

漢字文献  
情報処理研究

第9号

漢字文献情報処理研究会 編

好文出版

# 漢字文献情報処理研究 第9号

## 目次

---

論文	4 文字とアーカイブについての問題提起 ——序論的ところみとして——	當山日出夫
	11 人文系データベースを構築するとはどういうことか？	白須 裕之
	20 送り仮名・返り点付き漢文資料からどのような言語学的情報が得られるか 山崎 直樹	山崎 直樹

---

特集 1：今だからこそ！ 日本語 IME を見直そう	29	
	30 IME について	二階堂善弘
	32 MS-IME	山田 崇仁
	39 ATOK	山田 崇仁
	47 ことえり	上地 宏一
	52 ウェブベース日本語 IME	秋山陽一郎

---

特集 2 文献画像のデジタル化をめぐる諸問題	61	
	62 より包括的な「漢字文献情報処理」をめざして 特集「文献画像のデジタル化をめぐる諸問題」によせて	師 茂樹
	66 寺院文化財調査におけるデジタル写真撮影 花園大学歴史博物館の調査活動を事例に 志水 一行	志水 一行
	71 文献のデジタル画像化 ——『周作人書簡集』影印本製作（失敗の巻）——	小川 利康
	80 報告：学会誌論文のデジタル化	山崎 直樹
	83 ハーバード大学における蔵書デジタル化支援体制 ——韓国古典籍デジタル化事業を中心に——	江上 敏哲

---

---

## 中国語 CAI 実践レポート 91

92 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! 清原 文代

103 Adobe Captivate を用いた教材作成 Flash ムービー教材を手軽に作る方法  
鈴木 慎吾

---

## ソフトウェア レビュー 107

108 アプリケーションソフト iPhone で中国語 / iPhone & iPod touch 2.0 Software / WWW ブラウザ / Office 2008 for Mac / フォント / 楽々中国語 V5 / Google ピンイン IME / 一太郎 2008  
内田慶市 / 秋山陽一郎 / 上地宏一 / 師茂樹 / 金子真也 / 千田大介 / 山田崇仁

---

## 学術リソース レビュー 137

138 学術サイト RSS リーダーで購読すべき人文系のブログ・Web サイト / 中国哲学 / 中国史 / 大蔵経データベース / 東大寺文書目録データベース / 中国のネットと IT 事情 / 図書館・OPAC・漢籍 / 中国語教員必見サイト / 「龍源」(Dragon source) ——中国語雑誌オンライン購読サービスを試す——  
岡本真 / 齊藤正高 / 佐藤仁史 / 石井公成 / 照井武彦 / 千田大介 / 小島浩之 / 金子真也 / 小川利康

172 学術ソフト・製品 『漢語大詞典』CD-ROM ver.3 / 唐長安の数碼重建 (A digital reconstruction of Tang Chang'an)  
千田大介 / 小島浩之

---

書評 178 『ICT を活用した外国語教育』/ 『キーボード配列 QWERTY の謎』

---

コラム 89 フリーソフトで出来る画像処理 山田 崇仁

---

お知らせ 171 漢字文献情報処理研究会 入会のご案内

---

180 漢字文献情報処理研究会彙報 / 著者紹介

---

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

---

# 文字とアーカイブについての 問題提起

## ——序論的こころみとして——

當山 日出夫（とうやま ひでお）

---

本年(2008)7月25日、情報処理学会「人文科学とコンピュータ研究会」(2008-CH-79)が、神奈川県立金沢文庫において開催された。この研究会において、筆者は、『文字とアーカイブ——デジタルアーカイブの視点からの問題提起——』と題する発表を行った。

このときのCH研究会は、他にもアーカイブについての発表が多数あった。

まず、特集としては、『特集セッション「アーカイブズとデジタル技術の未来を考える」』(2)が、プログラムに組まれた。そのタイトルは、次のとおり。

- 「デジタルアーカイブズ」における永久保存の概念  
五島敏芳（国文学研究資料館）
- パネル討論「アーカイブズとデジタル技術の未来を考える」(2) ——アーカイブズ  
の概念とデジタルアーカイブ——  
鈴木卓治（国立歴史民俗博物館）、五島敏芳（国文学研究資料館）

また、この他にも、

- 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——  
後藤真（花園大学）

などの発表があった。

本稿には、筆者の発表を基本にしつつ、上記のアーカイブ関連の発表についてふれながら、現時点における、「文字」と「アーカイブ」について、いささかの考察をこころみるものである。基本的には、CH研究会の原稿を基本としつつ、上述の他の研究発表との関連づけながら、適宜、加筆・削除をほどこしたものである。

---

## 回 はじめに

この原稿は、Windows Vista、MS-Word 2007、JIS X 0213:04、MS 明朝・MS ゴシック・Century メイリオ、という環境で書いている。では、この文書『情報処理学会研究報告 2008-CH-79-4』（金沢文庫）が、なにがしかの形態でデジタル化保存されたとき、将来（すくなくとも数十年後）、同じ「文字」の文書として見ることが可能であろうか。技術的な問題もさることながら、まず、このような問題点の発想を、今のわれわれは、共有し得ているであろうか。

※この箇所は、CH-79 の発表原稿について述べたものである。あえて、同じ文章を掲載する。これと同じことが、本稿が掲載となる『漢字文献情報処理研究』についても、いえるはずである。筆者が、この原稿を書いている時点では、最終的に、どのようなフォントで印刷されることになるのか、関知していない。しかし、この原稿のデータがデジタル保存されたときも、同様の現象を考えなければならない。さらには、もし紙の本の形態で残ったとしても、それを OCR で読み込んだとき、どのような文字として認識し出力されることになるのか、本稿のような論考のみならず、古典籍の翻刻においても、同様のことを考えておかなければならない。

---

## 回 文字とアーカイブの二つの論点

文字とアーカイブについては、基本的に、二つの方向がある。

第一に、現在の文書が、アーカイブとしてデジタル保存された場合、同じ文字（字体・グリフ）で、見ることが可能であるかどうか。

第二に、現在、われわれが使っている文字、21 世紀初頭のコンピュータ環境で見ている文字について、将来にわたって残し、記録・保存する必要性について、どう考えるべきであるのか。

それ加えて、現在（2008 年）では、コンピュータの文字としては、JIS X 0208 と JIS X 0213:04

とが共存している。現実的には、Windows XP と Windows Vista の問題でもある。また、さらに、現在、「(新) 常用漢字表」の改訂が問題になっている。「常用漢字表」の改訂は、JIS 漢字規格に影響を及ぼす可能性がきわめて大きい。いま（2008 年）からしばらくは、コンピュータの文字において激変期にあると言っても過言ではない。

(新) 常用漢字表の改訂の問題については、2008 年 7 月 19 日、「ワークショップ：文字——(新) 常用漢字を問う——」（花園大学）が開催された。この WS において、安岡孝一・小形克宏氏より、詳細な資料の紹介と問題提起がなされている。この WS は、漢字文献情報処理研究会が、協賛として、開催された。

---

## 回 デジタル・アーカイブの文字

CH-79 研究会が、アーカイブ特集（第 2 回目）となっているように、現在の日本社会において、「デジタル・アーカイブ」の構築は緊急の課題である。それには、以下のようないくつかの問題点がある。

### ※ (1) デジタルアーカイブの保存性の問題

いわゆる「アーカイブ」「アーキビスト」の側の人たちが、「デジタル」（特にその保存性）にどのように取り組んでいるかという問題。

この問題については、現時点では、きわめて悲観的な印象をもたざるをえない。日本アーカイブズ学会を中心とする人たちが、文書のデジタル保存について、きわめて否定的な見解を持っていることは、関係者にはよく知られている。その否定的根拠の主な理由は、デジタル化資料の、保存の安定性にある。

具体的には、以下のような理由による。紙（中性紙）であれば、100 年以上の耐久性がある。しかし、デジタルのデータは、どうであるか。CD-R や DVD は、そのまま 100 年の耐久性があるであろうか。あるいは、ハードディスクを 100 年間にわたって連続的に稼働させることは可能であろうか。これらは、現在のコンピュー

タ技術においては、否、であろう。

しかし、この問題点については、次のように反論できる。デジタル化データは、メディア変換とコピーを繰り返すしか保存が保証できない。そうである以上、これを認めた上で、そのコスト（人的・資金的）を、アーカイブの運用・保存に、組み込んで考えればよいのである、と。

また、現在、デジタル環境で発生し流通する文書・記録がある。例えば、国や地方自治体のHPなど、である。これは、公文書、と認定すべきと筆者は考える。このようなものは、もはや、デジタルでしか保存し得ないものである。さらに言えば、アーカイブの基本原則である、文書の原秩序維持は、デジタル文書（HPなど）において可能であるかどうか、再検討の余地がある。

デジタルの環境で発生する文書について、CH-79における五島敏方氏の発表では、「born digital」と表現されている。

## ※ (2) デジタル公文書

「デジタル公文書」と称さなければならぬものが現実に存在している。

たとえば、「(奈良県) 葛城市」「(東京都) 葛飾区」のHPは、インターネットでの閲覧を前提に作成されている。このなかには、市・区の正式名称としての文字についての言及がある。また、「(愛知県) 岡崎市」には、ポルトガル語のHPがある。これは、現在、この市における人々と行政のあり方（ブラジルからの労働者の存在への配慮）を端的にしめす公的な記録と考える。ポルトガル語のHPを行政として用意しておく必然性があったということ、およびその記録として、デジタル公文書としてアーカイブの必要がある。

## ※ (3) 2種類のデジタル化文書

以上、デジタル化文書について、その保存を中心に考えてみた。しかし、さらに考えると、文書のデジタル化には、2種類あることに気づく。

第一は、既存の文書の画像データ化である。この代表的なものとしては、国立公文書館のデジタルアーカイブや、国会図書館の近代デジタルライ

ブラリーもある。

これは、いくら文書作成がコンピュータ（ワープロ）に依存するようになっても、残ることである。少なくとも、文書に、印鑑（あるいはサイン）が必要とされる限りは、最終的な文書は、紙の形態で残ることになる。そのデジタル画像化と保存、ということになる。

第二は、印鑑やサインが必要とされない文書であるならば、ワープロ文書データのままで残すということが可能になる。あるいは、PDF化である。

この場合には、使用の文字コード、あるいは、使用のフォントが、問題となる。例えばCH-79の原稿であれば、筆者は、MS明朝で、書いている。そして、プリントアウトし、論集のカメラレディ原稿となる。だが、これを、別のフォントで表示・印字することは、いとも簡単である。つまり、文書の文字については、フォントを指定しない限り、同じ見え方を保証することができない。

## ※ (4) デジタル化文書と文字

コンピュータ環境による文字の見え方が違う、このようなことは、ワープロ文書に限ったことではない。今の社会で日常的に必須となっている電子メールでも、その表示フォントの設定によって、見え方は異なる。等幅のMSゴシックで表示するか、プロポーショナルのMSPゴシックで表示するかで、文書レイアウトとしての、見え方は同じではない。

さらに、文字コードが異なれば、同じ文字（字体）が見えない、ということが起こる。

筆者の研究対象とした文字として、「祇」「葛」がある。

- 「祇」は、0208では「ネ氏」、0213:04では「示氏」となる。
- 「葛」は、0208では「ヒ」であり、0213:04では「人」である。

これらの漢字は、地名としてよく使う。「祇」は、京都の祇園が有名である。しかし、地名としては、

祇園信仰の全国への広がりにもなって、全国各地に「祇園」の地名・駅名・学校名など、がある。

「葛」は、奈良県葛城市では「葛（ヒ）」を公式な文字とし、東京都葛飾区では「葛（人）」を公式な文字としている。それぞれのHPには、地名文字についての解説がある。HPなどは、アーカイブの対象ではない、などと言うことはできない。これらのHPで伝えている内容は、この市や区が、どのような漢字（字体）で書くべきか、その根拠と判断を示した、公的な文書であると、認定すべきである。

では、これらの漢字をふくむ地方自治体の文書（公文書）は、どのように見ればよいのであろうか。あるいは、将来、どのように見えるのであろうか。ここで、デジタルアーカイブの文字が、課題となる。

アーカイブの基本理念が、現在の文書を記録として未来に残す、ということであるならば、今、現実になれわれがパソコンのディスプレイで見ている、その文書・HPを、将来に残して保存することが重要課題である。文字情報（文字概念のレベル）として同じ文書であればよい、というものではないはずである。人間が、文書を書くとき、手書きであれ、ワープロであれ、なにがしかの価値判断のもとに、文字を選んでいく。複数の字体（異体字）があれば、そのうちどれを選んで文書を書いたか、このことも、また、文書に内在する歴史資料としての価値である。

「葛（ヒ）」か「葛（人）」か、そのどちらを選んでいるのか、それを、HPでどのように説明しているのか、また、それを見る人は、どのように見ているのか、これは、文書の文字についてアーカイブすべき事項である。

以上を総合して述べるならば、アーカイブ（デジタル）においては、その表示の文字（字体・字形）をもふくんだものでなければならない。

## ㊦ 文字のアーカイブ

次に、逆の方向から考えてみる。デジタルの環境における文字のアーカイブである。

第一に、今、われわれが使っている文字を、アーカイブとして、将来に残す必要性。

第二に、文字を、記録し・残す必要があるとするならば、それは、どのような手段によって可能であるのか。

デジタル化文書は、それを「テキスト（画像ではなく）」として残す場合、それに使用した、「文字」とともに保存しなければならない。でなければ、将来にわたって、同じ文書を、同じように見える、ということの保証が得られない。

文字について、主に次の三つの視点がある。

第一にコード系：現在のコンピュータによる文書データであるならば、このレベルで語られることが多い。

第二に文字セット：冒頭に述べた通り、筆者は、この原稿を、「JIS X 0213:04」の環境で書いている。日常的に書く文書や、論文、あるいは、電子メール、ブログの文章、などである。だが、筆者の方針として、基本的に、JIS第1・2水準の範囲内で書くようにしている。第3・4水準の文字は、原則使わない方針でいるし、ユニコード（Ext.A）も、使用しない。どうしても使用するときには、その旨を、明記した上でメッセージを書く。（この文書では、後述の「鷗」だけが例外的に、第3水準文字になる。）

つまり、Windows Vista（Ultimate）で、Word 2007で書く環境では、JISの第3・4水準、さらには、Ext.Aまで、確実にあつかえる。しかし、潜在的にその可能性があるということと、現実には、どの範囲の文字を使用するかは、別次元のことがらである。これは、文字コードにかかわる範囲での「文字セット」についてのことである。

だが、文字コードにかかわらない「文字セット」もある。教育漢字・常用漢字・人名漢字、などである。もし、公文書として、「常用漢字」の範囲内で書く（一部、固有名詞を除いて）という方針であるならば、文書について、その旨の属性情報が必要になってくる。もし、常用漢字に厳しく限定するならば、「イタリヤ」を「伊」とは略記できない。「伊」は常用漢字外である。しかし、第1水準漢字である。

このようなことは、文書のアーカイブにおいて、その属性情報として、記載し記録に残す必要がある。どのような「文字セット」の範囲で書いた文書であるか、ということは、その文書内のことば（語）の表記に影響する。

第三にフォント：筆者は、CH-79の論文（カメラレディ原稿）を、MS明朝・MSゴシック・Century・メイリオ、で書いた。だが、この論文の文書データの文字列を、エディタにコピーすれば、このようなフォントにかんする情報は、消えて無くなってしまふ。単なる、コード化された文字の連続でしかない。デジタルのままて保存して、はたして将来、同じように見える保証があるだろうか。それを確実にするためには、フォント全体を保存しておかなければ不可能である。

---

### 回 紙の文書とデジタル文書

以上に指摘したことがらを考えると、文書の保存において、「紙」というものの優位性が見えてくる。「紙」は、それ自身で、デジタル文書における、ハードウェアとソフトウェアのほとんど全部の機能をふくんでいる。CPUであり、メモリであり、記録媒体（ハードディスクやCD-Rなど）であり、ディスプレイでもある。だが、「紙」のものだけを残せばよい、というわけにはいかない。その論点は、次の三つであると考える。

第一に、はじめからデジタルでしか発生しない文書というものがある。これは、デジタルで保存するのが妥当である。

第二に、「紙」の文書の保存と利活用のためのデジタル化の必要。通常のアークバブの概念では、その対象となる文書は、それ一点きりである。この点において、大量の複製があるものを対象とする図書館とは性格が異なる。したがって、その保存においては、可能な限り慎重でなければならない。しかし、その利活用（閲覧）は、文書を傷めることでもある。「紙」の原本の保存と同時に、デジタル閲覧も、十分に考慮にいれるべきである。デジタル複製（画像化）である。

