# 漢字文献情報処理研究

第11号

漢字文献情報処理研究会 編

好文出版

# 漢字文獻情報處理研究　第 11 号

# 目　次


| | | | |
|---|---|---|---|
| 論文 | 4 | 検索エンジン（Lucene）による中国語用例抽出最適化<br>── 北京大学中国語語料庫を利用して | 詹　善斌 |
| | 14 | 漢字構造の正規化処理 | 川幡　太一 |
| 漢情研 2010 年公開講座報告 | | 電子出版の動向と諸問題 | 23 |
| | 24 | 電子書籍の動向と検閲問題 | 田代　真人 |
| | 34 | 電子書籍とソフトウェアの自由 ── 電子書籍の永続化のために ── | 守岡　知彦 |
| | 41 | 電子書籍（e-book）とプラットフォーム<br>── 競争法上懸念される二つの問題 | 石岡　克俊 |
| | 48 | 全体討論のまとめ | 師　茂樹 |
| 特集 1 | | Windows 7 時代の『電脳中国学』と人文系情報処理テキスト | 53 |
| | 54 | 『電脳中国学 III』はいかにあるべきか | 千田　大介 |
| | 56 | 『電脳中国学 III』千田案　情報化時代における総合中国学入門ガイド | 千田　大介 |
| | 62 | 導入教育としての電脳中国学　千田大介案へのコメントとして | 師　茂樹 |
| | 66 | 中国史学むけデジタルリソースマニュアルとしての目次案 | 山田　崇仁 |
| 特集 2 | | 陳腐化するデジタル資料 | 71 |
| | 72 | デジタル情報の劣化と陳腐化・化石化 | 小島　浩之 |
| | 76 | 失われた文字コード | 安岡　孝一 |
| | 82 | レガシーとの付き合い方　東洋学文献類目の場合 | 守岡　知彦 |
| | 96 | 永続性のある歴史資料デジタル・アーカイブへの試論<br>──「アーカイブズ」への接近とデジタル応用の可能性 ── | 後藤　真 |

## ソフトウェアレビュー　105

**106**　iPad

研究者のための iPad /
iPad で見る、聞く、読む、学ぶ ── 中国語教育を中心に ──
師茂樹 / 清原文代

**125**　ソフトウェア

Windows 7 & Office 2010 / フォント・ブラウザ・多漢字 /
文字コード技術 / IME / 一太郎 2010 & ATOK2010 / Adobe InDesign CS5
千田大介 / 上地宏一 / 師茂樹 / 小川利康 / 山田崇仁

## 学術リソースレビュー　151

**152**　学術リソース

中国・台湾の電子書籍の動向 / 図書館と OPAC /
ベトナムの漢字・字喃文献 / 中国古典学術リソース、回顧と展望 /
大学授業への Twitter 導入の事例報告
── 大妻女子大学「情報社会論」での経験からの 14 の教訓
安藤一博 / 小島浩之 / 矢野正隆 / 秋山陽一郎 / 岡本真

**182**　学術ソフト・製品

DVD 版内村鑑三全集 / 講談社『中日辞典』第三版 CD-ROM /
『四部叢刊』と書同文社のオンラインデータベース
當山日出夫 / 田邊鉄 / 千田大介

## 書　評　**190**　『中日大辞典』第三版 /『現代漢語規範詞典』第 2 版

金子眞也

## お知らせ　**22**　会員の皆様へ / 年会費納入についてのお知らせ

**70**　漢字文献情報処理研究会　入会のご案内

**192**　漢字文献情報処理研究会彙報

**194**　著者紹介 / 編集後記


◦ 本誌記事中のソフトウエア名、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おりません。

◦ 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。

# 検索エンジン（Lucene）による中国語用例抽出最適化

## ――北京大学中国語語料庫を利用して

詹　善斌（せん　ぜんひん）

### 回 1. はじめに

北京大学漢語語言学研究中心[7]の語料庫（コーパス）は、中国でも有数規模の中国語語料庫であり、現代漢語だけでなく、古代漢語も収録している。現代漢語は中国のニュース、現代文学作品、日常会話、映画、テレビ（字幕）等、多彩なソースから収集されている。古代漢語は、四書五経から明清の白話に至るまで広く経史子集がまんべんなく収められている。本研究は北京大学漢語語言学研究中心の語料庫（現代漢語）から、HSK[1]常用単語に関する中国語例文を抽出し、抽出済の1437万例文を検索エンジンでインデックス化し、教育向けに最適化された例文検索を可能にしようとするものである。

本研究ではsolr[5]という検索システムを利用する。solrはLucene[3]によってApache Groupが開発したオープンソースの全文検索システムである。Solrは管理画面やキャッシュ機構を取り入れた使いやすい商用検索エンジンである。北京大学漢語語言学研究中心語料庫から収集した例文は数が多く、例文を適切に選択するのも難しいため、検索エンジン技術の導入によって、中国語例文の選択を容易することを目指している。

本研究ではictclas[2]という中国科学院計算技術研究所が開発された漢語分詞システムも導入している。この漢語分詞システムを利用することで、収集済みの例文を品詞分解し、検索語の関連語も抽出が可能になった。さらに抽出した関連語に重さを加味し、検索語と関連語をクエリーとして、solrで検索することで、最適な中国語例文を推薦できるようになった。

本研究は1）solrをベースに検索理論を活用し、2）ictclasの品詞分解技術も利用することで、北京大学漢語語言学研究中心語料庫から収集した例文の有効利用を目指すものである。

### 回 2. コーパスの収集

本研究は、北京大学漢語語言学研究中心の語料庫を利用し、整理したHSK単語をクエリーとして、「口語」（話し言葉）類別の中国語例文を収集した。

これは、早稲田大学総合教育研究所の企画研究部会「中国語コーパスを活用した中級語彙の選定および教材開発」（2009-2010年度）で必要とする語彙選定資料として、筆者が技術面で全面的に協力した作業の成果である。

HSKとは「漢語水平考試」の略称で、中国大学留学に必要な語学力が十分かどうかを判定する試験であり、中国のTOEFLとも呼ばれる。この

HSK には受験に必須の語彙が甲乙丙丁の四段階に分けて設定されている。これらの語彙は 1992 年当時の言語使用状況に即して制定されたものであるため、昨今の利用頻度に合わせた改訂が必要となっており、その単語の使用状況調査の一環として行われたものである。

### ◈ 2.1 北京大学漢語語言学研究中心の語料庫

北京大学漢語語言学研究中心の語料庫（現代漢語）は、中国のニュース、現代文学、話し言葉、映画、テレビ等から 728,909,261 字を収集したもので、この語料庫を活用した『现代汉语语法信息词典』（俞士汶編　清华大学出版社 1998 年）はコーパスによる語法研究は先駆的存在として知られる。この語料庫は北京大学漢語語言学研究中心よりウェブインターフフェースの検索システムとして提供されており、誰でも自由に検索閲覧できる。本研究ではクローラを用い、北京大学漢語語言学研究中心の語料庫検索ウェブページで現代漢語の「口語」類のみを収集した。

### ◈ 2.2 HSK 単語の整理

HSK は漢語水平考試という、中国政府教育部の認定する中国語語学検定試験で、世界全域で実施されている。HSK は 4 レベルに分け、入門級、初級、中級、高級である。本研究ではそれぞれレベルに常に出てきた単語を集めて、入門級は 2,037 単語、初級は 1,043 単語、中級は 3,577 単語、高級は 2,206 単語、合計 8,863 単語を収集された。

### ◈ 2.3 収集手法

例文の収集流れは図 1 に示す。最初に整理した HSK 単語それぞれクエリーとして、北京大学漢語語言学研究中心の語料庫ウェブページで現代漢語の「口語」類別のみ選択して、検索を行う。検索結果はそれぞれ例文の長さは 1,000 字を取り、最大 2,000 例文を収集する。収集した例文は MySQL データベースに書き込む。

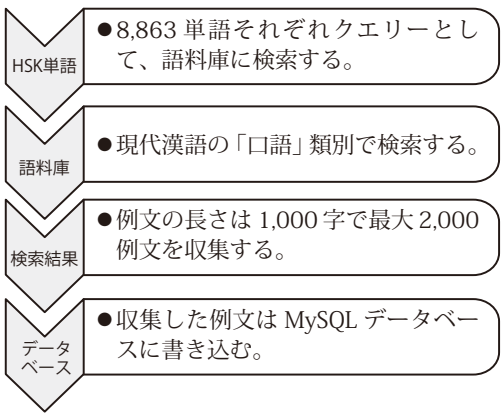



図 1：例文の収集流れ

### ◈ 2.4 例文収集

北京大学漢語語言学研究中心語料庫の検索結果は図 2 のようなものである。この検索結果ウェブページで「検索結果数」と「検索結果」を収集する。Ubuntu Linux で php 環境をインストールし、php の libcurl を利用して、例文の収集を行う。

### ◈ 2.5 品詞分解

クエリーとした HSK 単語及び収集した例文それぞれについて、中国語の文法構造を明らかにするために品詞分解を行う。品詞分解には ictclas という漢語分詞システムを利用する。ictclas とは中国科学院計算技術研究所が開発した品詞分解システムである。

本研究では ictclas の Linux C/C++ 版のソースコードをダウンロードし、C/C++ のソースコードを Linux 環境でコンパイルした。コンパイルした ictclas は Linux のコマンドライン環境で運行でき

図 2：検索結果のウェブページ

表 1：MySQL Table の設定

| Field | Type | Example |
|---|---|---|
| keyword | VARCHAR | 中文 |
| hits | INT | 6,398 |
| pos | VARCHAR | 中文 /nz |
| content | TEXT | ( 港澳台 ) 香港中文大学成立人类基因实验室 |
| content_pos | TEXT | (/w 港澳台 /j )/w 香港 /ns 中文 /nz 大学 /n 成立 /v 人类 /n 基因 /n 实验室 /n |
| content_hash | CHAR | 01e275b164ebffbffa66ddbb6142d430 |
| content_pos_hash | CHAR | 08d22111f7b78a3aff1520f41226fc4b |

る。ictclas の実行した結果はこの様である。例えば、下記のような文章に対して、品詞分解を行うとする。

> 随着越来越多的英国人开始对汉语和中国文化感兴趣 , 中文已被列入英国大学入学考试的选考科目。

その結果は以下のように与えられる。

> 随着 /p 越来越 /d 多 /a 的 /u 英国 /ns 人 /n 开始 /v 对 /p 汉语 /nz 和 /c 中国 /ns 文化 /n 感 /vg 兴趣 /n , /w 中文 /nz 已 /d 被 /p 列入 /v 英国 /ns 大学 /n 入学 /vd 考试 /v 的 /u 选 /vd 考 /v 科目 /n 。/w。

それぞれの記号は、n：名詞、v：動詞、d：副詞、p 前置詞、a：形容詞、w：句読点、u：助詞を意味している。

### ◈ 2.6 例文保存

収集した例文を保存するため、Linux 上の MySQL に例文をデータベースに書き込む。MySQL の中に corpus というデータベースを作成して table を作り、keyword、hits、pos、content、content_pos、content_hash、content_pos_hash というカラムを作る。keyword、pos、content_pos_hash 三つカラムを合わせて、unique key として設定する。コーパス収集量が多いので、MySQL の select 高速化の為に、この三つカラムもインデックスしていた。カラム keyword とは HSK 単語を保存する。カラム hits とは HSK 単語をクエリーとして、北京大学漢語語言学研究中心語料庫の検索結果数を保存する。カラム pos とは HSK 単語の品詞分解結果を保存する。カラム content とは HSK 単語をクエリーとして、北京大学漢語語言学研究中心語料庫のそれぞれ検索結果を保存する。カラム content_pos とは北京大学漢語語言学研究中心語料庫それぞれ検索結果の品詞分解結果を保存する。カラム content_hash とはカラム content の MD5 ハッシュを保存する。このカラムは重複の削除する為に作ったものである。カラム content_pos_hash とはカラム content_pos の MD5 ハッシュを保存する。このカラムも重複を削除するために作ったものである。

作成した MySQL Table は表 1 のようである。北京大学漢語語言学研究中心語料庫から収集した例文は 14,375,381 であった。

## 回 3. コーパスのインデックス

中国語例文抽出速度を向上させるために、収集した語料庫をインデックス化する必要がある。本研究では Lucene を用い、収集した語料庫をインデックス化し、Solr で検索できるようにする。

### ◈ 3.1 Lucene 紹介

Lucene とはオープンソースの検索エンジンで、Apache Group が JAVA によって開発した検索エンジンである。Lucene が主に用いる検索エンジンアルゴリズムは tf-idf[6] である。その特徴は以下の通りである。

> tf-idf は、文章中の特徴的な単語（重要とみなされる単語）を抽出するためのアルゴリズムであり、主に情報検索や文章要約などの分野で利用される。tf-idf は、tf（単語の出現頻度、term frequency）と idf（逆

出現頻度、inverse document frequency）の二つの指標で計算される（WikiPedia より加筆引用）。

これをコーパスに応用すると、単語の出現頻度 $tf_{i,j}$ は以下の方程式のように導かれる。$n_{i,j}$ はある文章 $d_j$ にある単語 $t_i$ の出現数である。$\sum_k n_{k,j}$ はある文章 $d_j$ に全て単語の数量である。

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

逆出現頻度 $idf_i$ は以下の方程式に示す。$|D|$ はコーパスの中に総文章数である。$|\{d:t_i \in d\}|$ はコーパスの中に $t_i$ がある文章数である。コーパスの中に総文章数割るコーパスの中に $t_i$ がある文章数の対数は逆出現頻度 $idf_i$ である。$n_{i,j}=0$ の時、$|\{d:t_i \in d\}|$ の部分は $1+|\{d:t_i \in d\}|$ になる。

$$idf_i = log \frac{|D|}{|\{d:t_i \in d\}|}$$

そして、ある単語 $t_i$ がコーパスの中にある文章の重要性 $d_j$ は以下の方程式に示される。

$$(tf\text{-}idf)_{i,j} = tf_{i,j} \times idf_i$$

ここで具体例で説明しよう。ある文章の全て単語数は 100 個がある。その文章は単語「牛」が三つである。その文章に「牛」の出現頻度 *tf* はで 3/100=0.03 ある。コーパスの中に総文章数は 10,000,000 で、「牛」がある文章は 1,000 なら、逆出現頻度 *idf* は $ln\left(\frac{10{,}000{,}000}{1{,}000}\right)$ =9.21 である。その結果、その「牛」単語がある文章の *tf-idf* は 0.03 × 9.21=0.28 となる。

*tf-idf* スコアはコーパスの文章をランキングする為に計算している。

### ❖ 3.2 Solr 紹介

Solr は Lucene に基づき、Apache Group が開発されたオープンソースの全文検索システムである。Solr は管理画面やキャッシュ機構を取り入れた便利な商用検索エンジンである。日本国外では小中規模のニュースサイトだけでなく、極めて大規模なソーシャルニュースサイト Digg や、インターネットアーカイブでなどで利用されている。日本国内では SHOOTI において約 2 億の Web ページのインデキシングに利用されている（WikiPedia より加筆引用）。

本研究では Solr のバージョン 1.3.0 と Tomcat のバージョン 6.0.18 を利用した。インストール手順は以下で紹介する。

**■ ダウンロード Solr**

以下の URL で Solr をダウンロードする。

**http://www.apache.org/dyn/closer.cgi/lucene/solr/**

**■ ダウンロード Tomcat**

以下の URL で Tomcat をダウンロードする。

**http://tomcat.apache.org/download-60.cgi**

Tomcat では、Solr 管理画面の為と Solr 検索結果 API の XML/JSON 出力の為に利用される。

**■ Tomcat の解凍**

```
> tar xvfz apache-tomcat-6.0.18.tar.
  gz
> mv ./apache-tomcat-6.0.18 /home/
  search/tomcat
```

**■ Tomcat の server.xml を編集**

UTF-8 サポートのため、Tomcat の server.xml の編集が必要となる。

```
> cd /home/search/tomcat
> vi conf/server.xml
```

以下の様に server.xml を編集する。

```
<Connector port="8080"
protocol="HTTP/1.1"
connectionTimeout="20000"
redirectPort="8443"
URIEncoding="UTF-8"
    useBodyEncodingForURI="true" />
```

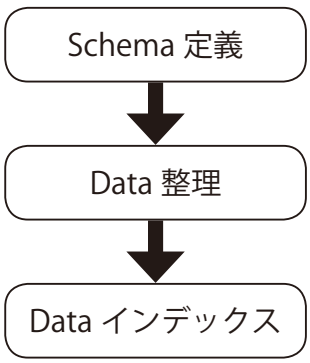



図 3：Solr の使い方

**■Solr に JAVA 環境の設定**

Solr は Lucene に基づき、開発されたアプリケーションである。Lucene は JAVA で開発された検索エンジンである。それで、Solr を使う前に、Linux で JAVA 環境の設定が必要になる。

```
> echo "export CATALINA_BASE=/home/
  search/tomcat" >> ~/.bashrc
> echo "export CATALINA_HOME=/home/
  search/tomcat" >> ~/.bashrc
> echo "export JAVA_HOME=/usr/lib/
  jvm/java-6-sun/jre" >> ~/.bashrc
> echo "export PATH=${JAVA_HOME}/
  bin:${PATH}" >> ~/.bashrc
> source ~/.bashrc
```

**■Solr の解凍**

```
tar xvfz apache-solr-1.3.0.tgz
```

**■Solr の設定**

```
> cp apache-solr-1.3.0/dist/apache-
  solr-1.3.0.war /home/search/
  tomcat/webapps/solr.war
> cp -R apache-solr-1.3.0/example/
  solr /home/search/solr
> echo "export JAVA_OPTS="$JAVA_OPTS
  -Dsolr.solr.home=/home/search/
  solr" >> ~/.bashrc
```

**■Tomcat の起動**

```
> cd /home/search
> /home/search/tomcat/bin/
  startup.sh
```

**■Solr の運用テスト**

色々設定が終わったら、運用試験のため、下記の URL で Solr の動作確認をする。

**http://localhost:8080/solr/admin**

### ※ 3.3 コーパスのインデックス

コーパスの例文を検索できるようにするには、インデックス化が必要である。本研究は MySQL データベースを利用し、データベースの table で keyword、hits、pos、content、content_pos、content_hash、content_pos_hash というカラムに北京大学漢語語言学研究中心語料庫の検索結果を保存している。本研究は MySQL データベース table の content カラムのみインデックス化した。

Solr の使い方は図 3 に示す。まず Solr の検索 Schema を定義する。定義内容は検索したい field、field のタイプ（テキスト、数字等）、重複削除、品詞分解等である。本研究は mmseg4j[4] という solr の品詞分解 plug-in を利用している。インデックス化する時とクエリーする時、品詞分解は必要である。日本語、中国語、韓国語等、文章中に space がないので、tf-idf 等検索アルゴリズムをランキング計算する為に、品詞分解は必要である。例えば、ある文章「早稲田大学は模範国民の造就を本旨と為すを以て…」、インデックス化しない場合は「早稲田大学」をクエリーとして検索しても、この文章は出て来ない。この文章を品詞分解して「早稲田　大学　は　模範　国民　の　造就　を　本旨　と　為す　を　以て…」、クエリーも品詞分解して「早稲田　大学」、クエリー「早稲田　大学」と文章の「早稲田　大学」

にマッチングできて、初めて検索結果に出てくる。mmseg4j を使用する前に、以下の設定は solr の schema.xml に書き込む。

```
<fieldType name="textComplex"
class="solr.TextField"
positionIncrementGap="100" >
  <analyzer>
    <tokenizer class="com.chenlb.
mmseg4j.solr.MMSegTokenizerFactory"
mode="complex" dicPath="dic"/>
    <filter class="solr.
LowerCaseFilterFactory"/>
  </analyzer>
</fieldType>

<fieldType name="textMaxWord"
class="solr.TextField"
positionIncrementGap="100" >
  <analyzer>
    <tokenizer class="com.chenlb.
mmseg4j.solr.MMSegTokenizerFactory"
mode="max-word" dicPath="dic"/>
    <filter class="solr.
LowerCaseFilterFactory"/>
  </analyzer>
</fieldType>

<fieldType name="textSimple"
class="solr.TextField"
positionIncrementGap="100" >
  <analyzer>
    <tokenizer class="com.chenlb.
mmseg4j.solr.MMSegTokenizerFactory"
mode="simple" dicPath="dic"/>
    <filter class="solr.
LowerCaseFilterFactory"/>
  </analyzer>
</fieldType>
```

品詞分解 plug-in を設定した後、検索されたい field と field のタイプを設定する。本研究は二つ field「content」と「content_id」を設定する。content とは北京大学漢語語言学研究中心語料庫から収集した例文をインデックス化する。content_id とはインデックス化する時、それぞれ例文の MD5 ハッシュを生成して、重複例文を削除する。以下の設定は solr の schema.xml に書き込む。

```
<fields>
  <field name="content"
type="textComplex" indexed="true"
stored="true" multiValued="true"/>
  <field name="content_id" type="text"
indexed="true" stored="true"/>
</fields>
<uniqueKey>content_id</uniqueKey>
<defaultSearchField>content</
defaultSearchField>
```

検索 schema を定義した後、検索されたいデータを整理する。XML フォーマットのファイルに生成する。ファイル内容は以下の様に書き込む。MySQL の中にそれぞれ例文を抽出し、例文はタグ field 中に書き込む。

```
<add>
  <doc>
    <field name=“content”>
      Contents that will be indexed.
    </field>
  </doc>
</add>
```

XML フォーマットの例文データファイルを生成した後、以下の方法で solr にインデックス化する。

```
java -jar post.jar feed_data.xml
```

本研究は北京大学漢語語言学研究中心語料庫か

ら収集した例文は 1,437 万ぐらいである。インデックス化する時、重複削除して、インデックス化した例文は 347 万ぐらいである。インデックス化した後、solr で検索できる。検索結果は二つフォーマット（XML と JSON）で出力できる。本研究は XML フォーマットの出力を運用している。図 4 は solr の web api を利用し、XML フォーマットの出力結果である。検索結果数、例文等、色々な情報を載せている。

```
<?xml version="1.0" encoding="UTF-8"?>
<response>
  <lst name="responseHeader">
    <int name="status">0</int>
    <int name="QTime">3</int>
      <lst name="params">
      <str name="indent">on</str>
      <str name="start">0</str>
      <str name="q">研究</str>
      <str name="hl">true</str>
      <str name="rows">1</str>
      <str name="version">2.2</str>
    </lst>
  </lst>
  <result name="response" numFound="92707" start="0">
    <doc>
      <arr name="content">
        <str>対策研究</str>
      </arr>
        <str name="content_id">46fcf6463fde7a42</str>
    </doc>
  </result>
</response>
```

図 4：Solr web api の検索結果

図 4 によって、パラメーター「q」は検索目的語「研究」である。パラメーター「numFound」は「研究」の検索結果数である。パラメーター「content」は検索結果の一つである。パラメーター「content_id」は検索結果の MD5 ハッシュである。本研究は以上四つパラメーターを利用して、例文推薦システムの検索インタフェースに組み込む。

## 回 4. 語料庫の検索

北京大学漢語語言学研究中心語料庫から収集した例文数は大変多く、重複数量も大変多いので、このままでは教育用に例文を選択することはできない。そこで検索エンジン技術を利用することにして、solr を導入した。

### ❖ 4.1 MySQL Search V.S. Solr

MySQL にも検索機能にも係わらず、なぜ solr を利用するかといえば、その検索速度にある。本研究では、MySQL 検索機能の性能と solr の性能を比較してみた。使ったサーバーの CPU は Six-Core AMD Opteron(tm) Processor 2427、RAM は 32GB である。「中文」をクエリーとして検索すると、solr では僅か 0.136 秒ですむのに対して、MySQL の場合は 16.95 秒もかかってしまう。この検索では、solr の検索に、ランキング機能を付けていても MySQL よりはるかに速い。そこで solr を利用することとした。

### ❖ 4.2 例文推薦システム

HSK 単語のみクエリーとした solr で検索結果を「科学」を例に挙げよう。

検索アルゴリズムの関係上、短い例文ほどランキングは高くなる傾向にあり、下記のようになる。

1. 科学种植，科学养殖。
2. 科学画报
3. 《科学画报》
4. 空间科学
5. 科学喂养
6. 科学万岁！
7. 科学时代
8. ⑺热爱科学，崇尚科学。

9. “科学无国界，科学家有国籍。”
10. 要提倡科学，靠科学才有希望。

明らかに提示される例文はどれもセンテンスとして不完全なものが多数を占め、教育用には不適切なものがほとんどを占めることになってしまう。

そこで本研究では、検索目的語の関連語を探し、例文検索のクエリーに関連語を補うことによって、中国語教育用に適切な中国語例文を推薦するシステムを提案したい。例文推薦システムのフローチャートは図５に示す。

図５：例文推薦フローチャート

- 最初に、MySQL データベースから HSK 単語を含む例文を抽出する。
- データベースから抽出した例文はそれぞれ HSK 単語の関連語を抽出する。
- 抽出関連語の重さを計算する。
- 最後に、HSK 単語と重さの付いた単語の関連語を Solr で検索し、最適な例文を推薦する。

### ※ 4.3 関連語の抽出

HSK 単語「科学」を例として、関連語を抽出する方法を説明する。 HSK 単語「科学」を含む文として次のような例を想定する。

> 研究农作物与气象条件关系的科学，称为农业气象。

この例文は品詞分解した結果は以下である。

> 研究 /vn 农作物 /n 与 /c 气象 /n 条件 /n 关系 /n 的 /u 科学 /n ，/w 称为 /v 农业 /n 气象 。/w

単語「科学」は名詞で、関連語は句読点の間に一番近い動詞である。例によって、句読点の間に単語「科学」があるセンテンスは「研究 /vn 农作物 /n 与 /c 气象 /n 条件 /n 关系 /n 的 /u 科学 /n」である。このセンテンスの中に、単語「科学」の一番近い動詞は「研究」であるから、単語「科学」の関連語は「研究」と見なすことが出来る。

従って、関連語を抽出するシステムのルールは以下の通りとなる。

- HSK 単語は「名詞」の場合、「動詞」関連語を探し出す。
- HSK 単語は「動詞」の場合、「名詞」関連語を探し出す。
- 探索範囲は句読点の間のみである。

### ※ 4.4 関連語の重さ計算

単に関連語をクエリーに加えるだけで適切な例文を検索できない場合もあるので、関連語の重さを計算に入れた方が良い。ここでは単語「科学」を例として、関連語重さの計算方法を説明する。

北京大学漢語語言学研究中心語料庫から収集した例文の中に、単語「科学」がある例文数は 2,000 がある。その 2,000 例文の中に単語「科学」の関連語「研究」がある例文数は 500 がある。それで、関連語「研究」の重さは 500/2000=0.4 になる。

関連語重さの意味は検索目的語がある文章の中に関連語の重要性である。高い重さは関連語出現の比率が高い。言い換えると、高い重さ関連語と検索目的語と一緒に使う場合は多い。

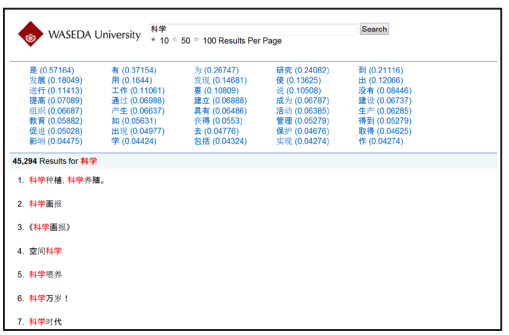


図 6：例文推薦システムの検索インタフェース

### ◈ 4.5 検索方法

Solr で検索する時、検索単語が二つ以上の場合は、検索単語に重さを付ける機能がある。本研究はこの機能を利用し、HSK 単語とその関連語が重さを付き、solr で検索し、最適な例文を推薦する。単語「科学」を例として、検索単語に重さを付ける機能の使う方法を説明する。

あらかじめ計算した関連語の重さは 0.4 で、HSK 単語の重さは常に 1 で、solr で検索式は「科学 ^1 ＋研究 ^0.4」になる。重さは 1 の場合は省略可能だ。そこで検索式は「科学 + 研究 ^0.4」になる。単純に HSK 単語「科学」はクエリーとした場合の検索結果は先に見たとおりだが、今度は「科学 + 研究 ^0.4」はクエリーとして、solr で検索すると、結果は以下のようになる。

1 研究数量和形状的科学
2 现代科学哲学的代表性著作之一，论述科学思想的起源和科学的研究方法，探讨科学中的若干基本概念。
3 要科学地研究市场、研究项目，不要急于求成。
4 【军事科学】研究战争和战争指导规律的科学。
5 军事科学】研究战争和战争指导规律的科学。
6 要研究制定科学合理的支出定额。
7 他认为，研究国学必须实事求是，科学地对待文化遗产。
8 【专攻】专门研究（一门科学）：他是～水利工程的。
9 专攻】专门研究（一门科学）：他是～水利工程的。
10 【力学】1 研究物体机械运动规律及其应用的科学。

「科学」と「科学＋研究 ^0.4」をクエリーとして、両方の検索結果を比較してみれば、「科学 + 研究 ^0.4」の検索結果の方が明らかに教育用例文として適切なものが多くなっている。

### ◈ 4.6 検索インタフェース

図 6 は本研究開発した例文推薦システムの検索インタフェースである 。このウェブインタフェースは三つモジュールがある。

一番上のモジュールは検索目的語を入力する所である。検索結果の表示数も選べる。

二番目のモジュールは重さを計算した関連語である。括弧の中に重さのスコアである。それぞれ関連語はハイパーリンクで、クリックすると、4.5 紹介した検索方法でもう一度検索する。

一番下のモジュールは検索結果である。この検索インタフェースをサーバーにインストールして、誰でもこの例文推薦システムを利用できる。

## 回 5. まとめ

北京大学漢語語言学研究中心語料庫から収集した例文はノイズが多いため、本研究は検索システムを利用して、収集した例文を重複分削除し、中国語例文を検索するようにした。更に本研究では品詞分解の技術を利用して、検索目的語の関連語を抽出し、その重みも計算し、検索目的語と関連語を合わせて検索するシステムを考案した。その結果、教育用中国語例文として適切な例文を効率的に検索することが可能になった。

ただし、現時点では例文中の品詞解析から「動賓構造」を発見し、関連語を抽出する方法をとっているため、例文検索の最適化としては、まだ改善の余地がある。今後、形容詞をサポートできるようにシステムを改善してゆきたい。また、現状

では例文中の単語を学習者の語彙量に対応させて、未習単語を含まない例文を抽出することはできない。だが、技術的には例文単語のフィルタリングは困難ではないので、この点についても改善を図りたい。

## 謝辞

この研究は早稲田大学教育総合研究所研究部会「中国語コーパスを活用した中級語彙の選定および教材開発」の研究成果の一部である。

## 参考文献

[1] HSK：http://www.hskj.jp/
[2] ictclas：http://ictclas.org/
[3] Lucene：http://wiki.apache.org/jakarta-lucene/FrontPage
[4] mmseg4j：http://code.google.com/p/mmseg4j/
[5] solr：http://wiki.apache.org/solr/
[6] tf-idf：http://en.wikipedia.org/wiki/Tf%E2%80%93idf
[7] 北京大学漢語語言学研究中心：http://ccl.pku.edu.cn/

# 漢字構造の正規化処理

川幡　太一（かわばた　たいち）

## 回 1 序論

情報交換用に符号化された漢字の数は増加の一途をたどっている。今年末には ISO/IEC 10646［JTC03］(Universal Coded Character Set .. UCS) にて拡張漢字 D の制定が予定され、総計で７万５千程度の漢字が符号化される見込みである。

大規模な漢字集合は多くの場合、漢字の重複が避けられない。Unicode/UCS でも、多数の漢字が誤って重複符号化されている［川幡 05］。

漢字の字形を管理し、必要な漢字を字形から検索し、漢字の重複符号化を避けるには、漢字の「字形・構造」をデータ化し、それに対する検索・比較機能が必要である。そのためには漢字の「構造」の表記方法と、それを比較可能にするための「正規化」を定めなければならない。

UCS では漢字の構造を記述する**漢字記述文字**（Ideographic Description Character ... IDC）が、ISO/IEC 10646 追補 F にて定められており、またこれを利用した、大規模な漢字構造データが整備されている［守岡 08］［川幡 08］。

本論文では、IDC の性質および IDC による漢字構造記述の課題を述べ、あわせてこれを解決するための手法として、漢字構造記述の正規化手法を提案する。

## 回 2 漢字の構造とその記述

### ◈ 2.1 漢字の構造

「説文解字」の説く「六書」の定義によれば、漢字は、その成立過程から象形・指事・会意・形声文字等に分類できる。その中でも会意・形声文字は、既存の漢字を組み合わせて、あるいは組み合わせた漢字を簡略化して生成される。このプロセスを繰り返すことにより、漢字は方形領域に無数のパターンをつくり出せる。

実際、漢字の大部分は会意・形声文字であり、これらは、既存の漢字の組み合わせとして表現することが可能である。すなわち、漢字の大部分は、他の漢字を部品として組み合わせた形で記述できる。

### ◈ 2.2 漢字構造の記述

漢字構造の表現手法として、国際文字標準規格 ISO/IEC 10646:2003 追補 F では、漢字記述文字（Ideographic Description Character）を 12 文字（U+2FF0 ～ U+2FFB ⿰⿱⿲⿳⿴⿵⿶⿷⿸⿹⿺⿻）、定めている。

この漢字構造記述文字を用いると、例えば「海」という漢字の構造は、「⿰氵毎」と表現できる。以下に、いくつかの例を示す。

| 漢字 | 漢字構造記述 |
|---|---|
| 海 | ⿰氵毎 |
| 草 | ⿱艹早 |
| 街 | ⿰彳圭亍 |
| 句 | ⿵勹口 |
| 進 | ⿺辶隹 |
| 旌 | ⿰方⿱𠂉生 |

「⿰氵毎」というIDCとそれに従う部品の文字列は、漢字構造記述列（Ideographic Description Sequence ... IDS）と呼ぶ。また、「氵」や「毎」などの漢字部品は、漢字部品記述文字

（Description Character ... DC）と呼ぶ。DC には通常、符号化漢字が用いられる。

上記の「進」の例で見られるように、必ずしも IDS は筆順に従って漢字を並べるわけではない。また「旌」のように、漢字を IDS へ展開する際に、単純に２分割できない漢字がある場合は、DC の部分に IDS を入れ子で用いることができる。

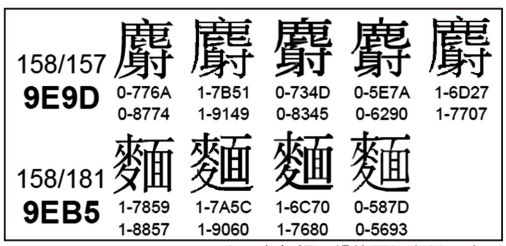


図１：麝と麺の規格票における字形

## 回 3 漢字構造記述の性質

### ◈ 3.1 IDS の展開と簡約

複雑な IDS はその一部分を**展開**したり、または**簡約**することができる。例えば、「峠」という漢字を IDS に展開すると、「⿰山卡」となるが、この右辺はさらに展開して、「⿰山⿱上下」と表記できる。展開の逆の操作として、IDS を簡約して、より単純な形にできる場合がある。IDS の簡約を繰り返した結果、単独漢字になれば、その IDS は当該漢字で**符号化可能**である。

### ◈ 3.2 DC を３つ取る IDC

IDC には、DC を２つ取るものと、DC を３つ取るものがある。DC を３つ取る IDC には「⿲・⿳」があるが、これらは２つの「⿰」・「⿱」に置き換えることができる。

⿲ xyz ＝⿰ x ⿰ yz ＝⿰⿰ xyz
⿳ xyz ＝⿱ x ⿱ yz ＝⿱⿱ xyz

例えば「儲」という漢字は、「⿲亻言者」と表記できるが、これを２つの「⿰」に置き換えた表現を簡約すると、「⿰亻諸」とも「⿰信者」とも表記できる。

⿲亻言者 → ⿰亻⿰言者 → ⿰亻諸
⿲亻言者 → ⿰⿰亻言者 → ⿰信者

### ◈ 3.3 「繞」「垂」「構」の構造記述

一般に、ある漢字部品が「繞（⿺）」の位置にある場合、これを「偏（⿰）」の位置に置き換えても、UCS の文字符号上は原則として区別されない。これは「垂（⿸）」と「冠（⿱）」の関係にもあてはまる。例えば､「麺」は「⿰麦面」と「⿸麦面」の両方で表記でき、また、麝は「⿸鹿射」と「⿱鹿射」のどちらでも表記可能である（図 1）。

同様に、「⿸・⿹・⿺・⿵」も単純な上下・左右関係とみなすことができ、これらは以下のようなルールで「⿱」・「⿰」に置き換えられる場合がある。

⿸ xy → ⿱ xy
⿹ xy → ⿱ xy
⿺ xy → ⿱ yx
⿵ xy → ⿰ xy

### ◈ 3.4 合成を示す IDC「⿻」

IDC「⿻」は、他の漢字部品にない以下のような特徴がある。

- ⿻ xy と⿻ yx は区別されない。そのため、正規化の際は、どちらかに片寄せする必要がある。
- ⿻を使った IDS から、漢字の実際の字形の予測が一般に困難である。
- 同じ漢字が、何通りもの合成方法で表現可能な場合が多い。

これらの特徴は、⿻を含む IDS は正しく比較できない可能性を示唆しており、一般には漢字の部品表現において、「⿻」の利用は極力避けるべきである。

## 回 4 漢字構造記述の曖昧性

同一の符号を持つ漢字でも、その IDS は必ずしも一意であるとは限らない。それは以下の５つの理由による。

### ◈ 4.1 分割点による曖昧性

例えば、「儲」という漢字の構造は、「⿰信者」、「⿰亻諸」、または「⿰亻言者」のいずれでも記述できる。同一の漢字でも、どこで分割するか、どのように分割するかで、漢字構造記述列による表現が異なる。

また、例えば「焱」という文字は、「⿱双双」とも、「⿲爻爻爻」とも分割可能である。このように、縦横両方で分離可能な文字もまた、異なる漢字構造記述列で表現できる。

### ◈ 4.2 分解レベルによる曖昧性

同じ漢字でもどこまで分解できるかによって、IDS 表現は異なる。例えば「峠」という漢字は、「⿰山⿱上下」と「⿰山⿱上下」のどちらの構造記述でも表現可能である。

### ◈ 4.3 「繞」「垂」「構」の構造記述の曖昧性

先述したように、「麺」は「⿰麦面」と「⿺麦面」の両方で IDS 表現可能であり、IDS としてこのまま比較しても、一致しない。このようなケースが、「垂（⿸・⿰）」や「構（⿰・⿺・⿱）」でもありうる。

### ◈ 4.4 統合可能な字形

漢字は符号化にあたって、少々の筆画が異なっても、同一の意味であれば同一の符号が与えられる。JIS X 0208 では、同一視する文字の規準を**包摂規準**と呼んでいる。同様の規定は UCS にも「追補 S: CJK 統合漢字の統合と配列の手順（Procedure for the unification and arrangement of CJK Ideographs）」で定めている。この規定に従えば、たとえば「平」と「平」、「僧」と「僧」や、「海」と「海」は漢字符号としては区別しない [1]。

しかし、部品としては区別しない字形でも、単独漢字としては区別する場合は多い。例えば、「海」と「海」は両方とも同じ U+6D77 だが、「毎」と「毎」はそれぞれ、U+6BCE と U+6BCF である。また、「納」と「納」は両方とも U+7D0B だが、「內」と「内」はそれぞれ U+5167 と U+5185 と、異なる符号を持つ。その結果、「海」という符号化された抽象漢字は、IDS としては「⿰氵毎」と「⿰氵毎」の２つの符号化表現を持つことができ、曖昧となる。

### ◈ 4.5 意味的 IDS と表示的 IDS の相違

漢字を説文等の字源説に従って意味から IDS に展開するのと、字形から IDS に展開するのとでは、IDS が異なる場合がある。例えば「勝」の文字は説文の字源説に従えば「⿰朕力」と展開できるが、字の形だけを見れば、「⿰月券」でも展開できる。

## 回 5 漢字構造記述の正規化

### ◈ 5.1 正規化の必要性

前章では、同じ漢字が、複数の異なる IDS で表現できる可能性を示した。そのため、単に漢字を構造表現するだけでは、漢字を適切に検索・比較することができない。

符号列を比較可能な形式に変換することを**正規化**と呼ぶ。例えば、Unicode に文字合成機能があるため、同一の文字を、合成文字と、合成済み文字の２通りまたはそれ以上の方法で符号化できる場合がある。これらの文字を比較するため、Unicode は正規化アルゴリズム［DWD09］を定めている。

IDS を用いて符号化された漢字を管理し、適切に比較・検索するためには、上述のような曖昧さを無くす**正規化**が必要である。

### ◈ 5.2 字形に基づく IDS

前述の「勝」の例で見たように、**字形**に基づく IDS と、**字源**に基づく IDS は異なる場合がある。

このように字源と字形で異なる IDS が構成される場合は、**字形だけに基づき** IDS を構成することをルールとして明記することで、曖昧さを排除する。

### ❖ 5.3 IDC の正規化

前述の「儲」や「麺」のように、異なる IDC を取りうる符号化漢字に対処し、または「⿲・⿳」のような 3 つの DC を取る IDC を除去するために、まず最初に以下の IDC 変換処理を施しておき、IDS で表れる IDC の種類を、「⿰・⿲・⿴・⿵」の 4 種類に制限する。また、⿵については、⿵ xy と⿵ yx は同じと見なせるため、x と y を画数が少ない順番に並べ直すことで一意性を確保する。

| IDC | 正規化 IDC |
|---|---|
| ⿳ xyz | ⿰ x ⿰ yz |
| ⿲ xyz | ⿲ x ⿲ yz |
| ⿰ xy | ⿲ xy |
| ⿱ xy | ⿲ xy |
| ⿺ xy | ⿰ xy |
| ⿴ xy | ⿲ xy |
| ⿵ xy | ⿲ yx |
| ⿸ xy | ⿰ xy |

### ❖ 5.4 分割点の正規化

前述のような「同じ漢字が⿰ x ⿰ yz と⿰ xy ⿰ z の両方で表現できる」IDS の曖昧性の問題を解決するため、正規化された IDS においては、下記のように同じ IDC が、上位ノードと、下位の最初の子に表れるのを禁止する。

⿰⿰ xyz（禁止）→ ⿰ x ⿰ yz
⿲⿲ xyz（禁止）→ ⿲ x ⿲ yz

この規定を守るために、IDS の正規化処理として、以下のような手順で分割点の変換を行う。

1. IDS の最上位のノードからスタート。
2. 処理対象のノードの下位の最初のノード（⿰の場合は左側、⿲の場合は上側）のノードが IDS の場合は、処理対象のノードと下位の最初のノードの IDC を比較する。
3. この 2 つの IDC が同じ場合、下位ノードの最初のノードを、上位ノードの最初のノードへと移動し、下位ノードの次のノードを、上位ノードの次のノードの最初のノードへ移動する。（図 2）
4. この処理の後、再度、上位ノードと下位の最初のノードの IDC を比較し、これが同じ場合は 3 を繰り返す。
5. 上位ノードと下位の最初のノードの IDC が異なった場合は、2 の処理を下位の 2 つのノードで再帰的に繰り返す。
6. 最終的に DC に到達した時点で処理を終了する。

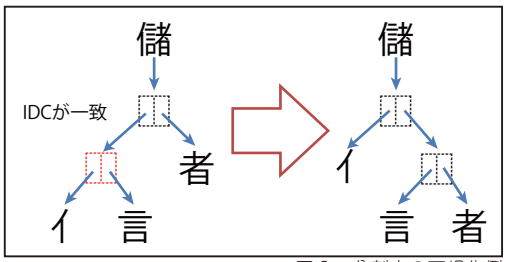
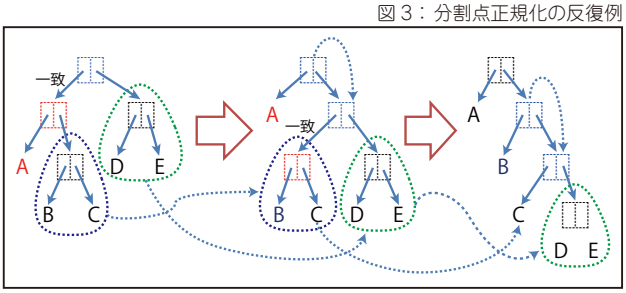


図 2：分割点の正規化例

この処理を行うことで、上位ノードと下位の最初のノードが一致することはなくなる。すなわち、同方向で複数個の部品に分割できる漢字において、常に最初の部分が単独で切り出される形で IDS が表現されることになり、分割点による曖昧さは無

図 3：分割点正規化の反復例



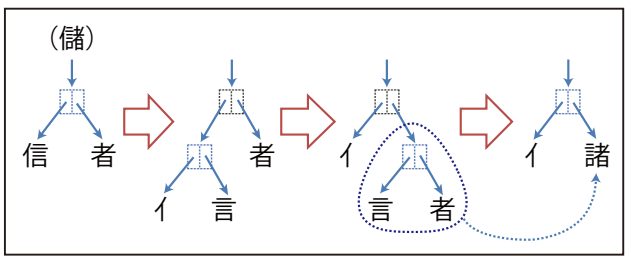



図 4：「⿰信者」から「⿰亻諸」へ

くなる（図 3）。

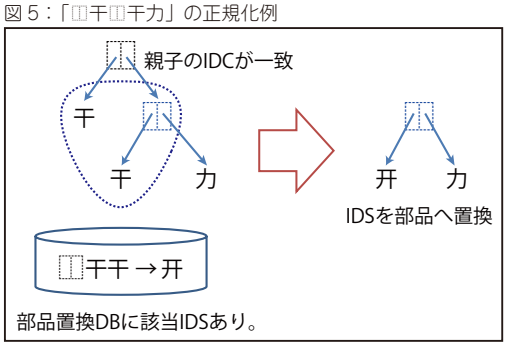
### ❖ 5.5 分割レベルの正規化

前章の「峠」の IDS 表現に見られるような、分割レベルの差違による曖昧さをなくすために、各ノードにおいて簡約可能な文字は、必ず簡約する。ただし、前述の分割点の正規化との矛盾が生じないよう、簡約を行った後で分割点の正規化を行うこととし、その際に、以下の規則を追加する。

- 処理対象のノードに対して、その下位の最初のノード（⿰の場合は左側、⿱の場合は上側）が DC であり、かつその DC が IDS へ展開可能で、その IDS の最上位ノードの IDC が、処理対象のノードの IDC と一致するようにできる場合は、一旦、その DC は IDS に展開した上で、前述の分割点の正規化処理を行う。

この規則を追加することにより、「⿰信者」と「⿰亻諸」とを比較しても同一と見なされない、という問題を防止できる（図 4）。

図 5：「⿰干⿰干力」の正規化例




### ❖ 5.6 漢字部品の正規化

前述したように、UCS 漢字は、「毎」と「毎」は単独漢字としては別符号だが、漢字部品としては、同一視される。このように同一視される部品[川幡 07]は事前にデータベース化しておき、画数が多い漢字部品を、**画数が少ない漢字部品**に置換するようにする。

画数が少ない方に置換するのは、同一視される漢字部品の一方が他方を包含する場合、正規化・分解のプロセスにおいて、無限ループが発生することを防ぐためである。例えば「犬」と「大」を同じ部品と見なす場合、「大」を「犬」に置換すると、置換後に「犬」を分解したら「大」が表われ、これ置換すると「犬」が表れる……という無限ループが現れる可能性がある。画数が少ない漢字に置換するようにすれば、このような問題は発生しない。

漢字部品の正規化は、分割点の正規化より優先する。例えば、「⿱干」は「开」に置換されるべき部品である。しかし、IDS「⿰⿱干力」を分割点での正規化を優先すると、「⿰⿱干⿰干力」となるが、漢字部品の正規化を優先すると、「⿰⿱干力」は「⿰开力」となり、一致しない。この場合は、「⿰开力」を優先させることで、矛盾を解消する。

ただし「⿰⿱干⿰干力」という形で分割された IDS を、「⿰开力」という形へ正規化するには、やや複雑な処理が必要になる。「⿰」が IDC である IDS ノードの左側と、右側の「⿰」ノードのさらに下位の左側ノードの両方を合わせて、置換対象となる漢字部品を構成するかをチェックする必要がある。

そのため、以下のようなデータベースとアルゴリズムで正規化を行う（図 5）。

1. IDS の正規化処理を行う前に、置換対象となる漢字部品で、上下（⿱）・左右（⿰）のいずれかで分割可能な全ての漢字について、IDC をキーとしてデータベース化する。
2. IDS を正規化する際、前述の分割点の正

規化を行った後で、処理ノードの IDC と、そのノードの右側のノードの IDC が一致し、かつその右側のノードの左側のノードの DC と、処理ノードの左側のノードの DC で構成される漢字が、データベースにある場合は、それを置換対象の漢字部品とみなして、置換する（図 5）。

3. 置換後は、漢字の構造が変化する可能性があるため、再度、分割点の正規化と簡約を実施する。

### ❖ 5.7 分割方向の正規化

縦・横の両方に分割できる漢字は必ず、縦方向への分割を優先する。例えば、「焱」という文字は、「⿰炎炎」ではなく「⿱双双」と分割する。この規則は、分割点の正規化よりも優先する。例えば、「⿰焱刂」という漢字を正規化する際、「焱」は「⿰炎炎」と見なすと、正規形は「⿰炎⿰炎刂」となるが、「焱」を「⿱双双」と見なすと、正規形は「⿰焱刂」のままとなり、２種類の正規形ができてしまう。これを回避する。

ただし、一旦「⿰炎⿰炎刂」（または「⿰炎刻」）という形で分割された IDS を「⿰焱刂」の形に戻すためには、前述の部品の正規化同様の処理が必要になる。すなわち、「⿰」が IDC である IDS ノードの左側と、右側の「⿰」ノードのさらに下位の左側ノードの両方を合わせて、上下・左右両方向で分割可能な漢字に変換できるかをチェックする。

そのため、以下のようなデータベースとアルゴリズムで正規化を行う。

1. IDS の正規化処理を行う前に、UCS で符号化された漢字のうち、上下（⿱）・左右（⿰）の両方で分割可能な全ての漢字について、事前に左右分割（⿰）された形を、左側の DC をキーにしてデータベース化する。
2. IDS を正規化する際、前述の分割点の正規化・簡約を行った後で、処理ノードの IDC が「⿰」であり、かつその左側の DC が前項のデータベースにある場合は、以下の手順を実行する。ない場合はここで処理を終了する。
3. 処理ノード（⿰）の右側をチェックする。右側が IDS 表現可能であり、かつその IDS 表現の最上位ノードが⿰である場合は、その下位ノードの左側の DC をチェックし、それが処理ノードの左側の DC と合わせてデータベースにある漢字と一致する場合（すなわち、⿰ xy が、⿱ ab とも表現可能な場合）は、⿰ x ⿰ yz の形を、⿰⿱ abz の形に置き換える。

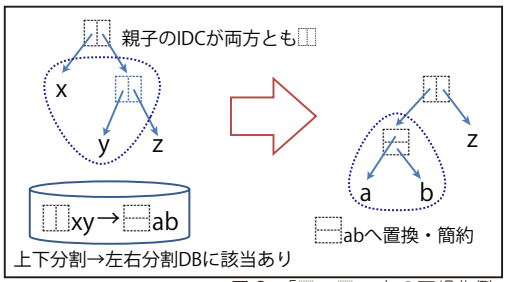


図 6：「⿰ x ⿰ yz」の正規化例

上記の手順で正規化を行うことで、正規化処理における、分割方向の曖昧さを解消することができる（図 5）。

### ❖ 5.8 正規化アルゴリズム

以上をまとめると、IDS の正規化処理は、以下の手順で行う。

1. 事前に、UCS 漢字の IDS データベースを正規化しておく。
2. 漢字部品の正規化において、置換対象となる漢字が上下・左右に分割可能な場合は、置換対象の IDS をキーに置換後の文字でデータベースを用意する。
3. 上下・左右の両方向に分割可能な漢字について、左右分割された形の左側の DC/IDS をキーにしたデータベースを用意する。
4. 比較対象となる IDS を、IDC や漢字部品で正規化し、簡約する。
5. 簡約した上記の IDS を、分割点で正規化し、簡約する。

6. 分割点で正規化したIDSを、**漢字部品の正規化**または**分割方向の正規化**ができないかを、2,3で用意したデータベースでチェックし、正規化可能ならば、各正規化処理を実施後、再度、5に戻る。
7. 6において、これ以上、**漢字部品の正規化**または**分割方向の正規化**ができないことを確認したら、そこで処理を終了する。

なお、5の簡約化において、IDSが単独漢字に簡約することができたならば、そのIDSはすでにUCSで符号化済みであることが分かる。

## 回 6 今後の課題

### ◈ 6.1 提案した正規化手法の評価

今回提案した正規化手法は、上記で述べたような曖昧さは解決できることは確認されているものの、網羅的な比較評価はできていない。今後はきちんとした評価手法を確立して、本正規化手法に問題はないかを確認する必要がある。

### ◈ 6.2 IDS表記の不完全性

UCSで符号化されている複雑な字形を持つ漢字は、その大部分が、IDSで表記可能である。また漢字部品がUCSにない場合でも、台湾の中央研究院が開発したCDP漢字部品外字[2]を利用することで、符号化できる場合が多い。

しかし、未だにUCSの一部の漢字、例えば「俿」や「辰」等は、分割した際に、一方は部品として独立した符号を持たない。

ここで、ある漢字（部品）集合$I$が部品表現として**完全**とは、どう定義できるかを考えてみる。漢字には奇妙な形を持つものも多くあり、それらを全て、分解・部品表現できることに実用性は見出しがたい。一つの仮説として、完全性を「符号化された任意の漢字をIDSで展開する際に、一方が康熙字典部首ならば、もう一方は部品符号として存在する（または符号化済みの漢字のIDSで表現できる）」と定義するならば、今のUCS漢字集合は完全ではなく、CDP外字を加えても、前述のような用件を満たさない漢字が若干、残っている。

今後は、IDSにおけるより完全な漢字部品集合の整備を目指して、必要な漢字部品の抽出・明確化と、GlyphWiki（**http://glyphwiki.org**）による作字等の整備を進めていきたいと考えている。

### ◈ 6.3 異体字処理への応用

漢字文献の情報処理においては、漢字字形表現の自由さと、検索の精度の向上を両立させるのが重要であり、それには、網羅性の高い異体字情報のデータベース化が重要である。異体字データベースはまた、漢字の入力を効率化にも役立つ。

UCS漢字の異体字データベースについては、すでに[処学会07]等があるものの、これら既存の異体字表は、過去の文献等に記された異体字に関する記述に基づくものにしかすぎなかった。

しかし、IDSのデータベースを利用すれば、例えば「参」と「參」の異体字関係から、「惨」と「慘」、「滲」と「滲」、「鯵」と「鰺」のように、異体字を部品として利用する漢字の異体字関係を機械的・網羅的に導出できる。

また「稟」と「稿」のような、漢字の部品の位置関係のみが異なる漢字の異体字関係も、IDSのデータベースを活用すれば網羅的に導出できる。

このように、IDSを利用して異体字関係を類推し、それに基づいて異体字データベースを充実させることで、より漢字文献における入力の自由度、検索の精度を向上させることが期待できる。

## 回 7 最後に

現在、UCSでは漢字符号は既に7万5千近くが符号化されている。規格の利用者にとって、これらの膨大な漢字集合から、必要な漢字を確実に探し出すことは段々困難になってきている。また、もし必要な漢字が存在しなかった場合は、将来の新漢字規格の制定に備えて、字形を何らかの形で符号化しておくことも重要である。

規格の制定側にとっても、たとえば新しい漢字

集合を数千個を既存の漢字集合に追加したい場合、その数千個の漢字の 1 つ 1 つが、既存の漢字集合とは重複しないことを確実にチェックするのは極めて困難になってきている。

本論文では、漢字構造の情報化手法として、IDS を紹介し、IDS が曖昧さを持つために、そのままでは比較等の処理ができないことを説明した。そして、その問題を解決するための、正規化アルゴリズムを提案した。この手法を活用することで、必要な漢字をより確実に見つけ出せることが期待できる。

本論文で提案したアルゴリズムは、その一部はすでに実装され、国際漢字符号の標準化組織 ISO/IEC JTC 1/SC 2/WG 2/IRG において、新しい漢字が提案された際に、それが既にある漢字と重複していないかをチェックするのに利用されている。

今後は、IDS を利用した異体字データベースの拡充や、漢字部品の拡充を図り、より簡単に大規模漢字集合を使いこなせるような環境の実現を目指してゆきたい。

## 注

[1] これらの差違は、漢字字形選択子（Ideographic Variation Selectors ... IVS）を用いて区別することができる場合がある。IVS の詳細については、[HM09]を参照のこと。

[2] 詳細は **http://glyphwiki.org/wiki/Group:CDP%E5%A4%96%E5%AD%97** を参照。

## 参考文献


- ［DWD09］ Mark Davis, Ken Whistler, and Martin Dürst. Unicode Standard Annex # 15, Unicode Normalization. 2009. URL：http://unicode.org/reports/tr15/.
- ［HM09］ Hideki Hiura and Eric Muller. Unicode Technical Standard # 37, Unicode Ideographic Variation Database. 2009. URL: **http://unicode.org/reports/tr37/**.
- ［JTC03］ ISO/IEC JTC1/SC2, ed. ISO/IEC 10646:2003. ISO/IEC JTC 1, 2003.
- ［川幡 05］ 川幡太一 . “国際的な漢字符号の規格化における問題点について .” In: 東洋学へのコンピュータ利用第 16 回研究セミナー（2005）.
- ［川幡 07］ 川幡太一 . “UCS の統合漢字から見る包摂規準 .” In: 東洋学へのコンピュータ利用第 18 回研究セミナー（2007）.
- ［守岡 08］ 守岡知彦 . CHISE / 漢字構造情報データベース . 2008. URL: **http://www.chise.org/ids/**.
- ［川幡 08］ 川幡太一 . 字形ＩＤＳデータ（漢字データベースプロジェクト）. 2008. URL: **http://kanji-database.sourceforge.net/ids/ids.html**.
- ［処学会07］ 情報処理学会試行標準委員会 WG5, ed. 大規模漢字集合の異体字構造 . 情報処理学会試行標準 IPSJ-TS 0008:2007, 2007. URL: **http://www.itscj.ipsj.or.jp/ipsj-ts/ts0008/toc.htm**.

## 会員の皆様へ

従来からの会員の皆様で、BBS 会員から一般会員への変更を希望される方は、会員資格変更の届けを下記フォームから送信していただけますようお願いします。

また、登録情報の変更がありました方も、同様にお願いします。

❖ **会員資格・会員情報変更フォーム**
**http://jaet.gr.jp/JAET-BBS/change.html**
※アクセスには漢情研 BBS の ID・パスワードが必要です。

## 年会費の納入についてのお知らせ

年会費が未納の方は、下記口座にお振り込み下さい。年会費は 2003 年度分が 1,000 円、2004 年度以降は BBS 会員は 1,000 円、一般会員は 3,000 円になります。

郵便振替
口座番号：00120-1-91107
口座名称：漢字文献情報処理研究会

三菱東京 UFJ 銀行 京都駅前支店
口座番号：普通 4571788
口座名称（漢字表記）：漢字文献情報処理研究会代表　師茂樹
（機械振込等カナ入力）：カンジブンケン　モロシゲキ

ゆうちょ銀行
ゆうちょ銀行 〇一九（ゼロイチキュウ）店
（金融機関コード：9900　店番：019）
預金種目：当座
口座番号：0091107
カナ氏名：カンジブンケンジョウホウショリケンキュウカイ

漢情研 2010 年公開講座報告

# 電子出版の動向と諸問題

本年は、Amazon の kindle や iPad の登場によって「電子書籍」が社会的に大きな関心を呼んだ年となった。

本会会員の多くは、なにかしらモノを書くことを生業としておられるかと思う。そのような方にとって、今後「電子書籍」に否応なく対応させられる事態がくるであろう。

本会でも、電子書籍で何が変わるのか・或いは変わらないのかなど、お招きした三人の報告者と共に、会場の参加者を交えて公開討論会を開催した。当日は、こちらの予想を上回る参加者に驚き、また活発な意見のやりとりが行われた。

本記事は、当日の各報告者の報告、及び全体討論をまとめたものである。

当日参加できなかった会員諸賢や、この問題に関心がある非会員の方には、本記事や Twitter のまとめ記事（全体討論まとめに掲載）を是非ご参照いただき、何ほどか資する所があれば幸甚である。

※漢情研公開講座の日程・会場等の詳細については、彙報（P.192）を参照していただきたい。

Contents


電子書籍の動向と検閲問題……田代　真人 … 24

電子書籍とソフトウェアの自由 ── 電子書籍の永続化のために ── …守岡　知彦…… 34

電子書籍（e-book）とプラットフォーム

── 競争法上懸念される二つの問題……石岡　克俊…… 41

全体討論のまとめ ……師　　茂樹…… 48

2010年公開講座報告
電子出版の動向と諸問題

# 電子書籍の動向と検閲問題

田代　真人（たしろ　まさと）

## ◇ はじめに

はじめまして。田代でございます。今日は、電子書籍と検閲というようなテーマで、小一時間ぐらいお話をさせていただきたいと思ってます。

まずは自己紹介からさせていただきます。

先日『電子書籍元年』という書籍を出版させていただきました。その理由は昨年末から、ツイッターやブログで、「印税が70%だ」「いや、80じゃないの？」「90ぐらいもらえるよ」というようなお話が出てきたことによります。そこで、出版社と編集者が著者と読者の間に立って全部暴利をむさぼってるような、そういうようなお話まで出てきました。にもかかわらず、そのとき、出版社の人はツイッターなどで何も発言していないというような感じに思えたのです。そのうち、編集者などが中抜きされて、著者がイコールそのまま本を出せるというのが当たり前のような雰囲気になってきたので、そこで一つ「いやいや、そんな簡単じゃないよ」と。一応私も20年以上の編集歴がありますので、現場というのはそういうものではないと本書によって伝えたかったのです。

例えば、以前、いろいろな本を出している著名な大学教授の方にコラムを書いていただこうと思ってご連絡を取り原稿を書いていただいたら、ほとんど日本語になっていなかったことがありました。今までの彼の本は、出版社の編集者が全部リライトしていたのです。だからちゃんとした文章が書ける人だと私は思っていたわけです。そこで、その方に「ちょっと手を加えさせていただいてもよろしいですか？」とお訊ねしたら、「好きなようにどうぞ」と言われまして、内容を変えずにほとんど原型を留めない文章にして掲載したのです。そういうようなことも実はままあるのです。

私自身は、電子書籍に興味があったというよりは、もともと大学でロボット工学、自動制御の研究をやっておりました。ロボットとは最高のマンマシンインターフェースなんですね。マンとマシンのインターフェース。現在皆さんが毎日のように使っているマンマシンインターフェースはキーボードです。人間のアナログな思考をデジタルに変えるのがキーボードなんです。逆にデジタルなものをアナログな人間の目で見える、網膜に映して見せるというのはモニターの役割です。その二つがマンマシンインターフェイスの代表的なものです。耳に音を伝えるスピーカーもそうですね。これらがどんどんと発達していくと、究極はロボットになって、人に話しかけるようにロボットに話しかければデジタルな情報になり、彼らがまた話しかけてくれるというようになっていくということなのだと思います。このようなことが私のバックグラウンドにはあり、それからファッション女性誌やビジネス雑誌のエディターをやったり、書籍を作ったりしておりました。

ダイヤモンド社には1996年に入りました。当時ダイヤモンド社はdiamond.co.jpというURLドメインだけは持っていたのですが、内容はワードで作ったような素人がアップしただけのようなものだったのです。そこで私が初代ウェブマスターとして、いまのダイヤモンドウェブの前身になるものを1年かけて作りました。

厳密には1年かけて作ったわけじゃないんですけど、リサーチも含めるとそのくらいの時間がかかりました。リサーチの結果わかったことはウェブの本質は検索だということです。

検索することによって、出版社の倉庫の奥の方にある書籍を消費者の前面に出すということが必要だと。それを消費者の前に出したうえで販売できるという仕組みが絶対に必要だと思ったのです。そこでデータベースを構築することにしました。予算もないので安くデータベースを構築して、書籍を買えるようにするというところまで1年かけて作り、97年の9月2日にオープンしたのです。

そうやってオープンしましたが、とはいえ、その時、ウェブの広告と比べて、まだまだ雑誌のほうが広告単価が高いなと思いました。ウェブの広告は全然安い。同じ人間が同じだけ動いて、片方は20万しか取れなくて、片方は100万円取れると。だったら、100万のほうだよねという感じでした。そこで私はウェブと連動型の雑誌媒体を作ったりしていました。その傍ら97年にアクロバット日本語版が、バージョン3.0になって初めてやってくるというので、じゃあ、7割くらいDTP化されていた『週刊ダイヤモンド』の最後3割を一気にDTP化して、そのままPDFにして電子版で売ってしまおうと考えたのです。

それでアドビの方に協力を仰いで、結局1年分くらいをPDF化して、『週刊ダイヤモンド』の定期購読者だけに、無料サービスとして全文検索できてPDFで読めるようにしたのです。これは、結構評判もよかったのですが、データベースも作りたてのものを改良しながらどんどん使うということをやっていたので、なかなかコスト的に維持するのが難しくなっていって、それで結局2年か3年ぐらい運用してなくなってしまいました。

ただ、その時おもしろかった経験を一つお話しすると、当時、著者の皆さんの権利をどうしようという話になったのです。そこで、実は去年グーグルで話題になったオプトアウトという方法をとりました。「私のは出さないでくれ」と著者が言わない限り承諾したとみなして出しますよというものです。そしてダイヤモンド社の著者データベースに入ってる著者の皆さんに案内の文章をお送りしたのです。一斉に。結構乱暴なことをやったのです。トータルすると大人数になります。しかし何もしないというわけにはいかない。だったらもうこの方法しかないだろうと。発想はグーグルさんと同じなんです。そういうことを97年にやりました。そうすると、たった1人だけ「出してくれるな」という返事が来たんですね。現在でも活躍している著名なジャーナリストの方です。彼女だけは「出してくれるな」と返事が来ました。

そこで「どうしてですか？」と、連絡を取って会いに行ったわけです。たった1人ですから。しかも、人気コーナーだったので、それを出さないというのもちょっと困るなと思って。

そうして本人に確認したら、「いや、わからないのよ」と。「わからないから取りあえず駄目」みたいな、そういう感じだったんですね。それで私が「いやいや、ちょっと待ってください」と。「ほんとに悪いようにはしませんから」といって、交渉して結局OKになったのです。

結果的にそういったかたちで、ある種電子書籍と言えば電子書籍、というか電子雑誌ですけれども。現状やってる、みんながいま本当にやろうとしていることを当初やっていたのです。

## ◇ 電子書籍〝騒動〟

今回の電子書籍の騒動……〝ブーム〟という声もありますが、ブームになるより、まだ騒動の域を出てないなというのが私の個人的な印象です。「ブームになっているわけではなくて騒いでるだけでしょう」という印象です。

まだ、いろんな会社がさまざまな電子書籍リーダーを発表し、サービスを発表しているという状況です。あまりに多く、出版社としても何を使って、どのようなことをやればいいのか全然わからなくなってきてしまっている。たぶん何かに収束していくだろうとは思っています。

まず今回の電子書籍騒動と出版ビジネスの現状をお話しますと、もう皆さんご存じのように2007年にアマゾンがキンドルを発売しました。そのころは、あまり大きな話題になるということでもなかったのですが、昨年2月、そのキンドルがバージョンアップしたのです。発表会では、ベストセラー作家のスティーブン・キングがキン

ドルだけでしか読めない小説を発表するなどと話題性もあったのです。そして秋に日本でも売り出すという話になり、それに加え、ご存じのようにiPad が今年発売になるというウワサも出てきたわけです。

先ほど申し上げたように、日本でも 10 年以上前からみんな電子書籍の発売などはやっていたわけですが、今回なにが違うのかというと、やはりこういう端末の登場がポイントかなと思うのです。つまり、これらは何が違うのかというと、最初にマンマシンインターフェースのお話をしましたが、iPad もキンドルも……キンドルは下の部分にキーボードのようなものがあるのですが、iPad に至っては、キーボードもなくて画面だけですから、ソファなどでリラックスしながら読めるという端末なのです。読むよりは動画を見たりとかネットを見たり、メールを見たりという、そのついでに本を読むという状況が現実には起こると思うのです。

そこで「書籍を電子化して読めるよね」「読むことが違和感なくなってきたね」というふうに消費者の方々が気付いて、それで先ほどの印税の話やいろんなこと、自分でも出版できるというような……アマゾンが DTP という自分で出版できるシステムを作ってますが、そうやってセルフパブリッシングができる。アップルでもできるようになるだろうというような期待感で、騒動になっていったと思います。

「騒動」というのは、出版社も同じで、急きょ 31 社で日本電子書籍出版社協会というのを立ち上げるなどしたわけです。この協会も以前からやっているサービスの〝電子文庫パブリ〟というのを前身としてますので、突然やったというよりは、そこにいる人たちが手を挙げて「みんなで結束しよう」といって協会を作ったというのが実情です。

## ◇ 委託販売制と再販制に守られた出版ビジネス

ここでおさらいです。現状の出版社ビジネスは出版社と書店、その間に取次があります。通常、出版社というメーカーが取次という卸会社にできあがった本を納品します。

日本の出版の場合は、再版制度と委託販売制度という二つの制度に守られてると言われてます。委託販売制度というのは、つまり「売れ残ったら返品ができますよ」という制度ですね。 この制度では、出版社と取次会社、取次会社と書店、それぞれに契約を結んでいます。その契約の下、出版社が取次会社に本を納品すると、返品できるので買い取りではありませんが、一応取次は出版社に代金を支払ってくれるのです。しかし全部じゃない。売れるか売れないかわかりませんし返品できますから。出版社によって時期や金額は異なるのですが、その本の売り上げが立たないうちに、翌月から半年くらいあとに半分ぐらいの金額を入金してくれるのです。これがものすごく大きな意味を持っている。要は、本が売れようが売れまいが入金してくれますから、これは大きい。

次に取次は全国の書店に配本します。書店も入荷分のお金を幾ばくか払う。しかし、同時に書店は売れなかった本を返品をする。そして、その返品分のお金も取次から、また返ってくるのです。それを相殺してお金のやり取りをするわけです。そのあたりの事情で、本を疑似貨幣という言い方で表現する方もいらっしゃいます。つまり本が現金のように扱われていて、書店にとっては返品してお金が入ってくるというような状況にあります。

取次に返ってきた本は出版社にまた戻っていきます。これら返品されてきた本は、出版社の中で在庫として眠ることになるわけです。そのとき出版社は最初に半額ぐらい入金されていますから、ここでもその分と相殺するのです。返本分と売れた分と。売れなかったら返金しなきゃいけないし、売れれば追加で取次から代金を回収できます。回収できるのは、本がほとんど売れなくなった 1 年ぐらいあとになります。このように出版ビジネスは動いているのです。

次に書籍のコスト計算をしてみます。これは私の本にも載ってますけれども、ざっくりお話をすると、紙の本の場合は、例えば単価 1,000 円の

書籍を初版で 5,000 部出したとしましょう。最近は小さな出版社はもう 5,000 部も刷れないんですが、いまだに大手や中堅どころは 5,000 部から 6,000 部くらいの初版を印刷します。そうすると、売り上げは 500 万円です。そのうち、書店は 22％ぐらい取ります。この辺の数字は大体と考えてください。取次が 8 ％、印税が 10％。加えて、印刷製本費、デザイン DTP 費、出版社の経費を入れると 5,000 部を売り切ってトントンなんです。ただし、ご存じのように返本率が現在 40％と言われています。しかし平均なんですね。平均で 40％ということは、村上春樹の『1Q84』も入ってると考えると、実はほとんどが 40％以上の返本率です。

私の感覚では、３割以上実売で売れてる本は３割ないです。結局それぐらいの量の本が大量に返品されている。先ほどの計算からすると、単価 1,000 円で、例えば 60％が本当に売れたとすれば 3,000 部実売ですから 300 万円の売り上げ。書店と取次はロイヤルティー契約ですから、金額はそのパーセンテージを支払います。印税も 10％と言いましたが、この印税、初版部数の 10％を 3,000 部しか売れなくても最初に支払うのです。出版社が。実売などの結果が出る前に。ですから、初版印税は、実は固定費なんです。固定費として 10％がかかっているのです。その他、印刷、製本、デザイン、出版社の経費というのは、売れようが売れまいが変わりませんから、これらも固定費。そうすると、大体 1 冊当たり 140 万円の赤字になるのです。ですから、1 か月 10 冊本を発行して、全部が返本率 40％だとすると、ひと月で 1,400 万円の赤字になるのです。

それはどこで賄うのでしょうか？　10 か月経てば、1 億 4,000 万円の赤字になります。実はこれは、資本を崩すしかないのです。崩しながらも発行して、発行した半金は取次から現金が入ってきますから、それに足して少しずつ少しずつ回していってるという感じです。

それなのにどうしてやめないかというと、実は、ベストセラーが出ると大きいのです。例えば 1,000 円で 10 万部売れると 1 億円売り上げます。そうすると、書店、取次、著者は、キッチリとロイヤリティのパーセンテージ分もらえます。とくに著者にとっては、初版分の印税はもらえて、それ以上売り上げると追加でもらえるので、本当にリスクが少ないビジネスになっています。

印刷、デザイン、出版社の経費は、10 万部売り上げてもそんなに上がりません。もちろん印刷はしますから、10 万部印刷すると 2,000 万ぐらいかかるのでその分は上がりますが、出版社にとってはトータルで 4,000 万円弱ぐらいの利益になるのです。このようにベストセラーが出れば大きい。だからやめられないというわけです。

もちろんビジネスだけで出版事業をおこなっているわけではない。ある種の使命感をもってやってる会社も数多くありますから、そうはざっくりとは言えないのですが、やはりベストセラーが出ると大きいのです。言ってみれば 1 枚 140 万円の馬券を毎月買っているようなもの。しかし当たれば大きい。そうすると、借金してでも馬券を買う人がいっぱいいるように借金してでも本を出す。そういうような状況になっているのです。現状は。

しかも前述のように取次に本を納品すると一応売り上げの半分ぐらいが入金される。となると、やはり幾分かは回していけるので、取りあえず本を発行しようという状態にもなるのです。1 タイトル出すと、例えばさっき言った 500 万円という売り上げの半分、250 万が翌月もしくは半年後ぐらいにいただけるとなると、やはり大きい話なのです。

出版社にとっては事業計画として、例えば編集者を何人雇っているので、というふうに人件費を考えると、1 か月の間に絶対に 10 冊出さなきゃいけないとか決まってしまうのです。そうすると、例えば出版できる書籍が 8 タイトル分の企画しかないと「どうしよう」と焦るわけです。あと 2 タイトル。絶版本から探し出して著者に話をして、前書きと後書きを書き直して、それで新刊として出すわけです。そうすると取次は、売れようが売れまいが入金してくれるので、取りあえずそういったかたちで書籍をそろえるようなことをやってるのです。

## ◇ 電子書籍の売り上げとコスト

次に電子書籍における売り上げとコストを考えてみましょう。電子書籍になったからといって極端にコストがかからなくなると言うことはないんですね。紙の場合、書店と取次に支払うのは合計30％ですが、電子になってもアップルに配信料を 30％支払わなければならない。これは紙と同じですから変わりありません。デザイン DTP 費や出版社経費も必要です。

となると、電子書籍になっても印刷代と製本費しか値下げできないのです。もちろん倉庫代など、ほかにいろいろなものも必要ないのですが、それらの占める割合はそんなに大きくないので、１冊当たりで考えると、最も割合が大きい印刷製本費の分だけ値下げできると考えられます。

そうすると 1,000 円の本は、印刷製本費にあたる 250 円分だけ抜いて 750 円の単価にできるということです。こう考えると「電子書籍になれば半額だ」とか「１／３だろう」という声もありますが、そんなに安く出せないのです。これがなぜにこう計算せずにみんな簡単に安く出せるような幻想を抱いてしまうんだろうかと思うのですが、こうやって考えるとそんなに安く出せないのです。

一方、紙の場合著者の印税は 10％ですが、電子化になったら「電子化になるんだから印税５割ぐらいもらえるだろう」と著者の方は考える方も多いのですが、そんなに支払えないのもわかると思います。しかし、その声に応えて「印刷製本費分の 25％だけを上乗せして 35％にします」という提案はできます。

しかし、著者に対する印税を 25％上げると、結局定価を紙と同じ 1,000 円にしないと出版社としては合わなくなってくるのです。結局のところ、紙の本と同じ値段でしか出せないということになります。著者にとっては印税率が増えますからいいとは思うのですが……後述しますが別の理由で手放しで喜べるとはいえません。

しかし、通常の消費者の声を聞くと「電子書籍は 450 円ぐらいが適正な値段でしょう」というのですね。「そのぐらいが一番買いやすいでしょう」と……。

では、仮に 450 円で電子書籍を販売するとしましょう。そうすると 3,000 部が売れたとしても私の試算では結局 130 万近くの赤字になるのです。赤字を回避するためには固定費を下げるしかないのです。しかし出版社の経費、デザイン DTP 費も半額にしても 40 万円の赤字になります。では、定価を 600 円にしたとしましょう。

余談ですが、アップルのアップストアというのは、好きな定価を付けられるというふうな言い方をしてる人もいますけれども「アップルが決めた定価の中から好きな定価を選ぶことができる」というのが正確です。一番安いのが無料。次は 115 円です。次は 230 円です。次は 350 円。そして 450 円。その次が 600 円なんですね。500 円という値段は付けられないのです。それが国ごとの値段に関連づけられていて、例えば 115 円はアメリカで 99 セントというふうに事前に決まっているのです。230 円は 1 ドル 99 セントです。ですから、その間の数字というのは選べない。だから 450 円でちょっと足が出るからと思って、日本的な感覚でワンコインで 500 円にしようと思っても、それはできないのです。

さて、そこで定価を 600 円にしましょう。そうすると「3,000 部完売した時にやっと利益が３万円出る」という計算になります。

しかし、電子書籍で 3000 部の売り上げと考えるのはあまり現実的ではありません。そこで非常に現実的なことを考えてみましょう。

そう考えると、やはり定価は 450 円が妥当です。そして 1,000 部だけ売れたと仮定しましょう。その場合、売り上げが 45 万円。出版社の経費は５万円、デザイン DTP 費は２万円とします。そうすると利益が８万 7,500 円となります。この場合の経費やデザイン DTP 費は、もはや外注する値段じゃありません。つまり、SOHO（Small Office Home Office）なのです。しかも著者も兼ねる。そのときに初めて 31 万 5,000 円という、７割の印税がもらえるという計算になります。すべて１人でやって印税が 70％もらえるわけです。

しかし、とはいっても自分で取材などもして、本を書いて、デザインのDTPの作業もしてるわけです。それを1か月やって31万5,000円です。そのくらいしか稼げないわけです。そうすると毎月1冊出してやっと30数万円入る生活です。生活はできると思います。デフレですし……。生活できる。だけれども、毎月、本を書けますか？ということになるわけです。書けないですよね、やはり。大変だと思います。ブログを書くのとは違うのです。本を書くということは。お金をいただきますから質も問われます。このあたりを考えると、最近の議論は、ちょっと現実的ではないものが本当にいっぱい出てきてるなっていうのが、私の感想です。

## ◇ 電子書籍でなにが変わるのか？

さて、電子書籍で何が変わるのでしょうか。先ほどお話しした取次は、電子書籍でも専門の会社はありますが、売り上げに関わらず出版社に前金を払うということはありません。売れた分だけの支払いになります。では、もし出版社が著者に執筆をお願いしようとしたとき、著者にどうやって仕事してもらうのでしょうか。紙の書籍とは異なり「初版印税はありませんよ」と言ったら、やはり著者は書かないと思うのです。いくら印税率が35％であっても、完全実売制で、「今月は10冊売れたので35％の印税率で3500円の入金です」と言われても、著者は困りますよね。であれば、紙の出版社で出版できる著者は、そちらで出します。紙の出版社で出せない人は、結局インディーズとしてSOHOで著者として自主的に書くしかない。がんばってガンガン書く。これは前述のように、そんなに簡単ではない。しかし、それを簡単に出そうとする。つまり、質が低下してくるだろうと想像できる。いま、ブログがいっぱい出てますが、あのブログが全部電子書籍になって定価が付けられたというふうに考えていただければいいと思います。そういうふうに考えていただければ、どのような電子書籍があふれるのか想像が容易に付くと思います。

## ◇ 書籍、雑誌、新聞の電子化とは？

では書籍、雑誌、新聞の電子化というのはどういうことなのでしょうか。消費者視点でニーズを考えてみましょう。まず消費者は読みたいときにすぐに手に入れたい。そして、紙の書籍のように場所を取りたくない。保存スペースを最小にしたい。そして、先ほどお話ししたように安い値段で手に入れたい。

しかし、現状の日本の本はとても安いですよね。再版制があるから。日本書籍出版協会などが再版制の話をするときに「再版制なくなると本の値段上がりますよ」と言うのですが、そう言うと一般の方々は「またすぐそんな言い訳して」っていうふうに怒ったり、過激に攻撃をしてくる方がいるのですが、本当に再版制があるから安いという側面も否定できないのです。出版社は、値引き販売されないので明確に計算ができる。計算ができた上でビジネスができるので利益が確保しやすいのです。

再販制がなければ値引き販売されることを考えなければならないので、高い値段を付けて「このぐらいまで値引きされても大丈夫」という値段を付けていくわけです。それは普通の小売りを考えてもそうだと思います。値引き幅を考えながら原価計算から何からしていきますから。ですから、再販制のないアメリカの場合はハードカバーで1冊当たり26ドルくらいすると。

