」などのJISテキストならば読み込んで電子ブックとして使える機能を備えたものがすでに存在する。CASIO XD-GT7350やSONY EBR-800MSなど。

「電子辞書座談会」レポート

田邊 鉄（たなべ てつ）

📎 「電子辞書座談会」盛會に終わる

昨年12月7日に慶応大学日吉キャンパスで開催された日本中国語CAI研究会・漢字文献情報処理研究会合同企画「電子辞書座談会」は、多数の参加を得て、盛會のうちに終了した。

講師にセイコーインスツル株式会社（SII）パーソナル機器事業部の木村一彦氏をお迎えして、電子辞書事業全体、また、中国語を含む外国語電子辞書についてメーカーはどのように考えているのか、お話いただいた。

なお、今回の座談会は漢情研とCAI研究会のジョイント企画として行われた。ともに電腦で中国語を扱う「同志」とはいえ、電子辞書や電腦中国語に対する見方・考え方には、異なる部分も多い。両研究会の交流は、個々の見識を広げてくれるまたとない機会であり、有意義な座談会であったと思う。講師の木村氏、お世話いただいた両研究会関係者、参加者各位に誌面を借りて御礼申し上げます。

以下、事前質問を含めて座談会の内容を、教育関連の話題を中心に、メモ的にレポートする。

📎 事前質問

木村氏には、漢情研BBSやCAI研究会メーリングリストで募集した「メーカーへの質問・要望」を事前に渡し、それに沿ってお話頂くようお願いしていた。

こちらの質問は次のようなものである。

(1) 技術・ハード

- 大規模文字セットへの対応はあるのか。
- コンテンツの自作やダウンロードに対応できないか。
- PCとの連動による串刺し検索や学習支援機能を実装しては？

(2) コンテンツ

- 辞書をもっと自由に選べるようにならないか（小学館だけじゃなく）
- 辞書を自作しようというコミュニティを支援する気はないか
- 出版社とメーカーが両方とも利益を得るビジネスモデルはないか

果ては「辞書専用ハードの存在意義はどこにあるのか」という、メーカー的には微妙であろう問いまであった。

電子辞書ユーザ全体から見れば微々たる数だろうが、中国語教師や中国研究者の中に、これほど熱狂的なユーザコミュニティが形成されているとは、木村氏の想定外であったのだろう。これらの質問全てに、明快にお答え頂けたわけではない。それを不満に感じなかったのは、少数の要望を門前払いするのではなく、なるべく多くの意見に耳を傾けて、よりよい製品開発を行いたい、というメーカーの姿勢を確認できたからである。

もとより商売であるから、何の声もないところに対応することはできない。教員や研究者の声をなるべくたくさんメーカーに投げて欲しい、一定規模の需要があると確認できれば、メーカーは動けるのだから、と、最後は木村氏からハッパをかけられた格好になった。

中国語教育と電子辞書

電子辞書市場で、教育関連の需要はどれぐらいあるのだろうか。明確な数字は出ていないが、家電店などで見る限り、中高生向け・英語に限定すれば結構な数が出ているように見える。これは進学祝いの一部が、紙の辞書から電子辞書に流れたということだろう。SIIでも、高校生の英語学習にはかなりの需要あり、と見ているようで、学習機能を充実させた電子辞書をリリースしている。

では、中国語はどうか。1・2年生のいわゆる「第二外国語」の授業で電子辞書を使っている学生は皆無に近い。強く指示しなければ紙の辞書も買わないのだから、無理もない。電子辞書を使うのは、専攻が中国関係であるか、キャリアとして中国語を必要とする志望を持っている、意識の高い学生に限られる。

そうすると、大学生の学習に特化した入門用電子辞書、などという製品は出しにくいだろう。以前は「第二外国語用」と、はっきり大学生向けであることを謳ったメーカーもあったが、やはり中心になるのはビジネス用途のユーザなのである。

中国語教育への配慮を求めるなら、クラス全員に持たせるぐらいのことは考える必要があるのかもしれない。

「使い方指導」の有効性

これだけ電子辞書が普及すれば、今さら「電子辞書は是か非か」などという議論をしても始まらない。むしろ正しい使い方を指導した方がいい、という考え方は、確実に広まりつつある。

それはそうなのだが、じゃあ私たちは辞書の「使い方」を教わったことがあるのか、というと、実は無かったりする。大学で最初に中国語を教わった先生が、結構マニアックな人だったので、私が最初に手にした『新華字典』は四角号碼索引のものだった。だから「引き方」はさんざん教わったが、学習の過程で辞書をなぜ使うのか、どう使えば学習効果が上がるのか、といったことは教わっ