第三に、「紙」のまま保存し、利活用に供する

としても、そのためには、検索のためのメタデータが必要になる。少なくとも現時点において、これはコンピュータに依存することになる。

---

### 回 検索と文字

そして、検索データのデジタル化にあたって、どのような「文字」で記載するのか、重要な課題となる。

その端的な例として、国立公文書館（デジタルアーカイブ）「森鷗外」を検索してみる。結論からいえば、「森鷗外」（区鳥）では検索可能。しかし、「森鷗外」（區鳥）では、何も出ない（本稿執筆時点）。

「鷗」（区鳥）：いわゆる「拡張新字体」、第1水準にある。83JIS漢字体。

「鷗」（區鳥）：いわゆる「正字体・旧字体」、第3水準にある。78JISとの互換性のために追加。

また、第2水準までの漢字であっても、「熙」「熙」の区別はかなりやっかいである。「近衛忠熙」「近衛忠熙」「近衛忠熙」それぞれに、検索結果が異なる（国立公文書館）。

---

### 回 アーカイブと文字セット

アーカイブにおいて、文字は重要である。この場合、一番重要なのは「文字セット」である。

第一に、どの文字セットで記述したか。まず、現行の常用漢字では、公文書アーカイブの検索データの作成は不可能であることを確認しておく。なぜなら、現行の常用漢字では、府県名ですら、書けない。岡山・熊本・大阪・栃木・山梨、などがそうである。府県名の文字を排除して、国、および、これら府県にかかわる公文書が、あつかえるはずがない。

第二に、では、それを拡張する範囲として、どの範囲までを許容したか。第2水準までか、第4水準までか。

第三に、いずれの「文字セット」内であっても、



その内部に、各種の異体字・新旧字体の組み合わせが複雑に存在する。これらは、統合するのか、あるいは、検索段階で、文字シソーラスを利用することにするのか。

第四に、そして、アーカイブにとって重要なことは、そのアーカイブ自体が、どのように形成されたのかのプロセスもまた、アーカイブの対象である、という視点である。アーカイブのプロセスそれ自体の記録というのも、アーカイブにとっては、きわめて貴重な概念である。それを、記述の文字セットの選択にも適用する必要がある。文書・資料の整理・検索のために、どのような文字セットを、どのような判断のもとに利用したのか、これも、記録され、公開されなければならない。

## 回 文字のアーカイブのために必要なこと

現在のコンピュータ社会、また、アーカイブの利活用のためには、コンピュータにおける文字というものが、いかに重要であるかは述べたとおりである。では、その文字を、どのようにアーカイブすることができるであろうか。

第一に、規格票そのものの保存。日本に限定してであるが、コンピュータの文字は、JIS（日本工業規格）によって、決められている。そして、それは、原則的に5年ごとに更新されることになっている。この規格票の保存が急務である。5年で更新ということは、5年以上経過して、新しい規格が作成されれば、用済みで廃棄されることになる。そして、この規格票は、意外なことに、ほとんど保存されていないことが判明している。

第二に、フォントデータの保存。しかるべき機関・組織が、フォントデータを残さなければならない。現在、一般社会では、Windows XPが、いまだに多く使用されている。XPでは、MS明朝といっても、規格としては「0208」によっている。XP（0208）と、Vista（0213:04）では、見える文字の字体・字種が異なることは既に述べた。デジタルで残せるものであるならば、可能な限り、文書データと同様に、フォントデータも、記録・

保存の対象として、策を講じるべきである。

## 回 アーカイブから学ぶべきこと

これまで、このCH研究会をはじめとする、各種の人文学とコンピュータにかかわる研究会などでは、さほど意識することなく「デジタル・アーカイブ」の用語を使用してきたように、筆者には思われる。だが、現在においては、「アーカイブ」の理念から学ぶべきこと（特に未来への責任という倫理観）があると感ずる。

特にCH-79研究会での、五島敏芳氏の発表で指摘されている、「アーカイブズ」とは、本来、永久保存の概念をふくむものであるということ。その理念のもとに、デジタル環境で発生した文書（born digital）をどのようにあつかうかを考えねばならない。

また、後藤真氏の発表で指摘された、資料の知的完全性の保証という視点も重要である。氏は、正倉院文書の刊行を事例に、誤植・誤記があっても、あえてそれを、そのまま残す（新しい翻刻は、それとは別に追加する）ことの意義を提言された。この観点から、どのような文字セットで記述したのか、ということは、きわめて重要である。0208でとどめたのか、0213まで使用したのか、あるいは、とにかく字が必要ということで、Unicodeの拡張領域まで使用するのか、これは、まさに文書の解釈の問題である。文字セットは有限であり、規格によって、変動・増減する。このことを逆に言うならば、別の文字セットの視点からは、「誤り」と認定されることになる。アーカイブにおける「文字」においては、使用文字セットと規格の情報は、絶対不可欠である。

これをふまえて、今の、デジタル技術で何が寄与できるか、さらには、現在のデジタル技術で実現している各種の事象を、どのように「アーカイブ」として残すべきか、本格的に考えねばならない段階にさしかかっていると感ずる次第である。そのなかで「文字」をどのように将来に残すことができるのか、あるいは、残すべきか否か、新たな課題である。

## 参考文献・HP

- 青山英幸. 『電子環境におけるアーカイブズとレコード』. 岩田書院. 2005
- 大濱徹也. 『アーカイブズへの眼』. 刀水書房. 2007
- 小川千代子・高橋実・大西愛(編著). 『アーカイブ事典』. 大阪大学出版会. 2003
- 小川千代子. 『電子記録のアーカイビング』. 日外アソシエーツ. 2003
- 小川千代子ほか(編著). 『アーカイブを学ぶ』. 岩田書院. 2007
- 後藤忠彦(監修)・谷口知司(編著). 『デジタル・アーキビスト概論』. 日本文教出版. 2006
- 三上喜貴. 『文字符号の歴史 アジア編』. 共立出版. 2002
- 安岡孝一・安岡素子. 『文字符号の歴史 欧米と日本編』. 共立出版. 2006
- 野村雅昭. 『漢字の未来 新版』. 三元社. 2008. (※旧版『漢字の未来』は、筑摩書房. 1988)
- 師茂樹. 「デジタルアーカイブ」とはどのような行為なのか. 『CH-66』. 情報処理学会. 2005
- 永崎宣研. 「デジタルアーカイブの弁証法」. 『CH-68』. 情報処理学会. 2005
- 當山日出夫. 「京都における「葛」と「祇」の使用実例と「JIS X 0213:2004」」. 『CH-70』. 2006
- 奈良県葛城市 市名の「葛」の字について  
<http://www.city.katsuragi.nara.jp/katsuragi/katsu.html>
- 東京都葛飾区 「葛飾区」を表記するときの「葛」の字の不思議  
<http://www.city.katsushika.lg.jp/aisatu/katsushikakunituite.html#katunoji>

## 付記

拙論の原稿を入稿の後、次の研究会が行われることになっている。参考までに記載する。これを見ると「デジタルアーカイブ」について、人文情報学の側と、アーカイブズの側との対話の回路ができつつあることを予感する。詳しい開催趣旨説明などは、日本アーカイブズ学会(JSAS)のHPを参照していただきたい。

<http://www.jsas.info/>

(なお、インターネット検索では「JSAS」で検索可能である。)

日本アーカイブズ学会 2008 年度第 1 回研究集会、2008 年 10 月 4 日(土)、学習院大学。内容は、以下のとおりである。

### ●報告

1. アジア歴史資料センターから見たデジタル・アーカイブズの現在と展望  
平野宗明(代表)、相原佳之、石田徹、蔵原大、黒木信頼、中村元、牧野元紀(国立公文書館 アジア歴史資料センター調査員)
2. 「デジタルアーカイブ」と記録資料——“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——  
後藤真(花園大学文学部文化遺産学科 専任講師)

### ●コメンテーター

岡本真(ACADEMIC RESOURCE GUIDE (ARG) 編集長)  
近藤靖之(株式会社 DNP 年史センター)

---

# 人文系データベースを構築する とはどういうことか？

白須 裕之（しらす ひろゆき）

---

---

## 回 1 はじめに

筆者は昨年度までの3年間、「京都大学 21 世紀 COE 東アジア世界の人文情報学教育据点」において、漢字文献知識ベースの構築に関する研究に従事した。本稿ではそのプロジェクト（以下、漢字文献知識ベースプロジェクト）の経験を振り返り、知識ベース構築の問題点と人文情報学の理論的な基礎の必要性を、筆者の視点から整理していこうと思う。その整理を通して、人文系データベースを構築するとはどういうことかを明らかにしたい。ここで人文系データベースという用語の意味が問題となるが、今は人文学研究を支援するための計算機システムであると便宜的に定義しておく。本稿の目的の一つがそれを明確化することである。

昨今、人文系のデータベースやデジタルアーカイブ等の様々な成果が提出されてきたが、その成果が人文学研究を本質的に支援しているかと言えば、まだ不十分な状況にあると言えるだろう。その原因は人文情報学の理論的な基礎が何かを明確にしないまま、情報ツールとしてのデータベースやアーカイブを構築していることにあると思われる。

人文学と情報学という異なる学問的な伝統を持つ二つの分野の出会いが、個々の分野の中だけでは起りえない概念的な反省を促し、人文情報学という新しい分野を創造しつつある。情報システム

による人文学研究の支援が、如何なる意味を持つかを再考することが本稿の目的の一つであるが、このことがこの分野の基礎を確立する上で貢献できれば幸いである。

---

## 回 2 本稿の背景

まず、漢字文献知識ベースプロジェクトの課題から概観しよう。本プロジェクトの目的と実施計画を明かにした文献 [高田] には、プロジェクトの課題である「情報学によって漢字文化の発展を目指す上での課題」を、以下のように二点にまとめている。

1. 「多くの情報の中から必要とするものを素早く的確に取り出すための仕組みを考えること」
2. 「我々が傳承してきた漢字文献を過不足なくデジタル化して後世に伝えることが出来るような方法を考えること」

これらの課題は一見すると、漢字文献を利用してなされる人文学研究を計算機によって支援するという意味で、方法論的に自然のように思える。我々はこの課題を解決すべくプロジェクトを進めてきた訳であるが、研究を進めるに伴って、この課題の大きさに圧倒されることになる。後から振り返って考えてみると、根本的な問題を解決しないままにプロジェクトを進めたことに原因があっ

た。筆者にとっては少なくとも、その原因を明確化するために、本プロジェクト全体の経験が必要であったのである。

文献[白須1]では、知識ベースを構築するとはどのようなことかを扱った。知識表現を情報化の手段とし、歴史記述を扱うための枠組みを提出した。本稿は文献[白須1]と非常に関係が深いが、考察の対象を人文系データベース一般に広げ、人文学の知識を表現する様々な枠組みについて考えることを目的とする。

---

## 回 3 ソフトウェア工学の視点から

前節で漢字文献知識ベースプロジェクトにおける二つの課題について述べた。その内容をソフトウェア工学の視点で眺めると、それは「情報の利用と蓄積」をどのように設計するかという問題であることが分かる。これらの問題を計算機を使って解決するためには、必要とする情報を計算機で利用できるように形式化しなければならない。ここで問題となるのは、必要な情報が何であるかということと、どのような情報を計算機上で形式化すれば良いかということである。即ち、ここでの情報は要求仕様と対象領域の概念モデリングによって得ることができる。

これらのモデリングさえ順調に進めば、ある程度の情報システムを構築できるのではないかと筆者は目論んでいた訳である。しかし、筆者の予想に反して、プロジェクトは暗礁にのりあげてしまうことになる。

漢字文献知識ベースプロジェクトが扱う対象領域は、中国唐代の歴史的な知識であったが、知識ベースに対する要求仕様は非常に曖昧であった。即ち、知識ベースの利用者が必要としている情報がどのようなものであるのかということが曖昧のままにされていた。この曖昧さの原因には宜からざるものがある。それは一般に歴史学の知識とはどのような情報であり、どのように形式化できるかということが知られていなかったからである。

この点については歴史学のみならず、他の人文学の知識についても同様な状況ではなからうか？

この二つの問題を明確化することは、そのまま人文情報学の課題を明確化することに繋がるであろう。これは知識を表現する言語と、表現できる情報の二者が表裏一体をなしているからであり、後の節で示すことになるが、言語と知識の問題を探索しない限り、本プロジェクトの二つの課題は到底解決できなかったのである。

---

## 回 4 情報検索の視点から

本節では様々な人文系データベースを統一的に理解するために、情報検索の視点から考察する。人文学研究を支援する計算機システムとしては、情報検索以外にも、情報抽出や分類、テキストの自動要約、自動分類等、様々なものがあるが、ここでは情報表現との関係に注目し、その関係が分かり易い情報検索について議論しよう。特に人文学の知識が、情報検索においてどのように利用されているかを明確化できるように、その枠組みを提出することが目的である。

### ◆ 4.1 情報表現

人間の操る表現、即ち、書き言葉、話し言葉、身振り、手振り、絵画、音楽、そして形式言語等を総称して「言語」と呼ぶことにする。「言語」はその基本機能として、環境世界の事物をシンボル化する機能と、何かについて語るという志向性の機能を持つ。「言語」についての考察は、対象となる「世界」がどのようなものであるか、「真理」をどのように扱うか等、哲学的な問題を引き起こすが、本稿は計算機の内部で情報を表現するということが目的であるので、これらの問題には深入りしない。以下、計算機による情報表現のみを対象とする。

人間はこの「言語」によって世界を情報化していく訳であるが、この情報を計算機で扱うための方法として、「形式体系」と「意味論」を明示的に構成するという立場で、情報を表現することを考えよう。

人文学研究においては、分野ごとの違いは存在するが、文献資料に非常に重きを置いているので

はないだろうか。本稿は如何に人文学研究を支援するかということを議論しているので、文献資料から作られたデジタルテキストがあることを前提とする。このデジタルテキストの蓄積を操作する仕組みを、文書リポジトリと呼ぶことにする。

文書リポジトリの例としては以下がある。ここでは主にデジタルテキストを対象とし、その内容としては自然言語で表現されているものを扱う。

- 論文、文献、歴史資料等
- 画像、音声、動画のコメント情報
- インターネットの各情報頁

人文学で扱う媒体は、デジタルテキスト以外にもデジタル画像・音声・動画等があり、その内容も遺跡・石刻史料・絵画等、多くのものが考えられるが、これらを直接、意味論的に扱うことは将来の研究に譲る<sup>[1]</sup>。但し、デジタルテキスト以外の媒体に、自然言語による内容についてのコメントが付加されているものも対象とする。

次に様々な人文系データベースを統一的に理解する仕組みとして、文書リポジトリに形式言語で表現したデータを付加する。即ち、自然言語による文書と形式言語によるデータを組としたものをここでの情報表現とする。この形式言語は後で議論する形式体系の一部分である。以下で述べることを多少先回りしてコメントしておく、形式体系と意味論が個々の文書に情報を付加する方法を提供する。この付加情報を利用して、情報検索を実現する。

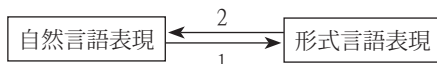


図1：情報表現

例えば、文書リポジトリに文書群が格納されているとき、メタ情報やキーワードが問合せ機構を担う形式言語表現に対応する。この対応関係は矢印1のような統計的な解析や自然言語処理によって構築するもの、矢印2のようなマークアップ技術等を使用するもの等がある。この対応関係に

ついては後程、詳しく議論する。

#### ※ 4.2 情報検索の枠組み

情報検索とは蓄積された様々な情報から、必要となる情報（の在処）を検索することである。ここでは情報検索の実現のために、「形式体系」を利用した枠組みを述べる。図は情報検索の枠組みを示したものである。但し、一般的な情報システムを表現しているため、全ての構成要素が具備されているとは限らない。

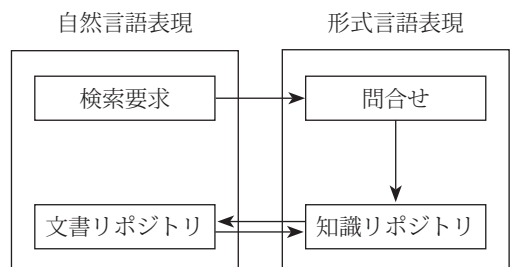


図2：情報検索の枠組み

図の右半分は「形式体系」を示していて、情報検索においては「問合せ」、「知識リポジトリ」共に形式言語で表現されている。左半分は自然言語によって表現された情報を含む「文書リポジトリ」を示している。「知識リポジトリ」と「文書リポジトリ」との関係は、情報表現で述べたように、「文書リポジトリ」中の文書に対して形式言語で付加した情報、知識を、「知識リポジトリ」に格納している。