また、アメリカの本はどうしてハードカバーが多いのでしょうか。本というものは作るときに、われわれもそうなんですが、厚みなども考慮するのです。

例えば「2,000円ぐらい値段を付けたい」と思うと、「やはりハードカバーのほうがいいよね」と考える。あまりに内容量が少ないと、「これで1,500円の値段付けられないからもっと厚くしなきゃいけない。じゃあ文章を増やしてください」と著者に頼んだりします。それでも「もう無理ですよ」と著者の方が言うと「じゃあ紙を厚くしましょう」と。やはり、ある種、商品を作っている

という意識があるのです。そのようにして本の値段は付けられているわけです。だからアメリカの場合は、ハードカバーにして重くして「BOOK」を制作する。そういうものを出せば定価が高くても「これなら 26 ドルという定価付けてもいいでしょう」となるわけです。そういう多少まやかしめいたやり方になっているんですね。

さて、読者のニーズ、もう一つです。検索して、欲しい情報をすぐに閲覧したい。このあたりが電子化に求めるところだと思います。もちろん電子雑誌になってマルチメディアで動画を見たいというような欲求もあると思います。楽しいと思います。そういうものが出てくると。ただし、創り手にとっては大変です。私は、月刊誌の編集をずっとやっていましたが、月刊誌を 1 冊作るだけでもすごく大変なのです。月刊誌はだいたい 3 か月ぐらい前から企画を始めるのです。 3 か月前から始めるということは、いつも 3 冊分の企画を抱えてるということです。3 号分です。それを同時にやっています。企画して取材してカメラマンに撮影をお願いして、と、それだけでも忙しいのに、編集長が「ビデオカメラで今度の対談撮ってきてくれ」とか「撮影風景撮ってくれ」と言われてもできませんよね、はっきり言って。しかも、撮って終わりじゃないです。動画は撮り終わったら編集しなければなりません。到底できないのです。となるとどうするか？　専用の人を雇わなきゃいけないわけです。雇う。つまり、コストがかかるんです。コストがかかる。じゃあそのコストをどこに転嫁するんですかと……。そうすると値段？　電子雑誌の定価を上げられるかというとそうはいかないだろうと思うのです。

「創り手が恐れること」は、書籍や雑誌の低価格化です。そうなると売り上げは減少、利益も減少します。これは仕事単価の低下に繋がります。

また、一度上がった値段は低価格化という波に乗ってしまうと上げられないわけです。たとえ制作費としてコストが上がっても。

ちなみに、アメリカに『TIME』という雑誌がありますが、彼らは iPad 発売と同時に iPad 版 TIME を出しました。 1 冊 4 ドル 99 セントです。iPad 版です。『TIME』の紙バージョンがあります。 1 年間に 56 冊出ます。56 冊で 20 ドルです。それで宅配してくれる。それなのに「iPad 版が 4 ドル 99 セントってどういうこと？」みたいな気持ちは普通に生じると思います。実際、アップルのアップストアで『TIME』iPad 版の評価は散々なものです。みんながコメントしているのですが、「あり得ない」とか「この値段はなんだ」と。やはりそうなるわけです。となると、先ほどの、上がったコストをどこで回収するのか？　回収できるのでしょうか？　広告費としてそれほど入るのでしょうか？　それも疑問です。『TIME』iPad 版の 1 号目は相当入ったと言いますけれども、毎週毎週入るのでしょうか。そういうことを考えなければならないわけです。であれば、現実的には何か違うところで稼がない限りは、電子雑誌のマルチメディア化というのはそんなに広がらないというふうに個人的には思っています。

## ◇ 電子化でジャーナリズムはどうなるか？

今日の主題にちょっと近づいていきます。

紙媒体の電子化でジャーナリズムは変化するのかというお話です。ジャーナリズムとはなんですか？　というところから話さなければならないと思います。

報道やジャーナリズムはなんなのかと考えるとき、新聞は社会の木鐸（ぼくたく）と言われてますけれども、ではその社会の木鐸とは何かというと、世の中がおかしい方向にいっているときに警鐘を鳴らす役割としてあるわけです。ですから、第 4 の権力というか、ジャーナリズムというのは、権力をもった国家などがおかしい方向に行こうとするときに、やっぱり警鐘を鳴らす役目なんですね。それはある程度影響力を持っていないと、警鐘も遠くに鳴り響かないだろうと……。その役割をマスコミがいままで担ってきたわけです。

ただ、そのマスメディアというのが最近は力としてそろそろどうなんだろうと思います。つまり、インターネットが出てきたうえで、だれでも見る

ことができ読むことができるブログなど媒体があるといった状況で、はたして紙の媒体が電子化になってジャーナリズムというのは変化するのであろうかと。佐々木俊尚さんや上杉隆さんみたいに、自分でツイッターでどんどんと書いていって、それが多くの人にリツイートされることによって、ある種影響力を持ちながら社会に対する警鐘を鳴らすことができるのであれば、もはや新聞の役割は、必要ないとまでは言いませんが、多少弱ってきている。それに代わるものがあるというのも確かなのかなというふうには思ってます。

あと客観報道について。新聞の内部ではよく「客観報道して、右にも左にも偏らないような中立な報道を心がけましょう」みたいなことを言うんですね。ただし、客観報道とはなんだろうかというところから考えると、つまり、この題材を取り上げるということがすでに主観ではないだろうかと思うわけです。となると、ほとんどの報道に客観というのはあり得ないんじゃないかとは私は思っているのです。以前、朝日新聞の本多勝一さんが『貧困なる精神』で相当書いてましたけれども、結局テレビであったって、テレビは真実を映してると言いますが、そのカメラを構えてどこを映そうと思うかというと、カメラマンの主観が入ってるはずなんですね。そうすると、激しい戦争を起こしている後ろ側の穏やかな山だけをテレビで映したら、観ている側から「ここではなにも起こってない」と思われても、それは客観報道をしたことにはならないわけです。主観になるだろうと。そういったことを考えると、実は今後ますます主観報道が出てくるだろうなとは思ってます。で、その主観報道というのは、どうやって発信されるんだろうかと。といったとき結局ビジネスつまりジャーナリズムで食べていけるのかと考えるわけです。

アメリカではすごい勢いで新聞が消えていってます。これは食べていけなくなったからですね。つまり、儲からないということです。アメリカの新聞は、大きく広告に頼っています。ですからこれだけ経済が弱ってしまうと、なかなか新聞を維持できないというところがあります。また、これは、雑誌も同じですが、広告を入れると、結局のところスポンサーのことを配慮しなきゃいけないので、そこにはジャーナリズムがあるのか？　という話になります。そこが非常に難しい。

ただ、日本で最近出てきたちょっと大きめなニュースサイトあるんですが、そこはスポンサーとして大きな外食産業が付いてるんですね。しかし、Web を見る限りどこにもそんなことは書いていない。私はそういうかたちがもしかしたら今後のジャーナリズムの姿になるのかなと思っています。そのようにお金を持ってる会社や個人がジャーナリズムを支える。ただし、我々が知っておかなきゃいけないのは、そこのニュースサイト、ジャーナリズムの裏側には、「ここがお金出してるよ」っていうことも本来は明らかにしなきゃいけない。肉中心の外食産業が、魚業界をもしかしたら「おかしい」っていうかもしれない。そのときには、そういう肉中心の外食産業の影響で報道しているのかなと。そういうことまで判断しなきゃいけない。つまり、読み手がメディアリテラシーというものを持っていなきゃいけない。そういう裏側までわかったうえで情報の中身を判断する。そういう時代になってくると思うのです。日本の場合、メディアリテラシーの教育をちゃんとしないので、そこをちゃんと誰かが教えていかなきゃいけなくなるんだろうと思います。

では、紙媒体の電子化で言論の自由はどうなるのでしょうか。いままでの紙の出版社は、言論の自由を実現するために取次とともに存在するという側面もありました。しかし、これだけブログが出てきてだれでも発言できるのであれば、憲法で保障されている言論の自由は、ブログの登場以来ほとんど叶えられているといえます。

## ◇ アップルの〝検閲〟は問題か？

アップルの検閲問題。私は個人的にはアップルの検閲問題といわれるものは、〝検閲〟じゃなくて〝選択〟だろうと思っています。自分のお店にどういう本を置くかは店長が決めればいい。そこに置きたくないというのであれば、それはしょう

がない。例えば、ディズニーランドで「どうしてポルノ雑誌を置いてないんだ」と文句言っても仕方ない。ディズニーランドなんだから。アップルをディズニーランドと考えれば、そこに文句を言っても仕方ないと。もちろん言論を扱うから問題にしていることもよくわかるのですが、公権力ではなく、一私企業ですからそこに対して文句言ってもしょうがないなと思います。

ただ、ビジネスをしている人にとっては大きな問題ですね。大変だと思います。開発費を数百万円かけて、それで申請して OK だったのに、自分たちの判断基準を変えていきなりそれが削除されるわけですから。開発費を回収する前に削除されると、その被害は大きいと思います。基準をしっかりと示してほしいという気持ちは理解できます。

ただ、こと電子書籍に関して、アップルが卑猥（ひわい）なものを出さないというのであれば、それはそれでしょうがないんじゃないかなとは思ってます。

一方アマゾンです。昨年、彼らはジョージ・オーウェルの『1984』を削除したという事件がありました。これは、再版権をもっていない出版社が販売していたので、アマゾンは単純に削除して、ユーザーにはギフトカードを贈ろうとしたわけです。しかし、アマゾンのキンドルというのは、アマゾンのサーバー上に電子書籍を購入した個々のデータ、購入したデータを保存してるんですね。それを有無を言わさず全部削除しちゃったということですが、ここで一番の問題は、それだけではないのです。キンドルは、その電子書籍上に自由にコメントを入れることができるのです。メモを書ける。そこで、高校生から訴えられました。「『1984』の本がほしいんじゃない」と。「中に書いてるコメントがほしいんだ」と。「どうして僕のメモまで消しちゃうんですか」と。そういう裁判を起こした。それでアマゾンは負けたわけです。1,500 万円の賠償命令が出ましたが。それと同時に「ごめんなさい」と、ユーザーみんなにアマゾンは謝って「二度とこういうことをしません」といって、みんなに本を返したのです。「いらない」という人にはクーポン券を渡すなどした。元のメモが付いた本もそのまま復元させたらしいです。

アマゾンには最初にお話しした DTP というセルフパブリッシングの仕組みがありますが、彼らは検閲は絶対しないと言っています。ただし、成りすましとか著名な著者を語ってアップロードしたりとか、人のものをアップロードしたりするようなことを防ぐために、そういう嘘については相当厳しく見ているそうです。暴力的な表現や猥褻なものについても、よほどおかしいことがない限り削除はしないそうです。ただし、DTP システムでの利用契約を見ると、公序良俗に反するものなどは、いつでもアマゾンが削除できると書いていますので、削除の権利は彼らにあるようです。

さて、グーグルです。グーグルは基本的に紙になった書籍のみをスキャンして Google エディションという電子書籍販売システムで販売します。日本では来年の初めになるようです。

彼らは、紙で発行された書籍が、ある種の品質を担保していると考えています。世に出していい品質を持っていると。一種のフィルターとして考えている。ですから、自由に電子のまま電子書籍を売るということはしないそうです。

## ◇ 日本独自のシステムは必要か？

あまり言われてないのですが、アメリカには愛国者法という法律があります。2001 年の 9・11 事件のあと、たった 1 か月間ぐらいの審議で米国愛国者法というのがアメリカにできました。これは国家権力によって簡単に企業のサーバーの中身を見ることもできるし、利用の停止をしたりすることもできるそうです。大きな公権力によってテロを防止する目的で制定されました。

ここで一番重要なことが、いまお話をしたアップル、アマゾン、グーグル、全部アメリカの会社ということです。すべてサーバーがアメリカにあります。アメリカにあるサーバーというのは、愛国者法の適用範囲に入っているんです。ですから、ダウンロードする書籍は問題ないとは思うのですが、アマゾンは、常にサーバーの自分の書籍と同期をとっていますし、グーグルに至ってはダウンロードしません。となると、いつ自分が購入した

電子書籍が読めなくなっても文句が言えないわけです。キンドルで書いたメモも、差し押さえられてアクセスできなくなるということが起こってもおかしくない。そう考えると、個人的にはやはり日本独自のプラットフォームというのは必要かなと思います。日本国内にサーバーを置いている電子書籍がないと、すべてアメリカに握られますというのは、日本愛国者の１人としてやはり好ましくないと思うのです。

アップル万歳、アマゾン万歳、グーグル万歳と言って、「全部乗っかっちゃえばいいじゃない」と結構乱暴に言ってる方もいらっしゃいますが、「そうじゃないだろう！」と思います。日本独自のプラットフォームの必要性というのはあります。

そのような事情もあって、アゴラブックスという電子書籍専門出版社に参画したのです。これはアスキー創業者の西和彦さんと経済学者の池田信夫さんと立ち上げたのですが、こういうことをすることによって、日本独自のプラットフォームができるんじゃないだろうかと思ってやったのです。アゴラブックスはクラウド型の電子書籍です。ストリーミング型です。iPad でも見られます。Ajax リーダーという、JAVA スクリプトを使ったリーダーを使用してブラウザで読めます。また、国内にサーバーがあります。我々は、こういうインフラも含めて皆さんに提供できたらいいなと思いながら、一番読みやすいものにすべく、検証や実験をおこないながら最適化しています。

しかし、クラウド型だけではなくて、ダウンロードして読みたいという声もあるので、PDF ファイルをダウンロードできる書籍もあります。ただし、PDF 型式で電子書籍を販売する方法は、著作権が守れないと言って出版社が許可してくれないので、それらはクラウド型で提供しています。

## ◇ 最後に

電子書籍に関してはネガティブな発言もあったかと思いますが、私は「ニーズのあるところにビジネスは芽生える」と考えています。ですから、成功する電子書籍ビジネスは絶対に生まれると思っています。ただ、変化しないとどうしようもない。創り手も変わらなければならない。ではどう変わるか？　創り手にとって、最も変えなければならないのはコスト構造です。

私は、出版社の方にも「変わらないと生きていけませんよ」「残っていけませんよ」と言います。彼らからは「何を変えればいいのですか。発想ですか？」と聞かれますが、そういうときにこう言います。「明日から月給が半分、年収が半分でになるとあなたは生活を変えますよね？」と訊ねます。そのくらい変わらなければ、実はやっていけないのです。それほどの変化を時代に求められる。しかし、大手の出版社は年収を 1,000 万以上もらってますから、ほとんどの人は「ノー」と言う。彼らは「ノー」と言ったまま、たぶん変わらない。紙の書籍があれば、取次がお金を入れてくれる。著者にとっても、紙の出版社と付き合っていたほうが初版印税をもらえる。リスクを回避できる。だから、そう簡単には変わらないと思います。

ではいつ本当に変わっていくのか？　iPad が今年出てきました。だから電子書籍元年と言っています。iPad が出た今年生まれた赤ちゃん。20 年後に成人になります。「そのぐらいでようやく」と考えたほうがいいかもしれません。紙の書籍が出せる著者であれば、お金にならない電子書籍だけを出版することはないと思います。よほど資産家じゃない限りはなかなか電子書籍に行かないと思ってます。儲かっている IT 企業が初版印税分にあたるお金を電子書籍の著者に払ってくれるなら別ですが。

となると、紙の出版社からお声がかからない著者、インディーズから電子書籍は火が着くと思います。しかし、インディーズがいったん出てきはじめたら、その後は早いでしょう。それが５年後なのか 10 年後なのかわかりません。

よく私は「出版社にいたから保守的な出版社側に肩を持った意見を言う」と言われますが、別に電子書籍の発展を否定しているわけではありません。むしろ肯定したいのですが、現状を冷静に分析するとこのような結論になるのです。その点、ご理解ください。本日はありがとうございました。

2010 年公開講座報告
電子出版の動向と諸問題

# 電子書籍とソフトウェアの自由

## ――電子書籍の永続化のために――

守岡　知彦（もりおか　ともひこ）

ご紹介していただきました守岡と申します。私は今回、ちょっとこの場には場違いかもしれないですけど、ソフトウェア、特にフリー・ソフトウェアを作る立場からのお話ということでオファーを受けて、そういう話をさせて頂こうかなと思います。

電子書籍自体は結構前からある訳ですが、Kindle とか iPad（この会場でも iPad を持っている方が結構いらっしゃるようですが）とかの登場で「電子書籍がどうも今度はくるんじゃないか」っていうような予感につながってるんじゃないんじゃないかと思うんですが、個人的には実はちょっと憂鬱なんですね。さっきのお話とはまた別の観点でなんですが、収入が減るっていうのも結構困りますけど、私の場合、自由じゃないっていうか、今まで紙の本だとできたことができなくなったら困るよなっていうのがあります。あと、もう一つは、iPad って本質的にコンピューターな訳ですよね。だけど、今までの PC なんかでできたことが、かなり制限されてるという面がある訳ですね。で、それが制限されたらなんで困るかっていうのが、今日のテーマなのです。

ところで、同様に本質的にコンピューターなんだけどいろいろ自由じゃないデバイスとして、ゲーム機とか携帯電話とかがある訳ですが、携帯電話だったら、例えば電話番号っていうのはオープンですよね。少なくとも最低限のインターフェースは開放されてるから、電話としては使える訳です。あるいは、メールのプロトコルっていうのもオープンな訳です。ゲーム機はそれに比べたらはるかに閉鎖的なんですが、でもまあ、ゲームしなければ良い訳ですよね。だけど、iPad の場合、私は中国の古典籍とかの電子化みたいなのに関わるような仕事をしてたりするのですが、iPad がブレイクすると、なんか自分がスルーしてても仕事が降ってくるような予感があって。そうすると、必然的に iPad のいろんな制約に巻き込まれそうな気がして。それでちょっと憂鬱だったりするのです。

具体的にその iPad がなんで問題かっていうと、iPad だけではなくて iPhone も含めた iOS, かつて iPhone OS といっていたものの SDK（ソフトウェア開発キット）のライセンス[1] の問題なんですが、これ、いくつかポイントがあるんですけど、まず一つ目は開発言語が制限されてるということですね。具体的には、Objective-C, C, C++, JavaScript 以外の言語で開発することができない訳です。で、百歩譲ってこれは良いとしましょう。っていうのは、多くの今日的なシステムは、大抵、C とか C++ とかで書かれてるから、例えばそれで言語処理系を作れば、載せられる訳ですよね。ところが、それも閉じられてて。つまり、イ

ンタープリターとか互換レイヤー、あるいはプログラムを変換してこれらの言語にすることっていうのも禁止しちゃった訳です。これは非常に困る訳ですね[2]。

一つは、直接的な問題として、好きな言語を使う自由がないということがあります。世の中にはいろんな言語、Lisp とか Scheme とか Perl とか Prolog とかいろいろあって、例えばウェブサービス作る時だったら好きな言語を使えば良い訳なんですけど、iPad の上で動かそうと思ったら、さっき言った C か C++ か Objective-C か JavaScript で書かないと駄目なんですよね。ということももちろんありますけど、より本質的な問題っていうのは、大抵のソフトウェアっていうのは、程度の差があれ、仮想計算機、仮想機械、つまり、「インタープリターとか互換レイヤーと一緒じゃないの？」ってことなんですよね。

つまり、コンピューターっていうのはそれそのものは何でもないものな訳ですよね。で、それが何かの有用な道具として使えるということは、つまり何かをエミュレートしているから使える訳です。で、直接人間に対する何か、例えば音楽プレーヤーのまねをするとかっていうのは良いけど、機械に対して何かのプロトコルをサポートするような仮想機械を作るっていうことが本質的にインタープリターな訳ですけど、それをやっちゃいけないっていうことは、実は今のネットワークっていうのが、例えばセマンティックウェブのように、人間に対する情報提供をするだけじゃなくて、機械に対して情報提供してやり取りすることによって、より賢いサービスを全体として作っていこうっていう風になってきてる訳ですが、こういう流れに対して、本質的にちょっと矛盾してるような、矛盾をはらむようなことを含んでいる訳ですね。

で、そういう理念的な問題はさておき、もうちょっとベタな問題として、まず一つ目は Flash 使えないっていうのは、結構今日のウェブサービスの上では困る訳ですよね。実はこれがアップルの本来の狙いだと言われてますけれど。また、例えば iPad をパソコンみたいにして使おうと思った時に、拡張可能なエディター、例えば UNIX 系 OS ではメジャーな Emacs っていうエディターがありますけど、これは中に Emacs Lisp というインタープリターを含んでいて、それでいろんな機能を拡張していく訳ですから、当然駄目っぽい訳ですよね。で、あと、GIMP というグラフィックエディターがあるんですけど、これプラグインで、例えばインタープリターとかでフィルターを拡張するとかいうことができるんですけど、これもまぁちょっと駄目っぽいですよね。それから、特に人文系とかでいろんなテキストデータベースとかをマークアップ言語でマークアップする場合（電子テキストって言った時に、人間が読むための見かけを表示するようなものだけではなくて、データの内容の論理構造を機械に理解させるためのいろんな情報を入れるっていうことが結構古くからなされてて、そういうふうに論理構造を書いておくと、例えばそれをベースに加工することでいろんな形式を作ったりとか、そういうこともできる訳ですし、何より意味の情報を付与することができる訳で、より賢い処理ができる訳です）を考えると、それって本質的にインタープリターと一緒なので、やっぱり駄目っぽい気がする訳ですよね。あるいは、ゲームの場合だと、最近では、いきなりゲームをスクラッチから書くんじゃなくて、しばしばスクリプト記述言語とかのゲーム用のミドルウェアみたいなのがあって、それでいろんな場面の移り変わりとかの情報を記述して書くというようなことをしたりするようなんですが、それも本質的にインタープリターですから駄目な訳ですよね。あるいは、気の利いた、ある程度複雑なデバイスというのは、いろんな設定ができたりしますよね。あるいは、内部状態を保存したくなるかもしれません。電源を切って起動する度にまた前の設定を全部クリアされてたら嫌だから、やっぱりそれは保存したい訳ですけど、それっていうのは本質的に形式言語で、それがある程度複雑になってくるとやっぱりインタープリターとみなされるかもしれない訳ですよね。結局コンピューターにとって、コンピューターが理解可能な情報っていうのはすべて形式言語で書かれ

てて、それを解釈するものっていうのがインタープリターという訳なんで、インタープリターが禁止されてしまうと、いろんなことができなくなるっていうか、ぶっちゃけて言えば、内部状態を持つような、ある程度複雑なソフトウェアシステムっていうのは、ほとんど原則禁止っていうことに……？ちょっとやや曲解かもしれないけど、そういう可能性がある訳ですね。

で、次に互換レイヤーの問題です。こちらもアップルが提供したもの以外のフレームワークの使用は本質的に禁止されています。例えば、Windows の世界でよく使われている .NET Framework というのがありますが、これのマイクロソフト以外の実装、例えば Linux とかその他 UNIX 系の OS で使える Mono という .NET の実装もあるんですけど、そういうのをポーティングしてやって使うこともできない訳ですよね。で、ミドルウェアも駄目。あるいは、自前のライブラリーっていうのは良いのかっていうことが当然疑問としてあって。結局、何ていうのか、ソフトウェアっていうのは普通ある程度複雑になってくると、何階層かのレイヤーを持ってモジュール化して設計していくことになる訳ですけど、その各レイヤーを提供するものっていうのが、フレームワークだったり互換レイヤーっていうものな訳です。これらがある意味ソフトウェアを抽象化して、複雑なプログラムを設計していくための肝なんですけど、そういうことを阻害しかねないということな訳ですね。

こういう、どう考えてもいろいろ矛盾をはらんだことなので、当然かなりいろいろ反対の声がいっぱい上がって、結局、その後、若干条件が緩和されたらしいんですね [3]。これまでは例外なしの禁止だったのが、アップルの承認がない限り禁止っていうことに変わって、だから、アップルが認めれば許されるかもしれないということにはなった訳です。でも、それはやっぱりちょっと問題ですよね。つまり、結局アップル次第な訳ですよね。で、ライセンス見ても「アップルが認めない限り許されない」って書いてあったら、結局自分が作りたいものが許されるかどうかっていうのは判らない訳ですよね。結局、何が良くて何が駄目かっていうことの線引きは見てもよく判らない訳です。第三者には判らなくて。そういう技術的なことだけじゃなくて、例えば、これは検閲ではないという話はあるんですけど、やっぱり事実上 jailbreak ということをしない限り（ jailbreak（脱獄）というのは、こういういろんな制約をクラックして外すっていうこと（Apple の作った監獄から逃げること）で、そうすると何でもできちゃうんですが）、Apple が恣意的に決めた制約から逃れることはできず、また、ソフトウェアを容易にに配布するためには、事実上、AppStore を使わないといけない訳で。そうすると、何ていうのか、やっぱり事実上、アップルが認めないと流通させることができないという問題が iPad にはある訳ですね。

もう一つ、これはちょっと心情的な気持ちの問題なんですけど。iOS とか、あるいは Mac OS X っていうのは、実はそれ自体は proprietary な、つまり商業的でかつ占有的なソフトウェアになってる訳ですけど、実は、全体としてはそうだけど、その基盤になってる部分っていうのはかなり自由ソフトウェアに依存してる訳ですね。例えば、カーネルであるところの Mach は、もともとカーネギーメロン大学で作られたフリーソフトウェアなんですよね。あるいは、開発言語、さっき言った C とか C++, Objective-C っていうのは、具体的には GNU Compiler Collection（GCC）を開発言語としてる訳です。これは、Free Software Foundation（FSF）が配布している自由ソフトウェアの代表例の一つなんですよね。Linux とかのいろんなフリーソフトウェアの開発に使われてるものと同じな訳です。で、カーネルとか開発言語、あるいはそれ以外の、カーネル以外の OS の（BSD サブシステムとかユーザーランドなどと呼ばれる）GUI とかに関わらないような部分は、Free BSD をはじめとする BSD 系の OS から引っ張ってきたもので埋めてたりする訳です [4]。こういう OS の根幹部分のかなりの部分が、…っていうか、そのほとんどが実は自由ソフトウェアに頼っている訳ですね。

で、そういうこともあって、MeCab という形

態素解析エンジン（例えば、iPad とか、あるいは iPhone, あるいは Mac OS X の Spotlight とかで、日本語検索する時には実は下でインデックスを作ったりとか、キーワードを切り出すために形態素解析っていう技術を使っているんですけど、それのエンジンにはこの MeCab を使っています。この MeCab も実はフリーソフトウェアの一種で）の作者の工藤拓さんが、「MeCab が iPhone, OS X に載っていると言うのは止めようと思う」っていう日記 [5] を書かれていて、ちょっとご紹介したいと思います。これは iPhone の SDK の最初の変更が行われた頃 [6] に書かれた日記なんですが、どうも、知らなかったんですけど、MeCab はもともと GPL と LGPL のデュアルライセンスで配布されてたんですが、アップルからの要請を受けて、修正 BSD ライセンスを含むトリプルライセンスにした、…ある意味、アップルのためにより使いやすくしてあげたのに…、こんな風に、実は、アップルはフリーソフトウェアのコミュニティーに対して結構働きかけをして、自分の都合に合わせてそういう風に要求をしてたりしていて、それに対して、割と多くの人がそれに応じてきた訳なんですけど、ある意味、それを裏切られた気持ちみたいなのを持っている人が結構いる訳ですね。で、この日記はその一例という訳です。で、GNU General Public License（GPL）というのは、さっき述べた GCC や MeCab とかで使われているライセンスなのですが、これの考え方のあらましを述べると、一つは多くのフリーソフトウェアにあることですけれども、まず利用する自由がある。それから、再配布する自由がある。そういう利用者サイドの自由というのが一方にあるんですけど、もう一方の側の自由として、ハックする自由、つまり、ソフトウェアのソースコードを入手する権利と、それを改変する権利というのがある訳ですね。GPL の特筆すべき特徴は、GPL のもとで配布されたソフトウェアそのものだけでなくて、その派生物に対してもこの 4 つの自由を強制するという点なのです。

この派生物っていうのは、GPL で配布されてるソフトウェアを改変したものだけでなくて、GPL で書かれたソフトウェアをリンクするソフトウェアに対しても適用されるということで、これがある意味ちょっとショッキングな訳です。しばしば GPL を批判する立場の人から『GPL 汚染』などとも揶揄されたりするんですけれども。ある意味、非常に強力に自由を強制するライセンスであった訳です。これによって、例えばサーバーとかを作ってる企業、例えば IBM とか NEC とかが Linux に対して改良を加えた時に、その成果物を自由ソフトウェアとして公開するということがかなりされてきて、これによって Linux が発展してきたという面がある訳ですね。結果的に、GPL には競争力のある大企業にとって、例えば、標準化しようと思った時にはある程度技術を出さないといけない訳だけど、それにただ乗りされたら嫌な訳ですよね。そういう意味では GPL っていうのは、ある意味、抜け駆けなしみたいな、そういうライセンスになってるので、企業にとってもメリットがあった訳です。で、結果的に標準化とかの点でもいろいろ意味があった訳です。

ところが今日のサービスの時代っていうことになった時に、GPL の弱点っていうことがちょっと出てきたという風に思われます。これはどういうことかっていうと、GPL はサービスにおける自由を強制することができないということなのです。例えば、地デジのテレビだとかハードディスクレコーダーとかの類には、実は、結構中で Linux を使ってるものがある訳ですよね。つまり GPL で書かれたソフトウェアを使って動作してるデジタル家電っていうのがいっぱいある訳です。で、そういうものだと、説明書の後ろの方をよく注意して見てると、実は GNU General Public License の全文が載っていたり、ソースコードの入手先が書いてあったりすることがあるんで、もしお持ちの方は家でじっくり御覧になってください。こういう風に、GPL が要請してる、例えばソースコードを入手する権利だとかを、デジタル家電の類は保証している訳です。だけど、だからといって、例えば、地デジの B-CAS カードの問題だとか、あるいは『ダビング 10』とかで孫コピーが作れない問題、そういう私的録画が制限されてるような

問題っていうのが、自分で直せる訳ではない訳ですよね。つまり、ソースコードが公開されてるからといって、なんか自由になった気にはならないですよね。実質的に何かができる訳じゃないですから。この種の、例えば放送システムの機器だとかの場合。

で、二つ目ですけど、今いろんなウェブサーバー、ウェブに限らず、インターネットのかなりのサーバーは Linux ベースで動いている訳なんですが。だけど、例えばウェブサービスをやったとして、そのソフトウェアを配布してる訳ではない訳ですよね。つまり、どこかからソフトウェアを取ってきてそれを動かしてるだけな訳ですよね、ウェブサービスっていうのは。で、例えば自分でそれに対して手を加えたとしても、ソフトウェアを配布してる訳ではないので、GPL が要請してるソース公開の義務とかが生じない訳ですね。で、結果的に例えばそのウェブサービスの API とかがもし公開されてなくて見たいと思ったとしても、そのソースコードを見てそれを調べるとかいうことができない。あるいは、そういうことが公開されてたとしても、実際のウェブサービスにロックインされてる問題だとか、いろんなウェブサービスにおいて享受したいと思われるような自由というのは、GPL だけではそういう自由を強制することはできない訳ですよね。そういうこともあって、ちょっと今、自由ソフトウェア、フリーソフトウェアっていう観点からすると、ウェブサービスっていうのにはいろいろ悩みがある訳です。

それからちょっと話は変わりますけど、iPad みたいなデバイスっていうのは、コンピューターと思うからある意味腹立つ訳で、一種の端末だと思えば良いじゃないかっていう考え方はある訳ですよね。で、ある意味、今のウェブサービスっていうのは、今までだったらパソコンの上で動かしてたようないろんなアプリケーションをサーバー側で動かすっていう側面があって、そういう意味では集中型のコンピューティングだっていう風に考えることができます。

大昔、コンピューターが非常に高価だった時代には、大きなコンピューターの下に端末がいっぱいぶら下がるような集中型システムが多かった訳ですよね。一方、インターネットっていうのはもともとどちらかというと分散型のシステムを志向していて。今でこそ IPv4 のアドレスが枯渇してきたというようなこともあって、インターネットに接続するコンピューターの全てにインターネットに直接つながる IP アドレスが振られてる訳ではない訳ですけれども、かつてはインターネットにつながるすべてのコンピューターにはグローバルな IP アドレスが振られていて、たとえ研究室の中でしか使わないようなローカルなコンピューターであったとしてもそうだったんですよね。基本的にインターネットにつながるすべてのコンピューターがサーバーでありクライアントでもあるような、そういう Peer to Peer（P2P）なシステムだった訳ですね。で、実際ウェブ時代以前のプロトコルは、大体バケツリレー型で、サーバーからサーバーにデータをバケツリレーしながら津々浦々に配送していくようなタイプのものが多かった。そういう Peer to Peer なプロトコルが多かった訳ですね。

ただ、こういう本来のインターネットのやり方っていうのは、結構管理上の手間がかかったりして、パソコンが普及してみんながインターネットにつながるようになると、すべての利用者が自分のパソコンの管理者として十分に振る舞える訳ではないっていうことがあって、それでセキュリティー問題とかが生じてしまったという面がある訳ですね。そういう面倒さを減らすという側面もあって、ウェブサービスっていうのが流行（はや）ってるっていう面もあると思うんですが、これは、ある意味最初に言ったように、集中型への回帰という側面があるかと思います。ただ、大昔と違って、今のウェブサービスでは、サーバーが 1 台っていうことはもう普通なくて、中でクラスタ化されてたりして、実は内部は分散的なシステムになってることが多い訳ですが。ただ、利用者から見た時に全体として一つのコンピューターシステムのように見えるっていう点で見ると、ある意味集中型で。実際例えば Twitter とか使ってると、Twitter のシステムっていうのは一つだから、例えばワー

ルドカップとかで中継が盛り上がったりするとTwitterのサーバーが落ちちゃったりする訳ですよね。これと同様な問題がmixiだとか他のいろんなサービスにある訳です。

ただ、こういうウェブアプリケーションの時代が永遠に続くかっていうと、私はたぶんそんなことはないだろうと思っています。というのは、計算機アーキテクチャっていうのは、これまでの歴史を見ると、10年とか20年ぐらいでゴロッと変わってる訳ですよね。ついこの間までダウンサイジングとか……、ついこないだでもないですけど、10年とか20年ぐらい前はダウンサイジングだとか、もっとネットワーク型……、大きなコンピューターを、小さなコンピューターをネットワークでつないだものに置き換えることによってコストを下げるんだとか言ってた時代もあった訳ですよね。それが気が付くと、ウェブサービスを使って大きなコンピューターにサービスを集約するとか、仮想計算機技術を使って集約するとか、そういうある意味集中型に回帰してる訳で。こういうことは、次またハードウェアの要請だとか新しいサービス上の要請とかによって変わる可能性は十分あり得て、現実問題、10年後とかには今のようにやってるとは思えない訳です。

さらに言えば、企業のサービスの寿命はおそらくもっと短いでしょう。例えば今から10年ぐらい前とか5年ぐらい前に始まったいろんなDRM付きの音楽配信サービスのほとんどは、もう今生き残ってない訳ですね。で、企業がサービスを停止したら、そのサービスにロックインしたデータとかDRM付きのデータって読めなくなる訳ですよね。で、これとはちょっと別の話なんですけど、アカデミック方面ではしばしばいろんなプロジェクトでデータベースを作ったりする訳ですけど、こういうプロジェクト型の成果物っていうのはしばしばそのプロジェクトが終わってしまうとそのデータベースをメンテナンスするためのお金が出なくなって、よくあるパターンだと、そのサーバー、とりあえず今は電源入って動いてるけど、そのうち担当者がいなくなって、マシン壊れたら終わりみたいな、そういうパターンがよくある訳ですね。

こういうことを考えるとDRM付きの電子書籍って心配ですよね。今後、電子書籍がある程度普及してきた時のことを考えると、やっぱりそれをずっと使いたいような気がします（ある日突然使えなくなるとか困ります）。紙の本だと劣化はしてきますけど基本的にその本はずっと読める訳ですよね。紙の本っていうのが、ある意味、スタンドアローンであるおかげなのですが。一方で電子書籍っていうのは、今のところ必然的にプラットフォームに依存しちゃうという面がある。あるいはフォーマットに依存しちゃうという面がある訳ですよね。で、そうするとそのプラットフォームとかフォーマットが終わっちゃうと、もう読めなくなる訳ですね。そういうことを考えると、電子書籍を永続化させるためには、プラットフォームを渡るための『渡る自由』っていうのがやっぱり必要だっていうことが言えるかと思います。つまり、新しいプラットフォームができたらそれに簡単に変換できるとか、乗り移れるっていうことが結構重要なんですよね。

こういうことを考えると、もしかすると特定のプラットフォームに過剰適応しない方が良いのかもしれないということが言えるかもしれません。で、iOS SDKのライセンスの本質的な問題というのはこのプラットフォームを渡る自由を阻害してるということなのです。iOS SDKのライセンスがなんであんな風になってるかっていうと、要するにiOSが動く機械、すなわちiPhoneとかiPadにユーザーを縛りつけたいからですよね。そのためにああいうライセンスを使ってる訳で。本質的にプラットフォームを渡る自由を制限しようというライセンスになっている訳ですよね。だけど、これはおそらく電子書籍の永続化という点ではやっぱりマイナスだという風にいえると思います。

より永続的なデータ……、例えば、古くから作られてるデータベースっていうのは、もうすでに20年とか30年の寿命を持ってるものがある訳ですけれど、そういうのが今日まで使えてるっていうのは、プラットフォームを渡ってきたからです。そのための、一例としては、例えば大昔

PC-98 時代に作ったようなものであっても、プレーンテキストで書かれてたら今も読めたりする訳ですよね。あるいは、構造化テキストみたいなのでいえば、Text Encoding Initiative (TEI) というところが人文科学向けのマークアップのガイドラインを作ってて、それに基づいて作られたマークアップテキストであれば、細かな注釈やバリエーションに関する情報も含んでいるような、ある程度複雑なデータであったとしても、割と寿命が長い訳ですね。あるいは、LaTeX なんかもそうですね。こういう中立的な形式を使っておくと、例えば必要に応じてその時々で流行ってるプラットフォームに変換してみるということができる訳です。そうした方が、ある意味永続化するためには良いかもしれないということは考えられます。

ちょっと雑駁な話になりましたけど、以上で発表を終わらせていただきます。

## 注

[1] iOS Developer Program License Agreement

[2] 2010 年 9 月 9 日にライセンスが改定され、この制限は大幅に緩和され、スクリプト、コード、インタープリターがアプリケーション内にパッケージされている場合、これらの利用が認められるようになった（なお、ダウンロードしたコードの実行が認められるのは WebKit（Safari のエンジン）で実行可能な JavaScript に限定される）。よって、この日の講演内容における iOS のライセンスに関わる問題はほぼ解消したと考えられる（但し、Flash アプリケーションをダウンロードして直接実行することは依然として認められないと思われる。この Flash の制限の『副作用』として生じてしまった諸々の問題を解消しようとしたのが今回の改定なのかも知れない）。

[3] 2010 年 6 月 7 日の改定。2010 年 9 月にさらに条件が緩和された [2]。

[4] シェルの bash をはじめ GNU（FSF）由来のものも使われている。

[5] http://chasen.org/~taku/blog/archives/2010/04/mecabiphoneosx.html

[6] 2010 年 4 月。

# 電子書籍（e-book）とプラットフォーム

## ── 競争法上懸念される二つの問題

石岡　克俊（いしおか　かつとし）

### ◇ 一、プラットフォームによるスマートフォン・アプリの配信

近年、RIM 社製の Blackberry やアップル製の iPhone の大ヒットで、スマートフォンは携帯電話端末の新しいスタイルを確立した。スマートフォンは、それぞれ対応するアプリケーション・ソフトウェア（以下、アプリ）の搭載によって、簡単に新たな機能を付け加えることに大きな特徴がある。これに搭載できるアプリは、「App Store」（iPhone ／ iPad 等アップル製 OS 搭載端末向け）や「Android Market」（グーグル製 OS Android 搭載端末向け）といったオンラインストアを通じてもっぱら配信が行われている。

2010 年 4 月時点において、「App Store」で提供されている iPhone ／ iPad 向けアプリケーションの数は 18 万 5000 本、一方、「Android Market」では 3 万本弱のアプリが提供されている。現在、この 2 社の存在感は圧倒的であり、Windows モバイルや BlackBerry といった他のスマートフォン向けアプリの本数を大きく引き離している。このように、最近では「App Store」や「Android Market」など、OS やスマートフォン端末の開発者が展開しているオンラインストアが、各々の携帯端末に対応するアプリの配信及び認証・課金・決済を一手に握り、コンテンツ流通のチャネル・リーダーとしての地歩を固めるに至っている。

ここでは、「App Store」や「Android Market」といったオンラインストアが有するアプリの配信及び認証・課金・決済機能を「プラットフォーム」といい、かかる機能を有する事業者を「プラットフォーム事業者」と呼ぶことにするが、いま、このプラットフォーム事業者が、川上に位置するコンテンツ供給事業者（コンテンツ・プロバイダー）や川下の一般消費者に対して占める突出した地位、あるいは、その地位がもたらす取引上の力に注目が集まっている。殊に、プラットフォーム事業者によるアプリの選別や、出版社とプラットフォーム事業者との間で見られる電子書籍（e-book）の価格設定をめぐる一連のせめぎあいが、しばしば取りざたされている。

以下では、この二つの問題に対する現状と考え方について述べていく。

## ◈ 二、プラットフォーム事業者によるコンテンツの選別あるいはコンテンツ供給事業者に対する取引拒否の正当性とその限界 ── 「App Store」のアプリ掲載拒否問題

### ⊙ アップルの審査

プラットフォーム事業者によるアプリの選別の問題がしばしば指摘されるのはアップルが運営する「App Store」においてである。「App Store」にアプリが掲載されるためには、「Android Market」の場合と異なり、アップルの審査を経なければならない。この審査は iPhone や iPad をウィルスなどの不正アプリから守る上で重要な役割を果たしているとの指摘もあるが、審査はこうした不正アプリ排除のためだけに行われているわけではないようだ。かつて、iPhone の画面が割れたようなアニメーションを表示する「イタズラ」アプリが開発されたが、「製品が壊れたとの誤解を招くおそれがある」との理由で「App Store」での販売を拒否されたことがある。また、「Skype」など IP 電話系のアプリも当初はパートナーである携帯電話キャリアに配慮して承認してこなかった経緯も知られている。そして、何よりプラットフォーム事業者によるアプリの選別が、世間一般の注目を集めたのは、2010 年 2 月に行われた 5000 本に及ぶアダルト系アプリの一斉削除だろう。

「App Store」によるアプリの掲載・不掲載の選別が、非公開ながらも一定の審査基準に基づくものとはいえ、アップルの裁量に委ねられていることに対するコンテンツ供給事業者の懸念はきわめて強いものがある。アップルの純正 web 閲覧アプリ「Safari」と競合するため、「App Store」の掲載を危惧したオペラが自社 web 閲覧アプリ「Opera mini」の「App Store」への申請を自社ホームページにおいて広く周知し、審査から公開までの期間を日々数え上げることでアップルの判断を衆人監視の下に置こうとする動きも見られた。

### ⊙ 「App Store」の影響力

ただ、素朴に考えると、アップルが運営する「App Store」にどんなアプリを掲載するか、つまり、どのコンテンツ供給事業者と取引するかは、商業ビルの管理者がいかなる企業をテナントとして迎え入れるかにつき自由な選択が許容されているのと同様、「App Store」の掲載・不掲載の選別は本来アップルに委ねられるべきもののはずである。

しかし、一方でスマートフォン・アプリの配信におけるプラットフォーム事業者の果たす決定的な役割は否定できない。しかも、非常に大きな成功を収めた iPhone や iPad を擁する「App Store」の影響力、殊にコンテンツ供給事業者に対するそれは無視することはできない。現に「App Store」の取扱いアプリ数は他のプラットフォームを圧倒している。

### ⊙ プラットフォームに対する関心と対応

競争政策の視点から、このプラットフォーム問題に初めてアプローチしたのは、公正取引委員会が 2003 年 3 月 31 日に公表した「デジタルコンテンツと競争政策に関する研究会報告書」だと思われる。この報告書では、いわゆるプラットフォーム事業者の中でもコンテンツ配信業者に注目し、まずその意義について「コンテンツをインターネット等のネットワークを利用して視聴者に提供するために必要なコンテンツ配信、データセンターサービス、認証、課金・決済……等のサービスを行っている事業者」とし、これらの事業は「インターネット接続サービス事業者、インターネット検索サービス事業者等が行っていることが多い」と、当時の状況を踏まえ説明されている（同報告書・7 頁脚注 10）。

また、この報告書が作成・公表された当時は、コンテンツ配信事業がいまだ未成熟で、有力なプラットフォームが存在しない中、活発な競争が展開されていることも示唆されていた（同報告書・25–26 頁参照）。

その後、プラットフォーム問題は、こうした当時の状況認識を踏まえ、公正取引委員会と総務省の共同ガイドライン（「電気通信事業分野における競争の促進に関する指針」2001 年 11 月 30 日）に結実する。

### ⊙ プラットフォーム事業者によるコンテンツ掲載基準をめぐるガイドラインの考え方

このガイドラインがコンテンツ配信について並々ならぬ関心を抱いているのには、もう一つ背景がある。NTT ドコモの i モード・サービスにおける「公式サイト」と「勝手（非公式）サイト」の選別をめぐる問題がそれである。当時、NTT ドコモは i モードのヒットでプラットフォーム事業者として大きな存在感を示すようになっていた。ガイドラインの「簡易端末情報サービスシステムを管理・運用している電気通信事業者（以下、システム運用事業者という。）は、その管理・運用の適切性等を確保する観点から、簡易端末の簡単なキー操作によってアクセスすることができるメニューを設定し、一定の基準の下に、コンテンツを掲載していることが一般的である」（同ガイドライン第 4-1-(1)）との指摘は、明らかに i モード・サービスを念頭に置いている。

このガイドラインにおいて、プラットフォーム機能を担うのは「システム運用事業者」だが、これが「自己のメニューリストへのコンテンツの掲載基準を独自に定めている」ことについて触れ、この掲載基準が公開されていなかったり、公開されていても明確さに欠ける場合には、「システム運用事業者による掲載基準の恣意的運用が行われ、他のシステム運用事業者との取引を制限することとなりやすい」旨の指摘が行われている（同ガイドライン第 4-1-(2)（注 51））。

### ⊙ ビルは店子を選べるか？

以上を踏まえると、プラットフォーム事業者によるコンテンツの審査ないし選別基準の設定それ自体は当然に認められている。

ガイドラインにも、「不適切なコンテンツが流れることを防止する観点から、メニューリストに掲載すべきコンテンツプロバイダーの選択基準を明確にした上で、公正に実施する場合は問題とならない」との記載が見えるが（同ガイドライン第 4-1-(2)（注 53））、これは、プラットフォームにより設定される審査・選別基準が、まずそれ自体を公開し且つ明確なものにし、その上で実施の公正性を担保すること、つまり実施にあって差別性や恣意性を排除したかたちで行われる限りで、当該行為は合理的なものと考えられ、問題はないとの理解である。

### ⊙ プラットフォーム事業者による審査・選別の限界

では、プラットフォーム事業者によるコンテンツの審査・選別の限界はどこにあるのだろうか。その答えは、独占禁止法（競争に悪影響を及ぼすかどうか）の理屈の中にある。ガイドラインによれば、まず、プラットフォーム事業者が「市場において相対的に高いシェア」を有し、その事業者がコンテンツ配信業者と他のプラットフォーム事業者の取引を制限するために審査・選別基準を設定・適用・実施することで、「競争事業者の取引の機会を減少させ、又はコンテンツ提供市場における価格競争を阻害するおそれを生じさせる」場合、つまり、コンテンツ配信業者を囲い込むとき等は、プラットフォーム事業者の行為は問題になる（第 4-1-(2)）。既に指摘した「App Store」の市場における地位を前提とすれば、この審査・選別行為が「コンテンツ配信業者と他のプラットフォーム事業者の取引の制限」に向けられている場合、ある人気のあるアプリを「Android Market」等の他のプラットフォーム事業者が掲載しようとすることを制限するようなとき、独占禁止法上問題となる可能性がある。また、ガイドラインは専らコンテンツ配信業者と他のシステム運用事業者の取引への影響に注目しているが、プラットフォーム事業者がその取引上の地位を利用し、コンテンツ配信業者に不当な不利益を課すようなかたちで審査・選別基準を設定し、適用し、実施する場合も、ガイドラインでは触れられてい

ないが、独占禁止法上問題（優越的地位の濫用）となる可能性があり注意が必要である。

## ◈ 三、誰が本の値段を決めてきたのか？そして、これから誰が決めていくのか？── 電子書籍（e–book）の価格設定権の所在

### ⊙ 電子書籍（e–book）元年！？

今年になって、そして、にわかにではあるが、いま、わが国では電子書籍（e–book）に熱い期待が注がれている。長く出版業界が不況にあえぐ中、新しい販路として電子書籍（e–book）が注目されるのは、当然といえば当然の話である。だた、本のデジタル化とそのビジネス化の試みは、これまで幾度となく繰り返されてきたものの、そのいずれもが成功とはかけ離れた結果であっただけに、今回も同じ轍を踏むのではないかとの冷めた見方があることも事実である。

もちろん、本稿は電子書籍（e–book）がわが国に根付くか否かを予想し論評するものではないし、筆者にそうした能力が備わっているわけでもないので、これ以上、あえて踏み込むことはしない。しかし、ただ一点だけ、今回のブームが、これまでの「期待」や「注目」と少しだけ違うとすれば、それは、米国における電子書籍（e–book）市場の活況ぶりであり（米国における書籍売上高のうち電子書籍（e–book）の売上高は 5 パーセント以下ではあるが、2008 年から 09 年の伸びは 3 倍にも及んでいるという。）、また、それを支えるすぐれた電子書籍（e–book）端末の登場だということは指摘しておくべきであろう。

今年、わが国においても、アップルからは iPad、アマゾンからは Kindle、ソニーからは Reader と評判の高い電子書籍（e–book）端末が相次いで発売される。もちろん、すぐれた端末の普及だけでは、わが国に電子書籍（e–book）が根付かない。当然、魅力的なコンテンツはもちろん、ネットワークを介して配信されるこうした商品にあっては、その配信だけではなく、認証・課金・決済を担うプラットフォーム事業者の存在が欠かせず、これがなければビジネスとして成り立たない。その意味でも、電子書籍（e–book）の流通過程における取引の検討は、早晩わが国においても不可避となる。

そこで、ここでは電子書籍（e–book）をめぐる取引上の課題―とりわけ、電子書籍（e–book）の価格設定のあり方について検討を加える。この問題は、旧来の書籍においては、著作物再販制度により半ば当然に出版社に認められてきたものであるが、同制度の範疇外にある電子書籍（e–book）の場合には、かかる保護（？）は妥当せず、取引上の力関係や競争関係によって左右されることとなる。ここでは、出版社を典型とするコンテンツ・プロバイダーとプラットフォーム事業者間の取引に着目し、わが国に先行する米国の経緯に触れながら、その競争法上の論点を指摘することとしたい。

### ⊙ アマゾンの戦略・アップルの選択

米国のアマゾンは、その電子書籍（e–book）端末の販売とともにオープンした同社の電子書籍（e–book）ストア（Kindle Store）の展開にあっては、これに先行し成功を収めたアップルの音楽配信サイト・iTune Music Store（iTMS）の販売戦略を徹底的に研究し、そのビジネスモデルに倣ったといわれている。

読者は、アマゾンの専用端末である Kindle から、Wifi もしくは 3G 回線を通じて簡単に電子書籍（e–book）を手に入れることができるわけだが、その電子書籍（e–book）（新刊一般書）の値段を、アマゾンは、当初、一律 9 ドル 99 セントで売り出していた（因みに、iTMS の楽曲の販売価格は一曲一律 99 セントであった）。これは、電子書籍（e–book）が安いというイメージを顧客に植え付け、この市場を育てるための戦略として採用された。

実際、米国での印刷版書籍の価格帯はおおむね 25 ドルから 30 ドルで、しかもアマゾンは版元からの卸価格も印刷版と同じ条件で仕入れていたといわれているので、仮に正味が 5 割程度だっ

たとしても、この 9 ドル 99 セントという販売価格は逆ザヤとなっている。つまり、1 冊につき 2・5 ドルから 5 ドル程度の損を出しながら売り続けられていたということになる。

一方、電子書籍（e-book）については後発のアップルは、電子書籍（e-book）の売り上げから一定率のマージンを徴収する方式を採用することによって、アマゾンとは異なる有利な条件を提示し、先行するアマゾンに揺さぶりをかけた。アップルは新端末 iPad の発売を控え、電子書籍（e-book）で先行するアマゾンに追いつくためには、より多くの出版社を味方につけ、コンテンツ面での充実を図る必要があったのである。

かねてより、米国大手出版社 6 社（ただし、ランダムハウスを除く）は、アマゾンの販売政策に不満を持っており、米国出版大手のマクミランは、2010 年 1 月、このアマゾン社の販売価格に異を唱え、自分たちの裁量の範囲、すなわち 13 ドルから 15 ドル程度での販売を認めるよう申し入れていた。マクミランとアマゾンの交渉は難航し、1 月末には、一時マクミランの電子書籍（e-book）がアマゾンから姿を消すという事態にまで発展した。

アップルは、こうした一連の経緯を受け、これら大手出版社の要求を飲んで、いわゆる「代理店システム（エージェンシー・モデル）」を導入した。これは、出版社が電子書籍（e-book）の小売価格を決め、代理店（エージェント）であるアップルはマージンとして売価の 30 パーセント受け取るものである。

最終的には、アマゾンもアップルに追随してマクミランの要求を受け入れ、値上げを決定した。その後、マクミランに同調する出版社も現れ、いわゆる「9 ドル 99 セント・ルール」は崩れ始めた（ランダムハウスは従来どおりである）。結果、アマゾンもこれら出版社との間で「代理店システム（エージェンシー・モデル）」を採用するに至っている

また、マクミランやアシェットは、他の大手出版 3 社とともに、アップルと iPad 向けブックストア（iBook Store）への電子書籍（e-book）提供について合意をしたらしい。その条件は公開されていないが、9 ドル 99 セントの固定ではなく、段階的な価格設定を出版社側の裁量で認めるものとなるらしい（「App Store」における価格選択方式が採用されると思われる）。

なお、グーグルは、「グーグル・エディションズ」という書籍販売・取次ぎサービスの準備を進めている。販売価格について、グーグルは公式に明らかにしてはいないが、報道によると、アップルやアマゾンと同様、価格決定権を出版社に帰属させることを検討し、「代理店システム（エージェンシー・モデル）」を採用するようである。また、自社販売では出版社の取り分を 63 パーセント、グーグルを 37 パーセントに設定し、取次ぎサービスにおいては出版社の取り分を 45 パーセントとして、自信はほとんどマージンをとらず、残り 55 パーセントの大半をオンライン書店側の取り分にするとのことである。

### ◉ 出版社の思惑

ただ、前節のマクミランを始めとする大手出版社の行動は、一点だけ不自然なところがある。すなわち、アマゾンは確かに電子書籍（e-book）の販売価格を 9 ドル 99 セントとしているが、仕入は印刷版と同様であり、仮に卸正味が 5 割程度であってもアマゾンの販売価格を割り込むことはない。マクミランとしては、仮に 9 ドル 99 セントで販売されたとしても、印刷版並みの利益はすでに確保できているわけである。

では、マクミランはアマゾンによる 9 ドル 99 セントの何が不服だったのか？何故、アマゾンの価格政策に異を唱えたのだろうか？

不満の理由は、実は彼らが獲得する利益などではなく、出版社に「価格決定権がないこと」だったのだ。米国ニューズ・コーポレーションのルパート・マードックは、「アマゾンの 9・99 ドルという価格モデルは好ましくない。この価格は書籍の価値を下げ、ハードカバー書籍の小売業者すべてを苦しめるものだ」と意見を表明している。印刷版の書籍は 25 ドルから 30 ドルで売られている一方、キンドル向けには、同じものが 9 ドル 99

セントで売られている。確かに、出版社側はこれで利益が得られる。しかし、消費者には「電子書籍（e-book）は 9 ドル 99 セントの価値である」との認識が広がることになる。つまり、このまま電子書籍（e-book）の存在感が増していくとすると、25 ドル以上であったはずの「本の売価」が成立しなくなる、というわけである。

### ⊙ ガイドラインの前提と視点

コンテンツ・プロバイダーが提供するコンテンツの料金設定については、必ずしも電子書籍（e-book）を想定したものではないが、かつて競争法の視点から検討されたことがある（デジタルコンテンツと競争政策に関する研究会『デジタルコンテンツと競争政策に関する研究会報告書―デジタルコンテンツ市場における公正かつ自由な競争環境の整備のために―』（平成 15 年 3 月））。これによれば、「コンテンツプロバイダーがプラットフォーム事業者を通じて配信されるコンテンツの視聴料金をコンテンツプロバイダーが定めることについて」、まず、「コンテンツプロバイダーが、配信事業者であるプラットフォーム事業者のポータルサイトという場所や、認証、課金・決済という機能を利用して、コンテンツを自ら視聴者に提供している場合であって、実質的にみて配信事業者がコンテンツプロバイダーのコンテンツの配信代行を行っているに過ぎないと認められる場合には、コンテンツプロバイダーがコンテンツの視聴料金を含めた取引条件を設定し、配信事業者と取引を行っていたとしても、独占禁止法上問題とはならない」とする（同報告書第 3-2-(2)-エ-①）。これは、先に触れた「代理店システム（エージェンシー・モデル）」に対応するものと考えられる。

一方、「配信事業者が、料金を支払い、コンテンツプロバイダーからコンテンツ配信の許諾を受け、自らのリスクでコンテンツ配信を行う場合に、コンテンツプロバイダーが、配信事業者が視聴者に請求する視聴料金を設定し、配信事業者の視聴料金の自由な決定を拘束する場合には、独占禁止法上問題となると考えられる」と述べている（同報告書第 3-2-(2)-エ-②）。これは、いわゆる「卸売システム（ホールセール・モデル）」といわれるものである。

この報告書の特徴は、あくまでコンテンツ・プロバイダが配信を担うプラットフォーム事業者よりも取引上優位にあることが想定されており、現在、さまざまな取引分野において取引上有力な地位を占めつつあるプラットフォーム事業者に関する問題点に何ら触れていないことである。

確かに、アップルの「iTMS」を学んだアマゾンは当初後者の「卸売システム（ホールセール・モデル）」を指向したが、出版社の反発を受け、やがて前者に移行した。この動きは、単に多くの米国大手出版社の利益が一致し、その力がアマゾンに向けられることでそれとの交渉において対抗可能であったこともあるが、一方で、電子書籍（e-book）のプラットフォームとして、アマゾン以外にアップルの iBook Store などが存在し、それらとの競争が存在したことも大きかった。プラットフォーム事業の草創期においては、有力なコンテンツの獲得をめぐってプラットフォーム事業者間で競争が存在するため、同事業者の力がコンテンツ・プロバイダに直接行使されることは少ないかもしれないが、この市場における競争が進展していく過程で、プラットフォーム事業者が力を持ち、競争法上の問題を惹起する可能性もある。

この後、公正取引委員会・総務省によって「電気通信事業分野における競争の促進に関する指針」が公表され、この問題は、インフラを有しつつプラットフォームとしての機能も担う事業者が、「コンテンツをメニューリストに掲載させる条件として、コンテンツ・プロバイダーと顧客との間におけるコンテンツ提供に係る料金の設定に関与する」場合と、いささか対象は限定されるものの、競争法ないし事業法上問題となりうることを指摘している（同指針 II- 第 4-③）。

### ⊙ まとめ：著作物再版の本質 ── 「本の値段は誰が決めるのか？」という問いかけによってもたらされるもの

ここで取り上げたように、電子書籍（e-book）の価格設定は、コンテンツ・プロバイダ（出版社）

とプラットフォーム（電子書籍（e-book）ストア）の競争状況に依存する。著作物再販制度の適用がない電子書籍（e-book）においては、出版社に価格設定権が当然に帰属するものではなく、取引場裡において力が裸のまま行使されることとなる。

確かに現状においても、出版業界は取次寡占の状況にあり、出版社は取引上劣位におかれることもしばしばあるとの話だが、件の再販制度がその保護膜として機能しているとの指摘がこれまでなかったわけではない（正田彬『全訂・独占禁止法[Ⅱ]』（日本評論社、1981 年）286 頁）。この立場からすると、再販制度は、発行事業者（出版社等）の価格設定の自由を直接に保障しているものということになる。

ここでの指摘が、電子書籍（e-book）の価格設定権の確保のために何らかの措置が必要だということにはならない。現在のところ、いま現下に起こっている問題に対し報告書の検討やガイドラインは、いまだ不十分ではあるものの、いずれも競争法の適用によって問題を解決することが可能だからである。むしろ、著作物再販の目的を文化の普及と位置づける主張に対し、実は再販の問題は取引上の議論にすぎず、価格設定権という出版社の自由をめぐる議論にすぎないことを白日の下に晒すことにもなるのである。

2010 年公開講座報告
電子出版の動向と諸問題

# 全体討論のまとめ

師　茂樹（もろ　しげき）

## ◇ はじめに

ここでは、「2010 年度公開シンポジウム 電子出版の動向と諸問題」で行われた全体討論の内容を、話題ごとにまとめた形で報告する。この全体討論についても、前の報告者の場合と同様に、各発言を文字に起こした形で掲載する方法もあったと思うが、ここではそれをとらなかった。その理由は、発言者が多数であったことと、オフレコ発言が多かったためである。当日の討論の雰囲気については、Twitter 上の実況 [1] からも感じとっていただけると思うので、ご参照いただければ幸いである。

なお、本シンポジウムは、石岡克俊氏をお招きして毎年夏に行われてきた一連の公開講座（以下「著作権講座シリーズ」と呼ぶ）を受けてのものである。

- 東洋学情報化と著作権問題（2003 年）
- 東洋学情報化と著作権問題Ⅱ（2004 年）
- 東洋学研究と著作権問題（2005 年）
- 国際化時代のデータベースとコンプライアンス（2006 年）
- 〝版面権〟とは何か（2007 年）
- 著作権侵害訴訟と裁判制度（2008 年）
- 著作権をめぐる新動向 ── Google ブック検索と著作権法改正案（2009 年）

この中でも特に 2004 年（電子図書館における著作権処理）、2005 年（研究活動と法）、2007 年（版面権）、2009 年（書籍の電子化）の内容は、本シンポジウムと関係が深く、継続して著作権講座シリーズに参加している筆者を含めた参加者は、少なからずこれらの議論を参照しつつ、発言していると思われる。一方で、会場の参加者の中には、必ずしもこれらの議論をふまえないで発言している方も少なからずおられる。もちろんそれ自体はまったく問題ないのであるが、以下のまとめにおいては、そのような方々の発言についても従来の議論（特に研究活動や学術情報の流通についての議論）の文脈でまとめてしまっている。この点についてもまた、ご了解いただければ幸いである。

## ◇ 電子書籍時代の著作権

では、全体討論について見てみよう。

電子書籍の作成には大きく分けて二つのルートが考えられる。一つは既存の出版物の電子化で、もう一つは新作である。作者が著したテキストが書籍として読者の手に届くまでには、出版社を中心として作者以外の様々な人々が関わっていることは言うまでもないが、それらの人々の作業をどのように捉えるのかについては、特に既存の出版物を電子書籍化する際に大きな問題となってくる。電子書籍の出版はしばしば紙で出版したときの出版社と異なることがあるが、田代氏によれば、元の出版社が版面をそのまま電子化して欲しいと要求する場合もあれば、表紙は使わないで欲しいという出版社もあるなど、出版社によって対応が全然違うというのが現状であるという。

このような違いが生じる背景の一つとして、著者の権利が著作権法で保護されている一方、それ以外の出版に関わる仕事の多くが、編集者やデザ

イナーなどのクリエイティビティが高い仕事も含めて、著作権法では保護されていないことがあげられる（それは「版面権」というものが法律で認められていないことに象徴される）。加えて田代氏は、出版活動における著者に近い活動として、日本語になっていない大学教授の原稿を編集者が匿名ですべて日本語に直して出版した例や、経営者などのインタビューをいかにも本人が書いたようにゴーストライターが直して出版している例を紹介していたが、石岡氏によれば、共同著作者として名前を明示していない編集者やゴーストライターの著作権は認められない可能性が高いという。

もっとも、これまでの著作権講座シリーズでも再三再四強調されているように、出版を含めた知的な情報の公開や流通に関する法的問題については、著作権法以外の法律（民法、独占禁止法など）に関わる面が大きい。既存の出版物の電子化についても、たとえ出版に関わる多くの作業が著作権法で保護されていないとしても、出版社と著者との間で個々に結ばれた契約があれば、それに基づいた対処をすることが可能である。もっとも、田代氏によれば、出版業界で契約を交わすようになったのは 21 世紀に入ってからのことだと言うから、今後の電子化に際しても出版社ごとの場当たり的な対応が続くことが予想される [2]。

例外として田代氏が紹介していたのは、PHP 研究所である。ここは、同じく松下幸之助が創設したパナソニックが電子書籍リーダー「シグマブック」を出した際に、著者に連絡をしなくても PHP 研究所が OK を出せば電子化をすることができる、という契約形態に変えたとのことである。今後、このような契約が一般化していくのかもしれないが､著作の電子化を抑制するような方向（他の会社からの電子出版は認めない、と契約で縛っておきながら、自社での電子出版に消極的な場合など）にならないことを強く望みたい。

## ◇ 電子書籍のコスト(1)

さて、フロアの一般参加者からは、主に電子書籍のコストの問題について質問が出され、活発に議論が行われた。電子書籍をめぐる議論では、しばしば出版社や取次が中間搾取業者として（不当に）槍玉にあげられ、セルフパブリッシング（に近い形態）による著者の収入（印税）の増加が有り得べき未来として語られている。一方で田代氏がその著書や本シンポジウムにおいて述べられたのは、電子書籍になっても全体のコストが大きく下がることはない、という実際の数字を示した上での議論であった。

これに関連して、電子書籍の場合、1 冊（という数え方が妥当かどうかはともかく）あたりの単価を下げることによってより売れるようになるのではないか、という質問がフロアから出された。例えば従来の紙の本であれば、1 冊 1,000 円として 5,000 部しか売れなかったものも、電子書籍として 1 冊 100 円で売れば 1 万部、5 万部と売れるのではないか、というものである。それに対する田代氏の答えは、安いからといって売れるわけではないだろう、というものであった。その根拠として、⑴日本人は平均するとそれほど本を読んでいるわけではない、⑵本の売れる売れないは、価格ではなく内容の良し悪しで決まることが多い、⑶端末が普及していない（一例としてベストセラーの『もし高校野球の女子マネージャーがドラッカーの『マネジメント』を読んだら』が、紙の書籍だと 100 万部近く出ているのに対し、半額のダウンロード版が 4 万部程度であることが指摘された）、などの点があげられた。

また、特に⑵に関して、田代氏は編集者の重要性を強調している。田代氏によれば、編集者とは、世の中から著者の原石を探し出し、それに磨きをかけることで売れる本を作るのが仕事だという。確かに、Youtube やニコニコ動画のような場ができたことで多くのアマチュア映像作家が（それなりに）良質な作品を発表することができるようになったように、電子書籍もまたセルフパブリッシング環境が普及することで良質な作品が生まれるであろうことは否定出来ない。一方で松田聖子のような国民的スターは、レコード会社がその才能を見抜き、ボイストレーニングなどで磨きをかけたからこそ生まれたのであって、Youtube のよう

な場だけでは登場しなかったと思われる。出版も同様に、それなりのコストをかけるからこそ良質なものができるのではないか、と田代氏は言うのである。

## ◇ 電子書籍のコスト(2)

ここで注意しなければならないのは、質の高いコンテンツを生産するための編集（エディターシップ）の重要性については多くの賛同が得られるとしても、それが既存の編集者という職業を守ったり、出版業界のシステムでなければ良質の本を出せない、という議論にはなったりはしない、ということである。その一例（フロアから出た問題提起）として、フリーランスで活動している編集・校正・デザイン・装丁などのプロフェッショナルたちが、本を作るためにその場その場でチームを結成している例が紹介された。このようなやり方での出版活動では、もちろん大手出版社のようなギャラを得ることはできないが、やりようによっては小規模のメリットを活かしてかなりいい本を作ることができるのではないか、という。これまで、紙ベースの出版においてこのような形態で出版しようと思うと印刷費用を積み立てる必要があったが、電子書籍の場合にはそのような必要がないので、フリーランスの人々にはビジネスチャンスではないか、という。

また、学術出版の世界では、大学・研究機関などが出版助成のような形でコストを負担している場合も少なくない。学術的価値は高いとされるものの、世界でまともに読む（読める）人が 50 人もおらず、市場に 100 部も出回らない学術刊行物は山ほどある（そのような専門性の高い学術書の場合、出版社から出版される場合であっても、編集者が上に述べられているような編集者らしい仕事をしていないケースも少なからず見られる）。また、研究者がボランティアで編集作業や組版を行って出版される同人誌のような学術雑誌なども少なくない（本誌もその中の一つである）。フロアからの指摘もあったように、一部のジャーナル、学会誌等では、論文の投稿者が査読料や出版費用などのコストを負担する、ということも行われている。言い換えれば、編集という作業は相変わらず重要であり、実際に編集作業はしているのであるが、専門職としての編集者を雇うコストをかけずにすませている業界が少なからずあるのである。

こういった世界では、ビジネスとして出版活動をする場合とは自ずとコストの感覚が異なってくるであろう。守岡氏は、これまでマイナス（持ち出し）であった学術出版が電子出版によってゼロに（とまではいかないにせよそれに近く）なるかもしれない、という点に期待を寄せている研究者は少なくないのではないかと述べている。

## ◇ 電子書籍の永続性に向けて

また、特に人文科学の学術書の場合、100 年前、200 年前の本が、現在でも参照されたり、研究対象となったりすることがしばしばある。紙の本は（守岡氏の言葉を借りれば）スタンドアローンであるため、紙が物理的に劣化したり損傷したりしない限りはいつまでも読み続けることができる。

一方、電子書籍の場合、プラットフォームの永続性については疑問符をつけざるを得ない。田代氏は、新しいプラットフォームが出るたびに、古いプラットフォームから移行するサービスが登場するので大きな問題はないのではないか、という。例えば音楽 CD が出た際にはほとんどのレコードが移植され、されなかったものは移植してもビジネス的に成立しないものばかりであったとされる。

しかし、先に述べたように、そもそも学術書の多くはビジネス的に成立しないものであり、本来的に非商業的な出版物である。そういったものについては、まさにビジネスにならないという理由で、移植業者が出ずに滅びてしまうこともあるだろう（これまで販売されてきた学術データベースについても、同様の問題が繰り返し指摘されてきた）。

ここで守岡氏が強調するのは、プラットフォームを移り変わる自由である。守岡氏は先の報告で「ハックする自由」について述べていたが、自由ソフトウェアを象徴するライセンスである GPL

の考え方は、論文などの学術情報と親和性が高いという。すなわち、論文などで公開された知識は、世界人類全体の共有の知識であり、先人の業績をもとにしてそれを改良したり、新しい見解を積み重ねていったりすることが自由にできるようになっている。このような自由な考え方と正反対のものが電子書籍を支配してしまうことは、学問の自由にとって驚異になるのではないか、と守岡氏は懸念を表明しているのである。

また、このような学術出版を含めた非商業的な表現活動においては、ビジネスとして成立するかどうかよりも、他者から評価される場があることの方が重要ではないか、と守岡氏は言う。動画における YouTube やニコニコ動画は、表現活動を発表する場であると同時に、作品に対する批評活動をする場としても機能している。そのような場を特定の誰かが恣意的に制御するようなことになるのは、学問の自由という観点からも問題なのではないかと言うのである。

## ◇ 電子書籍と文化

ところで、学術書のような〝売れない〟書籍の出版をいかに維持していくか、という議論をする際にはしばしば「文化を守るため」という主張がなされる。そして書籍の再販制度についても、文化を守るということが謳われている。

> 出版物再販制度は全国の読者に多種多様な出版物を同一価格で提供していくために不可欠なものであり、また文字・活字文化の振興上、書籍・雑誌は基本的な文化資産であり、自国の文化水準を維持するために、重要な役割を果たしている [3]。

逆に再販制度がなければ「専門書や個性的な出版物を仕入れることのできる書店が今よりも大幅に減少します」と言われ、学術界にも広く受け入れられている考え方である [4]。

しかし、石岡氏によれば、再販制度は文化を守るための制度としてではなく、あくまで独占禁止法という経済法をベースとした価格設定権の問題として考えるべきであるという。文化の問題として論ずるのであれば当然、著者のことを外して考えることはできないが、再販制度は出版社・取次・書店間の制度であって、著者はその中に含まれない。また、もし再販制度が文化の維持に関わる制度だとすれば、ある特定の事案について、これは文化の維持に関わるので独占禁止法的な競争を否定してもかまわない、というような判断をするのは公正取引委員会ということになるが、公正取引委員会が文化について判断できるとは思えない。実際、著作物再販制度の改廃論議の際、文化の問題は国会では議論されていなかったという。

学術情報やマイナーな文化を後代に継承していくにはどうすればいいのか、という問題は、独占禁止法とは別の枠組みで考えるべきであると石岡氏は述べる。そういう場合には国からの補助金などで出版を維持するのが正当であって、再販というどの出版社にも当てはまるような制度を持ち出すことは筋違いであるという。しかしながら、再販制度について議論される場合には、しばしば「文化」というキーワードが出されて、問題の本質が隠蔽されるという。

## ◇ おわりに

以上、総合討論で議論されたことについて、テーマごとに簡単かつ乱暴にまとめてみた。多岐にわたる議論を一部削除しつつ、半ば強引にまとめたものであるため、当日参加された方には物足りないものになっているのではないかと恐れるが、先に述べたような事情があるということでご容赦いただきたい。

最後に、報告者の先生方、議論に参加していただいたフロアの方々、その他当日お世話になった方々に、心から御礼申し上げる。

### 注

[1] 「Togetter ── 「漢情研 2010 年度公開シンポジウム「電子出版の動向と諸問題」実況まとめ」」（**http://togetter.com/li/34725**、2010 年 9 月 26 日最終確認）

[2] これに関連して（全体討論では大きな話題にはならなかったが）電子図書館の問題を参照することができるのではないかと思う。国立国会図書館の場合、著作権法の改正以前は著作権者との個別の交渉によって進められていた電子図書館事業が（小島浩之「法理論と実務の狭間 ──「東洋学情報化と著作権問題 II」から ──」『漢字文献情報処理研究』第 5 号、2004 年）、先の著作権法の改正によって所蔵図書の一括電子化が可能になった（大場利康「著作権法改正と国立国会図書館のデジタル化を巡って」『漢字文献情報処理研究』第 10 号、2009 年）。電子書籍の場合も恐らく同様で、出版社との個別の交渉を前提とすれば、既存の出版物の電子化には膨大なコストがかかり、電子書籍の普及にとって大きな阻害要因となるだろう。Google のような強引なやり方と対向するためには、国会図書館のための著作権法改正と同様、電子書籍化を後押しするような包括的な法的枠組みが必要なのではないだろうか。

[3] 「再販制度｜社団法人　日本書籍出版協会」（**http://www.jbpa.or.jp/resale/**、2010 年 9 月 27 日最終確認）

[4] 一例として日本学術会議会長である伊藤正男氏が平成 8 年 10 月 17 日付で出した「著作物再販制度の存廃問題について（会長談話）」（**http://www.scj.go.jp/ja/info/kohyo/16/kohyo16-d2.html**）参照。

特集 1

# Windows 7 時代の『電脳中国学』と人文系情報処理テキスト

1997 年に『電脳中国学』、2000 年に『電脳国文学』、2001 年に『電脳中国学 II』と、本研究会が立ち上がってから数年は、出版ラッシュと言ってもよいほど立て続けに本を出している。これは、熱意ある執筆メンバーや、好文出版をはじめとする多くのサポーターの努力の賜物であることは間違いないが、一方であの時代の産物であったことも否定出来ない。現在『漢字文献情報処理研究』以外の出版活動は（水面下を除くと）行われていないが、それは、メンバーがそれなりに忙しくなって時間がとれなくなった、という理由だけではない。

OS が年とともにバージョンアップしていくように、『電脳中国学』もバージョンアップしていくものなのではないか。そのように期待されている方の声を時折耳にする。確かに巷のパソコン入門書は、OS やアプリケーションのバージョンを追いかけるように版を重ねている。しかし、『電脳中国学』が扱う対象はそのような汎用的なものだけではなく、例えば日中関係が時代と共に状況を変えていくように、出版を計画するたびにコンセプトから練り直さなければならないようなものである。『電脳中国学 III』のあるべき姿はどのようなものなのか。それを考えることは、中国学の現在、人文学の現在を考えることにもなるのだろう。そして、それを議論だけでなく、形にして世間に提示することが、現在求められているのかもしれない。

Contents


『電脳中国学 III』はいかにあるべきか……千田　大介 …54
『電脳中国学 III』千田案
　情報化時代における総合中国学入門ガイド ……千田　大介 …56
導入教育としての電脳中国学　千田大介案へのコメントとして ……師　　茂樹 …62
中国史学むけデジタルリソースマニュアルとしての目次案……山田　崇仁 …66

# 『電脳中国学Ⅲ』はいかにあるべきか

千田　大介（ちだ　だいすけ）

## ◇ 決まらないコンセプト

2009 年度の大会で、『電脳中国学Ⅲ』（仮）の企画案について討議した。筆者は、初級から上級までをカバーし、分野横断で中国学情報化を扱うという、『電脳中国学』シリーズのオールインワン路線を継承するものとして、電脳マニュアル＋中国学デジタルリファレンスマニュアルという案を提案したが、それに対して、内容を詰め込みすぎている、前半と後半を別々の書籍にすべきだ、パソコンにおける多漢字・多言語マニュアルという基本に立ち返るべきだ、など、様々な意見を賜った。

同大会では、師茂樹氏の導入教育教科書案、山田崇仁氏の中国学情報処理マニュアル特化案も提起された。しかしいずれも、広範な読者の獲得が難しい、あるいは読者の中国学および情報処理のレベルが想定しにくいなどの問題点を抱えている。

現在、鋭意『電脳中国学Ⅲ』の刊行に向けて検討を重ねているが、なかなかコンセプトが絞りきれず、難航している状況だ。

## ◇ 十年の歳月

振り返れば、1998 年に『電脳中国学』を出したとき、このような迷いは無かった。Nifty の歴史フォーラムから有志が独立してメーリングリストを立ち上げたのが 1997 年、やがて『電脳中国学』の企画話が持ち上がり、タイトル・目次案から本文まで、すべてを ML 上に流しては意見を聴取し、ブラッシュアップして作り上げたものだった。

当時は Windows 95/98 とともにパソコンの普及が進む、いわゆる IT バブル時期にあたり、上から下まで社会を上げて情報化の熱に浮かされていた。しかし多言語・多漢字処理という方面では、Win 95/98 はローカルコードベースの OS であったために何かと困難がつきまとい、パソコンで中国語や Unicode の 2 万字以上の漢字が使えることもほとんど知られていなかった。

『電脳中国学』が迷い無くオールインワンのマニュアルであり得たのは、こうした時代的背景ゆえである。Windows の多漢字・多言語機能と中国学学術リソースのガイドを一つにまとめて出版することは、技術・知識の普及をはかり、また来るべき本格的情報化時代における東洋学研究のあり方を示す、という意味があった。それはまた、当時の人文学界に存在していた、情報化へのバラ色の夢と漠然とした不安にもキャッチアップしていた。

## ◇ 拡散する電脳中国学

しかし今やワープロ専用機は姿を消し、Windows は Unicode 内部処理化してデフォルト機能だけで多言語を扱えるようになった。ネット上の解説ページを読めば、すぐに 7 万字以上の漢字を入力したり、中国語を扱うことが出来る。学生や院生たちは、家にパソコンがあり、学校で情報処理の授業を受けてきているし、教員も学校

の事務作業でパソコンを否応なしに使わなくてはならなくなった。

そうした中にあって、情報化への意識もだいぶん様変わりした。かつてのような切迫感が消えた反面、かつて一部の人文学研究者が抱いていた、情報化によって自らの研究が圧迫されるのではという誤解に基づく不安も解消し、逆に情報化を軽視するような風潮も生まれている。

かつて情報化に託されたバラ色の夢は、確かに多くが幻想であったが、幻想が潰えたことと夢が消えたこととは同義ではない。情報化が進展し特別のことでなくなったからこそ、基礎から理解を深める必要があるはずだ。しかし現実問題として、『電脳中国学』時代のような中国学の情報化に対する飢餓感は消えている。

このため、『電脳中国学Ⅲ』では読者のニーズを喚起する戦略が重要になろう。少なくとも授業・研究における教科書・参考書ニーズを確実に取り込まなくてはなるまい。ところがそのニーズさえも、昨今、各大学が様々なカリキュラム改革の努力を行ってきたことで、初年次教育・情報処理入門教育・人文学情報処理教育、さらには語学や教養授業の一環としての中国語・多漢字処理教育、専門授業の一環としてのデータベース検索実習など、拡散してしまっている。

これが、コンセプトを一つに定めることが難しい、最大の理由である。

## ◇ カウンターであるために

我々は、『電脳中国学』の出版によって、多くの若手研究者が中国学情報化の必要性に目覚めてくれるものと予測していた。しかしあれから十年、その予測が的外れだったことは、もはや明らかだ。電脳派の研究者は、いつまでたってもほんの一握りの見知った人たちばかりで、裾野はいっこうに広がらない。

多くの研究者の情報化に対する態度も、メディアの論調に踊らされる庶民のそれとさしたる違いはない。文部科学省の中学・高校情報教育の迷走ぶりからすると、政府すらも同じような感覚なのかもしれない。