た覚えがない。おそらく「辞書を引くのは当たり前」と思いこんでいたのだと思う。

最近その前提が通用しなくなっていることは、中国語教員の多くが感じている通りである。具体的な課題を提示し、辞書を使う必然性・必要性を明示することが必要になっている^[1]。

SIIは高校生対象の「洋楽翻訳コンクール」を実施している。これは「辞書をどう使うか」の実践トレーニングと言える。また、「電子辞書講習会」のような形で、教育現場への協力が可能だという。教員とタイアップして行う「講習会」は、単なる操作説明・辞書の引き方説明にとどまらず、生徒に電子辞書ならではのメリットに気づかせ、辞書の「使い方」に目を向けるきっかけともなるのではないだろうか。

「電子辞書→紙の辞書」もあり？

高校生の場合、電子辞書以前に、辞書引き自体の習熟度に格差があるという。普段から辞書を引き慣れているほど、電子辞書も普通に引けるだろうし、逆に普段ろくに辞書を引かない生徒に、電子辞書を与えたら突然辞書を引くようになる、というものでもない。まず紙の辞書をきちんと使う習慣を身につけた上で電子辞書を使うように、という忠告はいつも善良である。

けれども、中には電子辞書で使える様々な“裏ワザ”を駆使することに魅力を感じ（「悪知恵」働かせて授業でラクをするのは甘い体験には違いない）、そこから辞書そのものに興味を持つ、「電子辞書から紙の辞書へ」という流れもあるのではないか。そして、今後「電子辞書を最初に買った」という生徒／学生が現れることまで想定して、コンテンツに意識的である必要があるのだろう。

注

- [1] この議論は、田邊鉄「電子辞書は授業で使えるかーオンライン討論から」『漢字文献情報処理研究』第5号を参照頂きたい。

W-ZERO3 で中国語

内田 慶市（うちだ けいいち）

📎 W-ZERO3 について

昨年暮れ、日本で初めての Smartphone である W-ZERO3 (WS003SH, Shrap/Willcom) が登場した。PDA の機能に加えて、QWERTY 配列キーボードを備え、電話 (PHS)、カメラ、インターネットさらには Microsoft office (Mobile) や PDF 閲覧、MP3、Windows Media Player まで扱える優れものである。その後、内蔵メモリを増やし、英和・和英・国語辞典を収録し、ボディの色を変えた 2 代目 W-ZERO3 (WS004SH)、そして、よりスリムな 3 代目 W-ZERO3[es] (WS007SH) も発売されている。

OS はいずれも Windows Mobile 5.0 for Pocket PC 日本語版であり、以下のようなソフトウェアがあらかじめインストールされている。

- Microsoft Outlook
- Microsoft Office Mobile
(Word Mobile / Excel Mobile / PowerPoint Mobile)
- Windows Media Player 10.0 Mobile for Pocket PC
- Picsel PDF Viewer
- Microsoft Internet Explore Mobile / Opera Mobile

📎 W-ZERO3 を中国語化

さて、このような携帯端末を手にする、どうしてもやりたくなるのは「どこでも中国語」「モバイル中国語」ということになる。これまでも、

Apple Newton Message Pad や HP200LX に始まり、Palm や Clie などでも筆者は必ずこれを実現してきた。もちろん、これは「メーカーの保証外」であり、今回とりあげる W-ZERO3 も例外ではなく、メーカー (Sharp/Willcom) の公式見解では「中国語および韓国語はサポートされていない」ということになっている。つまり「自己責任」において行うことになるが、しかしながら、これまで扱ってきた PDA に比べたら実に容易であるし、ZERO3 に限らず Windows Mobile 5.0 搭載機であれば他の機種 (たとえば EX600: 後述) でも同様の手順で「中国語環境」が実現可能であり、ここで紹介することとした。

📎 1. インストールの前に

まず、以下のソフトをあらかじめダウンロードしておく。

(1)「e-china No5」(フリーウェア)

http://www.geocities.jp/asukal_trader/Localize/localizecab.htm

(2)「Maction Monster SIP 怪獣輸入法」(980 台湾元)

<http://www.mobuy.com.tw/>

なお、これを購入する際に、ついでに、以下のものも購入しておくといいかも知れない。

「PocketGenius 口袋天才」(980 台湾元)

(3)「梅花輸入法 (増強版) PlumSIP 4.0」(18 人民元)

http://www.new-e.net/4/4_1.html

(4)「Pocket の手 for W-ZERO3」(フリーウェア)

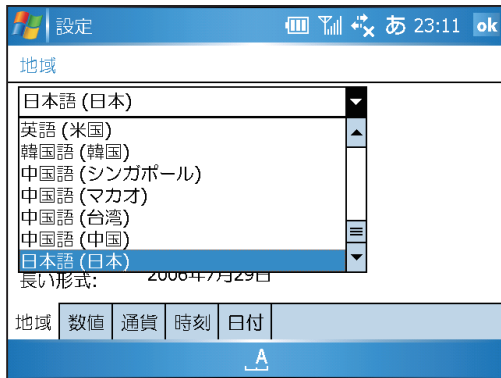


図 1 中国語が追加された

<http://smart-pda.net/useful7+index.id+7.htm>

(5)「GSFinder+ for HTC Universal」(フリーウェア)