形式体系は情報検索で使用できるように、以下のような性質を持つものでなければならない。

- 構文論——形式言語の構文は形式的に定義されていて、その構文を計算機で扱うことができなければならない。
- 操作系——問合せ等をモデル化できるように、操作の機構を供えていなければならない<sup>[2]</sup>。情報検索では、この操作系の問合せ機構を利用している。

以下は実際に人文系のデータベースで利用され

ている「形式体系」の例である。

表 1：形式体系の例

形式体系	知識リポジトリの内容
文字列操作系	キーワード、インデックス等
メタデータ操作系	メタデータ
関係代数	関係データベース
XQuery	XML データベース
書換系	書換規則
論理体系	公理（論理、対象領域の）

### ◆ 4.3 人文系データベースの例

人文系データベースの内でも比較的成功している例として以下がある。これらのデータベースを本稿で提出した情報検索の枠組みから理解し、人文学の知識がどのように使われているかについて概観しよう。

1. テキストデータベース
2. 文献・論文データベース
3. 人文系情報の Web 閲覧システム

テキストデータベースは、対象となる文書をデジタルテキストとして格納し、必要とする情報が書かれている文書及び情報の位置を検索するものである。問合せには、必要とする情報をキーワードとするもの、或いはキーワードの論理演算（AND, OR 等）が使えるものがある。キーワードを含む文書（またはその位置）を全文検索するもの、文書を適当な構造に分割して扱い、予めインデックスを作成しておくもの等があり、これらを総称して全文テキスト検索と呼ぶ場合がある。このときキーワードやインデックスは文書のメタ情報になっている。

また、文献・論文データベースや Web 閲覧システムでは、対象となる文書にメタ情報を与えて、それを利用して検索を行なう。メタ情報の抽出にはマイニング技術、統計処理などが利用されている。

情報検索の枠組みに照して見たとき、これらの

データベースでは文書リポジトリから、検索のための形式言語のリポジトリを如何に作成するかということがポイントになる。次節では知識リポジトリ作成の支援について考えるために、知識表現と自然言語処理について考えてみる。

## 回 5 言語の意味と知識

本節では、人文学研究を本質的に支援するためには如何にすべきかについて考えてみよう。前節で情報検索の一般的な枠組みについて述べたが、図の左右の間の関係については、例を見ただけに滞っていた。通常、検索要求は利用者が作成するが、これは問合せ言語の文法に従って行なわれる。

これに対応して、文書リポジトリに対するメタ知識はどのように抽出されるであろうか？このような技術としては、現在データマイニングや WEB マイニング、統計処理等、多くのものが提出されている。しかし、対象を人文学に限った場合でも、その意義について人文学的に確立できるものは少ないであろう。人文学研究においては、そのような知識の抽出に関しても人文学的に正当化できる手法が望ましい。本稿ではそのような手法を研究するために、まず最初に自然言語における形式的な意味論に目を向けてみよう。

### ◆ 5.1 自然言語の形式的意味論

対象となる文書データ（自然言語で書かれたデジタルテキスト）に対して、情報や知識を付加するための技術を模索するために、ここでは自然言語の意味論について考えてみる。本稿では計算機による支援を前提としているので、意味論としては形式的なアプローチを採ろう。

まず世界の意味論的な構造を表現する形式言語を設定することによって、世界の意味論的な分節化を表現する。自然言語の表現には構文論を通して構造が与えられるが、その各構造に対応する自然言語表現に対して、形式言語の表現を対応付ける。この対応を決めることによって、自然言語の意味解釈は、この形式言語を媒介として、世界へと対応付けられることになる。この様子を示した

ものが以下の図である。

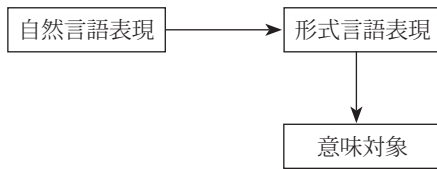


図3：意味解釈の仕組み

自然言語に構文論的な構造を与える際に、形態素解析や構文解析等で使用する辞書が必要になる。このような辞書には言語についての知識とともに、対象領域についての知識が埋め込まれている。従って、自然言語表現から形式言語表現への対応関係を構成するには、知識リポジトリのデータを活用することが考えられる。理想的なシステムを構築する上では、このような循環的な構造についても将来的に考慮する必要があるであろう<sup>[3]</sup>。

### ◆ 5.2 言語の意味について

自然言語の意味論は通常、日常的な言語使用を前提として、様々な場面での言葉の解釈を目標としてきた。従って、解釈の難しい特殊事例を設定して、その場面での自然言語の意味を決めるには、どのような意味論が必要であるかを研究するという面が多かったように思う。

自然言語の意味論の構築には、意味を解釈すべき世界と、その世界についての情報が予め与えられているという前提で、理論を構築しているようにみえる<sup>[4]</sup>。これは情報や知識を辞書という形態でデータ化して名辞に意味を与えれば、様々な言語現象を解釈できるとする言語理解の方法論に依拠する結果であろう。しかし、情報や知識は、それを表現しようとする理論や言語体系に依存している。自然言語の意味論をそのような立場で捉え直し、様々な学問分野で必要とする世界を記述できるような意味論を、構築すべきではないだろうか？<sup>[5]</sup>

### ◆ 5.3 意味論の必要性

ここで本稿の今迄の議論を纏めておこう。図1

には文書に対する情報、或いは文書から導出される知識としての「情報表現」を示し、図2には「情報検索の枠組み」を示した。また、自然言語の意味解釈の仕組みを図3で示した。実は情報検索で使用する情報や知識の意味を提出するためには、そのための意味論が必要であるという理由から、図1に含意される形式体系は図3と共通のものを考える必要があり、図3の意味論が「情報表現」の意味を与えることになる。即ち、図1は図3の形をもって完結すると言える。

従って、図2と図3を人文系データベースの一般的枠組みとする。情報検索は形式体系によって実現され、検索に使用される情報、知識は形式体系に対する意味論によって、その意味を与えられる。即ち、本稿の枠組みは意味論の構築の必要性を強く主張するものである。

対象領域の意味論が構築できた暁には、計算機で扱うことのできる知識の領域が明らかにされ、また、知識ベースを含む計算機システムへの要求仕様を、利用者の要求に従って明確化できるであろう。しかし、自然言語の意味論の研究の実情を見ても、このような枠組みをすぐに実現するのは難しいであろう。

次節では、このような人文系データベースのための意味論を構築するヒントとして、漢字文献知識ベース・プロジェクトにおいて、筆者が今まで研究対象としてきた歴史記述の場合を取り上げる。

## 回 6 歴史記述の情報表現に向けて

### ◆ 6.1 歴史記述の意味とは？

ここでは歴史記述の情報論的な意味論が構築できたと仮定して、対象となる知識をどのように扱わなければならないかについて議論する。その議論を通じて、歴史記述の情報論的な意味論を構築するためのヒントとしたい。

歴史記述の意味を考える上で様々な立場があるが、本稿では「物語り論的歴史理解」[鹿島][野家][Danto]に、歴史記述を解釈するための足場を設定したいと思う。まず前節で述べた自然言語理解

との関連で、Dantoの「物語り文」という概念を手掛かりにして考えよう。文献[Danto]では、「物語り文」を提出する目論見を以下のように述べている。

私はここで、あらゆる種類の物語に現れ、ごく自然な日常の話し方のなかにさえ入り込んではいくが、歴史叙述において最も典型的に生じるように見える種類の文を、分離し分析してみようと思う。(Danto)邦訳174頁)

Dantoは「物語り文」の概念を導入する前に、人々の歴史に対する暗黙的な見方に反省を促すために、「理想的編年史」という概念を導入する。その暗黙的な見方とは、出来事を歴史的に語る場合に、《出来事を「それが生じたまま」に語る現場の証言に歴史記述の理想を見る通念》([鹿島]16頁)に基づくものである。このような通念を具現化した超越的な能力を備えた編年史家の存在を考え、その編年史家が記述するものとして、「理想的編年史」という概念を導入する。

彼はたとえ他人の心のなかであれ、起こったことすべてを、起こった瞬間に察知する。彼はまた瞬間的な筆写の能力も備えている。「過去」の最前線で起こることすべてが、それが起こったときに、起こったように、彼によって書き留められるのである。その結果生ずる生起しつつある叙述を、私は「理想的編年史」と名付けることにしよう。(Danto)邦訳181頁)

過去における出来事の全ての記述が書かれているという意味で、「理想的編年史」は歴史家の叙述が目標とすべきものになろう。従って、「理想的編年史」が存在すると仮定したとき、実際の歴史家が行なうことができるのは、以下の修正規則を適用して、歴史の叙述を改訂していくことだけになる。

1. 理想的編年史には現れているが、歴史家の叙述にはまだない文をつけ加える。
2. 歴史家の叙述には現れているが、理想的編年史にはない文を削除する。
3. 歴史家の叙述に残っているすべての文の位置を、理想的編年史におけるそれに相当する文の位置に合致させるため、相互に交換する。

しかし、果して歴史家の営為はそのようなもので尽されているだろうか？ Dantoは上で引用した「物語り文」の目論見でも分かるように、歴史叙述の分析を通して、出来事を歴史的に語る場合の時間的な構造に注目する。「物語り文」という歴史記述は、

ふたつの別個の時間的に離れた出来事、E1 および E2 を指示する。そして指示されたうち、より初期の出来事を記述する。(Danto)邦訳185頁)<sup>[6]</sup>

より初期の出来事 E1 を目撃した人物は、出来事 E2 について知るすべを持たないので、この文は「理想的編年史」に含まれない。このように出来事 E1 の歴史的な意味は、出来事 E2 との関係によって与えられることになる。

物語り文を解釈するための意味論は、このように出来事 E1 および E2 を記述する状況（実際にはこれらについて語る歴史家の談話状況）と、この物語り文が書かれた状況との関係を記述できなければならない。また、その状況には出来事間の時間構造を含める必要がある<sup>[7]</sup>。

更に Danto は「理想的編年史」に含まれない機能として、指示機能を問題にしている。

理想的編年史にはある指示機能が欠落している。つまり、「…の場所」とか「…の人」といった関係代名詞を用いることによって、ある特定の出来事、人物、場所をさし示すような表現が脱け落ちており、このブランクは、その個物が存するときよ



りも、時間的にのちに起こるような出来事を指示する表現によって埋められるのである。([Danto] 邦訳 192 頁)<sup>[8]</sup>

言語の意味解釈において、指示の問題はしばしば議論となるが、「物語り文」が出来事の歴史的な意味を与えていることを考慮するならば、指示についても「物語り文」の解釈とそこから構成される状況、時間構造についても検討する必要がある。

文献[野家]では、指示機能を様々な言語レベルで分節化された世界との関連で議論している。指示機能の定式化は対象化された世界がどのように構成されたのかに依存している。

これに関連して、文献[野家]は理論的对象について以下のように述べている。

理論的对象を一つの存在者として指定し、それへの指示を可能にするのはそれらを包括する「理論的背景」以外にない。([野家]222 頁)

理論的对象としての歴史的対象の場合、背景となるのは物語りの歴史理論であり、歴史的対象への指示を可能にするのは「物語り文」であることになる。この点を考慮すると、歴史記述の意味論を構築する上では、自然言語の意味論で通常用いられる実在論的な考え方も、反省してみる必要があるであろう。

## ※ 6.2 歴史的時空間の構成

通常、自然言語の意味論では、解釈に使用する状況において時間及び空間はパラメータに過ぎない。これは従来の時空間の通念として、物理的に構成されたものを前提とするからである。

しかし、ここまでの議論で分かるように、時間及び空間は歴史記述において構成されるものである。即ち、時間や場所という時空間自身が、歴史学の概念枠としての歴史的な出来事によって構成される。出来事と概念枠の成立は一種の解釈学的な循環の中にあると言える。こうして時空間には

以下のような様々な構成があることになる。

1. 歴史的な事実から構成される歴史的な時空間<sup>[9]</sup>
2. 曆書、地理書から構成された時空間<sup>[10]</sup>
3. 物理的な時空間

多くの人文系の時空間データベースが、物理学によって形作られた直感に基づく時空間を無反省のまま採用している。その分野の知識はそのための意味論によって表現されるべきで、意味論が必要としている時空間を本来使用すべきであると思う。

## ※ 6.3 歴史記述の枠組み

歴史記述のための情報システムは、対象領域である歴史情報についての問合せとともに、その情報の典拠となったテキストを示す必要がある。この二重の要求を充すために、歴史を扱う情報システムの構成は以下の二つの部分からなる。

1. 対象領域である歴史記述についての情報システム
2. テキストの構造、校勘情報についての情報システム

これらは各々その対象となる知識を扱い、また協調して動作する必要がある。しかし、情報検索のためには、システム1の歴史記述についての情報システムが中心であり、そのシステムの文書リポジトリを管理するのがシステム2になっている。

情報表現としての文書と知識の関係は、以前にも述べたように意味解析によって抽出できれば理想的であるが、現在はそのようなシステムを構築する前段階であるという認識のもと、文献[白須1]では、知識の根拠を示すために、文書の場所を提示するという枠組みを提出した。文献[白須1]では、このような仕組みをメタ言語の立場から記述している。システム1が「語彙」を、システム2が「証拠」を扱い、二つのシステムの

関係を「表明」と呼んだ。

## ◆ 6.4 プロジェクトを振り返って

本稿の考察を前提として、漢字文献知識ベース・プロジェクトについて振り返ってみよう。最初に述べたプロジェクトの二つの課題は、知識ベースとは何かを考察するうちに、対象領域の意味論を構築する試みへと発展した。特に二つ目の課題中にある「過不足なくデジタル化」するということは、情報を表現する枠組みが提出できて、初めて判定できることであり、その枠組み次第でその意味も変わってしまう。筆者はこの「過不足なく」の意味にプロジェクトの間、悩まされ続けてきたのである<sup>[11]</sup>。今やその意味論は、ここで述べた歴史記述の意味を表現できるようなものへと発展されねばならない。

## 回 7 おわりに

人文学の研究において、比較的良く使われている情報ツールに grep と関係データベースがある。これらのツールを使ったときの人文学の知識の有り所を見てみよう。

grep では探したい文字列を指定して、その文字列が文書に含まれるかどうか、含まれるならば何処に含まれるかを調べる。検索された文字列の意味は、その文字列が文書中のどのような文脈にあるかによって決まる。即ち、文字列の意味は検索した研究者が文脈から読みとることになる。この段階では文字列の意味は、検索になんらの情報も提供していない。

関係データベースを使用する場合は、データベーススキーマに対して、ある種の人文学の知識をモデル化できる。しかし、文書リポジトリ中の文書に付加するモデルとして、関係データベースが適切であるとは限らないし、対象領域の知識が関係データベースで表現できるとは限らない。例えば、歴史記述における時間表現は、既存の Temporal Database では表現できない [白須 2]。

本稿では情報検索の一般的な枠組みを提出し、そのための情報表現として、形式体系と意味論を

使用することを述べた<sup>[12]</sup>。従って、情報システムを研究の支援として使う場合、我々は対象となる研究分野の意味論を構築していかなければならない。ここには人文学研究の営為を意味論として表現することが含まれている<sup>[13]</sup>。人文系データベースの研究者は、データベース構築にあたって、どのような意味論を構築したのかということを意識されてはどうだろうか？

## 謝辞

本稿を書く機会を頂いた山田崇仁さんに感謝いたします。京都大学 21 世紀 COE 「東アジア世界の人文情報学教育拠点」<sup>[14]</sup>において、高田時雄さん、安岡孝一さん、守岡知彦さん、永田知之さん、C. Wittern さんにお世話になりました。また、同プロジェクトで共同研究者であった秋山陽一郎さん、牛根靖裕さん、山田崇仁さんには日頃の議論を通して多大の影響を受けました。東京大学大学院人文社会系研究科次世代人文学開発センターの下田正弘先生に感謝しております。永崎研宣さんには多くの助言を頂きました。また、お世話になった多くの方の助力に感謝いたします。最後に妻留美と家族に感謝します。

## 注

- [1] 本稿は論理表現を手掛かりに考察を行なっているが、文献 [守岡 1] では、一般キャラクター論について、「(深層構造としての) 論理表現から (表層構造としての) 視覚的表現が生成されるという (計算機の世界で根強い) 「論理表現中心主義」では不十分な問題系である」としている。
- [2] 対象領域によっては、その操作が Turing 完全性を持つとは限らない。
- [3] 文献 [守岡 2] は古典中国語用の形態素解析を扱っている。漢字文献についての知識ベースという観点から、次世代の文字処理環境である CHISE 及び文字オントロジーと、自然言語処理との連携について、今後の研究が期待される。
- [4] 例えば、文献 [徳永] を参照。