しかし、情報化が話題にならなくなったのは、むしろ情報化が一般化し、もはや取り立てて言うべきことではなくなったからであり、その重要性はいっこうに減じていない。むしろ、社会のインフラという意味では、水や空気のように欠くべからざるものになっていると言えよう。

人文学情報化についても同様だ。台湾中央研究院を嚆矢とする各種古典文献データベースや様々な専門データベース・コーパス・OPAC・電子ジャーナルなどを何一つ利用しない研究者は、もはやほとんど居ないと思われるが、一方でデータベースごとの漢字処理のクセを把握している人、あるいは正規表現等による検索条件の指定ができる人も限られている。まして、自らあらたな学術リソースを提供する人は寥寥無幾である。

一時期、世界最先端を誇った日本中国学は、諸橋『大漢和』やさまざまな逐次索引の作成などを通じて、中国学のインフラを整備してきたが、今や地盤沈下が著しい。デジタル時代になり、次世代中国学インフラたる大規模文献データベースの構築は台湾・中国の独壇場となり、日本中国学が学術インフラ整備に関わるケースはめっきり少なくなった。こうしてみれば、情報化が言われなくなった今であるからこそ、人文学情報化の意味を改めて訴えることが、本質的な意味でカウンターたり得るのではないか。

『電脳中国学Ⅲ』は、是非ともそのような存在になってもらいたいものである。限られたマンパワーでそれを実現するには、いかなる戦略が有効だろうか？ 2009 年度の大会を振り返りつつ、考えてみたい。

※本稿は、漢情研メールマガジン第一六二号（2010.7.1）掲載のコラムを手直ししたものである。

# 『電脳中国学Ⅲ』千田案

## 情報化時代における総合中国学入門ガイド

千田　大介（ちだ　だいすけ）

### ◇ はじめに

本稿は、2009 年 12 月 20 日の漢情研大会で発表した『電脳中国学Ⅲ』コンセプト案と目次案を改めて整理・紹介するものである。発表の際には、本案について諸賢から様々な貴重なご意見を頂戴し、またその後、執筆者の個人的な考えも変化しているが、以下では特集構成上の都合もあり、発表時の案を改めることなくそのまま紹介し、後に問題点や変更の方向性について、思うところを付記することとしたい。

### ◇ 『電脳中国学』・『電脳中国学Ⅱ』を振り返る

漢情研の正式な発足のきっかけともなった『電脳中国学』は、1998 年に刊行された。キャッチコピーは「Windows で漢字・中国語を扱う方法から、最新東洋学サイト情報までを満載した、究極のマニュアル!」、PC での多漢字・多言語処理マニュアルにとどまらず、PC・ネットによる情報の収集から発信まで、幅広く網羅した内容となっていた。想定読者は、パソコンで多漢字・多言語が使いたいのにやり方がよくわからない、専門課程に進んだ学生や研究者である。

『電脳中国学』の編集に際して、とりわけ意識されていたのが『人文系論文作法──Windows95 版』（中尾浩・伊藤直哉著、夏目書房、1998 年）である。同書は、PC・ネットを駆使した論文執筆の手順を解説しており、情報時代の研究リテラシー本とも言うべきものである。『電脳中国学』の情報収集・分析・発信という構成は、研究リテラシー本としての同書のあり方の影響を受けたものである。

『電脳中国学』では多漢字・多言語処理のテクニックについて多くの紙幅を割いたが、それは当時が Unicode の普及と発展が始まったばかりであり、また Windows 9x の内部処理が Unicode 化されていないなど過渡的な時期にあたり、多漢字・多言語処理に一定の難度があったことによる。

『電脳中国学』は幸いにして好評を博した。類書が存在しなかったこともさることながら、時に日本は IT バブルのさなかで、中国学研究者の間にも強迫観念にも似た IT 化への焦りがあったことが大きかったのではなかろうか。

2001 年に刊行された『電脳中国学Ⅱ』は、Windows 2000 ＋ Office XP に対応した、『電脳中国学』シリーズ第二弾である。同書では『電脳中国学』について寄せられた感想──難しすぎる、多漢字・多言語処理から Word の使い方まで一冊で全て分かる本がほしい、などなど──に応えてほしい、とのリクエストを受けて、入門レベルの解説をより厚くしている。Word のマニュアル的内容が手厚いのは、このためである。

しかし、『電脳中国学Ⅱ』は売り上げ部数が伸びなかった。原因の一つには、〝Ⅱ〟というネーミングが、より高度な内容を想起させたことがあろう。また、せっかく厚くした入門レベルの解説が単なるテクニックの解説にとどまっていたことも、

反省すべき点である。パソコンを上達したいというモチベーションがあれば、『できる』シリーズなどの入門書を購入するはずであり、それを購入せずに他のマニュアルを求めている人に対しては、むしろモチベーションを喚起する動機付けの仕掛けこそが必要であったのではなかろうか。

## ◇ 『電脳中国学Ⅱ』後の変化

### ⊙ PC 環境

『電脳中国学Ⅱ』から 9 年、中国学情報化を取り巻く環境は大きく変化している。

PC 環境については、当時の Windows 2000 と現在の Windows 7 とでは、多言語の設定方法は特に変化していないが、Unicode の CJK 統合漢字 Ext.B までの約 7 万字を網羅したフォントが標準で搭載され、Windows の多言語インターフェイスがネット経由で提供されるなど、より簡便になっている。多言語・多漢字処理を解説した Web ページも増えている。

古典文献データベースについては、中国基本古籍庫をはじめ、多くの大規模データベースの開発が進み、数十億字規模の文献が一気に検索出来るようになっている。

何よりも大きな変化は、インターネットリソースの充実であろう。ネット上に古典から現代まで多種多様なテキストデータ・画像データが ── 違法なものも含めて ── あふれかえるようになり、また Google・百度などのサーチエンジンの普及と成長によって、そうしたデータへのアクセスが容易になった。

情報の発信という面でも、ブログ・ツイッター・クラウドコンピューティングなど、インターネットの発展に伴って登場した新たなサービスが脚光を浴びている。その結果、HTML 言語を意識する必要のないサービスが主流になってきている。

教育環境という面でも、さまざまな変化が見られる。今や小学校でもコンピュータを使った授業が行われており、大学でもコンピュータ教室や情報対応教室の普及が進んだし、教員にとっても各種業務に PC は必需品になっている。また家庭でも PC が普及しており、家に PC が無いという学生は今や少数派である。

学生・教員にとって PC はもはや〝普通の道具〟になっている。以前のような過剰な〝電脳幻想〟が消滅し、PC にフラットに向き合えるようになったのは良いことではあるが、一方で PC はモチベーションを喚起するツールではなくなっている。

### ⊙『電脳中国学Ⅲ』が使われる授業とは

『電脳中国学』の成功と『電脳中国学Ⅱ』の失敗、さらに昨今の PC の一般ツール化を考えあわせると、『電脳中国学Ⅲ』のコンセプトを策定する上での鍵が、モチベーションの喚起にあることは明白である。このモチベーションとは、読者に本の購入を喚起するような直接的なものでもかまわないが、昨今の若者の内向き傾向を考えるとむしろ、大学教員が教科書・参考書指定したくなるようなモチベーションの喚起を考えるべきであろう。

では、『電脳中国学Ⅲ』を教科書・参考書として指定する可能性があるのは、いかなる授業であろうか。

第一に考えられるのが、中国学の入門科目である。専門課程に進んだ一・二年生を対象に、資料調査や研究史を教授したり、対象となるジャンルのアウトラインを概説する授業は、大半の中国学選考過程に設置されていることだろう。中国学情報処理講座も、そうした授業の一形態であると言えよう。

また、より専門に特化したデータベースやテキストの分析手法などの解説は、学部・大学院の専門授業の参考書ともなり得よう。

昨今、専門課程の入門科目とは別に、主に一年生を対象に、大学における専門学習・研究の前提となるスキルを教授する、いわゆる〝初年次教育〟（導入教育）が脚光を浴びている。『電脳中国学』シリーズは『人文系論文作法』の影響もあり、当初から初年次教育的な内容を含んではいたが、読者対象を専門課程の学生・研究者に置いていたため、その方面の授業に合わせた構成にはなっていなかった。検索技法や論文データベースの使い方、

MS Office のスキルなどの解説を、知的作業の過程に位置づけることで、初年次教育ニーズに応えることもできよう。

このほか、語学授業や一般教養での参考書指定も考えられる。例えば、中国語の授業の一環として、中国語のタイピングやブラウズ、Google の中国語検索方法を実習するとともに、広く中国学への興味を喚起するようなケースもあり得よう。

『電脳中国学Ⅲ』では、これらをはじめとしたさまざまなニーズを想定し、その全て、乃至一部を、確実にキャッチアップしていく必要があろう。

### ⊙ デジタルとアナログのミックス

前述のようにコンピュータやネットワークが社会の隅々にまで浸透したことで、PC に対する意識も魔法のアイテムからごく一般的な道具へと変わりつつある。それと同様に、中国学研究のツールとしてのパソコン・インターネットも、他のアナログ的手法とフラットに扱われ、教授されるべきものになってきている。

『電脳中国学』シリーズでは、Web リソースやデータベース製品など、中国学デジタルリソースの解説を一つの柱としてきたが、こうした時代の変化を受けて、その記述方法をも改める必要があろう。

考えてみると、広く中国学全般の調査方法や工具書をまとめた参考書籍としては、『中国学レファレンス辞典』（潘樹広編、松岡榮志訳、凱風社、1988）があり、また『中国文学語学文献案内』（第四次修訂版。早稲田大学中国文学会、1994）・『漢文研究の手引き』（第四次増補版、中国詩文研究会、2000）などもある。

このうち『中国学レファレンス辞典』は、中国の各種工具書のガイドブックというべき内容で、文史哲の各ジャンルを網羅しており、画期的な意味を持ったが、いかんせん 1980 年代の刊行と古く、中国における 1990 年代以降の工具書ブームを反映していないため、今となっては時代遅れになっている部分も多い。また、後二者についても 20 世紀で更新がストップしてしまっている。

このように、現在、各研究ジャンルごとの研究方法に特化した入門書はいくつも見られるものの、中国学全般を横断的に眺め渡した調査方法のガイドブックは、欠けた状態にある。今の時代における中国学の調査方法として、ネット・データベースなどを駆使するデジタル的手法を欠くことはできないが、一方で、調査手法がデジタル化されていない分野も多いし、デジタル化の過程で抜け落ちてしまうアナログ情報もあり、現状ではアナログ・デジタル、どちらの手法も一長一短である。

『電脳中国学』シリーズでは、研究ジャンルごとに有用なデジタルリソースを紹介してきたが、広く中国学全体の状況を考えると、文史哲の諸分野を横断的に扱い、デジタル・アナログ両方の調査・研究手法を公平に比較して、最善の調査・研究方法を提示するマニュアルこそが必要なのでは無かろうか。

## ◈ 『電脳中国学Ⅲ』のコンセプトと構成

『電脳中国学』シリーズは、初級・入門レベルから専門研究レベルまでをカバーし、また、文史哲、さらに仏教学までも網羅した幅の広さを特徴としてきた。『電脳中国学Ⅲ』でもそうした路線を踏襲するという前提に立つと、〝情報化時代における総合中国学入門ガイド〟というコンセプトが浮上する。

『電脳中国学』シリーズで解説されているスキルのうち、Windows や Office の多言語設定、入力方法、ブラウジングやメールなどは、OS・アプリケーションの基本操作にあたるので、第 1 章にまとめて解説する。

一方、Google 等の検索スキルや、MS Office の使いこなしなどは、いずれも、論文・レポートのためのライティングや情報収集といった文脈で取り扱うべき、アカデミックの基本スキルに属するものである。第 2 ～ 5 章では、論文・レポート作成を念頭に、テーマ選定→調査→執筆→プレゼンテーションという手順に沿って解説する中に、アナログ・デジタルの各種スキルを解説する。

以上は全体を教科書的に構成し、1 課 4 ～ 6 ペー

ジ、4～5課を1セクションとし、全5～6セクションで構成し、授業の形態に合わせて、各セクションを取捨選択して利用できるようにする。

第6章、テーマ別中国学リファレンスガイドとする。人名・地名、さらには出土文献情報の入手、唐詩の解釈、宋詞の典故、清代地方社会の調査などなど、具体的な調査項目について、1項目4ページほどで解説する。前述のように、アナログ的技法・デジタル的技法を同列に扱い、現時点における最善の調査方法を提示するものとする。

章の冒頭には台湾中央研究院・電子版『四庫全書』・基本古籍庫などの、あらゆる分野で参照される大規模データベースの操作方法を掲載する。また、各学問ジャンルの入門書についても章の末尾で、簡単に紹介する。

## ◇ 目次ラフ案

0. はじめに
  - 0.1. 本書の目的
  - 0.2. 本書を読む前に

1. パソコンの準備と基本
  - 1.1. 文字コードとフォントについて
  - 1.2. Windows7 で多漢字・多言語を使う
    - 1.2.1. Windows7 のバージョンによる違い
    - 1.2.2. Widnows7 の多言語設定
    - 1.2.3. Windows7 で海外ソフトを使う
    - 1.2.4. XP モードを使う
  - 1.3. Office2010 の設定
    - 1.3.1. Office2010 の多言語設定
    - 1.3.2. MLP を使う
  - 1.4. 多漢字と中国語を使ってみる
    - 1.4.1. 変換できない語句・文字を入力する
    - 1.4.2. 中国語を入力する
    - 1.4.3. 中国語ウェブページを見る
    - 1.4.4. Win7・Office2010 の対応アジア言語
  - 1.5. メールでコミュニケーション
    - 1.5.1. Firefox の使い方
    - 1.5.2. Gmail の使い方
    - 1.5.3. Skype で国際コミュニケーション
  - 1.6. セキュリティに注意しよう

2. テーマの設定と絞り込み
  - 2.1. 論文・レポート執筆の第一歩
    - 2.1.1. 情報交換・共同研究と研究の戦略
  - 2.2. まずあたりをつける
    - 2.2.1. Google してみよう
    - 2.2.2. 百科事典を使う
  - 2.3. 図書館蔵書を探す
    - 2.3.1. 図書館と OPAC
    - 2.3.2. 漢籍を探す
    - 2.3.3. 中国書籍を探す
    - 2.3.4. 電子図書館を使う
  - 2.4. 書籍を購入する
    - 2.4.1. 書店とオンライン書店
    - 2.4.2. 中国書籍を購入する
    - 2.4.3. 古書を買う
  - 2.5. 論文を探す
    - 2.5.1. 論文イモヅル検索
    - 2.5.2. 日本語の論文を探す
    - 2.5.3. 雑誌を探す
    - 2.5.4. 新聞資料を探す
    - 2.5.5. マイクロフィルムを見る
  - 2.6. 国内外の主要中国学関連図書館・収蔵機関

3. 資料の読解と調査・研究の深化
  - 3.1. 研究深化のワークフロー
    - 3.1.1. グループ研究の進め方
    - 3.1.2. ネットで共同作業
  - 3.2. まずは辞書から
    - 3.2.1. はじめに
    - 3.2.2. 諸橋大漢和
    - 3.2.3. 漢語大詞典
    - 3.2.4. 漢語大字典
    - 3.2.5. 現代中国語の辞書

3.3. 出典・用例を探す
3.3.1. 出典・用例探しの重要性
3.3.2. サーチエンジンを使う
3.3.3. テキストデータベース（コーパス）を使う
3.3.4. エディタでテキストデータベース
3.3.5. 必要な文献の電子テキストを作る
3.3.6. さまざまな分析ツール・技法
3.4. 用例・事例の整理方法
3.4.1. カードデータベースを作る
3.4.2. カードを取る
3.4.3. Excel でデータ管理
3.4.4. 画像・映像の収集と整理

4. 論文・レポートを書く
4.1. 構想をまとめる
4.2. 論文・レポートのための Word マニュアル
4.2.1. 本文と見だしの体裁
4.2.2. 引用の方法
4.2.3. 箇条書きとリスト
4.2.4. 表を作る
4.2.5. 脚注と文字カウント
4.2.6. 図の挿入
4.2.7. グラフの挿入
4.2.8. ルビを振る
4.2.9. 目次を作る
4.2.10. Word で訓点文
4.2.11. スタイル機能の使い方

5. 研究をプレゼンテーションする
5.1. はじめに
5.2. PowerPoint を使う
5.3. Word でレジュメを作る
5.3.1. 聴衆の立場に立ったレジュメ作成
5.4. Google Earth で地理的プレゼン
5.5. 研究情報をインターネットで発信する
5.5.1. ブログで研究情報発信
5.5.2. リポジトリへの登録
5.5.3. 学術雑誌のデジタル化・ネット公開

6. ジャンル・目的別リファレンスガイド
6.1. 古典総合データベース
6.1.1. 漢籍電子文献
6.1.2. 四庫全書
6.1.3. 四部叢刊
6.1.4. 中国基本古籍庫
6.2. テキストデータの入手
6.2.1. ウィキソース
6.2.2. 国学
6.2.3. 西陸
6.3. テーマ別調査方法
※各項目原則 1 ～ 2 ページ
※各ジャンルで共通性の高いもの
※各ジャンルの主要文献・工具書
6.3.1. 清代以前の人物の調査
6.3.2. 官職名と官制
6.3.3. 地理の調査
6.3.4. 地方誌を探す
6.3.5. 唐代の詩文を探す
6.3.6. 宋代の詩文を探す
6.3.7. 詩文の典故を調べる
6.3.8. 古典小説・戯曲の語彙を調べる
6.3.9. 古典小説を検索する
6.3.10. 戯曲の出典を調べる
6.3.11. 仏典を調べる
6.3.12. 道教経典を調べる
……
6.3.**. 近現代人物の調査
6.3.**. 公文書を探す
6.3.**. 現代中国の統計資料
6.3.**. 中国の新聞を探す
6.3.**. 中国のテレビ・ラジオの視聴
6.3.**. 中国映画を調べる
6.3.**. ポップカルチャー
6.3.**. 現地調査を手配する
……
6.4. ジャンル別入門書・概説書・ウェブサ

イトガイド
- 6.4.1. 書誌学
- 6.4.2. 文学
- 6.4.3. 言語学
- 6.4.4. 史学
- 6.4.5. 哲学・宗教学
- 6.4.6. 美術
- 6.4.7. 演劇学

## ◈ 問題点

本目次案の問題点は、ボリュームの問題に尽きる。

入門レベルから研究レベルを網羅し、中国学諸分野に跨るため、盛り込まれる内容のボリュームは膨大なものとなる。それに比例して、物理的ボリュームもふくれあがり、総ページ数は 500 ページを下るまい。これでは、気軽に教科書・参考書に指定するわけにもいくまい。大会でも、前後を二つに分けて別個の書籍として出すべきだ、との意見を頂いた。

『電脳中国学』シリーズの総合性という特色を守るとしたら、内容を刈り込むしかない。しかしそれで、現在の状況を踏まえた、先鋭的な本ができるとも思えない。やはり、シリーズのあり方から見直さなくてはならない時期にさしかかっているのだろう。

# 導入教育としての電脳中国学

## 千田大介案へのコメントとして

師　茂樹（もろ　しげき）

### ◇ はじめに

1998 年に初代『電脳中国学』（以下「初代」）が刊行されてから、すでに 12 年が経とうとしている。その間に、パソコンやインターネットをめぐる状況、中国と日本との関係、そして大学などでの中国語や中国文化に関する教育も様変わりしたように思う。新しい『電脳中国学』がそのような状況の変化に対応したものでなくてはならないのは言うまでもないが、（広い意味での）中国学のコンピュータ利用については全員がスタートラインに並んでいた初代の時とは異なり、現代は読者の関心もコンピュータのスキルもバラバラで、どの層に照準を絞ればよいのか難しい。

したがって正直なところ、以下に述べようとしている初年次教育を念頭に置いた目次案については、筆者はまるで自信が持てないでいる。大学関係者だけではなく、社会人にも広く手にとってもらうことができた初代とは異なり、筆者の目次案は、大学の、しかも中国・東アジア関係という限定されたコースでの、さらにしぼって初年次教育における利用を想定しているものなので、明らかに初代より読者人口が減ることも予想されるからである。

一方で、新『電脳中国学』が中国に限定されず、近接する日本文学や日本史学などを含めたより広い読者に認められたならば、勝算はあるのではないかとも皮算用をしている面もある [1]。とは言え、筆者は自分に経営センスのようなものがないことを自覚しているので、目次案の是非は読者諸賢に委ねたい。

なお、第十二回大会（2009 年 12 月 20 日）の企画討論会「Windows7 時代の〝電脳中国学〟と〝人文系情報処理テキスト〟」において筆者は、後藤真・田中正流・師茂樹『情報歴史学入門』（金壽堂出版、2009 年）のコンセプトに基づいたコメントを行ったが、その後の討論等を経て本稿のような目次案になったことを付言しておく。

### ◇ 大学教育の現状

近年、18 歳人口の減少などを背景にして、大学全入時代あるいは大学のユニバーサル化とよばれるような状況を迎え、多様な学生が大学に進学するようになってきている。それにともない、「リカレント教育」などと称される高校以前の科目の再教育、補習授業を行う大学が増加し、また大学での主体的・専門的な学びにスムーズに移行できるようにするための初年次教育あるいは導入教育とよばれる教育プログラムがほとんど大学で行われるようになってきている。

初年次教育においては、インターネットによる情報収集・ワープロ・表計算・プレゼンテーションの四つが取り上げられることが多い。一例として、初年次教育の教科書としてよく知られている

学習技術研究会編『知へのステップ改訂版』（くろしお出版、2006 年）の目次を見てみよう。

第Ⅰ部　はじめに
　第１章　スタディスキルズとは
第Ⅱ部　聴く読む
　第２章　ノートテイキング
　第３章　リーディングの基本スキル
　第４章　より深いリーディングのために
第Ⅲ部　調べる整理する
　第５章　大学図書館における情報収集
　第６章　インターネットによる情報収集
　第７章　情報の整理
第Ⅳ部　まとめる書く
　第８章　アカデミックライティングの基本スキル
　第９章　効果的なアカデミックライティングのために
　第10章　パソコンによるライティングスキル
第Ⅴ部　表現する伝える
　第11章　プレゼンテーションの基本スキル
　第12章　わかりやすいプレゼンテーションのために
第Ⅵ部　巻末資料

ジャンル・目的別リファレンスガイドの部分を除くと、上の目次が千田氏の目次案と少なからず共通している点は注目しておきたい。

しかし『知へのステップ』は、学部・学科の専門課程を超えた、汎用的な初年次教育プログラムとして設計されたものである。一方、学部・学科における専門的な教育を前提とした初年次教育のあり方についても、多くの大学で実践されている。中国学をはじめとするアジア系の学部・学科においても、そのようなニーズはあってもおかしくない。筆者の提案する新しい『電脳中国学』の目次案は、このようなニーズに答えることができるような教科書であることに重点を置いて組み立ててみた。

## ◇ 読者の想定

教科書が想定する主な読者はもちろん学生である。しかし、最近の学生は授業で使わない書籍はどんなに教員が勧めてもなかなか買わないので、シラバスの中で教科書指定されることが、学生が手に取るようになる最大の近道であろう。ということは、学生が必要と思うかどうかではなく、教員に必要だと思われなくてはならない。

では、教員のニーズを分析してみよう。初年次教育・導入教育はその性質上、語学教育のように同一の科目を複数クラスに別れて行うことが多い。したがって教員は複数いることになる。しかし、語学教育と異なり、教員自身が学生だった頃にはなかったタイプの授業であるため、実際にどのように授業を行えばいいのかとまどう教員も少なくない。加えて、自分で論文を書いたりプレゼンをしたりするための情報処理能力はそれなりに持っていたとしても、それを学生に教える経験や自信を持っている教員も多いとは言えない。

したがって教員のニーズとしては、

- 導入教育や情報処理の経験に乏しい教員でも、それなりの授業ができるような内容。
- 15 回（× 2）の授業（＋テスト）という構成になっていることが望ましい（目次がそのままシラバスになる）。
- パソコンの参考書などで補足する場合も考慮すると、教室のソフトウェアやハードウェアは標準的であることが望ましい。

という具合になるだろうか。

## ◇ 目次案

以上の現状分析をふまえて、以下に目次案を提示する。初年次教育や人文情報処理教育についての筆者の乏しい経験から作ってみたものであるので、不十分な点が多いのではないかと思う。また、通年の授業の割には内容が盛りだくさんで、３～

4 週かけてもよいようなものを 1 ～ 2 週でやってしまう、というような窮屈なものになっているかもしれない。これらの点も踏まえて、ご批判をいただければ幸いである。

### ⊙ 前期

中国語能力を前提としないがネタが東アジア系

| | |
|---|---|
| 1 日目<br>大学での学びとパソコン | • 大学での「学び」と高校までの「勉強」との違い |
| 2 日目<br>情報モラル | • 情報セキュリティについて<br>• 著作権などの法令について |
| 3 日目<br>インターネット・コミュニケーション・スキル | • ビジネス（風）メールの書き方<br>• 携帯電話のメールとの違い |
| 4 日目<br>発表・レポートに向けた計画を立てる | • 必要な手順（テーマ決定・問題発見→先行研究調査→実際の調査→レポートなどの作成）の説明 |
| 5 日目<br>巨人たちの肩に乗る(1) | • 先行研究調査の重要性の説明<br>• 大学図書館における情報収集の基礎（一次資料と二次資料、書籍と雑誌など）<br>• OPAC、CiNii などの使い方 |
| 6 日目<br>巨人たちの肩に乗る(2) | • 中国関連のリファレンスブックの使い方<br>• 国会図書館アジア情報の調べ方（**http://rnavi.ndl.go.jp/asia/entry/research-guide-asia.php**）など |
| 7 日目<br>巨人たちの肩に乗る(3) | • サーチエンジンを使いこなせ（AND 検索など）<br>• Wikipedia を使うときの注意点 |
| 8 日目<br>巨人たちの肩に乗る(4) | • ウェブ以外の情報源 |
| 9 日目<br>巨人たちの肩に乗る(5) | • 集めた先行研究を読み、比較したりしながら問題点を抽出<br>• 専門辞書を使いながら読む<br>• ノートを取る（Excel やエディタによる文献ノート、コラムでマインドマップ、Evernote など） |
| 10 日目<br>調査(1) | • 調査の方法（先行研究で言われていることが、他の資料でもあてはまるか確認してみる／先行研究が言っていることと違う証拠を探してみる etc ……（学生にとっつきやすいテーマで）<br>• ※人文学の場合、問題発見→資料調査→問題解決みたいな流れより、資料調査→問題発見→問題解決みたいな流れなので、このやり方は教員にとって馴染みがないかも知れない。 |
| 11 日目<br>調査(2) | • 実際の調査の例<br>• 日本の A 新聞では「中国の若者は○×だ」って書いてあった。ほんとかな？<br>• → B 新聞／中国の新聞の日本語版を検索してみたら同じネタについての記事があった！<br>• →読んでみたら、A 新聞とぜんぜん違うことが書いてあった！ |
| 12 日目<br>発表・レポートに向けて(1) | • レポートの書き方、レジュメの作り方（レジュメとは何か？など） |
| 13 日目<br>発表・レポートに向けて(2) | • 脚注や参考文献の書き方 |
| 14 日目<br>発表・レポートに向けて(3) | • ワープロの様々な機能（構造化とかスタイルとか） |
| 15 日目<br>発表・レポート | • 反省、自己評価、教員による評価 |

### ⊙ 後期

初等中国語能力を前提

| | |
|---|---|
| 1 日目<br>中国学とパソコン | • 中国系の勉強をするときにはネットは不可欠 |
| 2 日目<br>中国のネット事情 | • 中国のネットの常識（検閲、海賊版など）<br>• パソコン以外のネットワーク（携帯電話、QQ など） |
| 3 日目<br>中国語入力の基礎 | • 中国語 IME の使い方（ピンイン入力など）<br>• 読めない漢字の出し方 |
| 4 日目<br>巨人たちの肩に乗る 6 | • 中国語の書籍、論文を探す（各種サービス）<br>• 入手方法（国会図書館の複写サービスなども含む）、本の買い方など |
| 5 日目<br>調査 3 | • 中国のサーチエンジンを使いこなせ（中国の検索サイトの違いなど） |
| 6 〜 7 日目<br>調査 4 | • 各分野ごとの資料調査方法（主要データベース紹介）<br>• 代表的な古典籍・文献史料などのデータベースの使い方<br>• ※検索結果を見ても読めない場合があるので、訳注本なんかへのリンクが必要<br>• 新聞記事や統計データなどなどのデータベースの使い方￥<br>• ※中国の現代語について調べる方法、翻訳サービスの使い方なども必要 |
| 8 日目<br>調査 5 | • 実際の調査の例<br>• 先行研究では「○×」という語は〜という意味で使われていると書いてあった。本当だろうか？<br>• →文献データベースで検索してみたらヒットした<br>• →訳注を見てみたら、違う解釈が書いてあった！<br>• 日本の新聞では「中国の若者は○×だ」って書いてあった。本当だろうか？<br>• →中国のニュースサイトを検索してみたら記事があった！<br>• →自力で読んでみたら／翻訳してみたら、日本の新聞とぜんぜん違うことが書いてあった！ |
| 9 日目<br>発表・レポートに向けて 4 | • 中国語・多漢字をワープロで使うための様々な機能（フォント漢文入力などの基礎テクニック） |
| 10 日目<br>発表・レポートに向けて 5 | • PowerPoint の使い方 |
| 11 日目<br>発表・レポートに向けて 6 | • 効果的なプレゼンのために（しゃべるスピード、スライドの数などのテクニック） |
| 12 日目<br>発表・レポート | • 反省、自己評価、教員による評価 |
| 13 日目<br>次のステップへ向けて 1 | • 中国語情報処理の基礎知識（文字コードとか）<br>• 書籍案内 |
| 14 日目<br>次のステップへ向けて 2 | • 中国語メールの書き方 |
| 15 日目<br>総括 | |

### 注

[1] 実際、筆者は某大学の歴史学専攻（日本史・東洋史・西洋史を含む）の学生に対して『電脳中国学Ⅱ』を使った授業をしたことがある。ちなみにこの時は、本来の担当教員が急に担当できなくなったための年度ギリギリに決まった代打であったため、シラバスを書いたのも『電脳中国学Ⅱ』を教科書に指定したのも筆者ではない。つまり、『電脳中国学Ⅱ』の共著者でもある筆者が売上を伸ばそうとしてむりやり教科書指定をした、というわけではなく、歴史学の教育において『電脳中国学Ⅱ』が選択されていたということなのである。

# 中国史学むけ デジタルリソースマニュアル としての目次案

山田　崇仁（やまだ　たかひと）

## ◇ はじめに

2009 年の漢字文献情報処理研究会大会で筆者が担当したのは、大学で古典系中国学（いわゆる文史哲分野）を専門として学んでいる学習者に向けて、実際に利用するアプリケーションや Web サイトの使い方を学習するために使う教科書、という内容であった。

ここで掲載する目次案は、その時の報告を下にして目的や到達目標、実際の授業案などを前提に作成したものである。

そのため、人文学や中国学の初学者教育に電脳分野を組み合わせた目次案を提示するであろう千田氏や師氏に比べ、筆者の目次案はむしろマニュアル本に近いものになっている。そのため、版型についても、従来の『電脳中国学』より大きめのもの（B5 変型・A4）を想定している。

これは、授業で専門的な内容の一環としてデジタルリソースの利用方法を教える必要性が出てきた場合、上記のような包括的な入門書とは別に、実際の授業で使用・紹介可能な副読本的なマニュアル集やレジュメが必要となるだろうと目論んだからである。従って、筆者の提示するものは、書籍と言うよりも雑誌のムック版に近い。そのため、利用形態としては、レジュメ代わりに使用していただければと考えており、内容についても単にソフトや Web サービスを紹介するのではなく、多少冗長になってもその使用方法についても懇切丁寧な解説を行うことにした。

## ◇ 全体構成

### ⊙ 本目次案の概要

上記のように、本目次案は中国学向けのデジタルリソースの具体例を列挙するだけではなく、各 Web サイトの特徴やその利用方法をマニュアル本式に紹介する内容を予定している。

対象とする読者層は、中国学の学習者でかつデジタルリソースを利用して学習活動を行いたい人であり、加えてそのような授業を設定した教員の教科書・参考書の需要を見込んでいる。

ただし、この分野の専門向けな学習内容の提示や書籍の紹介については省略する。これは、本書はあくまでインターネットリソースのマニュアル本に特化することを念頭に置いているからである。

本書を使用して授業を行う場合は、各担当者の必要性の優先順位や、勤務先に導入されているソフトウェアやブラウザなどの対応状況に応じて本書から内容を適宜セレクトしていただければよい。環境の状況のため授業では使用できない部分についても、受講生が興味を持った範囲で自ら利用すればよいと考えている。

### ⊙ 到達目標

以下の内容で、本目次案で学習・到達可能な内容を設定した。基本的に紹介するデジタルリソースは、読者層の負担を考えて無料あるいは機関利用可能なものを中心に据えているが、当該分野のより深化した学習を行う際に必要と思われるものについては、有料であっても積極的に紹介する予定にしている。

- ●多漢字の入力や閲覧など、デジタルリソースの読み書きレベルから学習し、その能力の向上を目標とする
- ●その能力を前提にしつつ、Web サービスの利用方法について学ぶ
- ●実習系（ソフトウェア・Web リソース）以外に必要となる知識（文字コード・デジタルテキスト・著作権など）の解説

## ◈ 目次案

〝はじめに〞で述べたように、本目次案は中国学向けデジタル情報基礎知識をその基本コンセプトに持っている。

その中で、学習すべき内容についていくつかの章を設け、その中で必要な知識の提供やソフトウェアや Web リソースについて利用方法を中心に紹介するスタイルをとる。

### ⊙ 読み書きの基本

この章では、デジタルリソースを利用する前に、最も利用機会が多いと思われるデジタルテキスト（含む文字コード）について事前学習をすることを目標としている。

そのため、本書の中では例外的に実践的な内容ではない。

しかし、この部分はデジタル読み書きの最も基礎であるので、マニュアル本であっても最低限の学習は必要だと考えている。

また、それに加えて、テキスト情報と並んで利用機会が多いと思われる画像や PDF についても採り上げることにしている。

**■デジタル化された文字について**

- ◆文字コード
  - ●日本の文字コード規格
  - ●中国（GB, CNS(Big5)）の文字コード規格
  - ●Unicode について
- ◆デジタルテキスト

**■フォントについて**

- ◆多漢字対応のフォント
- ◆フォントを作ろう
  - ●FontForge
  - ●グリフウィキ

**■IME を使いこなせ**

- ◆日本語 IME を使おう
- ◆中国語 IME を使おう
  - ●ピンイン IME
  - ●WG 拼音輸入を利用してピンイン入力
  - ●注音 IME
  - ●Chinese Writer を使ってみよう
  - ●簡体字←→繁体字変換
- ◆IME で出てこない文字を検索・入力
  - ●繁体字 IME の手書き入力
  - ●CHISE IDS FIND
  - ●教育部異体字字典

**■画像によるデジタル情報**

- ◆画像の種類
- ◆閲覧用ソフトウェア
  - ●Irfan View
  - ●Leeyes
  - ●DjVu
- ◆PDF を使ったデジタルデータ
  - ●Adbe Acrobat, Adobe Reader で
  - ●Adobe 社以外の PDF 作成・閲覧ソフト

### ⊙ インターネットリソースを使いこなせ！

本章が、本目次案のメイン部分となる。

まず、利用の大前提となる言語を指定しての

Web サイトの閲覧方法から解説し、中国のネット事情や著名どころの Web サイトの特徴についても簡単に解説する。

その次に、サーチエンジンを紹介する。世界的なシェアが高い Google と中国でシェアが高い百度を紹介している。

その後は、初学者教育向けに必要な項目毎に、利用機会が多いと思われる Web サービスの利用方法について解説している。

■WEB ページを見てみよう

◆言語を指定した Web サイトの見方
- ●中国語 Web サイトの見方
- ●多漢字 Web ページの見方

◆中国について知るために
- ●公的情報を利用（中华人民共和国中央人民政府门户网站・台湾総統府・日本国外務省の公開する情報）
- ●中国のネット事情（金盾など）

◆ニュースサイトを見よう
- ●人民日報・新華網の見方（中国語版・日本語版の違いを知る）

◆掲示板（BBS）を読み解く
- ●BBS を眺めて、ネット世論を知る
- ●ネットスラングを知る

■サーチエンジンを使いこなす

◆Google を利用した下調べ
- ●Google 日本語版の使い方
- ●オプション検索を使いこなせ
- ●画像を調べる
- ●Google 中国語版

◆百度を使いこなせ
- ●使い方解説
- ●百度画像検索

■インターネット辞典を利用しよう

◆Wikipedia を使いこなせ
- ●各言語版の違いを知る（内容や表記から翻訳語を知る）
- ●Wikipedia をどう評価するか

◆JapanKnowledge を使ってみよう
- ●辞書（辞典）を指定した検索
- ●串刺し検索で情報比較
- ●東洋文庫を見てみよう

◆辞書・翻訳サイトを使ってみよう
- ●ポータルサイトの翻訳サービス
- ●英辞郎
- ●北辞郎
- ●漢典
- ●知識詞典
- ●複数翻訳エンジンを比較してみよう

■図書・文献を調査する

◆図書を調べる
- ●OPAC で図書館の蔵書を調べる
- ●NACSIS WebCat で全国大学の蔵書状況を知る
- ●北京国家図書館の蔵書を調べる
- ●台湾国家図書館の蔵書を調べる
- ●NBINET で台湾蔵書を調べる
- ●ネットで書籍を調べる
- ●Google ブック検索
- ●超星数字図書館
- ●Amazon 中み！検索
- ●Scribed

◆漢籍を調べる
- ●漢籍データベース
- ●機関独自の漢籍データ（東大・早稲田・京大など）
- ●中國古籍善本書目導航（北京国家図書館）
- ●數字方志（北京国家図書館）
- ●台湾の漢籍系データベース

■論文を手に入れよう

◆論文調査のこつ

◆論文データベースを使いこなせ
- ●Cinii
- ●学術リポジトリ
- ●CNKI
- ●万方数拠庫
- ●龍源

●INBUDS
●簡帛

■デジタルテキストを入手する
◆Wiki リソース
◆中華系デジタルテキストサイト
◆Project Gutenberg

■典拠調べ
◆文字・語彙を調べる
●オンライン版字典・詞典
◆漢籍系データベース
●漢籍電子文献
●寒泉
●中国哲学書電子化計画
●唐宋文詞資料庫（元智大学）
◆仏典を調べる
●SAT
●CBETA
◆文学を調べる
●中国通俗小説書目データ検索
◆出土資料を調べる
●文物図像資料庫（台湾）
●京都大学人文科学研究所所蔵　甲骨文字
●甲骨世界（北京国家図書館）
●敦煌遺珍（北京国家図書館）
●西夏碎金（北京国家図書館）
◆その他デジタルデータ
●年画擷英（北京国家図書館）
●碑帖菁華（北京国家図書館）
●京都大学人文科学研究所所蔵石刻拓本資料
●東京大学東洋文化研究所所蔵アジア写真資料集成データベース
◆地理を調べる
●中国行政区画網
●【宋人與宋詩】地理資訊系統（元智大学）
●Google マップ検索
◆サーチエンジンを利用した典拠調べ

### ⊙ デジタル情報の整理

本章では、調べた後の情報をどのように二次利用するかについて解説している。

単に他のアプリケーションにコピー＆ペーストするだけではなく、Evernote などのクリッピングサービスを利用した独自の情報蓄積の方法についても採り上げる。また、収集した情報を二次利用する場合の注意点についても、著作権法の紹介などを通じて採り上げている。

■デジタル情報の二次利用
◆Word に貼り付ける場合の注意事項
◆テキストファイルに貼り付ける
◆Excel でまとめる
◆Web リソースを利用
●Evernote でクリッピング
●Google ドキュメントでクリッピング
◆情報を保存する
●DropBox
●SkyDrive
◆利用しやすいデジタルデータとは

■デジタルデータと著作権

### ⊙ ソフトウェアを使いこなせ

ここでは、ローカルな環境にインストールして使うソフトウェアについて紹介する。

それらのソフトウェアの中には、数十万円から百万円を超える高価なものもあり、個人レベルではおいそれとは手を出しかねる物もある。

しかし、それらの多くはこの分野でよく使われる文献をデジタル化した物であり、その利用価値は計り知れない。そのため、その存在を知ってもらうためにも採り上げることにした。

■大規模叢書を利用してみる
◆『四部叢刊』・『四庫全書』
●Windows7 へのインストール（Windows Xp モード , Vitual Box を利用）
●Web 版の利用
◆愛如生製品を利用する
●中國基本古籍庫
●その他（愛如生数字古典・愛如生大型数拠庫・愛如生数字叢書・愛如生系列数拠庫）

●拇指数拠庫を使ってみよう

■新聞系データベース

◆人民日報図文電子版
◆光明日報図文電子版
◆文物報

■その他の調べ物

◆金文データベース
◆全宋詩分析系統（北京大学）

## ◇ おわりに

以上、筆者の担当したコンセプトと、それにもとづいた目次案について紹介した。

筆者は中国古代史を研究の中心分野としており、どうしても現代中国や語学系に弱い。その辺りは、読者諸賢がそれぞれご存じのものを紹介していただきたい。また、本誌（『漢字文献情報処理研究』）の既存刊にも数多く紹介されているので、それらを併せて参考にしていただきたい。


## 漢字文献情報処理研究会　入会のご案内

漢字文献情報処理研究会（略称：JAET）は、下記の活動目的に賛同し、大学院生以上の研究者、教育者、もしくは本会と関連する業務・活動に携わる社会人であれば入会することができます。

**⊙東洋学（日本・中国・韓国など）分野におけるコンピュータ利用方法の研究・紹介および関連情報の交換**
**⊙研究・教育現場でのコンピュータ活用・普及の促進**
**⊙関連諸分野の人材交流**
**⊙海外における同種の学会、プロジェクトとの積極的な交流・協同活動**

会員には

**⊙一般会員（BBS 利用＋『漢情研』購読）：年会費 3000 円**
**⊙BBS 会員（BBS 利用のみ）：年会費 1000 円**

があり、どちらか一方を選択できます。『漢字文献情報処理研究』を定期購読されるならば一般会員が便利です。

◘ 入会は下記 URL から手続きできます
**http://www.jaet.gr.jp/guiding.html**

# 特集2

# 陳腐化するデジタル資料

本誌ではこれまで人文科学研究・教育におけるデジタル情報の活用や、学術情報のデジタル化についての理論的研究から方法論、技術論まで幅広く取り扱ってきた。そこでは、デジタル情報の作成と作成されたデータの利用に主眼がおかれてきたといっても過言ではない。しかしデジタルデータもモノだということを考えれば、この世に「永遠の命」が無いように、いつの日かデータが利用できなくなることは容易に想像できよう。

本特集はこのデジタルデータの終焉について取り扱うものである。これまで本誌が作成・利用について紙幅を割いてきたことからすれば、違和感のある読者もおられるかもしれない。しかし学術資料を永く保存し伝えていくためには、やはりデータの終わりについても無視することはできないし、最低限の保存について考えるのは作成者の責務でもある。

本特集では、デジタルデータの終焉を劣化と陳腐化に分け、なかでもデジタルデータに特徴的な後者について紙幅を割いてみたい。デジタルデータの陳腐化とは、デジタルデータじたいではなく OS、アプリケーション、データ形式などデータを取り巻く諸環境の変化によって生ずるデジタルデータの終焉のことを指す。

全体構成としては、まず劣化と陳腐化について全体を俯瞰し、次に後者の諸例についての研究を配した。読者諸氏の御批正を乞う次第である。

**Contents**


| | | |
|---|---|---|
| デジタル情報の劣化と陳腐化・化石化 | 小島　浩之 | 72 |
| 失われた文字コード | 安岡　孝一 | 76 |
| レガシーとの付き合い方　東洋学文献類目の場合 | 守岡　知彦 | 82 |
| 永続性のある歴史資料デジタル・アーカイブへの試論 ――「アーカイブズ」への接近とデジタル応用の可能性―― | 後藤　　真 | 96 |

# デジタル情報の劣化と陳腐化・化石化

小島　浩之（こじま　ひろゆき）

## ■ はじめに

本特集では、様々な理由から使えなくなった、もしくは使えなくなる可能性の高いデジタルデータの問題を取り上げる。いわばデジタルデータのレッドデータブックに載せる、絶滅種と絶滅危惧種に関する諸問題を論ずるものである。

デジタルデータが読み取れない理由は大きく分けて二つある。第一には劣化や破損など物理的にデータが破壊される場合である。ただし、これは正確に言えばデジタルメディアの問題、すなわちデジタルデータを記録する媒体の劣化や破損であって、デジタルデータは原則として劣化も破損もしない。ただし、後述のように、デジタルデータにおいて、媒体と記録材は不可分であって、媒体の劣化はデジタルデータの劣化と見なされる。

第二には、データじたいもしくは再生装置等の改良や更新によって、記録されたデータが再生できなくなる場合がある。一般にはこれをデータやフォーマットの陳腐化などと表現するが、データがあっても使えないという意味では化石化とも言えるだろう。

具体例は枚挙に暇がないが、データフォーマットの陳腐化では、使えない文字コード問題（後掲の安岡論文を参照されたい）や、ワープロ専用ソフトで作成されたデータの問題などが、再生装置の陳腐化では、レーザーディスク(LD)プレーヤー、フロッピーディスク(FD) [1] ドライブ、などがすぐに思い浮かぶだろう。

またメーカーが対応を打ち切ったなどの理由から、特定世代の OS に依存したアプリケーションやプラグインを必要とするデータは、データの陳腐化、再生機器の陳腐化の双方を兼ね備えているといえる。

このように、デジタルデータが読み取れない原因は、実はデータそのものの劣化ではなく、媒体や再生装置、フォーマットといったデジタルデータを取り巻く諸環境に求められねばなるまい。

## ■ 裸眼可読と機械可読

デジタル情報の劣化、陳腐化・化石化は、デジタルデータの機械可読という特性が大きく影響を与えている。

一般に記録は裸眼可読であるものと、機械可読であるものに二分される。裸眼可読とは、再生装置を使わずに肉眼で読み取れる記録のことで、紙やフィルムへの記録はその代表格だろう。一方の機械可読とはその名称の通り、再生装置を媒介しなければヒトが認識できない記録のことで、デジタルデータはその最たるものと言える [2]。また、再生装置は、指定の初期化形式（フォーマット）に準拠した媒体に記録されたデータだけが再生可能である。つまりヒトがデジタルデータを認識するには、再生装置やデータフォーマットに大きく

依存せねばならいことが解るだろう。換言すれば、デジタルデータは、再生装置やフォーマットの存在が担保されてはじめて、一人前の記録として存在し得るのである。したがって、再生装置やフォーマットの陳腐化・化石化は、モノとしての劣化以上に深刻な問題となっている。

機械可読の記録では、媒体と記録材料が不可分となる。一般に記録は記録材料と媒体つまり、書くモノと書かれるモノから成り立っている。墨と紙、インクと羊皮紙、トナーとコピー用紙というようにである。多くのアナログ媒体において、両者は肉眼で識別できる。劣化ひとつを見ても、媒体の劣化と記録材料の劣化があり、当然その対処法も異なっている。例えば、紙の記録の劣化には、酸性紙という媒体の問題に起因する劣化現象もあれば、インク焼といって、インクに含まれる鉄分が酸化して紙を腐食させる記録材料による劣化もある。

これに対してデジタルデータの場合は、媒体のみ劣化するが、記録媒体と記録材料を肉眼で識別することは不可能である。この意味で、媒体が劣化しても複製を作成して記録内容、すなわち記録材料の部分だけを残せる紙の記録とは決定的に異なっている。デジタルデータでは媒体が劣化すれば、記録されたデジタルデータだけを取り出すことは技術的困難を伴うからである。つまりデジタルデータにおいて記録媒体と記録材料は不可分の関係であって、媒体の劣化はイコール、デジタルデータの劣化だと言っても過言ではない。

## ■ 劣化、陳腐化・化石化の諸問題

さて、ここからは、劣化や陳腐化・化石化に関するいくつかの興味深い例を挙げてみたい。

### ❇ 媒体の劣化に関する話題

現代で最も身近な記録媒体の一つに、CD や DVD などの光ディスクがある。光ディスクには再生専用形（CD-ROM, DVD-ROM）、追記形（CD-R, DVD-R）、書換形（CD-RW, DVD-RW）などがある。光ディスクはトラック上のピット（くぼみ）にレーザー光を反射させ、その変化を読み取って記録を再生している。このうち、ROM は金型からプレスによって物理的なピットを作ったものであるのに対し、R や RW は記録層に化学変化を起こさせ、いわば擬似ピットとして認識させている。このため物理的にデータの刻み込まれた ROM は最も保存性に優れている。ただし ROM の作成は金型の製作から入るため、専門業者に依頼せねばならず作成料金も高価なものとなる。

一方、R や RW はパソコンや専用レコーダーがあれば誰でも安価で作成できる。しかしこれらの長期保存については、温度 10℃～ 25℃、相対湿度 40 ～ 60％が推奨され、一般的なオフィス環境で保管する場合は、定期的なデータ移行（マイグレーション）がなければ長期保存は難しいと言う [3]。さらに、この寿命は保存環境だけでなく、媒体、記録装置、記録速度の組み合わせに大きく依存することが解っている [4]。つまり我々が日々使用している CD-R や DVD-R は大変不安定な媒体であって、温湿度管理された保存庫を持たない限り、一定年数でマイグレーションをしなければ、長期保存は保証されない [5]。

ここまでは一般に言われていることだが、実はこの ROM なのか R もしくは RW なのかということに大きな落とし穴がある。まず、インディーズものの CD には自作しているものも多く、この場合は CD-ROM ではなく CD-R である可能性が高い。しかし一見しただけでは両者の区別はすぐにつかない [6]。

また DVD については、市販されているデータのほとんどに DVD-ROM である旨のロゴが印刷されているが、実際は DVD-R であることも多いという。これはロゴの使い分けが、先述したデータ書き込み方法の相違に拠らず、購買時点でのデータ追記の可否からなされていることによる。確かに R であってもデータが既に書き込まれていれば追記はできないから、ROM（Read Only Memory）だとされても嘘ではない。

しかしながら、先に述べたように R の場合は、ROM に比べて媒体の劣化が早く進む。つまり購入した CD や DVD が ROM だと思っていたら、実

はRで、ある日突然データが再生できないなどということが起こる可能性がある。こういったROMの定義における業界のダブルスタンダードは情報の保存にとっては迷惑千万というほかない。

### ❊ 陳腐化・化石化に関する話題

マイクロフィルムは裸眼可読ではあるが、読み込むには拡大する必要があり、機械可読と裸眼可読の中間的な性質を持つ。リーダープリンターと呼ばれる拡大機器で閲覧し、必要に応じてプリントアウトするという使い方が一般的だろう。

ところが近年、フィルムの読み込みに小型CCDを使い、フィルム画像をデジタルデータに変換して直接パソコンで読み込めるデジタルリーダーが出現した。これは大変便利なものでアナログデータをデジタル変換して加工の便利度や自由度を高めただけでなく、従来のリーダープリンターより安価となっている。

当然、パソコン接続での利用ということは、アプリケーションがOS依存ということになる。なかでも多くのシェアを占めるMT社は先頃、最新バージョンの機器を発売したが、一つ前のバージョンについて、Windows7への対応を保証しないと宣言する。安価といっても百万以上するものである。発売されて僅か数年の製品がこの有様とはいかがなものであろう。アナログの再生装置であれば、機械が動き部品があるまでは保証されていたものが、デジタル再生装置に切り替えたがために、OSの寿命に左右されるという皮肉な結果が生じてしまったのである。

さらに、筆者の関わった某プロジェクトでは次のようなこともあった。MS社という業者の言うがまま、業者が独自開発した検索ソフトを入れた（そもそもこのソフトは使いづらく、性能もよろしくないため筆者らは最初から導入を反対したが）。これも最近になって、Windows7対応はしないとあっさり切り捨てられることとなった。この間、導入から対応放棄まで5年とたっていない。これをOS依存するものだから仕方ないと言い切れるだろうか。OSのバージョンアップが、業者側の対応放棄にとって都合良い言い訳になっていると感じるのは筆者だけだろうか。

また、この10年ほどの間に開発されたブラウザのプラグインで、OSやブラウザのバージョンアップに伴い対応できなくなってきているものがある。図書館や資料館などは、大量の画像情報発信のために、一時期競って様々な画像フォーマットのプラグインを導入している。当時ライセンスという形で利を得た業者が、新しいOSに対応するプラグインの提供を拒んだり、開発費用を求めたり等で、今後化石化するデジタルデータは増えると予想される。

## ■ おわりに

本特集の前振りとして雑多な議論を繰り返してきたが、最後に劣化や陳腐化・化石化にどのように向き合えばよいか私見を述べて擱筆することとしよう。

この世に永遠なるものは存在せず、形あるものはいつか朽ち果てる。これはデジタルデータであっても例外ではない。紙の記録では、写本や印刷で形を新たにして伝える方法をとってきた。これと同様に、デジタルデータも、複製を作成したり、媒体の寿命が来る前にマイグレーションやデータ変換を行うのは理に適っている。後世まで保存したいと思う情報は、媒体の種類を問わず手間暇かけて対策をとらねばならないのである。逆に言えば、紙の記録であっても保存状態を顧みず放置していたなら、後世まで残ることはない。要は、残すべきか否かはヒトが決めることであって、その意志決定ができなければ、モノは決して残らないということである。

陳腐化や化石化についても、実はデジタルデータに特有というわけではない。例えば中国の本の装幀ということを考えても、巻軸装→経折装→旋風装→胡蝶装→線装という変化をたどる。変化の時間的な流れは緩やかではあるが、進化の陰にはそれ以前の型式の陳腐化・化石化がある。ただし、現在の陳腐化・化石化は進化の陰というのではなく、業者の利益誘導による部分もあり、この点は遺憾である。

歴史的に見れば、人類は様々な場面で劣化や陳腐化・化石化に突き当たり、それを乗り越えてきている。本特集がこの前例に倣い、今後のデジタル情報の劣化や陳腐化・化石化を乗り越える起爆剤の一つになればありがたいと思う。

## 注

[1] 8インチや5インチのものは言うに及ばず、最近では3.5インチFDドライブも化石化しつつあると言えよう。

[2] 機械可読にはカセットテープ、ビデオテープといったアナログ記録も含まれる。この意味で、再生装置やフォーマットの陳腐化・化石化はデジタルデータに特有のことではなく、むしろ機械可読の記録に特有のものだと言える。

[3] JIS Z6017: 2006「電子化文書の長期保存方法」附属書2(規定)「主なCD・DVDディスクによる電子化文書の長期保存方法」

[4] 森島英行「DVD(CD)の劣化について」上・下『月刊IM』47(2), 47(3), 2008

[5] こういったデジタルメディアの劣化問題については、様々な論考があるが、最近の総論としては大島茂樹「デジタル情報保存のリスク：記録メディアの劣化・陳腐化とファイルフォーマットの陳腐化」『情報の科学と技術』60(2)をまずは参照されたい。

[6] CD(DVD)−ROMなのか−Rなのかの区別は、光ディスク裏面を丁寧に観察することで解る場合も多い。−ROMが裏面の色変化が無いのに対し、記録層に化学変化を起こさせる−Rは、データ記録された部分の色が変化するため縞模様に見える。

# 失われた文字コード

安岡　孝一（やすおか　こういち）

## ■ はじめに

「失われた文字コード」とタイトルに付けたものの、実は、文字コードそのものが失われているわけではない。失われているのは、ある文字コードを読むための環境であって、その文字コードで書かれたファイルそれ自体は、古い CD-R や磁気ディスクやその他の記憶媒体の中に、ひっそりとしかし厳然と存在していたりする。そのような「失われた文字コード」で書かれたファイルが、何かのはずみで我々の前に姿を現す。実際、読めない文字コードで書かれたファイルほど、始末の悪いものは無い。しかも、完全に読めないならまだしも、平仮名や片仮名など一部が読めるのに他が文字化けしていたりすると、何としても残りも読みたくなってしまうのが人のサガである。

本稿では、そのような「失われた文字コード」のうち、本誌の読者の手元に残っていそうな文字コードとして、特に「シフト GB」と「文字鏡コード」を取り上げる。また、これらに加えて「住基コード」に対しても、警鐘を鳴らしておくことにする。

## ■ シフト GB

GB 2312[1] に収録された簡体字を、日本語版 MS-DOS や日本語版 Windows 3.x で使うために開発された文字コードである。本来 GB 2312 は、94 × 94 の文字表を A1A1 ～ FEFE に割り当てて用いるが、日本語版 MS-DOS では FD や FE のコードは使えなかった。そこで、JIS X 0208 をシフト JIS に変換するやり方をまねて、GB 2312 の 94 × 94 の文字表を 47 × 188 に折り曲げ、1 バイト目に 81 ～ 9F・E0 ～ EF を、2 バイト目に 40 ～ 7E・80 ～ FC を使うことで、いわばシフト JIS の上に GB 2312 を載っけてしまうのが、シフト GB の基本的アイデアである。

当然、同じ文字コード上にシフト GB とシフト JIS が重なることになり、同じ「889F」という文字コード（図 1・2 の 16-01 にあたる）であっても、シフト GB では「啊」を、シフト JIS では「亜」を、それぞれ意味することになる。たとえば

図 1　GB 2312 の 16 区（シフト GB の 889F ～ 88FC）

| 16区 | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 01-19 | 啊 | 阿 | 埃 | 挨 | 哎 | 唉 | 哀 | 皑 | 癌 | 蔼 | 矮 | 艾 | 碍 | 爱 | 隘 | 鞍 | 氨 | 安 | 俺 |
| 20-39 | 按 | 暗 | 岸 | 胺 | 案 | 肮 | 昂 | 盎 | 凹 | 敖 | 熬 | 翱 | 袄 | 傲 | 奥 | 懊 | 澳 | 芭 | 捌 | 扒 |
| 40-59 | 叭 | 吧 | 笆 | 八 | 疤 | 巴 | 拔 | 跋 | 靶 | 把 | 耙 | 坝 | 霸 | 罢 | 爸 | 白 | 柏 | 百 | 摆 | 佰 |
| 60-79 | 败 | 拜 | 稗 | 斑 | 班 | 搬 | 扳 | 般 | 颁 | 板 | 版 | 扮 | 拌 | 伴 | 瓣 | 半 | 办 | 绊 | 邦 | 帮 |
| 80-94 | 梆 | 榜 | 膀 | 绑 | 棒 | 磅 | 蚌 | 镑 | 傍 | 谤 | 苞 | 胞 | 包 | 褒 | 剥 | | | | | |

図 2　JIS X 0208 の 16 区（シフト JIS の 889F ～ 88FC）

| 16区 | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 01-19 | 亜 | 唖 | 娃 | 阿 | 哀 | 愛 | 挨 | 姶 | 逢 | 葵 | 茜 | 穐 | 悪 | 握 | 渥 | 旭 | 葦 | 芦 | 鯵 |
| 20-39 | 梓 | 圧 | 斡 | 扱 | 宛 | 姐 | 虻 | 飴 | 絢 | 綾 | 鮎 | 或 | 粟 | 袷 | 安 | 庵 | 按 | 暗 | 案 | 闇 |
| 40-59 | 鞍 | 杏 | 以 | 伊 | 位 | 依 | 偉 | 囲 | 夷 | 委 | 威 | 尉 | 惟 | 意 | 慰 | 易 | 椅 | 為 | 畏 | 異 |
| 60-79 | 移 | 維 | 緯 | 胃 | 萎 | 衣 | 謂 | 違 | 遺 | 医 | 井 | 亥 | 域 | 育 | 郁 | 磯 | 一 | 壱 | 溢 | 逸 |
| 80-94 | 稲 | 茨 | 芋 | 鰯 | 允 | 印 | 咽 | 員 | 因 | 姻 | 引 | 飲 | 淫 | 胤 | 蔭 | | | | | |

「白雪皚皚」という文字列は、シフト GB では「88D5 9948 88A6 88A6」という文字コードになるが、これをシフト JIS だと思って表示すると「易僣始始」となってしまうわけだ。逆に言えば、「易僣始始」とシフト JIS で書かれたファイルを、シフト GB のフォントで表示すれば、見事「白雪皚皚」になるという仕掛けだった。これにより、フォント切り換えの可能なワープロソフト上では、日本語と中国語を混在して表示できたのである。

### ✲ Windows 98 以降のシフト GB

ところが Microsoft は、Windows 98 において全てのフォントを Unicode 化することを決定、同時に MS-Word の内部コードを Unicode 化したため、それまでのシフト GB は使えなくなってしまった。シフト JIS の文字コードのうち、CP932[2] において Unicode との対応を規定されている部分しか、使えなくなってしまったのである。たとえば、CP932 は 9873 ～ 989E（図 3 の 47-52 ～ 47-94）に、文字を収録していない。一方シフト GB は、この部分に「显」～「晓」の 43 字（図 4 の 47-52 ～ 47-94）を収録していた。この結果、Windows 98

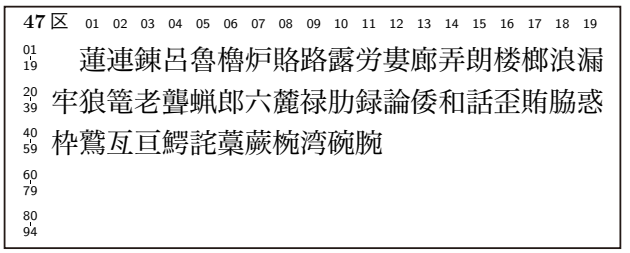



図 3　JIS X 0208 の 47 区（シフト JIS の 9840 ～ 987E・9880 ～ 989E）

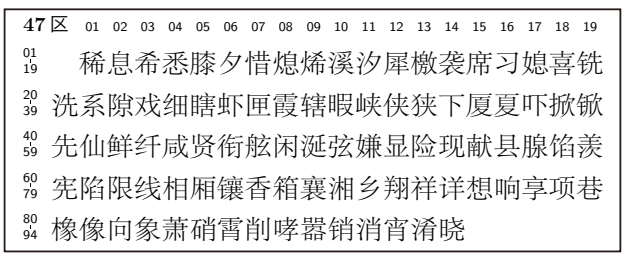



図 4　GB 2312 の 47 区（シフト GB の 9840 ～ 987E・9880 ～ 989E）

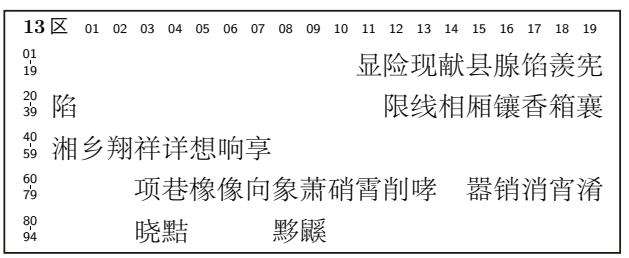



図 5　Unicode 対応シフト GB の 13 区（8740 ～ 877E・8780 ～ 879E）

表 1　シフト GB の Unicode 対応

| オリジナルのシフト GB | | Unicode 対応シフト GB | | |
|---|---|---|---|---|
| GB 2312 | シフト GB | 移動先 | NEC 拡張文字版シフト GB | IBM 拡張文字版シフト GB |
| 47-52 ～ 47-61 | 9873 ～ 987C | 13-11 ～ 13-20 | 874A ～ 8753 | 874A ～ 8753 |
| 47-62 ～ 47-77 | 987D ～ 987E・9880 ～ 988D | 13-32 ～ 13-47 | 875F ～ 876E | 875F ～ 876E |
| 47-78 ～ 47-80 | 988E ～ 9890 | 13-63 ～ 13-65 | 877E・8780 ～ 8781 | 877E・8780 ～ 8781 |
| 47-81 | 9891 | 13-66 | 8782 | FA59 |
| 47-82 | 9892 | 13-67 | 8783 | 8783 |
| 47-83 | 9893 | 13-68 | 8784 | FA5A |
| 47-84 ～ 47-88 | 9894 ～ 9898 | 13-69 ～ 13-73 | 8785 ～ 8789 | 8785 ～ 8789 |
| 47-89 ～ 47-93 | 9899 ～ 989D | 13-75 ～ 13-79 | 878B ～ 878F | 878B ～ 878F |
| 47-94 | 989E | 13-83 | 8793 | 8793 |
| 84-01 ～ 84-94 | EA9F ～ EAFC | 89-01 ～ 89-94 | ED40 ～ ED7E・ED80 ～ ED9E | FA5C ～ FA7E・FA80 ～ FABA |
| 85-01 ～ 85-94 | EB40 ～ EB7E・EB80 ～ EB9E | 90-01 ～ 90-94 | ED9F ～ EDFC | FABB ～ FAFC・FB40 ～ FB5B |
| 86-01 ～ 86-94 | EB9F ～ EBFC | 91-01 ～ 91-94 | EE40 ～ EE7E・EE80 ～ EE9E | FB5C ～ FB7E・FB80 ～ FBBA |
| 87-01 ～ 87-78 | EC40 ～ EC7E・EC80 ～ EC8E | 92-01 ～ 92-78 | EE9F ～ EEEC | FBBB ～ FBFC・FC40 ～ FC4B |
| 87-79 | EC8F | 13-84 | 8794 | 8794 |
| 87-80 | EC90 | 13-88 | 8798 | 8798 |
| 87-81 ～ 87-90 | EC91 ～ EC9A | 92-81 ～ 92-90 | EEEF ～ EEF8 | FA40 ～ FA49 |
| 87-91 | EC9B | 13-89 | 8799 | 8799 |
| 87-92 ～ 87-94 | EC9C ～ EC9E | 92-92 ～ 92-94 | EEFA ～ EEFC | FA55 ～ FA57 |

や Windows 2000 では、シフト GB の「㬎」～「暁」を表示できなくなってしまった。

この問題を回避するため、シフト GB を使用していた各社は、これらの文字を 13 区に移動させることにした（図 5）。CP932 の 13 区にあたる部分（8740 ～ 877E・8780 ～ 879E）は、NEC 特殊文字と呼ばれる独自拡張文字を収録していたが、この部分をこれまでのシフト GB は使っていなかったからだ。また、GB 2312 の 84 区以降に関しても、NEC 拡張文字のある 89 区以降に移動させた。移動の詳細を表 1 に示す。なお、NEC 拡張文字（ED40 ～ EEFC）に対しては、全く同じ文字が多少順序を変えて IBM 拡張文字（FA40 ～ FC4B）にも収録されており、ファイル入出力の際に相互に自動変換される。このような自動変換がおこなわれた場合の文字コードを、表 1 に「IBM 拡張文字版シフト GB」として示しておく。

### ❊ シフト GB のサルベージ

前世紀末には一世を風靡したシフト GB も、今世紀に入ってからは衰退の一途をたどっている。開発の一翼を担ったオムロンも、2010 年 5 月で『楽々中国語』（cWnn）のサポートを完全に打ち切ってしまった。現時点でシフト GB をサポートしている製品は、わずかに高電社の『ChineseWriter』のみ [3] となってしまっている。では、何かの拍子にシフト GB のファイルを発掘してしまった場合、『ChineseWriter』を購入して読むしかないのだろうか。

実は、シフト GB のサルベージに関しては、針谷壮一氏が公開している「中国語コンバータ [4]」というフリーウェアが、現時点において最も強力なツールである。各社固有のシフト GB テキストを GB 2312 の A1A1 ～ FEFE に変換してくれると同時に、シフト GB・シフト JIS 混在の rtf 形式も変換可能というスグレモノだ。すなわち、doc 形式のシフト GB が発掘された場合でも、MS-Word の「リッチテキスト形式」でそのまま保存しなおすことができれば、「中国語コンバータ」によって、フォント名も含め変換可能 [5] となっている。

ただし、サルベージ変換後のテキストや rtf は、現在は読み書きが出来ても、いずれ遠い将来には、また「失われた文字コード」となりかねない。これらのテキストや rtf は、プリントアウトして紙の形にしておくか、せめてフォント埋め込みの PDF にしておく必要があるだろう。

## ■ 文字鏡コード

『今昔文字鏡』で使われている 10 進 6 桁の文字鏡番号に対し、000001 ～ 005640、005641 ～ 011280、011281 ～ 016920、……、と 5640 個ずつに区切り、各 5640 個を JIS X 0208 の 16 ～ 45・48 ～ 77 区に割り当てた文字コードである。文字コード上は完全にダブった形となるが、000001 ～ 005640 は「Mojikyo M101」フォントに、005641 ～ 011280 は「Mojikyo M102」フォントに、011281 ～ 016920 は「Mojikyo M103」フォントに、……、それぞれ収録することにより、フォント切り換えによって全ての文字を扱うことができる、というのがウリであった。逆に言えば、文字鏡フォントが実装されていない環境下では、文字鏡コードで書かれた「一」「声」「鴪」……は、全て「亜」に文字化けすることになる（図 6 ～ 8 および図 2）。

文字鏡番号の 000001 ～ 049964 は基本的に『大漢和辞典』の検字番号を踏襲しているが、それ以降の番号は文字鏡研究会が独自に付けたものである。管理を担当する文字鏡フォントセンターは、2000 年 9 月の使用許諾書において、文字鏡番号と他の文字集合との対応表および、その対応表を用いて文字の変換をおこなうソフトウェア等の作成・配布を、明確に禁止している。

### ❊ 文字鏡コードの現況

文字鏡コードで書かれた MS-Word や Excel あるいは一太郎などのファイルは、文字鏡フォントが実装されていない環境では文字化けが起こるが、文字鏡フォントセンターは 2008 年 3 月以降、文字鏡フォントの無償ダウンロードを打ち切っている。文字鏡番号の付与をおこなってい

たはずの文字鏡研究会は、ほぼ同時期に活動を凍結し、そのWWWサイト[6]も「再構築中」の状態が続いている。また、現時点で文字鏡コードをサポートしている製品は、エーアイ・ネットの『今昔文字鏡』のみである。ただし、各文字鏡番号に対応する字形に関しては、ISO/IEC 10036のグリフ登録サイト[7]で閲覧可能だ。ISO/IEC 10036のグリフ番号10000001～10576000が文字鏡番号000001～576000に対応しており、128ドットのgif画像が入手できる。

このような状況において、文字鏡コードで書かれたファイルが発掘された場合、どうやってサルベージすればいいのか。残念ながら文字鏡フォントセンターが、そのようなサルベージをおこなうソフトウェアを明確に禁止してきたため、現時点では全く妙案がない。化けてしまった一字一字に対して、その字のJIS X 0208における区点番号を調べて、フォント名と共に文字鏡番号に変換[8]し、さらに10000000を加えて、ISO/IEC 10036のグリフ登録サイトからgifを入手[9]するしかない。もし、フォント名がわからない場合は、とりあえず「Mojikyo M101」の文字鏡番号を計算し、そのグリフ番号に順に5640を加えていって当該グリフを入手、というのを繰り返すことになる。こんなのを手作業でおこなうのは、全く馬鹿げているとしか言いようがないのだが、ソフトウェアを作成できない以上、どうしようもない。

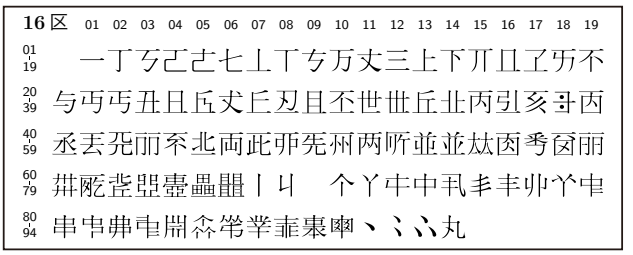



図6 「Mojikyo M101」の16区（文字鏡番号000001～000094）

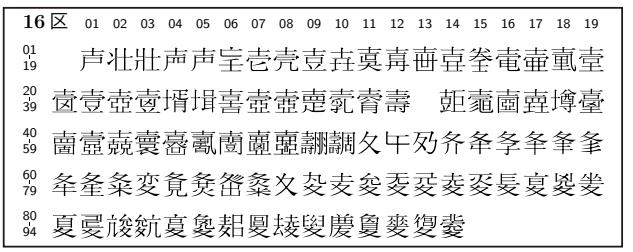



図7 「Mojikyo M102」の16区（文字鏡番号005641～005734）

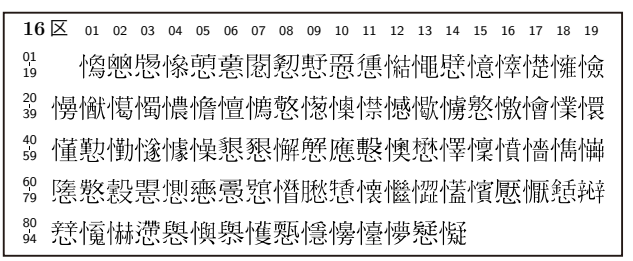



図8 「Mojikyo M103」の16区（文字鏡番号011281～011374）

## ■ 住基コード

住民基本台帳ネットワークにおける統一文字コードとして、2001年2月に検討版が作成され、2001年10月に制定、2002年8月より運用開始された文字コードである。Unicodeの基本多言語面（U+0000～U+FFFF）を独自拡張した2バイトの文字コードであり、0000～9FFFおよびE000～FFFFに関してはUnicodeと互換である（図9）。ただし、ハングルにあたる領域（AC00～D7A3）に関しては、ハングルではなく独自の拡張漢字や変体仮名などを、勝手に割り当てる形で設計された（図10）。これに加え制定版では、JIS X 0213の第3・4水準漢字のうち漏れていた漢字を、AAA1～ABCFに追加している（図11）。

住基コードの仕様は一般には非公開となっており、現時点で唯一のフォントである「KAJO J明朝」を購入して、文字コードを調べるしかない。ただし、2009年3月発行の『汎用電子情報交換環境整備プログラム成果報告書別冊』（日本規格協会）には、住基コードの漢字部分が収録されており、漢字に関しては一応の調べがつくようになった。

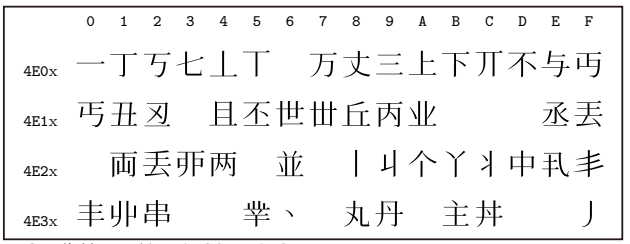


図 9　住基コードの 4E00 ～ 4E3F

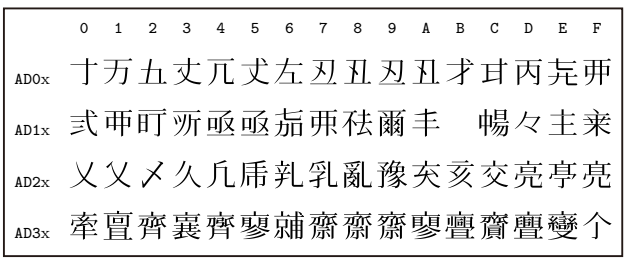


図 10 住基コードの AD00 ～ AD3F

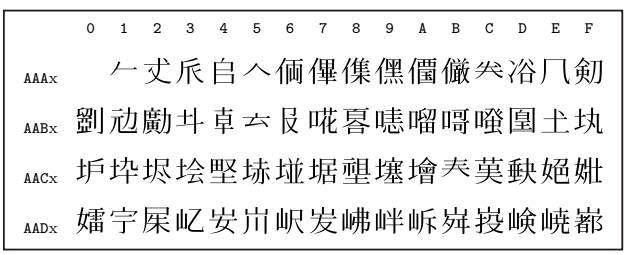


図 11 住基コードの AAA1 ～ AADF

### ❊ 住基コードの崩壊

2008 年 12 月 に Tai Viet[10] が U+AA80 ～ U+AADF に収録されたことから、住基コードは急速に崩壊が進んでいる。というのも、たとえば U+AABE は「TAI VIET VOWEL AM」という非前進文字なので、通常の Unicode 処理では直前の文字に重ねて表示される。この結果、住基コードの AABE にあたる「圭」は、直前の文字と重なって表示されてしまう [11] のである。

今後、Unicode の U+AAA1 ～ U+ABCF に新たな文字が追加されるたび、住基コードの崩壊が加速していくのは間違いない。本来、これらの独自拡張漢字は、U+20000 ～ U+2FFFF の拡張漢字面にちゃんと収録するか、さもなくば IVS[12] を使って見分けるべきなのだ。元々の設計が完全に誤っている上に、住基コードの仕様を公開しないという方針を堅持し続けているため、もはや誰からも救いの手を差しのべようがない、というのが現状である。

## ■ おわりに

シフト GB、文字鏡コード、住基コード、の 3 つの文字コードについて概要を述べ、これらの文字コードがどのようにして「失われていった」のかを解説した。これらの文字コードに共通する特徴は、特定のベンダやグループによってややもすれば閉鎖的に使われており、標準化からは程遠い立場にあったことである。言い換えれば、標準化をおこたった文字コードは、いずれ「失われる」ということだろう。

もし、読者諸氏が今現在、仕様が一般に公開されていない文字コードを使っているのなら、悪いことは言わない、さっさと別の文字コードに乗り換えるべきだ。閉じた仕様の文字コードは、いずれ誰にもサポートできなくなるのだから。その点を最後に指摘して、本稿をしめくくることにする。

### 注

[1] 中華人民共和国の国家標準局が 1980 年に発布した漢字コード。94 × 94 の文字表の中に、非漢字 682 字、第 1 級漢字 3755 字、第 2 級漢字 3008 字、あわせて 7445 字を収録していた。現在は GB 18030 に、とって代わられている。

[2] Microsoft の独自拡張シフト JIS。

[3] 高電社版のシフト GB である「CW コード」では、1 ～ 8 区の非漢字部分は GB 2312 ではなく JIS X 0208 である。すなわち、非漢字部分はシフト JIS で、漢字部分だけが Unicode 対応シフト GB だと言える。

[4] **http://www5b.biglobe.ne.jp/~harigaya/chcnv.html**

[5] 『ChineseWriter』 の「CW-GB Mincho」 や「GB 中 国

明朝」、『楽々中国語』の「OM 中国 GB 明朝」、あるいは『NiHao2』の「NH 簡体宋」などのフォントを、Microsoft Windows の「SimSun」に変換し、同時に文字コード変換をおこなう。ゴシックなども同様。

[6] http://www.mojikyo.org/

[7] http://10036ra.org/glyph-index.html

[8] 16 ～ 45 区の場合は、区番号から 16 を引いて 94 倍し、点番号を加える。48 ～ 77 区の場合は、区番号から 18 を引いて 94 倍し、点番号を加える。さらに、フォント名の番号部分（「Mojikyo Mxxx」の xxx の部分）から 101 を引いて 5640 倍した値を加えれば、文字鏡番号が得られる。

[9] グリフ番号 10050001 の gif が http://10036ra.org/retrieve.php?gid=10050001 で入手可能。

[10] 黒タイ語や白タイ語に用いられる文字。

[11] U+2123D に収録されている本物の「𠂤」を使えば、このような重なり表示の問題は発生しないが、住基コードでは U+10000 以降を使うことができない。

[12] Ideographic Variation Sequence。Unicode において複数の字体を見分けるために付加する文字コード。たとえば「丈」と「丈」は、それぞれ「U+4E08 U+E0100」と「U+4E08 U+E0101」という IVS で見分けられる。

# レガシーとの付き合い方

## 東洋学文献類目の場合

守岡　知彦（もりおか　ともひこ）

### ■ 1 はじめに

コンピューター上のデータはそれが作られ、運用される環境に多かれ少なかれ依存してしまう。賢明な本誌の読者の中には特定のシステムやソフトウェアにロックインされてしまう問題を避けるために、なるべく、プレイン・テキスト、CSV やタブ区切りの表データ、XML に基づく標準的な形式等を用いるようにしている方も少なくないかも知れない。少なくとも、プレイン・テキストであれば、一般的な OS であればテキストエディターで編集できるだろうし、Unicode が収録する文字数の増加とともに外字の必要性も減ってきており、プレイン・テキストの表現力は以前よりも向上しているといえる。また、HTML や TEI などの XML に基づくタグセットを用いることで、より表現力が高いテキストをマルチプラットフォームで利用することもできる。これらは Unicode のような標準的な文字符号とプレイン・テキストという最低限の共通基盤を前提に、その上に構築されたものといえる。

しかしながら、こうしたプレイン・テキストの恩恵は、汎用機上においては自明ではない。UNIX では、ASCII に基づくプレイン・テキストという共通形式を前提に、あらゆるデバイスやリソースを原則としてファイルという抽象によって扱うという『ファイル抽象』という仕組みを備えている。Linux や Mac OS X のような UNIX 系 OS だけでなく、Windows 等も含めて今日普及しているさまざまな OS は多かれ少なかれ UNIX の影響を受けている。インターネットも BSD 系 OS とともに成長してきた結果、UNIX の影響を多大に受けている。このような結果、ファイルやファイル名、プレイン・テキストというものは意識するまでもないような自然なものとなっているといえる。また、今日普及しているさまざまな文字符号の多くは ISO 646 に対する上位互換性を有しており、ISO 646 の共通部分内を使えばほぼ全てのシステムで情報交換できるようになっている。こうしたことはある意味において UNIX が我々にもたらした恩恵のひとつであるように思われる。言い替えれば、それ以前においてはこうした恩恵を得ることは出来なかった訳である。

東洋学文献類目の場合、編纂作業の電子化が始まったのは 1980 年代初頭であり、富士通製の汎用機が用いられたので、文字コードが EBCDIC と JEF 漢字コードであるというだけでなく、今日的なファイルシステム上のテキストデータというものとは思想的に全く異なるものとして電子化データの蓄積が始まった。また、類目システムは、入力対象を選定し指示するといったコンピューターの外での作業を含むワークフロー全体から成り立っており、また、この全体の構造は電子化以前の頃のワークフローの影響を受けていると考えられ、データ形式もまた電子化以前の紙のカードの