<http://mkawa.dyndns.org/Soft0kiba.htm#GSF-U>

なお、これらのうち、(2)と(3)は「中国語入力ソフト」(IME)であり、いずれか1つでもよい。(5)はいわゆるファイル閲覧ソフトなので、標準装備の「ファイルエクスプローラー」でもよい。

また、これら以外に、インストール後「リセット」が必要になる場合があるので、それを行うソフトウェア(「Powercheck」など)をインストールしておく方が便利である。というのは、W-ZERO3では[es]以外の機種では「リセット」をする場合、裏のカバーを開けてハード的に行わなければならないからである。

中国語化の手順

「中国語化」の手順は以下の通りである。

1. 「e-china No5」のインストール

W-ZERO3に新しいソフトをインストールするには、PCと連携させるか、あるいはMini SDカードを使う必要がある。

普通はPCと連携させるが、この場合、PCには「Activesync」(最新バージョンは4.2)をあ

らかじめインストールしておかなければならない。これによって、PCとW-ZERO3との同期(アドレス帳や予定表)とファイル交換を行うわけである。W-ZERO3のインストーラーは「cab」ファイルによって行われるので、これをMini SDカードに入れてW-ZERO3本体でインストールを行ってもよいのだが、Activesyncを通してしかインストールできない場合もあるので、こちらを利用することをおすすめる。(なお、ActivesyncはPC側にウイルスチェックソフトが入っていると起動しない時もあるので、syncさせるときには、それを念頭に置いてほしい)

さて、ダウンロードしておいた「e-china No5」を本体にインストールすると、「リセット」の後に、これまでの環境と大きく違う現象が現れる。

それは「地域」の設定画面であるが、インストール前にはアジア系の言語は「日本語(日本)」だけだったものが、インストール後には「中国語(中国)」「中国語(台湾)」「韓国語(韓国)」などが表示されるようになる。

これはちょうど、Windows XPでの「言語と地域のオプション」での「言語」の設定が「e-china No5」で行われたと考えてよい。同時に、ユニコードベースの中国語フォント(mUniFontM)もインストールされている(これを確認するには、レジストリエディタである「PocketTweak」(<http://tillanosoft.com/ce/ptweakj.html>)や「TRE Pocket PC」(<http://www2r.biglobe.ne.jp/~tascal/download/pocketpc/tre.htm>)などを使って、システムフォントを見ればよい)。

これで「中国語環境」の基本はできあがっているが、これではまだ中国語の入力はできない。

2. 「怪獣输入法(Monster SIP)」のインストール

次に、「中国語IME」ソフトのインストールである。

ここでは、台湾から出ている「怪獣输入法」を使用することにする。このソフトのベースは「繁体字」であるが、「簡体字」も入力・表示可能であり、しかも、この「手書き入力」の威力はすごいもの

がある。唯一の欠点は、ハードキーボードが使用できないことで、ハードキーボードを使いたい場合には、中国で開発された「梅花输入法」を使うことになる。

さて、この「怪獣输入法」のインストールは、PC本体から Activesync を使用しないとインストールはできない仕様になっているが、「怪獣输入法」をダウンロードしたら、その中の「Monster SIP 5.cab」をまずクリックして手動でインストールする。これで「Maction SIP」がシステムにインストールされることになる。これは、「繁体字」のみの、しかも「注音」入力であるが、これでは我々は使えない。そこで、次に、他の「输入法」を順次インストールすることになる。

これ以降は、PC上で、「setup 输入法」をクリックすれば、メニューが表示されるので、そこに現れる输入法をクリックしてインストールすればよい。

おすすめは、以下のものである。

- 「天才手写繁体版」
- 「繁体通用拼音输入法」
- 「天才手写简体版」
- 「简体拼音」

これらをインストールして、「リセット」をすれば、中国語入力が可能となる。

なお、「設定」→「入力」→「オプション」で、レジスターの画面になり、そこに購入した際に与えられるシリアルナンバーを入力する。

以上で、「中国語化」は完了だが、しかしながら、たとえば、ウェブブラウザを立ち上げて中国語のページを開くと、図3のように文字化けする場合がある。もちろん、これはOperaなどの設定ファイル=opera.iniのフォントの部分を書き換えれば直るのではあるが、Mobile Exploreではiniファイルの書き換えなどは面倒であるので、システムフォントに手を加えることにする。これによって、メールの中国語送受信も可能となるのである。その方法は以下の通りである。