- [5] 本稿は対象領域に依存しない言語現象を否定するものではなく、先に述べた言語理解の方法論への反省を志向している。
- [6] Dantoによる「物語り文」の例を見よう。科学者Sが理論Tをt1時に発見し、その後t2時に別の科学者S'が理論Tを発見する。Sはこの理論を公表しなかったために、S'が理論Tを発見したことが、一般の科学理論の体系に組込まれているとする。その後、ある科学史家がSがS'より以前に理論Tを発見していたことに気づき、「SはS'によるTの発見を先取りしていた」と言う。この文はSがt1時にしたことの記述であるが、理想的編年史には記載されていない。
- [7] このような解釈は、例えば状況意味論で記述できるであろう。しかし、歴史記述の意味論では、時間の構造も含めて、従来提出されてきた状況よりも詳細な構造をもつ状況を記述できる必要がある。
- [8] Dantoの提出している例を見よう。ニュートンは「プリンキピア」を1687年に出版した。従って、それ以前には「プリンキピアの著者」についての記述は理想的編年史にはない。しかし、「プリンキピアの著者が1642年に生れた」なる文は通常、歴史記述として扱われる。
- [9] このような時間の構成の例として、まだ不十分ながら文献[白須2]を参照して欲しい。
- [10] 地理書から構成された空間の例として、例えば文献[牛根]を参照して欲しい。
- [11] 筆者は当初、この「過不足なく」の意味を、知識の集積によって実現できる「理想的編年史」のようなものを構築することである、と漠然と考えていた。
- [12] 人文系の知識には多くの階層が存在する。情報検索の枠組みにこの階層という面をどのように表現すべきか、という問題は今後の研究課題とする。

[13] 逆に対象とする研究分野の意味論の構築が、その研究分野の方法論に变革を迫ることも有り得よう。

[14] <http://coe21.zinbun.kyoto-u.ac.jp/>

## 参考文献

- [牛根] 牛根靖裕, 白須裕之, 山田崇仁: 唐代行政地理の地図情報について, 人文科学とコンピュータシンポジウム「じんもんこん」2007, 2007.
- [鹿島] 鹿島徹: 可能性としての歴史. 越境する物語り理論, 岩波書店, 2006.
- [白須1] 白須裕之: 歴史記述に対する概念分析の試み, 情報処理学会研究報告, 2007-CH-74, 2007.
- [白須2] 白須裕之: 歴史記述における時間概念について, 人文科学とコンピュータシンポジウム「じんもんこん」2007, 2007.
- [高田] 高田時雄: 漢字文化の全き継承と発展のために、『漢字と文化』創刊号, 3-4, 京都大学人文科学研究所, 2003.
- [徳永] 徳永健伸: 情報検索と言語処理, 言語と計算 5, 東京大学出版会, 1999.
- [野家] 野家啓一: 物語の哲学, 岩波現代文庫, 岩波書店, 2005.
- [守岡1] 守岡知彦: キャラクターを考える, 人文情報学シンポジウム, 京都大学人文科学研究所, 2007.
- [守岡2] 守岡知彦: MeCabを用いた古典中国語の形態素解析の試み, 情報処理学会研究報告, 2008-CH-79, 2008.
- [Danto] A.C. Danto: Analytical Philosophy of History, Cambridge UP, 1965. (邦訳河本英夫訳: 物語としての歴史, 歴史的分析哲学, 国文社, 1989.)

# 送り仮名・返り点付き漢文資料からどのような言語学的情報が得られるか

山崎 直樹 (やまざき なおき)

## 回 1 この研究の目的

本研究は、返り点や送り仮名等を付された漢文(下の画像参照)から、どのような言語学的情報が引き出されるか、そしてその情報が有益であるかどうかを検証しようとするものである(ここでいう「言語学的情報」とは、特に統語構造に関する情報のことである)。

秦之攻韓魏也無有名山大川之限稍蠶食之傳國都而止音傳  
燕固弱國不足畏也秦之所害於天下者莫如趙然而秦不敢  
舉兵伐趙者何也畏韓魏之議其後也然則韓魏趙之南蔽也

この研究は、京都大学人文科学研究所附属漢字情報研究センター「東アジア古典文献コーパスの研究」共同研究班(班長・安岡孝一)内のプロジェクトである「漢文自動解析システムの開発」の一環である。このプロジェクトは、班内で“自動「点」打ちプロジェクト”とも呼ばれていることからわかるとおり、句読点などが一切施されていない中国古典語の白文を機械的に分節して句読点を施すシステムを開発することが目的である。現在、このシステムの学習・検証用データとして、すでに著作

権の切れた返り点・送り仮名つき漢文資料を、言語学的なメタ情報を付した電子コーパスに機械的に変換することが検討されている。

この文章の目的は、上述のプロジェクトに資するために、(1)返り点・送り仮名つき漢文から、自動的・機械的にどのような言語情報を導きだすことができるか、(2)その情報が、「漢文自動解析システム」の開発に有益であるかどうか、を検証することである。

## 回 2 注釈つき漢文の中の利用できるもの

漢文に対して施された種々の注釈は、原文の中国古典語に付された一種のマークアップであると考えてよい。このマークアップのうち本研究の目的に使えるものを、以下で概観する。

### ◇ 2.1 点と丸

まず、「、」「。」について。

この標点符号は、古典中国語の文がどこで切れるか、つまり、「この語が出現したら、必ず文/句が終止する」という学習をするのに役立つ(そのような法則が存在すれば、だが)。しかし、古典中国語は文末助詞が発達した言語であるので、常用の文末助詞のリストは、文法書などから簡単

に得られる。そのようなリストがあれば、漢文資料からわざわざ文終止の学習をする必要はない。

文末助詞以外にも、句や文の切れ目に来やすい字（形態素）は存在するかもしれない。それは、他の字（形態素）に前接して複合することのない字（形態素）である。そのような字（形態素）が存在するのであれば、「<sub>レ</sub>」「<sub>ル</sub>」は、その学習に役立つ。

しかし、この「接続確率で句や文の切れ目を判断する方法」のみでは、高い精度を期待できないことがすでにわかっている<sup>[1]</sup>。

### ※ 2.2 送り仮名と振り仮名

振り仮名は、日本語での訓読みが使用されることが多いので、原語の意味に対するメタ情報になる。ただ、振り仮名は、資料によっては付されることが極めて少ないので、体系的な資料にはならない。よって、ここではこれ以上検討しない。

送り仮名として付されるのは次のような情報であろう。

- 用言の活用語尾
- 体言に付着する格助詞・副助詞・複合助詞

これらがどのように活用できるかは、§3～4でやや詳しく検討する。

### ※ 2.3 返り点

山崎（2007）では、漢文に付された返り点を入力として、古典中国語の統語構造をある程度反映した樹構造を自動的に取得できることを論じた（枝の交差しない樹構造はXMLで表現できるので、得られた統語構造はXML文書としてデータ化される）。同時に、得られる情報には限界があることも論じた。

例えば、次の(1a)の返り点が付された文は、(1b)のXML文書に置き換えられ、そこから(1c)の樹構造が得られる。

- 1a) 此<sub>乙</sub>非<sub>下</sub>所<sub>下</sub>以<sub>下</sub>跨<sub>下</sub>海内<sub>下</sub>制<sub>中</sub>諸侯<sub>上</sub>之術<sub>甲</sub>也。  
《史記・李斯》

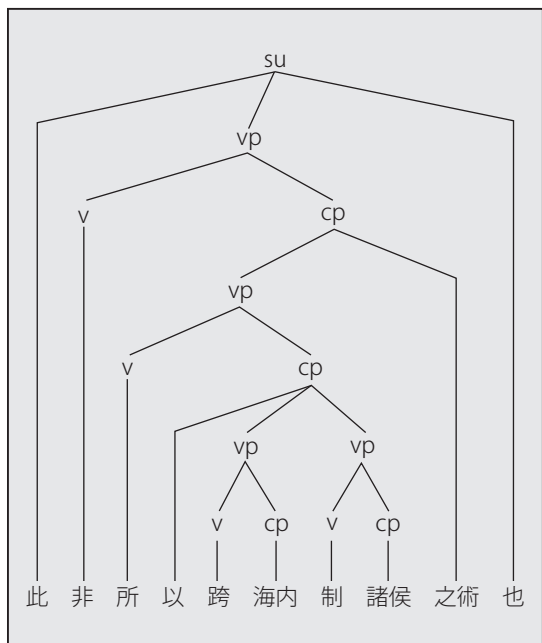
1b)

```

<su> 此
  <vp>
    <v> 非 </v>
    <cp>
      <vp>
        <v> 所 </v>
        <cp> 以
          <vp>
            <v> 跨 </v><cp> 海内 </cp>
          </vp>
        </vp>
      </cp>
    </vp>
  </su>
  之術
</cp>
</vp>
也
</su>

```

1c)



この樹構造は、下記の書き換え規則から得られる極めて単純なものである（返り点のみからでは、これ以上の情報は得られないので、単純な構造にならざるを得ない）。

1.  $su \Rightarrow x \text{ vp}^+ y$  (+ = 重複生起、 $x, y =$  決定不能な要素)
2.  $vp \Rightarrow v \text{ cp}$
3.  $vp \Rightarrow v \text{ vp}^+$
4.  $cp \Rightarrow x \text{ vp}^+ y$

$su$  は「文」である。「文」は、決定不能の要素（語彙範疇かもしれないし句範疇かもしれない） $x, y$  を含む可能性がある。それは、 $vp$  のどちら側にも出現しうる。 $vp$  は、 $v$ （これは、動詞的な性質をもつ語彙範疇と解釈していただきたい）を主要部とする句範疇（ $p$  で示す）である。ある句範疇  $vp$  の中に、語彙範疇と句範疇があれば、語彙範疇が、自動的に、その親節点で示される句範疇の主要部になる。 $vp$  の補部は、 $vp$  であることもあるし（否定辞、助動詞、句目的語をとる他動詞が主要部のばあい）、 $cp$  であることもある。 $cp$  は補部を形成する句範疇という意味である。その内部は、もし返り点をもつ構造があれば、 $vp$  と他の未定義要素を含んでいると解析できるが、返り点がなければ、それ以上解析しない（＝できない）。

$v$  の補部が  $vp$  であるか  $cp$  であるかの判断は容易である。 $v$  と判断される要素（ $\leftarrow$ 点か、返り点のうち、一、上、甲、天……などの序列の最初の記号を除いた記号が付されている要素）に直接後続する要素が、 $v$  と判断される要素であれば、 $vp \Rightarrow v \text{ vp}$  という構造であり、でなければ、 $vp \Rightarrow v \text{ cp}$  である。

### 回 3 送り仮名から何が得られるか(1)

この章と次の章では、送り仮名から得られる情報と、それが有益かどうかを検証したい。

#### ◇ 3.1 多機能語の抽出の可能性

送り仮名から、ある語が多機能語であるかどうかを知ることができる。次の例を見ていただきたい。

- 2a) 客有遠為吏而其妻私於人者。《史記・蘇秦》
- 2b) 易王母、文侯夫人也、與蘇秦私通。《史記・蘇秦》

(2a) の「私」に「スル」という送り仮名があり、(2b) の「私」に「ニ」という送り仮名があったとする。これは、(2a) は「(ワタクシ) スル」という動詞として読んでいることになり、(2b) は「(ヒソカ) ニ」という連用修飾語として読んでいることになる。もちろん、振り仮名がなければ（ ）内の情報を得ることはできないが、送り仮名として付されている活用語尾から、この「私」は述語動詞と連用修飾語の2つの用法があることがわかる。

次も同様の例である。

- 3a) 日暮路遠。《尉繚子・兵教下》
- 3b) 叟不遠千里而來。《孟子・梁惠王上》

同じく、(3a) の「遠」に「シ」という送り仮名があれば、これは「(トオ) シ」という形容詞、(3b) の「遠」に「シトセ」という送り仮名があれば、「(トオ) シトス」という動詞として読んでいる（「不遠」で「トオシトセズ」）ことになり、この語は、述語形容詞と述語動詞の多機能語であることがわかる。

このように、同一の語彙に付された複数の活用語尾から、多機能語のリストを作成することができる。

#### ◇ 3.2 しかし我々はすでにこの情報をもっている

§ 3.1 で、その可能性を示唆した多機能語のリストは、有益であるように思える。しかしごく大雑把に言うと、我々はすでにこの種の語彙情報がある程度持っている。

というのは、古典中国語の語彙がすべて日本語に入っているわけではないが、よく使われる語彙は、すでに日本語になっているものが多い。つまり、それらの語彙は日本語の辞書に登録されているわけである。以下で、もう少し具体的に述べる。

日本語の形態素解析ソフトである『茶筌 (ChaSen) <sup>[2]</sup>』『和布蕪 (MeCab) <sup>[3]</sup>』は解析のために電子化された日本語辞書を用いる。例えば、国立国語研の提供する『UniDic<sup>[4]</sup>』などがそれである。

このような辞書には、漢文の「私」の連用修飾語用法に当たる「私 (ひそかに) に」や、動詞用法に当たる「私 (わたくし) する」という語彙が収録されていても不思議ではない。よって、これらの辞書に登録された語彙情報を用いれば、わざわざ漢文資料から多機能語リストを作らなくても……ということになる。

日本語辞書の有用性をもう 1 つ付け加えれば、日本語の辞書における漢字に対する訓 (例: 「私」に対する「ひそかに」など) は、古典中国語において当該の文字で表される語に対する意味注記 (=日本語で、意味についてのメタ情報を付与している) になっているということである (ごく簡便な対訳レベルの注記であるにせよ)。

上で述べたことは単なる推測ではない。現実、「東アジア古典文献コーパスの研究」共同研究班のメンバーである守岡知彦氏は、上述の『和布蕪 (MeCab)』と、日本語解析用 (!) の IPA コーパスに基づく IPA 辞書 (※ ただし、守岡氏自身が独自にチューンしている) を用い、古典中国語解析を行って興味深い成果を挙げ、この方法が有効であることを示している <sup>[5]</sup>。

上で例示した多機能語「私」と「遠」について、日本語との関わりの中で、もう少し考えてみたい。

「私」は語彙的なレベル (レキシコンのレベル) での多機能語であるので、この種の情報は、当該の漢字の「訓」として日本語辞書に登録されている可能性は高い。

しかし、「遠」は、統語的なレベル (シンタックスのレベル) で、「形容詞 > 動詞」の転換が行われているので (むろん、古典中国語のシンタッ

クスにおいて、である)、日本語の辞書はこの情報を収録していないだろうと推測できる。「とおしとする」というフレーズは、日本語では、漢字の訓として成立しないからである。

ここから、漢文資料から多機能語のリストを作成する時、このようなシンタックスのレベルの現象に着目してそれを行うのであれば、それなりの価値があるかもしれないという結論を出せる。

ただ、ここに 1 つ興味深い資源がある。それは漢和辞典の訓読索引である (まだ、電子資源化されたものはなかりうと思われるが)。「とおし/とおい」の他に「とおしとする」を見出しにして、どちらからも「遠」を得ることができる訓読索引があれば、「遠」の多機能性に関する情報が得られる。そして、そのような漢和辞典は現実存在する。

『学研漢和大学辞典』(藤堂明保編、学習研究社)の音訓索引は、「とおい」「とおしとする」を見出しにして、「遠」が得られる (参考までに、同索引では「ひそかに」「わたくしする」も見出しになっている、「私」が得られる) <sup>[6]</sup>。

## 回 4 送り仮名から何が得られるか(2)

### ※ 4.1 格フレームの抽出の可能性

山崎 2007 で述べた (概略は上述) ように、返り点を手がかりに、動詞・前置詞・助動詞・否定辞の作用域を、機械的に決定することは可能である。

加えて、動詞の作用域の中に送り仮名として「ヲ」「ニ」などを付されている語があった場合 (語の切れ目は、これらの助詞が送られていることによって認識できる)、それを、当該の動詞の取る項として抽出できる。

さらに、その取り出した項を、助詞 (送られた日本語の助詞は、ある種の「格」を表していると考えることができる) とともに記述することにより、格フレームができあがる。