影響を受けている。コンピューター・システムはこうしたワークフローや電子化時代以前のものも含めたレガシーから自由ではないのである。

本稿では、類目システムでの経験を中心に、レガシーなシステムとの付き合い方について議論してみたい。

## 2「東洋學文獻類目」とは

「東洋學文獻類目」（以下では、「類目」と呼ぶことにする）は、東洋学研究に関する内外の文献について、その対象年の日付のある雑誌・論集所載の論文と単行本とを、それぞれ分類・排列し、著者索引を付けたもので、1935 年に東方文化学院京都研究所の事業として 1934 年度版「東洋史研究文獻類目」の編纂が始まって以来逐年編纂されている（1963 年度版から現在の名称を用いている）。

類目にはさまざまな言語で書かれた文献が収録されており、その電子化システムは日本語、中国語、朝鮮語、英語、フランス語、ドイツ語、ロシヤ語、その他欧州諸語、ベトナム語、タイ語、サンスクリットやアラビア語等のラテン転写、IPA 発音記号、甲骨文字、金文、楚系文字、および、それらの隷定文字等のさまざまな言語・用字系（script）をサポートする必要がある。

このことは Unicode が登場してさまざまな文字が符号化された今日においても必ずしも容易なことではないが [1]、類目の電子化が始まった 1980 年代初頭においては現在以上に難事業であったに違いない。

また、類目は電子化以降も今日に至るまで冊子体の出版を維持しており、類目の電子化システムは単なるデータベースではなく、冊子体の電子編集・製版システムとしての側面を持つこととなった。

## 3 システムとデータ形式の変遷

### ✲ 3.1 汎用機での電子化

類目の電子化システムの開発が始まったのは 1981 年のことである〔2〕。このシステムは京都大学大型計算機センター（当時）に設置された富士通製の汎用機 FACOM M382 上の検索システム FAIRS を使用して実装され〔4〕、1984 年 3 月にはこのシステムを用いた最初の電子製版による冊子体類目である 1981 年度版類目が発行された。以降、2003 年に発行された 2000 年度版類目までこのシステムを用いた電子編纂が行われた。

#### ◆ 3.1.1　文字表現

汎用機時代の類目データでは、文字コードとして EBCDIC / EBCDIK と JEF 漢字コードに独自の外字を追加したもの [2] を用いていた。

EBCDIC は英数字と記号からなる文字コードであり、ASCII とは互換性がない。EBCDIK は EBCDIC にカタカナを追加したものである。

一方、JEF 漢字コードは EUC に似た JIS X 0208 : 1978 の GR 表現 [3] であるが、2 バイト固定長であり、空領域に外字領域を設けている。また、基本的に、EUC-JP のようにビットパターンで EBCDIC / EBCDIK と JEF を切替えるようなことはせず [4]、データ構造の定義において、この欄（このバイトからこのバイトまで）は英数字（EBCDIC）、この欄はカタカナ（EBCDIK）、この欄は漢字（=jef-china3）ということを予め決めておくようになっている。逆にいえば、データの実体はどこに区切りがあるのかも判らないようなバイナリーデータとなっている訳である。

#### ◆ 3.1.2 データ形式

電子化された類目のデータ形式は、主に、電子化時代以前に用いられていた編纂用のカードと書誌情報の交換用形式であった UNIMARC の影響を受けていると考えられ、編纂用のカードを FAIRS 上に載せたような形式だと看做すことができる。

| | 行番号 | タグ | | サブフィールド | |
|---|---|---|---|---|---|
| 雑誌カード | 00003930 | 010 | | Ａ８２０３５５ | ID |
| | 00003940 | 020 | | ＄ｊ０４７２００ | 雑誌名コード |
| | 00003950 | | | ＄ｋｌ６ | 巻 |
| | 00003960 | | | ＄ｄｌ９８２（９） | 出版年（月） |
| 論文カード | 00003970 | 100 | 00001 | ＄ｔ漢字の「一字一音」に對する變則 | タイトル |
| | 00003980 | | | ＄ｕ中國語についての基本概念 | 副題 |
| | 00003990 | | | ＄ａ江英居 | 著者名 |
| | 00004000 | | | ＄ｚ２ | 著者名の排列種別 |
| | 00004010 | | | ＄ｐ６７－８０ | ページ |
| | 00004020 | | | ＄ｂｌ５２４ | 分類コード |
| | 00004030 | | | ＄ｃｌｌＡ０ＸＸｌ | 排列のための手がかり |
| 論文カード | 00004040 | 100 | 00002 | ＄ｔ漢魏の詠史詩 | |
| | 00004050 | | | ＄ｕその成立と發展 | |
| | 00004060 | | | ＄ａ市川 桃子 | |
| | 00004070 | | | ＄ｆイチカワ モモコ | 著者名のよみ |
| | 00004080 | | | ＄ｚｌ | |
| | 00004090 | | | ＄ｐｌ－ｌ８ | |
| | 00004100 | | | ＄ｍ（右） | ページの注記 |
| | 00004110 | | | ＄ｂｌ０３Ｘ | |
| | 00004120 | | | ＄ｃｌｌＢＩｌ０Ｅ：１２２ＥＸＸｌ | |

図 1：汎用機時代のマスターデータの例（雑誌＋論文）

各文献は、単行本や雑誌、論文、細目論文、書評論文などの種別に応じたカードで表される。各カードはカードの種類によって決まる幾つかのフィールドからなり、各フィールドは行番号、タグ、サブフィールドという３つの欄からなる。また、サブフィールドは、親カードにおける子番号（以下、単に『子番号』と呼ぶ）、$＋英字１文字からなるサブフィールド名、および、サブフィールド値（文字列）からなる（図 1）。なお、実際には、子番号はサブフィールド名 / 値とは別の欄になっている。そして、行番号欄、タグ欄、子番号欄は EBCDIC, サブフィールド名 / 値欄は JEF 漢字コード（=jef-china3）で符号化されている。また、サブフィールド値中の漢字はいわゆる『康熙体』に正規化されている。

雑誌–論文–細目論文や単行本–書評論文等の親子関係やリンクを表現するために、階層構造やリンク構造を持ったデータ形式となっており、雑誌カードの後には幾つかの論文カードが続き、論文カードの後に複数の細目論文カードを置くこともできる（図 2）。書評論文カードや批評論文カードでは対象となる文献に対するリンクを $l<ID> という形式で表現している（図 3）。

### ◆3.1.3 検索システム

汎用機上の類目の電子化システムはどちらかといえば、検索用のデータベースというよりも、冊子体の編集のためのシステムとしての性格が強かったといえる。このことを顕著に示すのが「排列のための手がかり」（$c）というサブフィールドである。これは文献の内容に関するメタデータの一種であり、地域コード、時代コード、事項コード、内容コードという 4 種類の情報を結合したものなのであるが、この $c の値でソートすると類目の冊子体の排列に近い結果になるよう工夫されている。この工夫の結果、コードの符号化が複雑化し、$c 内の各コードや分類コード（$b）との間で依存関係が発生したりして、編集や検索がしづらいものとなってしまっている。また、漢字をいわゆる『康熙体』に正規化していることも一般の利用者にとっては検索文字列が入力しづらく問題であったようである。

こうしたことから、類目の検索システムを実現するための試みが始まり、1987 年に〝CHINA3〟というデータベースが公開された。

〔4〕で述べられている「東洋学文献類目データベース」（RD）では、マスターデータの持つカー

| | 行番号 | タグ | | サブフィールド | |
|---|---|---|---|---|---|
| 雑誌カード | 00021070 | 010 | | Ａ８２０９１５ | ID |
| | 00021080 | 020 | | ＄ｊ１３２８００ | 雑誌名コード |
| | 00021090 | | | ＄ｋ２０ | 巻 |
| | 00021100 | | | ＄ｄ１９８２（３） | 出版年（月） |
| | 00021110 | 030 | | ＄ｎ創立二十周年記念特集號 | 雑誌・著者の注記 |
| 論文カード | 00021120 | 100 | 01000 | ＄ｔ佛教の體系と展相の研究（２） | タイトル |
| | 00021130 | | | ＄ａ武邑 尚邦 | 著者名 |
| | 00021140 | | | ＄ｆタケムラ ショウホウ | 著者名のよみ |
| | 00021150 | | | ＄ｚｌ | 著者名の排列種別 |
| | 00021160 | | | ＄ｐ１２８－１６１ | ページ |
| | 00021170 | | | ＄ｂ０７４３ | 分類コード |
| 細目論文カード | 00021180 | 110 | 01001 | ＄ｔ佛教の體系的同一性と歴史的多様性について | |
| | 00021190 | | | ＄ｕインドの場合 | 副題 |
| | 00021200 | | | ＄ａ武邑 尚邦 | |
| | 00021210 | | | ＄ｆタケムラ ショウホウ | |
| | 00021220 | | | ＄ｚｌ | |
| | 00021230 | | | ＄ｐ１２８－１３９ | |
| 細目論文カード | 00021240 | 110 | 01002 | ＄ｔ說一切有部における遍知（承前） | |
| | 00021250 | | | ＄ａ加藤 宏道 | |
| | 00021260 | | | ＄ｆカトウ ヒロミチ | |
| | 00021270 | | | ＄ｚｌ | |
| | 00021280 | | | ＄ｐ１３９－１５２ | |
| 細目論文カード | 00021290 | 110 | 01003 | ＄ｔ龍樹における知の問題（１） | |
| | 00021300 | | | ＄ｕ動詞ｊñā とその派生語の使用の檢討を通して（上） | |
| | 00021310 | | | ＄ａ田丸 俊昭 | |
| | 00021320 | | | ＄ｆタマル トシアキ | |
| | 00021330 | | | ＄ｚｌ | |
| | 00021340 | | | ＄ｐ１５３－１６１ | |

図２：汎用機時代のマスターデータの例（細目論文）

図３：汎用機時代のマスターデータの例（リンク）

| | 行番号 | タグ | | サブフィールド | |
|---|---|---|---|---|---|
| 雑誌カード | 00024590 | 010 | | Ａ８１０２４０ | ID |
| | 00024600 | 020 | | ＄ｊ２７７０００ | 雑誌名コード |
| | 00024610 | | | ＄ｋ５５ | 巻 |
| | 00024620 | | | ＄ｇ３ | 号 |
| | 00024630 | | | ＄ｄ１９８２（３） | 出版年（月） |
| 書評論文 | 00024640 | 200 | 00001 | ＄ｒ千葉 德爾 | 評者名 |
| | 00024650 | | | ＄ｆチバ トクジ | 日本語著者名のよみ |
| | 00024660 | | | ＄ｚｌ | 著者名の排列種別 |
| | 00024670 | | | ＄ｐ１９８－１９９ | ページ |
| | 00024680 | | | ＄ｌＢ８１０１１９ | リンク |
| 単行本カード | 00035090 | 510 | | Ｂ８１０１１９ | ID |
| | 00035100 | 520 | | ＄ｔ神話と傳說の旅 | タイトル |
| | 00035110 | | | ＄ａ川喜田 二郎／加藤 千代 | 著者名 |
| | 00035120 | | | ＄ｆカワキタ ジロウ／カトウ チヨ | |
| | 00035130 | | | ＄ｚ１１ | |
| | 00035140 | 540 | | ＄ｅ古今書院／／東京 | 出版者//出版地 |
| | 00035150 | | | ＄ｄ１９８１ | 出版年 |
| | 00035160 | | | ＄ｐ２４６ | |
| | 00035170 | | | ＄ｓネパール叢書 | シリーズ名 |
| | 00035180 | 550 | | ＄ｍ圖１７表２ | ページの注記 |
| | 00035190 | 560 | | ＄ｂ１４４Ｘ | 分類コード |
| | 00035200 | | | ＄ｃ４２Ａ０ＸＸ１ | 排列のための手がかり |

| | サブフィールド名 | サブフィールド値 |
|---|---|---|
| レコード | ADD A82091501001 | |
| | -T | 佛教の體系と展相の研究（2） |
| | -T | 佛教の體系的同一性と歴史的多様性について |
| | -T1 | 仏教の体系と展相の研究（2） |
| | -T1 | 仏教の体系的同一性と歴史的多様性について |
| | -T2 | (2) |
| | -U | インドの場合 |
| | -U1 | インドの場合 |
| | -A | 武邑 尚邦 |
| | -A | 武邑 尚邦 |
| | -A1 | 武邑 尚邦 |
| | -A1 | 武邑 尚邦 |
| | -J0 | 132800 |
| | -J | 龍谷大學佛教文化研究所紀要 |
| | -J1 | 竜谷大学仏教文化研究所紀要 |
| | -K | 20 |
| | -P | 128-161 |
| | -P | 128-139 |
| | -D | 1982(3) |
| | -N | 創立二十周年記念特集號 |
| | -B | 0743 |
| | -B2 | 教学 |
| | -F | ﾀｹﾑﾗ ｼｮｳﾎｳ |
| | -F | ﾀｹﾑﾗ ｼｮｳﾎｳ |
| | -Z | 1 |
| | -Z | 1 |
| | -KE | ﾌﾞｯｷｮｳ ﾀｲｹｲ ﾃﾝｿｳ ｹﾝｷｭｳ |
| | -KE | ﾌﾞｯｷｮｳ ﾀｲｹｲﾃｷ ﾄﾞｳｲｿｳｾｲ ﾚｷｼﾃｷ ﾀｲｳｾｲ ｲﾝﾄﾞﾊﾞｱｲ |
| レコード | ADD A82091501002 | |
| | -T | 佛教の體系と展相の研究（2） |
| | -T | 說一切有部における遍知（承前） |
| | -T1 | 仏教の体系と展相の研究（2） |
| | -T1 | 説一切有部における遍知（承前） |
| | -T2 | (2) |
| | -A | 武邑 尚邦 |
| | -A | 加藤 宏道 |
| | -A1 | 武邑 尚邦 |
| | -A1 | 加藤 宏道 |
| | -J0 | 132800 |
| | -J | 龍谷大學佛教文化研究所紀要 |
| | -J1 | 竜谷大学仏教文化研究所紀要 |
| | -K | 20 |
| | -P | 128-161 |
| | -P | 139-152 |
| | -D | 1982(3) |
| | -N | 創立二十周年記念特集號 |
| | -B | 0743 |
| | -B2 | 教学 |
| | -F | ﾀｹﾑﾗ ｼｮｳﾎｳ |
| | -F | ｶﾄｳ ﾋﾛﾐﾁ |
| | -Z | 1 |
| | -Z | 1 |
| | -KE | ﾌﾞｯｷｮｳ ﾀｲｹｲ ﾃﾝｿｳ ｹﾝｷｭｳ |
| | -KE | ｾﾂ ｲｯｻｲｺｳﾌﾞ ﾍﾝ ﾁ ｼｮｳｾﾞﾝ |
| レコード | ADD A82091501003 | |
| | -T | 佛教の體系と展相の研究（2） |
| | -T | 龍樹における知の問題（1） |
| | -T1 | 仏教の体系と展相の研究（2） |
| | -T1 | 竜樹における知の問題（1） |
| | -T2 | (2) |
| | -T2 | (1) |

図 4：CHINA3 における細目論文の例

ド間の階層構造を維持しつつも、「排列のための手がかり」（$c）を検索しやすい形に展開した情報を追加するとともに、タイトルや著者名等に対して『康熙体』に正規化したフィールドとは別に常用漢字に正規化したフィールドを追加することで一般利用者が検索しやすいように工夫している。

しかしながら、階層構造を維持したままで検索システムを実現するには複数回検索を繰り返す必要があり、計算時間、あるいは、課金コストなどの点で不利であったようで、CHINA3 では階層構造を展開して、フラットな 1 枚の表の形にしたデータベースをマスターデータとは別に構築

することで、検索を容易にしていた。しかしながら、この結果、冊子体編纂用のマスターデータとCHINA3のデータベースという2つのデータセットに分岐してしまうこととなった。ここで、CHINA3と区別するために、冊子体編纂用のマスターデータを『旧RUIMOKU』と呼ぶことにする。また、後のWWW版のCHINA3と区別するために、この汎用機上のCHINA3を『旧CHINA3』と呼ぶことにする。

前述のように、CHINA3では階層構造を展開しており、細目論文の親論文のレコードはなくなり、その情報は各細目論文のレコードに入る（例えば、図2の場合、図4のようになる）。この結果、細目論文のレコードにその細目論文のタイトルや著者名と親論文のタイトルや著者名が入ることとなり、親論文の著者が細目論文も書いたように見えたり、親論文のタイトルと細目論文のタイトルの両方がタイトルとして見えたりすることになるなどの問題も生じることとなった。

#### ◆3.1.4　端末の変遷

当初は、汎用機用の専用端末があり、外字（=jef-china3）も文字として表示することができ、また、必要に応じて外字を作成することができていたが、後に、パソコン上の端末エミュレーターを用いるようになり、端末上で外字を見たり新たな外字を作成することができなくなったらしい。このため、冊子体の作成時に新たな文字が必要になった場合、電子製版した版下に文字を手で張り付けるという運用をしていたらしい。当然、データベース上にはその文字の情報は存在しないことになってしまう。また、当初はJEF漢字コードのJIS X 0208部分は1978年版であったのだが、Windows上の端末エミュレーターでは1990年版ベースになっていたようで、字体化けが生じたり、文字の運用に混乱が生じていたようである[5]。

### ❖3.2 現行の類目

旧RUIMOKU及び旧CHINA3が稼働していた富士通製の汎用機が利用できなくなったため、冊子体の編纂及び検索サービスをUNIXサーバーやパソコン上で再構築することになった。

この時、まず、旧CHINA3をベースにCHINA3のWWW上の検索サービスを開発し公開したのだが、この時点では、著者は旧CHINA3のことしか理解しておらず旧RUIMOKUの存在や、旧RUIMOKUと旧CHINA3の関係を理解していなかった[6][7]。このために、今思えば、いろんな失敗や遠回りをすることになるのだが、それとは別に、環境や要求事項の変化等によって、幾つか試

図5：現行類目のマスターデータの例（初期型）

| | | |
|---|---|---|
| 雑誌レコード | A006001 | ID |
| | $j200800 | 雑誌名コード |
| | $g1 | 号 |
| | $d2001 | 出版年 |
| 論文レコード | 10000001 | タグ+子番号 |
| | $t大寨奇遇记 | タイトル |
| | $a马烽 | 著者名 |
| | $z2 | 著者名の排列種別 |
| | $p84-93 | ページ |
| | $b1076 | 分類コード |
| | $q创业, 山花 | 内容の注記 |
| | $n(ma feng) | 著者名のピンイン表記 |
| 論文レコード | 10000002 | タグ+子番号 |
| | $t一九三五年一月的鲁迅 | |
| | $a李书磊 | |
| | $z2 | |
| | $p99-103 | |
| | $b1072 | |
| | $xlu xun | 学者名 |
| | $y4 | 学者名の排列種別 |
| | $x$鲁迅 | 学者名の漢字表記 |
| | $n(li shu lei) | |

| | | |
|---|---|---|
| 雑誌 | A2007-00471 | ID |
| レコード | $j040200 | 雑誌名コード |
| | $k59 | 巻 |
| | $g2 | 号 |
| | $d2007(9) | 出版年（月） |
| | $i通巻第569号 | 巻号の注記 |
| 論文 | 10000001 | タグ+子番号 |
| レコード | $t前漢皇帝陵の再検討:陵邑, 陪葬の変遷を中心に | タイトル |
| | $A 村元　健一;ja(ムラモト　ケンイチ)//著 | 著者名（日本語） |
| | $p38-60 | ページ |
| | $b12XX | 分類コード |
| | $c111000B111EXX1 | 排列のための手がかり |
| 論文 | 10000002 | タグ+子番号 |
| レコード | $t韓半島南海岸新石器時代の埋葬遺構 | |
| | $A 任鶴鐘;zh[ren he zhong]//著 | 著者名（中国語） |
| | $A 平郡　達哉;ja(ヒラゴオリ　タツヤ)//訳 | 訳者名（日本語） |
| | $p127-145 | |
| | $b12XX | |
| | $c18B010EXX1 | |
| | $q東三洞貝塚, 金谷洞栗里貝塚, 欲知島, 山登貝塚, 礼安里, 煙台島, 凡方貝塚 | 内容の注記 |

図 6：現行類目のマスターデータの例（現行形式）

行錯誤を繰り返すことになり、それが形式の変化としても現れることになった。

### ◆3.2.1　文字表現

現行類目では文字符号として UTF-8 を用いている。また、Unicode にない文字を表すために外

図 7：遡及入力されたデータの例

| | | |
|---|---|---|
| 雑誌エントリー | A34S00025 | ID |
| 論文エントリー | 10000000 | 論文タグ (100)+子番号 (00000) |
| | $t游俠に就て | タイトル |
| | $a宮崎　市定 | 著者名 |
| | $j歴史と地理 | 雑誌名 |
| | $k34巻4·5號合刊 | 巻・号 |
| | $p20 | ページ |
| | $d1934年 | 出版年 |
| | $b一般史 時代史 | 分類名 |
| 雑誌エントリー | A34S00026 | |
| 論文エントリー | 10000000 | |
| | $t史記會注考證 | |
| | $a瀧川　龜太郎//著 | |
| | $e東方文化學院東京研究所刊 | 出版者 |
| | $h菊版2冊 | 規格 |
| | $k卷第四、第十 | |
| | $p940 | |
| | $d1934 年 | |
| | $b一般史 時代史 | |
| 書評論文エントリー | 20000001 | 書評論文タグ (200)+子番号 (00001) |
| | $r梁隱 | 評者名 |
| | $j圖書季刊 | |
| | $k1卷1期 | |

| | | |
|---|---|---|
| 雑誌エントリー | A35S02263 | ID |
| 論文エントリー | 10000000 | タグ (100)+子番号 (00000) |
| | $t黔縣方言調查錄 | タイトル |
| | $j國學季刊 | 雑誌名 |
| | $k4卷4號 | 巻・号 |
| | $d1935年 | 出版年 |
| | $b言語文字學 音韻 | 分類名 |
| 細目論文エントリー | 10000001 | タグ (100)+子番号 (00001) |
| | $t記事 | |
| | $a魏建功 | 著者名 |
| | $p3 | ページ |
| 細目論文エントリー | 10000002 | タグ (100)+子番号 (00002) |
| | $t方言同音字輯錄 | |
| | $a舒耀宗 | |
| | $p23 | |
| 細目論文エントリー | 10000003 | タグ (100)+子番号 (00003) |
| | $t方言同音字讀音 (欠) | |
| | $a胡□ | 入力できなかった文字が「□」に |
| 細目論文エントリー | 10000004 | タグ (100)+子番号 (00004) |
| | $t方言同音字記音 | |
| | $a劉復/白滌洲/魏建功 | |
| | $p7 | |
| 細目論文エントリー | 10000005 | タグ (100)+子番号 (00005) |
| | $t方音音素之整理 | |
| | $a魏建功 | |
| | $p21 | |

図 8：遡及入力されたデータの例（細目論文）

字（現行類目外字 [8]）を用いている。

### ◆3.2.2　マスターデータの形式

現行類目のマスターデータの形式は、基本的に、旧 RUIMOKU の形式を踏襲したものとなっている（図 5）。但し、漢字の正規化はやめて、原則として、原表記で入力することになった。また、現行類目に移行後の途中から、中国語著者名にピンインを入力することになり、著者名欄として、原表記とともに任意の言語タグと必要に応じて対応する漢字 / カナ / ラテン文字表記を付与可能な新形式を導入した（図 6）。この他、段階的に幾つかの拡張が加わっており、幾つかのバリエーションが存在する。

## ✲3.3 遡及入力されたデータ

現行類目の開発が行われていた頃、電子化される前の類目を対象に、データを遡及入力する計画が立ち上り、数年をかけて最終的に 1934 ～ 1980 年度の全データの遡及入力作業が行われた。しかしながら、前述のように、この頃、著者は旧 RUIMOKU についてあまりちゃんと理解しておらず、データ形式の設計を適切に主導することができなかった。また、編纂作業に使われた原データが既に廃棄されていたために、冊子体の情報を入力することとなったため、旧 RUIMOKU や現行類目にあるようなさまざまなメタデータの情報が入力できなかった。そして、論文・批評論文とその掲載誌の親子関係が消失し、また、批評論文は批評対象の単行本の子供として表現されることになってしまうという問題も生じた。そして、分類は年度毎に異なっていたりする [9]。こうしたこともあって、この遡及作業では旧 RUIMOKU のマスターデータに似ているものの、互換性のない形

式となってしまった（図 7, 8）。

また、プロジェクト管理がうまく行かず、校正作業がちゃんと行われず、ひどい場合には Shift_JIS で保存してしまってデータを壊してしまうというようなこともあったようである。独自形式であるために、XML のようにバリデーションがかけれる訳ではないから、構文上のエラーも少なくなかった。このため、データは出来上がったもののその品質は低く、機械処理しづらいものとなってしまった。

後に、岩井茂樹氏が XML ベースでの校正作業に取り組んだ。当初、形式の理解が十分でなく、また、データの品質が極めて低かったこともあり、作業は難航し、相当の手作業も必要だったようであるが、少なくとも機械可読性を持ったデータを実現した。但し、内容面での校正はできていないので、特に、日本語と中国語以外のデータの品質は低いままとなっている。

## ■ 4 どう付き合うか

3 節で述べたような事情によって、現在、大別すると

1. 旧 RUIMOKU: 汎用機時代のマスターデータ（1981 ～ 2000 年度版）
2. 旧 CHINA3: 汎用機時代の検索用データベース（1981 ～ 2000 年度版）
3. 現行の類目（2001 年度版以降）
4. 遡及入力されたデータ（1934 ～ 1980 年度版）

の 4 種類の互いに互換性がない類目のデータが存在している。これらは文字コードやデータ構造が異なっており、単一のパーサーで扱うことが困難である。また、これ以外にも細かなバリエーションや運用の変化による意味の変化があると考えられ、現在の所、まだその全貌は把握できていないので、全体を網羅するようなスキーマをきっちりと定義するのはあまり容易ではないといえる。特に汎用機時代のデータについては 30 年ぐらい前の事情だということもあり、ある意味において、考古学や文献学に似たような調査を要し、新たな『発見』や『研究』の進展によって、スキーマの見直しを迫られることも少なくない。形式や分類・運用等の変遷を辿り、それらを現行形式と比較し、対応関係を確定させるためには、各時代のコンピュータ・システムの事情とともに、目録学的な知識や、場合によっては対象文献に対する知識も必要となるなど、なかなか大変な作業だといえる。

利用者にとって、これらがひとつの類目に見えることが望ましい訳だが、現状では、残念ながら、このことは実現しておらず、2 と 3 を統合したもの（類目 Ver.6）、および、2 と 4 を統合したもの（類目 Ver.5）が実現できただけである。そういう訳で、まだ類目におけるレガシーの問題は解決の途上にあるので、『こうすればうまくいく！』というようなことは言い難いのであるが、ここでは問題を避けるための幾つかの方策について述べたい。

### ✽ 4.1 調査と文書化

レガシーデータを使えるようにするには、移植対象となるモダンなシステムとのマッピングをどうとるかが問題となるが、このためにはデータの形式・意味論の把握が必要である。過去のシステムやデータ形式、運用規則等について十分に文書化されていれば良いのだが、そうでない場合もあるし、また、文書の所在が良く判らなくなっていることもある。そして、文書自体のデータ形式の問題もある [10]。また、説明の前提となる環境や知識等が変わりすぎていて、背景知識の理解に手間取ることもある。そうしたことから、レガシーシステムの開発者や実際にデータを運用していた人の話を聞くことができることはとても重要だといえる。人はデータのセマンティクスを保存するための最後の砦だといえる [11]。しかしながら、人は異動するし忘れもするので、30 年前の事情を知る人を確保して情報を得ることはそれほど容易なことではない。なので、異動する前の文書化はとても重要なことだといえる。

システムに関する説明書や各種文書はシステム

の利用や運用等に焦点を当てて書かれていることが多いので、何故そのようになっているのかというような設計思想に関するような情報はあまり記述されていないことも少なくない。一方、システムの開発者が論文を書いている場合、こうした情報を補完する上で有用である。論文内の記述は時に実際に残されているデータ等と食い違っていることもあるが、その時点での開発者にとっての理想像を考える上でのヒントにもなるし、また、その食い違いを考察することで実装上の問題やシステムの変遷、運用上の問題等々を推測する上での重要なヒントを得られることもある。

いずれにせよ、後継システムを設計する上で、過去のシステムやデータの設計思想や具体的な設計、実装、データ形式・意味論、実際の運用等を理解することは非常に重要であり、これができない限り他の作業もできないし、また、不十分な理解のままにデータ形式やスキーマ等を設計した場合、データの移植や統合等で問題が生じるといえる。しかしながら、この作業にはしばしば多大な労力がかかる上、この作業の間はデータ入力作業やシステムの実装、データの公開ができないので、周りの理解が得られにくいこともある。この作業で必要となる背景知識の中にはコンピューターシステムに関するものもある一方、データの対象分野・領域に関するものもある。類目の場合、具体的には、当時の汎用機の知識と、目録学や各分野の研究史に関するような知識も必要となるし、場合によっては、各分野の専門用語に関する知識も必要となり、各領域の人文学研究者と司書、情報科学研究者・情報技術者の共同作業が必要になるといえる。このため、人文情報学的プロジェクトで良く見られるような対話の問題が生じ得るが、類目の場合、プロジェクト型ではなく、毎年冊子体を刊行しながら、あるいは、データベースを運用しながら行わなければならないので、一般的な研究プロジェクトとは別種の問題もある。例えば、図書系の事務職員である司書と教員である研究者が共同で通常の業務とは別に共同作業を行う場合、職制上の問題も存在し得るし、近年の非プロジェクト型予算の削減やそれに伴う人員の減少の影響で一人あたりの業務が増加する傾向にあるので、通常業務とは別にこの種の共同作業を立ち上げることは難しかったり、あるいは、複雑で難しい問題の調査や量的に大きい作業は参加者の負担等の点で厳しかったりする。私見ではあるが、この種の作業はなるべく『宿題』なしで、ミーティング中に実際に参加者全員でデータを見ながら共同で作業するのが、心理的な負担を軽減するという観点で良いのではないかと思う（文書化もなるべくミーティング中に行った方が良い）。これは異なる分野からの視点を突き合わせることで、一人の参加者だけでは解決が難しい複雑な問題を解決することにも繋がり、問題の共有という点でも有用である。その半面、参加者が自分の知らない（不得意な）分野の背景知識や専門用語、議論を聞くことになり、それによる心理的負担が生じることもある。いずれにせよ、難しい問題ではあるが、この部分がその後の成否を大きく左右するといえる。

### ✻ 4.2 揺れへの対処

前節で述べたように、過去のシステムやデータ等を十分に調査して理解を深めることは非常に重要なことであるが、しかしながら、完璧な調査を実現して完全な理解を得ることは容易いことではなく、しばしば、新たな『発見』や過去の理解の誤りに直面したりする。また、例外や運用規則の変化の発見等によって、新たな形式化やモデルが必要になることがある。

一方、大規模なデータベースの形式や意味論を変更することはあまり容易なことではない。特に、現在一般的に使われている関係データベースの場合、一度作ったスキーマを変えるのは困難であるといえる。

このため、RDBMS を使いながらも、正規化は行わないとか、単なる Key-Value Store 的に使うことで拡張の余地を考慮した設計を行うことが考えられる。また、複数の形式の混在を許すような設計にしておくことも有用である。いずれにせよ、この種の形式や意味論の把握が不十分でスキーマの変更や例外への対処が要求されるような分野に

おいては関係データベース（関係モデル）の利用はあまり有利ではないかも知れず、RDBMS を使ったとしても単なるバックエンドとしてのみ使うとか、場合によっては、それ以外のデータベース・システムの利用を検討しても良いかも知れない。

### ❇ 4.3 リファクタリング

より積極的に改良や拡張を行うことを考えた場合、データベースをリファクタリングすることが考えられる。リファクタリングにおいて重要なのは実装を改変した時に仕様やインターフェイス（データへのアクセス仕方）が変わっていないことを機械的な検証によって保証することである。これによって、機械的検証が通る限り、実装を自由に改変できる訳である。しかしながら、データベースに対してリファクタリング手法を適用するためには、データベースの仕様やインターフェイスを実装と分離し、実装が仕様を満たしておりインターフェイスの不変性が維持されていることを検証するためのプログラム（テストケース）が記述できなければならない。これは即ち（少なくともテストケースが書ける程度に）仕様が形式的に定義されていなければならないということであり、言い替えればデータベースの実装たる具体的なデータ形式に対してその意味論（インターフェイスや振舞いなど）が形式的に定義されていなければならないということである。

データベースを作るためには、対象物をどのようにモデル化するかという作業が必要であるが、レガシーなデータベースの場合、その作成者がどのようにモデル化したかは良く判らないことが少なくない。データにおける実装は、概ね、対象物を指示するための表現、例えば、数字や記号列等による code や ID, 名前、その他属性や説明等であるが、code や ID の場合、マッピングテーブルが必要なのは当然だとして、例えば人名の場合、表しているのが人なのか、名前自身なのか、あるいは、作者のようなある役割を担った者を表しているのか、等の可能性が考えられる。これにより、同名の別人を分離すべきかどうかや、別名を扱うかどうか、同じ人でも役割毎に分離するのか、あるいは、法人等の人じゃないものを扱うかどうか、等の差異が生じて来る。こうした差異は、しばしば、レアケースにおいて顕著となるので、データを漠然と目で眺めても良く判らないことも少なくない。このように、レガシーなデータベースの場合、仕様はあくまで推測に過ぎず、残されたデータそのものを『一次資料』として扱わざるを得ないために、データ形式等の改変がしづらくなるという問題がある。よって、データの改変が安全に行えたのか（改変後のデータは改変前のものと構造等価か？）、あるいは、推測したデータモデルが実際のデータと矛盾してないか、などといったことを保証しながら、推測したデータモデルの確実性を高め、仕様の形式的記述を進めることで、改変の自由度を高めて行くのが望ましい。

仕様を形式的に定義することは必ずしも容易なことではないが、もし、この時はこうなって欲しいとか、こういう場合にこうなっては困るというような、データやインターフェイスが満たすべき条件が判っていれば、それを検証するテストケースを書けば良い。テストケースは必要に応じて（例えば、問題が判明する度に）増やして行けば良く、これを積み重ねることで結果的に仕様を明確化することができる。

意味論の形式的定義が十分に定義できていない場合でも、マスターデータをデータベース・システムに入力し、その取り込んだデータをマスターデータの形式で出力し、マスターデータと比較するという手法が利用可能である。マスターデータをデータベース・システムに取り込むためには、マスターデータの形式に対するパーサーが正しく機能し、その結果得られた構文木が正しく評価され、仕様通りの内部表現として正しく格納されている必要がある。そして、データベース・システムにおける内部表現の表現力がマスターデータの表現力と同等以上でなければ、データベースからマスターデータを再現することができない。その上で、データベースをマスターデータの形式で出力するプログラムが正常に動作している必要がある。この全ての条件が満たされて始めて、マスターデータとそれを取り込んで再出力したデータが一

致する訳である。このような手法はコンパイラーのチェックで良く用いられている。例えば、C 言語で書かれた C コンパイラーの場合、

1. ソース S を既存の C コンパイラー $C_0$ でコンパイルして実行形式 $C_1$ を得る
2. ソース S を $C_1$ でコンパイルして実行形式 $C_2$ を得る
3. ソース S を $C_2$ でコンパイルして実行形式 $C_3$ を得る
4. $C_2$ と $C_3$ のバイナリーを比較し、両者が一致したら成功

という検証手法が用いられている。

このような入力を機械処理した結果を比較する手法を用いた場合、チェックが通らず、やむなく人手でチェックしなければならない場合でも、両者の差分をチェックするだけですみ、チェックする範囲を絞りこむ上で有用である。

データ形式を変更する場合、手作業でデータ形式を変更するのではなく、データ形式を変更するプログラムを作成して、機械的に変換することが望ましい。この時、変更前の形式と変更後の形式の表現力が等価であれば、変更後の形式から変更前の形式に逆変換するプログラムも作成して、新形式に変換したデータを旧形式に逆変換して、元のデータと一致するかを調べることで、うまく変換できたかを機械的にチェックすることができる。データの追加や変更の際も、同様な差分チェックを行うことが望ましいといえる。

### ✲ 4.4 機械処理を前提とすること

4.3 節で述べたように、データのメンテナンス性を向上させる上で機械的検証手法の利用はとても重要であるといえる。これはリファクタリングを行わない場合でもいえることであり、特に、大規模なデータベースの場合、ほぼ必須要件なのではないかと思われる。XML や S 式のような機械処理が容易な形式を利用することのメリットもここにあるといえ、独自のタグセットよりも TEI 等の標準的なものを使った方が良い理由もここにあるといえる。即ち、XML 以外の独自形式の場合、パーサーやバリデーター等を自分で書かない限り、構文上の問題も検証できない訳である。XML を用いたとしても、独自のタグセットを用いた場合、DTD やスキーマ等を自分で正しく定義しなければならない。一方、標準的なタグセットの場合、各種定義や周辺ツール等が用意されており、この種のコストを下げることができる訳である。また、第三者が理解しやすいという点も重要である。もし、長期にわたって使われるデータを作ろうと思うなら、TEI のような標準的形式を用いることが望ましいといえる。

しかしながら、レガシーデータを扱う場合、データ形式やデータモデル等の理解の不十分さから、データの内容を損なうことなく安全に TEI 等の標準的な形式に変換することが容易ではないことがある。その場合、レガシーなデータ形式に対応する XML 等の新形式を定義するとともに、4.3 節で述べたように、旧形式から新形式への変換プログラムと新形式から旧形式への変換プログラムを同時に作り、差分チェックによって旧形式での安全性を維持しつつ、XML 関連ツールを利用するということが考えられる。

いずれにせよ、機械的検証手法がないままでデータ入力を進めるのは望ましくない[12]。

### ✲ 4.5 オントロジー化

データ形式やデータモデルに関する知見がある程度蓄積してくれば、それらを機械可読なオントロジーとすることで、複数のデータ形式やデータモデル間の関係をコンピューターに理解させたり、それによって複数のデータベースを跨いで検索を行ったり、リファクタリングや各種メンテナンス用のプログラムを書きやすくしたり、複雑な検索や処理を記述しやすくできる。

例えば、CHISE 文字オントロジーの場合〔1〕、幾つかの対象に対して素性名や素性値の形式の変更を行ったり、素性名の階層化や継承による差分的文字定義の導入等のかなり大規模な変更も行って来たが、

1. 素性名の変更には別名（alias）を利用する

2. プログラムによって、新旧両方の形式をサポートするとともに、具体的な素性名や素性値（データの実装）を隠蔽し、段階的な移行を実現する
3. 変更は原則としてプログラムによって行い、人手によるチェック・修正は元データと変更後のデータの差分箇所（およびその前後）に限る
4. データをチェックするプログラムを利用する

といったことにより、全体として形式の統一が取れていない状態でもとりあえず使える状態を維持し、システムを動かしながらリファクタリングしつづける体制を作っている。

類目においても、CHISE と同様なメンテナンス手法の導入を目指して、CHISE と同様なオントロジー化を試みている [13]。

## ■ 5　おわりに

東洋学文献類目の例を題材に、著者が遭遇したレガシーの問題やそこでの経験、対処のための方策等について雑駁に述べた。ここで判ることのひとつは、レガシー問題とは単なるデータの問題ではなく、また、コンピューターシステムだけの問題でもなく、人的システムも含めた複合的なシステムの問題だということである。そして、コンピューターシステムも人もずっと同じではありえず、移り変わって行くということを鑑みれば、レガシー問題は大規模データベース等が宿命的に抱えてしまうような問題だといえる。

レガシー問題に対処するためには、ベンダーロックインを避けることはいうまでもないことであるが、恐くていじれなくなるような状態を避けるという、より実質的な意味での『いじる自由』の確保が重要であるといえる。このためには、システム全体に対する理解がきちんと継承されることがとても重要だといえ、このためには、適切な文書化やコミュニケーション、情報の共有をきちんと行うことが重要だといえる。また、これらをなるべく機械可読化し、メンテナンスの自動化を計ることは極めて重要だといえる。また、標準的な技術を利用することはデータの可読性を高めるとともに、知識が失われた後でも理解しやすくする上でとても重要であるといえる。その一方で、標準的なフレームワークを利用するために不自然な形式化や運用をすると、後に禍根を残す可能性があることも想像しておいた方が良いかも知れない。データのモデルの寿命はその実装たる具体的なデータ形式よりも長い（あるいは、変更のためのコストが高い）ので、データにおいても、適切に仕様と実装を分離し、データ抽象によって実装を隠蔽することで、モデルの寿命を長くするとともに、実装の変えやすさを高めるような設計や体制が望ましいといえる。

大規模なデータベースのリファクタリングは、現在の所、技術的に容易な問題ではないし、また、前述のように、この問題が人的システムも含めた複合的なものであることを鑑みれば、この問題に対して『こうすればうまく行く！』と簡単にいえるような魔法の道具はおそらくないのではないかと思うが、失敗のパターンやうまくやるためのヒントのようなものはあるのではないかと思う。本稿でそれをうまく述べることができたかは心許ないが、何かのヒントになれば幸いである。

### 注

[1] 類目に含まれる文字の大部分は Unicode に収録されたものであるが、甲骨文字（例： 、 ）やその他先秦期の文字（例： 、 ）およびその隷定字（例： 、 ）等で毎年必ず外字が発生する。そのため、Unicode にない字も扱う必要がある〔3〕。

[2] CHISE の =jef-china3

[3] 各バイトの最上位ビットに 1 を立てた 8bit 表現

[4] 切替え用のシーケンスはあるが

[5] 著者の着任以前の出来事であるので、詳しい事情はあまり良く判っておらず、伝聞なのであるが…

[6] というか、当時、全貌を理解していた人は誰もいなかった。

[7] というか、CHINA3 の WWW 上の検索サービスを作るだけのはずが、いつのまにかなしくずし的に冊子体の編纂システムの開発や冊子体データの作成や組版・

印刷業務まで降り掛かってきたのである。そして、類目の冊子体の組版は多言語・多字種度が高くて、それだけでも非常に大変な作業だった。

[8] CHISE の =ruimoku-v6

[9] 日本語・中国語・朝鮮語の部と欧文の部で分類が異なる年度もあった。

[10] 例えば、古いワードプロセッサー（専用機）等で書かれていた場合、読めなかったりする。

[11] システムの設計思想の把握は少し話を聞いただけでは容易ではないし、人文学者と情報学者の対話がそうであるように、背景となる知識の差異などから対話によって理解に至るには時間がかかることもあるが、データを見てもよく判らないことが理解できることは確かである。

[12] 類目の遡及入力はこの点で非常に問題があったといえる。

[13] 類目 Ver.7。

## 参考文献

〔1〕 Tomohiko Morioka. CHISE: Character processing based on character ontology. Large-scale Knowledge Resources（LKR2008）, LNAI No.4938, pp.148–162, 2008 年 3月.

〔2〕 安岡孝一 .『東洋学文献類目』の編纂の歴史 ── CHINA3. センター所蔵資料の活用と人文社会科学 , 全国文献・情報センター人文社会科学学術セミナーシリーズ No.13, pp.63–70, 2003 年 11 月 .

〔3〕 守岡知彦 . 類目外字における “Old Hanzi” . 東洋学へのコンピューター利用第 20 回研究セミナー , 全国文献・情報センター人文社会科学学術セミナーシリーズ , pp.115–133, 2009 年 3 月 .

〔4〕 星野聰 , 勝村哲也 . 東洋学文献類目データベースの研究と開発 . 情報処理学会論文誌 , Vol.25, No.2, pp.187–193, 1984 年 3 月 .

# 永続性のある歴史資料デジタル・アーカイブへの試論

## ―「アーカイブズ」への接近とデジタル応用の可能性―

後藤　真（ごとう　まこと）

## 1. はじめに

本稿は、日本におけるデジタル・アーカイブの長期保存について論じるものである[1]。

筆者は、以前データベースがアーカイブズとして機能するための様々な可能性について論じた（本稿では、この拙稿を「前稿」と仮に呼称する）。その中での論点をいくつか確認しておきたい。特に問題となるのは、データの「再現性」の問題であった（前稿ではアーカイブズの課題にひきつけて「完全性」と表現した）[2]。

デジタル・アーカイブは更新がある。この更新こそがデジタルデータの最大の利点でありながら、一方で、多くの研究者が同じものを見て検証することができない、という重要な課題を抱えるものであった。

このような論点を提出したそもそもの意図は、人文科学の研究者にとって「使える」デジタル・アーカイブとはどのようなものなのか。もしくは、人文科学の研究者は、なぜ「デジタル・アーカイブは信用できない」と考えるのかという問題意識から始まったものである。

人文科学が真に使えると考えるデジタル・アーカイブとは何かを、作成者の側、そして作成者と利用者の橋渡しをする立場から考察したことが主たる狙いであった。

また、同時期には、アーカイブズとデジタルデータの可能性について、議論がなされることもあった[3]。

それらの動向に対しては、白須が記号体系としてのアーカイブズとデジタル・アーカイブという形でひとつの可能性を見出している[4]。

また、最近では研谷がデジタルカルチュラルヘリテージという視角からデジタル・アーカイブの総合的なガイドラインを作成した。このガイドラインは、極めて周到に用意されたものであり、現在のデジタル・アーカイブ行為の理想的な形態といえる。研谷のガイドライン作成の意図は、品質や方向性を可能なかぎり明瞭にすることを目指したものと理解できる。その点ではデジタル・アーカイブがいかに長期的な資源足りうるかという文脈の延長線上にある[5][6]。

これらは、いわば「アーカイブズとして」、もしくは、長期保存・永続的な保存可能性としてのデジタルデータという文脈（物理的なハードが永続するという意味ではなく、コピーを繰り返しつつ、継承し続けるという意味づけ）での議論であると理解できよう。

一方、いわば「古典的」とも理解できるような

「デジタルアーカイブ」の方向性は今も連綿と続いている。特に文化財をいかに保存し、公開するかという文脈での議論もいまだに継続されている。「文化を継承するためのデジタル化」という議論は、ある種の「正解」ではあるものの、直接的な「アーカイブのため」という方向性は、もはや素朴に過ぎるという断じ方をせざるをえないであろう[7]。

たとえば、京都市の「京都デジタルアーカイブ研究センター」は、平成15年度をもって解散し、http://www.kyoto-archives.gr.jp/ のURIももはや用いられなくなった。その後継リンクとして張られている http://www.archives.astem.or.jp/ もすでに機能しておらず、移転している。さらにその移転後の http://archives.kyo-bunka.or.jp/（財団法人京都国際文化交流財団）のサイトも、一部の文化財閲覧以外がすべて404エラーを返す。なぜなら京都国際文化交流財団は「特定非営利活動法人京都文化協会」に組織移管を行っており、http://www.kyo-bunka.or.jp/ にURIが変更されている。なお、このサイト上でも「デジタルアーカイブ事業」の説明は404エラーを返してくるため、どのようなデータが引き継がれ、公開されているのか。どのような更新がなされているのか、いないのかなどの、諸情報を一切得ることができない。京都デジタルアーカイブ研究センターの仕事は、現在の京都の「デジタルアーカイブ」とは「継続」が疑わしい位置におかれている。

このような内的な論理で方向性の見えないデータの「継承」が行われている状況では、すくなくとも学術情報として、もしくは文化情報として継続的な「デジタルアーカイブ」を構築しているとは言いがたい。少なくとも、単純な現物は劣化するからデジタルで、という「素朴な」議論は、その推進者の側から崩壊しているといってもよい。

現物の代替たる「デジタル・アーカイブ」を目指すのであれば、「アーカイブズ」のありようを参考にすべきであろう。

そこで、前稿ではデジタル・アーカイブ論から「アーカイブズ」という用語にこだわりその論を展開した。しかし、前述のとおり、その状況は少しずつ変わりつつある。現時点では、もう少し整理が必要な研究状況にいたっているといえるであろう。

そこで、本稿では、少しひねりをいれて、デジタル・アーカイブ、とりわけ時空間情報などでよく語られる5W（Who（誰）What（何）When（いつ）Where（どこ）Why（どうして））をやや「もじった」ものを整理の軸として考えてみたい。これらの状況と、私が作成したいくつかのデジタル・アーカイブをあてはめることで、デジタル・アーカイブの長期保存の可能性について述べてみたい。

## 2. 何のためのデジタル・アーカイブか

前稿では、「アーカイブズ」という文脈に注目し、人文科学研究者が参照すべきデジタルデータに着目して論を述べた。

これは、あくまでも人文科学研究に参照するためのデータベースと、その課題というところに論点を絞ったものであった。しかし、デジタル・アーカイブはすべて同様の目的をもって作成されるものではない。

特に、近年はデジタル・アーカイブをその目的で分ける論点が提出されている[8]。解析系と蓄積系でわける論点である。解析系に用いる場合には、そのデータの量・質ともにより高いものが求められる。無論、最終的な成果についてのエビデンスとしてのデータ提示は必要となるであろうが。一方、蓄積系では、特にデータを後世に伝える・共有するということが目的となる。したがって、いかに真正性を保ちつつ、つたえるかということが焦点となってくる。それは、データそのものが正しくなければ出せない、ということを意味しない。誤っているものも含めあらゆる段階のデータにアクセスできることこそが「真正性」を担保する第一歩になるであろう。

ここで、私の作成したいくつかのデジタル・アーカイブに上記の整理を当てはめてみたい。

正倉院文書データベース（SOMODA）に関しては、復原の提示という意味では解析系に属する

ものの[9]、正倉院文書の画像情報を広く公開する、という意味では蓄積系に属する部分が考えられる。

しかし、画像情報は必ずしも質のよいものが提供できていないため、現時点で蓄積系アーカイブとしての機能を果たしきれているのかという課題は残る。

一方上田貞治郎写真コレクション[10]は、まず古写真のアルバム単位、資料単位での閲覧を主眼としている。そのため、画像情報は可能な限り検証可能性の高いデータの提示を試みている。また、可能な限り共有可能なメタデータを用い、多くの人に上田貞治郎古写真アルバムの存在を知ってもらい保存のために役立てると同時に、写真を用いた諸研究に役立てるようなシステム設計を心がけている。

しかし、上田コレクションに関しても、現物資料の保存を前提としており、デジタルデータそのものが、現物資料に代わるものとして位置づけられてはいないことも付記しておく。

## 3. 誰のためのデジタル・アーカイブか

デジタル・アーカイブを評する際に言及される「問題点」の一つに「デジタル化されている資料情報がわかりにくい」というものがある。これは、「ヘルプがない」ということとも関連する。

この問題は、デジタル・アーカイブが誰のために開かれているか、という点と関わってくると考えられる。そのデジタル・アーカイブが専門の研究者を主たる対象とする場合、資料の詳細な情報は必ずしも必要ではないかもしれない。しかし、初学者向けであったり、非専門者向けであるデジタル・アーカイブであれば、基礎的な資料情報が必要となってくるであろう。

前項との関連で述べるならば、解析系は比較的専門の研究者に向けたものが多く、蓄積系は、その性質からも広く開かれたものである場合が多い。

例えば、東京大学史料編纂所のデータベースはむしろ日本史学を中心に専門に学ぶものを対象としていると理解できる（無論、どのようなデータがあるのかもわからないような状況が必ずしも好ましいわけではないが）。また、SOMODA についても、あくまでも正倉院文書の資料・学術情報を中心に検討する目的であることを考えれば、正倉院文書の特性を理解し、その課題を解決することをまず主眼においたデータベースであると理解できる。正倉院文書の画像を閲覧し、文字検索は可能だが、少なくとも歴史系研究者以外が直感的に理解することは難しい設計になっている。特に、復原と現状の６つのフェーズを効果的に理解するのはやや困難であるだろう。

一方、上田貞治郎写真コレクションは、直感的にアルバムを選択することができる。また、アルバムをページ単位で閲覧し、その資料の状況を直感的に理解することができる。つまり、写真を「楽しむ」ことからはじめ、資料の情報がアルバム全体から写真の一枚一枚にまで付されるという意味では、多くの人々にデジタル・アーカイブを理解してもらうことを可能としたものである。

同時に、関連の研究論文を掲載し上田貞治郎の学術的な位置づけを明らかにすると同時に、資料の特質を効果的に表現することも考慮している。その点では、研究者のみならず、多くの人々にむけたアーカイブであるといえるであろう。

## 4. いつまでのどこにあるデジタル・アーカイブか

これは、そのデジタル・アーカイブが必要とされる期間の問題である。プロジェクト型であり、一時的に研究データを発見するためのデジタル・アーカイブであれば、時限性を持って作成されるものであってもよい。また、「発見型」データベースであれば、将来的には大幅な変更ないし、消滅を許容するものでありうる。

もしくは、先ほどの「誰に向けたアーカイブか」という課題に即し、一時的に閲覧し「楽しむ」データベースであれば、問題はないかもしれない。

しかし、通常のデジタル・アーカイブであれば少なくとも、可能な限り精緻な再現性とまでは言わずとも、ある程度「再発見」の可能性を模索し

なければならないであろう。

この点は、本稿の主題とも深くかかわってくる。そのため、どこにあるアーカイブか、という問題と併せて、多少位相を変えて後述したい。

どこにあるか、という問題は、物理的なサーバの位置・データの主たる権利者が誰かなどの問題が複雑に絡み合うが、ここでは、ある研究者集団に単純に即すのか、もしくは永続的な機関に属すのか、という位相で考えてみたい。

一般的に解析系データベースは、多くの場合研究者（グループ）の研究に即してデジタル化される。また、大学では蓄積系のデジタル・アーカイブも基本的には研究者の興味・関心に即してデジタル化される場合が多い。しかし、これらは研究グループの解散や、研究者の所属変更、引退などに際し、長期的な維持・管理が困難になる場合が往々にして生じている。

最近では COE プロジェクトの成果の一環として作成されたデジタル・アーカイブ群の「処理」に苦労しているところもあるようである。

SOMODA に即して述べる。SOMODA は正倉院文書データベース作成委員会の指揮のもと入力を行った。委員長は大阪市立大学教授（作成当時）栄原永遠男氏であった。栄原氏のもと入力を行った SOMODA は、大阪市立大学学術情報総合センターのデータベースサーバ内におかれている。

しかし、栄原氏は 2009 年度末をもって停年退職し、2010 年現在は特任教授となっている。現時点では SOMODA のサーバは維持できているが、今後の保証が必ずしもあるわけではない。一人の研究者・研究グループがその問題意識に即しデジタル・アーカイブを作成することは、その方向と目的が明瞭であるため、予算の獲得や、推進力となりやすい。しかし、作成には予算がついても、研究組織である大学の場合デジタル・アーカイブのメンテナンス・改善・向上などには予算がつきにくい原状がある。

SOMODA も入力には科学研究費（研究成果公開促進費）の予算が付されたが、その後の効果的な改善・向上に資するための予算獲得の方法がいまだに見つかっていない。それが、データベースそのものの新規性を徐々に奪い去り、さらに維持管理を困難にするという悪循環に陥りかねない状態にある。

上田貞治郎写真コレクションも同様の懸念は常に指摘されている。現時点では、大阪市立大学都市研究プラザのサーバ上に置くことが決まっており、組織が続く限りはサーバの維持は可能であろうが、本データベース構築のための組織「上田貞治郎写真史料アーカイブ編纂室」[11] は、現時点では大阪市立大学専任の教員がおらず、長期的な安定運用が危ぶまれる状況にある。今後の現物資料との連関を考慮しつつ、維持管理の方法を早々に決定する必要がある状態である。

大学の、研究室の研究関連資料のアーカイブズ一般についても、その保存方法が議論されつつある状況だが [12]、データベースについても同様の指摘を行う必要がある。URI の長期的な維持という意味でも、可能な限り「どこにおくか」はデジタル・アーカイブの長期的な保存について深くかかわっていると考えるべきであろう。

また、入力の際には、長期的な維持・保存が必要なデジタル・アーカイブなのか、それともある段階で明確に使命を終えるべきデジタル・アーカイブなのかを整理しておくべき必要がある。

## 5. なぜ、デジタル・アーカイブなのか

これらを総合して勘案した場合、デジタル化の際には、目的をより明瞭化しなければならないことがわかる。

特に、単純に保存を意識する場合には、デジタル化は第一選択ではない [13]。デジタルデータは、今までても述べてきたように、漫然と作成していても、長期維持が可能なしくみとはならない。また、現物の写像であるかぎり、相対的な「劣化」は起こりうることを前提とすべきである。

むしろ、どのような目的なのか、誰に向けるのか、そしてそのためにはいつまで、どこにデータを置き、維持・管理を行うのかを勘案し、「なぜデジタルでなければならないのか」を再考慮する

段階にあるのは間違いない。

デジタルである優位性は、少なくとも現在の技術水準では、通時性よりは共時性にある。そのため、その共時性と通時性がバランスよく発揮できなければ「デジタル・アーカイブを選択する」という理由にはなり得ないだろう。

以降、さらに別に節を立て、正倉院文書データベース（SOMODA）と上田貞治郎写真コレクションをそれぞれの資料の特性とともに「なぜデジタル・アーカイブにするのか」を細かく述べていきたい。

## 6. 正倉院文書データベースの特徴とアーカイブズとしての機能

正倉院文書データベースは、その名のとおり、正倉院文書に関するデータベースである。正倉院文書そのもののテキスト・画像、そして接続情報や料紙情報などの「メタデータ」的な情報、さらに関連論文リストや写経事業リスト（正倉院文書の多くを造東大寺司写経所の関連文書であるため）などまでを格納したデータベースである。

そのため、単純に「復原」だけではなく、正倉院文書の総合的な情報が得られるツールであることを目指している。しかし、これらの情報が長期的な「アーカイブズ」として機能することを必ずしも目指しているわけではない。

それは、現時点での SOMODA の技術的な限界もひとつの要因にある。SOMODA は、データの更新の際に、過去版を参考できるようなシステムを持っていない。また、復原情報なども再検証できるシステムを持っていない。後者については後述する。しかし、それだけではなく、SOMODA が対象としている史料の問題もあるのである。

SOMODA が対象としているのは「正倉院文書」である。正倉院文書は正倉院宝庫に伝来し、現代でも厳重に「勅封」で管理される「御物」である。

管理主体は宮内庁であり、SOMODA は宮内庁正倉院事務所の許諾の上で公開されている。しかし、SOMODA 作成の際には実際の文書そのものにアクセスして作成してはいない。SOMODA はあくまでも、正倉院文書に関する「学術情報」を蓄積したデータベースなのである。

また、正倉院文書は、ある意味で「散逸の心配の（相対的に）ない」史料である。現物史料の真正性の保障、史料の長期保存などについては、他の文書史料に比しても圧倒的に「安全」な環境にあるといえる。また、デジタル・アーカイブを用い（少なくとも現時点では）多く知らしめることで、その保存をはからねばならない状況ではない。いわば、史料そのものの永続性については、デジタルデータを用いる必要が少ないタイプの史料であるといえる。

しかし、この保存環境を維持するために犠牲になっている点が一点のみある。それは、公開の問題である。正倉院文書の原史料を閲覧できる機会はきわめて限られている。正倉院宝物の調査を行う関係者ですら、長期計画の中での調査の一環でしか触れることはできない。それ以外のものは、たとえ正倉院文書を専門に研究するものであっても、その原史料に触れる機会は持ち得ないのである。

そのため、実際の古文書から料紙情報などを必要に応じ情報を得ることはできない。無論、復原の可能性を実際に検討することなど一切不可能である。また、調査に応じ公開される情報も、分散されて出てくるという問題がある。

SOMODA では、この「公開の問題」を解決することを目指した側面が大きい。料紙情報に可能な限り触れられるように、画像に可能な限り触れられるように、そして、復原の可能性にさまざまな形でアクセスできるように目指したものである。無論、公開については宮内庁が一義的に行うものであるし、その点で宮内庁正倉院事務所の活動になんらの妨げを行うものでもない。

しかし、アーカイブズが長期的な保存と同時に公開をしなければならないものであるとしたら、学術的な側面ではアーカイブズの位置づけを果たしているともいえる。SOMODA は、正倉院文書という史料の特殊性に即したデータベースであるといえる。

## 7. 上田貞治郎写真コレクションの資料の特徴と、デジタルアーカイブ

上田貞治郎コレクション（正式名称は『上田貞治郎全国名所写真貼コレクション』関連デジタル・アーカイブ）は、20 世紀初期写真業界の領袖で、古写真蒐集家であった旧上田写真機店社主上田貞治郎（1860-1944）が蒐集・作成したアルバム群のデジタル化である。

これらの写真群の存在は、昭和 50 年代に読売新聞社が展覧会をやるなど、一部では知られていた。しかし、関連文書とともにその全体像が知られるようになったのは、ごく最近の「編纂室」の仕事であり、関連のまとまった資料情報や研究もごく最近になってあらわれてきた [14]。

こちらは SOMODA とは対照的に「新出資料」である。また、あくまでも個人蔵であるため長期的に安定した保存環境を維持できる状態ではないのが現状である。また、写真という化学変化を起こしやすい物質であり、かつ、アルバムには（しかたのないことではあるが）セロハンテープなどでの決して良好とはいえない修補が加えられており、劣化の進行状況が不安な状態にある。

また、写真一点一点は、複製も多く、全てが「一級資料」とは言い切れない。さらに昭和の写真（写真史の分野では「古写真」とは定義されない）も混じっており、時代的な希少性も決して高いとは言えない。上田貞治郎写真コレクションの最大の価値はアルバムの編成やその蒐集の特性であって、今までの「デジタルアーカイブ」論では、むしろ積極的に「後回し」にされるタイプの資料であるといえるだろう。

結果的に現物の保存にも多くの予算が割けず、温度・湿度の管理もきわめて困難な状況での資料の「管理」を強いられているのが現状である。

そのようないわば「現物の限界」を抱えた資料であるところに上田貞治郎コレクションは置かれている。

そのため、まずは「保存のためのデジタル・アーカイブ」を考慮する必要が生じている。

何を行ったか。まずは、資料の存在を広く公開し、存在を知ってもらうところからはじめた。デジタルデータ、とりわけ web での公開を前提としたデジタルデータの最大の利点は、4 でも述べたように、「共時性」である。多くの人々の目に触れる資料がいきなり「消える」ことは少ない。そのため、まずはデジタルデータでその価値を知らしめることを第一段階とした。公開による「保存」である。この場合にはデジタルデータは「永く保存されるべきもの」として強く意識されていないことには留意しておくべきであろう。

次のステップとして、可能な限り高精細のデータを作成し、劣化状況を公開する段階、そして資料構造とそのメタデータを記すという段階に進む。特に劣化状況を示すことは、どの程度「危ない」かを専門の研究者に示し、助言をいただくためのものである。その意味でもよりきれいな情報を発信するという方向性は一切指向しない。常にある段階のアクセスすれば同じものが見られるためのデータ提供を第一義に考えている。

上田貞治郎古写真コレクションは、むしろ、多くの近代資料と近い特性を備えているといえるであろう。この資料の特殊な部分は化学的な劣化の早い写真であることと、アルバムという複数の構成要素であるといえる。

上田貞治郎写真コレクションは、多くの人が（資料の劣化とのバランスで許容できるかぎり）資料に実際に触れることができる。しかし、一方で保存に関してははなはだ心もとないところが多く、そのためにデジタル・アーカイブを作成することを選択した。

SOMODA と上田貞治郎写真コレクションは、公開と保存のバランスがそれぞれに異なっている。しかし、デジタル化を積極的に行うべきであったという点では共通していた。

それでは、この両者を「アーカイブズ」として見た場合、どのような位置づけを考えるべきであろうか。前稿でも述べたが、SOMODA では復原情報の提示を含め、再現可能性を保証していない。その点には今後の課題があることは事実である。

表：SOMODA と上田貞治郎写真コレクションそれぞれのデジタル・アーカイブの位相の比較

| | SOMODA | 上田貞治郎写真コレクション |
|---|---|---|
| なんのための | 主として「復原」情報の提示<br>（解析系） | 主として現物資料の保存と閲覧<br>（蓄積系） |
| 誰のために | 正倉院文書の研究者 | 写真を用いた研究者をはじめ、市民にまで広く閲覧 |
| いつまで | ある程度の永続性をもって。データの更新は（諸条件が許す限り）積極的に更新を行う | 現物資料とともに可能な限り永く同一のものを使用。ただしメタデータには更新の余地あり |
| どこに | 大阪市立大学学術情報総合センターのサーバ<br>もとの資料の権利関係は、宮内庁正倉院事務所をはじめ、正倉院文書を扱う諸機関に。デジタル入力したものは正倉院文書データベース作成委員会に | 大阪市立大学都市研究プラザ。ただし、近い将来には可能な限り永続するしくみのある場所に移動。<br>原資料は上田家が保存。学術関連の情報については大阪市立大学都市研究プラザに |
| なぜ | デジタル環境でないと自由な復原を行うことが困難であったたため | アーカイブズとしてのアルバム群の公開の最適解であった |

とりわけ、「いつの段階での復原情報か」は、正倉院文書の帳簿理解にとっては重要である。しかし、それ以外の情報については、（諸条件が許すかぎり）積極的に更新を行っていくべきものであると理解している。それは、多くの資料に「もととなる現物」があるためであり、最終的なアクセスはそちらに任せることができるからだ。

一方、上田コレクションは、蓄積系である。そのため、特に資料へのアクセスを（現物への可能な限り接触を減らす（＝保存）という意味で）減らすことができ、学術情報に資するためのアーカイブとしては、データの永続性と参照可能性はできうる限り維持したい。しかし、現実としては、「現物が適切に保管されている」ことを前提にすれば、必ずしも永続的である必要はないかもしれない、とも考えるのである。

## 8. まとめにかえて

以上、論点を散らかしたままではあるが、まとめにはいる。本稿がこのような整理方法を試みたのは、以前からあったデジタルアーカイブ論のいわば「素朴な議論」から、近年の「解析系」「蓄積系」という理解の方向性を再整理するとともに、デジタル・アーカイブの目的の多様化を提示したものである。

いわば「素朴な」デジタルアーカイブ論では、多くの場合、現物資料の代替をも可能とする議論が多かった。しかし、実際には文字コードの限界、画像データや音声データのフォーマットや解像度の維持、サーバや URI 維持コストなどの問題からデジタルデータの永続性が疑問視されるようになった昨今、現物資料の代替としてのデジタルアーカイブは最早不可能に近い存在となっている。

無論、ボーンデジタルな資料は、維持コストの問題を除けば、ある程度の安定した保存の可能性を模索できよう。しかし、現物資料の写像としてのデジタル・アーカイブは、どこまで行っても現物資料の「一部」であり、全てとはなりえない。デジタル技術の進歩とともに現物資料の一部がより広がることを考えれば、相対的には一段階前のアーカイブは「劣化した状態である」と言ってよい面すらあるだろう（図 1）。そして、デジタル技術の進歩に鑑みた場合、デジタルデータの方が

現物資料よりその劣化は早いと言えるのだ。

事実、資料保存機関の中には、（実物を多くアーカイブする予算がないこともあるが）目録情報を「変更前提」で作成し、「誤りを含むがまずは公開」という意識を持って作成をするにいたっているところも現れてきている[15]。

現物資料の劣化速度より、デジタル・アーカイブがより「忠実」に資料をアーカイブできる技術の向上速度が速ければ、時間軸Bの段階で最も「劣化」しているのはデジタル・アーカイブAである

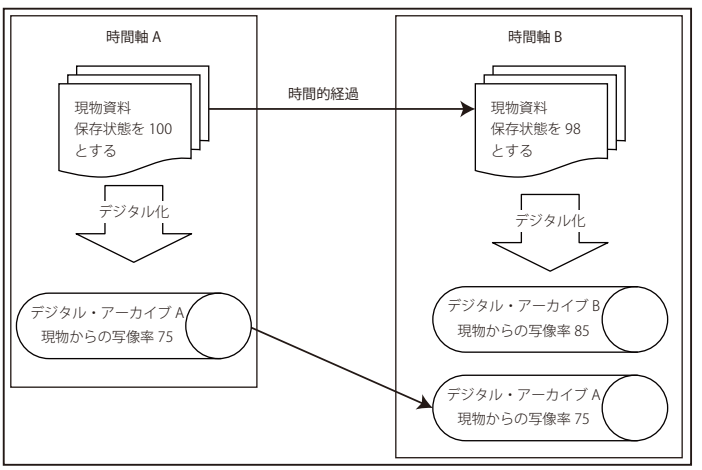


図 1：デジタル・アーカイブの相対的劣化概念図

したがってデジタルデータが須らく長期保存可能な存在であるという意識は、最早通用しないといってよい。

その意識をふまえたうえで、何を・誰のために・いつまで・どこにデジタル・アーカイブを作成するのか、そして、なぜそれを作成するのかの射程を明瞭にしたうえで、作成する必要がある。

むしろ、その際には積極的に「アーカイブズ」という概念を用いるのかどうか、再考が必要であろう。どこまでの情報を維持すればよいのか、そのための長期的な予算、場合によっては作成者がいなくなった後までの長期的なケアの枠組みまで含めるのか含めないのか。場合によっては、「時限的なデジタル・アーカイブ（と呼称すべきはわからないが）」という可能性を探ることを考えるべきであるのかもしれない。時限的なデジタル・アーカイブである場合には、その旨を明示するなどの工夫も求められるであろう。

デジタル・アーカイブは万能ではない。それは、もはやこの数年で充分に理解されたことのはずである。デジタルの優位点を生かすアーカイブとは何か、もう一度再整理を行う段階に来ているといえるであろう。

## 注

[1] なお、本稿では、美術品・優品などを「保存・継承」などの文脈でデジタル化する議論を「デジタルアーカイブ」それ以外は「デジタル・アーカイブ」と呼び、区別をしている。また、「資料」を通常の場合には用いているが、歴史学を中心に資する資料については「史料」の用語を用いた。

[2] 後藤真 2008「文化遺産学における「デジタル」序説 ── 保存と共有・活用と表現 ──」（『情報処理学会研究報告』CH-79、情報処理学会）

[3] 五島敏芳 2008「デジタルアーカイブにおける永久保存」（『情報処理学会研究報告』CH-79、情報処理学会）

[4] 白須裕之 2008「記号機能としてのアーカイブズ」（『人文科学とコンピュータシンポジウム』2008、情報処理学会）

[5] 研谷紀夫・北岡タマ子・高橋英一 2009「文化資源の電子化における記銘情報管理を重視したガイドラインの提案とそれに基づく Digital Cultural Heritage の構築」（『人文科学とコンピュータシンポジウム』2009、情報処理学会）

[6] 研谷紀夫 2010『デジタルカルチュラルヘリテージ構築のためのガイドライン（評価版 ver1.0）』（**http://www.center.iii.u-tokyo.ac.jp/guideline**、東京大学大学院情報学環・凸版印刷株式会社）

[7] 笠羽晴夫 2010『デジタルアーカイブ 基点・手法・課題』）文化とまちづくり叢書、水曜社）

[8] 原正一郎の「文化とコンピューティング国際会議（2010. 02. 24 開催）」の発表での指摘などによる。

[9] 後藤 2008「正倉院文書データベースと「復原」」（『アジア遊学』113 号、勉誠出版）など

[10] 緒川直人・後藤真 2008「デジタル化による写真の史料学的研究の方法論的進展── 上田貞治郎写真コレクションのデータベース化を通じて ──」（『人文科学とコンピュータシンポジウム』2008、情報処理学会）

[11] **http://www.ur-plaza.osaka-cu.ac.jp/archives/ueda.html**

[12] 五島敏芳 2010「京都大学研究資源アーカイブの構築」（『人間文化研究情報資源共有化研究会報告集』1、大学共同利用機関法人人間文化研究機構 研究資源共有化事業委員会）

[13] 緒川 2010「戦前期写真資料学の誕生 ──「上田貞治郎日本全国名所写真帖コレクション」の方法史的考察 ──」（『文化資源学』第 8 号、文化資源学会）

[14] 後藤真 2009「「デジタル化」とアーカイブズ ── 〝正倉院文書データベース〟と近代史料のデジタル化を通して ──」（『アーカイブズ学研究』第 10 号、日本アーカイブズ学会）

[15] 国立公文書館は、誤りを前提とし、更新を前提としたシステムを構築している（**http://www.jacar.go.jp/uvoice/uvoice.html**）。なお、その点において、誤りの多さを指摘する新聞報道があった（毎日新聞　2010 年 9 月 6 日　大阪夕刊）が、前提が異なるので的確な記事とはいえないであろう。

# ソフトウェア レビュー

2010 年度のレビューをお届けする。

本号では、iPad を採り上げた。単に新しい情報端末という観点からだけではなく、多漢字環境を中心に、どのような（本会会員ならではの）使い方があるのか、についてレビューをお願いした。是非ともご一読いただきたい。

ソフトウェアについては、IME が元気である。これは主にサーチエンジンを基盤とするソフトウェア会社からのリリースが活発だったことによるが、情報を検索する前段階部分として IME を捉え、それにアプローチする新興 IME と、ATOK などの規範的な文章語とは何か、と問い続けるアプローチの違いが興味深いところである。

その他、ソフトウェアや規格についても、本年は将来的な展開を見据える興味深いレビューが書かれている。読者諸賢におかれては、一読後何かしら益するところがあれば幸いである。

## Contents


| | | | |
|---|---|---|---|
| **iPad** | 研究者のための iPad | 師　茂樹 | 105 |
| | iPad で見る、聞く、読む、学ぶ ── 中国語教育を中心に ── | 清原　文代 | 112 |
| **ソフトウェア** | Windows 7 & Office 2010 | 千田　大介 | 125 |
| | フォント・ブラウザ・多漢字 | 上地　宏一 | 127 |
| | 文字コード技術 | 師　茂樹 | 138 |
| | IME | 小川　利康 | 140 |
| | 一太郎 2010 & ATOK2010 | 山田　崇仁 | 145 |
| | Adobe InDesign CS5 | 山田　崇仁 | 148 |

❖ iPad

# 研究者のためのiPad

師　茂樹

## ❖ 持ち歩き用コンピュータとして

最近、多くの研究者[1]が、日常的にノートパソコンを持ち歩くようになった。ノートパソコンは、昔と比べれば格段に軽くなったとはいえ、それでもそれなりに重くてかさばる。それを裏付けるかのように、テキストを入力するだけのデジタルメモ「ポメラ」(KING JIM）がヒットしたり、「ネットブック」などとよばれる小型のノートパソコンや、「スマートフォン」とよばれる高機能な携帯電話に注目が集まったりしている。いずれも、外出先で“仕事”をするための軽くて高性能なデバイスが欲しいという虫のいい要求の現れなのだが、これといった定番が確立されたわけでもなく、ノートパソコンが手放せないというのが現状ではないかと思われる。

iPad は、Apple 社が 2010 年 5 月（アメリカでは 4 月）に発売したタブレット型コンピュータである。発売日にはテレビのニュースなどでも紹介され、電子書籍ブーム（あるいは騒動）とも相まって、多くの人々の関心をひきつけている。研究者の中にも、ノートパソコンにかわる新しい持ち運びコンピュータとして、iPad に注目している人も多いのではないだろうか。ここではまず、研究者の持ち歩きマシンとしての iPad について、簡単に検討してみよう。

| | |
|---|---|
| メモリ | 16、32、64GB |
| ディスプレイ | 9.7 インチ／ 768 × 1024 ピクセル |
| 通信 | Wi-Fi または Wi-Fi ＋ 3G ／ Bluetooth など |
| 寸法 | 242.8mm × 189.7mm × 13.4mm |
| 重量 | 680 グラム（Wi-Fi モデル）／ 730 グラム（3G モデル） |

### ■ ディスプレイと操作環境

iPad の概要についてはよく知られているところであるが、一応確認しておこう（2010 年 9 月 2 日現在）。

ディスプレイは概ね A5 版程度の大きさで、縦に持てばハードカバーの書籍や論文の抜き刷りの 1 ページぐらい、横に持てば文庫本見開き 2 ページぐらいに相当する。充分に読みやすい大きさである。ただし、軽いとは言えない重さなので、文庫本のように片手で長時間持ち続けることは難しいだろう。

通常の操作はタッチディスプレイと 3 軸加速度計（傾きセンサー）によって行う。例えば電子書籍の場合であれば、指でディスプレイをなぞることでページをめくり、iPad 本体を傾けたりすることで縦持ち（1 ページ表示）・横持ち（見開き 2 ページ表示）を変化させるのである。傾きを変えても縦横を変化させないようにするために、画面固定用のスイッチも存在する。また、少数ではあるが、アプリケーションの中には傾きによる操作に対応していない（画面が固定されている）ものもある。片手で持ちながら操作をすることは無理なので、手に持って操作をする場合には両手を使わざるを得ない。

文字入力もまた、基本的にはタッチパネルディスプレイ上に表示されたソフトウェアキーボードによって行う。画面が大きいので、両手のタッチタイピングに近い感覚で入力することができる。

ソフトウェアキーボードは通常のキーボードとは異なり、たとえば英数字の入力時においてさえ、数字や記号とアルファベットが複数のキーボードに分割されたりしているので、切り替えながらの入力が面倒といえば面倒である。また、カーソルキーが存在しないのも、少し不便である。テキスト入力時にカーソルを特定の位置に移動させたい場合には、その場所を直接タッチするのであるが、小さい文字のあいだを指で指

定するのはけっこう難しい、という印象があった（英語入力の事情に合わせて、日本語でもひらがなと漢字の境目などを“単語”の境界として認識し、自動的にカーソル位置を補正しているような印象がある）。

ソフトウェアキーボードに不満がある場合には、別売りにはなるが Apple iPad Keyboard Dock や Apple Wireless Keyboard などの外部キーボードを利用することもできる。ちなみに後者は、iPad の場所とキーボードの場所の両方が必要なので、ノートパソコン以上に場所をとる場合がある（図 1）。

多国語キーボードでサポートされている入力可能な言語は、以下のものに限られている [2]。

> 英語（米国）、英語（英国）、フランス語（フランス、カナダ）、ドイツ語、日本語（QWERTY）、オランダ語、フラマン語、スペイン語、イタリア語、簡体字中国語（手書き文字、ピンイン）、ロシア語

本誌の読者としては、繁体字中国語や韓国語など、アジアの言語の多くが含まれていない点に不満をおぼえるのではないだろうか。本稿執筆時点では、本誌発売後の 11 月ごろに iPad が iOS 4 にバージョンアップするのではないかという報道がなされているが、その際により多くの言語をサポートしてくれることを希望したい。

### ■ 日常的な仕事のための機能

iPad には、メール、ウェブブラウザ（Safari）、カレンダー（ToDo 機能も含む）、アドレス帳、メモ帳などのアプリケーションが最初から含まれており、パソコンと簡単に同期することができるので、手帳的な利用に便利である。いずれも、Unicode ベースの多言語に対応しているので、閲覧等には問題がない（入力は先に述べたように不可解な制限がある）。

ウェブブラウザによる情報収集ももちろん可能である。パソコンで集めた情報は、Evernote などを使って蓄積、整理している研究者も少なくないと思うが、iPad 用の Evernote も iTunes Store で無料配布されている。

原稿などを執筆するためのアプリケーションも多数入手することができる。また、Mac 用のオフィススイートである iWork（プレゼンテーションの Keynote、

図 1 ：Apple Wireless Keyboard を使った文字入力

文書作成の Pages、表計算の Numbers）の iPad 版が、各 1,200 円で販売されている。iWork は Microsoft Office のファイルを読み書きすることができるので、それなりの仕事をすることは可能であろう。VGA アダプタを使ってスライドを外部のプロジェクターに出力し、プレゼンテーションをすることも可能である。

## ❖ 読書環境としての iPad

### ■ 電子書籍の入手方法

現在 iPad には、電子書籍の配布形式が二つ存在する。ひとつは一作品が一アプリケーションになっているタイプ（以下、アプリ型）、もうひとつは作品データと閲覧アプリケーションが分離しているタイプである（以下、データ型）。アプリ型の電子書籍はゲームを買ったりする場合と同様、App Store での入手になるが、データ型の電子書籍は何らかの形で書籍データをダウンロードした後に、それを iPad の閲覧ソフトに取り込む作業が必要となる。

データ型の電子書籍ビューワとして代表的なのは、以下のものであろう。いずれも App Store から無料で入手可能である。

1. iBooks
2. Kindle for iPad
3. ebi Reader
4. i文庫 HD
5. Stanza
6. Apabi Reader

1.は Apple の電子書籍アプリケーションであるが、基本的に英語の本が中心で、中国語の古典籍がいくつ

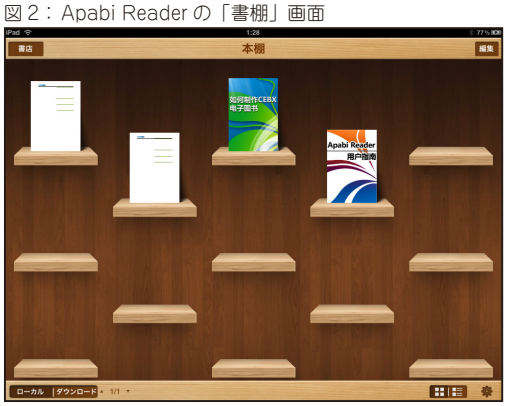
か見られるものの、日本語の書籍は執筆時点では配信されていないようである。

2.はマルチプラットフォームで展開しているAmazon Kindle の iPad 版 で あ る。Amazon.com で Kindle 用のデータを購入すると、登録されている Kindle に自動的にダウンロードされるという仕組みである。Amazon.com なので基本的に英語の本しか販売しておらず、また英語の本であっても “This title is not available for customers from your location in: Asia & Pasific” と表示され、購入できない商品がある点には注意が必要であろう。なお Kindle の仕様を継承しているためか、横向きにしても見開き 2 ページ表示はならず、横長の 1 ページとして表示される。