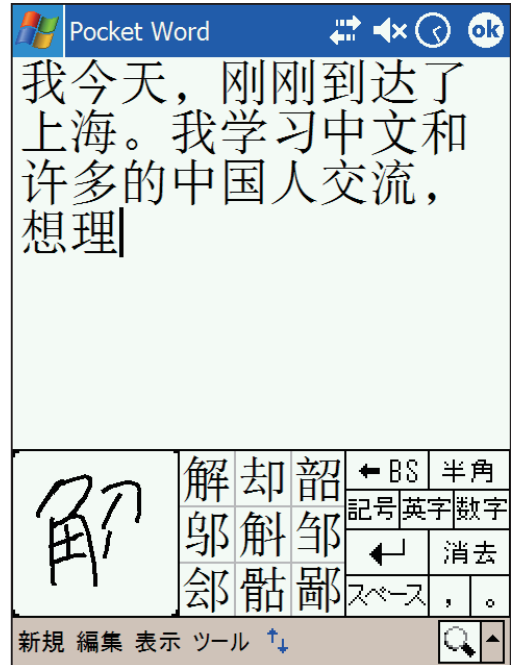


図2 Pocket Word で中国語入力

図3 Web 閲覧—文字化けしている



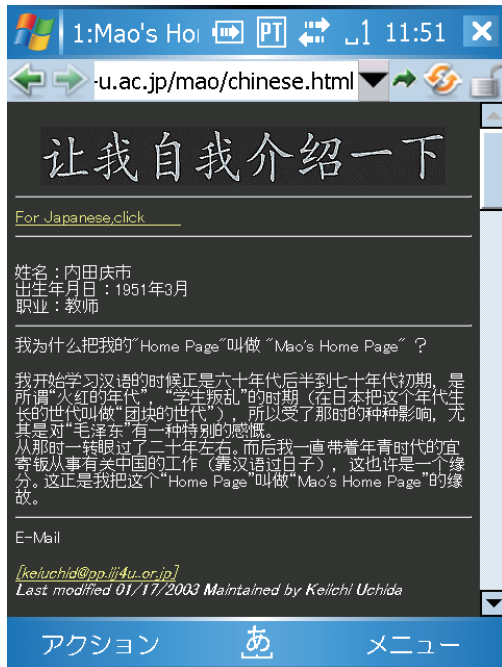


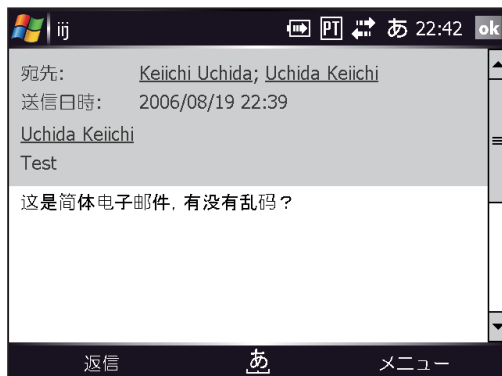
図4 文字化けが解消された

3. 「Pocketの手 for W-ZERO3」のインストールと設定

「Pocketの手 for W-ZERO3」をインストール後、起動させ、以下の部分を書き換える。



図5 電子メールでも中国語



フォント：mUniFontM

これで「リセット」して、ブラウザを立ち上げると、文字化けは解消される（図4）。メールも問題ない（図5：ただし、プロバイダのメールでは問題ないのだが、Willcom メールはどうもうまくいかないようである）。

おわりに

以上のような環境の構築は上でも述べたように「Windows Mobile」搭載機であればすべて同じであり、筆者はすでに香港で発売された日本語 Windows Mobile 搭載 GSM 携帯である「EX600」でもこれを実現している。「EX600」は日本では携帯電話としては使用できないが、中国や香港、欧米では携帯として使い便利である。

Windows Mobile に対応したソフトも数多くあるが、先の「怪獣輸入法」を開発しているメーカーの「天才口袋」は、英漢・漢英字典、医学事典、電子図書、パイブルなどを資料庫からダウンロードして利用でき、重宝なものである。

また、日本のメーカーとしては、高電社から「翻訳ウォーカー-j 北京 V2」（¥31,290）が「日中・中日翻訳+小学館日中・中日辞典」という機能を備えて販売されており、これもおすすめである。これをインストールすれば、先の手順は踏まなくとも中国語入力・表示が可能となる（ただ、フォントはどうか Unicode ではなく、GB あるいは Big5 フォントを使用しているようであり、メールやブラウザでの表示については検証していない）。

今後は、我が国でも Smartphone が普及して行くことが予想されるが、「多言語環境」を自前で用意しなくとも、最初からその機能が備わっていることを強く望むものである。