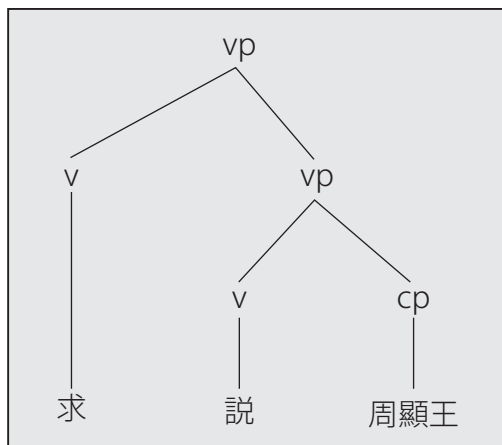
以下の例は、『史記列傳・上・増補版』(重野安繹校訂、漢文大系第 6 巻、富山房編集部編輯)

による（現在の学校教育で用いられる返り点の用法とはいささか異なる方式を用いていることに注意）。返り点を下付添字、送り仮名を上付添字で示した。

- 4a) 少<sub>レ</sub>ト<sub>シテ</sub>之ヲ (原文)
- 4b) 少：\_\_ヲ (動詞：出現した格)
- 4c) (之)ヲ(( )内は、その格を伴った名詞)
- 5a) 説<sub>ク</sub>周ノ<sub>ニ</sub>顯王<sub>ニ</sub>
- 5b) 説：\_\_ニ
- 5c) (周ノ<sub>ニ</sub>顯王)ニ
- 6a) 號<sub>ス</sub>奉陽君<sub>ト</sub>
- 6b) 號：\_\_ト
- 6c) (奉陽君)ト

(a)で示された文が与えられた時、点を手がかりに動詞とその作用域を決定し、その中で、送り仮名の助詞を手がかり（※助詞「ノ」は体言を連結する機能を持つので、この助詞を動詞の項の標識としない処理が必要である）にして、(b)で示したように、その動詞がどのような格（現代日本語の文法でよく使われる「ヲ格」「ニ格」などの用語が便利か）を持つ項をとるかを抽出する。そして、必要なら、それらの項にはどのような名詞が現れるかを示すこともできる（(c)の( )内の名詞がそれに当たる）。

埋め込み構造を持った文（樹形図で示せば下図のようになる）は次のように解析される。



- 7a) 求<sub>ム</sub>説<sub>ク</sub>周ノ<sub>ニ</sub>顯王<sub>ヲ</sub>
- 7b) 求：\_\_ヲ
- 7c) vp[説[周ノ<sub>ニ</sub>顯王]ヲ]

「求」という動詞は、ヲ格を伴う動詞句（vp[ ]で示した）を項として取る、と解析され、この後、vp[ ]の内部は、上述の(5)と同じ構造なので、上記のように解析される。

また、次のような文も、同様に解析される。

- 8a) 事<sub>ヘテ</sub>師<sub>ニ</sub>於<sub>レ</sub>齊
- 8b) 事：\_\_ニ, \_\_ φ
- 8c) (師)ニ, vp[於[齊]]

「事」という動詞は、ニ格を伴う名詞句と φ 格を伴う動詞句（=格助詞を伴わない動詞句）を項として取る、と解析される。もちろん、「於」が古典中国語の前置詞句であることを知識として持っていれば、第二項は、“前置詞「於」を伴う名詞句”として解析をここで止めることも可能である。でなければ、次のように、vp[ ]内が更に解析される。

- 9a) 於<sub>レ</sub>齊
- 9b) 於：\_\_ φ
- 9c) (齊) φ

#### ※ 4.2 このような格フレームは必要か

このような格フレームは有益であると思う。しかし、§ 3.2（「しかし我々はすでにこの情報をもっている」）で述べた見解を、ここでも適用することができる。つまり、この格フレームが持つ情報が他から得られるのであれば、わざわざ時間をかけて漢文資料から格フレームを構築する必要はあるまい、ということである。

古典中国語に使われている動詞は、漢字を媒介にして、日本語の類似の意味をもつ動詞とリンクさせることができる。以下の通りである。

- 10a) 説>説(と)く



10b) 求>求（もと）める

10c) 事>事（つか）える

意味が類似しているからといって、取る項の数と性質が同一であるとは限らないのだが、そう大きく食い違わないであろうことは推測できる。

好都合なことに、日本語動詞に関しては、Web上の言語資料から格フレームを自動抽出し格フレーム辞書を構築する仕組みが、すでに実用化されている<sup>[7]</sup>。よって、「漢字を媒介にして、日本語の動詞の格フレームを古典中国語の動詞に適用する」という手法は、これを試行する条件はすでに整っていると考えられるし、また試行する価値があるように思える。

もしも、これらの既存の資産を活用して成果が挙げられるのであれば、わざわざ漢文資料から格フレームを構築する作業は、緊急の要があるものとは言えない。少なくとも、省力化のためには、既存の資産から得られる情報と得られない情報とを正確に見極める必要があると言える。

前節で挙げた例で言えば、「求」が「ヲ格」の項を取ること、そして、その項が動詞句であったりすることなどは、語彙レベルでの情報であり、日本語の資源から容易に得ることができる。

しかし、いっぽう、日本語との対比では得られない、古典中国語のシンタックス特有の構造というものも存在する。次に、その一例を示す。

(8) a で例示した、動詞「事」が構成する動詞句の構造は、返り点だけを手がかりにする山崎(2007)の方式によれば、右上図のように構造化される。

この構造化は、言語学的に見て、少し不十分である。返り点に拘らず、ごく常識的な枠組みでIC分析（直接構成素分析）をすれば、右下図のようになろう（P=Phrase, n=名詞、v=動詞、prep=前置詞）。

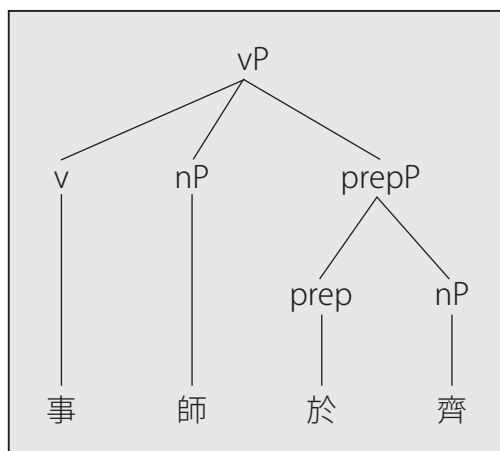
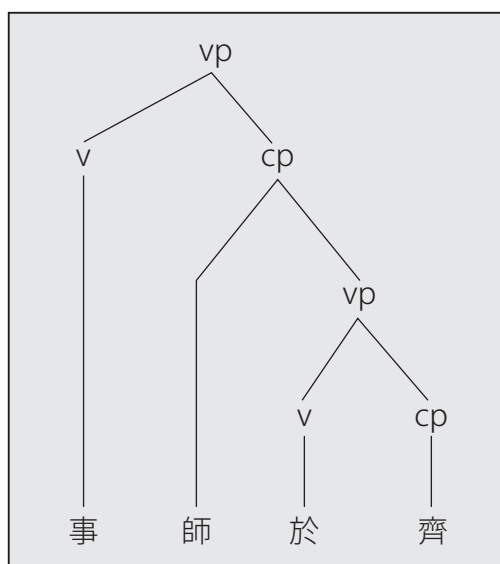
つまり、この構造は、「師」が、動詞「事」や前置詞句の姉妹要素であると解析するのが正しいということである。しかし、これも、「師」に付された送り仮名「ニ」を手がかりに、この語を1つの項として切り出す操作を設定することは可能

である。

ここで言いたいのは、古典中国語の動詞「事」が構成しうるこのような動詞句の構造は、日本語の「事（つか）える」からは得られない、ということである。このようなシンタックスのレベルでの情報は漢文資料からでないと思われ得ない。

## 回 5 語彙的信息と統語的信息

ここまで、送り仮名・返り点付き漢文資料からどのような情報が得られるか、そしてその情報は有益か、を概観してきた。そして、同時に、その



有益な情報の多くは、既存の資産（例：日本語の辞書、漢和字典など）から得ることが可能であることを示した。

ただ、上述したように、日本語の辞書（IPA 辞書、格フレーム辞書……）から得られるのは、個々の語彙の特性に関わる情報、もう少し具体的に言えば、その語彙がどのような意味であるか、どのような項を取りうるか、などの情報である。構造的に上位の統語論レベルの情報になると、既存の資産からそれを得ることは難しいということも示した（§ 4.2）。

以下では、語彙レベルではない、シンタックスレベルの情報、漢文資料からどのように得られるのか、「述語の階層性」に的を絞り、否定辞の作用域と助動詞の作用域を例にとって考えてみたい。

## 回 6 述語の階層性を学習する資料としての漢文資料

### ◆ 6.1 否定辞の作用域と助動詞の作用域

否定辞の作用域と助動詞の作用域、つまり、否定辞や助動詞を機能的な主要部として形成される句構造の範囲を決定することは、「漢文自動解析システム」にとって大きな意味をもつ。これが決定できなければ、SVO の基本語順を持つ（＝述語が文末に来ない）古典中国語を句に分割することはできないからである。

以下の例のように、句末に多用される助詞「矣、也」などがあれば、作用域は少なくともその直前までに閉じられるので、範囲の決定は容易である（先行する否定辞・助動詞の作用域を [ ] で示した）。

- 11a) 燕固弱國、不[足畏]矣。《史記・蘇秦》  
 11b) 此可[以說當世之君]矣。《史記・蘇秦》

しかし、次の例は、範囲の決定が難しい。

- 12a) 乃閉室不[出]、出其書遍觀之。《史記・蘇秦》

- 12b) 秦甲必不敢[出於函谷以害山東]矣。《史記・蘇秦》  
 12c) 未嘗[倍泰山絕清河涉勃海]也。《史記・蘇秦》  
 12d) 則諸侯莫不[西面而朝於章台之下]矣。《史記・蘇秦》

12a の最初の「出」は自動詞であり目的語を持たないので、否定辞がどこまでかかるのかを判定する手がかりに乏しい。12(b) は「出於函谷」「以害山東」という 2 つの動詞句に助動詞「敢」がかかっているが、この作用域を途中で閉じてしまわないようにする判断が難しい。12(c)(d) は否定辞が複数の動詞句を作用域にしている例である。

### ◆ 6.2 述語の階層性

正しい作用域の決定のためには、大雑把に言って、次の 2 つの方法が考えられる。

方法 1：語彙的な情報を元にして、考えられる作用域をいくつか決定し、それぞれの「意味」を考え、前後と矛盾を来さない候補を選ぶ。

方法 2：語彙的な情報を元にして、考えられる作用域をいくつか決定し、作用域の中に「形式的に許容されない構造」が存在する候補を排除する。

素朴に考えれば、方法 2 が容易である。以下で、方法 2 の実現の見通しを検討してみたい。

南（1974）は、日本語において述語の階層を提案している。南（1974）によれば、日本語の文構造には A 類 / B 類 / C 類 / D 類の四つの段階があり、そのうち、A～C 類が従属節に現れるとしている。簡単な例を次に示すが、この階層自体は古典中国語の分析とは関係ない。

- [玉乗りをしながら]<sub>A 類</sub> 火を吹いた。  
 [玉乗りをしたので]<sub>B 類</sub> 見物人が集まった。  
 [玉乗りをしたが]<sub>C 類</sub> 誰も注目しなかった。

南の分類で興味深いのは、この分類自体よりも、A類従属節内にB類・C類従属節が生起できず、B類従属節内にC類従属節が生起できないという、構造上の制約としての階層性が主張されていることである。

このような階層性——ある構造  $\alpha$  の内部には、別の構造  $\beta$  は生起できない——を発見できれば、上述の方法2の手がかりが見いだせる。

例えば、下記の例のゴシック体の語（接続詞やある種の副詞が多い）は、否定辞「不」の作用域を閉じるシグナルとして使えるのではないかと思われる（筆者が試みに《史記・蘇秦列傳》をざっと見てみただけの印象なので、科学的な根拠はない）。

- 13a) 而不[能以取尊榮]雖多亦奚以為。《史記・蘇秦》
- 13b) 奉陽君妒而君不[任事]是以賓客游士莫敢自盡於前者。《史記・蘇秦》
- 13c) 不[與]則棄前功而受後禍。《史記・蘇秦》
- 13d) 韓魏不能[支秦]必入臣於秦。《史記・蘇秦》
- 13e) 臣竊量大王之國不[下楚]。然衡人怵王交彊虎狼之秦以侵天下、《史記・蘇秦》
- 13f) 卒有秦患、不[顧其禍]。夫挾彊秦之勢以內劫其主、《史記・蘇秦》
- 13i) 秦之不[能害燕]亦明矣。《史記・蘇秦》

しかし、次のような例もあり、単純な文字列だけでは決定できない場合もあることが予測される。

- 14a) 則諸侯莫不[西面而朝於章台之下]矣。《史記・蘇秦》
- 14b) 民雖不[佃作]而足於棗栗矣。《史記・蘇秦》
- 14c) 不[戰]而地已削矣。《史記・蘇秦》

(14a) は、接続詞「而」が否定辞「不」の中にある例、(14b)(14c) は、その直前で作用域が閉じている例である。

いずれにせよ、これらを量的データとして学習

するためには、§2.3 で示したような、返り点を手がかりに漢文資料に樹構造のメタ情報を付与したコーパスが必要ではないかと思われる。筆者がここでイメージしているのは、“The Penn Chinese Treebank”（現代中国語の文に、詳細な品詞情報と樹構造情報を付与したコーパス）<sup>[8]</sup> のようなコーパスである。

## 回 7 結び

この文章では、送り仮名・返り点付き漢文資料から機械的に生成されうる電子資源にはどのようなものが考えられるか、そして、それは「漢文自動解析システム」の開発にどう役立つか、また、それは既存の資源で十分ではないのか、などの諸点を検討した。

その結果、「漢文自動解析システム」に役立つ多くの情報（特に、語彙レベルの情報）は、既存の資源から得られるのではないかという感触が得られた。しかし、より統語的レベルに近い情報（例：否定辞や助動詞の作用域）を得るためには（そして、できあがった解析システムを評価するためには）、樹構造を持ったコーパスは一定の用途があるのではないかとの見方を示した。

## 謝辞

本論考で述べた考えをまとめるにあたっては、「東アジア古典文献コーパスの研究」共同研究班の諸氏との議論が非常に助けとなった。特に、守岡知彦氏にはいろいろとご教示をいただいた。ここに記して感謝したい。ただし、この文章の誤り等の責任はすべて筆者にある。

## 注

- [1] 安岡 2008 (§ 4.1 および 4.2) を参照。
- [2] <http://chasen.naist.jp/hiki/ChaSen/>
- [3] <http://mecab.sourceforge.net/>
- [4] <http://www.kokken.go.jp/lrc/index.php?UniDic>
- [5] 守岡 (2008) による。

- [6] 膨大な情報量を誇る『大漢和辞典』（諸橋轍次編、大修館書店）では、この情報は得られない。
- [7] <http://www30.atwiki.jp/corpus-ling/pages/35.html>
- [8] <http://www.cis.upenn.edu/~chinese/ctb.html>

## 参考文献

- 南不二男 . 1974. 『現代日本語の構造』, 東京 : 大修館書店 .
- 守岡知彦 . 2008. 「MeCab を用いた古典中国語の形態素解析の試み」, 『情報処理学会研究報告』, 2008-CH-79, Vol. 2008, No. 73, pp. 17-22.
- 安岡孝一 . 2008. 「漢字情報学の構築」 共同研究班報告 .
- 山崎直樹 . 2007. 「訓点付き漢文の返り点から統語情報を導出し XML で構造化する試み」 『漢字文献情報処理研究』, Vol. 8, pp. 73-82.