3.の ebi Reader は、日本企業による eBookJapan[3] というサービスの電子書籍用のアプリケーションで、こちらもマルチプラットフォームとなっている。前二者と異なり、日本企業だけあって、日本語の書籍、特にマンガが充実している。iPad 用の ebi Reader は iPad のワイド画面に最適化されたものではなく、iPhone ／ iPod touch 版の画面の大きさを規準にして作られているので、標準の大きさだと iPad の画面に比して小さく、画面全体に引き伸ばせば大きくなったドットが丸見えの荒い画面になってしまう。ただし、書籍を表示する部分に関しては、拡大しても荒いという感じはしない。

4.は上の三つとは異なり、現在のところ青空文庫を中心とした電子書籍リーダーである。しかし、上の三つと比べても遜色がない、場合によってはそれらを上回る表示品質と、わかりやすいインターフェースを持っている。最初から青空文庫に収録された文学作品などが多数内蔵されているので、すぐに読み始めることができる。

図 2： Apabi Reader の「書棚」画面




5.は数多くのフォーマットに対応しているビューワとして有名である（下表参照）。デフォルトでプロジェクト・グーテンベルクなどがライブラリとして登録されている。

6.は中国発の電子書籍フォーマットとして注目されている CEBX を開発する方正技術研究院数字出版分院による iPad アプリケーションである。iBook と同様のインターフェースを持っているが、日本語訳がこなれておらず、若干完成度は低いという印象である。また、電子書籍の入手先として「中華数字書苑」と「番薯网」が利用出来るようになっているが、前者は法人向け、後者は執筆時点では 20 冊程度しかダウンロードができず、実用段階とは言えない状況である。ただし、中国語圏から先陣を切っての公開であり、また CEBX の今後を占うアプリケーションであるとも言えるので、完成度の向上と書籍データの充実を期待したい。

なお、香港企業の Ouiivo studio 社が公開している iPhone 用の電子書籍ビューワ Ouiivo eReader for iPhone（億万書庫）は、執筆時点では iPad 未対応である。また、電子文庫パブリのアプリケーションは iPad 上で（iPhone の画面の大きさで）動作するものの、電子書籍をダウンロードして閲覧することには対応していない。

### ■ 自前ファイルの取込み方

以上のアプリケーションは、有料・無料はともかくも外部で作られた書籍データを入手して読むタイプのものであった。一方で最近「自炊」という言葉を耳にした読者も多いと思うが、自分でスキャンしたり作成したりした PDF などを iPad 上で読みたいという要求も当然ながら存在するだろう。iPad には PDF ビューワのほか、先に紹介した電子書籍アプリケーションのなかにも外部ファイルに対応しているものがあり、それらを用いれば大量の文書を iPad に入れて持ち歩くこともできる。以下に主なアプリケーションが対応しているフォーマットとファイル転送方式をまとめてみたので参照されたい。

これ以外にもブラウザ経由で HTML 化されたテキストを閲覧するなど、様々な方法が存在する。また変

| アプリケーション | 対応フォーマット | ファイル転送方法 |
|---|---|---|
| iBooks | EPUB、PDF | iTunes |
| i 文庫 HD | 青空文庫形式テキスト、PDF | iTunes、Dropbox、FTP、Web ダウンロードなど |
| Stanza | EPUB、eReader、PDF、DjVu、CBR/CBX | iTunes、Web ダウンロード |
| Apabi Reader | CBEX、PDF | iTunes |
| GoodReader | PDF | iTunes 、Dropbox、Web ダウンロードなど |
| iAnnotate PDF | PDF | iTunes 、Dropbox、Web ダウンロード、Aji PDF Service など |

則技としては、iBooks や Kindle for iPad の場合、個人出版によって自作ファイルを（広い意味での）ファイル転送を行うことも可能である。もっとも、「自炊」したファイルの転送はできないし、出版ということで様々な制約があるのは言うまでもない。

さて、ファイル転送方式を見てみると、iTunes や Web 経由のダウンロードは当然として、ファイル共有サービスの Dropbox に多くのアプリケーションで対応している点は注目される。またファイルフォーマットでは PDF がやはり強いが、iBooks や Stanza で採用されている EPUB や i 文庫 HD の青空文庫形式テキストは、PDF とは異なるリフロー型（ウェブブラウザにおける HTML のように、画面の大きさや文字の大きさに応じて表示を変化させる形式）の電子書籍フォーマットとして注目される。なお、青空文庫形式は返り点などに対応している点で（本誌の読者に少なくないであろう）漢籍・古典などの研究者としては要注目であるが、残念ながら i 文庫 HD は返り点などの表示に対応していない（図 3 は AIR 草子 [4]、図 4 は i 文庫 HD の表示例）。

## ■ ブックマーク・注釈機能

紙の書籍や論文などを読む場合、重要な箇所に線を引いたりマーカーで強調したり、ページの端を折ったり、余白にメモを書いたりすることは、多くの人がやっていることであろう。電子書籍の場合も、そのようなことができなければ「読書環境」としては失格であろう。多くのビューワでブックマーク機能やしおり機能が搭載されているが、なかでも注目されるのは iAnnotate PDF の注釈機能である。これは、通常のコメント機能のほか、手書きで書き込みをすることもで

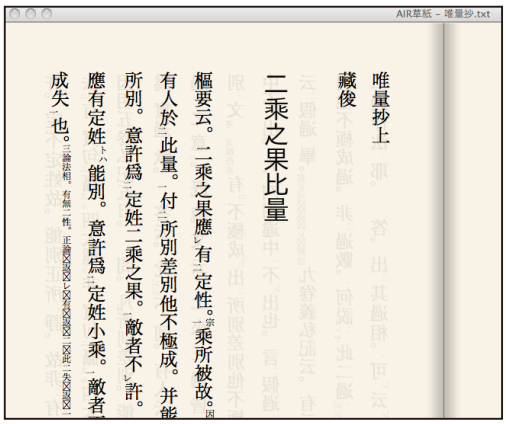



図 2：返り点入りテキストを AIR 草子で表示した例

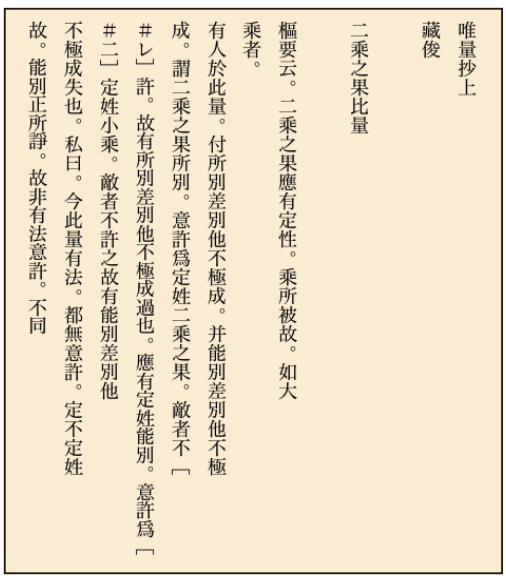
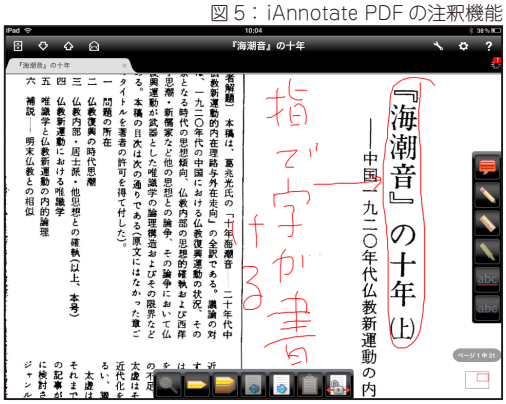

唯量抄上
藏俊
二乘之果比量
樞要云。二乘之果應有定性。乘所被故。如大乘者。
有人於此量。付所別差別他不極成。并能別差別他不極成。謂二乘之果所別。意許爲定姓二乘之果。敵者不「#レ」許。故有所別差別他不極成過也。應有定姓能別。意許爲「#二」定姓小乘。敵者不許之故有能別差別他不極成失也。私曰。今此量有法。都無意許。定不定姓故。能別正所諍。故非有法意許。不同

図 4：図 3 と同じテキストを i 文庫 HD で表示した例

図 5：iAnnotate PDF の注釈機能



きる。

GoodReader も同様の機能を盛り込む予定とのことで、今後、各アプリケーションで注釈機能が充実することが期待できる。

## ※ 学習プラットフォームとしての是非

近年、電子教科書、デジタル教科書という言葉を頻繁に目にするようになってきた。文部科学省が平成22 年 4 月より「学校教育の情報化に関する懇談会」を開催したが、そのなかで「デジタル教科書・教材」が検討課題としてとりあげて話題となった [5]。同様に IT 戦略本部第 53 回でも「児童 1 人 1 台の各種情報端末」「デジタル教科書・教材などの教育コンテンツの充実」が謳われている [6]。民間でもデジタル教科書教材協議会（DiTT）[7] が発足したりするなど、デジタル教科書をめぐる議論は活発化している。

このような動きに対しては脊髄反射的な批判 [8] も見られる一方、これをビジネスチャンスとうかがう企業もいれば、電子書籍に対する批判と同様「書店を殺す」というような危機感を煽る言説も見られ、いまだ評価は定まらないといった状況である。一方、海外や国内の複数の大学が学生に iPad の配布ないし貸与を発表し、「売名行為だ」という揶揄もあるが、大学教育での iPad の利用も現実味を帯びてきている。

いずれにせよ、iPad のようなデバイスを教育・学習の現場に導入する際には、そのメリット・デメリット、そしてコストの問題をよく検討すべきであろう。たとえば教育現場に iPad などが導入される際にしばしば「教員のやり方次第だ」といった（教員から見れば無責任にも見える）発言が見られるが [9]、教員が学生と読書会や討論会をしたり、生徒といっしょに課外活動をしたりといった他の時間を割いてまで iPad を使った授業の準備に時間をかけるべきなのか、あるいは人件費等の他の経費を削っても iPad 導入にコストをかけるべきなのか等々、論点はたくさんあるはずである。ここでは、そのための材料を提供できればと思う。

### ■ デジタル教科書デバイスとして

まず、狭い意味でのデジタル教科書のデバイスとしての iPad を考えてみたい。ここで「狭い意味」と述べたのは、iPad に代表される携帯端末を用いた「デジタル教科書」と呼ばれるものの射程が、従来の「教科書」よりも幅広いという点である。それは、たとえば先のデジタル教科書教材協議会が掲げる「3 つの目標」に、

1. どこに住んでいても世界中の知識に触れる機会を。
2. 創造力、表現力、コミュニケーション力を育む最高の環境を。
3. 友人、先生、家族とつながる手段を。

と書かれていることからもわかるだろう [10]。ここでは従来の教科書が持っていた機能を携帯端末上で実現しようとするものを中心に考えてみたい。

デジタル教科書のメリットとしてしばしばあげられているのが、教科書の中に静止画だけでなく動画や音声を埋めこむことで、学習者の理解を助けようという従来の視聴覚教育の延長線上にある発想である。これについては、「動画等による問題の解説が豊富になることによって、問題文、つまりテキストとして書いてある問題文から、自力で問題設定をイメージし、問題を解く力というのが弱まるのではないか」などといった批判的な意見も出されている [11]。いずれにせよ、この点については、従来の視聴覚教材、マルチメディア教材に関する議論を反復している印象がある。

一方、教科書のデジタル化独自のメリットとしては、学習者の能力や状況にあわせて教科書が表現形式を柔軟に変化させること可能になる点があげられるのではないかと思われる。適切に構造化された文書として教科書がデジタル化されれば、例えば以下のようなことが可能になるのではないだろうか。

- 視覚障害者・読字障害者に対する拡大表示や配色の変更、テキスト読み上げサポートなど
- 海外生活などで日本語のフォーマルスキーマを知らず、日本語は理解できるのに日本語教科書を読むことができない言語的弱者のためのサポート [12]
- 学習の進展に応じた補足情報（ふりがなや発音記号、英単語の意味など）や発展的内容の表示・非表示

これに加えて、検索機能なども学習者の助けになる

だろう。

ところで、通常の読書は一人対一冊が前提であるのに対して、教育や自主学習の場面では一人対複数冊という状況もあり得る。図書館の机に何冊も参考書や辞書をならべて、調べ物をしたりレポートを書いたりした経験は、インターネットやパソコンがまだそれほど発達していないかったころには、日常的なものであった。先に見たように iPad には書籍一冊分の大きさしかないため、小説などの一般的な電子書籍を読む場合には問題なかろうが、教科書のプラットフォームとしては不十分になる可能性もあろう。複数の書籍を閲覧するためには、文字通り「デスクトップ」的な広さが必要なのである。一冊の大きさしかない iPad で複数の教材を並べて利用できるようにするためのフォーマットやシステムを開発、あるいは iPad を前提とした教授法の開発などが必要になってくるかもしれない。

### ■ 学生・教員間のコミュニケーションツールとして

近年、教員と生徒・学生との間の双方向的な授業が「よい授業」の条件として言われるようになり、またグループワークのような授業での生徒・学生相互の学び合いにも注目が集まっている。先にみた「3つの目標」にも「友人、先生、家族とつながる手段を」と挙げられているように、iPad などの端末を授業で利用することにおいても、このようなコミュニケーションを重視した授業・学習・コラボレーションのためのツールになることが期待されている。

もっとも、この手の議論は決して新しいものではなく、インターネットが登場した頃から繰り返し主張されてきたことである。最近の学習管理システム（Learning Management System、LMS）にはたいてい教員と生徒・学生間のコミュニケーション機能が搭載されおり、人と人とのつながりをより重視した教育機関向けの SNS（Social Networking System）も開発されている。そしてこれらのシステムは、現代の学生のニーズに合わせて携帯電話に対応しているものがほとんどで、〝携帯端末によるネットワークを介した教員と生徒・学生間のコミュニケーション〟はすでにある程度実現されていると言って良いのではないかと思う。その意味で、iPad などを使うことでより双方向的あるいは協同的な新しいコミュニケーションが生まれるのかどうか、期待したい。

### 注

[1] ここでいう「研究者」は、大学教員のような狭い分野の人々を指すのではなく、何かのテーマに関心があり、情報の収集、分析、発信、教育活動などを日常的に行なっている人々のことを指す。

[2] **http://www.apple.com/jp/ipad/specs/**（2010 年 9 月 2 日最終確認）

[3] **http://www.ebookjapan.jp/ebj/**

[4] **http://www.mememomo.com/u/sato/f1892/**（2010 年 9 月 2 日最終確認）

[5] **http://www.mext.go.jp/a_menu/shotou/zyouhou/1292783.htm**（2010 年 9 月 2 日最終確認）

[6] **http://www.kantei.go.jp/jp/singi/it2/dai53/siryou.pdf**（2010 年 9 月 2 日最終確認）

[7] **http://ditt.jp**

[8] 田原総一朗『緊急提言！ デジタル教育は日本を滅ぼす』（ポプラ社、2010 年 8 月）など。

[9] たとえば小寺信良「先生が変わる、授業が変わる iPad がもたらす変化」（**http://plusd.itmedia.co.jp/mobile/articles/1008/30/news030.html**、2010 年 9 月 2 日最終確認）など。

[10] **http://ditt.jp/about/aim**（2010 年 9 月 2 日最終確認）

[11] **https://jukugi.mext.go.jp/archive/188.pdf**（2010 年 9 月 2 日最終確認）における新井紀子氏の発言。

[12] 中尾桂子・森下淳也「小学校教科書の学習者に応じた電子的再構造化について」（『情報処理学会研究報告』2003-CH-058、2003 年 5 月）

# iPad で見る、聞く、読む、学ぶ
## ――中国語教育を中心に――

清原　文代

## ❖ 1. はじめに

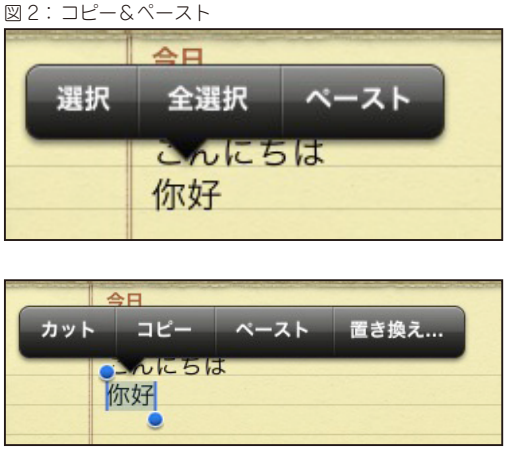
液晶画面を直接触って操作する iPad[1] は直感的な操作が可能で、起動にかかる時間も非常に短く、持ち歩ける学習端末として大きな可能性を持っている。

レビューに使ったモデルは iPad の Wi-Fi モデル（32GB）、OS は iOS 3.2.2 である。iPad は OS のバージョンアップによって機能が追加されていく。今年度中には OS のバージョンアップによってシングルタスクからマルチタスクへなどの大幅な機能向上が行われることがすでに予告されている。また紹介するアプリも同様にバージョンアップによって機能が向上したり、既存の機能に変更があることがある。したがって以下の内容は執筆時（2010 年 8 月）の状況であることをご承知おきいただきたい。

本稿でまずは iPad における文字入力、画面出力など入出力や表示機能を概観し、次に中国語を中心にモバイルラーニング端末としての iPad について見ていく。

図 1：日本語入力キーボード

図 2：コピー＆ペースト




## ❖ 2. 文字入力

### ■ 2.1 キーボード入力の〝作法〟の違い

iPad の液晶画面上のキーボードを使った入力は、パソコンのキーボード入力と若干勝手の異なる部分がある。例えば、パソコンで Undo は Windows であれば Ctrl+z、Mac OS ではあればコマンドキー＋ z であるが、iPad のキーボードの場合は、数字入力画面に切り替えると、そこに Undo 用のキーがある。これは液晶画面に表示されるキーボード上でキーコンビネーションを使って入力するのは難しいためだろう。

文字入力で頻繁に使うコピー＆ペーストも操作方法が異なる。文字をしばらく押さえていると、選択か全選択を選ぶウィンドウが文字のすぐ側にポップアップする。指でなぞって選択範囲を指定すると、次はカット、コピー、ペースト、置き換えといったような選択肢が表示される。

iPad は Bluetooth の外付けキーボードも使用することもできるが、iPad の機動性を犠牲にすることになり善し悪しである。先にパソコンのキーボード入力に親しんでいると、確かに最初のうちは若干とまどうが、要は慣れの問題だと思われる。

### ■ 2.2　入力可能な言語

「設定」→「一般」→「キーボード」→「各国のキーボード」で設定する。入力できる言語は以下の通りである。

- 英語
- イギリス英語
- イタリア語
- オランダ語
- スペイン語
- ドイツ語
- フラマン語
- フランス語
- フランス語（カナダ）
- ロシア語
- 簡体字中国語（手書き）
- 簡体字中国語（拼音）
- 日本語

### ●2.2.1 日本語入力

日本語についてはローマ字入力のみで、推測変換機能（最初の一文字で入力される単語を予想して候補が出現する機能）がついている。同じく iOS が搭載されている iPhone にある五十音を使ったテンキー入力はない。これは iPad の画面の大きさを考えれば、不必要という判断だろうか。確かに iPad を横位置で使えば、液晶画面上に表示されるキーボードを両手でタイピングすることも可能だ。筆者の体格は日本人男性の平均身長とほぼ同じなので、男性でも両手でタイピングが可能だろう。

### ●2.2.2　中国語入力

中国語入力については簡体字のみというのが非常に残念だ。iPhone の iOS は現在 ver.4 であるが、一つ前の ver.3 でも簡体字、繁体字両方が入力可能であった。なぜ iPad で繁体字入力が省かれているのかは謎である。iPhone（iOS 4）の入力方法は全部で 54 あり、それと比べると iPad には搭載されていない言語が多数ある。

中国語簡体字入力には、拼音入力と手書き入力がある。拼音入力は声調なしのローマ字のみ、推測変換に加え、連想変換（確定した単語の次に来る単語を入力前に予想して候補として表示）がある。手書きの方にも連想変換機能がある。

## ■ 2.3　単語登録機能は現時点では無し

日本語、中国語とも単語登録機能はない。一足先に

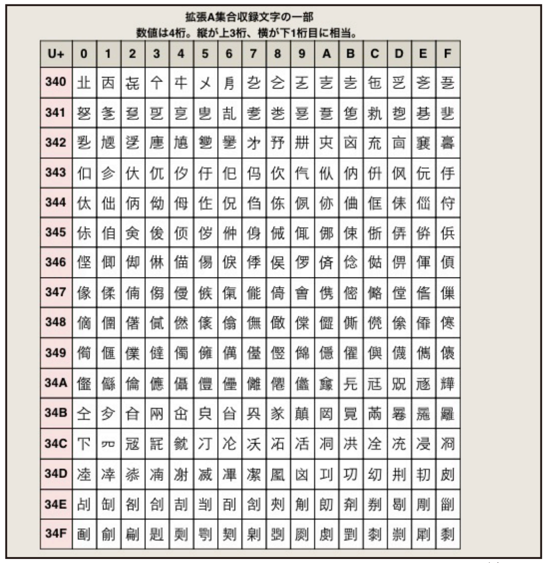



図 3：拡張 A

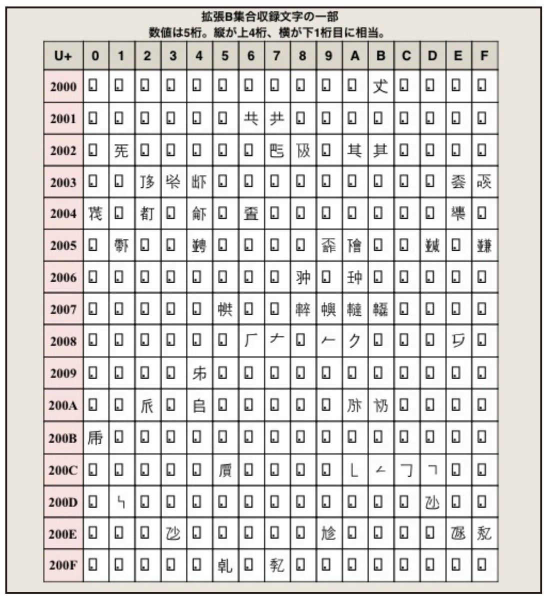



図 4：拡張 B

図 5：拡張 C

バージョンアップしている iPhone の iOS 4 では、単語登録が可能になっているので、iPad の iOS の次期バーションアップでは単語登録機能が搭載されることが期待される。

## ❖ 3. Web ブラウザ Safari

### ■ 3.1 Flash には非対応

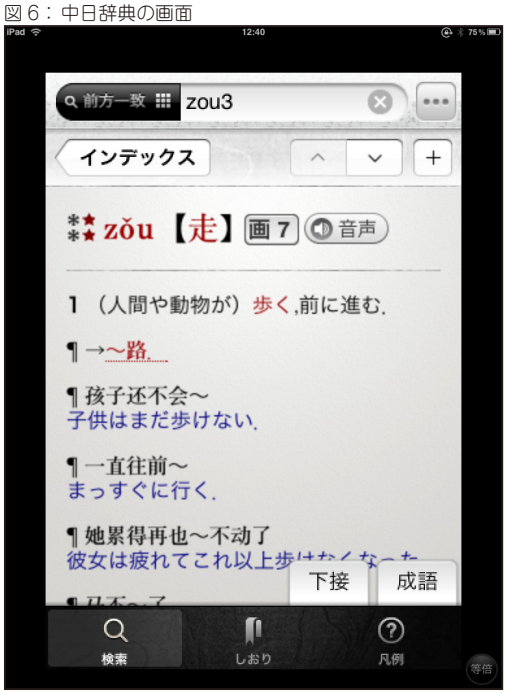
iPad には Web ブラウザの Safari が搭載されている。Flash が再生できないなど、パソコン用の Web ブラウザとは若干仕様が異なる部分があるが、その他はほぼパソコンの Web ブラウザと同等の機能を持っている。

### ■ 3.2 多漢字の表示

漢字文献情報処理研究会的視点で見れば、iPad がどの程度多漢字の表示に対応しているかも気にかかるところである。山田崇仁氏の Web サイトに Unicode CJK 統合漢字拡張 A、拡張 B、拡張 C のそれぞれ一部を取りあげた Web ページ [2] があるので、それを iPad の Safari で表示させてみた（図 3・図 4・図 5）。

入力については、拡張 A に含まれるいくつかの漢字を簡体字中国語手書きで入力してみたが、認識されなかった。拼音入力でも同様であった。

図 6：中日辞典の画面




## ❖ 4. 画面の外部出力

iPad には別売の専用 VGA ケーブルがある [3]。プロジェクタを使ってプレゼンテーションもできるのだが、パソコンとは異なり画面をそのままミラーリングするのではない。VGA 出力に対応した特定のアプリケーションの画面を出力するものである。VGA 出力に対応しているアプリをいくつか例として挙げる。

- iPad にプリインストールされているアプリ
  - ビデオ
  - 写真（スライドショーの再生）
  - YouTube
  - Safari（一部のビデオコンテンツ）
- 無料アプリ
  - CloudReaders[4]
- 有料アプリ
  - Keynote[5]
  - 2Screens ── プレゼンテーションの専門家 [6]
  - Good Reader for iPad[7]
  - i 文庫 HD[8]

## ❖ 5. モバイルラーニング端末としての iPad

iPad の大きさはおおよそ B5、厚さは約 1.3 センチ、重さは約 700 グラムである。もちろん携帯電話ほど気軽ではないが、持ち歩くのはさほど負担ではなく、ノートパソコンを持ち歩くのに比べれば格段に楽だ。iPad は携帯電話に比べて画面が大きい分、文字が見やすい。ノートパソコンの場合、出先でのバッテリー切れが常に気になるものだが、iPad のバッテリーは最長 10 時間持つ（公称値）。このようなハードウェアの特徴はモバイルラーニング端末に向いている。以下、筆者の専門である中国語教育を中心にモバイルラーニング端末としての iPad を見ていく。

### ■ 5.1 中国語電子辞書アプリ

日本語を母語とする中国語学習者にとって待望の中国語辞書アプリが発売された。小学館の『中日辞典』『日中辞典』（いずれも第二版）を 1 つにまとめた「中日・

日中辞典」[9] である。例文検索がないなど機能面で単体の IC 電子辞書にかなわない部分もあるが、電子辞書として基本的な機能は備わっている。

価格は 4100 円と iPhone アプリとしては高額の部類に属するが、紙の『中日辞典』『日中辞典』を 2 冊買うことに比べれば約 3 分の 1 だ。紙の辞書にはないネイティブスピーカーの音声（中国語の見出し語 1 万 6 千語）もついている。

このアプリは iPhone 用であるが、iPad でも使用できる [10]。iPad 用にアプリ画面を 2 倍に拡大すると表示に若干ジャギーが出るが、読むのに支障があるほどではない。iPhone より広い画面で使える分、操作性は iPad で使用する方が良いように感じる。

なお、このアプリは容量が大きいので、iPad の 3G 回線（ソフトバンクと契約する携帯電話回線）では直接購入することはできない [11]。パソコンにインストールした iTunes[12] の中にある AppStore のコーナーで購入し、パソコンの iTunes と iPad を同期してインストールする。

### ●4.1.1　検索方法

#### ・(1)　見出し語検索は 3 モード

- ●前方一致曖昧検索（ひらがな・カタカナ、清音・濁音、直音・拗音・促音、大文字・小文字、全角・半角などを気にせず検索する）
- ●前方一致厳密検索（入力した文字どおりに検索する）
- ●部分一致検索

#### ・(2)例文検索は不可

残念ながら検索は見出し語のみで、例文検索には対応していない。例文検索は紙の辞書にはできない電子辞書ならではの機能なので、ぜひ対応してほしい。

#### ・(3)　検索語の入力

検索欄は一つで、日本語を入力すれば日中辞典を、中国語を入力すれば中日辞典を検索する。「学校」のように日中同形語で字体も同じ場合は、両方を検索する。

中国語については拼音入力と手書き入力ができる。いずれも簡体字のみ。拼音入力は声調有り無しいずれも可で、声調入力のためのボタンが画面上に表示される。声調は「hao 三声 ren 二声」のように音節ごとにつけていく。

簡体字手書き入力は iPad が内蔵しているものを使用する。日本で購入した iPad は最初は簡体字手書き入力が ON になっていないので、自分で設定を変えて追加する必要がある。

拼音入力も手書き入力もインクリメントサーチに対応しているので、1 文字入力していくごとに自動的に検索が始まる。

### ●5.1.2　リンクとジャンプ

#### ・(1)　リンク

関連語彙にはあらかじめリンクが貼られていて、点線の下線が施されている。「中日・日中辞典」は紙版のセールスポイントの一つであるコラム欄も収録しており、当該単語の本文の末尾からコラム欄へのリンクが貼られている。

#### ・(2)　ジャンプ機能

ジャンプ機能は辞書の引き直しを容易に行える機能で、電子辞書には欠かせない。例えば、作文のために日中辞典を検索し、そこで見つかった中国語の単語の詳しい用法を確認するために、中日辞典を引き直すといったことが簡単にできる。

「中日・日中辞典」では、あらかじめリンクが貼ら

図 7：コラム欄へのリンク

図８：ジャンプ機能

れていない部分でも、指で触れて単語を選択すると「コピー」と「この場で検索」という選択肢が現れる。「この場で検索」を選ぶと、当該単語を検索する。「中日・日中辞典」は IC 電子辞書にありがちなジャンプボタンなどを押してから操作するのではなく、いきなり指で単語に触れて操作する。指で直接画面を触って操作する iPad のインターフェースを生かした操作体系である。

●5.1.3　検索履歴としおり機能

検索履歴は自動的にしおりのコーナーに保存される。任意の単語にしおりをつけることもできる。検索履歴は一括削除のみ。しおりは個々に削除が可能である。

## ■ 5.2　中国語学習アプリ

iPad で使用できる中国語学習用のアプリは無料、有料、様々なものがある。パソコンの iTunes の iTunesStore、iPad の AppStore で「中国語」や「Chinese」などをキーワードにして検索すると数多く見つかる。筆者はその全てを把握しているわけではないので、筆者が実際ダウンロードしたものをいくつか挙げておく。なお、アプリの価格はしばしば改訂される。無料のものが有料になることもあれば、有料のものが無料になることもあるので、AppStore でダウンロードする前に必ずご自身で確認していただきたい。

●アルクの動画でニイハオ 220 ～シーンで覚える中国語・多言語字幕付き～ [13]

「ネットアカデミー２中国語」の会話部分をアプリ化したもの（有料）。字幕は簡体字中国語・日本語・英語、字幕無しの設定もできる。VGA 出力に対応しているので、プロジェクタに投影して学生に見せるといった使い方も可能だ。

●Pinyin Trainer by trainchinese[14]

拼音を憶えるためのアプリ。音声を聞いて、選択肢の中から正しい拼音を選ぶ、声調符号を手書きで入力するといったドリルである。無料だが追加の有料コンテンツ有り。

●Chinese Number Trainer by trainchinese[15]

中国語の数字の読み方を訓練するアプリ。入力した数字を中国語で読み上げたり、音声を聞いて選択肢の名から正しい数字を選ぶ。無料だが追加の有料コンテンツ有り。

●パンダと学ぶ中国語旅行会話編 [16]

長崎外国語大学の三枝裕美氏の開発したアプリ（無料）。三枝氏の Web 教材 [17] を iPhone アプリ化したもの。iPhone 用アプリだが、iPad で使用できる。中国語教員自身による中国語学習アプリの開発ということで注目される。三枝氏は他にもパンダと学ぶ中国語シリーズを iPhone アプリ化しており、いずれも AppStore から無料でダウンロードできる。

## ■ 5.3　中国語圏の今を知るためのアプリ

中級レベルになれば、学習者向けの教材だけではなく、中国の生の素材に触れる機会を増やしたい。iPhone、iPad アプリの中には中国語圏からのものもあり、その中から放送やニュースを扱った無料アプリの例を挙げる。筆者が中国語を学び始めた頃、日本で中国のテレビ放送を見ることは容易ではなかった。そのことを思えば、画像はやや荒いところもあるが、iPad で無料で中国のテレビ放送を見られるのは隔世の感がある。

・動画ニュースアプリ

回線状態によってはつながりにくいことがある。

●CCTV mobile television[18]

中央電視台が見られる。iPhone アプリだが、iPad でも使用可能。

●新华社电视 [19]

新華社通信による動画ニュースが見られる。iPhone アプリだが、iPad でも使用可能。

●凤凰移动台 HD[20]

香港を拠点にした中国語（普通话）衛星放送である凤凰卫视の iPad 専用アプリ。

・文字によるニュースアプリ

●新民网 [21]

上海の夕刊紙《新民晚报》のニュースアプリ、iPhone 用だが、iPad でも使える。

●iNFZM News Reader ── 南周阅读器 [22]

広州を中心とした週刊誌《南方周末》のニュースアプリ、iPhone 用だが、iPad でも使える。

## ■ 5.4　Podcast、iTunes U

5.2 及び 5.3 で紹介したのはアプリであるが、iPad には無料で利用できる教育コンテンツが他にも膨大にある。それが Podcast と iTunes U である。

### ●　5.4.1　Podcast

・(1)　ポッドキャストとは？

ポッドキャストはダウンロード型のインターネット放送である。ポッドキャスト自体はアップル社が始めたものではなく、自然発生的に始まったものだが、iTunes Store の中にはポッドキャストのポータルが設けられており、無料で検索・視聴・ダウンロードが可能である。ポッドキャストの内容は娯楽、ニュース、教育など多岐に渡る。iTunes Store の中にある Podcast ではカテゴリー分けがなされており、教育カテゴリーには外国語学習番組が多数登録されている。英語ほど多くはないが、中国語学習番組もある。以下に日本から配信されている中国語学習ポッドキャストの例を挙げる。もちろん中国語圏から配信されている中国語学習番組もある [23]。

●徹底学習中国語会話 ── Chinese Station3[24]

沈国威、氷野善寛、紅粉芳惠、関西大学中国語教材研究会によるビデオ＋音声＋ PDF ポッドキャスト。

●Chinese Station 2 ── ポッドキャストで中国語 [25]

関西大学中国語教材研究会によるビデオポッドキャスト。

●ガチャピン＆ムックの中国語教室 [26]

有名な幼児番組のキャラクターが中国語に挑戦している音声ポッドキャスト。

図 9：iPad の iPod アプリ。ダウンロード済みの Podcast の番組を表示したところ。

●大阪府立大学中国語講座 ── ぜひ憶えておきたい 100 の表現 ── [27]

清原文代、顧春芳による音声ポッドキャスト。

●大阪府立大学中国語ポッドキャスト 旅行会話編 [28]

清原文代、顧春芳による音声＋ PDF ポッドキャスト。

●接客のための中国語 [29]

清原文代による音声＋ PDF ポッドキャスト。

・(2)　iPad におけるポッドキャストの視聴方法

iPad におけるポッドキャストの視聴方法は 2 つある。

(A) パソコンの iTunes にダウンロードしてから、iPad に同期して転送する。

(B)iPad の iTunes アプリで iPad にダウンロードする。

(A) はパソコンの iTunes でポッドキャスト番組を登録して無料購読する方法である。購読すると、インターネットに接続していて iTunes が起動していれば、更新分がパソコンに自動的にダウンロードされる。その後、パソコンの iTunes と iPad を同期してダウンロードされたファイルを iPad に転送する。転送された動画・音声は iPad の iPod アプリで再生する（図 9）。

ポッドキャストは動画・音声だけでなく PDF も配信できる。ポッドキャストで配信された PDF は、アップルの無料アプリ iBooks[30] に自動的に転送されて読むことができる（図 10）[31]。但し、本稿の初稿提出後に iTunes が ver.10 にバージョンアップした。ver.10 ではポッドキャストで受信した PDF が自動的に iPad に転送されない。これが恒久的な仕様変更なのか、それともバージョンアップに伴う不具合なのかは現時点では不明である。ver.10 の場合、iTunes の Podcast で受信した PDF をドラッグ＆ドロップで一旦デスクトップにコピーし、デスクトップにできた PDF を iTunes の「ブック」にドラッグ＆ドロップして入

図 10：iBooks の PDF 表示。2 本指でピンチして文字の拡大縮小ができる。

図 11：iPad の iTunes アプリで Podcast を表示したところ

れれば、iPad に転送することができる。

(B) は iPad だけで直接動画と音声を視聴し、ダウンロードできるが、PDF は見られない。また、動画や音声ファイルを個々にダウンロードすることは可能だが、パソコンの iTunes と異なり、番組の自動継続配信を受けるための無料購読ボタンは用意されていない。ダウンロードした音声や動画は iPad の iPod アプリで再生する。

ポッドキャストを本格的に継続利用するのでれば、(A) の方法が良いが、(B) は iPad だけで動画や音声をダウンロードし再生できるので簡便である。

ポッドキャストの作成や配信方法については、清原（2006）・清原（2008a）・清原（2008b）を参照されたい。

### ●5.4.2　iTunes U

#### ・(1)　iTunes U とは？

iTunes U はアップル社が提供するサービスである[32]。アメリカを中心とする大学などの教育機関の講義、公開講座を無料で視聴、ダウンロードできる。

#### ・(2)　iPad における iTunes U の視聴方法

iPad における iTunes U の視聴方法は 5.4.1 の Podcast と同様である。パソコンの iTunes の iTunesStore の中にある iTunes U からダウンロードしてから iPad に同期して転送するか、或いは iPad の iTunes アプリの中にある iTunes U から iPad に直接ダウンロードする。

これまではアジアの大学は参加していなかったが、最近日本、中国、香港の大学が加わった。現時点で参加しているのは、東京大学[33]、慶應義塾大学[34]、早稲田大学[35]、明治大学[36]、北京广播电视大学[37]、中山大学[38]、香港中文大學[39]、香港公開大學[40]である。

#### ・(3)　Podcast と iTunes U の違い

5.4.1 で紹介した iTunesStore の Podcast との違いは、iTunesStore の Podcast のコンテンツは各配信者のサーバ上にあり、iTunesStore はそこへのリンクを提供しているだけである。iTunes U はアップルがコンテンツを管理するためのシステムを提供している。

iTunesStore の Podcast では配信者自身がカテゴリーを決めているため、教育カテゴリーに入っているポッドキャストであっても、その配信者は教育機関とは限らない。iTunesStore の Podcast では、個人から

公的機関まで多様な配信者による多彩なコンテンツから教育リソースを探すことができる。一方、iTunes U では教育機関がコンテンツを提供するため、一定の質が保たれていると考えられる。

### ■ 5.5 教材を自作する

iPad で利用できる教材には、上述のように無料、有料様々なアプリ、無料で利用できる Podcast や iTunes U があるが、教員自身が iPad 用の教材を自作することはできるのだろうか？

教員自身が iPad 用の教材を作成し公開するには、まず以下のような方法が考えられる。

(A) iPhone Developer Program[41] に参加し、iPhoneSDK を使ってアプリを作成し、AppStore に登録する。
(B) HTML、CSS、JavaScript を使って、Web アプリ[42] を作って、Web 上で公開する。

いずれもプログラミングのできない筆者にとっては技術的に敷居が高い方法でちょっと手が出せない。(A) や (B) で制作可能なインタラクティブ性のある教材でなくてもかまわないので、iPad で閲覧可能な、テキストに音声や動画がついたデジタル教材を作る方法はないか？

その方法として、筆者は PDF と EPUB を提案する。現在のところ後述するように iPad における再生環境に若干の制限はあるものの、作成のためにプログラミング技術は必要ない。ファイルの規格にも栄枯盛衰があるが、PDF と EPUB は両者ともオープンな規格であり、将来新しい規格のファイルが主流になった場合でも、コンバータが用意されるなどして、データが再利用できる可能性が高いことも、教材を作るファイルとして適している。

#### ●4.5.1 PDF

・(1) PDF は紙と電子教材の両方 OK

PDF は OS の壁を越えて文書を同じように表示する規格として知られており、公文書などの配布によく使用されている。PDF は紙にも同じ版面で印刷できることができるのが大きな特長である。それだけでなく、PDF には Web へのリンクを含むことができ、更に音声や動画などのマルチメディアファイルを PDF に埋め込むこともできる。

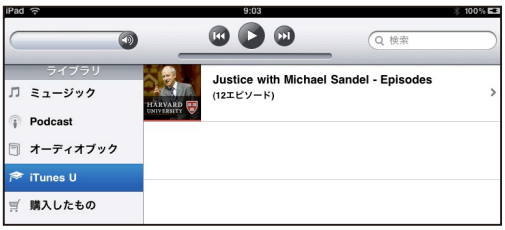



図 12 : iPad の iPod アプリ。ダウンロード済みの iTunes U の番組を表示したところ。

図 13 : iPad の iTunes アプリで iTunes U を表示したところ

紙の教材への需要はそうそうなくなることはないと考えられるので、同じ版面を持つデジタル教材と紙の教材を一つのファイルから作成可能な PDF はワンソース・マルチユースを実現するファイルであると言える。

・(2) テキストと URL を含む PDF の作成方法

テキストと URL を含む PDF の作成は容易である。Mac OS X は OS 自身が PDF への書き出し機能を持っており、各ソフトウェアの印刷画面から PDF の書き出しが行える。Windows の場合、Microsoft Office 2007 に PDF 書き出し機能がある[43]。

・(3) 音声や動画を含む PDF の作成方法

できあがった PDF に音声や動画などを埋め込むに

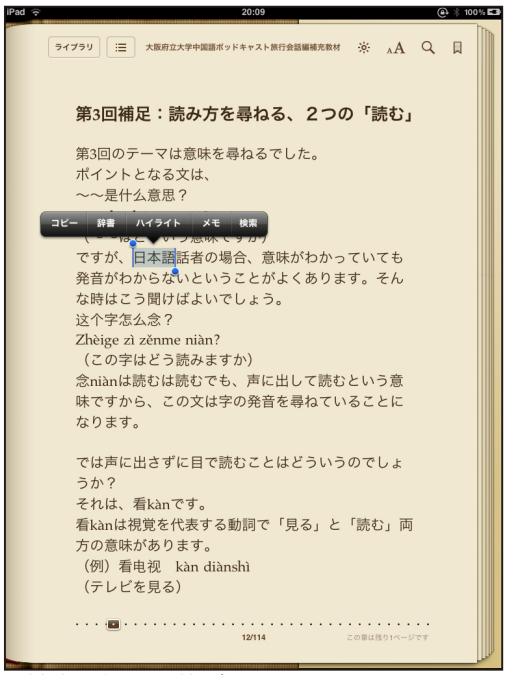
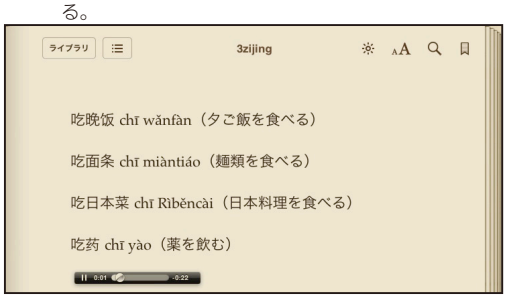


図 14 : iBooks の EPUB 表示

図 15 : 動画を含む EPUB の試作例、動画の全画面再生も可能。

図 16 : 音声を含む EPUB の試作例、音声の場合は再生を始めてから iPad をスリープさせても再生が続き、EPUB を一種のオーディオブックとしても使用できる。




は Acrobat Pro[44] を使用する。Acrobat Pro は安価なソフトではないが、幸いアカデミック版が用意されている [45]。音声や動画を埋め込んだ PDF 教材の作成については清原（2008c）及び清原（2010）をご参照いただきたい。

・(4)　iPad における PDF の閲覧と再生方法

iPad アプリには PDF ビューワーはいろいろあるものの、残念ながら PDF に埋め込んだ音声や動画まで再生できるアプリは今のところ存在しない。そのため音声や動画をネット上に置き、PDF の中にその URL を記述するという手段を取るしかない。音声や動画がネット上にある以上、再生時にネット接続は必須である。

iPad アプリで、PDF に記述された Web へのリンクが有効なものは以下の通りである。

- iBooks[46]（無料）
- GoodReader for iPad[47]（有料）

iBooks は PDF の中の URL をタップすると、一旦 iBooks を閉じて Web ブラウザの Safari で開く。再びテキストに戻るにはもう一度 iBooks を起動せねばならない。iBooks は読んでいたページを記憶しているものの、音声や動画を再生するためには一手間多くなる。

Good Reader for iPad は PDF 以外に多くの種類のファイルの閲覧・再生機能を持っており、同アプリの中だけでテキストの閲覧、音声や動画の再生、Web ページに閲覧が完結する。有料といっても非常に安価なアプリであるので（執筆時の価格は 115 円）、現時点では PDF 教材を見るには最適なアプリと言えるだろう。

### ●4.5.2　EPUB

・(1)　EPUB とは？

Google やアップル、SONY が電子書籍の規格として採用したため、にわかに注目を浴びているのが EPUB である。EPUB の特長は端末の画面の大きさに合わせてテキストが配置されることで、iPhone のような小さな画面でも、iPad のような大きな画面でも読みやすい。

EPUB のテキスト部分は XHTML で書かれており、その他必要なファイルを一つのフォルダに入れ、ZIP

で圧縮して拡張子を epub としたものである。

・⑵　EPUB の作成方法

パソコンで EPUB 作成するには、以下のようなものがある。

- Sigil[48]（無料、Windows 版、Mac OS X 版、Linux 版）
- calibre[49]（無料、Windows 版、Mac OS X 版、Linux 版）
- Pages[50]（有料、Mac OS X 版のみ）
- パブー[51]（無料、Web サービス）

Sigil は既存の HTML 書類を読み込んで EPUB に変換する機能もあるが、基本的には一から EPUB を執筆するためのソフトである。EPUB のテキスト部分は XHTML で書かれているが、Sigil はそのソースを表示し編集することもできる。

calibre は既存の HTML、RTF、PDF などを読み込んで EPUB に書き出す機能を持っている。但し、XHTML のソースは編集できない。

Pages は iWorks09 という Mac OS X 用オフィススイートに含まれるワープロソフトであるが、最近のアップデートで EPUB の書き出しに対応し、EPUB 用のサンプルファイルも公開された[52]。

パブーはブログを作成する要領で Web 上でテキストを書き込み、それを EPUB と PDF に無料で変換できるサービスである。PDF についてはどうやらパブーの仕様の関係で中国語簡体字が正しく表示されないが、EPUB は問題ない[53]。出力した EPUB や PDF は、手数料を払えばパブーで有料販売することもできる。

・⑶ EPUB のポッドキャスト配信

パブーは EPUB のポッドキャスト配信にも対応している[54]。アップルが正式に認めているわけではないが[55]、iTunes ver.9.2.1 及び ver.10.0 は EPUB のポッドキャスト配信に対応しているようで、それを利用した EPUB のポッドキャスト配信が始まっている[56]。筆者も実験してみたが、ポッドキャスト配信のための RSS の記述で、enclosure の type 属性を application/epub+zip とすればパソコンの iTunes で受信できた。

・⑷ iPad における EPUB の閲覧と再生方法

EPUB はパソコンの iTunes のブックに EPUB を入れて iPad と同期して転送する。なお、ポッドキャスト配信された EPUB は自動的にパソコンの iTunes の中のブックに入る[57]。

iPad では iBooks を使って EPUB を読む。iBooks の EPUB 閲覧機能は単にテキストを表示するだけでなく、検索機能、しおり機能、任意のテキストをハイライトする機能、任意のテキストにメモをつける機能、辞書引き機能[58]がある。しおりやメモ、ハイライトは紙の本の再現を目指した機能、検索や辞書引きは電子書籍ならではの機能である[59]。

・⑸音声や動画を埋め込んだ EPUB の作成方法

現在の EPUB の規格は、横書きのテキストと挿絵を表示するためのものである。EPUB への音声や動画の埋め込みは EPUB の次期の規格として検討されている最中である。

ただ、現時点では正式な規格ではないのだが、アップルは HTML5 の audio タグや video タグを使用する形で音声や動画を埋め込んだ EPUB を、iBooks で再生する機能を実装している。

・⑴有料ソフトの Pages を使用する

Mac OS X で音声や動画を含む EPUB を作るには Pages を使用するのが最も簡単である。

① Page の文書作成モードで書類を作る。レイアウトモードでは EPUB に書き出せない。Pages は Word の書類を読み込むことができるので、既存の Word 書類の利用も可能である。
② 音声や動画ファイルを書類にドラッグして埋め込む。
③ 音声や動画、図などがフローティング（テキストと共に移動しない）のままでは EPUB には書き出せない。当該の音声や動画、図をクリックして選択し、メニューバーにある「インスペクタ」の「回り込み」という設定を使って、インライン（テキストと共に移動）にする。
④ ファイルメニューの「書き出す…」から EPUB に書き出す。

・⑵無料ソフトのみで作る

Windows の場合や、Mac OS X でも有料の Pages を使用しない場合の制作手順はおよそ以下の通りである。

① Sigil などで EPUB のテキスト部分を作って、EPUB に書き出す。
② ①の拡張子を zip に変更し、解凍する[60]。

③ ②でできたフォルダに音声や動画ファイルを入れる。

④ ②のフォルダの中にある XHTML ファイルを適当なエディタで開けてソースを編集して、音声や動画ファイルへのリンクを書く。
音声のリンクの例、ゴシック部分がファイル名である。

```
<audio src="Ch001.m4a" controls="true"></audio>
```

動画へのリンクの例、ゴシック部分がファイル名である。

```
<video src="01-chijiaozi.m4v" controls="true"> </video>
```

⑤ フォルダを再び ZIP で圧縮し、拡張子を epub とする。その際にはフォルダの中にある mimetype[61] というファイルは圧縮せずに、フォルダ全体を ZIP で圧縮する [62]。

動画や音声を含む EPUB は、現時点では Mac OS X で Pages を使えば簡単に作成できるが、そうでなければ上記のように EPUB に音声や動画を埋め込むのに手間がかかる。更に再生環境が iBooks に限られるという欠点もあるが、マルチメディアを含む EPUB を教員が自作し、iPad で閲覧・再生することはすでに可能になっている。

## ❖ 5. おわりに

iPad はタブレット型コンピュータと言われるが、ノートパソコンの代用品になるものではない。そこを取り違えると、iPad に対して「あれもできない、これもできない、できるけれど不便だ。」といったような失望感や不満を抱くことになるだろう。もちろん iPad でメールを書く、twitter に書き込みをする、パソコンで作ったファイルの一部手直しをするといったことはできるが、論文を書くなど一から本格的に何かを作るのであれば、やはりパソコンを選択すべきである。ただ、そこを割り切ってしまえば、iPad は液晶画面に直接触れて直感的に操作できる非常に優れたビューワーであり、モバイルラーニング端末である。また、制限はあるものの、教員が音声や動画を含む iPad 用デジタル教材を自作する道も開けている。

デジタル教材の場合、どの機種を選ぶにしろ、コンテンツに加えて電気で動く端末が別に必要である。それに対して紙の教科書はモバイルラーニング端末としては、音が出るわけでもなく絵が動くわけでもなく、そしてインタラクティブでもなく非常に地味である。しかし、紙はコンテンツと再生装置と記憶装置（メモ、書き込み）が一体化した簡便で取り扱いやすい端末であり、そうそうその役目を終えるとは思えない。その一方で、人類がよりリッチなコンテンツを求めてきたという流れからすれば、デジタル教材への流れが止まるとも思えない。今後デジタル教材を再生する端末として何が普及するのか？文字通り日進月歩である ICT の世界の先を読むことは難しい。しかし、デジタル教材を再生するためのモバイルラーニング端末を考える時、iPad がその参照点となることは間違いないだろう。

### 注

[1] **http://www.apple.com/jp/ipad/**

[2] **http://www.shuiren.org/chuden/teach/code/main8.htm**

[3] **http://store.apple.com/jp/product/MC552ZM/A**

[4] **http://itunes.apple.com/jp/app/cloudreaders-pdf-cbz-cbr/id363484920?mt=8**
PDF や ZIP で圧縮した JPEG ファイル等を見られるビューワー。外部出力中も iPad のアプリ画面を表示するが、但し iPad のアプリ画面をピンチして拡大縮小しても外部出力画面には反映しない。ポインタ機能は無し。

[5] **http://itunes.apple.com/jp/app/keynote/id361285480?mt=8**
プレゼンテーション作成＆再生アプリ。外部出力中の iPad のアプリ画面はページ数が示されるだけになるため、外部出力画面を見ながらのプレゼンになる。ポインタ機能があり、追随速度は上々。

[6] **http://itunes.apple.com/jp/app/id370913954?mt=8**
既存のファイルを読み込んでプレゼンテーションするためのアプリ。外部出力中も iPad のアプリ画面を表示する。iPad のアプリ画面をピンチして拡大・縮小すると外部出力画面も追随。ポインタ機能もあるが追随速度がやや遅い。

[7] **http://itunes.apple.com/jp/app/goodreader-for-ipad/id363448914?mt=8**

PDF を初め様々な種類のファイルに対応したビューワー。外部出力中も iPad のアプリ画面を表示する。iPad のアプリ画面をピンチして拡大・縮小すると外部出力画面も追随。ポインタ機能は無し。

[8] http://itunes.apple.com/jp/app/id369111608?mt=8
青空文庫、プレーンテキスト、PDF を見られるビューワー。縦書き表示対応。外部出力は iPad を横位置にした状態のみ。外部出力中は iPad のアプリ画面は黒くなる。ポインタ機能無し。

[9] http://itunes.apple.com/jp/app/id376159451?mt=8

[10] カメラなど iPhone 固有のハードウェアを使用するものや、制作者側が iPad での使用を制限しているアプリを除き、iPhone アプリは原則的には iPad でも使用できる。iPhone と iPad では液晶画面の大きさが異なるので、そのままでは小さく表示されてしまうが、iPad の画面に合わせて 2 倍に拡大する機能がある。

[11] iPad の 3G 回線で購入できるのは 1 つのアプリあたり 20MB まで。それ以上大きなものは無線 LAN を使ってダウンロードするか、パソコンの iTunes を経由する。

[12] http://www.apple.com/jp/itunes/download/

[13] http://itunes.apple.com/jp/app/id387681860?mt=8

[14] http://itunes.apple.com/jp/app/pinyin-trainer-by-trainchinese/id376797304?mt=8

[15] http://itunes.apple.com/jp/app/chinese-number-trainer-by/id378767080?mt=8

[16] http://itunes.apple.com/jp/app/id374194454?mt=8

[17] http://saigusa.com/

[18] http://itunes.apple.com/jp/app/cctv-mobile-television/id331259725?mt=8

[19] http://itunes.apple.com/jp/app/id379043540?mt=8

[20] http://itunes.apple.com/jp/app/id379224841?mt=8

[21] http://itunes.apple.com/jp/app/id343790302?mt=8#

[22] http://itunes.apple.com/jp/app/id333953812?mt=8#

[23] ChinesePod、CSLPod、iMandarinPod、ChineseClass101 など。

[24] http://itunes.apple.com/jp/podcast/id374453946

[25] http://itunes.apple.com/jp/podcast/id265275552

[26] http://itunes.apple.com/jp/podcast/id163794678

[27] http://itunes.apple.com/jp/podcast/id146197712

[28] http://itunes.apple.com/jp/podcast/id251090043

[29] http://itunes.apple.com/jp/podcast/id312365350

[30] http://itunes.apple.com/jp/app/ibooks/id364709193?mt=8
iBooks は iPad 出荷時にはインストールされていないので、自分で AppStore からダウンロードしてインストールする必要がある。

[31] ポッドキャストすでにダウンロード済みの PDF は転送されない。一旦当該のポッドキャスト番組を iTunes から削除し、登録しなおして、ファイルを再ダウンロードし、iPad と同期すれば転送される。但し、番組によっては過去の PDF をすでに削除していることもあるので、ポッドキャスト番組を削除するときには注意が必要である。

[32] http://www.apple.com/jp/education/itunes-u/what-is.html
アップルは 2010 年 8 月に iTunes U におけるダウンロード件数が開設後 3 年で 3 億件を超えたと発表している。
http://www.apple.com/jp/news/2010/aug/25itunes.html

[33] http://itunes.apple.com/WebObjects/MZStore.woa/wa/viewiTunesUInstitution?id=382733273

[34] http://itunes.apple.com/WebObjects/MZStore.woa/wa/viewiTunesUInstitution?id=383151528

[35] http://itunes.apple.com/WebObjects/MZStore.woa/wa/viewiTunesUInstitution?id=383298529

[36] http://itunes.apple.com/WebObjects/MZStore.woa/wa/viewiTunesUInstitution?id=384120439

[37] http://itunes.apple.com/WebObjects/MZStore.woa/wa/viewiTunesUInstitution?id=380619263

[38] http://itunes.apple.com/WebObjects/MZStore.woa/wa/viewiTunesUInstitution?id=380397310

[39] http://itunes.apple.com/WebObjects/MZStore.woa/wa/viewiTunesUInstitution?id=384989860

[40] http://itunes.apple.com/WebObjects/MZStore.woa/wa/viewiTunesUInstitution?id=382733276

[41] http://developer.apple.com/jp/programs/iphone/

[42] アップルによる Web アプリの作例紹介がある。
http://www.apple.com/webapps/index.html

[43] http://office.microsoft.com/ja-jp/infopath-help/HA010064992.aspx?CTT=5&origin=HA010074224

[44] http://www.adobe.com/jp/products/acrobat/

[45] http://www.adobe.com/jp/education/

[46] 注 [30] に同じ。

[47] 注 [7] に同じ。

[48] http://code.google.com/p/sigil/

[49] http://calibre-ebook.com/

[50] http://www.apple.com/jp/iwork/pages/

[51] http://p.booklog.jp/

[52] http://support.apple.com/kb/HT4168?viewlocale=ja_JP

[53] 筆者がパブーを使って作った日本語、中国語簡体字、拼音が混在したEPUBのサンプルは以下のところにあり、無料でダウンロード可能である。音声については、パブの書き出し方法ではEPUBに埋め込むことはできず、ネットに置いた音声ファイルへのリンクの形で処理している。
http://p.booklog.jp/users/qingyuan

[54] http://booklog.jp/puboo/podcast

[55] http://www.apple.com/jp/itunes/podcasts/specs.html
アップルによれば、配信できるファイルは、「m4a」、「mp3」、「mov」、「mp4」、「m4v」、「pdf」である。

[56] http://tablemark.whizzo.jp/

[57] パソコンのiTunesのブックの中に入ったPDFとEPUBは両方とも同じ閉じた本のアイコンで示される。PDFはダブルクリックすると、Adobe Readerなどの PDF閲覧ソフトが自動的に起動して中身が表示されるが、EPUBはダブルクリックしても開かない。Sigilなどパソコン用のEPUBソフトを使って自分で開かなければならない。

[58] iBooksの辞書引きについては、初めて辞書を引いた時に辞書のダウンロードが始まる。辞書は国語辞典、英英辞典、英和辞典があり、どの辞書を引くのかは、選ばれている単語とEPUBの言語設定によって自動的に決定され、自分で選ぶことはできない。

[59] iBooksのPDF閲覧機能では、しおり機能と検索機能が使えるのみである。

[60] Mac OS Xでは、OSの持つZIP解凍機能では拡張子を変えてもうまく解凍できなかったが、StuffIt Expanderを使用するとうまく解凍できた。
http://www.act2.com/products/stuffit2010mac-exp.html

[61] mimetypeというファイルは、このZIPファイルがEPUBであることを表すもので、このファイルが他のファイルと一緒に圧縮されてしまうと、EPUBとして認識されない。

[62] Mac OS XのFinderのZIP圧縮機能ではフォルダ内の特定のファイルのみを除外してZIP圧縮をすることはできない。Mac OS Xのユーティリティの中にあるターミナルを起動して、コマンドラインでZIP圧縮を行う。以下はコマンドラインの例。ゴシック部分ができあがるファイルの名前である。

```
zip -0 3zijing.epub mimetype
zip -r 3zijing.epub * -x mimetype
```

Windowsの場合もコマンドラインによる圧縮を行う。EPUBの作成方法については、横浜工文社のWebページ「日本語Epubブックサンプル」に詳しい。
http://www.kobu.com/docs/epub/index.htm

## 参考文献


- 清原文代、2006、「ポッドキャストで中国語」、『漢字文献情報処理研究』第7号、好文出版、pp.100-109
- 清原文代、2008a、「Podcastによる多言語音声教材およびテキスト教材の配信――大阪府立大学の事例を中心に」、吉田晴世・上村隆一・野澤和典・松田憲編著CIEC外国語教育研究部会著『ICTを活用した外国語教育』、東京電機大学出版局、pp.209-231
- 清原文代、2008b、「Podcast制作の実際」、吉田晴世・上村隆一・野澤和典・松田憲編著CIEC外国語教育研究部会著『ICTを活用した外国語教育』、東京電機大学出版局、pp.242-249
  2008a及び2008bに出版後にiTunesのバージョンアップなどがあったため、CIEC外国語教育研究部会のWebサイトに訂正情報に掲載している。
  http://www.ciec.or.jp/~f_lang/wiki/
- 清原文代、2008c、「PDFで作る音声付き教材　紙でもパソコンでも使える教材をPDFで！」、『漢字文献情報処理研究』第9号、好文出版、pp.92-102
- 清原文代、2010、「紙とe-Learingを繋ぐ――PDFによる外国語学習教材――」、『CIEC研究会論文誌』vol.1、pp.13-19


本研究は科研費(22520574)の助成を受けたものである。

# ❖ソフトウェア

## Windows 7 & Office 2010

千田　大介

### ❖ マイナーチェンジ

Windows・Office のバージョンアップが漢情研的にあまり盛り上がらなくなって久しい。多漢字・多言語関連の機能は Windows 2000・Office 2000 で一通り完成し、その後はブラッシュアップや文字コード最新規格への対応などが繰り返されているに過ぎないからだ。

Windows 7 と Office 2010 についても同じである。多言語・多漢字機能は、Windows Vista・Office 2007 とほぼ同等で、簡体字・繁体字のフォントについてはバージョンも全く同じであり、Vista・2007 のブラッシュアップ版であると言えよう。

以下、例によって多言語・多漢字機能を中心に、Windows 7 と Office 2010 を簡単にレビューする。

### ❖ Windows 7

#### ■ エディションの差異

Windows 7 で提供されるのは、Starter・Home Basic・Home Premium・Professional・Enterprise・Ultimate の 6 エディションである。このうち、Windows XP Mode が利用できるのは Professional 以上、多言語インターフェイスの切り替えは Enterprise・Ultimate だけが対応している。

複数言語インターフェイスの切り替えが必要であれば、Enterprise がボリュームライセンスのみでの提供なので、個人ユーザーは Ultimate を導入することになる。ただ、XP や Vista と同様、「地域と言語」の「Unicode 対応ではないプログラムの言語」を切り替えることで、中国版ソフトなども動作させることはできるので、単に海外ソフトを使いたいだけであれば Ultimate の導入は必要ない。なお、多言語インターフェイスは Vista 同様、Windows アップデート経由で提供される。

なお、動作言語の切り替え無しに、ローカルコードベースのソフトを動作させることのできる AppLocale ユーティリティであるが、Vista 同様 7 でも、対応バージョンは提供されていない。ただ、XP 対応版をダウンロードして Windows 7 で動かすことは可能であるので [1]、Unicode 非対応の海外ソフトを愛用している方には、自己責任において試していただきたい。

#### ■ 64 ビット版・XP Mode

今バージョンから、32 ビット版と 64 ビット版が同時に提供されるようになった。近頃の PC は 64 ビット対応しているので、動作速度は 64 ビット版を使った方が 1•2 割向上する。また、64 ビット版であっても、大抵の 32 ビット版アプリケーションソフトが動作する。しかし、64 ビット版へのドライバやモジュールの対応が不十分なため、問題が発生することもある。両者の多言語・多漢字対応には差異がないので、いずれを使用するかは、愛用するアプリケーションソフトの対応状況に応じて決めることになろう。

Windows 7 Professional 以上で使用できる XP Mode は、Windows XP のバーチャルマシンを提供するものだ。デフォルトではインストールされておらず、MS 社のサイト [2] からダウンロードして使用することになる。『四部叢刊』などの文献データベースには、Vista 以降のバージョンで問題が発生するものが多く見られ、Windows 7 でそれらを使うには XP Mode の使用が唯一の選択肢となる。また、XP Mode の動作言語を中国語に切り替えておけば、ホストである Windows 7 の動作言語を切り替えることなく、中国語アプリケーションを使うこともできる。

従来、ハードルの高かったバーチャルマシンが、Windows 標準の機能として提供されたことは、多言

語処理の普及という意味でも画期的だと言えよう。

## ❖ Office 2010

### ■ MLP

Office 2003 までは、英語版 Office と各言語のインターフェイス・校正ツールなどをパッケージにした多言語版、それとフォント・校正ツールなどを追加する Proofing Tools という二種類の多言語関連パッケージが存在したが、2007 からは、各言語版の Office に、マルチランゲージインターフェイスや校正ツールを追加できる、マルチランゲージパック（MLP）のみが提供されるようになった。

Office 2010 でも 2007 と同様に MLP が提供されている。ただし、執筆時点では個別の言語のパックがボリュームライセンスサービスを通じて提供されているものの、全対応言語を一つにまとめたパッケージ版の MLP は未発売である。2007 の時も、ボリュームライセンスが先行し、パッケージ版の提供は Office 発売よりも一年ほど遅れたので、今回も同様の対応になるのであろう。

筆者は、ボリュームライセンス経由で入手した簡体字中国語・繁体字中国語の MLP を導入したが、2007 に比べていくつかの改善が見られた。

まず、Word では、リボンの「校閲」タブ以下に、「中国語の変換」が追加されるようになった。2007 ではリボンのカスタマイズが出来ないことから、クイックアクセスツールバーに中国語の翻訳関係のアイコンを登録しなくてはならず、非常に手間がかかったので、これは嬉しい仕様変更である。

また、Excel 2007 では多言語パック導入後、分析ツールなど一部のリボンメニューの表示が、後から追加した言語による表示に変わってしまうという問題が発生したが、2010 では解消されている。

Windows 7 上の Word 2010 で中国語のルビ機能を使いたい場合には、MLP の導入が必須となる。この点は Vista・2007 の組み合わせと変わらない。ただし、Office 2010 64 ビット版を導入した筆者の 2 台の PC では、いずれも簡体字中国語のルビが振れないという問題が発生している。32 ビット版では問題無くルビが振れるので、これは 64 ビット版固有のバグと思われる。

Office 2010 も Windows 7 と同様、32 ビット版と 64 ビット版が同時提供されるようになったが、インストーラーはデフォルトで 32 ビット版を導入する仕様であり、MS も 32 ビット版を強く推奨している。現時点では、それに従った方が安全であると言えよう。

### ■ Publisher

Publisher は、ポスター・チラシやラベルの作成のほか、ページものの作成もこなせる DTP ソフトであり、特に多言語への対応に優れている。2007 では旧来のメニューバーが使われていたが、2010 ではリボンインターフェイスに対応した。

インターフェイスが変わったものの、機能面については大きな変更は無いようだ。多言語対応に関しても、旧バージョンと同様、アラビア文字・モンゴル文字などの文字結合も表現できる。ただし、モンゴル文字の縦書き・左→右組みには、残念ながら 2010 でも対応しなかった。

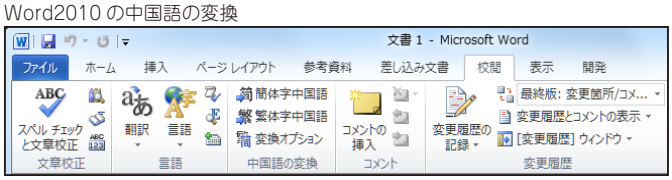
むしろ気になるのは、2007 までは「書式」→「文字幅と間隔」で文字幅の調整ができたが、2010 ではこの機能が見あたらなくなってしまったことだ。チラシ・ポスター作成に文字幅の調整は必須の機能であるので、復活を願いたい。

### 注

[1] **http://w7.vector.jp/softrev/detail.php?s_no=68923**、**http://iibox.blog123.fc2.com/blog-entry-336.html** など に関連情報が見られる。

[2] **http://www.microsoft.com/japan/windows/virtual-pc/default.aspx**

Word2010 の中国語の変換



# フォント・ブラウザ・多漢字

上地　宏一

## ❖ フォント

この 1 年の間に（多漢字）フォントに関して大きな動きは見られなかったが、本記事では特に漢字処理に関係のありそうなものをいくつか紹介する。

### ■ Mac OS X

いささか旧聞に属する情報となるが、2009 年 8 月に発売となった Mac OS X 10.6 (Snow Leopard) には標準で簡体字ヒラギノ角ゴシックフォント（Hiragino Sans GB）が搭載され、ファミリーとして W3 と W6 の 2 種類が利用できる（図 1、2）。収録文字については Mac OS に標準搭載される他の簡体字フォントと同様、GB18030-2000（国家認証取得済）および Adobe-GB1-4 相当である [1]。ヒラギノ簡体字フォント単体が 5 万円程度で販売されていることを考えれば、Mac OS は従来のヒラギノフォント搭載を含め非常にお得な OS であることは間違いない。もっとも、ISO/IEC 10646 規格の Ext.B 集合を満たすフォントが搭載されていないという点では、Windows に大きく水をあけられていることは確かであり残念である。

### ■ Microsoft Office 2010

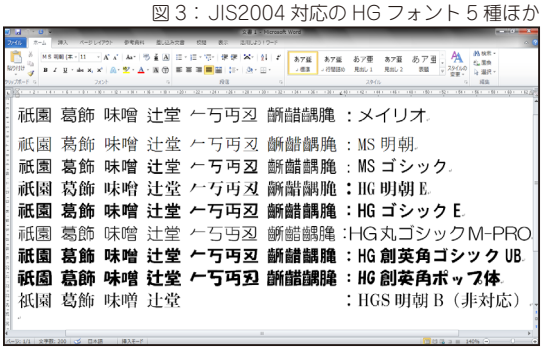
Microsoft Office には頭に「HG」という名前のついた日本語フォントが複数搭載され非常に便利であるが、この HG フォントは JIS 文字コード規格の 2004 年改正に対応していなかったため、2004 年改正に対応した環境において MS フォントとの字形の差が生じる問題が起きていた。今回 Office 2010 にバンドルされている HG フォントの一部（以下の 5 種）については、この 2004 年改正に対応して字形変更と JIS X 0213 規格への対応（文字追加）がなされている（図 3）[2]。

- HG 明朝 E
- HG ゴシック E
- HG 丸ゴシック M-PRO
- HG 創英角ゴシック UB
- HG 創英角ポップ体

図 1：Hiragino Sans GB W3

図 2：Hiragino Sans GB W6

図 3：JIS2004 対応の HG フォント 5 種ほか



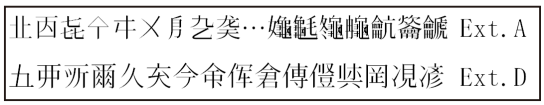


図 4：花園明朝の新規追加字の例

それと同時に、従来の Office において 2004 年改正字形を利用できる「Microsoft Office JIS2004 互換フォント」と、Office 2010 において従来の字形での利用を可能とする「Microsoft Office JIS90 互換フォント」がマイクロソフトから公開された。

JIS 漢字規格の 83 年改正時の混乱と比較すると、2004 年改正による字形の混乱はそれほど大きいとは感じないが、OS の標準フォントと Office 主力フォントの JIS2004 対応によって、この問題は収束に向かいつつあるのかもしれない。

## ■ 花園フォント

手前味噌で恐縮であるが、フリー（無償、自由）の漢字フォントである花園フォント [3] は本号が出るころに新しいバージョンとなっているはずである。本誌第 10 号「電脳漢字処理環境この 10 年」執筆時点の「花園明朝 2009 年 9 月 9 日版」と比較すると以下の点が拡張されている（図 4、一部予定を含む）。

- Ext.A の完全収録
- Ext.B 収録率が 50% を突破
- Ext.D 222 字の収録
- IVS に対応・IVD 収録
- ひらがな・カタカナの収録

Ext.B の完全収録にはまだ道遠しではあるが、今後も収録活動を継続していくところである。

### ●今後の花園フォント

ところで、新しいバージョンの花園明朝の収録字数は 6 万字に達する予定であり、そろそろ 6 万 5 千字というフォントファイルの収録限界が見えてくる。Ext.D が制定された時点での Unicode の総漢字収録数は 7 万 7 千余であるので、1 万 2 千字程度不足してしまう。

花園フォントは特に既存の日本語フォントが収録していない文字集合の表示にメリットがあると思われる。そこで、日本語フォントが収録する文字集合を独立させるとメリットが生かせると考える。Unicode 漢字 77,279 字から JIS 漢字 13,143 字を引くと 64,136 字となる（表 1、2）。実際には一部の非漢字や IVD（IVD については次項参照。現状で IVD の異体字は日本の漢字集合と結びつける必要がある）の追加、移動があるのでこの数字通りにはならないが、方針としては妥当と考えられる。

表 1：Unicode の漢字収録数リスト [4]

| 名称 | 字数 |
|---|---|
| URO | 20,902 |
| URO+ | 38 |
| Ext.A | 6,582 |
| Ext.B | 42,711 |
| Ext.C | 4,149 |
| Ext.D | 222 |
| Cmp. | 470 |
| Cmp.+ | 542 |
| IVD 追加 | 1,663 |
| 合計 | 77,279 |

表 2：JIS 文字コード規格の漢字収録数リスト

| 名称 | 字数 |
|---|---|
| JIS X 0208 | 6,356 |
| JIS X 0213 | 3,695 |
| JIS X 0212 | 5,801 |
| 0213 と 0212 の重複 | -2,743 |
| IBM | 360 |
| IBM と JIS の重複 | -326 |
| 合計 | 13,143 |

まとめると、花園フォントは以下の 2 つに分離することが妥当であろう。

- 日本集合：JIS 漢字集合＋後述の IVD
- その他集合：日本集合以外の全 Unicode 漢字

あるいは、将来的に Ext.E 集合のような文字集合の追加を見越して、文字面で分割する案も当然考えられ、しばらく頭の痛い日々が続きそうである。

## ■ IVS/IVD

IVS（Ideographic Variation Selector）は別名「異体字セレクタ」のことである [5]。Unicode に収録された漢字に対して専用の符号（異体字セレクタ）を後ろに結合することによってその異形を指定する。IVS は Unicode 4.0 で用意された [6] が、IVS 自体は字形を指定する仕組みだけであり、実際にどのような字形を指

定できるかを定義した IVD[7]（Ideographic Variation Database：異体字データベース）が必要となる。IVD の枠組みは 2006 年に制定され、2007 年に初めてのデータが決められた。なお、IVS は勝手に自分で異体字に番号を振るのではなく、Unicode が定めた IVD に対してのみ使用が認められているものである。

さて、初めて規定された IVD であるが、その実態はアドビ社が定義した文字集合である Adobe-Japan1-6（以降 AJ16 と記す）である。言い換えると、AJ16 の漢字テキストをほぼ [8] すべて Unicode テキストデータで表現できることになる。IVD については重複や親字の齟齬なども指摘されているが、DTP 方面で多用される AJ16 のデータをそのまま Unicode テキストとして扱うことができるというのはメリットが大きいと感じる。

問題は IVS に対応したフォントと OS、アプリケーションがほとんどないということである。現状の IVD は実質的には AJ16 であるのだから、AJ16 に対応したフォントがそのまま使えるのかというとそうではない。IVS に対応しているフォントは小塚明朝・ゴシック Pr6N と花園明朝に限られている [9]。また、OS としては Windows 7 が対応しているが、実質的に IVS を利用できるのは「メモ帳」のみ（！）であり（図 5）、Microsoft Office をはじめとする各種アプリケーションではまったく利用できない。一部のブラウザでは IVS への対応が始まっているが、それについては「ブラウザ」の項で述べる。

このようにまだ実用には程遠い IVS / IVD であるが、最終的には異体字問題を解決する手段となりうる可能性も秘めている [10]。特に文字コード未収録字がユニフィケーションの対象となるような場合には、文字コードに追加するのは困難であり、また仮に文字コードに追加できる場合でもその審議・制定には非常に時間がかかる。一方 IVD への追加は ISO への申請ではなく Unicode への申請であり、一定期間の公開レビューを経て追加が決定されるようである。現在のところ日本から「汎用電子情報交換環境整備プログラム」に基づく文字集合が登録申請され、レビュー期間が終了したところである [11]。

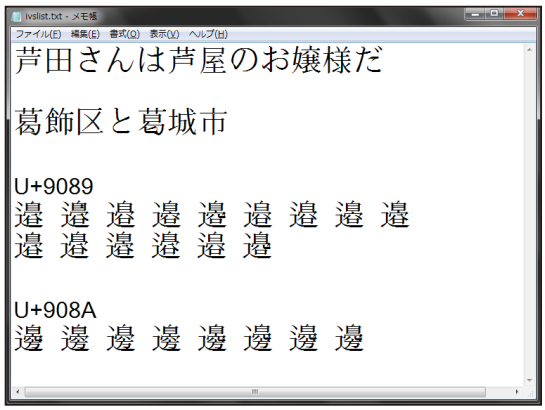



図 5：Windows 7 のメモ帳で IVS を使用

## ブラウザ

### ■ はじめに

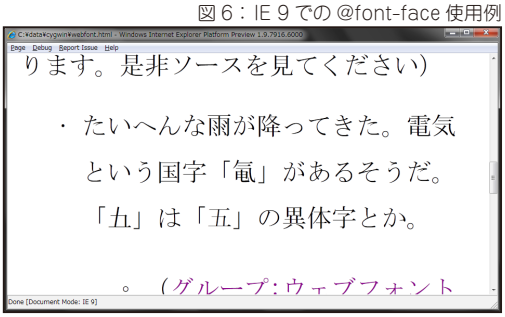
2010 年 9 月現在で各メジャーブラウザの安定版および開発版のバージョンは以下のようになっている。

- Internet Explorer 8（安定板、以降 IE）
- Internet Explorer 9（開発版）
- Mozilla Firefox 3.6（安定版、以降 Firefox）
- Mozilla Firefox 4（開発版）
- Google Chrome 6（以降 Chrome）
- Opera 10.61（以降 Opera）
- Apple Safari 5（以降 Safari）

### ■ @font-face の対応状況

本誌第 9 号、第 10 号のブラウザレビュー記事ではいずれも @font-face について取り上げた。昨年の記事では「実用段階間近」とあるが、さらに状況が進展したことについて紹介する。

図 6：IE 9 での @font-face 使用例



まず昨年の状況は次の通りであった。

- IE：eot 形式フォントのみ、BMP（基本多言語面）領域のみ対応
- Firefox：対応
- Chrome：対応せず

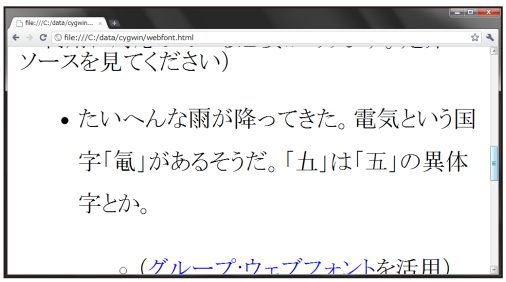



図 7：Chrome での @font-face 使用例

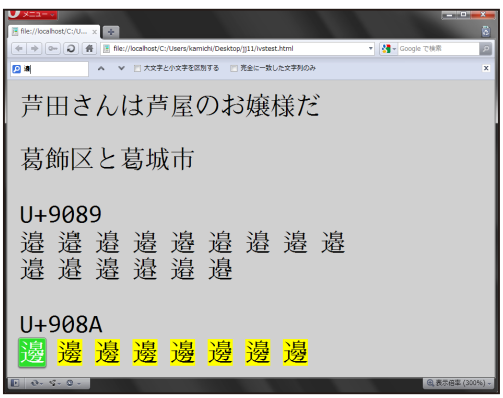
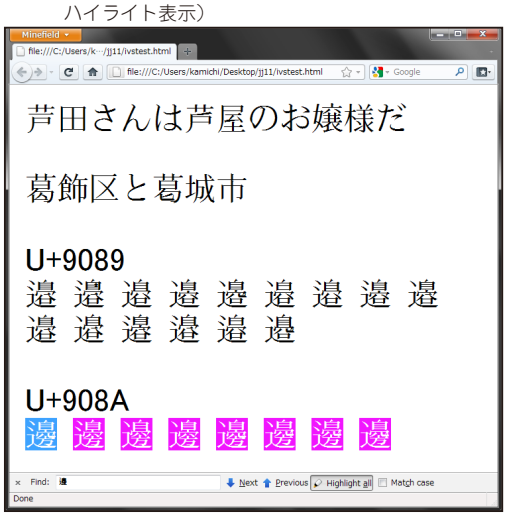



図 8：Opera での IVS/IVD 表示（検索のハイライト表示）

図 9：次バージョン Firefox での IVS/IVD 表示（検索のハイライト表示）




- Opera：BMP 領域のみ対応
- Safari：対応

その後 IE については、次のバージョンである IE 9 のプレビュー版において TrueType 形式によるフォント呼び出しができることを確認した [12]。また、BMP 領域だけでなく第 2 面（Ext.B 以降）でも可能であった。

また、Chrome についても @font-face への対応を確認できた。第 2 面以降も表示可能である。まとめると、次のようになる。

- IE（次バージョン）：対応（図 6）
- Firefox：対応
- Chrome：対応（図 7）
- Opera：BMP 領域のみ対応
- Safari：対応

このように、ほぼ完全に @font-face への対応が完了したと言えるだろう。従来 eot 形式にのみ対応していた IE が一般的なフォント形式に対応する予定であることは非常に望ましい。なお IE 9 は Windows XP には対応しないことになっている。Ext.B フォントの標準搭載も含め、そろそろ Windows XP から Windows 7 へ乗り換えるべきタイミングが到来したのではないだろうか [13]。