# 今だからこそ！ 日本語 IME を見直そう

日本語を入力するときに当たり前のように使っているのが、日本語 IME である。この必要だけども目立たない縁の下の力持ち的存在が目立つのは、思いもかけない変換結果を連発したときである。

Windows VISTA と Microsoft Office2007 が発売された際、不幸にもこの「IME が目立つ」現象が起こってしまった。これについては既に Web で様々な報告があがっているが、試してみると色々と感じになることができた。そこで本誌でも一つ特集を企画してみることにした。

昨今のバージョンアップのおかげで、IME の Unicode への対応は当然のように進んでいる。おかげで文字の入力は随分と楽になったように思う。しかし、そこばかりに目を奪われないで、各 IME が日本語 IME 本来の目的である「いかに日本語を効率よく入力するか」についても、もう一度見直してみよう機会だろうと思った次第である。

本誌の情報が読者諸賢の手助けとなれば幸いである。

## Contents

IME について	二階堂善弘	30
MS-IME	山田 崇仁	32
ATOK	山田 崇仁	39
ことえり	上地 宏一	47
ウェブベース日本語 IME	秋山陽一郎	52

# IME について

二階堂 善弘（にかいどう よしひろ）

## ◇ 意識されなくなった IME

パソコンを使う上での重要なアプリケーションであるにもかかわらず、IME（インプット・メソッド・エディター。あるいは単に、インプット・メソッド、また IM とも）はあまり意識されないツールになってしまっている。

初歩的な授業においても、IME を使って日本語なりの言語を入力していることについて、ほとんど知っている学生はいない。彼らは「パソコンで打つ」「ワープロを打つ」という言い方をする。つまり「MS-IME を使って、Word に文字を入力している」という風にとらえていないのだ。そのため、メモ帳や Excel に入力している場合でも、IME を使っているのにそれを意識することはない。下手をすると、その時ですら「ワープロを打つ」と称している。

これは、やはり OS 側の変化が大きいと思う。大半のユーザは、Windows パソコンを買ってきて、そこに付属している MS-IME を使う。その中身もほとんど調節せずに、そのままの状態を使っている。だから授業で MS-IME の辞書機能を使って、変換できる単語の候補を変えてみると、驚いたりするのである。そもそも、IME というアプリケーションを理解していないのだ。

また、IME の種類も少なくなったことも、意識されなくなった原因の一つであろう。いま Windows で日本語 IME を使おうとすれば、MS-IME か ATOK 以外の選択はほとんどないと言ってよい。ATOK とて、一太郎の付属ソフトであるとも思っていないだろうか。

かつて MS-DOS の時代は、まだ IME を FEP と呼んでいたと思うが、VJE があつたし、ATOK も強かつた。松茸もあつて、この三者を合わせて「ご三家」の称もあつたほどだ。WX シリーズも人気があつた<sup>[1]</sup>。Mac の世界でも、漢字 Talk から Mac OS にかけて、付属の IME である「ことえり」と、また優秀な変換エンジンである EGBridge などがあつた。ATOK もいまは選択肢の一つになっている。

これとは別に、中国語などを生業とする我々の業界では、GB や Big5 の漢字をアプリケーションに打ち込むための IME とフォントの組み合わせが重要であつた。そのために、かつては Chinese Writer や cWnn などの中国語入力ソフト、また南極星や双橋や加盟などのマルチリンガル IME をいろいろ試していたものである<sup>[2]</sup>。

そういった、OS とアプリケーション、それに IME の組み合わせというのは、一般ユーザにとっても大きな問題であつた。しかし、Windows が 95 / 2000 から XP / Vista と変化するにつれて、標準に搭載されている日本語 IME や中国語 IME で、大半の処理が可能になり、わざわざ IME を購入する必要がなくなった。IME を開発していた会社も、売れなくなったためか、すっかり発売を控えるようになってしまった。かくて IME は、OS に付属してくるのが当たり前になり、「ほとんど意識されないツール」になってしまったわけである。いまや ATOK をわざわざ購入して使っているユーザは、「日本語にこだわりのあるユーザ」とみなされるに至っている。いまや Mac でも、ほとんどことえりを使うだけで用は足りると思う。

## ◆ 近年の変化

ただその状況も、Windows Vista が登場した後に微妙に変化してきた。

Vista になってから、若干アプリケーションとの組み合わせが複雑になってきたことがその背景にある。本誌の前号にも書いたが、Vista は Home / Business などと、Ultimate / Enterprise などでは多言語対応を中心に、やや性能が異なっている。さらに MS Office system 2007 をインストールすると、組み合わせによりまた差が出る。Office には、Multi-Language Pack 2007 もあり、これに Proofing Tools が含まれている。これらをインストールすることにより、多言語対応の度合いも異なってくる。

この件で MS-IME については、漢情研 BBS でも話題になった「MS-IME ってそんなにダメなの？」問題がある<sup>[3]</sup>。まず ITmedia サイトに瓜生聖氏の「最近の『MS-IME』は目に余る——よろしい、ならば『ATOK』だ」<sup>[4]</sup>という記事が掲載され、そこに古川亨氏のブログ「MS IME さらに…お馬鹿になっていく」<sup>[5]</sup>が引用されていた。これらの記事によれば、MS-IME はバージョンアップして、むしろ日本語変換がおかしくなったと指摘されているのである。

そこに出ている例だが、確かに「腱鞘炎」といった単語が変換できなかつたりする。自分も普段 ATOK ユーザなので、あまりこの点については注意していなかった。

とはいえ、漢情研 BBS で清水哲郎氏が指摘しているように<sup>[6]</sup>、現在 MS-IME は二種類存在することが、そもそも意識されていない。

つまり、Vista に標準で付属している MS-IME

2007 と、Office に付属する MS Office IME 2007 とがあり、両者は異なるものなのだ。この両者は開発元が異なっているようであり、Office IME 2007 の方は中国で開発されているとのことである。「腱鞘炎」が変換できないのは、むしろこの Office IME 2007 の方だ。

この両者のどちらがデフォルトで設定されているかで、議論は相当に異なってくる。ただ大学などの現場でも、どちらがどうかという意識をあまりしていないのではないかと。要するに、Vista と Office の組み合わせが厄介になってしまったのだ。しかしいずれにせよ、やはりユーザの側で「IME はアプリケーションである」という認識を、もう一度持つ必要があろう。

マルチリンガル IME についても、最近多くのアプリケーションを作成している Google の Google IME などが開発されており、このあたりでも MS の牙城を崩せるかどうか、興味のあるところである。

## 注

- [1] このあたりの詳細については、「日本語入力プログラムの歴史 (MS-DOS ~ Windows)」([http://www4.airnet.ne.jp/koabe/com\\_inet/im/history1.html](http://www4.airnet.ne.jp/koabe/com_inet/im/history1.html)) 参照。
- [2] 漢情研編『電脳中国学』(好文出版) 46 ~ 56 頁参照。
- [3] 「自己紹介と雑談の部屋 4」2008 年 5 月 22 日の千田大介氏の投稿による。
- [4] <http://plusd.itmedia.co.jp/pcuser/articles/0805/22/news055.html>
- [5] <http://furukawablog.spaces.live.com/blog/cns!156823E649BD3714!9079.entry>
- [6] 注 [3] より始まる一連のスレッドへのレス。清水氏の書き込みは 2008 年 5 月 22 日。

# MS-IME

山田 崇仁（やまだ たかひと）

## ◆ 二つの Microsoft IME

Microsoft IME（マイクロソフト・アイエムイー）とは、Microsoft が開発し、Windows に標準搭載されているインプットメソッドエディター<sup>[1]</sup>である。Windows は多言語対応を念頭に置いて開発されているため、IME を必要とする各言語毎に IME が用意されているが、本稿では日本語入力システムとしての IME のみを対象とする。

元々 Windows 用の IME は、エー・アイ・ソフトが開発した日本語入力システムである WX II（Windows3.1 用）および III（Windows95 用）をベースとして開発されたものである。その後も次第に改良を重ね、Windows や Microsoft Office のニューバージョンリリースとともに MS-IME も新版が提供されてきた。現在の最新バージョンは、Windows VISTA に標準搭載されている Microsoft Input Method Editor（以下、MS-IME と略）と Microsoft Office 2007 のインストール時に導入される Microsoft Office Input Method Editor 2007（以下、OFFICE-IME2007 と略）の二種類ある。

前者は、MS-IME の系譜を引き継ぐバージョンであるが、後者は Microsoft Office 2007 の開発にあわせて、新たな変換エンジンを搭載した登場した新 MS-IME といってもよいバージョンである。

通常、Microsoft Office 2007 のインストール時に、MS-IME は選択可能な日本語 IME の一覧から消される（削除ではなく、テキストサービスとして使用するという選択が解除されている状態）ため、あまり気がつかれることがない。

ところが、この二つの IME の違いに注目され

るようになったのは、両者の振るまい（変換効率など）が明らかに異なる、特に OFFICE-IME2007 の振る舞いが、これまでの MS-IME とは（悪い意味で）一線を画したものであるとする評価が出てきたからである。

この二つの IME については、すでに本誌前号にて清水哲郎氏が報告しているが<sup>[2]</sup>、本稿ではそれぞれの機能と変換精度について報告したい。

## ◎ MS-IME の追加（復活）

上述のように、MS-IME は MS Office 2007 のインストール時に消されてしまう。ただし、ハードディスク上から削除されてのではなく、単に選択可能な IME の一覧から排除されているだけである。そのため、以下の手順で MS-IME を復活することが可能である。なお、MS-IME に限らず、Windows 上での IME の管理は [テキストサービスと入力言語] 上で行う。そのため、はじめにこのダイアログボックスを呼び出す必要がある。方法は以下の通り（Windows VISTA Ultimate で確認）。

1. IME の言語バーを表示した後、言語バー上で右クリックしてメニューを表示し、[設定] を選択する。コントロールパネルの [地域と言語のオプション] を選択し、[キーボードと言語] のタブをクリックして画面を切り替え、[キーボードの変更] をクリックしてもよい。
2. [テキストサービスと入力言語] のダイアログボックスが表示されたら、[インストールされているサービス] の [追加] ボタ



ンを押す (図 1)。

- [入力言語の追加] のダイアログボックスが表示されたら、日本語の項目にツリー表示されている各 IME から、Microsoft IME のチェックボックスにチェックを入れる (図 2)。
- [OK] をクリックして各ダイアログボックスを閉じる。
- 言語バーの IME 選択一覧に、Microsoft IME が表示されたら成功である。



## ◆ 二つの IME の辞書

### ◎ MS-IME の辞書

標準の辞書以外に、郵便番号・単漢字・話し言葉顔文字・人名地名・カタカナ語英語・記号の辞書がそれぞれ搭載されている。このあたりの辞書に関しては、ATOK でもほぼ同様の機能を持つ辞書を搭載している。ここから言えば、今の日本語 IME に求める語彙のカテゴリをまとめると、この位に整理されてしまうということなのだろう。

### ◎ OFFICE-IME2007 の辞書

OFFICE-IME2007 の辞書も、基本的には MS-IME と変わりがない。ただ、サイズはこちらの方が大きいので辞書のカスタマイズはそれなりに行っていると考えられる (OFFICE-IME2007 では、図 3 のように変換時に印刷標準字体・簡易慣用字体が明示されるが、辞書サイズの増加はこのような部分にも理由があるのだろう)。

OFFICE-IME2007 は変換候補の末尾に複数の辞書が列挙され、それを指定することで辞書を切り替えることができる (図 5)。ファンクションキーを使用して同様の機能を提供する ATOK とどちら

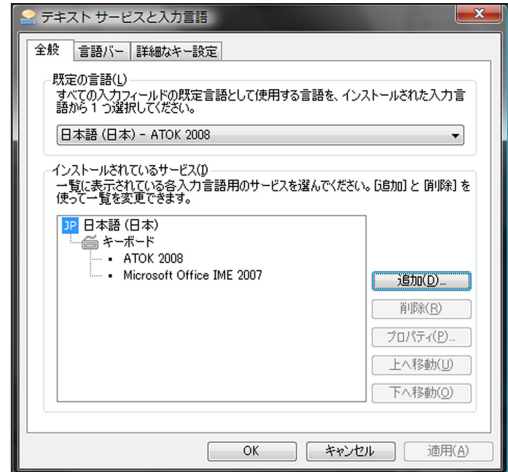


図 1

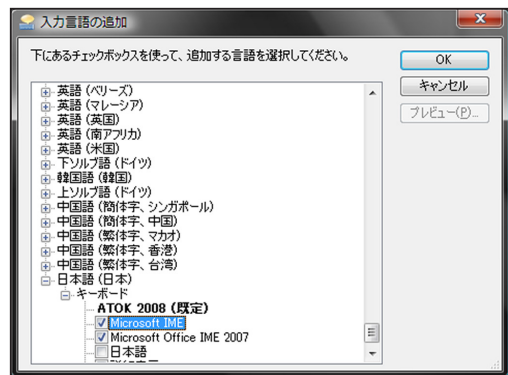
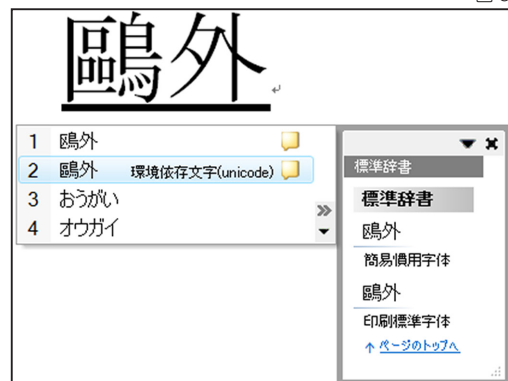


図 2

図 3



が使いやすいかは別にして、普段使わない言葉を別途辞書として分けておいて、必要に応じて切り替えて使用する MS-IME の方式も案外便利である。

辞書ファイルを対象としてみた場合、その大きな違いとして注目されるのが「Outlook 連絡先辞

同	撞
1 導	1 撞
2 懂	2 桐
3 撞	3 瞳
4 瞳	4 童
5 萄	5 藤
6 働	6 単漢字...
7 //	7 人名地名...
8 々	8 英語に変換...
9 全	9 記号...

図 4 (左)・図 5 (右)

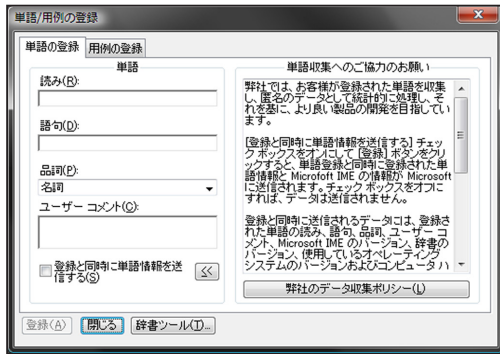


図 6



図 7

書「Outlook グローバルアドレス一覧辞書」など、Outlook の名称が冠された辞書群である。これらの辞書は、変換に際し Outlook の連絡先に登録した姓や名を優先順位が高い変換候補として表示するための辞書である。Outlook を利用しているユーザーの場合は便利な機能かもしれないが、筆者のように Outlook を全く使っていないユーザー

にとっては、レスポンスの低下を招くので注意を必要とする<sup>[3]</sup>。

## ◎ 辞書への単語登録

日本語 IME をより自分に近づけた使い勝手にするために、やはり辞書を鍛える必要がある。

MS-IME の単語登録は、単語登録ウィンドウを表示し (ATOK がオンの状態で Ctrl キーと F7 キーを同時に押すと表示される)、単語・読み・辞書・品詞などを指定して登録すればよい。JIS X 0208 収録字以外の文字、すなわち Unicode に登録されているすべての文字を登録することが可能である。ただし、Windows VISTA 環境上では、Internet Explorer7 上での単語登録に不具合が出るとのこと<sup>[4]</sup>。注意が必要である。

多くの単語を一括して登録したい場合には、辞書ツールを利用して一括登録することになる。

あらかじめ、以下の形式のテキストファイルを作成しておく。

- 先頭の行に、以下の文字列を入れる。
- 以下、「よみ」「単語」「品詞」「コメント (必要に応じて入力)」の順で入力する。各項目の間は水平タブで区切っておく。

おか [氷平] 丘 [氷平] 地名その他 [氷平] 丘の  
避諱字。もちろん、孔子その人の諱。

後はファイルの名前をつけて保存すればよい。

JIS X 0208 収録字以外の文字 (Unicode のすべての文字)を使用する場合は、保存する文字コードを、UTF-16 リトルエンディアンにする必要がある。

後は、辞書ツールを起動し、メニューの「ツール」→「テキストファイルからの登録」を選択し、ダイアログボックスが表示されたら、読み込む単語ファイルを指定して「開く」を実行すればよい (図 6: MS-IME の単語登録画面。図 7: OFFICE-IME2007 の単語登録画面)。

これとは別に、MS-IME, OFFICE-IME2007 では、誤変換や登録した単語を Microsoft に送付す



# 今だからこそ！ 日本語 IME を見直そう

また、ATOK の文字表のように、Unicode 文字表の一部のカテゴリ名が英語で表示されていることもなく、日本語に翻訳されて表示されていることは、利用者にとって便利だろう。

- Unicode (基本多言語面)
- Unicode (追加多言語面)
- Unicode (追加漢字面)
- Unicode (15 面)
- Unicode (16 面)
- シフト JIS
- JIS X 0208
- JIS X 0212
- JIS X 0213 (1 面)
- JIS X 0213 (2 面)