### ■ IVS / IVD

先述の IVS / IVD はすでに一部ブラウザにおいて対応している。そのブラウザとは Firefox（次バージョン）と Opera である。試しに先ほどの「メモ帳」と同様のサンプルを加工した HTML ファイルとして表示したものが図 8、9 である。花園明朝フォントの IVD グリフを正しく表示できている。なお、SVG フォント形式であれば Safari と Chrome の WebKit 系ブラウザも IVS の利用が可能である。

いずれも文字の検索を行うと、IVS を無視した形（すべての異形字）が抽出される。

このようにまだ対応ブラウザは少ないが、将来的には IVS を含んだテキストドキュメントを普通に処理できる環境がもたらされるであろう。

### ■ Firefox の新しい機能

次バージョンの Firefox ではさらに多言語に関する

興味深い機能が試験搭載されている。本誌第９号の拙稿[14]にて「Arial Unicode MS」フォントに日本以外の中国大陸および台湾・香港デザインのグリフが含まれていることに言及したが、現在開発中の Firefox では、この１つのフォントの中に含まれる複数デザインのグリフを CSS と lang 属性で切り替えることができる。実際にいくつかの漢字を書き分けたのが図10である。

具体的な HTML ソースを以下に挙げる。

```
<html><head><style>
@font-face {
  font-family: BodyText;
  src: local("Arial Unicode MS");
}
body { font-family: BodyText; }
:lang(ja) { -moz-font-language-
override: 'JAN'; }
:lang(ko) { -moz-font-language-
override: 'KOR'; }
:lang(zh-TW) { -moz-font-language-
override: 'ZHT'; }
:lang(zh-CN) { -moz-font-language-
override: 'ZHS'; }
</style></head><body>
<span lang="ja">骨述与差花餃 (ja)</
span><br>
<span lang="ko">骨述与差花餃 (ko)</
span><br>
<span lang="zh-CN">骨述与差花餃 (zh-
CN)</span><br>
<span lang="zh-TW">骨述与差花餃 (zh-
TW)</span><br>
</body></html>
```

現状では次バージョンの Firefox でのみ対応する CSS 記述であり、また複数地域デザインに対応したフォントはなかなか見受けられない[15]ため、残念ながら今すぐに実用できる技術ではない。

とはいえ、従来はそれぞれの地域のデザインが施されたフォントを CSS で組み合わせて多言語同時表示を行ってきた。そのためにフォントごとのデザインの違いや微妙な文字の大きさの違いが読み手に違和感を与えることがあった。グリフデザイン切り替え機能がこの問題の解決に役立つことに期待したい。

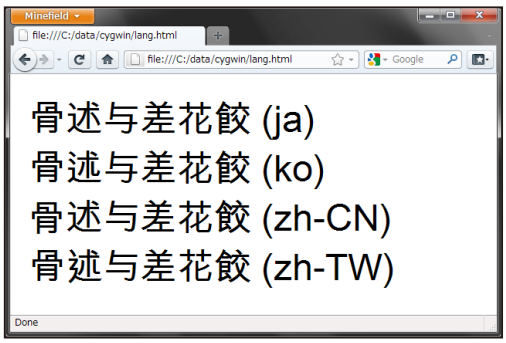



図 10：次バージョン Firefox での同一フォントによるデザイン書き分け

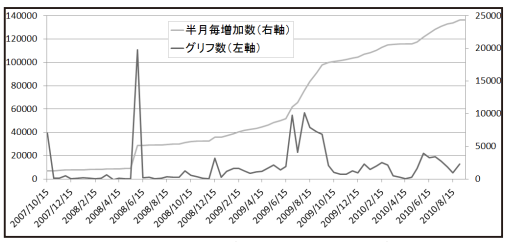
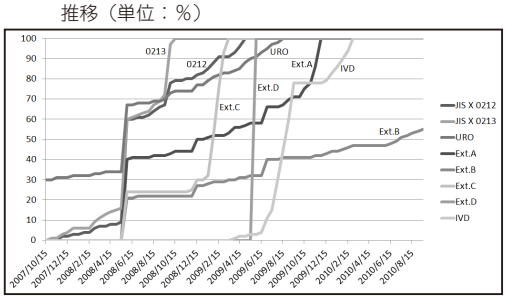



図 11：グリフウィキ登録グリフ数の推移

図 12：Unicode など各漢字集合のグリフウィキ収録率の推移（単位：%）




## ❖ 多漢字（グリフウィキ）

グリフウィキは本誌が出るころに運用開始３周年を迎えることとなる。グリフ数も順調に増加しつつあり、13 万 7 千グリフに達している（図11）[16]。Unicode 収録文字集合として、URO[17]、Ext.A、Ext.C、Ext.D を完全収録し、Ext.B については 55% にまで達した（図12）。このほかにも国字や戸籍統一文字集合、東洋学文献類目外字、大正新脩大藏經テキストデータベース外字などの収録が開始されている。

従来、グリフウィキに登録したデータは、TrueType

- たいへんな雨が降ってきた。電気という国字「𠃌」があるそうだ
  - （グループ:ウェブフォントを活用）
- 新元素コペルニシウムに充てる漢字は繁体字は「鎒」(U+93B6)で、簡体字は「镉」(文字コード未定義)です。
  - （copernicium-simplifiedを呼び出し）

図 13：　Web フォントの利用例（矢印が外字）

フォントとしてまとめるほか、画像ファイルとして HTML で参照するなどの利用方法があったが、ブラウザの Web フォント対応に合わせて、Web フォントとしての活用方法を新たに 2 種類用意した。また、登録されたグリフデータを利用した漢字検索サービスの試作を行った。

## ■ Web フォント

グリフウィキに登録したデータを Web フォントとして利用する方法を用意した。これは 2 つの形態からなっており、一つは従来のグループページを用いてフォント化した集合をまとめて呼び出すもの、もう一つはグリフ 1 字のみを単字で呼び出すものである。

特に後者は、グリフをグリフウィキに登録した時点で自動的に「〓」に対してそのグリフを割り当てた 1 字フォントが生成されるようになっているため、文字を登録したらすぐに Web フォントとして呼び出せるようになっている。

いずれの場合も HTML に以下の link 要素（<link> タグ）を記述することで利用できる。

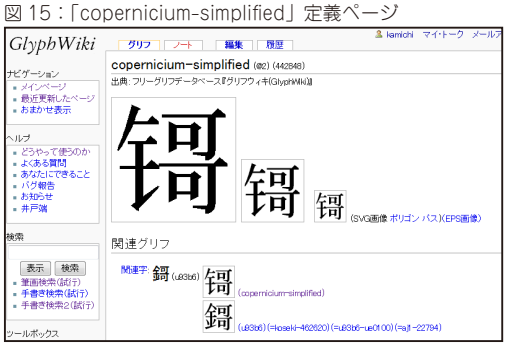
グループページ：<link rel="stylesheet" type="text/css" href="http://glyphwiki.org/style?page=(ページ名)&label=(任意のラベル)">
グリフ 1 字：<link rel="stylesheet" type="text/css" href="http://glyphwiki.org/style?glyph=(グリフ名)">

実際に利用した例を図13に示す。
このソースの一部は以下のとおりである。

```
1: <html>
2: <head>
3: <link rel="stylesheet" type="text/css" href="http://glyphwiki.org/style?page=Group:ウェブフォント &label=ame">
4: <link rel="stylesheet" type="text/css" href="http://glyphwiki.org/style?glyph=copernicium-simplified">
5: </head>
6: <body>
7: <ul style="line-height: 180%;">
8: <li> たいへんな <span class="ame"> 雨 </span> が降ってきた。電気という国字「<span class="ame"> 電 </span>」があるそうだ </p>
9: <li> 新元素コペルニシウムに充てる漢字は繁体字は「鎒」(U+93B6)で、簡体字は「<span
```

図 14：「Group: ウェブフォント」定義ページ

図 15：「copernicium-simplified」定義ページ



```
    class="copernicium-simplified">〓
    </span>」（文字コード未定義）です。</
    p>
10: </ul>
11: </body>
12: </html>
```

3 行目では「Group[18]: ウェブフォント」という名称のグループページ（図14）に文字の集合を記述し、フォント化している。この集合を「ame」というラベルのクラス名で呼び出す。8 行目に記述している。

4 行目では「copernicium-simplified」というグリフ 1 字（図15）をそのまま呼び出している。9 行目の「〓」の部分が外字に対応している。

このように用いる文字の定義情報を <link> で埋め込むことで簡単に利用できる。一方で環境によっては head 要素（<head> タグ）を操作できない場合がある。この時は Google Font API[19] を援用して <body> の中に以下のように記述すればよい。「urls:」に続けて呼び出したいフォントの CSS（を生成するため）の URL を指定する。複数ある場合には例のようにカンマで区切る。

```
<script type="text/javascript">
WebFontConfig = { custom: {
urls: ['http://glyphwiki.org/
style?page=Group:ウェブフォント
&label=ame', 'http://glyphwiki.org/
style?glyph=copernicium-simplified']
} };
</script>
<script src="http://ajax.googleapis.
com/ajax/libs/webfont/1/webfont.js"
type="text/javascript"></script>
```

この 2 つの <script> を埋め込むと、<link> と同様の使い方で好きな文字をウェブフォントとして呼び出すことが可能となる。

まとめると、以下の手順で文字コード未収録文字を HTML 文書として利用できる。

(1) グリフウィキに文字を登録する

| | |
|---|---|
| 著者 | 周宝宏 (zhou bao hong) 著 |
| 題名 | &RUI6-E018;公盨"天命禹専土&RUI6-E04C;山&RUI6-E04D;川"考释 |
| 雑誌 | 出土文献语言研究 1輯 |
| 発行年 | 2006 |
| 頁 | 49-55 |
| 分類 | 金石・古文書學 / 金文 (133000) |
| ID | A2007-10507-00028 (CHU195) |
| 著者 | 刘雨 (liu yu) 著 |
| 題名 | 师&RUI6-E04F;钟和姬寘母豆 |
| 雑誌 | 古文字研究 26輯 |
| 発行年 | 2006 |
| 頁 | 165-171 |
| 分類 | 金石・古文書學 / 金文 (133000) |
| ID | A2007-10507-00012 (CHU195) |
| 著者 | 王蕴智 (wang yun zhi) 著 |
| 題名 | 释甲骨文&RUI6-E04E;字 |
| 雑誌 | 古文字研究 26輯 |
| 発行年 | 2006 |

図 16：外字メタ記述「rui6」をキーワードとした検索結果の一部（「&RUI6-E」を色づけ）

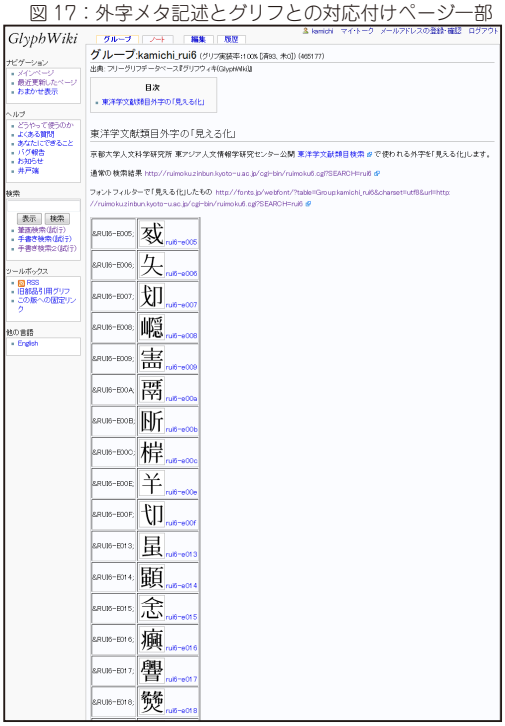
(2) HTML ファイル内に <link> または <script> を埋め込み、文字を記述する

(3) HTML を公開する

## ■ Web フォントフィルター

先に述べた Web フォントはこれから公開する Web

図 17：外字メタ記述とグリフとの対応付けページ一部



コンテンツに対して有用である。その一方で、すでに公開しているWebリソース内で文字コード未収録文字が使われているケースも多数存在する。多くの場合、この未収録文字はIDなどで区別されている（外字メタ記述）。そこでこの区別して記述された外字部分を、グリフウィキに登録した字形データに置き換えて可視化するWebフォントフィルターを用意した。具体的には以下の手順を経て利用できる。

(1) 外字字形をグリフウィキに登録する
(2) Webコンテンツ内の外字メタ記述とグリフウィキに登録したグリフ名との対応関係を記述したグリフウィキのページを用意する
(3) フィルターURLに目的のページURLを組み込んでブラウザで表示する

たとえば京都大学人文科学研究所で公開されている東洋学文献類目のWeb検索サービスでは外字が「&RUI6-Exxx;」の形で記述されている。以下のURLは「rui6」をキーワードに検索を実行するものであり、外字メタ記述部分が抽出される（図16）。

図18：フィルターを通した検索結果の一部（置き換わった部分を色づけ）

| | |
|---|---|
| 著者 | 周宝宏 (zhou bao hong) 著 |
| 題名 | 𡈼公簋"天命禹專土&RUI6-E00E;山&RUI6-E00F;川"考释 |
| 雑誌 | 出土文献语言研究 1輯 |
| 発行年 | 2006 |
| 頁 | 49-55 |
| 分類 | 金石・古文書學 / 金文 (133000) |
| ID | A2007-10507-00028 (CHU195) |
| 著者 | 刘雨 (liu yu) 著 |
| 題名 | 师&RUI6-E00C;钟和姬寏母豆 |
| 雑誌 | 古文字研究 26輯 |
| 発行年 | 2006 |
| 頁 | 165-171 |
| 分類 | 金石・古文書學 / 金文 (133000) |
| ID | A2007-10507-00012 (CHU195) |
| 著者 | 王蘊智 (wang yun zhi) 著 |
| 題名 | 释甲骨文&RUI6-E04E;字 |
| 雑誌 | 古文字研究 26輯 |

http://ruimoku.zinbun.kyoto-u.ac.jp/cgi-bin/ruimoku6.cgi?SEARCH=rui6

この外字群について、字形をグリフウィキに登録し、外字メタ記述とグリフとの対応関係をグリフウィキに登録する。

実際の対応関係の記述は1行1組で「,外字メタ記述,グリフウィキでのグリフ名」の形式で列挙すればよい。今回のサンプルでは「Group:kamichi_rui6」というページ（図17）に以下（一部抜粋）のような対応付けデータを登録した。

```
,&RUI6-E005;,[[rui6-e005]]
,&RUI6-E006;,[[rui6-e006]]
,&RUI6-E007;,[[rui6-e007]]
,&RUI6-E008;,[[rui6-e008]]
,&RUI6-E009;,[[rui6-e009]]
,&RUI6-E00A;,[[rui6-e00a]]
,&RUI6-E00B;,[[rui6-e00b]]
,&RUI6-E00C;,[[rui6-e00c]]
,&RUI6-E00E;,[[rui6-e00e]]
,&RUI6-E00F;,[[rui6-e00f]]
```

以上のような準備を行ったうえで、次のフィルター用URLに対象となるWebページのURLを指定する。フィルターの書式は以下のとおりである。

```
http://fonts.jp/webfont/?table=(対応関係を記述したグリフウィキのページ名)&url=(フィルターを通したいWebページURL)
```

東洋文献類目外字の場合は以下のようになる（文字コード指定の記述が必要なため、上記書式に1要素加えている）。

http://fonts.jp/webfont/?table=Group:kamichi_rui6&charset=utf8&url=http://ruimoku.zinbun.kyoto-u.ac.jp/cgi-bin/ruimoku6.cgi?SEARCH=rui6

すると、外字部分が実際の漢字字形に置き換わって

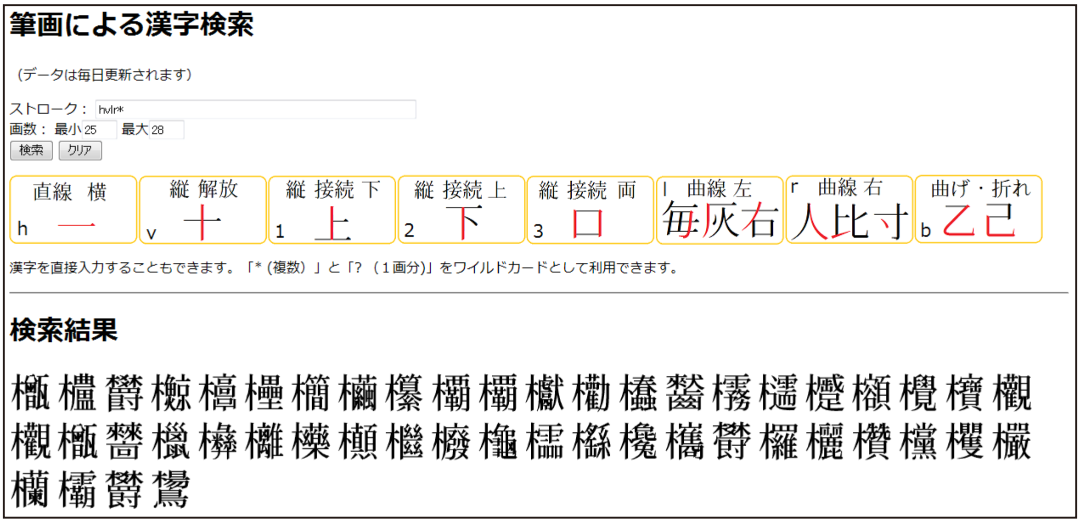



図 19：キーワード「hvlr*」の検索結果

表示される（図18）。置き換わった文字はいずれも「〓」のコードになっている。

残念ながら、東洋学文献類目外字は漢字（明朝体）以外の文字も含むため、グリフウィキと Web フォントフィルターを用いて完全に文字を表示することはできないが、それでも多くの外字について、実際の文字を表示することが可能となる。このように既存のテキストデータ資源 [20] について、元となるデータを変更することなく、グリフウィキに文字字形データの登録と対応付けデータを記述してフィルターを通すことにより外字部分を実際の文字で表示することが可能となっている。

## ■ 大規模漢字集合検索

大規模漢字集合に対する検索手段は、これまで本誌で紹介した『今昔文字鏡』、『超漢字検索』などの製品や、あるいは Web サービスとして「CHISE IDS Find」が挙げられる [21]。ところで著者が運用しているグリフウィキに登録されているグリフはすべて KAGE 形式によるストロークデータであり、どのような筆画がどのような位置に配置されているかがデータ化されている。そこで、ストロークデータを用いた文字検索サービスを試作・公開した。試作したのはストローク種別による検索と手書き検索の 2 種類である。

これらの文字検索サービスはグリフウィキに登録されているすべてのグリフを対象としているため、Unicode に収録されている以外の各種異体字等（戸籍統一文字集合、AJ16、国字）も検索できる。その一方、Unicode 収録文字であってもグリフウィキに登録されていない一部の Ext.B 集合は検索できない。

現在グリフウィキにはエイリアス（同じ字形で別の名称がつけられているグリフ）を除いた 6 万 7 千余字が登録されており、これが実質的な検索対象となる。

### ●ストローク（筆画）による漢字検索

ストロークによる漢字検索は、筆画を以下の 8 種類に分別して検索キーワードとして指定する検索サービス [22] である。

- ●横直線 (h)
- ●縦直線：両端が解放 (v)
- ●縦直線：起筆部が他画に接続 (1)
- ●縦直線：終筆部が他画に接続 (2)
- ●縦直線：両端が他画に接続 (3)
- ●曲線：左方向に運筆 (l)
- ●曲線：右方向に運筆 (r)
- ●曲げ・折れ (b)

直線については 5 種類とし、横線を単独で 1 種類、縦線は始筆部と終筆部が他画と接続するか否かで 4 種類に区別する。注意として KAGE データの特性上、「口」は 4 画で表現し、直角に折れる筆画は分解して考える必要がある。

曲線は運筆方向の左右の違いで 2 種類に区別する。

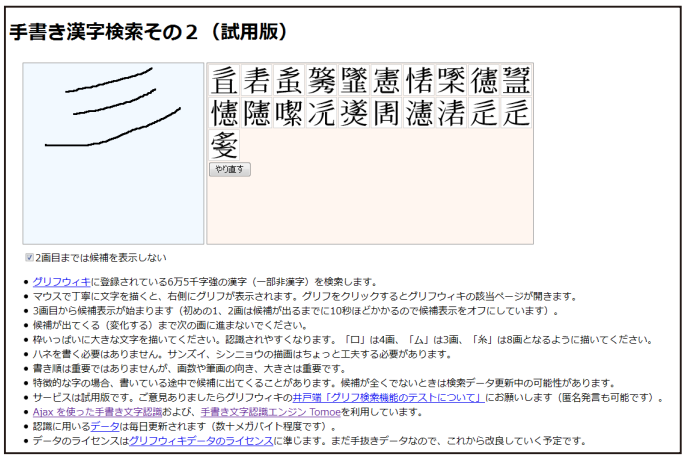


図 20：手書き漢字検索の例（その 1）

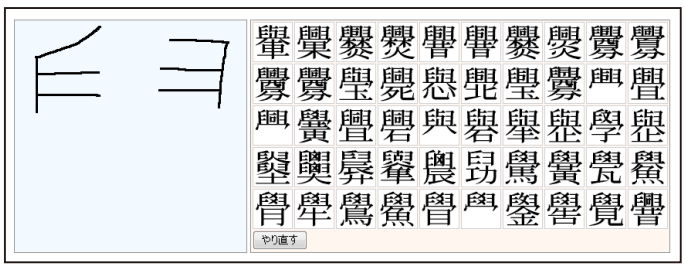


図 21：手書き漢字検索の例（その 2）

「曲げ」と「折れ」は統合して 1 種類としている。上記リスト末尾のアルファベット・数字は検索キーワードに指定するための略称記号である。

例として「木」という字は「横直線」、「縦解放直線」、「左曲線」、「右曲線」の「hvlr」で表現され、「口」は「両端接続縦直線」、「横直線」、「両端接続縦直線」、「横直線」の「3h3h」となる。このように 8 種類の筆画を数字及びローマ字 1 文字に代替した記号を検索キーワードに書き順に沿って指定する。

またワイルドカードを用意しており「?（任意の 1 画）」と「*（複数画、あるいはデータなし）」も利用できる。画数 [23] については最小画数と最大画数が指定できる。

たとえば「木偏」で 25 画～ 28 画の文字を探す場合、キーワードに「hvlr*」を、最小画数を 25 画、最大画数を 28 画とする。図19の通り、48 種類のグリフが表示された。その内訳は Unicode 収録字が 39 字、Unicode の IVD が 1 字、1978 年時の JIS 規格収録字形が 1 字、Ext.E に収録審議中が 2 字、その他外字が 5 字であった。各グリフはグリフウィキの当該ページへのリンクとなっている。

なお、「木」を指定する際に「hvlr」ではなく直接「木」と入力してもよい。この場合グリフウィキに登録されている「木」のデータが自動的に展開され、「木 *」と「hvlr*」は同じキーワードとなる。ただし「土偏」をキーワードに指定するときは「土」としてはいけない。なぜならば「土偏」の 3 画目は「横直線」ではなく「右曲線」に変化するためである。同様に日本デザインの「ウ冠」と中国スタイルの「ウ冠」は 1 画目が「下接続縦直線」と「右曲線（＝点）」とで異なるような部分にも注意が必要である。

また、ストローク検索による漢字検索は書き順が重要であるが、残念ながらグリフウィキに登録されているデータには書き順が正しくないものがあるほか、人によって書き順が異なるケースもある。このため実際の利用には画数指定による限定化とワイルドカードの活用が求められる。

### ●手書きによる漢字検索

筆画指定による漢字検索とは別に、過去に本誌で紹介した [24]「Ajax を使った手書き文字認識」を加工した手書きによる漢字検索サービスも試作・公開した（図 20）[25]。内部で利用している文字認識エンジンも同じであり、単純に文字認識用のデータをグリフウィキのものに差し替えたものと考えてよい。

前項のストロークによる検索と同様、「口」は 4 画に分解して描くほか、「于」の 3 画目は跳ねずに「干」と同じように描く必要がある。

検索対象が 6 万 7 千字であるため、認識スピードが遅いという問題があるが、書き順にとらわれずに文字を描画するだけで検索が可能である。一方、文字の大きさや筆画の向きが重要なため、慣れるまでは意図する文字が表示されないこともある。

また、認識エンジンの特性により、字形の一部分を描画するだけでも候補が現れる（図21）。

大規模文字集合に対する手書き文字認識エンジンとしては、富士通研究所の開発した技術 [26] が見受けら

れる。また中規模文字集合に対するものとしては、本誌で過去に紹介した Q-Pen[27] やその Web サービス版 [28] がある。今回著者が公開した手書き漢字検索は、大規模文字集合を対象とした Web サービスであり、珍しいのではないかと思われる。

●より有用な多漢字検索を目指して

今後、Unicode の Ext.B 集合や Ext.E 集合の完全収録により、さらに大規模の漢字集合に対応していく予定であり、また大規模漢字集合の部品検索で有用な CHISE IDS Find に匹敵するようなサービスを目指して改良を重ねていく所存である。

## ※ まとめ

特に後半は著者の研究内容の紹介に終始してしまったが、多漢字環境の実現に向けて Web フォントと IVS/IVD という 2 つの要素が普及段階に入りつつあり、今後の動向が大いに注目されるところである。それに合わせて著者自身も関連研究を進め、より実用的な成果が上がるよう努力したいと考えている。

## 注

[1] つまり、漢字について言えば Ext.A を含む基本多言語面（BMP）をカバーすることになる。

[2] 当会会員清水哲郎氏が執筆した IT Pro のニュース記事「MS が Office 2010 で日本語フォント環境を強化」 **http://itpro.nikkeibp.co.jp/article/NEWS/20100617/349377/**

[3] 花園フォント **http://fonts.jp/hanazono/**

[4] URO：CJK 統合漢字 20,902 字、URO+：URO の後ろに配置された追加集合、Cmp.：CJK 互換漢字、Cmp.+：CJK 互換漢字補助、IVD+：IVD については次項「IVS/IVD」を参照願いたい。IVD+ とは、この IVD 収録によって追加される異体字集合を指す。

[5] 厳密には「異体字セレクタ」は漢字以外の文字も想定している。

[6] 本誌では第 6 号「Unicode 4.1.0（ソフトウェア・レビュー：多言語情報処理）」で紹介されている。

[7] Ideographic Variation Database **http://www.unicode.org/ivd/**

[8] Ext.D 集合が制定された時点で 2 文字だけ未収録となる。

[9] 「IPA フォント」は IVS の仕組みは用意されているが IVD の文字集合は含まれない。このほか、手書き書体であれば「Y.OzFont」が IVS に対応している。

[10] 互換漢字との運用整合性も含め、将来混乱を招く可能性をも秘めていることに言及しておく。

[11] PRI 167 **http://www.unicode.org/ivd/pri/pri167/index.html**

[12] なお本稿では Windows 7 (32bit) において、グリフウィキで生成したフォント（Fontforge の生成による TrueType 形式）を使って試した。

[13] 著者は 2010 年 4 月に観念し、XP から Vista をスキップして Windows 7 に移行した。

[14]「ことえり（特集 1 ：今だからこそ！日本語 IME を見直そう）」を参照。

[15] 花園フォントでは将来的に複数地域デザインのグリフを収録する予定である。

[16] グリフウィキに登録されているグリフの一部は科研費（22700262）の助成を受けたものである。

[17] ここでは CJK 統合漢字 20,902 字の集合を指す

[18] グリフウィキのページ名で「Group:」と「グループ :」は同じものを指す

[19] Google Font API **http://code.google.com/intl/ja/apis/webfonts/**

[20] すべての外字部分を「〓」だけで置き換えてしまっているケースでは残念ながら利用できない。

[21] このほかにも「FindUniCode」のようなフリーの検索プログラムも公開されている。 **http://gishtec.com/FindUniCode.aspx**

[22] 筆画による漢字検索 **http://fonts.jp/glyphwiki/kensaku.cgi**

[23] 繰り返しになるが、一般的な画数とは異なるので注意が必要である。「口」は 4 画、「凸」は 8 画となる。

[24] 本誌第 9 号「ウェブベース日本語 IME（特集 1 ：今だからこそ！日本語 IME を見直そう）」を参照。

[25] 手書き漢字検索その 2（試用版）**http://fonts.jp/glyphwiki/hwr2.html**

[26] 8 万字を超える漢字の手書き入力技術の開発に成功 **http://pr.fujitsu.com/jp/news/2006/04/24-2.html**

[27] Unihan Q Pen（本誌第 5 号「中国語 IME（ソフトウェア・レビュー）」を参照。

[28] 書同文 Web 工具：手写輸入 **http://hanzi.unihan.com.cn/Tools/QPen/**

# 文字コード技術

師　茂樹

## ❖ Unicode 6.0

Google の統計によれば、ウェブ上で用いられている文字コードのうち、Unicode[1] の割合が 50％を超えたという [2]。今や名実ともに、Unicode が最もメジャーな文字コードとなったわけである。

その最新バージョンである 6.0.0 が 10 月に正式公開となる（はずであったが、まだ最終投票に至っておらず、もう少しかかりそうである。執筆時点では β 版）。6.0 で追加される予定のブロックの中には、東アジアに関係が深いものがいくつかあるので、ここで簡単に見ておきたい。

- CJK Unified Ideographs Extension D（CJK 統合漢字拡張 D、U+2B740……U+2B81F）
- Kana Supplement（仮名補助文字、U+1B000……U+1B0FF）
- Emoticons（顔文字、U+1F600……U+1F64F）
- Miscellaneous Symbols And Pictographs（様々なシンボルやピクトグラム、U+1F300……U+1F5FF）
- Playing Cards（トランプ、U+1F0A0……U+1F0FF）
- Transport And Map Symbols（乗り物と地図のシンボル、U+1F680……U+1F6FF）

図 1：CJK Extension D

### ■ Extension D

1 番目の CJK 統合漢字については今さら説明するまでもないだろう。Extension D では、日本と中国からの提案によって、新たに漢字 222 文字が追加されている（図 1）。

### ■ 仮名の追加

次の仮名補助文字は、1B000 KATAKANA LETTER ARCHAIC E と 1B001 HIRAGANA LETTER ARCHAIC YE の 2 文字である（図 2）。

この 2 文字については提案者と日本 National Body との間で見解の相違が見られ [3]、収録されることにはなったものの今後、禍根を残しそうな雰囲気である。

### ■ 絵文字

続く 4 つのブロック（Emoticons、Miscellaneous Symbols And Pictographs、Playing Cards、Transport And Map Symbols）は、すべて日本の携帯電話の絵文字に関わるものである（図 3）。これら絵文字の収録の経緯については小形克宏氏のレポート [4] が参考になるが、こちらも紆余曲折を経て収録されたものである（筆者も小形氏が結成したチームに参加し、Emoticons に対するコメントをいくつか出した）。文字数を数えればわかるが、Unicode に収録された絵文字の数は、日本の携帯電話のそれをはるかに上回る。その背景には、日本の携帯電話の絵文字を、Unicode という国際規格にふさわしいより一般性の高い絵文字にしたいという意志がある。一方で Unicode はどん

どんと「文字博物館」的色彩を強めており、上に見た Ext. D や変体仮名を見ても分かるように、新しく収録される文字は極めて一般性の低い文字ばかりになりつつある。このあたりのせめぎ合いが、今後の課題となるのではないかと思われる。

## ❖ 普及する IVS・IVD

Unicode の IVS・IVD（詳細は上地宏一氏のレビュー「フォント・ブラウザ・多漢字」参照）は、漢字の微妙な字体差を記述する方法として、最近、急速に普及しようとしている。まだ実用段階とは言えないものの、Windows 7 や Mac OS X などのメジャーな OS が徐々に IVS・IVD に対応しつつあり、IVS 対応のフォントも出まわるようになってきているのも、それを後押ししている。

今後、特に日本語の漢字処理においては、IVS はなくてはならない技術になるかもしれない。住民基本台帳ネットワーク・新戸籍システム・新登記システムが統合されてできた汎用電子情報交換環境整備プログラムの「文字一覧表」には、戸籍や登記簿で用いられている細かい異体字が多数収録されているが、これらはすべて IVS として Unicode に登録されている [5]。地名等を厳密にコンピュータ上で表現しようと思えば、IVS 技術が必須になるのである [6]。

### 注

[1] 本レビューでは ISO/IEC 10646 なども含めて一括して Unicode と呼んでいる。

[2] Twitter 上での Mark Davis 氏の発言（**http://twitter.com/mark_e_davis/status/22673110887**、2010 年 9 月 27 日最終確認）。

[3] 提案者側の見解としては「二つの「え」の話〜『古言衣延辨』」（**http://www.akenotsuki.com/eyeben/**、2010 年 9 月 27 日最終確認）を参照。一方、日本 NB の一員である高田智和氏は「ア行の /e/・ヤ行の /je/ を表わす仮名文字の標準化をめぐる諸問題」（漢字文献情報処理研究会第 12 回大会、2009 年 12 月 20 日、花園大学）において「この提案については、⑴提案文字をコンピュータで扱う目的や想定される使用層が必ずしも明確でないこと、⑵「変体仮名」の追加提案を今後誘発する可能性があること、⑶活字字形としての安定と洗練が不足しているため代表字形のデザインが難しいこと、⑷提案が日本からではなく他の National Body からなされたこと、などの問題がある」（発表概要より）と述べる。

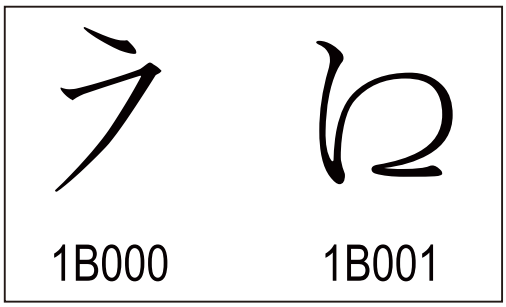


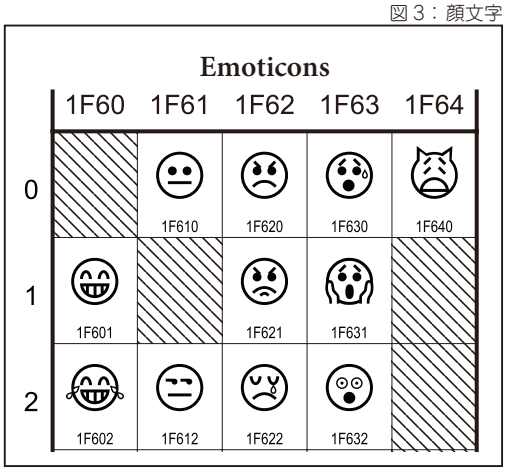

図 2： Kana Supplment

[4] 小形克宏「絵文字が開いてしまった「パンドラの箱」」（『東洋学へのコンピュータ利用 第 21 回研究セミナー』、2010 年 3 月）に、CNET Japan（http://japan.cnet.com/）で公開されている原稿が一括収録されている。

[5] “Combined registration of the Hanyo-Denshi collection and the first batch of its ideographic variation sequences.”（**http://www.itscj.ipsj.or.jp/domestic/sc02/hanyo-denshi/20100331/**、2010 年 9 月 27 日最終確認）。

[6] 安岡孝一「漢字 1 文字が最大 8 バイト、Unicode の「IVS」とは？」（**http://itpro.nikkeibp.co.jp/article/COLUMN/20100126/343783/**、2010 年 9 月 27 日最終確認）。

図 3：顔文字



# IME

小川　利康

## ❖ IME 花盛り

### ■ Office IME 2010 の無償公開

MS Office 2007 に標準搭載されていた Office IME 2007（Microsoft Office Input Method Editor 2007）の変換効率の悪さが話題になったことをご記憶の読者も多いだろう。惨憺たる評判に Microsoft も 2007 年 6 月に修正プログラムを配布したが、その後も評価は覆されることはなかったようだ。

この問題に関連して、本誌も「特集 1 ：今だからこそ！日本語 IME を見直そう」（第 9 号、2009 年 10 月刊）という特集記事を組んだことがある。冒頭の文章で二階堂善弘氏は、かつて多彩だった日本語入力システムの選択肢が減り、ほとんどのユーザが Windows 付属 IME を使うようになった結果、誰も IME を意識しなくなったと指摘し、Office IME 2007 を契機にもう一度「アプリケーションとしての IME」を見直すべきだと結んでいる。

今回、MS Office 2010 発売に当たり、その名誉回復を意図してか、Office IME 2010（日本語、中国語簡体字及び繁体字、韓国語）が旧バージョンの MS Office（XP, 2003, 2007, 2010）ユーザ向けにも無償公開された。[1]

Microsoft の公式ページで謳うところによれば、従来の IME 2007 と比較して 2 倍以上の変換速度を誇り、変換途中の文字を使ってネット検索もできるという。さらに新たな拡張性として、入力変換辞書が Microsoft Update を通じて更新されるだけでなく、拡張辞書も XML ベースのフォーマットによって、誰でも自由に作成配布することができるようになった。中国語簡体字版についても同様の機能が提供され、貧弱だった IME の印象を完全に一新した。

なるほど確かに目覚ましい進化だが、どの機能も実は他の IME では早くから実現されており、今さら目新しいものでもない。これまで MS Office のシェアのうえにアグラをかいてきた Microsoft も昨今のウェブアプリケーションの攻勢に相当危機感を募らせ、ようやく本気になってきた証ともいえるだろう。

微软拼音输入法 2010

### ■ IME 戦国時代の再来？

Input Method Editor、略して IME は文書作成効率を決定する重要なソフトウェアとして、MS-DOS 以来、長く研究開発が行われてきたが、二階堂氏が指摘する通り、Windows 全盛時代になってからこの方、存在を忘れられて久しい。かつては MS-DOS 上で動作する FEP（Front End Processor）として ATOK、松茸、VJE、WX シリーズなど多くの国産ソフトが存在した。それぞれ ATOK は一太郎、松茸は松というワードプロセッサのサブセットとして提供されたものだったが、変換効率の高さから、独自の発展を遂げてきた。

当時のソフトウェアハウスは少しでも変換効率を上げようと、開発に尽力したが、日本語変換効率は辞書登録単語と文節区切りのアルゴリズムに尽きるわけで、一般ユーザの目から見て、大きなブレイクスルーは余りなく、入力効率を上げるには、自分で多用する単語を登録し、「辞書を鍛える」ことが重要とされてきた。

だが、Windows95 が発売され、インターネットが普及すると、専門分野ごとの登録単語がネット上で公開配布されるようになり、状況は一変した。ATOK も 2006 以降、明鏡国語辞典や広辞苑など、各種国語辞典の参照機能に加え、専門辞書や流行語辞書を搭載し、ネット経由でアップデートできるようになった。現実問題として、いかに変換アルゴリズムを精緻にしても、入力したい単語が辞書になければ画餅に等しい。こうした事情は中国語 IME にも共通するため、今や日本でも中国でも登録単語の鮮度が重視される時代になりつつある。「かんじゃに」と打てば「患者に」の隣に「関ジャニ」が並び、中国語で「woju」と打てば「臥具」より「蝸居」が先に候補に出る時代だ。

実はそんな流行語の動向を最もよく把握している

のは検索エンジンだ。日本語ならば Google トレンド（**http://www.google.co.jp/trends**）や Goo ランキング（**http://ranking.goo.ne.jp/keyword/**）などで、中国語ならば百度搜索風雲榜（**http://top.baidu.com/**）搜狗熱搜榜（**http://top.sogou.com/**）などで、我々もその動向調査の一端を知ることができる。

その意味では数年来、搜狗、Google、百度が相次いで IME を公開配布し始めたのは当然の帰結ともいえる。検索と入力は実は一体の作業なのだ。欧米圏言語とは異なり、日本語と中国語はキーボード入力、変換作業抜きに検索語は入力できず、検索は漢字変換というフロントエンド処理なしには完成しないのだ。例えば、Google エンジニアの工藤拓は Google 検索の「もしかして」機能を担当していたが、実は該当するケースのほとんどが IME での誤変換によるスペルミスであることに気づいたところから IME 開発の必要性を確信したのだという [2]。無論、Google の場合、Chrome OS への搭載も念頭にあるのだろうが、いずれにせよ、この辺りに検索エンジンサイトが IME を提供する理由が存するといえよう。現在 Microsoft 以外に公開配布されている主な IME は以下の通りだ。

《日本語》

- ATOK2010（2010 年 2 月発売：有償）
  **http://www.justsystems.com/jp/products/atok/**
- Google 日本語 IME（2009 年 12 月公開：無償）
  **http://www.google.co.jp/intl/ja/ime/**
- Baidu Type（2009 年 12 月公開：無償）
  **http://ime.baidu.jp/type/**

《中国語》

- 搜狗拼音（ピンイン）輸入法（2006 年 6 月公開：無償）
- 搜狗雲（クラウド）輸入法（2009 年 11 月公開：無償）
  **http://pinyin.sogou.com/**
- 搜狗五筆輸入法（2008 年 8 月公開：無償）
  **http://wubi.sogou.com/**
- Google 谷歌拼音 IME
  （2007 年 4 月 Ver.1，2008 年 12 月 Ver.2：無償）
  **http://www.google.com/ime/pinyin/**

※公開時期はいずれもβ版も含む

このなかで無料配布の IME として最も早かったのは搜狗拼音（ピンイン）輸入法である。搜狗とは搜狐公司（SOHU.COM）の提供する検索エンジンで、他に五筆輸入法や拼音輸入法のサブセット版の Ajax IME の搜狗雲輸入法も公開されている。

搜狗拼音輸入法

Google は搜狗に遅れること 2 年、ようやく IME を公開したものの、Ver.1 については搜狗から登録辞書を盗用していると告発され [3]、その翌年全面的に改定した Ver.2 を公開して今日に至っている。この後、Google は中国からは撤退したものの、IME の開発は現在も継続している。日本語 IME は 09 年末から公開され、ちょうどバイドゥ・ジャパン（百度の日本検索エンジン）の提供する Baidu Type の公開時期と重なったが、バイドゥ側のインタビューによれば全くの偶然であるという [4]。

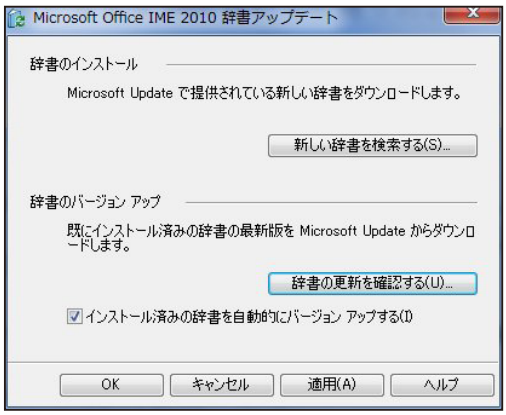
以下、日本語 IME・中国語 IME それぞれの特徴について簡単にレビューしたうえで、変換効率の比較をして読者諸兄の参考に供したい。ただし、ATOK2010 については別途レビューが用意されているので、そちらを参照いただくとして、ここでは割愛する。

## ❖ 日本語 IME

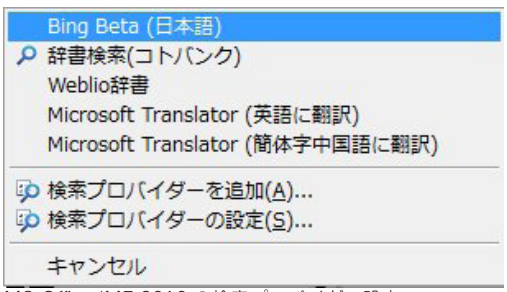
### ■ Office IME 2010 日本語

まずは Microsoft IME 2010 から始めよう。IME 2007 では「検証苑（腱鞘炎）」「部ログ（ブログ）」「社零（謝礼）」など誤変換が多数やり玉に挙げられたが、今回試した限りでは（さすがに）問題は出なかった。今回新たに追加された機能は「ツール」メニューの中にある。「追加辞書サービス」と「標準辞書の自動アップデート機能」がある。拡張辞書は現時点ではサンプル用の IT 用語辞書だけしかない。辞書のアップデート機能は標準設定では手動アップデートなので、図のように「インストール済みの辞書を自動的にバージョンアップする」にチェックを入れておくとよいだろう。

もう一つの新機能は変換中の文節を検索できる機能である。これは IME 辞書に未登録の語彙を入力する

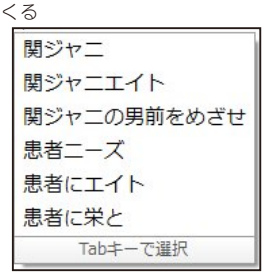


MS Office IME 2010 の辞書更新

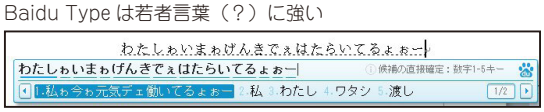


MS Office IME 2010 の検索プロバイダー設定

Google 日本語 IME ではサジェスト機能により予想外の変換候補も出てくる




Baidu Type は若者言葉（？）に強い




必要があるときに威力を発揮することになるだろう。現状では図のように Bing などいくつかのサイトに限られるが簡単な XML を書けば自分でも設定できるので、拡張性はありそうだ。

### ■ Google 日本語 IME

長年日本語変換の研究開発をしてきた Microsoft やジャストシステムに対抗して新たな IME を世に送り出すことは実はそれほど容易なことではない。それを可能にしたのは、やはり長年オープンソースソフトウェア開発に携わってきた人材を多数擁している Google だからこそだろう。

Google の IME は自動的に辞書がアップデートされるが、この辞書の内容は Google が膨大なウェブデータを統計的に処理した語彙データに基づいているという [5]。このため、規範的な日本文の作成を志向する ATOK とは目指す方向性が相当異なっている。

その特徴を最もよく表しているのが「サジェスト機能」である。入力する途中でユーザの入力したい内容を先取りして候補を示してくれるものだが、ウェブ上でのデータの実勢をよく反映しているものの、中には規範的とは言えない表現も多数含まれており、ユーザの望まない情報が表示される可能性もあるので、サジェスト機能をオフにすることも可能だ。

### ■ Baidu Type

中国の検索エンジンとして最大手の百度がリリースした Baidu Type は後発ならではの工夫に富んだ IME だ。その開発期間はインタビューによると、わずか半年であるという [6]。Baidu Type の特徴はまず第一には Google と同じくウェブデータをリアルタイムに反映する辞書だが、次に目を引くのはスキン機能であろう。確かに日本語 IME としては前例がない試みだが、後述する搜狗拼音輸入法や Google 拼音 IME にもあるもので、余り新鮮味はない。残るポイントは、「ケータイライクな入力」を PC でも実現するというものだ。例えば、変換候補には顔文字が多数登録されていたり、若者が携帯で多用する表現も変換でき、助詞「は」を小文字の「ゎ」で代用する文も変換できる。

だが、上図の通り、IME としては致命的なインライン入力ができないという欠点がある。これも携帯電話の入力スタイルを模しているとも考えられるが、広い PC 画面では目の前に変換候補が出てこないと不便きわまりない。特にデュアルディスプレイで利用する場合、変換候補はプライマリー画面にしか表示できないので、実用に堪えない。

### ■ 日本語 IME 変換文例比較

実際の変換効率を Windows 標準の IME も含めて比較してみた。[7] **標**準、**O**fficeIME、**G**oogle、**B**aidu、

ATOK の順で示す。変換ミスは網掛けで示す。

> 中朝国境の鴨緑江、大雨で氾濫、両国に洪水被害広がる（asahi.com 記事）[8]

標：中朝国境の追うりょっこう、大雨で氾濫、両国に洪水被害広がる
O：中朝国境の鴨緑江、大雨で反乱、両国に洪水被害広がる
G：中朝国境の鴨緑江、大雨で反乱、両国に洪水被害広がる
B：注腸国境のおうりょっこう、大雨で氾濫、/ 両国に洪水被害広がる
A：中朝国境の鴨緑江、大雨で反乱、両国に洪水被害広がる

変換効率評価のポイントは固有名詞が前後の文脈でどれだけ正確に変換できるか、更に文脈に応じた正確な変換ができたかどうかであるが、Baidu は Windows 標準よりも固有名詞が弱いが、大雨で「氾濫」と正しく変換できたのが Baidu と標準 IME だけで、他はどれも間違えている。文脈に応じた変換に強いと評価される ATOK も間違えている。ただし、Baidu は長文の一括変換には制約があるようで、ローマ字入力の場合 50 キー（打鍵数）ぐらいが上限のようだ。このため上記の文例も半分ずつ入力しているので多少割り引いて評価しなければならない。

以上からまとめると、ATOK、Office IME、Google は十分実用レベルに達しており、あとはユーザの好み次第だろう。Baidu はチャットなど利用目的に合致すれば使えるだろうが、汎用性が足りない。まだ実用レベルにはほど遠いのではないか。今後に期待したい。

## ❖ 中国語 IME

### ■ 微软拼音输入法 2010

Office IME 2010 に付属する中国語 IME には入力スタイルを二つから選択できる。以前の日本語 IME もそうであったように、自動的に連文節変換を行う新体験（New Experience）と手動でユーザが変換をする簡捷（simple fast）だ。基本となる辞書やアルゴリズムは同一なので、好みに応じて選べばよい。

登録辞書の自動更新や拡張辞書の提供、更に変換中

搜狗拼音輸入法の豊富なオプション設定画面

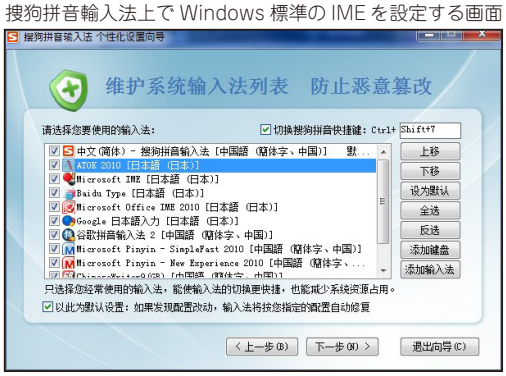
の単語のウェブ検索機能も日本語と変わらないが、検索プラグインは当然ながら追加する必要がある。提供されるプラグインは日中両言語一括で表示される。幾つか試した範囲ではどれも実用的とはいえない。自分の目的に合うものだけに精選した方が良いだろう。

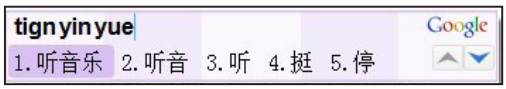
それでも従来の MS-Pinyin に慣れたユーザが使えば変換効率の向上に目を見張ることは間違いないだろう。蜗居、舟曲、织田裕二、星际争霸などの単語が簡単に変換できるだけでも大きな進歩なのだから。

### ■ 搜狗拼音輸入法

搜狗拼音輸入法は 2006 年に公開されて以来、幅広いユーザ層から支持され、事実上 Windows 付属 IME に代わるデファクトスタンダードとして愛用されてきた。公開当初から辞書の自動更新が行われ、こまめに新語の登録が行われる [9] 一方で、「細胞詞庫」と呼ばれる拡張辞書も非常に充実しており、汎用性の高いものからマニアックな辞書まで 14122 種ある。この点でも他の IME の追従を許さないところいえる。最近

搜狗拼音輸入法上で Windows 標準の IME を設定する画面





Google 拼音 IME は入力ミスを自動修正する

は変換中候補をサーバーに問い合わせ、最適解を追加表示する「雲計算（クラウド演算）」機能も追加された。これは下記の Ajax IME の機能を取り込んだもので、MS IME の変換中検索よりも使い勝手は遙かによい。

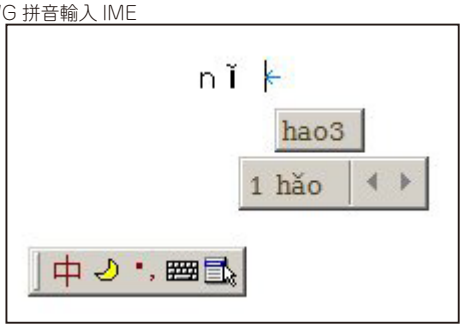
もう一つの特徴は豊富なスキンの存在だ。ほとんどがユーザ自作のものだが、アイドル系からアニメ、映画、ゲーム関連まで様々なデザインのスキンが提供されており、特に若いユーザから強い支持を受けている。現在は拼音輸入法のサブセット版として、Ajax IME の捜狗雲（クラウド）輸入法も公開されており、こちらは IME を全くインストールすることなく利用できるので、海外出張などで中国語環境が全くない時に利用でき大変重宝する。

ただし、この捜狗拼音輸入法のインストール時にオプション変更なしに標準インストールすると、捜狗を Windows 標準 IME として登録し、Windows 側で既定の言語を変更しても、捜狗側の設定に合わせて自動修復されるので、日本のユーザは若干注意が必要だ。

### ■ 谷歌（Google）拼音 IME

谷歌拼音 IME は捜狗より後発であるうえに、辞書の自動更新、拡張辞書の提供、豊富なスキンなど、あらゆる分野で先を越されてしまったので、中国では Google 撤退の一件も手伝ってか、はなはだ影が薄い。決して変換効率や操作性は捜狗に劣るものではないのだが、日本語 IME のようにウェブデータを大胆に取り込むものでもないので、どうしても捜狗の二番煎じという印象は避けられない。

WG 拼音輸入 IME




若干新味といえるのが、「主動下載個性化語言模型」で、ユーザごとの入力時の癖を記録し、それに合わせて変換候補を提示するというもの[10]。Google のアカウントごとに利用データを管理するので、複数のマシン間で入力環境を共有できるのも捜狗にはない利点だ。また、拼音のタイプミスを矯正して変換候補を提示してくれる機能もあり、細かい使い勝手は改善されている。

### ■ 中国語 IME 変換文例比較

実際の変換効率を **O**ffice IME、**S** 捜狗、**G**oogle の順で示す。変換ミスは網掛けで示す。

> 鸭绿江洪水迫使朝鲜新义州五千多人疏散转移（REUTERS 中国）[11]

O： 鸭绿江洪水迫使朝鲜新义州五千多人疏散转移。
S： 鸭绿江洪水迫使朝鲜新一周五千多人疏散转移。
G： 鸭绿江洪水迫使朝鲜心意周五千多人疏散转移。

Office IME は全くノーミスであった。前バージョンと比べると真に隔世の感がある。他の二つも僅かに新義州という地名だけが変換できなかったものの、他は問題ないという結果になった。なお、捜狗は、筆者が「鸭绿江」の「绿」を誤って lù とタイプしたところ、lǜ が正しいと「友情提示」付きで上記変換例を示し、第二候補に「雲計算（クラウド演算）」機能による正しい変換例を挙げていたことを付記して置かねば不公平だろう。

### ■ WG 拼音輸入 IME

最後になるが、ここで少々特殊な IME も紹介しておきたい。中国語の国語政策の要でありながら、その文字入力となると大変な手間がかかるもの、それが拼音（もちろん声調符号を含むもの）である。ここまで紹介してきた IME でも拼音は入力できるが、どの IME でもソフトウェアキーボードを呼び出して一つ一つ打たねばならず、ほとんど使い物にならない。

そこで開発されたのが千田大介氏による WG 拼音輸入 IME である。本来、IME 作成は容易ではないが、多多输入法生成器を利用して作成したものであるという。このソフトに変換テーブルを組み込むことで拼音への変換を可能にしてくれるので、ゼロからプログラ

ムを組む必要はなかったわけだ。

IME は氏のサイト「電脳瓦崗寨」（http://wagang.econ.hc.keio.ac.jp/）のページから WG 拼音 IME のページを開けばダウンロードできる。使用方法はごく簡単だ。入力したい拼音のローマ字に続けて、アラビア数字で声調を入力すればよい。例えば「ni3」なら「nǐ」という具合だ。ウムラウトが入る場合なら慣例通り v で代用して「lv3」とすれば「lǚ」に変換される。なお、教材作成では声調なしの拼音も必要になるが、その場合は「nve0」とすれば「nüe」と変換される。

最近は Word のルビ振り機能などを利用して一括して拼音をつけることも簡単にできるようになった。とはいえ、教材作りには、細かい軽声や異読音の修正作業は欠かせない。その際に手軽に拼音入力できるツールとして適当なものがなかった。筆者は同じく千田氏による Word のオートコレクト機能を利用した拼音入力マクロをかねてから愛用していたが、拼音入力のために Word をわざわざ起動するというのは、いかにも不便であった。この IME のおかげで拼音入力が更に簡便になった。有用なツールを次々と作成配布してくださる千田大介氏に改めて感謝の意を表したい。

### 注

[1] 《日本語》Microsoft Office IME 2010（2010年6月公開）
http://www.microsoft.com/japan/office/2010/ime/default.mspx
《中国語簡体字》微软拼音输入法 2010（2010 年 6 月公開）
http://www.microsoft.com/China/pinyin/

[2] Google 日本 Blog：「思い通りの日本語入力 ── Google 日本語入力」
http://googlejapan.blogspot.com/2009/12/google_03.html

[3] 新浪網科技時代：“搜狗和谷歌拼音輸入法抄襲之爭”
http://tech.sina.com.cn/focus/sg_pinyin/index.shtml

[4] IT メディアニュース：「『日本語を楽しく』27 歳ケータイ女子が企画した IME Baidu Type」
http://www.itmedia.co.jp/news/articles/0912/21/news020.html
および INTERNET Watch「Baidu Type で携帯電話での文字入力の便利さ、面白さを PC にも」
http://internet.watch.impress.co.jp/docs/special/20091222_338974.html
による

[5] ASCII.jp ×デジタル「IME として『変態』開発陣が語る Google 日本語入力」
http://ascii.jp/elem/000/000/484/484639/

[6] INTERNET Watch「Baidu Type で携帯電話での文字入力の便利さ、面白さを PC にも」（前掲注 4）

[7] 以下、日本語、中国語とも公平な比較のため登録辞書、学習効果はリセットしたうえで行ったが、現在の IME はオンライン更新が行われるため完全な再現性は保証できないことを予め了解頂きたい。

[8] asahi.com ニュース（国際アジア記事）
http://www.asahi.com/international/update/0822/TKY201008220187.html

[9] 更新には搜狗網の ID が必要になる。

[10] Google 谷歌拼音輸入法功能介绍
http://www.google.com/ime/pinyin/feature.html

[11] 路透财经报道
http://cn.reuters.com/article/cnBizNews/idCNnCN127206520100823?feedType=RSS&feedName=cnBizNews

# 一太郎2010 & ATOK2010

山田　崇仁

## ❖ はじめに

本誌10号公刊直後、ジャストシステムの創業者浮川和宣・浮川初子夫妻が、一身上の都合により突然の辞任を発表したことは [1]、驚きをもって迎えられた [2]。

その後両氏は、株式会社 MetaMoJi（メタモジ）[3] を立ち上げ、新しいビジネスに取り組んでいる。辞任のニュースを知ったときには、一つの時代が終わったと感じたのだが [4]、まだまだご夫妻の意気は盛んなようである。夫妻の新たなビジネスと、ジャストシステム新時代の成功を共に祈りたい。

## ❖ 一太郎 2010

### ■ 25 周年

一太郎も本年で発売 25 周年となった。ジャストシステムの Web サイトでは、記念スペシャルコンテンツ関連のページを公開している [5]。

また、初代一太郎が社団法人情報処理学会が制定する「情報処理技術遺産」に認定された [6]。しかし、一太郎は単なる博物館に陳列される過去の遺物には終わらない。現在もなお活きている商品なのだ。

現行の一太郎 2010 は、最新の Windows 7 に標準対応し、新しい機能も搭載している [7]。Twitter のアカウント @ichitaro_js を取得してユーザとの活発な交流を試みるなど、なお意気軒昂である。

しかし、現実はそれほど明るくはない。一太郎の市場シェアは Microsoft Word の後塵を拝して久しい。一太郎が強いと言われているはずの教育や行政分野でも、Word はもとより OpenOffice.org や PDF といった他のフォーマットの攻勢に押されている。また、Google ドキュメントや Office Web APP などのクラウド系文書作成ツールが虎視眈々と文書作成ツールの占めるスペースを狙っている。

無論ジャストシステムでも、行政文書作成専用の一太郎ガバメントや警察機関向けのパックをリリースするなど、国内メーカーの強みを活かした分野で失地回復に取り組んでいる [8]。行政文書のような特殊なフォーマットを持ち [9]、一定のシェアが見込まれる分野に特化した商品を展開するのは一つの戦略眼といえる。それがどこまでシェアの奪回に結びつくかどうか不明だが、レビュワーとしては、一太郎を応援したい。

### ■ 長年使っているが故の不満

レビュワーの一太郎使用暦は約 20 年である。一太郎は自分の論文書きには欠かせないツールだけれども、日常的な使用機会は減っているのが現状である。

例えば、事務的な文書類は Word の文書や PDF でのフォーマット指定が多いし、勤務先の授業で受講生に教える文書作成アプリケーションは Word であり、レジュメの作成には Word や InDesign を使用している。そのため、一太郎文書を他人に渡したり渡されたりする機会は、特定の人同士という非常に限られた状況でしかないのも事実である。

それでも、自分が論文を作成する際に使用するアプリケーションは一太郎である。これは、惰性の部分が大きいのかもしれないが、自分が作成する文書やレイアウトと一太郎がよくマッチするのがその理由である。

やはり、Word のルビを入れると行間が自動で調整されてしまう仕様や [10]、脚注番号の指定（番号の前後に括弧を指定できない）など、個人的な定番設定ができないのが不満なのである。この辺りに対応している一太郎は、自分の長文作成には相変わらず欠かせないツールであることは間違いない。

但し、万々歳というわけではない。もちろん足りないところもある。先ほどの Word との比較で言うと、以下の例を挙げることができる。

#### ●箇条書きへの対応が不十分である

これは、バージョン 2009 から一応対応するようにはなったが、デフォルトのツールバーには表示されないし、無論 Word や HTML の箇条書きや番号付きリストのような気軽な使い方とはほど遠いといえる。

#### ●Windows 標準のキー割付モードが欲しい

一太郎には、ESC キーでメニューを表示する伝統的キー割付や、ジャストウィンドウの系譜を引くキー割付が採用されている。無論それはそれで慣れている指にはありがたいのだが、そろそろ Windows 標準のキー割付のセットを標準で添付して選択可能にしてもよいのではないか。多くのアプリケーションがそれに対応している以上、孤高を守る一太郎の態度は、新規ユーザ（厳しいようだがもしそれが存在するのであれば）の敷居を高くする一つの要因ともなろう。

#### ●添付フォントを JIS2004 対応にして欲しい

国語審議会の答申からはや十年、JIS 漢字の改訂からも 6 年経過した。Windows の標準書体である MS 明朝やゴシックや、一太郎の標準フォントである JS 明朝やゴシックも無論印刷標準字形に対応済みである。しかし、それ以外の添付フォントは相変わらず旧来の字形である。Microsoft Office は 2010 で添付される HG 書体の多くが印刷標準字形準拠のものとなった。これにならって、一太郎の添付フォントも JIS2004 に完全対応していただきたい。

その他、細かいことを言うと、ナレッジウィンドウの［スタイル］に表示される段落スタイルのカスタマ

イズ（表示される項目指定やリストをスクロールして欲しいなど）、インターネットディスクへの保存・読み込みは可能だけれども、クラウド形の一太郎ってリリースされるのだろうか？　などあるが、今のところ使い勝手としてはそれなりに満足している。従って、一太郎は当分自分の手を離れることはないだろう。自分の中では愛着のある文具となっているのである。

## ❖ ATOK2010

### ■ たゆまざる進化

ここでは ATOK を採り上げるが、IME 全般については、本誌で小川氏によるレビューが掲載されている。そちらと比較しながらお読みいただきたい。

ATOK2010 では、根幹部分のハイブリッドコアを新しくし、同音異義語などの変換効率をアップさせている。また、変換効率を上げる重要なポイントである文節区切りを改善し、更には新しい語彙を積極的に取り入れるなど、より使いやすい IME を目指している。

加えて、従来から定評があった誤変換の指摘も、単に指摘するだけではなく、より妥当と思われる候補を表示するなど、変換・校正の機能も充実させている。

### ■ ライバルが台頭する IME 業界

ATOK の特徴は変換効率であると、よく言われる。無論これはよい意味での評価であり、キーボードから入力した文字列から、入力者の意図に近い語彙をスムーズに表示する効率性の高さを評価したものである。

ATOK の歴史は、一太郎の歴史とほぼ並行しており、その中で、MS-DOS 時代から続くライバルとの争いの中でそれを磨き洗練させてきた。

ここ数年は、ATOK のライバルは実質〝MS-IME〟と〝ことえり〟との二つであったが、昨年来、Gogole や Baidu など、インターネット分野からの IME 参戦が相次ぎ、特に Google IME の固有名詞を中心とした変換効率の高さが知られることにより、ATOK の牙城もかくや！という雰囲気になった。これについては、上記小川氏のレビューをお読みいただきたいが、ATOK はこれらの新興 IME とどのように違うのかについて少し考えてみよう。

### ■ 日本語の規範化を支える ATOK

上記レビューで、小川氏が朝日新聞の見出しを例に挙げて、ATOK の事例を報告されていた。

> 中朝国境の鴨緑江、大雨で氾濫、両国に洪水被害広がる（asahi.com 記事）[11]

小川氏の事例では、ATOK の変換結果は以下の通りとなった。これは筆者の環境（Windows 7 64bit 版＆ ATOK2010）でも変わりない。

> 中朝国境の鴨緑江、大雨で反乱、両国に洪水被害広がる

「反乱」→「氾濫」が正しい変換例であることはそのとおりだが、おもしろいことに、この文章を少しいじるだけで「反乱」が「氾濫」になる。

例えば、「中朝国境の鴨緑江、大雨で川が氾濫」と、氾濫した対象に「川」を与えると目的の「氾濫」が、また、「中朝国境の鴨緑江で、大雨で川が氾濫」と助詞を補ったり、「中朝国境の鴨緑江、大雨で氾濫。」と読点を句点に買えても同様に「氾濫」となる。

ここから考えるに、ATOK の場合、文章の中で目的語を適切に指定しない、或いは助詞や句読点の指定が適切に行われていないと、変換効率が落ちる可能性があると言うことである。確かに、日本語の文章として考えた場合、上記見出しは氾濫の後で句点をつけた方がふさわしく思えるが、これは新聞の見出しであり、そのため一気に読んで貰うことを前提にした視覚的配慮として、句点を採用しなかった可能性もある。

こういった事例を考えると、日本語の文書変換の規範性とユーザーの選択権の幅をどう考えるのか、それが IME の作り手側にとって腕の見せ所であり、難しいところでもあるのだろう [12]。

インターネットで使用される、いわば〝現在の言葉〟を基盤とする Google や Baidu に比べると、ATOK は最新の用語に疎い可能性は否定できない（標準辞書に胡錦濤は登録されるものの、習近平は未登録）。しかし、前者は、その性質上過去の言葉やたまたま入力されてしまったものに対して弱い部分がシステム上否定できない（その辺りの効率化こそ、鍵でもあるが）。ATOK はその部分に強みと蓄積がある [13]。

IME は表に出にくいアプリケーションでありながら、実は一番使用しているアプリケーションでもある。

あらゆる情報機器で規範的な日本語を届けようとす

〝にほんご〟と入力し〝日本語英語辞書〟で変換した結果

〝こきんとう〟と入力し〝標準辞書〟で変換した結果

る ATOK。iPad など新たな情報機器にも積極的に対応しているその姿にあらためてエールを送りたい[14]。

### 注

[1] **http://www.justsystems.com/jp/just/finance/j0910291.pdf**

[2] 風穴江のカッティングエッジ：ジャストシステムに何が起こったのか？ **http://pc.watch.impress.co.jp/docs/column/cuttingedge/20091104_326144.html** など。

[3] **http://www.metamoji.com/jp/**

[4] 2010 年 8 月、PC-9801 シリーズ全盛期に一太郎のライバルであった、ワープロ松シリーズやデータベース桐などで知られた、管理工学研究所の関根智明が亡くなられている。一つの時代の区切りという意味では、確かにそうなのかもしれない。**http://www.kthree.co.jp/obiturary.html**

[5] 一太郎ガバメント 2010：**http://www.justsystems.com/jp/camp/just2010/**
警察機関向けオフィス統合ソフト JUST Suite Police 2010：**http://www.justsystems.com/jp/products/justsuite_police/**

[6] **http://museum.ipsj.or.jp/heritage/ichitaro.html**

[7] 旧バージョンとの比較は、以下の URL を参照されたし。**http://www.justsystems.com/jp/products/ichitaro/comparing.html**

[8] **http://www.justsystems.com/jp/products/ichitarogov/**

[9] 公用文に関する諸通知（文部科学省：国語表記の基準）：**http://www.bunka.go.jp/kokugo/frame.asp?tm=20100916144201**

[10] 一応、Word でも行の設定で行間のサイズを固定したり、［1 ページの行数を指定時に、文字を行グリッド線に合わせる］のチェックを外せすことで、ある程度この辺りの振る舞いを指定可能となる。

[11] asahi.com ニュース（国際アジア記事）**http://www.asahi.com/international/update/0822/TKY201008220187.html**

[12] これについては、ジャストシステムの Web ページ「ATOK の辞書を語る」でも同様な話が掲載されている。**http://www.justsystems.com/jp/tech/atok/dic/**

[13] よく使われるネットスラング「ふいんき」を入力すると、ATOK2010 では「雰囲気《「ふんいき」の誤り》」と誤変換の指摘と正しい候補漢字が表示される。

[14] iPad 用の IME ではなく、メモ形式のアプリとして ATOK Pad をリリースした。

# Adobe InDesign CS5

山田　崇仁

## ❖ はじめに

Adobe 社のリリースする DTP ソフトである InDesign も、はや CS5（Ver.7）にまで至った。本誌も InDesign で DTP を始めてからずいぶん経つが、体感的にここ数年はおおむね二号につきバージョンが 1 つ程度上がっているかと思う。

無論、その間、DTP や印刷された本を巡る環境はずいぶん変化した。その辺りは本誌でも何度となく採り上げているし、本号でも電子書籍について特集している。そのような環境下において、DTP ソフトはど

のように対応しまた変化しようとしているのか。

### ■ アカデミックパックの販売形態変更

まず、ソフトウェア自体の変化について採り上げる前に、販売形態の変化について採り上げておこう。

本誌の読者層の多くは、いわゆるアカデミックな業界に関係する方々だと思われる。ご存じのように、ソフトウェア会社ではアカデミック業界を対象にした割引販売を行っている所が多い。InDesign をリリースしている Adobe 社もその一つであり、多少制限があるものの高価な同社のソフトウェアを一般市場価格より相当の安価で購入できる形態はありがたい存在であった。

この Adobe 社のアカデミックパックの販売形態が、2009 年 10 月 16 日より、変更となった。

従来は研究室単位で購入可能であったのが不可能となり、あくまで個人単位での購入となったのである。無論、研究室のような単位ではなく、それを差配する教員が責任を持って購入すればよいのだが、予算の関係上そうも行かないところもあるだろう。或いは、学校単位でのライセンスに切り替えたところもあるかもしれないが、レビュワーはそのような状況とは無関係なためよく分からない。

また、認証方法も変更になった。従来は購入時に、店頭でアカデミックパック購入資格を示す書類（学生証・教員証・在籍証明書など）を提示し、必要書類に記入してから購入できたのた。

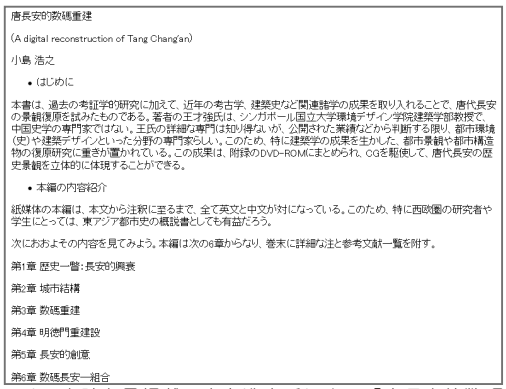
しかし、新しい購入形態では、購入時には何も必要とせず、認証時に Web 経由等で必要な手続き（フォームへの記入・証明書の画像データを送付など）を行った上で、シリアル番号を Adobe 社から送付して貰うという形に変わったのである。

これらの変更は、単に Adobe 社のマーケティング方針にしか過ぎないのだろうが、混乱している部署もあるかもしれない。ただ、そのうちになれることだろう。

## ❖ InDesign

### ■ 紙以外への出力

上記のように、昨今は電子ブックをはじめとする電子出版についての検討が盛んに行われるようになった。InDesign でも電子媒体への書き出しについては従来から力を入れており、特に同社に関連する規格である


唐長安的数碼重建

(A digital reconstruction of Tang Chang'an)

小島 浩之

• はじめに

本書は、過去の考証学的研究に加えて、近年の考古学、建築史など関連諸学の成果を取り入れることで、唐代長安の景観復原を試みたものである。著者の王才強氏は、シンガポール国立大学環境デザイン学院建築学部教授で、中国史学の専門家ではない、王氏の詳細な専門は知り得ないが、公開された業績などから判断する限り、都市環境(史)や建築デザインといった分野の専門家らしい。このため、特に建築学の成果を生かした、都市景観や都市構造物の復原研究に重きが置かれている。この成果は、附録のDVD-ROMにまとめられ、CGを駆使して、唐代長安の歴史景観を立体的に体現することができる。

• 本編の内容紹介

紙媒体の本編は、本文から注釈に至るまで、全て英文と中文が対になっている。このため、特に西欧圏の研究者や学生にとっては、東アジア都市史の概説書としても有益だろう。

次におおよその内容を見てみよう。本編は次の6章からなり、巻末に詳細な注と参考文献一覧を附す。

第1章 歴史一瞥：長安的興衰

第2章 城市結構

第3章 数碼重建

第4章 明徳門重建設

第5章 長安的創意

第6章 数碼長安一組合

図 1：本誌九号掲載の小島浩之氏による「唐長安的数碼重建」の一部を電子ブック化したもの（FireFox 用 Plugin EPUBReader で表示）

PDF や Flash や電子書籍の標準フォーマットの地位を目指す EPUB への書き出しも対応するようになった。

昨今、PDF 入稿が当たり前のようになってきた。本誌でも以前から入稿に際して採用しているが、やはり InDesign の PDF 書き出しは規格開発元だけあって、各種 PDF 形式への出力に対応している。

また、Flash への書き出しについても対応している。これは、Web 上で公開されるチラシや PDF での公開では少々おとなしいと感じられるようなコンテンツを対象にしているのだろう。細かい設定をしないで書き出しても、ページめくりのアクションがついてきたり、なかなか楽しい形に仕上がっている。

では、EPUB 形式はどうだろうか。これについては、規格自体がまだ発展途上と言うこともあるので、日本語組み版からの書き出しは “とりあえず対応している”

図 2：図 1 と同じモノを SWF 型式で出力し、Internet Explorer8 で開いたもの。

程度の印象しか持たなかった。特に縦書きは、規格にないこともあって当然未対応だし、元のレイアウトが反映されていない（それだけレビュワーのいい加減な素人 DTP がばれてしまうだけのような気もするが）部分もあり、まだまだ実用にはほど遠い。

今のところ、電子媒体への出力という意味では、PDF への書き出しがもっとも無難だろう。

### ■ 便利になった機能

InDesign 日本語版は、当初から Adobe-Japan を利用した多漢字（多字形というべきだろうか）への対応や、日本語組み版を意識した製品作りを行ってきた。そのため、DTP 分野では後発ながらそれなりのシェアを占めるまでに至った。

CS5 でも新機能の追加は続いている。

例えば、一つのデータファイルの中で複数のページサイズを指定可能になったのが挙げられよう。その他にも、新聞や雑誌の見出しでよく使われる複数の段をまたいで見出しを構築する［段抜き見出し］機能や、その逆の一つの段落を複数の列に分割する［段抜き］機能などが挙げられる。これは従来、フレームを一段或いは複数段並べて実現していたのだが、そのような面倒な部分が解消されたのは、作業の効率化にとっては喜ばしいところだろう。

しかし、まだまだ日本語組版との関係で言うならば物足りないところもある。例えば、文字や段落の周囲を線で囲むいわゆる［囲み線（囲み罫）］機能や、文字列への網掛けについて是非とも実装して欲しい。

### ■ 廃止された SING 外字

漢字文献を中心に文章を書くレビュワーにとって、CS5 ではとてもショッキングな仕様変更があった。それは、SING 外字の廃止である。

Adobe 社は CS2 以来、独自の外字規格である SING 外字ソリューションを提供してきたが、CS5 からはそれを取りやめた。Adobe 社の説明から判断するに、OTF の充実や他の外字フォントとシェア争いで敗れたことが原因と推測されるが、いきなり廃止されるのは困ったものである。作成機能はともかく、SING 外字を管理するグリフレットマネージャーだけでも添付して欲しかった。

レビュワーは、この規格で作字した外字が大体 200 字程度ある。そのうち半分くらいは、既存の OTF フォントに無い Unicode 収録文字で、残りの半分が出土文字資料を隷定したものを、既存のフォントのデザインをベースに作ったものなどがある。特にレビュワーの関連する分野の論文に出土文字資料を用いたものが多いため、Unicode 未収録字が大量に発生する原稿を DTP する際に、SING 外字は欠かせないものとなってきた。

確かに、SING 外字は必ずしも使い勝手がよい規格とは言えなかった。その意味では廃止もやむを得ないかな、とも思っている。

たまたまレビュワーは、規格が死んでも大丈夫な保険として、一時作字する毎に SING 外字形式とは別に Adobe 社の Illustrator 形式でも保存してあったので、それをインライングラフィックで使用するなり、他のフォントソフトを使って True Type Font、或いは Open Type Font として作成し直すことで、まあなんとかなるだろうと考えている。

本誌の特集 2 でも採り上げているように、デジタルなものであっても、規格自体が陳腐化・化石化することで、既存のリソースが無駄になってしまうことは避けられない。そのため、最低限、どのレベルが救われればよいのかを考えつつ、データを作成していく必要について改めて実感させられた次第である。

### 注

[1] 前者については、株式会社エステンナイン京都から InDesign 用のプラグインが提供されている。また、後者については下線機能を工夫することで擬似的に編みかけ状態にする裏技が知られている。
株式会社エステンナイン京都：**http://www.st-kyoto.co.jp/idPlugin/**
InDesign の勉強部屋／ No.50 下線を利用した網かけ。情報は InDesign　CS のものだが、CS5 でも利用可能。
**http://study-room.info/id/studyroom/cs1/study50.html**

[2] Adobe SING Glyphlet Manager について（InDesign CS5）
**http://kb2.adobe.com/jp/cps/850/cpsid_85080.html**

[3] 出土文字資料の字形をもとに、我々が日常的に用いる楷書体や明朝体に近い形にデザインしなおす作業。

[4] その場合、フォントを作成するならば、既存の文字表を上書きする形での独自文字表形式のフォントになるだろう。

# 学術リソース レビュー

2010 年度の学術情報レビューをお届けする。

ここでもキーワードとなりそうなのは「電子書籍」だろう。本コーナーでは、中華圏の電子書籍に関するレビューが掲載されている。これは単に中華圏のみの問題だけではない。編集中も気になる動きを見せる漢字文化を巡る電子書籍市場を理解する手助けとなれば幸甚である。

本年、メジャーな用語となった電子書籍だが、実は日本ではそれ以前から特定分野で電子書籍が大きな地位を占めてきた。それは辞書の分野である。昨今、同じ辞書を書籍とデジタルデータで平行して販売する動きも珍しくなくなっている。気になる方は是非とも本レビューをご一読いただきたい。

本誌読者の中には、Twitter のアカウントをお持ちの方も珍しくないだろう。それを授業に使うとどうなるのか（どうなったのか）。本号では誰しも関心を持ちそうなテーマについて取り組んだ結果についてレビューをお願いした。

学術関係のレビューでは、中国古典系のリソースの過去を振り返り、未来への展望はどうなりそうかという観点からレビューを行った。併せてスタンドアローンからクラウド型の変化を企図する、書同文のデータベースのレビューも参考にしていただきたい。

また、かつて漢字を主要な文字として採用し、またそこから字喃をうみだしたベトナムのデジタルリソースについても採り上げた。御参考になれば幸いである。

## Contents


| | | | |
|---|---|---|---|
| **学術リソース** | 中国・台湾の電子書籍の動向 | 安藤　一博 | 152 |
| | 図書館と OPAC | 小島　浩之 | 160 |
| | ベトナムの漢字・字喃文献 | 矢野　正隆 | 163 |
| | 中国古典学術リソース、回顧と展望 | 秋山陽一郎 | 170 |
| | 大学授業への Twitter 導入の事例報告<br>── 大妻女子大学「情報社会論」での経験からの 14 の教訓 | 岡本　真 | 176 |
| **学術ソフト・製品** | DVD 版内村鑑三全集 | 當山日出夫 | 182 |
| | 講談社『中日辞典』第三版 CD-ROM | 田邊　鉄 | 186 |
| | 『四部叢刊』と書同文社のオンラインデータベース | 千田　大介 | 188 |

# ❖学術リソース

## 中国・台湾の電子書籍の動向

安藤　一博

### ❖ はじめに

中国や台湾では電子書籍市場が急速な成長期と大きな転換期を迎え、巨大な華人市場が形成されつつある。今回は中国と台湾の電子書籍の動向についてこの1、2年の動向を中心に紹介したい。

### ❖ 中国

#### ■ 概要

中国図書商報社と書生読吧網が2010年4月に公開した『中国電子図書発展趨勢報告』2009年版[1]によると（以下、この項で特に言及のない場合は『中国電子図書発展趨勢報告』2009年版による）中国の電子書籍の利用者人口は2009年に1億を超えたということだそうだ。2009年の中国のインターネット人口が約3億8400万人なので[2]、インターネットユーザーの約4分の1が電子書籍を利用していることになる。ただし、この報告書で「電子図書（電子書籍）」と定義されているものは「系統性と一貫性を有するデジタル化されたテキストで構成されたもの」であり、必ずしも紙の書籍が電子化されたものを指しているわけではない。「1億人」という数字はその点を考慮する必要がある。