## ◎ 総画数・部首

漢和辞典などでおなじみの、画数から目的の文字を探すものである。前者は総画数順に、後者は「部首 (各週順に配置) →同一部首内画数順」に文字が配置されている。

図 9

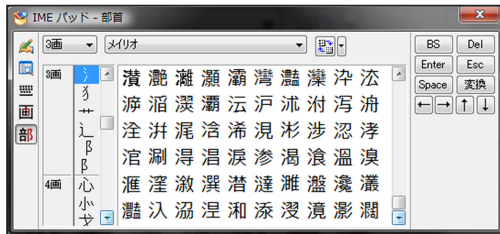
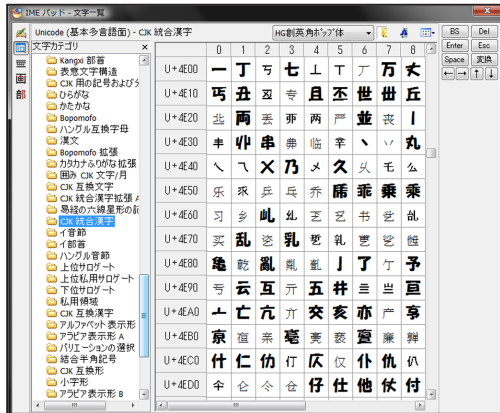


図 10



いずれも、漢字は JIS X 0208, JIS X 0212, JIS X 0213 に対応している。ただし、JIS X 0213 に関しては収録漏れとおぼしき文字が存在したり、部首内部が文字コード順(同一カテゴリ内画数順)に並んでいるため注意が必要である。

例えば、図 9 に図示した「さんずい」の項目では、CJK 統合漢字領域の「灑 (U+7064)」「灑 (U+704E)」の後に、拡張領域 A に属する「泚 (U+3CC3)」「浚 (U+3CD2)」が並び、その後の部分に「灑 (U+7069)」が配置されて、拡張領域 B の「汧 (U+00023C7F)」「湍 (U+00023D00)」がそれぞれ配置されている。

一文字や一筆ごとに認識関係を設定する必要がある手書き入力に比べ、こちらの方は単純かつ機械的に配置するだけのはずだから、収録漏れというのは単純に実装の不具合でしかない。このあたりの中途半端な対応の早急な訂正をお願いしたいところである。

Unicode 文字表・総画数・部首共々、文字表のフォントを切り替えて表示することができるのだが、選択したフォントによってはそのフォントに実装されていないグリフであっても代替フォントで表示されてしまうため、フォント毎に実装されているグリフを区別することが難しいという問題が生ずる(図 10。フォントに HG 創英角ポップ体を指定して、カテゴリを CJK 統合漢字に切り替えたもの。太字が HG 創英角ポップ体で実装されているグリフ。細字が FontLink 機能で他の実装されているフォントから代替されて表示されている部分。)<sup>16)</sup>

もっともこれは ATOK の文字パレットでも同様におこるため、実際には Windows の機能である FontLink に起因する問題だと判断されるため、IME を責めるわけにもいかないが、文字毎に実装されているグリフを確認したい場合には少々使いづらくなってしまうのも事実である。

## ◇ 入力テスト

そもそも、IME の特集を企画した理由は、MS-IME の挙動がどうもおかしい? というインター

読み	正しい変換	MS-IME	MS-IME2007
けんしょうえん	腱鞘炎	腱鞘炎	健勝苑
しゃれい	謝礼	謝礼	謝礼
しゃれき	社歴	社歴	洒落気
じょうきをさんしょうして	上記を参照して	上記を参照して	蒸気を参照して
かんじのごへんかん	漢字の誤変換	感じの誤変換	漢字のご変換
しょうかん	使用感	使用感	しょうかん
でんげんしゃだん	電源遮断	電源社団	電源社団
そうしきかいじょう	葬式会場	葬式会場	層式会場
だんしせいと	男子生徒	男子生徒	男子生徒
たしかに	確かに	確かに	確かに
かいげんしょう	怪現象	買い現象	買い減少
きーのながおし	キーの長押し	キーの長尾氏	キーの長尾氏
えいぞうくりえいたー	映像クリエイター	映像繰りえいたー	映像栗栄太一
かんじがこうへんかんされます	漢字がこう変換されます	漢字が高変換されます	漢字が高変換されます
ごくしょう	極小	ごく小	ごく章
きょうしつにてほしゅう	教室にて補習	教室にて補修	教室に手補修
さいかいごうかく	最下位合格	再開合格	再開合格

中島敦：名人伝より

原文	ちょうど彼等の真上、空の極めて高い所を一羽の鶯が悠々と輪を画いていた。
MS-IME	ちょうど彼らの真上、空のきわめて高いところを一羽のとびが悠々と輪を描いていた。
OFFICE-IME2007	ちょうど彼らの真上、空の極めて高いところを一羽のトビが悠々と輪を描いていた。

中島敦：名人伝より

原文	その人の前には我々の技のごとき児戯にひとしいと言った師の言葉が、彼の自尊心にこたえた。
MS-IME	その人の前にはわれわれの技のごとき時宜に等しいといった詩の言葉が、彼の自尊心にこたえた。
OFFICE-IME2007	その人の前には我々の技のごとき時宜に等しいといった詩の言葉が、彼の自尊心に答えた。

ネット上の各種記事であった。その代表が、古川享（元マイクロソフト株式会社社長・現慶應義塾大学教授）氏のブログに記された一連のエントリーである。

- ℃トウのご変換、館無料<sup>[7]</sup>
- IME 2007の甲斐減少<sup>[8]</sup>
- MS IME さらに ... お馬鹿になっていく<sup>[9]</sup>
- MS IME 野甲斐、さらに十章に ... (の怪、さらに重症に ... だってば)<sup>[10]</sup>

ここでは、MS-IME, OFFICE-IME2007の双方とも標準機能のみ（辞書もまっさらな状態）で、どこまで変換が可能か試してみよう。サンプルは、公平を期すために、ATOK・ことえりで共通のものを使用している。サンプルは、古川享氏のブログのエントリー「MS IME さらに ... お馬鹿になっていく<sup>[11]</sup>」などを参照して設定した。

何れの事例も、読みを最後まで入力してから、変換キーを一度押して表示されたものをそのまま

確定したものである。

以上のように、確かに一部の語彙や表現が（辞書や変換の癖を学習する以前の状態で）、MS-IME と OFFICE-IME2007 の間でさえ退化と判断されても仕方がない変換結果を見せることは事実である。ただし、これに対しては、学習を重ねることによってこの状態は解消され、それなりにユーザーの変換意図に応じた変換効率を見せるようになってくるとの指摘もあるので一概には言えないが<sup>[12]</sup>、少なくとも初期状態ではかなり癖がある IME であることは言えるかもしれない。

## ◇ おわりに

以上、MS-IME について紹介した。

筆者は日常的に ATOK を愛用しており、MS-IME は勤務先の共用パソコンで使用する程度のユーザーのため、ディープに使いこなしての変換効率を上げるような使い方はしていない。

そのためこの報告がいささか武断的に過ぎる形になってしまったと思っているが、共用パソコンや原稿を書くに際しての使用感や Web 上での記事から判断する限り、OFFICE-IME2007 は使いこなしてなんぼ！ という性質の IME なのだろう。

もっともこれは MS-IME の評価に対して昔から言われていることであり、初期状態の変換効率は（バージョンがアップしてまじにはなったものの）決して満足のいくものではないという実感がある。

特に共用パソコンのように、変換効率を高めるための学習行為や辞書への単語登録が制限されている環境では、OFFICE-IME2007 は常に初期状態なのであり、絶え間のない学習を繰り返すことで「あなた好みになってゆく」OFFICE-IME2007 の真の実力を見せる日は永遠にやってこないのだ。

そういう意味では宝の持ち腐れとも言えるかもしれないが、これは何も共用パソコンだけではない。パソコンを買い換える（オペレーティングシステムをバージョンアップする）度に、IME の学習結果を移行しなければ、またおかしな変換結果としばらくつきあわないといけないかもしれない。

MS-IME はこの状態を続けるのか？ あるいは

ATOK のように、初期状態でもそれなりに意図した変換結果が表示され、使いこなせば更に手にしっくりくるような IME を目指すべきなのか？

このあたりは IME の設計思想にも連なることであり、一概には言えないのかもしれない。おそらく次のバージョンアップではそのあたりの解決（変換エンジンの精度向上や、これまでの学習効率の反映や辞書のブラッシュアップ）がはかられるはずなので、それを期待することにしたい。

## 注

- [1] 文字入力システム。IME と略される。
- [2] 特集 1：Windows Vista の日本語フォント環境（続）と Vista / Office2007 の IME。
- [3] Outlook のアドレス帳に全く登録がない、あるいは登録が少ない状態で当該辞書を使用する設定にしておくと、レスポンスが極端に遅くなるとの報告が Web 上で散見される。
- [4] 前掲清水氏の文参照。ただし、回避用のアップデートファイルはすでに Microsoft から配布されている。
- [5] 【インタビュー】Vista と Office 2007、かな漢字変換の結果が違うのはなぜ？——IME 2007 の開発者に聞く。  
<http://itpro.nikkeibp.co.jp/article/NEWS/20061204/255812/>
- [6] 実装されていない文字については、微妙に廻りの実装されている文字とデザインが異なるため、完全に判読不能というわけではないが…。
- [7] <http://furukawablog.spaces.live.com/blog/cns!156823E649BD3714!7900.entry>
- [8] <http://furukawablog.spaces.live.com/blog/cns!156823E649BD3714!8312.entry>
- [9] <http://furukawablog.spaces.live.com/Blog/cns!156823E649BD3714!9079.entry>
- [10] <http://furukawablog.spaces.live.com/blog/cns!156823E649BD3714!9304.entry>
- [11] <http://furukawablog.spaces.live.com/blog/cns!156823E649BD3714!9079.entry>
- [12] 猿頁の記事：(25)IME2007 のしつけかた [http://salv.miscnotes.com/2008/02/25\\_ime2007.php](http://salv.miscnotes.com/2008/02/25_ime2007.php) など。

# ATOK

山田 崇仁 (やまだ たかひと)

## ◇ はじめに

ATOK (エイトック、Advanced Technology Of Kana-kanji transfer<sup>[1]</sup>) は、ジャストシステム社の開発・販売による日本語入力システム (以降、日本語 IME、あるいは単に IME と略す) である。

ATOK は日本語 IME の中でも、MS-DOS 時代から長く利用されてきた老舗である。Microsoft IME やことえりなど、各オペレーティングシステムにはそれなりに使える IME が標準でインストールされているにもかかわらず、現在でもなお新版がリリースされ、一定の指示を集め続けているのは、やはりその変換精度や使い勝手に対し、MS-IME に対する優位性を認識している層が存在するからであろう。

その優位性こそが、過去に存在し、またその多くがすでに淘汰されてきた日本語 IME の中で、現在もなお新版がリリースされ、また携帯電話や Windows Mobile といった携帯情報機器、更には Wii や PLAYSTATION(R)3 や PSP(R) といったゲーム機に至る、パーソナルコンピュータ用オペレーティングシステム以外への搭載・販売が幅広く行われている。そういった機器への ATOK の搭載がユーザーに対してのアピールとなり得るのも、やはり変換精度や使い勝手が一定の支持を集め続けているからといえる<sup>[2]</sup>。

では、ATOK のどのあたりが優れているのだろうか？ それについて、日本語入力システムとしての ATOK の特徴と、本会読者向けに、漢字文献を入力する際に、ATOK をどのように使いこなすより便利なのか？ について、書いてみることにしよう。

## ◇ 日本語 IME としての ATOK

### ◎ ATOK の入力システム

ATOK の日本語変換のコアは、ATOK2007 以降に搭載された「ATOK ハイブリッドコア」と呼ばれるシステムである。ジャストシステムによれば、このシステムは

ATOK がバージョンを経るごとに磨いてきた伝統的なアルゴリズムをバックボーンとする“かな漢字変換エンジン”に、「ATOK Lab. (エイトックラボ)」から生み出された日本語処理技術の研究成果を組み合わせた、最新の変換エンジン<sup>[3]</sup>である。

とのことであり、また「ATOK Lab.」とは、こちらもジャストシステムによれば、

最先端の自然言語処理研究を、製品版の ATOK の開発とは独立した形で進めている研究チーム

とのことである。すなわち、ジャストシステムの IME は日本語変換システムを言語の観点から捉え直した研究活動を踏まえた、日本語のための IME と呼ぶうるものである。

最新の ATOK2008 では、更にこのコアが洗練され、いくつかの新しい機能が実装されている。

例えば「次候補連動変換」がその代表である。例えば、「平成 19 年と 20 年」と入力したとき、前者の「19」を「一九」「十九」「壹拾九」に変

換すると、後者はそれぞれ「二〇」「二十」「貳拾」と連動して変換されるようになった。これは年月日を入力する際に重宝するだろう。

ATOK では、これらのシステムを基軸に、従来からの変換用辞書の充実や、ATOK 事態をコアとした各種外部情報とのリンク（これは「ATOK ダイレクト」と呼ばれる）を組み合わせることを試みている。日本語 IME は、コンピュータ用オペレーティングシステム上で日本語を扱う場合、必ず通らなければならない。単純にキーボードからの文字を受けつけて、日本語に変換するだけではなく、日本語 IME を他の環境とリンクさせることで、思考の手助けにするためのツールとして位置づけているようにも見える。

また ATOK は、インストールした直後であってもそれなりの使い勝手（日本語変換が入力者にとって不快感がない（比較的思い通りに変換される）、希望する変換候補が上位にくる確率が高いなど）を示し、使い込むことによって（文法の切り方の癖や、変換候補としてよく使用される漢字の登場順、推測変換の候補など）、更に利用者の手になじんだ形になるという性格を持つ<sup>[4]</sup>。

## ◇ ATOK の辞書

ATOK では、標準状態で複数の辞書を提供している。通常これらの辞書は以下に挙げた 4 つの辞書セットに分割して登録されている（各項目の説明は ATOK に記載されているものを引用した）。

通常は標準辞書セットを使用して変換をするが、読みを入力後に、ファンクションキーの 1（標準辞書）・2（人名変換辞書セット）・3（郵便番号辞書セット）・4（アクセサリ辞書セット）をそれぞれ押すことで、各辞書セットを指定して変換することができる。普段使わない特殊な語彙を利用する際に、このような辞書切り替え手法は非常に便利だろう。無論、自分で作成した辞書を各種辞書セットに追加もしくは新規に辞書セットを作成することも可能である（辞書セットは 10 個まで作成可能）。

## ◎ ATOK の標準の辞書セット・辞書一覧

### ● 標準辞書セット・人名辞書セット

- 標準辞書：ATOK の変換の中心となる辞書です。
- トレンド辞書：標準辞書の変換を補助する、人気スポット・著名人名・ヒット商品名など身近な単語を登録した 2008 年版辞書です。
- 人名辞書：標準辞書の変換を補助する、人名を登録した辞書です。
- 単漢字辞書：標準辞書の変換を補助する、単漢字を登録した辞書です。

### ● アクセサリ辞書セット

- カタカナ語英語辞書：カタカナ語から英語に変換する辞書です。例えば、「ぶろばいだ」から「provider」に変換できます。
- 記号辞書：いろいろな読みから記号に変換する辞書です。記号の名称などについても確認できます。
- フェイスマーク辞書（インターネット版）：いろいろな読みからフェイスマーク「(^)」に変換する辞書です。（単語内に半角カタカナなど機種依存文字を使っていません。）
- フェイスマーク辞書：いろいろな読みからフェイスマーク「(^)」に変換する辞書です。（C）1999 株式会社ビレッジセンター

### ● 郵便番号辞書セット

- 郵便番号辞書：郵便事業株式会社提供 2007 年 10 月 31 日更新のデータを使った郵便番号（3 桁 / 5 桁 / 7 桁）から地名に変換する辞書です。例「107-0061」から「東京都港区北青山」に変換できます。
- 郵便番号辞書：事業所：郵便事業株式会社提供 2007 年 10 月 31 日更新のデータを使った郵便番号（5 桁 / 7 桁）から大口事業所名に変換する辞書です。
- 町名住所変換辞書：町や村の読み（町域名の読み）から住所に変換する辞書です。郵



便事業株式会社提供 2007 年 10 月 31 日更新のデータを使用しています。例「ゆうらくちょう」から「東京都千代田区有楽町」に変換できます。

これらの辞書のうち、筆者は日常的に標準辞書に単漢字辞書を加え、独自の中国史用辞書を複数の辞書セットを作成して運用している。これは、日常の変換ニーズとして、カタカナ英語やらトレンド語やらを使用しないためである。このあたり、要するに「辞書をどう鍛えるか」というのは、変換システムの精度向上と並んで、伝統的に日本語入力システムの使い勝手を向上するための大きな部分を占めており、ユーザーの意図をより濃厚に反映可能な部分であるといえる。