その1億人が電子書籍をどのような手段で読むかを示したのが表1の統計である。

専用の電子書籍端末を使用する利用者は中国ではまだ少なく、デスクトップPCやノートパソコンのような、いわゆる普通のPCで読む割合が約9割を占めている。2010年には中国で300万台の専用端末が販売されるという予測もでているが、それでも1億人の利用者人口から見れば3%の増加にすぎない。

この状況は中国の電子書籍市場の構造に非常に大きな影響を与えている。欧米ではAmazonやApple、Barnes & Nobleのようなコンテンツと端末を紐づける垂直統合型のビジネスモデルが主流となっているが、コンテンツが端末に縛られない中国ではコンテンツプロバイダーが大小無数に存在しており、百度などの検

表1：電子書籍を読む手段[3]

| | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|
| デスクトップ | 92.1% | 86.9% | 85.9% | 76.1% |
| ノートPC | 4.3% | 6.5% | 7.2% | 12.3% |
| 携帯電話 | 2.7% | 5.9% | 6.3% | 10.7% |
| 専用電子書籍端末 | 0.8% | 0.4% | 0.3% | 0.6% |
| その他 | 0.1% | 0.3% | 0.3% | 0.3% |
| 合計 | 100% | 100% | 100% | 100% |

表2：主な電子書籍の販売収入[4]

| | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|
| 電子図書館 | 11000万元（74.3%） | 12100万元（71.4%） | 12200万元（53.9%） | 12250万元（42.7%） |
| 有料購読 | 3000万元（20.3%） | 3700万元（21.8%） | 6900万元（30.5%） | 9700万元（33.8%） |
| 携帯電話による読書 | 300万元（2.0%） | 650万元（3.8%） | 3030万元（13.4%） | 5760万元（20.1%） |
| 専用電子書籍端末による読書 | 500万元（3.4%） | 490万元（2.9%） | 500万元（2.2%） | 700万元（2.4%） |
| その他 | | | | 300万元（1.0%） |
| 合計 | 14800万元（100%） | 16940万元（100%） | 22630万元（100%） | 28710万元（100%） |

索エンジンで「電子書」（中国語で電子書籍の意味）で検索すると電子書籍を販売するサイトを大量に見つけることができる。専用端末はまだ普及する前の段階だが、それでも数十のメーカーがすでに市場に参入して新しい端末を次々と発表している。流通する端末の種類の多さなら中国はおそらくすでに世界一であろう。中国はコンテンツプロバイダーと閲覧機器（専用の端末だけではなく、PC も含む）メーカーが分かれている水平分業型に近い構造になっているのだ。

では、電子書籍市場の収益モデルはどのようになっているだろうか。それを示したのが前頁の表 2 の統計である。

表 2 の「電子図書館」というのは、図書館を対象とした法人契約によるものを指している。電子ジャーナルの電子書籍版と考えていただければよい。2006 年は電子書籍市場の 7 割強を図書館市場が占めている。図書館が電子書籍市場を支えていたことになる。しかし、図書館市場はすでに成長が止まっており、2009 年も依然として図書館市場は全体の中で大きな割合を占めているものの、他の収益モデルが成長を続ける中で図書館の割合は減っていく傾向にある。

表 2 の統計によると、2009 年には市場全体で約 2 億 9 千万元（約 37 億 7000 万円）の販売収入があったことになる。しかし、中国政府の新聞出版総署が 2010 年 7 月に公開した統計 [5] では 2009 年の電子書籍（電子雑誌を除く）の総売上が 14 億元（182 億円）となっている。統計のとり方はもちろんであるが、「電子書籍」の定義の違いもこのような数字の違いを生んでいるのだろう。

### ■ 中国の出版事情 ──「地下出版物」と海賊版 ──

公式の統計 [6] によると 2008 年の図書の出版点数は新刊本だけで 148,978 タイトルに上るが、その約 15 万タイトルを発行する出版社の数は 579 社しかない。この出版社の少なさは中国の出版業界の大きな特徴の 1 つだ。

しかし、上の出版社数と出版点数はあくまで公式の統計上の数字であり、中国政府（新聞出版総署）から公認を得た出版社とその出版物の数である。これら公認の出版社に加え、その数をはるかに上回る非公認の出版社が中国には存在すると言われている。その非公認の出版社が出版するいわゆる「地下出版物」が相当数流通しているらしい。海賊版、ポルノ、発禁された本、政治的に問題があるものなどが地下出版物として出版されているようだ [7]。地下出版物の電子書籍版が Web 上で流通することは当然ありうることで、海賊版の電子書籍は探せば簡単に見つけることができることから、すでに「地下出版物」の電子書籍も Web 上を大量に流通しているのではないかと思われる（もちろん個人が作成した海賊版も大量に流通しているだろう）。こうした地下出版物や海賊版の電子書籍は公式の統計には反映されないものであり、数字からこぼれた電子書籍が大量に存在しうることは認識しておく必要がある。

### ■ 電子書籍のコンテンツプロバイダー ── 方正と盛大文学 ──

前述のように中国には無数のコンテンツプロバイダーが存在する。そのすべてを紹介することはできないので、電子書籍のキープレーヤーとしてよく名前が挙がる方正と盛大文学を紹介したい。

#### ●方正

方正集団は 1986 年に北京大学の校弁企業として設立された企業グループである。傘下に IT 系の企業だけではなく、製薬業、不動産業、金融業などを営む企業などを抱えている。

その方正集団の中で電子書籍について重要な役割を果たしているのが方正阿帕比技術有限公司（以下、方正阿帕比）である。方正阿帕比は著作権保護技術の方正 DRM、電子書籍のフォーマットの CEB、CEBX など、方正集団が持つ電子出版の技術を出版社や新聞社などにソルーションとして提供している企業である。

方正阿帕比は権利処理を行い、正規版のコンテンツを提供することを売りにしており、60 万冊の電子書籍データを保有している [8]。これらのデータは法人契約している図書館や個人向けサービスの番薯網 [9]（同じ方正集団の北京方正網絡が運営）などの方正のプラットフォームを通じて提供されている。方正阿帕比にとって図書館は重要な顧客となっており、方正阿帕比の収益のほとんどが図書館との法人契約によるもので占めてられている [10]。中国の図書館が「所蔵する」電子書籍の 85％が方正阿帕比のデータであるそうだ。コンテンツが正規版であることを武器に方正阿帕比が電子書籍市場で大きな位置を占める図書館市場をほぼ独占してきたことが分かる。

しかし、表２のように電子書籍市場における図書館市場は成長が頭打ちになっており、図書館に収益を依存したままでは今後の成長が見込めないため、個人向けのサービスにも力を入れているようだ。３月には番薯網が「いつでも・どこでも・どの端末でも」というコンセプトの「雲閲読」（クラウドリーディング）というサービスを発表している。

方正集団は日本市場への進出も積極的で方正株式会社という日本法人を 1996 年に設立している。方正集団は縦書きなど日本の出版物の要件に対応させた JEBX というフォーマットを開発し、方正株式会社を通じて日本の出版社や新聞社に売り込みをおこなっているらしい。また、方正集団傘下の方正飛閲が開発した WeFound という電子書籍端末を日本市場への投入することも視野に入れている。

●盛大文学

盛大文学は個人からの投稿を受けてサイトにコンテンツを掲載するオンライン文学サイトを運営している企業である。上海のネット企業、上海盛大網絡の子会社である。

盛大文学は起点中文網、晋江文学、榕樹下など複数のオンライン文学サイトを運営している。110 万人の作家（一般のユーザー）が絶えずそれらのサイトに作品を投稿しており、盛大文学は現在、約 300 万タイトルの作品を保有しているそうだ。ほとんどが個人によって投稿された作品であるが、王蒙、莫言、阿来、蘇童などの有名作家約 200 人と契約するなどして、プロの作家の作品の収集にも力を入れている。数万部ではあるが、電子書籍のデータも保有しているようだ。

2010 年３月には「一人一書計画（One Person, One Book 計画）」という将来計画を発表し、新しい電子書籍のプラットフォームの構築を発表している [11]。また、８月には Bambook という電子書籍端末も発売し、端末市場への参入も果たしている。

海外への進出も積極的ですでに台湾、シンガポールにサイトを立ち上げ、アメリカ版のサイトの立ち上げも発表している。

2010 年３月には中国の検索エンジンの雄、百度が電子書籍市場に参入することを表明している。詳細はまだ明らかになっていないが、百度が参入すれば、有力なコンテンツプロバイダーになるかもしれない。

■ 電子書籍端末市場

専用端末で電子書籍を読むという習慣がまだ根付いていない中国だが、2010 年に入り、端末市場に続々と企業が参入している。2009 年に 30 に満たなかった企業数が 2010 年には 400 になるという見方がされている [12]。2009 年に約 80 万台であった端末の販売台数が、2010 年には 300 万台に伸びるという予測も出ている [13]。すでに中国は 2010 年上半期において出荷数で米国に次ぐ世界第２位の市場になっている [14]。世界市場のおけるシェアは米国が 58.1%、中国が 21.4％で米国市場との差はまだ開いているものの、中国の電子書籍の利用人口と専用端末の利用人口の現在の少なさを考えれば、中国の端末市場は成長する余地をかなり残している。

その中国の電子書籍端末市場で高いシェアを誇っているメーカーが漢王である。2010 年第一四半期では 65.6％のシェアを漢王が握っている。続いて台電 13.2%、翰林 4.0%、方正 3.8%、ソニー 2.4%，その他のメーカーが 11.0% と続く。また、いわゆる「山寨機」と呼ばれるノンブランドの電子書籍端末もすでに流通している。

中国の市場で流通している端末は以下のサイトで仕様も含めて確認することができる。中国で流通している端末の数の多さを見ていただきたい。

●中关村在线电子书频道

**http://ebook.zol.com.cn/**

ほとんどの端末が TXT 形式、PDF 形式、DOC 形式、HTML に対応している。これらのフォーマットに加え、CHM 形式や FB2 形式、その他 XLS 形式、PPT 形式などのドキュメントフォーマットに対応するものもある。MP3 などの音声形式や JPEG などの画像形式にはほとんどの端末が対応している。上のフォーマットに加えて、今年に入ってから EPUB に対応させる端末が次々と発売されるようになってきた。最近発表される新機種のほとんどが EPUB に対応させていると言ってよい。Apple の iPad の登場によって中国における EPUB の存在感が増したのだろうか。

価格帯は 1000 元を切るものから 3500 元（約 45,000 円）を超えるものまで幅広い。全体的に低価格化の傾向にある。電子ペーパーを採用した端末で 1,000 元（約 13,000 円）を切ることは少し前まで中

国でも難かったようで、1,000 元を切る端末のほとんどが TFT 液晶の端末だった。しかし、盛大が 8 月に発売した Bambook のように 900 元台の端末が現れてきている。7 月末に発表された Amazon の Kindle 3 がこの低価格化の動きに大きな影響を与えているのだろう。Kindle 3 WiFi モデルの 139 ドルという価格は単純に人民元に換算すれば 950 元を切る価格になる。これは中国市場でも電子ペーパー採用端末の最安値の価格帯に位置するのだ。

TFT 液晶採用の端末はさらに低価格化が進み、500 元（約 6,500 円）を切るものも現れてきている。

市場が拡大すればよりコストを抑えて生産することが可能になるので、この低価格化の傾向はしばらく続くことになるだろう。

前に少しふれたが、Apple が iPad を中国でも発売するという噂が流れている（Apple 中国のサイトには iPad を紹介するサイトがすでに掲載されている）。また、Amazon の中国法人、卓越亜馬遜が Kindle を中国市場に投入して、中国書の電子書籍を販売するという噂もある。この『漢字文化情報理研究』11 号が刊行される頃にはどちらも販売されているかもしれない。
※追記　9/17 に iPad WiFi 版が中国で発売された。

## ■ 乱立するフォーマット

様々な企業が電子書籍市場にコンテンツプロバイダーとして参入し、中には独自フォーマットでコンテンツを販売している企業も存在するため、中国は企業独自の電子書籍フォーマットが乱立している。電子ジャーナルなどで使用されているものを含めると、中国独自の電子出版フォーマットは表 3 のとおりになる。なお、現在利用されていることが確認できたフォーマットに限定している。

表 3：中国独自の電子出版フォーマット

| フォーマット | 企業名（提供サイト） |
|---|---|
| CEB/CEBX | 方正阿帕比（阿帕比電子書店、番薯網） |
| PDG | 超星公司（超星数字図書館） |
| SEP | 書生（書生読吧） |
| HEB | 漢王（漢王書城） |
| STK | 成都閲通互聯信息服務（愛捜書網） |
| MEB | 中国移動 |
| CAJ | 清華同方 (CNKI) 電子ジャーナル等 |

表 3 で挙げた独自フォーマットに加え、PDF 形式、テキスト形式、HTML 形式、CHM 形式、EPUB 形式が電子書籍、電子出版のフォーマットとして利用されている。フォーマットの乱立はコンテンツの流通を大きく妨げてしまい、市場の成長を阻害する要因になる。それをふせぐため中国では電子書籍フォーマットを含めた標準化の検討が始まっている。

## ■ 中国における標準化の動き

2010 年に電子書籍に関する技術の標準化を検討する組織が 2 つ成立している。

まず 4 月には全国信息技術標準化技術委員会の下に電子書標準工作組が成立した。全国信息技術標準化技術委員会は工業と情報を所管する行政部門である工業和情報化部と国家標準（GB 規格）を制定する国家標準管理委員会、この 2 つの政府機関の指導のもとで情報技術に関する規格の制定を担当している。その下で電子書籍に関する技術の標準化を検討する組織が電子書標準工作組である。組長は中国電子技術標準化研究所の王立建氏、副組長が漢王科技の董事長（社長）である劉迎建氏である。

検討内容はフォーマット、メタデータ、中国語（字体、文字記号、字形など）、ファイル形式（データストレージ形式、コンテンツ構造など）、著作権保護など多岐にわたる。最終的には国家標準（規格）の制定を目標としている。

電子書標準工作組の中では漢王科技がハードウェア、方正阿帕比がフォーマット、盛大文学がクロスプラットフォームでそれぞれ検討のリーダーシップをとるという報道がされている。これが事実なら中国における電子書籍の標準フォーマットとして方正の CEBX が採用される可能性がある [15]。しかし、複数の流通フォーマットに変換できる中間フォーマットを制定する可能性もあるという参加者の意見も報道されており、状況はまだ流動的のようだ。

7 月には新聞出版総署と出版関係者を中心に出版技術の標準化を検討する全国新聞出版標準化技術委員会によって電子書（内容）標準項目組が成立している。中国出版集団、上海世紀出版集団、江蘇鳳凰出版伝媒集団などの出版関係者のほかに漢王、方正阿帕比、盛大文学なども検討に加わっているらしい。

検討内容は用語、分類、レイアウトに関するフォーマット、コンテンツの品質基準、統計、コンテンツの

編集と校訂、出版プロセスの管理など出版社の業務に沿ったものになるようだ。電子書籍市場への企業の参入と退出の管理体制の構築など、政府がどう電子書籍市場を管理するかということも検討範囲に入っている。

全国信息技術標準化技術委員会と新聞出版総署の検討内容について明確に棲み分けがされているのかは報道される範囲ではわからない。電子書籍をIT技術、電化製品とみるか（全国信息技術標準化技術委員会）、出版とみるか（新聞出版総署）という問題が背景にあり、両者が電子書籍の検討で少し微妙な関係にあることはニュースを追っていくなかで感じるところではある。

なお、2009年4月に中国版式技術産業連盟（DIAA：Document Industry and Application Alliance）という団体が成立している[16]。政府機関も会員に加わっているようだが、電子出版に関わる民間企業を主とする民間団体という位置づけのようだ。方正のCEBXがこの団体の標準フォーマットとして採用されたようである[17]。このDIAAもCEBXが国家標準として認定されるように活動をしているらしい。

### ■ 変化するビジネスモデル

水平分業型の中国市場ではあるが、AmazonのKindleモデルのようにコンテンツと端末を紐づける垂直統合型のビジネスモデルを志向し、他のレイヤーとの結びつきを強化しようとする動きが以下のように出てきている。

- 方正がWeFoundという端末を発売した。
- 漢王が「漢王書城」という電子書籍を販売するサイトを立ち上げた。
- 盛大がBambookという端末を8月に発売した。
- 中国出版集団が自社ブランドで電子書籍端末を発売することを4月に発表した。
- 中国移動（CHINA MOBILE）が漢王科技、華為（Huawei）など4企業と提携して3G回線が使える端末を市場に投入した。
- 中国電信（CHINA TELECOM）が方正、浙江出版集団、中信出版社、盛大文学、新浪網、龍源期刊などのコンテンツプロバイダーと提携を発表した。

2010年に入ってコンテンツプロバイダーと端末メーカーが提携するという話が絶えず報道されており、プラットフォームにコンテンツ、端末というレイヤーを結びつける動きは今後も続くだろう。中国の水平分業型の産業構造がすぐに変わるとは思えないが、プラットフォームの構築で各企業がしのぎを削る状況が今年から来年にかけて見られるようになる。

## ❖ 台湾

### ■ EPUBに一本化する台湾

台湾は電子書籍フォーマットをEPUBにほぼ一本化していると言ってよい状況になっている。

2009年8月に「數位出版産業發展策略與行動計畫（電子出版産業発展策略及び行動計画）」[18]という5カ年計画（2009～2013）が行政院で承認された。5年間で21.34億台湾元（約60億円）を投じ、台湾の国際競争力の強化、電子書籍技術の標準化、電子書籍利用人口の増大などを目的とする計画である。翌月の2009年9月の數位出版産業推動座談會という会議で台湾はEPUBで一本化するという方針が決められたようだ。

それを受けて2009年12月に政府（經濟部）と財團法人資訊工業策進會の主導によって出版社、端末メーカー、通信業者など計200社以上が加盟する電子閲讀産業推動聯盟が成立した。台湾における電子書籍の標準化に関する検討は電子閲讀産業推動聯盟で行われることになる。

### ■ 電子閲讀産業推動聯盟における検討

電子閲讀産業推動聯盟に下に標準化の検討を行う「標準規範小組」が設置された。さらにその下に3つのワーキンググループが設けられて検討が進められている。

- 「中文EPUB Working Group（中国語EPUB Working Group）」
  検討内容：フォーマットの標準化
- 「閲讀器與應用平台Working Group（閲覧端末と応用プラットフォームWorking Group）」
  検討内容：著作権保護技術や電子書籍端末の標準化
- 「關鍵零組件Working Group（キーコンポーネントWorking Group）」

検討内容：電子ペーパー技術の標準化などのキーコンポーネント

### ■ 電子書籍フォーマット（EPUB）の検討状況

台湾の電子書籍フォーマットに関する検討は電子閱讀產業推動聯盟内の中文 EPUB Working Group が行っている。

最初の成果物として 2010 年 4 月に「EPUB OPF 電子出版品結構資料中文規範書 v1.0」[19] が公開された。これは EPUB の仕様書の 1 つである OPF（Open Packaging Format）[20] のガイドラインで、EPUB に格納されるメタデータの記述が出版社によってブレないように中国語で各項目の定義や記述方法を明確にしたものだ。

現行の EPUB の仕様では縦書きなど台湾の出版物の要件を十分に満たすことができない。そこで EPUB の仕様に台湾の要件を盛りこむため、IDPF（International Digital Publishing Forum）における EPUB 次期バージョンの仕様の検討に台湾は積極的に参加している。また、この原稿を書いている 8 月末時点の話になるが、中文 EPUB WG で現行の EPUB の仕様に含まれない独自仕様のようなものを作成している（した？）らしい。今後、これらの独自仕様が EPUB 次期バージョンの仕様に吸収されるかどうかは、EPUB の日本語対応にも影響を与える可能性があるので注視する必要があるだろう。

### ■ 台湾のおかれている状況と世界戦略

台湾が早期に EPUB に一本化し、官民で検討の体制を固めた理由は同じ中国語圏の中国の存在によるところが大きいだろう。台湾がおかれている状況をみるためにまずは台湾の出版状況を見てみたい。

2008 年の統計 [21] であるが、図書を発行する出版社は 10,002 社、図書の発行点数は 41,341 である。台湾では出版社を自由に設立できるので、実際に活動をしている出版社の数はこの数字よりかなり少なくなるが、台湾の人口が 2300 万人強であることを考えれば、台湾の出版業は十分盛んであるといえよう。しかし、台湾の 3 倍以上の出版点数を誇り、しかも物価が格段に安い同じ中国語圏の中国が台湾のすぐ隣に存在するのである。

繁体字圏、簡体字圏という違いがあるとはいえ、中国の出版物が台湾で自由に流通すれば、台湾の出版業界は価格と量で中国の出版社に圧倒されてしまう。台湾政府は台湾の出版社保護のために中国から台湾への簡体字図書の輸入に対してある程度の規制を課しているが [22]、電子書籍の台湾への流通を食い止めるためには、中国のサイトとのネットの接続を遮断するという強引な手段ぐらいしかないと思われるので、紙の書籍で行っているような輸入規制を電子書籍に課すことは困難であろう。

成長著しい中国市場で競争力をつけた中国企業が台湾に進出することに対し、台湾が強い警戒感を抱いていることは想像に難くない。漢王、方正などの中国企業は台湾市場への進出に積極的な姿勢をみせており、中国のコンテンツと端末が台湾で流通することは時間の問題になっている。

しかし、一方で電子書籍の普及は台湾企業にとって中国を含めた世界の華人市場に進出するチャンスでもある。台湾には華碩（ASUS）、明基（BenQ）、宏碁（Acer）のような国際競争力を持ったハードウェアメーカーがあり、元太科技（E Ink）のような電子書籍端末に不可欠な電子ペーパー技術を持つ企業があるという強みがある。年間 4 万点強という新刊図書の出版点数も決して少ない数字ではない。電子書籍に関する技術の標準化を台湾の中で迅速に進めるだけではなく、後述のように中国と標準化の協議を進める台湾の姿勢には世界の華人市場における台湾企業の進出を優位に運びたいという台湾の世界戦略の存在を感じる。

### ■ コンテンツプロバイダー

中国のようにコンテンツプロバイダーが数え切れないほど存在するという状況ではない。しかし、台湾でも電子書籍を販売するサイトが現れてきている。

【携帯キャリア】

台湾で携帯キャリア第 1 位の中華電信、第 3 位の遠傳が 2009 年に、第 2 位の台灣大哥大が 2010 年に電子書籍市場に参入している。

- ●中華電信　Hami 書城
  http://hamiweb.emome.net/categories/view/19/0
- ●台灣大哥大　行動書城（URL 不明）
- ●遠傳　遠傳 e 書城
  http://www.ebooktown.com.tw/

【出版社】

●遠流出版　eBook.com.tw
http://www.ebook.com.tw/

【端末メーカー】
●明基（BenQ）　eBookTaiwan
http://www.ebooktaiwan.com/ebooktaiwan/
日本の eBookJapan と技術・業務提携している [23]。

【書店】
●金石堂網路書店
http://www.kingstone.com.tw/

【その他】
2010 年 7 月には出版社だけではなく、個人からもコンテンツの投稿を受けて販売を行う Pubu 書城というプラットフォームがサービスを開始している。
●Pubu 書城
http://www.pubu.com.tw/

### ■ 電子書籍端末

台湾には華碩（ASUS）、明基（BenQ）、宏碁（Acer）など数多くのハードウェアメーカーが存在するが、台湾のメーカーから発売された電子書籍端末は 8 月末時点で明基電通（BenQ）の nReader ぐらいだった。この他、海外のメーカーから韓国 iriver 社の端末 iriver story がすでに台湾で販売されている。中国企業の漢王も 10 月に台湾で端末を発売すると発表している。台湾企業からも年内に華碩(Asus)と宏碁(Acer)が端末を投入すると発表しており、2010 年末には台湾でも多くの端末が流通するようになるだろう。

Apple の iPad は台湾でまだ発売されてはいないが、遠からず発売されると見られている。iPad に対抗する形で華碩（ASUS）が EeePad、微星科技（MSI）が Wind Pad というタブレット PC を発表している。電子書籍端末にはためらっていた他のハードウェアメーカーも PC の延長線にあるこのジャンルは華碩(ASUS)に続いて参入してくるかもしれない。

## ❖ 中国と台湾の標準化に関する協議

2009 年から中国と台湾の間で電子書籍に関する技術の標準化に関する協議が進められているようだ。中台の協議の状況について中国側の情報はほとんど見つけられないが、台湾側の報道や担当者の報告などで断片的な情報をつなぐ形になるが、ある程度の状況は知ることができる。

2009 年 7 月に中国の新疆ウイグル自治区ウルムチ市で行われた 2009 年海峡両岸中文信息技術標準論壇で中国と台湾が共通の中国語の電子書籍フォーマットを推進していくという「初歩共識（基本的な共通認識）」を持つことで合意したようだ [24]。ウルムチの会議でいうフォーマットがどのフォーマットを指すかどうかは不明だが、中国側の思惑としては中国で定めた標準フォーマットによる中国語圏のフォーマット統一というシナリオを考えているかもしれない。2009 年 12 月に中国と台湾が EPUB を中国語の電子書籍の標準フォーマットとして推進していくことで基本合意したというニュース [25] が PCWorld で報じられたので EPUB の可能性もある（ただし、電子書籍端末が EPUB に対応したというニュースを除けば、EPUB に関するニュースが中国側でほとんどない）。

電子閱讀産業推動聯盟　中文 EPUB Working Group の責任者である陳美莉氏による 2010 年 3 月の発表資料 [26] によると、期間を 5 月から 7 月と 10 月から 12 月の 2 回に分けて協議を行うことになっているようだ。3 月に中国と台湾が共催した 2010 海峡両岸標準合作研討會中文訊息組工作會議という会議で「術語比對（用語の比較対象）」、「元數據（metadata）」、「文件格式（ドキュメントフォーマット）」、「數位内容管理和交換（デジタルコンテンツの管理と交換）」、「電子閱讀器規範（電子書籍閲覧端末の仕様）」などの標準化について研究を進めていくという双方の共通の認識を持つことになったとされている [27] ので、5 月からの協議も上に挙げられた項目について検討が進めているはずだ。

## ❖ さいごに

台湾の動向で少し触れたが、最後に EPUB 次期バージョンにおける国際化の検討についても少し触れておきたい。EPUB は IDPF（International Digital Publishing Forum）が促進する電子書籍フォーマットで、国際的に電子書籍の標準フォーマットになりつつある。しかし、欧米圏で生まれたフォーマットであることから、例えば、日本語については縦書きができな

い、禁則処理やルビに対応していないなど、アジア言語の出版物の要件を満たしていないことが大きな問題になっていた。

EPUB 次期バージョンでアジア言語を含む国際化の対応が予定されている。IDPF の下に次期 EPUB の仕様検討のためのワーキンググループ（EPUB2.1/3.0 Working Group）が設置され、具体的な検討が始まっている。国際化の対応（アジア言語の対応など）については、上のワーキンググループの下に The EPub Enhanced Global Language Support sub-group（以下、EGLS）が 2010 年 6 月に設置され、この中で検討が進められている。台湾が EPUB の仕様の検討に積極的に参加しているのは前述の通りであるが、このサブグループにもメンバーをだして検討に加わっている。この EGLS の初めての顔合わせ会合が 8 月 3 日と 8 月 4 日に札幌で行われ、日本、韓国、台湾から電子書籍の関係者が集まった。次の顔合わせ会合は 10 月 5 日から 2 日間、台北で行われる予定になっている。この会合に中国の関係者が参加するか否かは注目すべきところだ。検討内容、会議の議事録、発表資料、インプット、各メンバーが集めた要件等はすべて公開されているので関心をお持ちの方はぜひご覧いただきたい [28]。

中国側の企業はコンテンツプロバイダー、端末のメーカーとも台湾市場への参入に積極的であり、また、台湾側も世界の華人市場への進出を視野にいれて官民が団結している。物量で勝る中国が台湾を飲み込むことになるのか、それとも標準化の検討で先んじて台湾が EPUB という国際的なフォーマットの器に乗る形で中国を含む世界の華人市場で乗り出すことになるのか。華人市場における電子書籍の動向はこの数年で決まると思われるが、その動向は同じ漢字圏の日本にも大きな影響を与えることになるはずだ。

この原稿を脱稿する直前の 8 月末に「中国が EPUB を採用する」という情報を入ってきた。電子書籍端末が EPUB に対応したというニュースを除けば、EPUB に関する報道は中国でこの半年ほとんどなかった。さらに方正の CEBX が標準フォーマットになる可能性を報じるニュースもあったので、中国が EPUB を採用するという情報には非常に驚いた。情報源は前述の EGLS で中心メンバーとして活動している村田真氏の Twiiter 上の発言である [29]。EPUB 次期バージョンの検討で最前線に立ち、日本、韓国、台湾（そして、おそらく中国）の間で調整を行っている村田氏の発言であるだけに非常に重要な発言だ。しかし、「採用」といっても中国の標準フォーマット及び国家規格としての「採用」なのか、それとは異なる次元での「採用」なのかで全く状況が異なってくる。この件について村田氏に尋ねたが、得ている情報が断片的なものでもあり、詳細はまだ公開できないとことであった。この『漢字文化情報理研究』11 号が刊行されるまで明らかになっているであろうか。非常に気にになるところである。

2010 年に入ってから中国と台湾の電子書籍に関する動きは激しくなっているが、特にこの数ヶ月はあたらしい動きが次々を生まれ、目が離せない状況になっている。今後の動向次第でこの文章が読者の目に触れる頃にはここで書いた状況とは異なったものになっている可能性も否定できない。脱稿した 8 月末以降の動向については筆者のブログ [30] で能力と時間が許す範囲で少し追っていきたいと考えているので、関心をお持ちの方はこちらもご覧いただきたい。

## 注

[1] 『中国電子図書発展趨勢報告』2009 年版：http://www.du8.com/html/zhuanti/ph2009/

[2] http://research.cnnic.cn/html/1263531336d1752.html

[3] 『中国電子図書発展趨勢報告』2009 年版「表 4.2　按类别划分的电子图书阅读终端」より

[4] 『中国電子図書発展趨勢報告』2009 年版「表 4.1　主要电子图书业务销售收入」より

[5] 2009 年新闻出版产业分析报告：http://news.xinhuanet.com/politics/2010-07/27/c_12376858.htm

[6] http://www.gapp.gov.cn/cms/html/21/464/200907/465083.html

[7] 地下出版物については、『中国・台湾の出版事情』（島崎英威著　出版メディアパル , 2007）が参考になる。

[8] http://www.apabi.cn/apabi.shtml

[9] http://www.fanshu.com/

[10] http://www.csmonitor.com/World/Asia-Pacific/2009/1231/China-book-market-As-e-books-grow-they-get-pirated-too

[11] http://www.qidian.com/ploy/20100310/

[12] http://www.spc.jst.go.jp/news/100103/topic_4_01.ht

ml

[13] China's e-book sales may exceed 3 mln units in 2010 http://www.marketreportchina.com/market/article/content/3377/201007/225725.html

[14] http://www.pcpop.com/doc/0/557/557804.shtml

[15] CEBX については以下が参考になる。
「中国の電子書籍フォーマット「CEBX」を方正に聞く ── builder by ZDNet Japan」: http://builder.japan.zdnet.com/sp/epub2010/story/0,3800103623,20417273,00.htm
方正集団の日本法人、方正株式会社も CEBX が中国の標準フォーマットになる可能性を示唆している。
「中国発の電子書籍フォーマット「JEBX」は黒船になるか？ ── builder by ZDNet Japan」: http://builder.japan.zdnet.com/sp/epub2010/story/0,3800103623,20417621,00.htm

[16] http://www.china.com.cn/tech/txt/2009-04/09/content_17585524.htm

[17] http://www.apabi.cn/html/201004/1004151155233781.shtml

[18] 杜紫軍「數位出版產業 發展策略與推動現況」『研考雙月刊』第 34 卷第 1 期　2010.2：http://bimn.rdec.gov.tw/lib/lib02/bimn/275/275-10.pdf

[19] http://www.oss.org.tw/getfile.php?id=01

[20] http://www.idpf.org/2007/opf/OPF_2.0_final_spec.html

[21] http://info.gio.gov.tw/Yearbook/98/c8.html

[22] 大陸地區出版品電影片錄影節目廣播電視節目進入臺灣地區或在臺灣地區發行銷售製作播映展覽觀摩許可辦法第 13 条

[23] http://www.ebookjapan.jp/ebj/info/news/2010/01/ebooktaiwan.asp

[24] 經濟部工業局「數位出版產業推動現況」p11: http://proj3.moeaidb.gov.tw/nmipo/content/getfile.aspx?sn=CEE83F166DAA438FA58F0A2783D56C76

[25] http://www.pcworld.com/article/183506/china_and_taiwan_agree_to_promote_epub_ebook_standard.html

[26] http://www.slideshare.net/taiwandigital/ss-3582426?from=ss_embed

[27] http://mag.chinatimes.com/mag-cnt.aspx?artid=4671

[28] the EPUB 2010-2011（2.1/3.0）revision.: http://code.google.com/p/epub-revision/

[29] MURATA Makoto（muratamakoto）on Twitter : http://twitter.com/muratamakoto
まだ断片的な情報ということで、該当する発言への直接のリンクはあえて控えさせていただくが、村田氏は Twitter 上で検討状況をリアルタイムに報告しているので、関心をお持ちの方はぜひ村田氏の Twitter を定期的にご覧いただきたい。。

[30] 筆者のブログ : http://kzakza.wordpress.com/

# 図書館とOPAC

小島　浩之

## ❖ はじめに

本稿では、書籍・論文等の検索やデジタルアーカイブに関する日本の状況をレビューする。個別図書館における膨大な量の電子化事業に目配りすることは事実上不可能なため、Google、国立国会図書館、国立情報学研究所の話題に集約してレビューを試みる。

## ❖ Google ブックス

昨年、著作権問題で世間を騒がせた Google ブックスのインターフェイスは、まだ改良途中であり β 版のままである。しかし、今年に入り Google 日本語版トップページのメニュー欄に、書籍としてリンクがはられるようになった。β 版とはいえ正式のサービスとしての位置を着々と固めつつある。

日本では慶應義塾図書館蔵書のうち、著作権切れのもの 12 万冊のデジタル化が予定されており、作業が徐々に進みつつあるようだ。ただし、現在の進行状況については公表されていない。

そこで、Google ブックスで全文が閲覧可能な日本語書籍のおよその数を知るため以下のような試験検索をした。検索オプションで、日本語で全文表示

の書籍を指定して、キーワードを「い」「ん」「か」「し」として検索したところ、それぞれ、91,500 件、111,000 件、86,000 件、130,000 件ヒットした。この 4 つのキーワードを指定したのは、これらが新聞で最もよく使われる平仮名の上位 4 文字だからである [1]。つまり平均して約 10 万件の日本語の書籍が Google ブックスで全文検索・閲覧可能となっていることが解る。もちろんこれはあくまで参考値である。

また全文閲覧可能な漢籍（古籍）もかなりあり、画像に写る蔵書印からハーバード大学燕京図書館と慶應義塾図書館のものが多いことが解る。

Google ブックスの閲覧用画像は PNG 型式だが、モノクロとカラーが混在してるので、撮影仕様が途中で変更されている模様。閲覧用のインターフェイスは、頁単位の画像を下方へスクロールしてゆく仕様のため幾分読みづらい。この点を解決してくれる閲覧ビューア が GooReader（ 配 布 元 URL：**http://gooreader.com/**）である。GooReader は Windows 上で Google ブックスの書籍を見開き単位で表示し、電子ブックの感覚で Google ブックスを利用できる。

データを一括でダウンロードする場合、Google ブックスの仕様では PDF 型式となる。PNG 画像のままで一括ダウンロードしたい場合や、特定頁の PNG 画像へ直リンクする URL を知りたい場合は、Google Book Downloader（配布元 URL：**http://userscripts.org/scripts/show/37933**）が便利である。ただし、こちらは Firefox 専用のアドオンとなっている。

この他、Google ブックスのインターフェイスには、クリップ機能がある。これを使用すると画像の任意部分を切り取ったり、テキスト化したりできる。

このように Google ブックスはデータ量の増大と機能の改善や、関連フリーソフトの公開が進んではいるが、前号の拙稿 [2] で指摘したような種々の問題点を内包したままで、まだまだ発展途上だと言えよう。

縦書右開の書籍については、横書左開のものと同様の画像排列になっているため、画像が裏表紙から始まり冒頭で終了する珍妙なものが多い。元データがおかしければ、当然 GooReader で再構成された画像は乱丁状態になり、ものによっては読めたものではない。また PDF のダウンロードボタンが表示されない不具合もある。Google はこれらの点について改善を進めていると聞くが、まだまだ追いついていない。

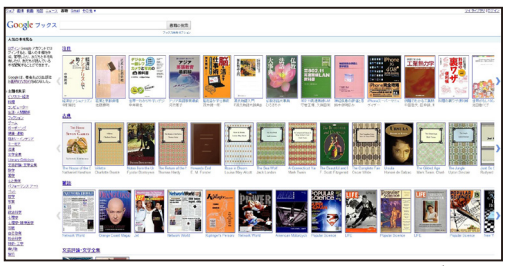


図 1：Google ブックス

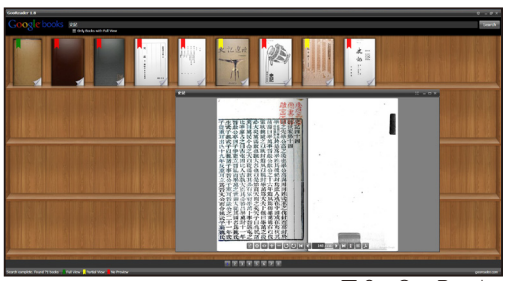


図 2:　GooReader

## ❖　国立国会図書館（NDL）

「私」のデジタル化の Google に対し、「公」のデジタル化を進める NDL の状況はどうであろうか。

●近代デジタルライブラリー

**http://kindai.ndl.go.jp/index.html**

近代デジタルライブラリーは、これまで著作権処理が完了もしくは著作権切れの明治・大正期刊行図書について全文をインターネット配信してきた。当然、権利処理ができていない図書については、書誌情報も含めて一切アクセスできなかった。

ところが、今年度よりこれら明治・大正期の著作権未処理の図書や昭和前期刊行図書は、NDL 館内でデジタル画像を提供するようになった。デジタル化され

図 3:　国立国会図書館サーチ　開発版

図 4：CiNii 検索結果詳細表示画面

図 5：機関リポジトリへのリンクアイコン（図 4 の○部分）

た書籍総数は約 39 万冊、このうち明治期 129,000 冊、大正期 41,000 冊余りがインターネット配信されている。館外においても 39 万冊分の書誌データは確認できるので、明治から昭和前期の国内刊行資料に関する大規模書誌データベースとしても機能するようになったと言えよう。

●国立国会図書館サーチ（NDL Search）開発版

**http://iss.ndl.go.jp/**

このほか NDL では、国立国会図書館サーチ（NDL Search）開発版が公開された。開発版とあるのは、平成 24 年の本格稼働までの試行ということらしい。この検索システムには、連想検索とメタデータハーベスト用 API を実装しているという。

既に NDL は、デジタルアーカイブのポータルサイトとして、POLTA を公開している。NDL Search では、POLTA の内容（ここには近代デジタルライブラリーも含む）に加え、NDL、都道府県立図書館、政令指定都市の市立図書館の蔵書の検索が可能となる。恐らく次世代 PORTA の位置づけなのだろう。

●PORTA

**http://porta.ndl.go.jp/**

なお、NDL Search は携帯版も公開されている。

●国立国会図書館サーチモバイル

**http://iss.ndl.go.jp/mobile/**

## ◈ 学術雑誌のデジタルアーカイブ

一方で、国内における学術雑誌のデジタル化はどのような状況にあるのだろうか。

学術雑誌のデジタル化にいち早く取り組んできたのは国立情報学研究所（NII）であった。日本における雑誌のデジタル化は諸外国に比べ遅れていると言われてきたが、既に CiNii には約 320 万件の論文が電子化されている。また CiNii じたいに本文収録の無い場合でも、各研究機関の機関リポジトリや J-STAGE（科学技術振興機構（JST）が主導する電子ジャーナル支援システムで、主に科学技術系の雑誌論文を電子化して公開している）に本文があれば、リンクにより誘導してくれる。

●NII 論文ナビゲーター CiNii

**http://ci.nii.ac.jp/**

●J-STAGE

**http://www.jstage.jst.go.jp/browse/-char/ja**

この二者に加えて、NDL も大規模な雑誌デジタル化作業に着手している。このプロジェクトにでは CiNii や J-STAGE との重複を避けてデジタル化が行われると聞いているので、近いうちに日本の学術論文の電子化の範囲が一気に広がることが期待される。

## ◈ おわりに

ここまでで述べた、統合検索や他者へのリンクの活用、電子化事業の棲み分けに見られるように、今後は既存のデジタル資源を有効に利用しつつ、分担・協力した電子化により公開の範囲の拡大を目指すことが一般的となるだろう。ただし、これが Google などの「私」の機関をも巻き込んでゆくかどうかは未知数である。

### 注

[1] **http://www.excite.co.jp/News/bit/00091179941794.html**

[2] 拙稿「図書館・OPAC・漢籍」『漢字文献情報処理研究』10 号 , 2009 年

# ベトナムの漢字・字喃文献

矢野　正隆

## ❖ はじめに

編集の依頼は、ベトナム漢字文献公開の現状をレポートし主要データベースを紹介する、というものであった。筆者としても利便性の高いガイドの必要性は日々痛感しているところであるが、この分野については現在ようやく整備が進み始めたところで、進行中・未公開のプロジェクトが多く、纏まった形での紹介はまだ難しそうである。また、本誌ではこれまでベトナムの漢字文献について全く触れられることがなかったとのことなので、本稿は過渡的あるいは些細な論点は極力素通りし、やや基礎的な話となることを予めご承知おき願いたい。

## ❖ 1. ベトナム古典籍とは

こんにち、東アジア漢字文化圏が話題になる場合、詳細はともかく、何らかの形でベトナムに言及するのが普通であろう。ベトナムが漢字遣いの伝統を持っていたという事実は、本誌読者のように漢字文献に親しく接している者にとっては、ごく常識的な知識であると思われる（[グエン 1987] など）。

ベトナムでは、自らの古典籍のことを「ハンノム Hán Nôm」と呼ぶ。これは「漢喃」のベトナム語読みで、「漢」はもちろん漢字のこと、「喃」はチューノムを指す。チューノムとはベトナム語を表記するために、漢字の構成要素を用いて作られた疑似漢字の一種で、纏まった形で現れるのは 13 世紀頃からとされる。その構造上、読解には漢字の知識が不可欠であったため、20 世紀初頭以降、フランス植民地政府による科挙廃止等の政策や、ベトナムの指導的文化人による「クオックグー」（アルファベットに補助符号を付した現代ベトナム語表記）使用の推進により、漢字の使用頻度が下がると、それに伴って廃れていった。こうした事情もあり、こんにち漢字文献とチューノム文献とは、個々別というよりは、今では用いられることのない古典語として、セットで取り扱われるのが普通である。

ベトナムで漢字が使われていたことを知る者も、この「ハンノム」という語は聞き慣れないかも知れない。これを「安南本」と言えばどうだろうか。「安南」は、もともとはベトナム地域の中国側からの呼称で、中国に限らず、日本やフランス等の研究ではごく一般に用いられていた。3.1 節で紹介する、日本で公表されたハンノム文献書目は皆、その対象を「安南本」と記している（[松本 1934] ほか）。

もちろん「安南本」と言い代えてみたところで、漢籍に触れる機会の多い者にとっても、やはり心理的に遠く感じる方が一般的であろう。では実際のところ、日本国内でこのハンノム文献はどれほど所蔵されているのだろうか。試みに全国漢籍データベース（**http://www.kanji.zinbun.kyoto-u.ac.jp/kanseki**　URL は 2010 年 8 月現在。以下同。）をキーワード「安南」で検索してみると、500 件以上ヒットする。このうちの 300 件あまりは、中国人や日本人が著した「安南」を内容とするタイトルで、ベトナム人による編著作に限ればおよそ 200 件となる（所蔵機関は約 30 館）。さらに、この 200 件の大半は、中国で刊行された叢書に含まれるもの、あるいは日本で翻刻されたものであり、明らかにベトナムの地で刊行・抄写されたもの、及び、その影印本・マイクロフィルム版に限ると、所蔵館は 5 館ほど、点数も約 30 件に絞られる。ただし、全国漢籍データベースには、纏まったハンノム文献のコレクションは採録されていない（代表的なものとして、東洋文庫の越南本コレクション（約 300 部）や慶應義塾大学の松本文庫（62 部）がある）。

ここは詳細なデータを提供する場ではないので、大雑把な推計になるが、中国版でも日本版でも構わないからハンノムのテキストを読みたいというのであれば、国内 30 以上の機関で総計 500 部以上、モノとしてハンノム本を見てみたい場合には、およそ 10 機関で 400 部ほど、これがひとまずアクセス可能な部数の見積もりになる。この数をどう評価するか。案外身近に思われるか、あるいは、やはり縁遠いものと感じられるだろうか。主観的な感想は多々あるであろうが、こうした事実をより客観的に位置づけるためには、もう少し視野を広げて、ハンノム文献の現状全体を確認しておく必要がある。

図 1：越南漢喃文獻目録資料庫系統・書目詳細資料

## ❖ 2. ハンノム文献公開の現状

ここではハンノム文献がいったいどういった内容を持つものか、またその資料自体の来歴と現在の所在について、近年公開されたデータベースや刊行物により、概要を紹介する。

### ■ 2.1. 「越南漢喃文獻目錄資料庫系統」

こんにちハンノム文献を探索する者にとって、台湾の中央研究院で作成された「越南漢喃文獻目錄資料庫系統」（**http://www.litphil.sinica.edu.tw/hannan/** 以下「系統」と略す）は、まず欠かすことのできない目録データベースである。

ここに収録されているのは、ベトナムのハンノム研究院とフランスの５つの機関（フランス極東学院 L'École Française d'Extrême-Orient、フランス国家図書館 Bibliothèque Nationale de France、アジア協会 Société Asiatique、ギメ東洋美術館 Musée Guimet、東方言語文化学院 Institut National des Langues et Civilisation Orientales）が所蔵しているハンノム文献、およそ 5000 タイトルである。台湾の機関が含まれないのは、このデータベースが、ハンノム研究院とフランス極東学院（以下 EFEO と略す）から刊行された書目『ベトナムハンノム遺産』（本編３巻補遺２巻、1993–2002、以下『遺産』と略す）を漢訳するというプロジェクトから生まれたことによる。

この書目『遺産』は、越仏合同の刊行であり、本文は現代ベトナム語とフランス語で記されているため、中国・台湾や日本の研究者にとって、ややアクセスしにくいものであった。この漢訳プロジェクトでは、より精確なデータを提示するために、ベトナムやフランスの収蔵機関とも連携して現物を逐一チェックしている。排列は、原書では現代ベトナム語のアルファベット順であったのを四部分類順に組み替えており、漢籍目録の伝統を活かして利用の便を図っている。漢訳書の書名は『越南漢喃文獻目録提要』（正編２巻補遺２巻、2002–2004）で、データベース「系統」はこの全体を含む。

詳細データは書誌情報（書名、作者、版本、頁数及版式、提要、版本分類、四部分類）に所蔵情報（館蔵編号）が付される形をとる（図 1）。検索は書名・作者・提要・館蔵編号等の項目から繁体字でおこなうことができるが、特に書名については、音標符号抜きの現代ベトナム語（クオックグー）で検索できるのが便利である。

この所蔵情報には、各収蔵機関独自の分類番号が記されており、これらを辿ることによって、各本の出自を知ることができる。以下、分類ごとに説明する。

#### ●2.1.1. EFEO

「系統」に採録されているハンノム本の淵源を辿ると、ハノイにあった EFEO の収蔵本に遡る。それまでの王朝時代にも宮廷の書庫は存在したが、こんにちに直接続く組織的収集事業は、20 世紀初頭のフランス植民地支配下に端を発したものであり、これが以降のハンノム文献収集史における主要な流れのひとつとなる（ハンノム文献収集史については［八尾 2004］）。

EFEO の収集対象は、一般書籍（版本、写本）だけでなく、地方村落の文書類を含んでいた。まず一般の版本・写本は漢文 (A)、チューノム文 (AB)、中国書籍のベトナム版 (AC) の３つに区分される。これが『遺産』の「本編」にあたる。以下、各記号の内容とその部数を示すが、数値はおおよその規模を掴むために、筆者が「系統」より抽出したデータを集計したもので、異本を含めた総部数は 10144 となる（分類記号のナンバリングがタイトル単位か物理単位かは必ずしも統一されていないが、ここは比較しやすくするためナンバリング単位で集計している）。

A： 漢文。「系統」には 2926 部を含む（ナンバリングは A.1 ～ 3231 であるが欠番を除く総数を記す。以下同）。

AB： チューノム文。「系統」には 569 部を含む。

AC： 中国書籍のベトナム版。「系統」には 564 部を含む。

AC は翻訳翻案ではなく、テキストは同一のベトナム版本である。線引きが難しいところであるが、これもハンノム文献に準ずるものとして位置づけるのが慣行になっている。内容としては経書や仏典が多い。

一方、『遺産』の「補編」に採録されている文書類は以下の 6 種類がある。

AD：神勅。ベトナム歴代王朝が各村落で奉られている諸神を認定する勅封。411 部。
AE：神蹟。乂安以北各村落の神の事蹟。568 部。
AF：俗例。「例」「券」「約」といった村落の掟書や契約文書類。647 部。
AG：地簿。土地台帳。526 部。
AH：古紙。「古字紙」の略称。「古字」とは漢字およびチューノムのこと。雑多な内容を含む。21 部。
AJ： 社誌。EFEO が各村落で行った調査の調査票。107 部。

●2.1.2. EFEO 撤退以後

1954 年のジュネーブ協定締結を承けて、ベトナムは南北に分断され、フランスはベトナムから全面的に撤退することになる。EFEO も 1957 年にパリに移転するが、所蔵していたハンノム文献については、その重要なものをマイクロフィルムとして本国に持ち帰り、現物はハノイに残された。これらは社会科学委員会の管理下に置かれ、収集活動も継続されることとなる。1970 年には社会科学委員会の直下にハンノム班が置かれ、これが、南北統一（1975 年）後の 1979 年にハンノム研究院となり現在に至っている［Phan Văn Các 1991］。この時期の収蔵本には EFEO 時代とは別の分類番号が附されている。『遺産』-「系統」に採録されているのは、1987 年までに収蔵されたものである［Trần Nghĩa 1993］。

VH：漢文。EFEO の分類では A に相当するもの。大きさで VHt(大)・VHv(中)・VHb(小)に区分している。併せて 2775 部。
VN：チューノム文。EFEO の分類では AB に相当する。大きさによる区分は VH と同様である。378 部。
HV：中国書籍のベトナム版。EFEO の分類では AC に相当。16 部。

●2.1.3. フランス国内蔵書

EFEO にはハノイから持ち帰ったマイクロフィルムの他にもわずかながら現物のコレクションがあり、「系統」で 64 部を確認することができる。（ここでは［Truong Đình-Hoè 1986］で紹介されている 66 部のうち 2 部が採録されていないが、これは合冊の可能性、あるいは紛失等現物の管理上の問題であることも想定され、データの不備とは限らない。こうしたズレは全ての数値について言えることを付記しておく。）

EFEO 以外では以下の 4 機関の所蔵データが採録されている。

・国家図書館（「系統」の略号は BN、以下同）

ペリオ Paul Pelliot、マドロール Claude Madrolle の旧蔵本を中心とするコレクション。「系統」では、漢字文献（分類記号 A）95 部、チューノム文献（同 B）113 部を確認することができる。

・アジア協会（SA）

マスペロ Henri Maspero（分類記号 HM）、ドミエビル Paul Demiéville（同 PD）、およびランド Antony Landes（同 Ms.b）の旧蔵本を中心とするコレクション。「系統」では、HM126 部、PD87 部、Ms.b が 15 部確認できる。

・ギメ東洋美術館（MG）

「系統」で確認できたところでは、およそ 50 部。

・東方言語文化学院（LO）

「系統」では 70 部ほど確認できる。

「系統」が含むカテゴリーは以上である。ベトナムおよびフランスの主要機関の蔵書を包括するデータベースであることが分かると思うが、同時に、その来歴を辿ることにより、ハンノム文献収集の歴史、その広がりがイメージできるであろう。

## ■ 2.2. その他

EFEO- ハンノム研究院関係の資料では、「系統」以外にも注目すべきデータベース、資料集の公開が進んでいるので二三紹介しておく。

図２：漢喃研究院・漢喃遺産

●ハンノム研究院

**http://www.hannom.org.vn/default.asp?CatID=1**

上に紹介した書目『遺産』がデータベース化されており、書名、作者、記号等の項目から検索やブラウジングをすることができる（**http://www.hannom.org.vn/DEFAULT.ASP?CATID=131**）。このデータ自体は「系統」のもとになったものである。

漢字による検索も可能だが、現代ベトナム語仕様であるためか、漢字使用者への配慮にやや欠ける（例えばキーワード「大越」ではヒットしないが、「大 越」という具合に文字間にスペースを挿入するとヒットする等）。また通常の多言語 OPAC のような音標符号の有無を無視した検索はできない。（ちなみに本稿で紹介するベトナムのサイトはすべて UTF-8 対応である。）

このように「系統」に比べると使い勝手の面でやや劣るが、漢訳段階での編集はデータの追補や誤植修正だけではなく、データの省略も含まれているため、こちらも参照できるようにしておきたい。

このほかに、機関の学術誌『ハンノム雑誌 Tạp chí Hán Nôm』が全文公開されており、ここでハンノム書誌学の基礎文献を閲覧することができる（**http://www.hannom.org.vn/default.asp?CatID=15**）。

図３：漢喃古籍文獻典藏數位化計畫・集部

●越南漢喃銘文拓片總集 Tổng tập thác bản văn khắc Hán Nôm

ハンノム研究院の収蔵資料のうち、「系統」に含まれないものでは、碑文の拓本コレクションが重要である。これも EFEO 時代に収集されたものをベースとしており、現在の収蔵数は 5 万点を超えるが、これらすべてを写真版として出版するという計画が 2005 年から進行している。全体で 40 冊になる予定で、2010 年 8 月時点では 21 冊目まで刊行されている。将来的にはベトナムの全拓本データベースを構築する予定であるという［Trịnh Khắc Mạnh, etc., ed. 2005］。

ハンノム文献はもちろん EFEO 関係に限らない。近年ではこれ以外にも多方面で様々なタイプの資料が発掘・公開されている。

●漢喃古籍文獻典藏數位化計畫

**http://nom.nlv.gov.vn/nlvnpf/vindex.php**

ベトナム国家図書館 Thư viện quốc gia Việt Nam はハンノム文献をおよそ 4000 部所蔵しているが、これを、チューノム遺産保存会 Vietnamese Nôm Preservation Foundation と共同で画像データベース化したもの。計画は 2006 年に始まり、「2 年のデジタル化作業で 1158 冊、78536 ファイルの画像が処理」（Nhân dân 06-11-2009）された。排列は経・史・子・集に区分されており、2010 年 8 月段階では、経 156 部、史 375 部、子 327 部、集 400 部の総計 1258 部が公開されている。全文検索等の機能は含まれないが、これによって、ベトナムに出向かずとも大量のハンノム文献をモノとして見ることができることになり、研究基盤の整備という意味において、誠に画期的な事業であると言える。画像形式は JPEG。検索は繁体字で可能である。

このほかに目録や資料研究が公表されているもの

としては、社会科学情報院の神勅［宇野 1999］、公文書館蔵の硃本（王朝時代の公文書類）［Vũ Thanh Hằng, et al. 2003］のほか、ハンノム研究院ほか複数の機関にある地簿のコレクション［Phan Huy Lê, et al., ed. 1995］等広範囲に渡り（概要は［八尾、岡田 2003］）、丁寧に紹介するためには別の紙面を要するであろう。本稿の残された紙幅では、こうした全般的な状況を踏まえた上で、日本とハンノム文献との関わりについていくつかの事実を述べる。

## ❖ 3. ハンノム文献と日本

「系統」の全データを構成する書目『遺産』の編纂が始まったのは 1983 年のことである。この事業は、言わば、全ハンノム文献の所在情報を集成しようとするものであったが、その前提として、この時点までのハンノム文献情報について、総括がなされている。具体的には、それまでに公表されたハンノム文献の書目と、収蔵機関がリストアップされているのであるが、これを見ると、日本の占める部分は決して小さくないことが分かる［Trần Nghĩa 1993：20–29］。

### ■ 3.1. 書目

実は、「系統」に採録されているハンノム文献は、その多くが日本では既に紹介されていた。

EFEO の蔵書のうち一般の刊本・写本（分類記号 A、AB、AC）については、［松本 1934］［山本 1938］および［山本 1953b］において、当時の原簿や閲覧室備付けの目録に拠ってリスト化されている。また、仏教関係については［陳 1943］が、A、AB、AC の中から 427 部を抽出しリストとしている。試みに「系統」やオンライン版『遺産』から同じ記号を持つものを抽出すると、少なからず欠番があることが分かるが（たとえば A は 1 から 3231 までナンバリングされているが、このうち「系統」で確認できるのは 2926 部）、これらのリストはこうした不明本を探索する大きな手がかりとなる。EFEO はハノイ時代にはハンノム文献の目録を公刊しておらず、これらの書目は過去の収集状況を知る上で非常に貴重な資料となっている。

フランス国家図書館については［山本 1953a］が 1951 年までに収蔵された 173 部（A1 ～ 65、B1 ～ 108）を、［藤原 1974］がそれ以降の収集分のリストを公表している（A66 ～ 105、B109 ～ 134）。［Berthier 2000］には収蔵数について A は 107 部、B は 134 部とあり、藤原が紹介して以降わずかに 2 部増えたのみであることがわかる。つまり、このふたつの書目で現在の蔵書をほぼすべてカバーできるということになる。また、「系統」と比較すると、このデータベースにはチューノム文献のうちベトナム国外で刊行されたものが省かれていること（B115 ～ 120 など）、また明らかにハンノム本であるが採録されていないタイトルも少なからずあることが分かる。

アジア協会に収蔵されているハンノム文献のうち、マスペロ（HM）およびランド（Ms.b）の旧蔵本については、［山本 1954］が全リストを公表している。このリストのうち山本が「安南本ではない」と注記しているタイトルについては、当然「系統」には採録されていないが、明らかなハンノム本で検索にヒットしないものもいくつかある。ランド旧蔵本に欠が多いのがやや気になる（「系統」採録は 23 部中 15 部）。

EFEO がハノイから撤退した後、その蔵書を引き継いだ社会科学書院は収集活動も継続することになるが、その新たに収蔵されたハンノム文献（分類記号 VH,VN）については、［川本 1971］が報告している。これは、同書院が編集していた書目『Thư mục Hán nôm』から、EFEO の記号（A,AB,AC）が附されている文献を除いた 581 部をリスト化したものである。ただし、排列は現代ベトナム語表記（クオックグー）のアルファベット順であり、また VH、VN といった分類記号も記されていないため、「系統」採録分との比較はやや困難である。しかし、EFEO 撤退後のベトナムにおけるハンノム文献の所在や、収集活動の状況を伝える数少ない記録として貴重である。

以上のように現在「系統」に含まれる 1 万部を超

図 4：東洋文庫越南本漢籍検索結果

Database List → 越南本漢籍 → 検索結果

東洋文庫　越南本漢籍　結果

A search found 2 entries.

検索にかかった時間：［2.446 秒］

＊叢書の一部のみ閲覧をご希望される際には、「詳細」ボタンをクリックしていただき、閲覧カウンター備え付けの「資料請求票」の「書名」欄に、「詳細画面」の叢書名・収録冊次・子目の書名をご記入下さい。

| No. | 詳細 | CallMark | 叢書名 / 収録冊次 / 書名・巻数 | 編者者 | 出版事項 |
|---|---|---|---|---|---|
| 1 | 詳細 | X-3-4 | 芸臺類語九巻 | 黎貴惇著 | 鈔本 |
| 2 | 詳細 | X-3-5 | 芸臺類語殘六巻　存巻第三至第八 | 黎貴惇著 | 鈔本 |

＊叢書の一部のみ閲覧をご希望される際には、「詳細」ボタンをクリックしていただき、閲覧カウンター備え付けの「資料請求票」の「書名」欄に、「詳細画面」の叢書名・収録冊次・子目の書名をご記入下さい。

Database List → 越南本漢籍 → 検索結果

［東洋文庫利用の仕方］

えるハンノム文献のうち、1970 年代初頭までに、日本では、その半数近くの存在が知らされていたことになる。

### ■ 3.2. 収蔵機関

『遺産』に掲載されている世界のハンノム文献収蔵機関リストのトップに掲げられているのが日本の東洋文庫である［Trần Nghĩa 1993：20-21］。このコレクションは、全国漢籍データベースには含まれていないが、ハンノム文献の纏まった資料群としては国内で最大の規模を誇るものであり、その内容は冊子の書目［東洋文庫古代史研究委員会(編)1999］、およびデータベースとして公開されている。(**http://61.197.194.11/open/KansekiEtunanQueryInput.html**)。

このコレクションの基礎となったのは、1930 年代にハノイ駐在総領事であった永田安吉による寄贈本 92 部であり（［岩井 1935］）、戦後 EFEO 蔵書等のマイクロ複製により拡充され（102 部）、近年では山本達郎の旧蔵本が収められている（山本旧蔵本については［Chu 2005］［Yano 2005］で 60 部近くが紹介されているが、その後に追加の寄贈があり総数は 100 部を超える）。因みに上記の書目とデータベースに山本旧蔵本は含まれていない。

東洋文庫に次ぐ規模のコレクションとしては、慶應義塾大学の松本文庫（62 部）がある。これは、山本達郎と並び 1930 年代から収集を行っていた松本信廣の寄贈によるもので、内容については詳細な目録が公表されている［和田 1992-1993］。

日本で公開されているハンノム文献としては、このふたつが主要なコレクションであり、おそらく国内全体の大半を占めると思われる。(慶應義塾大学斯道文庫にあるガスパルドン旧蔵本もハンノム文献を含むが、リストは公表されていない)。これ以外にも所蔵機関は複数あるが、前節でも触れたように中国や日本で出版されたものがほとんどである。昨今のベトナム本国における包括的な収集活動からすれば、こんご日本におけるハンノム文献の原本の所蔵が劇的に増えるということはまず考え難いであろう（それこそが民族の財産としての歴史資料の本来的なあり方でもあろう）。

したがって、日本を含む外国の研究者による資料収集活動は原本の複製が中心となるはずである。これについては、東洋文庫で公開されているコレクションの半数近くは EFEO 所蔵本のマイクロフィルム版であることがまず思い起こされるが、個人で収集した資料については公表されないのが一般的であろう。また、こうした複製の収集資料は、仮に公的機関に収蔵されるとしても、山本や松本のコレクションのような原本と違い、そのモノとしての保存方法（原本と同様の取り扱いをすべきか）や、原本の収蔵機関との権利関係等、取り扱いには不透明な部分が残る（東洋文庫にあるマイクロフィルムは寄贈時に電子複写不可との取決めがなされている）。

国内で公開されているハンノム文献の規模から考えると、個人が所蔵する複製資料の存在は決して小さいとは言えず、研究基盤として遺してゆく方策を考える必要があるが、大学図書館をはじめとして、こうした資料の受け皿が整っているとは言い難い状況である。

個別の研究者の活動としては［八尾 2009］が個人で収集したハンノム文献を積極的に紹介しており、連綿と続く、日本人による資料研究を今日に継承するひとつのあり方を示している。

### ■ 3.3. まとめ

ハンノム文献は、漢字文化圏の中では、確かに辺境に位置づけられる存在であろう。しかし、よく見れば、内容的な広範さはもちろんのこと、物理的な数量や収蔵地の広がりについても、案外大きな存在であることをイメージしていただけたかと思う。また日本におけるハンノム文献についても、表だった収蔵数に現れない部分での関連をいくつか紹介し、その存在が決して小さくないことを示した。

筆者が初めてベトナム前近代史の講義を受けた際、その冒頭で紹介されたのが、上述の日本人による書目であった。当時はちょうど『遺産』が出版された直後で、正直な所これら古い書目は、その存在を知識として知っておく必要はあっても、実際の資料探索に利用することがあろうとは考えも及ばなかった。しかしながら、今回、「系統」のデータを精査すると、前後のつながらない部分や疑問点が意外なほど多く現れたのであるが、その解決のための、唯一と言ってよい手がかりが、先学達が手づから写した書目であった。

ハンノム文献と日本人。縁遠いようであるが、注意して見ると存外近辺にあったりする。これはもしかすると、別に文献に限ったことではないのかも知れない。案外近くにベトナム人が居ながら、その国のこと言葉のこと文化のことを全く知らないというのと、パラレ

ルに論ずることができるのかも知れない。そこに気づくか否かは、ほんの些細な差に過ぎないのではなかろうか。

## 参照文献


### 和文・中文

- 岩井大慧 1935「永田安吉氏蒐集安南本目録」『史學』14 (2)：283-291
- 宇野公一郎 1999「書評 Viện nghiên cứu Hán Nôm. 1996. Bảng tra thần tích theo địa danh làng xã. Hà Nội: Nhà xuất bản Khoa học xã hội. 869 pp.; Viện thông tin khoa học xã họi. 1996. Thư mục thần tích thần sách. Hà Nội: Viện thông tin khoa học xã hội. ix+1276 pp.」『ベトナムの社会と文化』1：419-423
- 川本邦衛 1971「越南社会科学書院所蔵漢喃本目録」『慶應義塾大学言語文化研究所紀要』2：103-137
- グエン・タイ・カン（川本邦衛訳）1987「漢字文化とベトナム語：現代ベトナム語における漢字起源の要素」橋本萬太郎ほか編『漢字民族の決断：漢字の未来に向けて』443-468 頁　東京：大修館書店
- 陳文玾 1943『河内遠東攷古學院現藏越南佛典略編』東京：國際佛教協會
- 東洋文庫古代史研究委員会（編）1999『東洋文庫藏越南本書目』東京：東洋文庫
- 藤原利一郎 1974「パリ国立図書館新収安南本目録」『史窓』32：77-81
- 松本信広 1934「河内佛國極東學院所藏安南本書目」『史學』13 (4)：699-786
- 八尾隆生 2004「『大南一統志』編纂に関する一考察」『広島東洋史学報』9：1-35
- 八尾隆生 2009『黎初ヴェトナムの政治と社会』東広島：広島大学出版会
- 八尾隆生；岡田建志 2003「ベトナム史料」早瀬晋三；桃木至朗（編集協力）『東南アジア史研究案内』（池端雪浦他編『岩波講座東南アジア史　別巻』）124-129 頁所収　東京：岩波書店
- 山本達郎 1938「河内佛國極東学院所藏字喃本及び安南版漢籍書目」『史學』16 (4)：571-628
- 山本達郎 1953a「パリ國民圖書館所藏安南本目録」『東洋學報』36 (1)：87-107
- 山本達郎 1953b「河内佛國極東學院所藏安南本追加目録」『東洋學報』36 (2)：245-260
- 山本達郎 1954「パリ亞細亞協會所藏安南本書目」『東洋文化研究所紀要』5：310-352
- 劉春銀他（主編）2002；2004『越南漢喃文獻目録提要』上下冊；補遺上下冊（中央研究院中國文哲研究所圖書文獻專刊 7；8）臺北：中央研究院中國文哲研究所
- 和田正彦 1992-1993「松本信廣博士将来の安南本について（上）（中）（下）」『史学』62 (1)：165-183；62 (3)：125-158；63 (1-2)：127-150

### 欧文・越文

- Berthier, Annie, ed. 2000. *Manuscrits, xylographes, estampages : les collections orientales du département des Manuscrits : guide.* [Paris] : Bibliothèque nationale de France.
- Chu Tuyết Lan. 2005. Thông tin bổ sung về tài liệu Hán Nôm ở Đông Dương văn khố (The Toyo Bunko), Nhật Bản. *Tạp chí Hán Nôm*. 2005 (5) : 65-68.
- Phan Huy Lê, et al., ed. 1995. *Địa bạ Hà Đông (Từ Liêm, Đan Phượng, Thượng Phúc, Hoài An, Sơn Minh)*. Hà Nội.
- Phan Văn Các. 1991. Khánh thành trụ sở mới, kỷ niệm 20 năm thành lập Ban Hán Nôm, nay là Viện Nghiên cứu Hán Nôm. *Tạp chí Hán Nôm* 1991 (1) : 3-6.
- Trần Nghĩa. 1993. Dẫn luận : di sản Hán Nôm Việt Nam. In *Di sản Hán Nôm Việt Nam : thư mục đề yếu*, vol. 1, pp. 15-47. Hà Nội: Nxb. Khoa học xã hội.
- Trần Nghĩa; Gros, François, đồng chủ biên. 1993; 2003. *Di sản Hán Nôm Việt Nam : thư mục đề yếu*. I-III; Bổ di 1 thượng-hạ. Hà Nội: Nxb. Khoa học xã hội.
- Trịnh Khắc Mạnh, etc., ed. 2005-. *Tổng tập thác bản văn khắc Hán Nôm*. Hà Nội : Nhà xuất bản Văn hóa thông tin.
- Truong Đình-Hoè. 1986. Inventaire n° 1 : fonds Hán-Nôm de l'EFEO (Paris). *Bulletin de l'Ecole Française d'Extrême-Orient 75*: 270-294.
- Vũ Thanh Hằng, et al. 2003. *Châu bản triều Tự Đức 1848-1883*. Nxb. Văn học.
- Yano Masataka. 2005. Ghi chú về lưu trữ Hán Nôm của Thư viện Toyo Bunko. *Tạp chí khoa học xã hội* 2005 (7) : 84-86.

# 中国古典学術リソース、回顧と展望

秋山　陽一郎

## ❖　はじめに

中国古典系学術リソースのこの 10 年を振り返ってみると、大きく進化した部分と、本質的にほとんど変わっていない部分とがある。

大きく進化したのは、何といっても学術リソースの量的な充実だろう。ハードウェアスペックやネットワーク環境の劇的な進化も手伝って、各種叢書や、档案、石刻、出土資料等々、この 10 年の間に膨大な数量のデータが世に出た。

逆に、著作権や版権の問題、学術リソースの質的な充実（学術的な信頼性や汎用性）、制作・ランニングコストを誰が負担するかといった問題は、相変わらずくすぶり続けている。

## ❖　中央研究院 漢籍電子文献

中国古典文献の最初の本格的な全文検索データベースとして、台湾中央研究院の漢籍電子文献の存在は欠かせない。

漢籍電子文献は、1984 年に台湾の中央研究院の歴史語言研究所・情報科学研究所・計算中心と共同で開発に着手した漢籍全文検索データベースで、1995 年に WWW 上に公開されてより、今なお、国籍をこえて、多くの学生や、院生、研究者によって利用されている。

それまで索引や原文を冒頭から読んで用例を探していた、あるいは高価かつ場所をとる基本書籍をそろえていた当時、『二十五史』や先秦諸子の原文が、インターネットで無料で手軽に検索できるというインパクトは、とてつもなく大きかったのである。

新漢籍全文資料庫

漢籍電子文献は、その後も連年、学術リソースレビューに取り上げられているように、『十三経注疏』をはじめ、各種文史哲文献や類書など、数多くの古典文献を電子化・公開している。台湾政府の補助金で制作されたということもあって、納税者優遇措置で台湾内の登録者、あるいは提携・契約を結んだ海外の機関のみ利用可能な文献も多数あるものの、台湾以外の地域から利用できる分だけでも相当な量に上り、また登録利用者向けのコンテンツも、わずかずつながら開放されてきたのは、会員諸氏もご承知の通りだ。

なお、一昨年のレビュー（佐藤仁史氏。本誌 9 号、2009 年）でも触れられているように、漢籍電子文献のインターフェイスを新たにし、XML ベースでテキストを構築し直した「新漢籍全文資料庫」が公開された。

さらに、2009 年より台湾内の登録ユーザーに限り、漢籍電子文献が無料で利用できるようになった。

現在新版で日本からアクセスできるのは、『十三経注疏』や、二十五史、紀昀『閲微草堂筆記』、円仁『入唐求法巡礼行記』、数点の通俗小説などわずかだが、本文と注釈の区別、異体字や同義語検索機能という新機能実装された。

このうち「同義詞」検索は、後述の陳弱水氏も指摘していた、異称も含めた用例の検索ができないという、従来の漢籍電子文献の弱点に対応したと思われるもので、内部的にマークアップが施されつつあるものと思われる。

ただ、現時点では、まだ研究の用に堪えると言うにはほど遠い段階だ。「異体字」検索の方は、今のところ、各種俗字や通仮字を含む、高度な異体字同一視検索といったものにはなっていないようだが、「別録」で検索して「別錄」がヒットするなど、文字コード上の問題に詳しくない、日本人を含む一般的な外国人ユーザーにやさしくなった感はある。

中央研究院はこのほかにも、中国古代の出土資料を扱う歴史語言研究所の文物図象研究室資料庫や、内閣大庫档案、傅斯年圖書館蔵善本図籍など、目録や画像データも含むさまざまなデータベースを公開している。

●漢籍電子文献（バージョン 2.0）
http://www.sinica.edu.tw/~tdbproj/handy1/
●新漢籍全文資料庫
http://hanchi.ihp.sinica.edu.tw/
●文物図像研究室資料庫
http://saturn.ihp.sinica.edu.tw/~wenwu/search.htm
●数位典蔵資料庫
http://repository.digiarch.sinica.edu.tw/

なお、漢籍電子文献の経緯や詳細については、1994–1997 年にかけて同院の副所長・歴史学組主任としてプロジェクトに参与した、陳弱水氏の「中央研究院歴史語言研究所　漢籍全文自動化計画の発展、現状、未来」（野村英人訳。本誌 2 号、2001 年）、「デジタルアーカイブと東洋学 ── 中央研究院歴史語言研究所の経験から ──」（山下一夫訳。本誌 5 号、2004 年）に詳しい。

後者は 2004 年 3 月に花園大学で行われた本会の春期公開講座の内容をまとめたものだが、その時のレポートである師茂樹氏の「春期公開講座レポート」（本誌 5 号、2004 年）もある。また、初学者や学生にとっては、山田崇仁氏のオンラインマニュアル [1] も有益なリソースだろう。

## ❖ その他のウェブ系全文検索データベース

中央研究院の漢籍電子文献と共に、本誌創刊時分によく利用されていたのが、台湾・香港系の老舗ウェブ検索サービスだ。

まず陳郁夫氏の龍泉二号のウェブ版である寒泉。特に『資治通鑑』や『四庫全書総目提要』の検索で世話になった方は多いはず。このほかにも『全唐詩』・『宋元学案』・『明儒学案』・『朱子語類』など、基本的に読むより、工具書的に検索するような用途で利用する文献が多いのも特徴的だった。

ただし、寒泉については、台湾故宮のサイトで公開当初存在していた二十四史が削除されるなど、権利関係が理由と思われる、データベース利用形態の変更が行われている。冒頭でも述べたように、著作権や中国大陸と台湾との関係など、複雑な権利関係や政治状況がこのような場にも色濃く陰を落としている。

また先秦両漢古籍逐字索引叢刊・魏晋南北朝古籍逐字叢刊の元データをウェブから検索できるようにした、香港中文大学の華夏文庫（のち、2002 年に漢達文庫と改称）。ここは中国の古代・中世文献を扱う研究者に有益な独自データが多い。特に『全上古三代秦漢三国六朝文』が無料で検索できたことに驚いた会員も少なくあるまい。甲骨文や金文（『殷周金文集成釈文』）を含む出土資料類や類書の電子化を行っている点もこのサイトの特長といっていいだろう。これらのデータベースは、当初、データがフロッピーディスクや CD-ROM の形態でも頒布されていた。

古典文学（詩詞・小説・戯曲など）方面では、平仄・音韻・方言・音声・画像などのデータを所々含んだ、台湾元智工学院の網路展書読も老舗中の老舗ながら、今なお、そのコンテンツには見るべき特色がある。

●寒泉
http://libnt.npm.gov.tw/s25/
●漢達文庫
http://www.chant.org/
●網路展書読
http://cls.hs.yzu.edu.tw/

ほかに特定の分野や文献に特化したものの中で特筆すべきは、本誌創刊号（2000 年）で取り上げられた北京大学の『全唐詩』・『全宋詩』電子検索系統が挙げられよう。

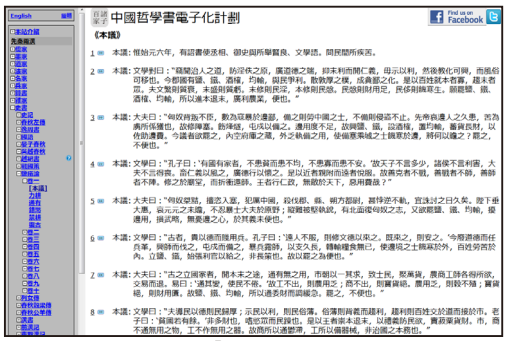
ここでは『全漢三国晋南北朝詩』のほか、『詩話総編』や『唐詩紀事』・『唐才子伝』・『全宋詞』・正史列伝など、関連する文献の記述も同時にウェブで検索できたため、利便性が非常に高かった。

ところが 2006 年にこれらのデータベースが、突然アクセスできなくなる。代わって置かれた全唐詩分析系統・全宋詩分析系統の試用版も、翌年の以下の CD-ROM 版の発売により、その後、跡形もなく取り除かれてしまった。

本誌 8 号の千田氏のレビューでも糾弾されているように、いずれも「事前・事後のアナウンス」なく一方的に削除された点は、一定の利用者を持つサービス提供者の責任を蔑ろにするもので、甚だ残念である [2]。

このように、研究者や研究機関は、論文同様、公開する成果物に対して一定の責任を負うべきことを、再確認させられた事例でもあった。

●北京大学数据分析研究中心『全唐詩分析系統』北京



中国哲学電子化計画で、『鹽鐵論』のテキストを表示

大学出版社、2007 年（CD-ROM 版）

●北京大学数据分析研究中心『全宋詩分析系統』北京大学出版社、2007 年（CD-ROM 版）

### ■ 中国哲学電子化計画

ボランティアベースでデジタルテキストデータベースを作成、公開しているサイトである。先秦から後漢・三国辺りを収録範囲としているが、他のデータベースが一般未公開としている書籍のデータを検索できるのは便利。ただし、版本の情報などが一切なく（おそらく下記のデジタルテキスト系サイト辺りから入手したものだと思われるが）、専門的な研究で使用する場合には、信頼できる標準テキストとの対校が欠かせない。

●中国哲学電子化計画
**http://ctext.org/zh**

## ❖ デスクトップ系全文検索データベース

現状、中国史系のデータベースにおいて、もっとも質と量を兼ね備えているのは、スタンドアローンタイプの全文検索データベースだろう。

### ■『四庫全書全文検索版』のインパクト

まず、2000 年に香港迪志文化出版有限公司より『四庫全書全文検索版』が発売された。全文検索部分が CD-ROM16 枚、書影の画像データが 166 枚（スタンドアローン版）分という大部のソフトウェアだが、紙媒体の影印本ですら六畳一間を埋め尽くさんばかりとでも表現される書架スペースを占有していたことを考えると、1 台の PC 端末に収まるという事実に、筆者も当初はなかなか実感が湧かなかったことを、今でもはっきり記憶している。

### ■『四部叢刊全文検索版』

続いて翌 2001 年末には、商務印書館の四部叢刊初編・続編・三編を電子化した『四部叢刊全文検索版』が、『四庫全書』電子化の実際の作業を担当した書同文数字化技術有限公司（以下、書同文）から出た。こちらは、全文検索部分が CD-ROM4 枚、書影データが 20 枚となっている。『四庫全書』に比べると、カバーしている文献の範囲は狭いが、四庫全書編纂時にも採集されていない、宋元明の善本を直接影印収録している点で、研究上の利用価値が高いのが最大の特長である（ただし、文献によっては、古逸叢書など、より優れた善本が存在する場合もあるので、かならずしも四部叢刊本が最善のテキストになるとは限らない）。

これら『四庫全書』と『四部叢刊』は、その後、西岡漢字情報工学研究所より日本語版も発売されている。大陸で電子化されるデータが簡体字で入力されることが多い中、書同文のテキストデータは、繁体字を使用している上、異体字検索も充実しているため、研究者には非常にありがたい。なお、書同文については、千田大介氏の取材による報告「中国における古典文献データベースの構築　書同文へのインタビューを通じて」（本誌 2 号、2001 年）がある。OCR や対校システムといった、同社の電子化技術の核心部にも触れられており、貴重な資料である。

更に書同文は、開発した高機能な OCR を用い、『歴代石刻史料彙編』『康熙字典』『十通』など、大部な漢籍叢書シリーズのデジタル化を一時期積極的に行ってきた。

しかし、書同文はその後この分野の新製品開発やアップデートが滞り、『四部叢刊』は Windows Vista、『四庫全書』は Windows 7 の 64bit 版で動作しない（32bit 版では動作する）という致命的な問題が生じてしまった。一応、対応パッチが書同文から配布されており、また Windows 7 の XP モードや VM WARE などの仮想環境を使用すれば回避可能だが、いつまたバージョンアップなどで動作不可になる（高価なソフトがゴミと化す）可能性が存在するのは不安である。

そのため、書同文では上記パッケージ版で販売したデータベースについて、Web 版の提供を始めている。これについては本号の千田氏のレビューを参考にしていただきたいが、少なくともネットワークに繋がる環

境であれば、OS を問わず使用可能になるので、Mac OS X や Linux などの利用者にはうれしい措置だろう。ただし、これについてもサーバーの設置場所など、インターネット接続環境次第で使用できないという問題は常につきまとうのだが。

### ■『中国基本古籍庫』

書同文のソフトウェアも、中国史研究に与えた影響は相当なものがあるが、更にそれを上回るインパクトを与えたのが北京愛如生数字化技術研究中心から登場した『中国基本古籍庫』である。

本ソフトウェアは、単に『四庫全書』の三倍という分量の書籍をデジタルテキスト化するだけではなく、底本となる版本を複数入力してそれを比較検討する機能を持つ、空前のデータベースとなった。もっとも、価格の方も 1,000 万円超えという個人ではまず手が出ない価格帯であり、研究機関であっても、高額の予算や研究費がつきにくい人文系の研究機関では、購入を躊躇するところもあるだろう。しかし、おそらく現時点でもっとも使いでのあるデータベースであることは間違いない。

さらに愛如生では、基本古籍庫のデータを中心に、それ以外のテキストも組み合わせて USB 形式のシリコンディスクの形でデータベースの提供を始めた。これについては、本誌十号の齊藤正高氏によるレビューを参照していただきたい [3]。

### ■ 国学

このほかにも大陸では、総集や類書・史書・筆記小説などが詰まった国学の基本典籍庫がある。上記のデータベースに比べてそれなりに優位な点もあるが、古典文献学としての中国学からみた場合、簡体字でデータが提供されているというのは、弱点としか判断し得ない。

### ■ スタンドアロン系データベース共通の問題

ただし、これらのデータベースは、膨大なデータと検索システムを提供してくれる反面、応用・発展性に欠ける閉鎖的なデータベースだともいえる。テキストデータの複製には厳しい制限がかけられ、そのままでは、ほかのデータベースとの連携や、独自の解析を加えることもできない。これだけの膨大なテキストデータを前に、初歩的な横断検索しかできないのは、何とも歯がゆいところである。

また、新しい OS に対するサポートの問題も表面化してきている。高額で購入したデータベース資産が OS や端末のアップデートと共に化石化するのはあまりに痛い。上記書同文のように Web 版への移行を進める動きもあるが、結局、継続的にサービスが保守・管理されなければ、多大な投資が無駄になるリスクは潜在化したままだ。

資産の化石化という意味では、昨今、流行しつつある電子書籍も同様だ。PDF、EPUB、AZW、CEBX（JEBX）等々、目下、電子書籍のフォーマットは乱立状態だ。著作権保護技術とも相俟って特定の端末・サービス・企業の寿命と共に、データ資産も化石化するようでは困る。特に最近の新しい著作物の場合、特定の媒体でしかアクセスできないとなると、後日、著作物の保存や利用が困難になる懸念もある。

## ◈ 古典文献全文データ（テキストデータ）

テキストデータについては、本誌創刊時以来の定番どころでいえば、台湾の中華文化網、大陸の国学・新語絲電子文庫・亦凡公益図書館が挙げられるだろう。

一昨年 9 号でレビューされた Michael Hart 氏の The Project Gutenberg も、1971 年以来の老舗プロジェクトだ（最近の iPad や iPhone 用電子書籍ビューワー、iBooks でダウンロードできる中国古典籍の多くも、この The Project Gutenberg 筋のデータ）。

同プロジェクトは、青空文庫のようなボランティアベースの電子化という点でも注目されるが、似たようなリソースとして、維基文庫（WikiSource）もある。また最近では掲示板サイト西陸網のフォーラム「古典小説之家」も、叢書や筆記・別史・通俗小説などで「他サイトには見られない学術的価値のある文献」がセレクトされてるとはいえ、他サイトや特に『四庫全書』からの転載が目立つ。

この辺りの著作権侵害問題は、古典テキストの著作権や校訂に関する権利などとの関係でグレーゾーンであることは言うまでもないが、テキストサイトの老舗である中華文化網が、そもそも漢籍電子文献のデータを利用したものであり、そこから派生して数多くの同様なサイトが乱立してきた歴史的経緯もある。

最近はそこからさらに簡体字化を行ったり、フリーな配布形態を標榜する Web サイトへの（故意・ある

いは善意による）転載され、結果としてテキストロンダリングとでも称しえるような過程を経て公開されている状態になっているところもあるようだ。この辺りまでくると、もはやデジタルテキストの一次作成者やその底本情報がすっかりわからなくなり、そのため（底本の本文の著作財産権が消滅しているものについては）、フリーな状態で配布されいても文句が言えない状況になっているのが現状である。

無論、出土文字を隷定する作業や、高度な文献学的知識を背後に持つ校訂作業などを経て生まれたテキストに対し、心情的なリスペクトは別にして著作権法などの規定が曖昧なのは事実である[4]。

日本では青空文庫のように、その辺りにきちんと対応しているところもあるが、中華系テキストサイトにはそのような状況はあり得るのか？　そのようなサイトがあふれているからこそ、上記のような自衛措置（テキストコピーに関する様々な制限）が施されているのだろうし、おそらくは今後もそのようなサイトは現れないと悲観的になってしまう。たとえ現れたとしても、あっというまにそこで公開されているテキストはコピーされ、さらにそこから増殖を始めてしまい、埋もれてしまう可能性が高いからである。

- The Project Gutenberg
  **http://www.gutenberg.org/wiki/Main_Page**
- 維基文庫（WikiSource）
  多言語用ポータルサイト
  **http://wikisource.org/wiki/Main_Page**
  中文版 WikiSource
  **http://zh.wikisource.org/zh/%E9%A6%96%E9%A1%B5**
- 国学
  **http://www.guoxue.com/**
- 新語絲電子文庫
  **http://www.xys.org/library.html**
- 亦凡公益図書館
  **http://www.shuku.net/novels/cnovel.html**
- 古典小説之家
  **http://club.xilu.com/wave99/**

## ❖　図書館所蔵本の検索・公開

中国史研究の材料として使用される書籍は、基本的にどこかの研究機関や個人で所蔵されている古典籍（そのものや、それをリプリント・活字化したもの）である。そのため、ある分野の研究をする場合には、どの図書館にどの本が所蔵されているのかを知る必要がある。

かつてのカード検索の時代から、現在では OPAC による検索が主流となった。この辺りの歴史的推移は、本誌の小島氏による「図書館・OPAC」に関係する断続的な記事を参照していただきたい。

昨今は、単に所蔵情報を検索するだけではなく、所蔵する本そのものも公開する所が増えてきた。無論、その中身は一部の抜粋から全部という違いがあるが、東京大学東洋文化研究所や早稲田大学が公開している漢籍類や、日本の国会図書館の近代デジタルライブラリー、更には Google ブックスの登場によって、漢籍のみならず明治以降の書籍まで利用可能になったことは驚きだろう。特に近代デジタルライブラリーや Google ブックスによって、これまで稀覯本や劣化によって読むことが難しくなってきた研究所などを気軽に扱えるようになったことは、大変に喜ばしいことである。

- 東京大学東洋文化研究所
- 早稲田大学
- 漢籍データベース
- 国会図書館近代デジタルライブラリー
- Google ブックス

## ❖　マッシュアップによる新たな展開

マッシュアップは本来音楽用語だが、Web の世界では既存の Web サービスの API を組み合わせて、新たな Web サービスを生み出すことを指す。

これまでは、Web サービスといっても、独自のインターフェイスやデータベースを利用したものが圧倒的であったが、一昔前の Web2.0 の流れの中から、この分野でもマッシュアップを利用した Web サービスの構築が試みられるようになってきた。

例えば中央研究院歴史語言研究所の提供する「殷周青銅器地理資訊系統」が挙げられる。この Web サイトは、Google マップをベースに、青銅器の出土地を単位としてマッピングし、検索できるようにしている。また、独自の地図をレイヤーとして重ねて表示することで、古代史の地図としても利用可能だ。

また、これはマッシュアップというわけではないが、

ToutTube を利用して音声による情報提供を行っている試みもある。

その一例として、大島正二『唐代の人は漢詩をどう詠んだか ── 中国音韻学への誘い』（岩波書店。2009年）が挙げられる。本書は、いわゆる中国中古音の音価がどのように復元されてきたのかについて、その研究史と復元過程をわかりやすく述べた本だが、このほんのおもしろい所は ToutTube 上に実際に復元した音で唐詩を朗読した音声をアップロードしている点だろう [5]。レビュワーは「あとがき」でこれを知ったのだが、これについてはもっと目立つところでアピールしてもよかったのではないかと思う。

音声を伝えるという行為は、紙媒体ではなかなか難しい。また、わざわざそのために Web サイトを立ち上げるのも大変だろう。そういう意味では、ToutTube を利用するというアイデアはなかなかである。

このような試みは、上古音で『詩経』を読んだ事例を ToutTube にアップロードするなど、いくつか行われているようだ [6]。また、単に人が読むだけではなく、初音ミクに読ませてみるというおもしろい試みもあるようだ [7]。

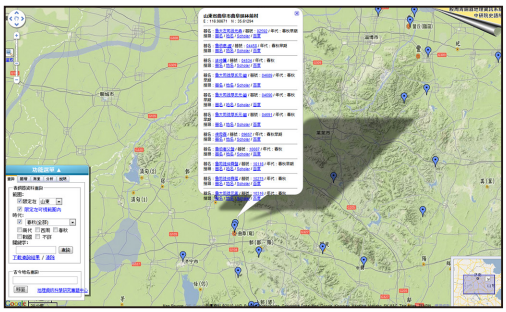
## ❖ 終わりに

以上、本誌刊行後の中国史分野の学術リソースについてつらつらと整理してみた。

基本古籍庫の登場で、大規模テキストデータベースについては一段落した用に思われるが基本古籍庫がハードディスクで提供されるように、あまりの大規模化はスタンドアローン系がもはや限界に来ていることをも伺わせる。今後は、書同文系データベースのようなクラウド化の方向や、愛如生のような USB 系のデータベース販売に移るのかもしれない。

## 注

[1] 漢籍電子文献の使い方
**http://www.shuiren.org/chuden/toyoshi/sinica/index-j.html**
文物圖象研究資料庫を使ってみよう
**http://www.shuiren.org/chuden/toyoshi/zuzou/index-j.html**



殷周青銅器地理資訊系統

また新漢籍全文資料庫については、以下のページにヘルプがある。
**http://hanchi.ihp.sinica.edu.tw/ihp/help.htm**

[2] 千田大介「学術リソース・レビュー／中国古典文学」（本誌 8 号、2007 年）

[3] 齊藤正高「USB 個人用小型データベース ── 愛如生拇指数據庫：『方以智著作集』を例に」（本誌 10 号、2009 年）

[4] この辺りの議論は石岡克俊「「校訂」の著作権法における位置」『慶應義塾女子高等学校 研究紀要』第 26 号、2009 年 3 月。を参照されたし。なお、当該論文は、以下の URL で公開されている。
**http://www.sanken.keio.ac.jp/publication/KEO-dp/116/KEODP116.pdf**

[5] 一例として、杜甫「春望」唐代長安音（推定）のリンクを紹介しておく。
**http://www.TouTube.com/watch?v=7-js1_JK_zw**
その他、李白や孟浩然などについてもアップロードされている。また、参考用に現代中国音でも同様の作業が行われている。

[6] 《關雎》上古漢語朗讀 Kroon Shyaa leiendo en chino antiguo
**http://www.TouTube.com/watch?v=Bqt3_02lxGo**
鄭張尚芳『上古音系』（上海教育出版社。2003 年。）の復元音に従ったもの。
《關雎》（王力系統）
**http://www.TouTube.com/watch?v=3wu6PS7nAF0**
音韻学者として著名な王力の復元音によるもの。

[7] 【初音ミク】詩経 ── 蒹葭（自作曲）
**http://www.TouTube.com/watch?v=pAdOkUQ5ym8**
どの復元音に基づいたのかは不明。

# 大学授業へのTwitter導入の事例報告
## ──大妻女子大学「情報社会論」での経験からの14の教訓

岡本　真

### ❖ はじめに

2010年の4月から7月にかけて、大妻女子大学社会情報学部で非常勤講師として、「情報社会論」の授業を担当した。この「情報社会論」では、一貫してインターネットを扱ったこともあり、試験的に授業の中で学生にTwitterの使い方を指導し、授業中や授業の前後にTwitterを使うことを推奨してみた。結果としては、当初の思惑通りにいかないことが多かったが、今後、同様の取り組みを行う教員に参考になると思われる知見も多々得られた。そこで本誌を借りて、今回の経験と得られた知見をつづっておきたい。拙い経験ではあるが、本稿が大学で教鞭をとる教員の方々の参考になれば幸いである。

それでは、以下、授業での導入プロセスから実際の活用方法をつづっていこう。なお、参考までに述べておくと、大妻女子大学での「情報社会論」は、社会情報学部の2年生から4年生、53名が受講し、授業は毎週火曜日の10時40分から12時10分にかけての90分間で合計13回行われた。授業の構成は以下の通りである。

### ❖ 授業中の一斉アカウント作成

Twitterの導入にあたっては、まず学生に授業時間の中でTwitterのアカウントを作成してもらうところから開始した。全13回の授業だったが、Twitterの

表1：2010年度「情報社会論」のカリキュラム

| | | |
|---|---|---|
| 第 1 回 | 4月13日 | インターネットとは何か──授業の概説とインターネットに関する基本的な説明 |
| 第 2 回 | 4月20日 | 情報・知識とは何か──私たちはこれまでどのように情報・知識に接してきたのか |
| 第 3 回 | 4月27日 | インターネットの特性(1)──検索エンジンの仕組み |
| 第 4 回 | 5月11日 | インターネットの特性(2)──検索エンジンでは検索できない情報 |
| 第 5 回 | 5月18日 | インターネットの特性(3)──インターネットで使われている様々な仕組み |
| ── | 5月25日 | 休講 |
| 第 6 回 | 6月 1日 | 変わるインターネット(1)──ソーシャルメディア |
| ── | 6月 8日 | 休講 |
| 第 7 回 | 6月15日 | 変わるインターネット(2)──メディアのインターネット対応 |
| 第 8 回 | 6月22日 | 変わり出した社会(1)──コミュニティーの変化 |
| 第 9 回 | 6月29日 | 変わり出した社会(2)──制度の変化 |
| 第10回 | 7月 6日 | 変わり出した社会(3)──情報・知識の変化 |
| 第11回 | 7月13日 | まとめ(1)──インターネットによる情報・知識の変化（教員の講義中心） |
| 第12回 | 7月20日 | まとめ(2)──インターネットによる情報・知識の変化（学生の発表中心） |
| 第13回 | 7月27日 | まとめ(3)──インターネットによる情報・知識の変化（教員・学生の討論中心） |

アカウント作成は第 6 回目(2010 年 6 月 1 日)に行った。表 1 の通り、それまでの 5 回の授業で検索エンジンを中心にインターネットの基本的な仕組みを解説し、この第 6 回では、Twitter に代表されるソーシャルメディアを解説した。Twitter の利用は、ソーシャルメディアの特性を頭だけで理解するのではなく、学生一人ひとりに実感を持って理解してもらいたいと考えたためである。

授業は第 1 回を除いて、一貫してパソコン教室で行い、学生一人ひとりがパソコンを使える環境を用意した。事前に学生たちに確認したところ、本授業以前から Twitter を利用していたのは出席していた学生約 35 名のうち、1 名のみであったため、授業の中で学生にアカウントを作成してもらうことにした。授業では前回の復習を 15 分ほどした後に、学生に Twitter の公式サイトにアクセスしてもらい、一斉にアカウントを作成してもらったのだが、まずここで大きなトラブルが発生した。同時に同一の IP アドレスから一斉にアカウントを作成しようとしたため、Twitter 側の不正利用の可能性があるアカウントの一斉取得防止プログラムにひっかかり、最初の数名を除いて、他の学生がまったくアカウントを作成できなくなったのだ。問題の発生当初は、教員である自分自身、アカウントが一斉に作成できなくなってしまった理由がわからず、相当困惑したことを覚えている。後述するが、事前に Twitter の自分自身のアカウントで、この授業で学生に Twitter を利用してもらうことを記し、約 3000 名のフォロワーの方々に協力を呼びかけていたのだが、呼びかけに応えてくれた数名のフォロワーの方々のご指摘によって、ようやく自分自身、問題の所在を把握するという状況であった。

結局、この日は時間をあけて、数名ずつアカウントを作成するよう学生たちに伝えたものの、結局 5 名ほどの学生がアカウントを作成できたのみであった。いま振り返ってみれば、完全に教員側のミスであったが、以下のような教訓が得られたのではないかと思う。

《大学授業での Twitter 利用の教訓》

1. Twitter のアカウント作成を一斉に行うのはリスクが高く、事前に学生にアカウント作成を依頼しておく。
2. 仮に授業内で一斉にアカウント作成を行うのであれば、携帯電話やスマートフォン等のモバイル機器経由で行う。
3. 上記が難しい場合は、2、3 名ずつ、時間をあけながら徐々にアカウントを作成する。

図 1：筆者のアカウントでの Twitter の画面

## ❖ Twitter の利用方法（第一段階）

1 回の休講を挟んで 2 週間後の第 7 回授業では、上記 2．3．の方法で、あらためて学生にアカウントを作成してもらい、Twitter の利用を開始した。なお、アカウント作成にあたって、学生にはアカウントに本名を用いないこと、プロフィール等にも本名を入力しないように伝え、プライバシーの保護を図っている。この点については後述したい。

学生全員がアカウントを作成したところで、まず、

1. つぶやく（Tweet）
2. フォローする／される（arg）
3. ハッシュタグをつけてつぶやく（#otsuma_is）
4. リプライしてみる／されてみる
5. RT してみる／されてみる

という 5 つの作業を順次行ってもらった。「1.つぶやく（Tweet）」では、何でもいいので、一言つぶやいてみるようにと指示したが、かえって困惑する学生が少なからず見られた。ここは「情報社会論の授業なう」といった定型文を用意し、初めてのつぶやきへのハードルを低くしておくほうがよかったかもしれない。

次いで、教員である私のアカウント(arg)をフォローしてもらい、順次私からフォロー返しを行った。また、ここまで進んだ学生には、隣に座っている学生のアカウントをフォローし、フォローされたら、逆にその学生をフォローするように指示した。だが、ここでも学生同士では、誰がどのアカウントなのかがわからず、

右往左往するさまが見てとれた。これも反省ではあるが、「3.のハッシュタグをつけてつぶやく（#otsuma_is)」を実践してから、学生同士のフォローを勧めるべきだったろう。ただし、厳密なことを言えば、ハッシュタグは学生以外の誰もが使えるものなので、最初の段階で学生が学生以外の方々のアカウントをフォローすることを避けるのであれば、私のアカウントをフォローした学生に、まず私あてにリプライしてもらい、そのアカウントにさらにリプライをする形で学生に一緒に授業を受けている学生のアカウントを一つずつ紹介するほうが望ましいかもしれない。

例：

arg はい！ありがとうございます！ RT @****
つぶやけました！

つぶやき、フォローする段階を経て、次にハッシュタグをつけてつぶやいてもらった。ハッシュタグはあらかじめ「#otsuma_is」とした。この際、「otsuma」と「is」の間にある「_」（アンダーバー）を入力できない学生が散見された。確かに、日常のパソコンやモバイル機器の利用シーンでは使うことが少ない記号であるため、ここは「#otsumais」のようなハッシュタグのほうが望ましかったかもしれない。だが、IT リテラシーの向上を図るという意味では、この機会に「_」（アンダーバー）の入力を覚えてもらうという意義もあるだろう。

ハッシュタグ「#otsuma_is」をつけて学生につぶやいてもらい、このハッシュタグでの Twittter の検索結果画面が学生のつぶやきであふれてきた頃を見計らって、学生たちに他の学生のつぶやき、あるいは一連の様子を見て、同じくハッシュタグ「#otsuma_is」をつけてつぶやいてくれている外部の方々にリプライをするように指示した。この段階まで来ると、大部分の学生が Twitter に馴染みだしてきていることが感じられた。しかし、一部には、勝手がわからない学生の姿も見受けられ、その支援に教員である私が奔走することになった。今回の授業は非常勤講師としての担当であり、TA は不在で、学生たちとの日常的なコミュニケーションもなかったが、本務校で授業を行う教員であれば、TA を手配する、あるいはすでに Twitter を活用している学生をヘルパーに起用するといった対策が求められるだろう。

最後に「5. RT してみる / されてみる」に取り組んだ。これは Twitter の大きな特徴である RT（Retweet）による波及力の大きさを実感してもらうことをねらったものである。しかし、授業の時点ではフォロワーが少ない学生たちにとっては、RT することはできても、RT されることは少ない。また、学生同士で RT しあっても、学生同士ではフォロワーが少ないため、結局その効果を実感できなかったと思われる。私のフォロワーを中心に外部の協力者たちが熱心に RT してくれたものの、ここはやはり事前に私が自らのフォロワーで、かつフォロワーが多い方に事前に協力を依頼しておくべきであったかもしれない。

以上の Twitter 利用の第一段階を振り返ると、以下の教訓が得られた。

《大学授業での Twitter 利用の教訓》

4. 最初のつぶやきにはあらかじめ定型文を用意しておく。
5. 学生同士のフォローはハッシュタグの使い方を教えてから行う。
6. 教員の方針次第だが、ハッシュタグは極力簡易なものにする。
7. TA や学生ヘルパーをあらかじめ手配し、授業の流れに遅れをとっている学生を支援する。
8. RT の効果を実感させるためには、事前にフォロワーの多い方に協力を依頼しておく。

## ❖ Twitter の利用方法（第二段階）

以上の利用方法を第 6 回、第 7 回の授業で、それぞれ授業時間 90 分のうちの 30 分程度をあてて指導したが、実際のところ、すべての学生が上記の 5 つの作業を円滑に行えたわけではない。そこで第 8 回の授業以降、毎回前回授業の復習を約 10 分行った後、約 20 分を Twitter 利用の復習にあてた。第 8 回授業では、上記 5 作業のうち、「3.ハッシュタグをつけてつぶやく（#otsuma_is)」以降を、第 9 回授業では、4.リプライしてみる / されてみる以降を、という案配である。

Twitter への学生たちの習熟が見られだした第 9 回授業からは「6.ブロックする」を、第 10 回授業から第 13 回授業にかけては、「7. #otsuma_is で意見を述べ、意見をもらってみる」を追加して指導した。特に

今回の授業が女子大学での授業であることを踏まえ、学生のプライバシーには十分に配慮し、学生に学外者のフォロワーが増えてきた頃を見計らって、「6. ブロックする」で意に沿わないフォロワーを排除する方法を教えたほか、授業中に繰り返し、ハッシュタグ「#otsuma_is」を一度でもつけてつぶやいた以上、学生一人ひとりのアカウントは見知らぬ第三者から大妻女子大学の学生である可能性が極めて高いと知られていること、日常の行動をつぶやく際は、自分の所在地や行動が第三者に知られても問題がないものかどうかを十分に留意するよう指導した。このため、授業時間外に、何度か学生一人ひとりのアカウントをチェックし、フォロワーの増加傾向や、私自身の経験上、危険性のあるアカウントからフォローされていないかを確認している。

以上のような取り組みを経て、最初のTwitter利用から約1ヶ月を経た第10回授業（2010年7月6日）から、私の講義に対する意見を述べ、ハッシュタグ「#otsuma_is」をつけて発言している他の学生や外部の協力者たちと幅広く意見交換するよう指導した。1ヶ月という時間はいささか長いように思われるが、授業開始前のTwitter利用者が1名のみという状況を考えれば、やむを得ないだろう。

講義内容を受けての発言をTwitterに書き込んでもらうことも意識して、授業では冒頭にその回のテーマに関する以下の問いかけを用意した。

- 第6回授業：　ソーシャルメディアを、どう定義しますか？
- 第7回授業：　メディアとは何か？
- 第8回授業：　コミュニティーとは何か？――あなたの身近なコミュニティーとは？
- 第9回授業：　「制度」という言葉から、何を思い浮かべますか？
- 第10回授業：　「情報」「知識」という言葉から、何を思い浮かべますか？

これらの問いかけには、授業の中で学生たちに口頭で答えてもらいつつ、口頭で答えている学生以外に、Twitterに自分の考えを書き込んでもらった。第6回授業から第8回授業あたりまでは、学生も手探り段階という印象であったが、第9回授業での問い「『制度』という言葉から、何を思い浮かべますか？」あたりからは、学生の発言が比較的活発に見られるようになり、なかなか挙手しない学生がTwitterに非常に優れた回答を書き込んでいるのをみて、その学生を指名し、さらに詳しく発言するよう促すことができた。

ただし、実際のところ、Twitterに熱心に書き込んだ学生は出席者35名のうち、10名弱といったところだ。この点は、もう少し授業全体での設計が必要だろう。特に大部分の学生にとっては、授業を聞きながら、その内容への反応を書き込むという行為そのものが未知の領域であることをもっと意識すべきであったという反省がある。教員がどのような振る舞いを求めているのかを明示的に学生に伝えるためには、たとえば、Twitterと連動したUSTREAMの番組を事前に見せるといった配慮が必要かもしれない。

また、学生たちが授業時間外にTwitterを使うようになるかどうかが、授業へのTwitter導入の成否を分けるように思われる。実際、授業中にTwitterで熱心に発言する学生の多くは、授業時間外でもTwitterでつぶやいている様子が観察された。では、Twitterを学生にとっての日常的なツールにするためには、何をすべきだろうか。これは仮説に過ぎないのだが、Twitterのプロフィールの入力やアイコン、デザイン、背景画像の設定に時間を早い段階で確保するべきだったかもしれない。これはウェブプロデューサーとしての経験知だが、コミュニケーションのためのウェブサービスは自分専用と感じられるカスタマイズを施すことによって、そのサービスへの愛着が高まり、利用の頻度が高まることがあるからだ。

さらに、これまでも触れているが、外部の協力者の存在は欠かせない。今回の取り組みでは、毎回の授業に数名の自発的な協力者が現れ、学生たちに有益な示唆を与えてくださった。授業の評価はレポートで行っ

図3：Twitterの背景設定画面

たのだが、レポート執筆に際して、Twitter で外部の方々のアドバイスを受けることを推奨したところ、数名の学生のレポートには、その痕跡が認められた。学生たちにとっては、私という一人の教員の意見や考えを相対化する装置として、Twitter を介した外部とのコミュニケーションが機能したように思われる。

このような外部の協力を得るためには、毎回の授業の都度、授業開始の 2、3 時間前に当日が授業日であること、Twitter の実習を行うこと、ハッシュタグ「#otsuma_is」がついたつぶやきに反応してほしい旨をつぶやいた。また、授業開始の 30 分前や授業開始時にも協力を呼びかけるつぶやきをしている。この際、重要となるのは、教員自身が一定数のフォロワーを有していることだろう。私の場合、約 3000 人のフォロワーには、研究者やライブラリアン、ウェブ業界関係者が多く、これらの方々の協力を得られたことは大きかった。その意味では、Twitter を授業に導入するにあたって、教員自身が Twitter 上に広範なネットワークをあらかじめ築いておくことが必要だろう。また、授業そのものは USTREAM で中継されているわけでもなく、外部の協力者にとっては、いま授業で何をしているのかがわかりづらい。また、教室内での学生の反応も当然見えてはこない。外部の協力者たちがフラストレーションを感じずに協力できるように、授業中に随時、教員が状況をつぶやくことや、授業後に授業終了と簡単な状況報告をつぶやくことが必要だろう。

《大学授業での Twitter 利用の教訓》

9. 学生の習熟度に依るが、毎回の授業で復習のための実習時間を確保する。
10. Twitter の利用は自身のライフログを不特定多数に公開するものであり、一定のリスクを伴うものであることを伝え、特定ユーザーのブロック方法を教える。
11. 授業内容に関連した書き込みを促すため、共通の質問を授業の中に織り込んでおく。
12. 授業を受けながらの書き込みの実際を想像できるように、Twitter と連動した Ustream の番組を事前に見せておく。
13. 学生の日常的な Twitter 利用を促すため、プロフィールやデザイン等のカスタマイズを行う時間を早い段階で確保する。
14. Twitter 上で外部の協力者を得るために教員自身があらかじめ、Twitter 上に広範なネットワークを築いておく。また、授業中や授業後に状況を共有する。

## ❖ Twitter 導入を振り返って

以上、大妻女子大学「情報社会論」での Twitter 導入の実際を概観してみた。最後に、大学の授業に Twitter を導入する意味を考えてみたい。

今回の Twitter 導入にあたって学生に口頭でアンケートをとったところ、すでに Twitter を使っている学生は 1 名のみだったが、同種のサービスである mixi ボイスは半数以上の学生が利用していた。この差は非常に象徴的であると感じている。携帯電話のメール機能を含め、少なくとも大妻女子大学の今回の授業履修者たちは、自覚的につながっている閉ざされたネットワークはすでに有している。しかし、Twitter のようにオープン性の高いコミュニケーションのネットワークには大部分が参加していない。単位認定を行う授業という強制力が働く場面に Twitter を導入することは、まずは学生たちを自らが自覚的に創り上げた環境から、見知らぬ第三者にあふれた外の世界に引き出すことを意味している。

これは教員についても言える。これまで、教員は教室という閉ざされた環境の中におり、好むと好まざると、いわば絶対的な存在になっていた。しかし、Twitter を授業に導入し、授業中に教員の講義内容や、それへの反応をつぶやくことを学生に推奨すれば、授業は極めてオープンなものへとなっていく。当然、教員の地位は、従来のような絶対的なものではなくなり、幾分相対化されていくだろう。

このような現象をどうとらえるべきだろうか。私自身は研究者ではなく、常勤の教員という立場でもないので、長年、教壇に立っている大学教員の方々とは異なる感覚であるかもしれない。だが、Twitter を授業に導入することで、大学における学びの空間はオープン化し、不特定多数の外部からの参画を得ることで学びそのものが多様化していくだろう。このような状況を是とするべきだろうか。あるいは、非とするべきだろうか。その答えは私自身、まだ模索段階である。だが、Twitter に象徴されるソーシャルメディアは今後ますます私たちの生活に浸透していくだろう。学生たちにとっても大学生活において、あるいは大学卒業後の職

業生活において、ソーシャルメディアとの関わりは次第に欠かせないものへとなっていくはずだ。そう考えると、現時点では、頭から否定することなく、まずは一度、一人でも多くの大学教員がTwitterの授業への導入に取り組み、試行錯誤の中でこれからの教育環境のあり方を考えていくことが望まれる。私の拙い実践とその報告が、その一助となれば、授業に参加してくれた大妻女子大学の学生たちも大いに報われるだろう。

## 追記

なお、本稿執筆にあたり、Twitterやブログ上で、Twitterの授業への導入事例について尋ねたところ、以下の情報をお知らせいただいた。他にも多数の事例があるが、参考までに挙げておきたい。

- 日本女子大学夏期スクーリング「情報メディアの活用」（2010年度）
  岡野裕行法政大学兼任講師による取り組み。通信教育のスクーリング授業。受講生はすべて社会人の女性。年齢構成は40代前後が中心で、60代から20代まで。全員がTwitterは初利用。講義終了後も、利用を継続している学生がいる。
- 慶應義塾大学「資料検索法」（2010年度）
  市古みどり講師らによる取り組み。授業用の公式アカウントとハッシュタグを設定。
  http://info-literacy.sfc.keio.ac.jp/
  http://twitter.com/sfc_infolit
- 慶應義塾大学sfc noteプロジェクト（2009年度）
  慶應義塾大学湘南藤沢キャンパスにて学生有志によって実施。授業ごとにハッシュタグを決め、学生が自発的に運用。中心となっていた学生の卒業によって終了。
  http://sites.google.com/site/sfcnote/
- 立命館大学「ゲーム製作実習」（2009年度）
  米光一成立命館大学教授による取り組み。約50名の学生が参加。
  http://www.nikkeibp.co.jp/article/nba/20091102/193002/
- 嘉悦大学「情報メディア論」（2009年度）
  松村太郎講師による取り組み。日本における教員主導のTwitter導入の先駆的事例としてテレビ等でも紹介。
  http://www.tarosite.net/2009/05/twitter---class-infomedia.html
- 京都造形芸術大学「情報メディア論」（2010年度）
  中西洋一准教授による取り組み。250名の大講義で実施。
  http://togetter.com/li/14186
  http://twitter.com/gintacat
- 京都外国語大学「情報社会論」（2010年度）
  村上正行准教授の取り組み。「twitterを活用した授業デザインと実践」として、教育システム情報学会第35回全国大会で報告もされている。
  http://www.murakami-lab.org/masayuki/blog/diary/

# ❖学術ソフト・製品

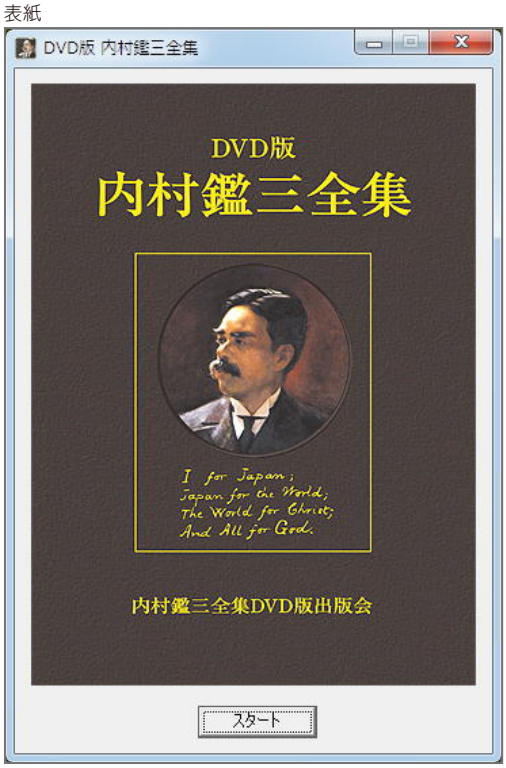
## DVD 版内村鑑三全集

當山　日出夫

### ❖ 00. はじめに

電子書籍とは……このことをめぐって種々に議論があることは承知している。本稿では、昨今の電子書籍の流れなかにあって、いかにも時代錯誤的にさえ見えるかもしれない、『DVD 版内村鑑三全集』について、簡単な紹介をこころみたい。そして、できれば、このDVD 版を考えることによって、近年の電子書籍論で見のがされがちないくつかの論点……知的生産のためのツールとしていかに利用可能であるか……についても、いささか言及することとしたい。

表紙




### ❖ 01. 筆者と DVD 版全集のかかわり

まず、最初に筆者と『DVD 版内村鑑三全集』（以下、適宜 DVD 版全集などと略すことにする）のかかわりについて、記しておきたい。DVD 版全集は、「内村鑑三全集 DVD 版出版会」により、製作・刊行されている。筆者は、直接には、この会のメンバーというわけではない。ただ、この会の事務局を担当し、製作作業の中核をになってきた、斎藤みち氏と、かねてより懇意であったため、折りにふれて相談をうけることがあった。そして、この企画が最終段階にさしかかったとき、いよいよ製品版の完成間近になって、特に文字処理のことについて相談にのって欲しい旨、要請をうけた。このことから、出版会のみならず、実際の製作にあたった精興社などとも、話しをする機会があり、いくつかの助言をしてきた。

そして、最終的には、2009 年 8 月に、製品版として世に出ることになった。また、これを記念して、「DVD 版内村鑑三全集刊行記念シンポジウム」を、翌年 2010 年 7 月 3 日に、国際基督教大学で開催することができたのである。このシンポジウムの企画立案は、筆者が担当した。

このような経緯であるので、直接、DVD 版の製作にたずさわったというわけではない。しかし、その直接の作業の中心にいた斎藤みち氏より、種々の相談をうけ、また、最終的な場面では、精興社との調整にもたちあっている。そしてまた、内村鑑三全集は、近代日本語資料としても貴重な側面をもっており、日本語の歴史的研究を専門とする筆者としても、この意味では、その恩恵に浴するユーザの一人でもある。

このような観点から、以下、DVD 版全集の企画から、実際の製品版の仕様にいたるまでの概略を説明することとしたい。そのうえで、電子書籍の観点から見

た DVD 版全集がいかなる価値を有するものであるか、簡単な考察をくわえるものとしたい。

## ❖ 02. DVD 版の書誌

まず、全集の書誌を簡単に確認しておく。『内村鑑三全集』は、従来、種々に刊行されてきている。1980 年、没後 50 年を記念して、決定版というべき全集が岩波書店より刊行された。全 40 巻という膨大なものであり、完全な編年方式の編集となっている。そしてその後、2001 年に、新発見の資料などを追加して〈第 2 刷〉が刊行された。これについては、岩波書店ではなく、教文館のあつかいとなっている。DVD 版全集は、この〈第 2 刷〉を基本にして、さらに新たに校訂を加えてある。おそらく、内容的には、もっとも信頼のおける版、いわば〈第 3 刷〉ということになるであろう。その入力は、業者委託などではなく、内村鑑三研究に自らたずさわっている研究者が、自分で行っているのである。

## ❖ 03. DVD 版の特徴

以下、DVD 版の機能・特徴について、概略をしるしてみたい。

### ■ (1). 書籍版に忠実であること

DVD 版全集の特徴はなんといっても、書籍版全集に律儀に忠実である、この一点につきるであろう。書籍版に忠実といっても、画像データとして（PDF など）ではない。全文をテキスト入力して、全集の本文を再現してみせているのである。一部、OCR を使ってはいるものの基本的には出版刊のメンバーによる手作業によるテキスト本文入力である。そのうえで、厳格な校正作業を加えてある。

しかも、その版面レイアウトは、全集を忠実に再現している。DVD 版（PDF）に表示される巻・ページ・行は、すべて、もとの岩波版全集を、まったくそのまま再現したものになっている。

ただし、使用した文字（コンピュータ）の関係で……この企画がスタートしたのは、10 年以上にさかのぼる……その当時の文字規格として、「JIS X 0208」（いわゆる 83JIS 系の文字）を採用している。そのため、一部、拡張新字体の採用、たとえば、「冒涜・祈祷」などが残るなど、若干の問題はある。しかし、このようなコンピュータの技術的・制度的な要因をのぞけば、他の文字はすべて原本（全集）に忠実に入力してある。

この、原本（全集）に忠実であるということは、DVD 版全集をつかっている限りは、さほど意味のあることではないかもしれない。しかし、後ほど確認するように、知的生産のツールとして見たときには、引用の典拠が書籍版によって保証されているということは、非常な意味をもつ。

この企画の当初の意図として、全集版の本文をより簡便に読者にとどけるということがあったときいている。その意図がどうであれ、結果として、アカデミックな電子書籍という観点から、きわめて価値の高いものになっているということを、まず、何よりも指摘すべきであろう。

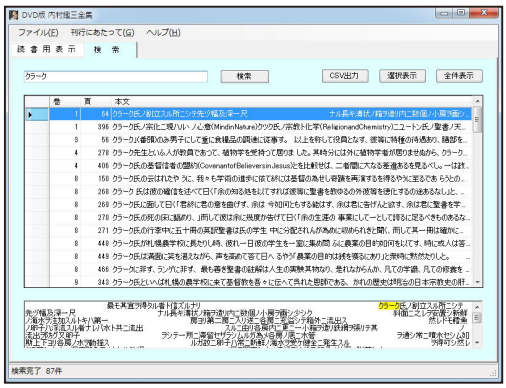
### ■ (2). 全文検索できること

全集の全 40 巻の本文データに対して、全文検索が可能である。これは、まさに、デジタル化された電子書籍、それもテキストとして入力されているが故の機能である。この機能のために、いいかえれば、内村鑑三コンコーダンスを作るということが、この DVD 版全集をつくることの当初の目的の一つでもあった。

「語」の検索ではなく「文字列」の検索として構築されているとはいえ、全集全体を総合的に、しかも、瞬時に検索してくれるということは実にありがたい。この機能のためにこそ、この DVD 版全集は存在するといってもよいであろう。

ただ、若干の難点としては、ルビの検索ができないこと、ページにまたがる文字列の検索ができないこと、など、幾分の問題点をかかえている。しかし、前述のごとく、この DVD 版は、これだけで独立して利用するというよりも、書籍版の全集の存在を前提にして作成してある。多少の問題点は、書籍版の全集を常に併用することで、解決されるべきことであるように筆者には判断される。

また、内村鑑三全集の表記に考慮してのことであるが、聖書名の同一視機能なども、検索システムに独自の工夫としてそなえている。たとえば、「マタイ伝」と検索ウィンドウに入力しても、「馬太」も同時に検索されるという具合である。この機能は、新・旧約聖書について、数十件におよぶ。



検索（クラーク）

### ■（3）. PDF が自由に利用できること

DVD 版全集の本体というべき部分は、書籍版全集の PDF（ワープロ「一太郎」で手作業により入力したものを PDF 化したもの）である。それに、LeafThrough（大日本スクリーン）で閲覧するようにしてある。検索のデータベースは、HiBase（ホロン）。

ここで、何よりも重要なのは、本体の PDF をフリーにしてあることかもしれない。LeafThrough で検索した結果の画面から、マウス右クリックで、該当箇所の PDF 画面表示に切り替えることが即座にできる。そして、その PDF の画面から、自由に本文データを、コピーして利用（引用）することができるのである。

また、DVD 版全集のファイルの構成としても、PDF は独立させてある。そのため、はじめから PDF の方の本文を見るということも可能である。そして、Acrobat などの PDF 用ソフトで、ダイレクトに PDF を検索することもできる。なお、PDF をダイレクトに検索する場合であれば、ルビの文字も検索対象とすることが可能になる。

そして、DVD 版全集全体については、コピープロテクトなどの処理はほどこしてはいない。個人の良識的な利用の範囲内であるならば、デスクトップマシンと、ノートパソコンと、複数にインストールして自在に利用ができるようになっている。

## ※　04. 知的生産のための電子書籍

昨今、電子書籍をめぐって、様々に議論がある。Amazon、Apple、Google、などの外国の企業の動きもあれば、日本国内でも、国立国会図書館の大規模電子図書館構想（長尾真館長）がある。出版社、図書館、印刷業、IT 企業、それぞれの思惑が交錯して混沌とした状態にあるといってよい。ここでは、えてして忘れられがちな、利用者の観点、特に、アカデミックな利用（研究者の利用）というところから、DVD 版全集を再考してみたい。

近年の電子書籍のうごきからすれば、DVD というパッケージになっている内村鑑三 DVD 版全集は、もはや古色蒼然とした趣さえ感じる。時代遅れの感はいなめない。

だが、はたして本当にそうだろうか。特に東洋古典学のような人文学研究の立場からした場合、電子書籍はどうあるべきなのか、たちどまって考えてみる必要はないであろうか。

このとき、

⑴. 資料としての電子書籍
⑵. 論文としての電子書籍

二つの方向があるにちがいない。ここでは、前者、⑴資料としての電子書籍を考えてみたい。

では、資料としての電子書籍、旧来の言い方をするならば、資料のデジタル化ということになるであろうが、これは、どのようにあるべきであろうか。ここで、これまで述べてきた DVD 版内村鑑三全集との関連で、論点を整理するならば、

⑴. 正確であること
⑵. 本文校訂の版の管理が明確であること
⑶. 引用・出典の典拠として所在情報が明確であること
⑷. 自由な検索が可能であること
⑸. 引用などの知的再利用が自由にできること

これらのことが要件として思い浮かぶことである。そして、このような観点から見たとき、DVD 版全集こそが、これらの要件を満たしていることに、改めて気づく。逆に、現在、世の中で取りざたされているところの電子書籍は、これらの要件を満たすものであろうか。

具体的に述べるならば、iPad 版、あるいは、Kindle 版の、内村鑑三全集があったとして、では、それは、

思想史研究・近代日本語史研究に、知的生産のツールとして利用できるものであろうか。

筆者は、ここで別に iPad や Kindle をおとしめる意図はない。ただ、あまりに電子書籍礼賛とばかりはいかないであろう、という側面を指摘しておきたいのである。俗に「読み書き算盤」という。リテラシの基本をさす。ここで考えてみるべきは、「読む」と「書く」とはその環境が連続して一体化したものでなければならない、という側面があることである。「読む」ということだけを考えて、「書く」（知的生産）のことを忘れてしまってはいけないであろう。

だからといって、旧態依然たるパソコン環境がそれにふさわしいというわけでもない。また、新しい電子書籍端末が、まだ、知的生産に十分に対応しているともいいがたい。これは、これからの、大きな課題という他はない。知的生産のシステム全体を視野にいれた、電子書籍がこれから望まれるのである、ということは言ってもよいであろう。

## ❖ 05. まとめ

以上、DVD 版内村鑑三全集について、その概略・機能について紹介してきた。また、一見すると古色蒼然とした DVD 版こそが、かえって、内村鑑三研究という知的生産のためには役立つシステムを提供してくれるのであるということも確認してみた。

DVD 版内村鑑三全集を、単に研究資料のデジタル化としてのみ見るのではなく、電子書籍というものとならべてみたとき、それは、単なるデジタル資料ではなく、これから進むべき電子書籍のあるべき一つの方向をしめしてはいないであろうか。

### 参考文献


- 當山日出夫（2009）.「『内村鑑三全集』デジタル版の文字処理について」.『東洋学へのコンピュータ利用 第 20 回セミナー』. 京都大学人文科学研究所附属漢字情報研究センター
- 當山日出夫（2010）.「DVD 版『内村鑑三全集』について」.『東洋学へのコンピュータ利用　第 21 回セミナー』. 京都大学人文科学研究所附属漢字情報研究センター
- 當山日出夫（2010）.「DVD 版内村鑑三全集と学術資料電子出版」. 情報処理学会研究報告（CH-86-1）. 於大阪大学
- 當山日出夫（2010）.「電子書籍と知的生産」.『丸善ライブラリーニュース』. 第 11 号


### 付記

2010 年 7 月 3 日（国際基督教大学）でおこなわれた、DVD 版内村鑑三全集刊行記念シンポジウムのプログラムは以下のとおりである。概略のみ記載する。

- 開会挨拶（田村光三．出版会）
- 会場校挨拶（千葉眞．ICU）
- 開会趣旨説明（當山日出夫．立命館大学）
- DVD 版全集の経緯（斎藤みち．出版会）
- 近年の内村鑑三研究の動向と DVD 版（大山綱夫．『内村鑑三研究』編集委員）
- DVD 版全集の実際の利用について（柴田真希都．東京大学大学院）
- DVD 版作成の技術的課題（木村素行．精興社）
- デジタル書籍の最新動向から考える内村鑑三全集 DVD 版の可能性と課題（岡本真．ARG）
- 全体討論（総合司会　當山日出夫．立命館大学）
- 閉会挨拶（斎藤みち．出版会）

### 謝辞

困難な大事業にたちむかってこられた、内村鑑三全集 DVD 版出版会の方々に、深甚なる謝意を表するものである。

### 補記

なお、現在の DVD 版の作動環境は、Windows XP（SP2）以上となっている。Windows VISTA および、Windows 7 でも、正常に動くことは確認してある。しかし、32 ビットに限定される。現段階では、64 ビットには対応していない。（この件は、精興社にといあわせて確認済み。）

# 講談社『中日辞典』第三版CD-ROM

田邊　鉄

## ❖ CD-ROM とネットで辞書を拡張

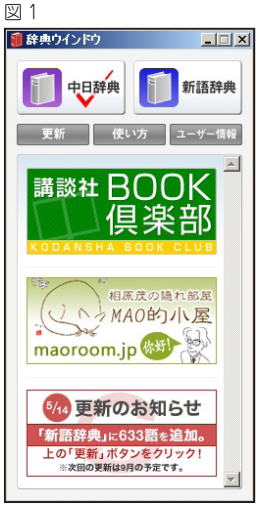
講談社『中日辞典』第三版が刊行された。第二版から約 8 年、中国語辞書の改版ペースとしては、早い方だろう。今回の改訂は、新語や慣用語を中心とした約 7500 語の増補がウリであるが、それ以上に目を引くのは、全文を収めた CD-ROM が付属していることであろう。この CD-ROM から辞書ソフトをインストールすることで、ネットを通じて『現代中国語新語辞典』全文をダウンロードすることができる。また、新語・流行語を中心とした辞書の更新データをダウンロードすることもできる。

同じ講談社の『日中辞典』には、既に CD-ROM が付属しており、また『中日』第二版ともども市販の IC 辞書に搭載されているので、電子化自体はさしたるニュースではない。ただ、ネットを利用して辞書の内容自体を増補していくシステムは、単なる付録 CD-ROM の域を超えた可能性を示唆している。ここでは、この CD-ROM 版についてレビューする。

## ❖ 体裁〜インストール

CD-ROM は語学教科書の付録のように、セロハン袋に入ったものが裏見返しの遊び紙に直接貼り付けられている。袋を開けたあと、赤色のヒモがぶら下がってみっともない。それだけでなく、取り出した CD-ROM の保存にも困る。もとの袋に入れておくと、辞書を逆さに持った時に落ちはしないか心配だし、第一辞書をひくときに固い板が挟まっているとページをぱらぱらめくれない。自分でケースに入れてやればいいのだろうが、整理が悪いと紛失しそうだ。函に工夫して収納スペースを設けるなどできないものか。

図 1




CD-ROM のインストール方法や使い方についての説明文も、裏見返しに直接印刷されている。最初のインストール時以外はほとんど参照することもないのだから、しおりにして挟み込む、CD-ROM の袋に印刷する、など考えてほしかった。

学習用辞書の装丁は、使いやすいというだけでなく、初学者の外国語学習に対するファンタジーに応えるものであってほしい、と思うのは私が古くさい外国語学習観しか持ち合わせていないせいか。むやみと重厚にする必要はないが、あまりカジュアルに過ぎると学習意欲が萎えそうに思う。

CD-ROM 版の辞書を利用するためには、「辞典ウィンドウ」という専用ランチャーソフトをインストールする。辞書のデータは全てハードディスクにコピーされるので、起動に CD-ROM は不要だ。後述の「更新機能」を利用するためには、インストール後「ライセンス認証」手続が必要になる。ライセンス認証は、1 枚の CD-ROM につき PC3 台まで可能になっている。ライセンス認証を解除すれば、別の PC で更新機能が使えるようになるが、元の PC 上にある更新済みのデータはそのまま利用を続けることができるので、事実上際限なくインストール、利用することが可能だ。

## ❖ 「辞書ウィンドウ」の操作

「辞書ウィンドウ」は、上部に操作ボタン、下部に編者のブログや講談社の辞書ページ等、関連サイトへのリンクバナーが置かれている（図 1）。

操作ボタンは「中日辞典起動」・「現代中国語新語辞典起動」・「辞典データ更新」・「ユーザー情報表示」・「使い方」の 5 つである。インストール先のフォルダには、中日辞典を起動するためとおぼしき実行ファイルがあるのだが、中日辞典、新語辞典とも、直接起動する事はできず、必ず「辞書ウィンドウ」から操作しなければならない。

「更新」ボタンを押すと、アップデートの確認と

インストールができる。講談社では 2012 年まで年 3 回程度不定期に更新データを提供する、としている。既に 5 月に一度データ更新が行われ、新語が 633 語追加されたそうだ。最初の更新時には、「オマケ」についてくる『現代中国語新語辞典』のダウンロードとインストールも行われる。なお、更新内容の詳細を知ることはできない。「更新のお知らせ」というバナーをクリックしても、リンク先は講談社の販売サイトになっている（2010 年 9 月現在。おまけにそこで紹介されている『中日辞典』は第二版である）。2012 年までに 5000 語の増補を目指すとしているが、それは同じ辞書を使っていながら、最大 5000 語の差がある複数のバージョンが生じる可能性がある、ということである。授業で学生に使わせる場合は注意が必要だろう。増補内容などの詳細を一覧できればありがたいところだ。

図 2 は辞典を起動したところである。検索は「見出し語検索」・「全文検索」・「メニュー検索」と、IC 辞書などでおなじみの検索方法を一通り用意してあり、操作に迷うようなところはないだろう。

見出し語検索は前方一致・後方一致・完全一致を選べる。ワイルドカードや正規表現を用いた高度な検索こそできないが、検索内容・速度とも実用レベルだと思う。なお、ピンインでの検索もできる。声調記号付きピンインと、声調なし（アルファベットのみ）に対応する。

全文検索は、検索結果の該当語の部分が黄色でハイライト表示されて見やすい。絞り込み検索ボタンを使えば and 検索も可能だ。単語を一つ入れて検索ボタン、もう一つ入れてまた検索ボタン、というのはちょっと面倒だが、使える機能ではある。

全文検索は単に本文の中で一致する部分を検索するだけなので、入力するのは語に限らない。たとえば「是学生」を検索する、といったことも可能である。これを利用した便利な使い方として「[介]」のように品詞を全角の〝[　]〟で囲んで全文検索をかけると、該当する品詞の語を全てピックアップしてくれる。「[成]」とすると、四字熟語がずらりと並ぶ。絞り込み検索と組み合わせると「白を含む成語」のような検索ができるので、なかなか使いでがあると思う。なお、全文検索ではピンインは使えない。

講談社『中日辞典』のウリのひとつは、巻末につけられた実用的な固有名詞リストである。CD-ROM に

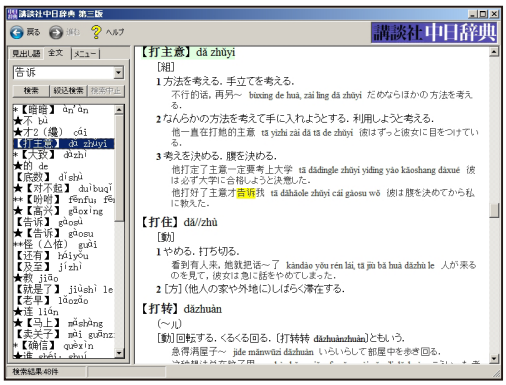



図 2

もももちろん取り入れられている。メニュー検索では、ブランド名、人名、地名などを、日本語表記の 50 音から参照することができるようになっている。『中日辞典』なのに、日本語から引くのをデフォルトにした理由については、編者が序文で述べているとおり、日本人がこれらの固有名詞を使う時には「圧倒的に日本語から発想する」からであり、実用を重んじた結果といえる。もっとも、身もふたもないことを言ってしまえば、電子辞書なのだからまとめて本文に放り込んでしまい、全文検索で探す方が手っ取り早いだろう。

## ◈　「オマケ」はどちらか

1998 年に小学館が「中日・日中統合辞典」として CD-ROM 辞書を発売した。その後も中国語入力 IME のオマケとして辞書が搭載されるなど、パソコンで使える中国語辞書はこれまでにもあった。講談社『中日辞典』のユニークな点は、全文を収めた CD-ROM を紙の辞書の「オマケ」としたことと、その「オマケ」を「CD-ROM 版」として単独で発売したことである。

『中日辞典』第三版は 7980 円である。CD-ROM がついていなかった第二版は 6825 円で、差額はわずか 1000 円強であり、この数字だけ見れば辞書の価格に対して CD-ROM はタダみたいなもの、まさに「オマケ」と考えられているように見える。

ところが、同時発売された CD-ROM 単体版は 4935 円である。「消費者目線」で見るなら、7980－4935＝3045 円が「紙版辞書だけの価格」ということになる。「紙版の方がオマケになっている」とまで言うつもりはないが、辞書の「本体」は紙や CD-ROM といった物理的なメディアではなく、データそ

のものであることを出版社自ら認め、来るべき電子書籍時代に備えているようにも見える。

ちなみに編者の相原氏の序文の中に、CD-ROM に関する言及は全くない。全く別のプロジェクトとして制作されたからか、それとも、言わずもがな、ということなのか、言いたくないのか。機会があればご本人に伺いたいものである。

# 『四部叢刊』と書同文社のオンラインデータベース

千田　大介

## ◈ CD-ROM 版の問題点

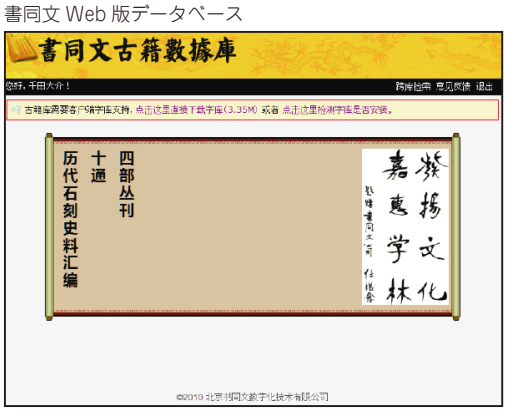
電子版『四部叢刊』は、中国学電子化の一つの画期となった電子版『四庫全書』の開発元である書同文社が開発した、大規模古典叢書全文データベースである。『四部叢刊』は『四庫全書』に比して収録版本の質が高いことから、特に古典・古代研究においては非常に有用なツールとなっている。

しかし発売から十年近くが経過して、電子版『四部叢刊』にはいくつかの問題が生じている。その最大のものが、Windows Vista 以降で動作できない、というものである。

本誌バックナンバーのレビューで触れたように、書同文社は近年、多角化を図っており、文献デジタル化に関しては技術開発部門を縮小し、旧来の CJK+（＝CJK 統合漢字＋ Ext.A ＋外字）技術による文献デジタル化サービスのみを提供している。近頃、Vista/7 対応パッチが出来たというが（筆者は試用していない）、あまりに時間がかかりすぎており、かつ、今後の OS アップグレードに際して、再び同様の問題が発生する可能性もある。

OS に依存したアプリケーションソフトとして提供されている以上、こうした問題の発生は不可避であると言えよう。

書同文 Web 版データベース




## ◈ Web 版データベース

書同文社では、こうした状況を踏まえ、またオンライン化の流れを受けて、同社が開発した電子版古典文献全文データベース製品の Web 版の提供を始めている。Web 版であれば OS のバージョンによる制限を受けないし、データベースにアクセスして認証して使用するため、ソフトをインストールすることなく手軽にさまざまな PC から利用できるメリットがある。

書同文社の Web 版データベース「書同文古籍数拠庫」は、以下の叢書を収録する。

- 四部叢刊
- 四部叢刊 09 増補版
- 歴代石刻
- 十通

このうち、「四部叢刊 09 増補版」は、張元斉の校勘記と『四部備用』を追加したバージョンである。他は、従来から CD-ROM 版が発売されていた。このほか明清実録や檔案などを収録した「明清古籍数拠庫」も提供予定であるという。

Web 版で特筆すべきは、インターフェイスの改善である。従来の『四庫全書』・『四部叢刊』等、書同文系のデータベースでは、検索結果画面がヒットした文

献名・巻数などの一覧になっており、リンクをクリックしてヒット箇所の本文を一々確認しなくてはならず、効率が悪かった。

Web 版では、検索結果画面に、ヒット箇所の本文が KWIC 風に表示されるようになった。1 ページに表示できる件数には 12 件の制限があるものの、ヒット箇所の本文をあらかじめ確認できるので、検索効率は格段に上がった。検索結果ページを Excel 等にコピーして、そのまま分析に使うことも出来よう。

更に、これまでは『四部叢刊』と『歴代石刻』・『十通』などのプログラムをそれぞれ起動して、別々に検索する必要があったが、Web 版では、契約している複数のデータベースを横断検索できる。ただし、検索結果は、各データベースごとに表示することになる。検索結果をクリックすると、各叢書のデータベースが開き、本文の画像・テキストを閲覧できる。テキストはもちろん、コピーすることもできる。

書同文の強力な異体字テーブルは健在であり、簡体字・繁体字・常用漢字体、さらには同音字なども一括検索できる。文字コードには相変わらず CJK+ が使われており、同サイトで配布している「HT_CJK+.TTF」の導入が必要になる。また、CJK+ は、外字を用いた汎用性のない独自規格であるので、本文テキストをコピーして利用する時などに問題が発生する可能性があるので、注意が必要である。

## ❖ 日本での発売

書同文社では上記各データベースを中国国内（台湾・香港・マカオを含む）の大学や研究機関に提供している。価格は以下の通り。

- ●四部叢刊：25,000 元
- ●四部備要：25,000 元
- ●歴代石刻：18,000 元
- ●十通：18,000 元

『四部叢刊』で日本円 35 万円ほど、この価格で無期限使用ライセンスが提供される。

書同文社では、日本での各種データベース発売に向けて、現在、準備を進めているとのことである。日本国内での販売価格がいくらになるかわからないが、中国国内と大差ない価格で提供されるのであれば、内外の学術論文データベースが年間百万円を超える値段であるのと比して、極めてリーズナブルであると言えよう。正式な国内発売が待望される。

『四部叢刊』検索結果一覧

上図：『四部叢刊』版本画面
下図：『四部叢刊』テキスト画面

# 『中日大辞典』第三版

愛知大学中日大辞典編纂所編　大修館書店　2010 年 3 月　ISBN978-4-469-01281-1　8,600 円＋税

金子　眞也

1968 年に『中日大辞典』が初めて刊行された当時、親字は 7876 字（併記されている繁体字・異体字を入れると 11,195 字）、1987 年の『増訂第二版』が 8812 字（同 13,166 字）であった。今回の『第三版』は 8921 字（同 13,840 字）であるから、親字の数は前回より若干増えている。

愛知大学中日大辞典編纂所ウェブサイト（**http://leo.aichi-u.ac.jp/~jiten/**）と『第三版』「凡例」によると、2009 年 8 月 12 日に、中国の教育部・国家語言文字工作委員会が発表した《通用规范汉字表》の中に、本辞典に未収録の漢字があり、詳細についてはウェブサイトで対応するとのことである。《通用规范汉字表》じたいが〝征求意见稿〟という位置づけであることを考えると、《通用规范汉字表》についてウェブサイト対応としたことは、妥当な判断であると評者は考える。

『第三版』の特徴は、ひとことで言うと「利便性の増大」にある。初心者にもなじみやすい作りに変わった。項目に分ければ五つ。

まず、第一に、「メリハリがあって内容を追いやすいレイアウト」があげられよう。これは、辞書をつい読み込んでしまう老眼の評者にとっても、ありがたいことだ。

第二に、「発音ごとに別々に立てた見出し」をあげよう。“得”を引く場合、〝dé〟は〝dé〟、〝děi〟は〝děi〟で引く。好みが分かれるかもしれないが、初心者にはこの方が便利だ。

第三は、「調べやすくするための工夫」である。たとえば〝不好意思〟は単独で見出し項目としてあがっているほか、〝好意思〟で引いた場合でも、語釈と例文を読めば、〝不好意思〟の意味がたどれるよう作られている。ちなみに、本書の“不好意思”のピンイン表記は、bù と hǎoyìsi の二つに分けない bùhǎoyìsi で、2004 年の《现代汉语规范词典》、2010 年の同第二版と同じ表記になっている。

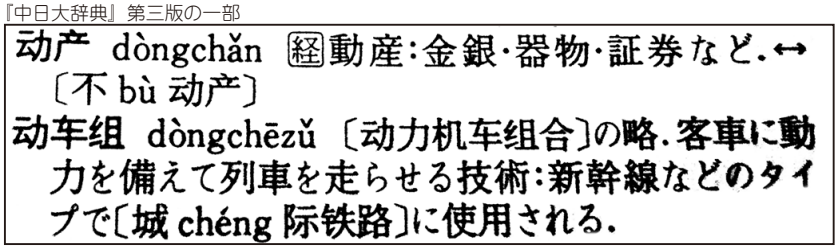
第四は、「財布にやさしい値段」である。税別 8,600 円で、学習者からみたら決して安いとはいえないものの、20 年以上前の『増訂第二版』がつい先日まで同じ値段で売られていたことを考えると、相対的に言ってずいぶん安くなった。評者などは、「『中日大辞典』も安くなったものだなぁ」という感慨を禁じ得ない。

第五は、「中国の現状への対応」である。〝公交车〟〝动车组〟〝八卦〟（ゴシップ）など新語・俗義はもちろん、〝和谐〟を引くと〝和谐社会〟が用例に出てくるなど、なかなか頑張っている。

レイアウトを見やすくし語彙を増やした分、どこか減らさなければならないのは、自然な道理だ。『第三版』から外された例として、〝二位〟（二人を表す尊称）をあげておく。こういう外国人に分かりにくい敬意の表現が省かれたのは残念である。

最後に、ひとこと。紙の辞書には「読み進む楽しさ」があり、電子媒体には「例文検索の便」がある。『第三版』が早期に電子化され、紙と電子媒体の両方で『第三版』が楽しめる日が来ることを、評者は強く望んでいる。

『中日大辞典』第三版の一部



# 『現代漢語規範詞典』第２版

李行健主编　外语教学与研究出版社　2010 年 5 月　ISBN978-7-5600-9518-9　83.00 元

金子　眞也

2004 年に《现代汉语规范词典》が刊行されたとき、新鮮だったのは「品詞を明示していること」、「実用性を重視していること」、「ピンインの扱い方が他の辞典とやや異なること」であった。《第２版》でも、基本的に方向性は同じである。

本辞典の親字の選び方は、〈凡例〉を読む限り、《第１版》《第２版》いずれも、1988 年発表の《现代汉语通用字表》をベースにしている。《第１版》は《通用字表》に収録されていない文字も含めて 13,000 字、《第２版》の方は、よく見かける文字、科学技術や古典教育に使用する文字等、繁体字・異体字も含めて 12,000 余字となっている。つまり、改訂の結果、親字の数自体は減っている。

《第１版》と《第２版》では品詞の分け方が一部異なる。《第２版》では〝形容词〟から〝区别词〟を独立させていて、この点が大きな違いとなっている。

ちなみに、〝区别词〟という語自体が《第１版》には収められていない。《第２版》で追加された〝区别词〟の語釈の部分を抜き書きすると、〝表示事物特征和分类、一般只用来修饰名词及名词性短语的词。区别词不能作谓语；也不能作主语、宾语和补语（少量可兼作状语）〟(以下省略)となっている。

《第１版》の凡例には、〝形容词〟の説明として〝形容词（包括非谓形容词）〟と断ってあったが、《第２版》では、〝区别词〟を独立させた結果として、〝形容词〟の説明から「述語にならない形容詞を含む」という注記が削除されている。一般の読者にとってはどうでもよいことかもしれないし、意見の相違もありうるが、これを本書のもつ「こだわり」として考えると、興味深いものがある。

実用性の重視は、《第１版》以来の〝小手提示〟(指さしマークのアドバイス) によく現れている。〝提示〟の例をいくつかあげよう。〝有求必应〟には「この〝应〟は〝yīng〟とは読まない」という〝提示〟、〝感染〟には「〝感受〟とは異なる。〝感染〟は他人に及ぶが、〝感受〟は自分だけに限られる」という〝提示〟、〝戍〟の字には「〝戍〟は別の字」という〝提示〟が、それぞれつけられている。

「こんなことまで書く必要があるのだろうか」と思う反面、学習辞典として本書を見るならば、面白い試みだ。

新語への対応ももちろんある。〝动车组〟〝低碳〟(二酸化炭素等の温室化ガス排出量の低い＝〝区别词〟) などを採ったことで新鮮味は感じるものの、〝公交〟はあるが〝公交车〟は用例にない、〝智能〟は「知恵と才能」だけで、英語の〝smart〟や〝intelligent〟に相当する意味は載っていないなど、採りあげ方にややばらつきがあるように感じられる。

本書は中国の読者を対象に編まれた辞典であるから、外国人が使う場合に注意すべき点がある。
〝轻声〟と〝儿化〟は、必ずそうなる場合のみ表記してある。どちらでもよい場合は、何も記載がない。これは、中国人にとってはなくてもよい情報だが、外国人には必要な情報であろう。

最後にピンイン表記について記す。〝一衣带水 yīyīdàishuǐ〟、〝有意思 yǒuyìsi〟など、他の辞典が「分かち書き」で書くような語を、続けて綴る傾向が、本書には見られる。

1996 年に出た〝汉语拼音正词法基本规则〟に反するように一見みえるが、よく考えてみると、〝基本規則〟の解釈の仕方次第で、おさまってしまう範囲内にあると評者は考える。

本書は教員や中級以上の学習者が二冊目、三冊目の辞書として手元に置くのにふさわしい読んで楽しい辞典である。

# 漢字文献情報処理研究会彙報

## 2009.10〜2010.9

2009 年 10 月 1 日
　会誌『漢字文献情報処理研究』第 10 号出版。
2009 年 12 月 20 日
　第 12 回大会、2009 年度総会開催。
2010 年 1 月 30 日
　文字研究会「第 4 回ワークショップ：文字 ── 言語生活のなかの文字」後援。
2010 年 8 月 11 日
　文字研究会「第 5 回ワークショップ：文字 ── 「現実」から見た改定常用漢字表」後援。

## 第12回大会・2009年度総会

日時：2009 年 12 月 20 日 ㈰
会場：花園大学　拈花館 202 教室

### ■ 第 12 回大会（13:00 〜 17:15）

1. 研究報告（13:00 〜 15:00）
- 「ア行の /e/・ヤ行の /je/ を表わす仮名文字の標準化をめぐる諸問題」
  高田智和（国立国語研究所）
- 「検索エンジン（Lucene）による中国語用例抽出最適化―北京大学中国語語料庫を利用して」
  詹善斌（早稲田大学）、小川利康（早稲田大学）
- 「IDS による漢字情報処理」
  川幡太一

2. 企画討論会（15:20 〜 17:00）：
- 「Windows 7 時代の〝電脳中国学〟と〝人文系情報処理テキスト〟」
  提題者：千田大介（慶應義塾大学）
  コメンテーター：師茂樹（花園大学）・山田崇仁（立命館大学）

### ■ 総会（17:00 〜 17:15）

- 2009 年度事業報告、会計報告（会計監査：清原文代・當山日出夫）
- 執行部改選
  代表：師茂樹
  副代表（兼会誌編集局長）：山田崇仁
  副代表（兼サーバ管理担当）：上地宏一
  幹事（会計・名簿担当）：小島浩之
  幹事：佐藤仁史
  幹事：田邉鉄
  幹事：千田大介
  幹事：二階堂善弘
- アルバイトの雇用に関して
- 投稿規定の改正について
- 2010 年度事業計画、予算案承認
- 新入会員勧誘・会誌販売促進のお願い

## 2010年公開講座

### ■ 2010 年度公開討論会

題目：電子出版の動向と諸問題
日時：2010 年 7 月 10 日㈯ 13:00 〜 17:30
会場：慶應大阪リバーサイドキャンパス Room1（大阪市福島区）

1. 報告の部（13:00 〜 15:30）
- 田代真人氏（㈱メディア・ナレッジ代表、㈱アゴラブックス取締役）
- 守岡知彦氏（京都大学人文科学研究所助教）
- 石岡克俊氏（慶応義塾大学産業研究所准教授）

2. ディスカッション（16:00 〜 17:30）

## 2010年後援講座

### ■ 後援講座

題目：第 4 回ワークショップ：文字 ── 言語生活のなかの文字
日時：2010 年 1 月 30 日㈯ 13:00 〜
会場：国立国語研究所　多目的室
- 當山日出夫「言語生活の視点からの文字 ── 景観文字研究の課題 ──」

- 鑓水兼貴「「略字・俗字」使用における場面差・属性差」
- 岡墻裕剛「『文字のしるべ』に見る明治期の外国人の漢字使用」
- 杉山元康「『活字離れ』論の実態と、私たちの触れている『カツジ』」
- 小形克宏「言語生活から見た絵文字の Unicode 提案」
- 全体討論

題目：第 5 回ワークショップ：文字
　　――「現実」から見た改定常用漢字表
日時：2010 年 8 月 11 日㈬ 13:00 ～
会場：東洋大学白山キャンパス 6 号館 2 階 6209 教室

- 趣旨説明
- 小形克宏「改定常用漢字表とは何か？」

第 1 部 「漢字調査」から考える
- 漢字文献情報処理研究会「文化庁の出現頻度数調査から改定常用漢字を考える」
- 萩原正人「ウェブ上における使用実態統計から改定常用漢字を考える」

第 2 部 「現場」から考える
- 前川孝志「国語教育の現場から改定常用漢字表を考へる」
- 比留間直和「新聞表記と常用漢字改定」
- 関口正裕「改定常用漢字表で情報システムはどうなるのだろう？」

第 3 部　みんなで考える
- パネルディスカッション

# 著者紹介

秋山　陽一郎（あきやま　よういちろう）

1975 年生。立命館大学大学院文学研究科博士後期課程修了後、2004 年度から 2 年間、京都大学人文科学研究所の 21 世紀 COE プログラムに技術補佐員・特別研究員として携わる。2006 年花園大学非常勤講師。専門は中国古代文献学。

安藤　一博（あんどう　かずひろ）

1976 年岡山県生。国立国会図書館所属。筑波大学第一学群人文学類卒業。大学時代は中国近代史を専攻。国立国会図書館に入館後、7 年間程アジア言語に関係する部署（アジア資料課→アジア情報課）に配属され、中国語資料を扱う仕事をしていた。その関係で中国語圏の出版動向に関心を持つ。
**http://kzakza.wordpress.com/**

石岡　克俊（いしおか　かつとし）

1970 年北海道生。専門は、経済法、知的財産法、消費者法。最近は、学術情報のディジタル化がもたらす著作権法或いは独占禁止法上の課題について関心を持っている。特に著作権に関する最近の論文に「『校訂』の著作権法における位置」『慶應義塾女子高等学校研究紀要』26 号がある。

岡本　真（おかもと　まこと）

1973 年東京都生。国際基督教大学卒業後、編集者、校正者、翻訳家等を経て、1999 年から 2009 年までヤフー株式会社にて Yahoo! 知恵袋等の企画・設計・運用に従事。2009 年同社退社後、1998 年に創刊したメールマガジン ACADEMIC RESOURCE GUIDE を母体に、アカデミック・リソース・ガイド株式会社を設立。同社の代表取締役、プロデューサーとして、インターネットの学術利用を促進する業務に従事。近著の『ブックビジネス 2.0』（実業之日本社、2010 年）がある。

小川　利康（おがわ　としやす）

1963 年東京生。早稲田大学商学学術院教授。専攻は現代中国文学。単著「周作人・松枝茂夫往来書簡」（戦前篇⑴⑵⑶及び戦後篇、『文化論集』30－33 号）など。
**http://www.f.waseda.jp/ogawat/**

金子　眞也（かねこ　しんや）

1955 年東京生。龍谷大学法学部教授。
**http://www.huaxia-info.com/**

上地　宏一（かみち　こういち）

1976 年生。大東文化大学外国語学部講師。コンピュータにおける多漢字処理、特に漢字字形生成エンジンや異体字データベースなどに興味を持っている。自身が構築・運営する漢字字形 Web データベース「グリフウィキ **http://glyphwiki.org/**」のさらなる発展を目指して試行錯誤中である。

川幡　太一（かわばた　たいち）

現在、某通信会社の研究所勤務。マークアップなどのテキスト情報処理システム、漢字・多言語の表示・編集処理や、ネットワーク分散オブジェクトシステムなどに関心を持つ。

清原　文代（きよはら　ふみよ）

1964 年大阪府生。大阪市立大学大学院文学研究科後期博士課程単位取得退学。現在は大阪府立大学総合教育研究機構准教授。近年のテーマはモバイルラーニング、Podcast による中国語教材の配信、PDF や EPUB による教材の開発など。iPhone と iPad、電子辞書が好物。
清原の Web サイト（更新停滞中）
**http://www.las.osakafu-u.ac.jp/~kiyohara/**
twitter：**http://twitter.com/wendai_Q**

小島　浩之（こじま　ひろゆき）

1971 年岐阜県生。東京大学大学院経済学研究科助手、経済学部資料室長代理。専門は中国史だが、最近は歴史資料の保存に関する研究で忙しい。主要論文に「東京大学総合図書館所蔵鴎外文庫『明代勅

命』管見」『漢字文献情報処理研究』10、「資料保存の考え方－現状と課題」『情報の科学と技術』60−2 など。

### 後藤　真（ごとう　まこと）

1976 年生。花園大学専任講師。博士（文学）。最近は知識情報と史料の結びつけ、構造化に興味を持っている。特に日本の奈良時代の知識情報を SOMODA の更新・発展とともに構造化する工夫を行い、プロジェクト推進に邁進する。また上田貞治郎古写真コレクションも、完成を目指し作業を続けている。より新しい形での「人文情報学」の可能性に資することができればと苦悶の日々を送っている。

### 詹　善斌（せん　ぜんひん）

1975 年台湾台中生。2005 年 9 月から 2008 年 6 月まで Yahoo! 台湾分社検索エンジン開発部配属。現在早稲田大学基幹理工学大学院博士後期課程二年生。専攻は特定言語（日本語、中国語、韓国語）クローラの性能改善とウェブで説明なし画像の概念推薦システムの研究。

### 田代　真人（たしろ　まさと）

1963 年福岡県生。九州大学工学部機械工学科（自動制御専攻）卒業。朝日新聞社、学習研究社、ダイヤモンド社と長年編集者として活動。近著に『電子書籍元年 iPad& キンドルで本と出版業界は激変するか？』がある。また、メール悩み相談マイ・カウンセラー代表として社会的企業も運営。
**http://m.mycounselor.jp/**

### 田邉　鉄（たなべ　てつ）

1963 年生。大阪外国語大学外国語学研究科東アジア語学専攻修士課程修了。北海道大学情報基盤センター准教授。専門は中国語教育・教育工学・デジタルコンテンツ・マルチメディア表現。

### 千田　大介（ちだ　だいすけ）

1968 年生。慶應義塾大学経済学部教授。専門は宋代民間芸能から現代のメイド喫茶・AV 受容まで、ここ一千年くらいの中国サブカル。**http://wagang.econ.hc.keio.ac.jp/**

### 當山　日出夫（とうやま　ひでお）

1955 年京都府出身。慶應義塾大学・大学院。本来の専門は、国語学・訓点語学。ただ、今は、「人文情報学」の方面にかたむきつつある。立命館大学グローバル COE「日本文化デジタル・ヒューマニティーズ拠点」客員研究員。
個人のブログ「やまもも書斎記」
**http://yamamomo.asablo.jp/blog/**

### 守岡　知彦（もりおか　ともひこ）

1969 年生。北陸先端科学技術大学院大学情報科学研究科修了。博士（情報科学）。京都大学人文科学研究所附属東アジア人文情報学研究センター助教。

### 師　茂樹（もろ　しげき）

1972 年生。花園大学准教授。「「公共の記憶」としての電子書籍」（『ユリイカ』2010 年 8 月号、青土社、2010 年 7 月）、「携帯電話の絵文字における semantics の問題」（『東洋学へのコンピュータ利用第 21 回研究セミナー』、2010 年 3 月）、後藤真・田中正流・師茂樹『情報歴史学入門』（金壽堂出版、2009 年）など。Twitter ID: **@moroshigeki**

### 安岡　孝一（やすおか　こういち）

1965 年大阪府生。京都大学人文科学研究所附属東アジア人文情報学研究センター准教授。京都大学博士（工学）。専門は人文情報学。著書に『キーボード配列 QWERTY の謎』(NTT 出版 )、『文字符号の歴史―欧米と日本編―』( 共立出版 ) などがある。

### 矢野　正隆（やの　まさたか）

1972 年福岡県生。京都大学大学院文学研究科修了。ベトナム前近代史。現在、東京大学大学院経済学研究科特任研究員として同学部資料室に勤務。近業は『ベトナム文化人類学文献解題』( 編集担当 )、「資料保存」(『図書館界』61-5) など。

### 山田　崇仁（やまだ　たかひと）

1970 年愛知県生。立命館大学院文学研究科修了。博士（文学）。滋賀大学・立命館大学非常勤講師。専門は先秦文献の成書時期の解明、先秦～前漢における語用法の変遷。**http://www.shuiren.org/**

## 編集後記

『漢字文献情報処理研究』第 11 号をお届けする。

今年は編集作業が大幅に遅れてしまった。早々に原稿を出していただいた執筆陣の方々や、好文出版の尾方社長には多大なご迷惑をかけしてしまった。これもひとえに編集子の怠慢である。この場を借りてお詫びしたい。

本号から表紙のデザインを変更した。これは今まで使用していた図案の中で本誌表紙に使用可能なものが無くなってしまったとためである。新しい表紙デザインについて、編集子は思ったほど変化がなかったかなと感じたが、読者諸賢はどのような感想をお持ちだろうか。

本号は電子書籍絡みの記事が多くを占めることになった。そういう意味でも今年は電子書籍の年だと実感している。さて、来年はどのような内容が多くを占めるのか。今から楽しみである。

本誌の刊行は、好文出版の尾方社長を始め、漢字文献情報処理研究会会員各位や原稿執筆陣からの有形無形の援助のたまものである。

末筆ではあるが、篤く御礼申し上げたい。(♬)


**漢字文献情報處理研究　第 11 号**

発行日　2010 年 10 月 1 日

定価　本体 2,000 円＋税



編集　© 漢字文献情報処理研究会
http://www.jaet.gr.jp/



編集委員　○山田　崇仁　金子　眞也
上地　宏一　小島　浩之
佐藤　仁史　田邉　鉄
千田　大介　二階堂善弘
師　茂樹　山崎　直樹

デザイン
DTP　睡人亭：http://www.shuiren.org/

発行人　尾方敏裕

発行所　株式会社好文出版
〒 162-0041
東京都新宿区早稲田鶴巻町 540
林ビル 3F
TEL:03-5273-2739
FAX:03-5273-2740
URL:http://www.kohbun.co.jp/


○本誌に関する訂正・補足情報は、漢字文献情報処理研究会サイト（http://www.jaet.gr.jp/）に掲載します。

○本誌の定期購読をご希望の方は、以下の項目につき明記の上、好文出版まで、書面・FAX もしくは電話にてお申し込みください（住所・FAX・電話は上記奥付参照）。

◦送付先住所　◦氏名　◦年齢　◦職業
◦勤務先　◦必要部数

○漢字文献情報処理研究会への入会をご希望の方は、http://www.jaet.gr.jp/guiding.html の趣意書および規約をよくお読みの上、同ページにリンクが掲載されている入会フォームよりお申し込みください。書面での申し込みは受け付けておりません。


ISBN978-4-87220-139-0
C3004　¥2000E