また、標準辞書には、変換時に印刷標準字体・簡易慣用字体の別や、フォントのデザインによっては区別しにくい括弧類の半角全角の別を指摘してくれるなど、より環境に応じたスムーズな入力を助けてくれる機能がある (図 1)。これは MS-IME にも実装されている機能だが、便利なものはアプリケーションの種別を問わず実装されていることはよいことだろう。

### ◎ ATOK の辞書登録機能

上述の通り、辞書を鍛える行為は、日本語 IME に対しユーザーの意図を色濃く反映可能な部分である。この機能を使いこなすには、やはりユーザー個々に独自の辞書を蓄積しておくのがよいだろう。

ATOK の単語登録は、単語登録ウィンドウを表示し (ATOK がオンの状態で Ctrl キーと F7 キーを同時に押すと表示される)、単語・読み・辞書・品詞などを指定して登録すればよい (図 2)。JIS X 0208 収録字以外の文字、すなわち Unicode に登録されているすべての文字を登録することが可能である。但し、字種によっては、登録時にグリフの関係で□で表示される場合もある (登録は正常に行われる) ので注意が必要である。

多くの単語を一括して登録したい場合や、単語にコメントや複数の置換候補を登録したい場合には、辞書ユーティリティィーを利用すればよい。辞



図 1

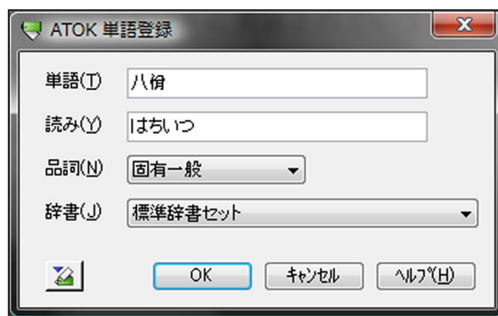
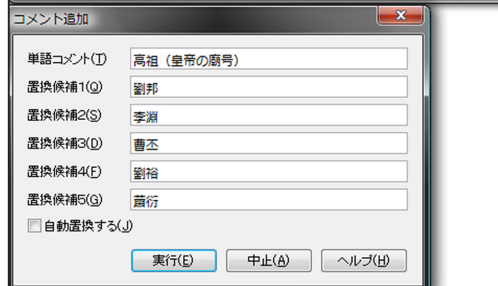


図 2



図 3





この機能の使い勝手はかなり上がったといえるだろう<sup>[5]</sup>。また、認識された文字に真いすカーソルを合わせると、文字に関する情報が表示されるのも便利だ。

### ◎ 文字パレット

手書き入力ですら認識できない文字、例えば JIS X 0213, JIS X 0212 に含まれていない簡体字などの入力はどうすればよいか？ 中国語用 IME を使え！ と回答するのは簡単である。しかし、中国語を解しない、でも簡体字を使う必要がある！ という場合は、どうしようか？ その場合に最後の助けとなるのが、文字パレットから探し出すことである<sup>[6]</sup>。

ATOK の文字パレットは、5つのタブに分かれている。そのうち、左三つのタブは、主に JIS 漢字コード（ただし、JIS X 0208 であることに注意）を中心とした文字種を、項目別に切り替えて表示することができる。左から四つ目の「Unicode 表」は、文字通り Unicode の文字表を各項目ごとに一覧表示する。CJK 統合漢字拡張領域 B の表示まで対応しているのはうれしいところである（図 5）。

また、「文字情報」のボタンをクリックすると、その字が持つ情報（読み・部首・各週・文字コードの文字表番号や小学生の学年別配当など）が表示される。小学生用の学年別の配当が表示されるのは、教育分野に強いジャストシステムならではの機能だろう（図 6）。

ただし問題なのは、文字の表示方法である。Unicode 表ではフォントを切り替えて表示することができるのだが、選択したフォントによっては、そのフォントに実装されていないグリフであっても代替フォントで表示されてしまうので、フォント毎に実装されているグリフを確認することが難しいという問題がある（図 7）。

もっともこれは、Windows の機能である FontLink が働いているためだと思われるので、一概に ATOK ばかりを攻められないのだが、BabelMap<sup>[7]</sup>のように、フォント毎の実装グリフを確認可能な文字パレットもあるのだから、個人的にはこのあたりは改善してほしい（図 8）。

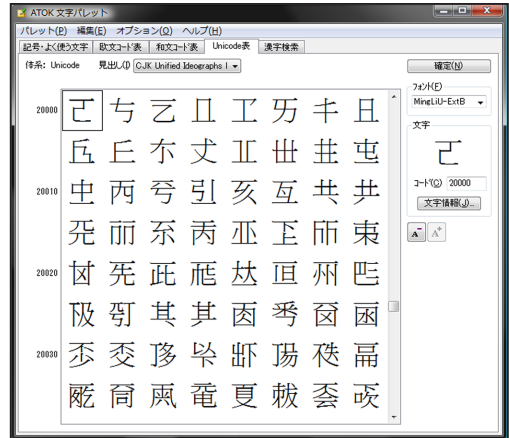


図 5

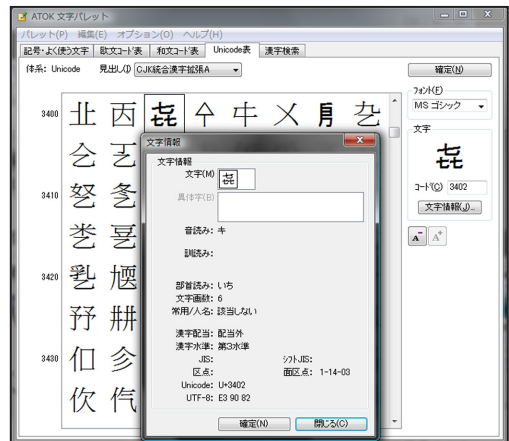


図 6

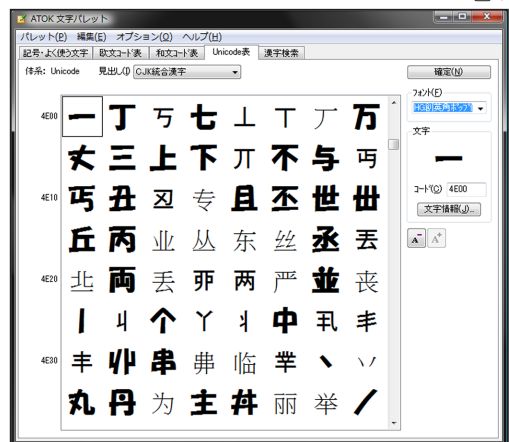


図 7

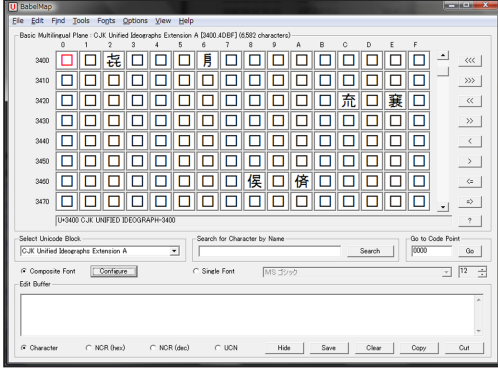


図 8

また、一番右側のタブは「漢字検索」である。これは、漢字のパーツ（部首など）や読み・画数という情報を組み合わせて目的の漢字を検索する機能である。ここで検索可能な漢字は、JIS X 0213, JIS X 0212 の範囲である。ATOK の文字パレットには、MS-IME の文字一覧に実装されている部首・画数のカテゴリが存在しないが、「漢字検索」を利用して同様 + α の機能を利用できる。少なくとも、絞り込みツールとしてはこちらの方が優秀だろう。

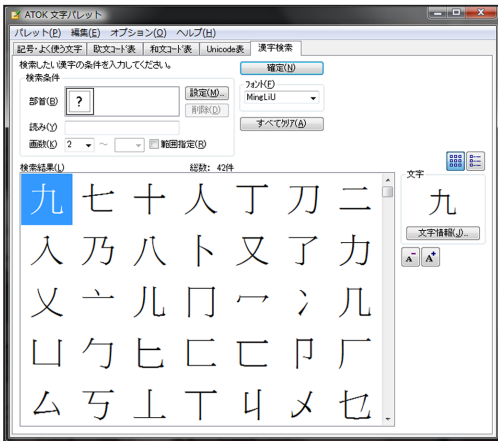


図 9

更に、「漢字検索」で検索した文字は、単純な文字一覧（図 9）と詳細情報（図 10）の二種類の表示方法が可能である。詳細情報ではその字が持つ情報（読み・部首・各週・文字コードの文字表番号や小学生の学年別配当など）が表示される。

## ◆ 入力テスト

ここでは、ATOK の標準機能のみ（辞書もまっさらな状態）で、どこまで変換が可能か試してみよう。サンプルは、公平を期すために、MS-IME、ことえりで共通のものを使用している。サンプルは、古川亨氏のブログのエントリ「MS IME さら

図 10

漢字	画数	部首読み	音読1	音読2	音読(他)	訓読1	訓読2	訓読(他)	常用/人名	配当	漢字水準	真体字	JIS	SJIS	区点	面区点	UNICODE	UTF-8
九	2画	いち	シチ	シツ		なな	ななつ	なの	常用	1年	第1水準		3C37	8EB5	02823	1-28-23	4E03	E4 B8 83
入	2画	いち	チョウ	テイ	チノトウ	ひと			常用	3年	第1水準		437A	929A	03590	1-35-90	4E02	E4 B8 81
又	2画	いち	コウ													2-01-02	4E01	E4 B8 82
ノ	2画	いち	ジョウ														4E04	E4 B8 84
ナ	2画	いち	カ														4E05	E4 B8 85
ル	2画	いち														2-01-01	20089(D840_	F0 A0 82 89
リ	2画	いりいり	ニュー	ジュ	ジュウ	いる	いり	いれるいりよ	常用	1年	第1水準		467E	93FC	03894	1-38-94	5165	E5 85 A5
ロ	2画	ろ	コン										5161	9981	04965	1-49-65	51F5	E5 87 85
リ	2画	えんがまえ	ケイ										5144	9963	04936	1-49-36	5182	E5 86 82
ル	2画	おつこよう	ク	キュウ		この	この		常用	1年	第1水準		3665	88E3	02269	1-22-69	4E5D	E4 B9 9D
レ	2画	おつこよう	バ														4E5C	E4 B9 9C
ル	2画	かたな	トウ			かたな			常用	2年	第1水準		4561	9381	03765	1-37-65	5200	E5 88 80
ル	2画	かたな	チョウ													1-14-58	5201	E5 88 81
ル	2画	かんだれ	カン										524C	99CA	05044	1-50-44	5382	E5 8E 82
ル	2画	さのひ	ヒ										5238	9986	05024	1-50-24	5315	E5 8C 95
ル	2画	じゅう	ジュウ	シュウ		とあ	と		常用	1年	第1水準		303D	8F5C	02929	1-29-29	5341	E5 8D 81
ル	2画	ちから	リョク	リキ		ちから	つとめ		常用	1年	第1水準		4E4F	97CD	04647	1-46-47	539B	E5 8A 9B
ル	2画	つくえ	キ			つくえ			印刷標準字体				515C	997B	04960	1-49-60	51E0	E5 87 A0
ル	2画	つくえ														2-03-18	20628(D841_	F0 A0 88 A8
ル	2画	つつみがま	ホ										5231	99AF	05017	1-50-17	53F9	E5 8B 89
ル	2画	なべふた	トウ										5035	86B3	04921	1-49-21	4E40	E4 BA A0
ル	2画	に	ニ			ふた	ふた		常用	1年	第1水準	式	4E73	93F1	03883	1-38-83	4E8C	E4 BA 8C
ル	2画	にすい	ヒョウ										5152	9971	04950	1-49-50	51AB	E5 86 AB

読み	正しい変換	ATOK
けんしょうえん	腱鞘炎	腱鞘炎
しゃれい	謝礼	謝礼
しゃれき	社歴	社歴
じょうきをさんしょうして	上記を参照して	上記を参照して
かんじのごへんかん	漢字の誤変換	漢字の誤変換
しょうかん	使用感	使用感
でんげんしゃだん	電源遮断	電源遮断
そうしきかいじょう	葬式会場	葬式会場
だんしせいと	男子生徒	男子生徒
たしかに	確かに	確かに
かいげんしょう	怪現象	怪現象
きーのながおし	キーの長押し	キーの長押し
えいぞうくりえいたー	映像クリエイター	映像クリエイター
かんじがこうへんかんされます	漢字がこう変換されます	漢字がこう変換されます
ごくしょう	極小	極小
きょうしつにてほしゅう	教室にて補習	教室にて補修
さいかいごうかく	最下位合格	再会合格

中島敦：名人伝より

原文	ちょうど彼等の真上、空の極めて高い所を一羽の鳶が悠々と輪を画いていた。
ATOK2008	ちょうど彼らの真上、空の極めて高いところを一羽の鳶が悠々永久を描いていた。

中島敦：名人伝より

原文	その人の前には我々の技のごとき児戯にひとしいと言った師の言葉が、彼の自尊心にこたえた。
ATOK2008	その人の前には我々の技のごとき児戯に等しいといった氏の言葉が、彼の自尊心に答えた。

に ... お馬鹿になっていく<sup>[8]</sup>を参照して設定した。

具体的な変換例は右表を参照していただきたい。短い語句に関しては、ほぼ完璧な結果を示している。わずかに「補習」が「補修」に「最下位」が「再会」となっている程度である。これは前項の MS-IME に比べて明らかに違う。この結果は実際の用例のごく一部にしか過ぎないが、使い始めたばかりでもそれなりにこう高率な変換を見せる ATOK の特徴がよく出ていると思われる。

対して文章の方はどうだろうか。上の例では「悠々と輪を」が、「悠々永久を」になっている。助詞の「と」と名詞の「輪」が「永久」に誤認されてしまっている。ここは、MS-IME は正し

く変換している。下の例では、MS-IME では「時宜」と誤変換されている「児戯」が正確に変換されている。ただし、「自尊心にこたえた」が「自尊心に答えた」になってしまっている。これは Microsoft Office Input Method Editor 2007 と同じ誤変換である。

MS-IME と比較した場合、単文レベルでの優位性はあるが、長文レベルでは差がそれほどでもなくなっている。要するに、ATOK はこまめに（短い文節単位で）変換を繰り返して入力することで強みを発揮する IME だといえるだろう。

無論、各 IME それぞれに学習を重ねることによって、よりユーザーの変換意図に即した振る舞

いをしてくれるようになると思われる。従って、どの IME を選ぶべきか？ という選択は、人それぞれであるとしかたええないのかもしれない。

## ◇ おわりに

以上、ATOK についてその機能と変換効率について解説した。筆者は、ATOK7 以来のユーザーであり、気がつくとも 20 年近く使用し続け、加えて一万語になる自作の辞書を使用しており、今更他の IME に乗り換える気はないのだが、今回 MS-IME と比較することで改めて ATOK の優秀さを実感するに至った。

日本語用 IME は、日本語を利用するものならば必ず使用するアプリケーションでありながら、あまりに関心が薄い。これは、IME が常に他のアプリケーションとの連動で使用されるものであるために注目が行きにくいためであろうし、MS-IME がそれなりに優秀であって、殊更他の IME に乗り換えようとする意志が出にくいということもあるのだろう（もっとも、本特集のように、Microsoft Office Input Method Editor 2007 の初期変換効率が??? というところから IME について話題になったというのも皮肉な話であるが）。

MS-IME しか知らないユーザーは不幸である。どんなものであれ、他との比較によってそれぞれの特徴を知ることができる。決して MS-IME を否定するものではないが、とりあえず ATOK も使ってみてはいかがだろうか。ATOK には 30 日体験版<sup>[9]</sup> や新バージョンへの無償アップグレード件も持つ月々 300 円からの定額サービス<sup>[10]</sup> もあるので、是非これを機に試してみたい。

## 注

- [1] 昔、「ATOK は Awa TOKushima の略である。」と言われたりしたもののだが、一応正式名称は本文中に明示したもののようである。
- [2] ただし、パソコン用 OS 以外の組み込み用 OS に供給される ATOK は、+ATOK という名称となっている。
- [3] [http://www.atok.com/2007/function/new\\_engine.html](http://www.atok.com/2007/function/new_engine.html) より抜粋。
- [4] Microsoft Office Input Method Editor 2007 では、初期状態での変換効率があまりよくないものの、使い込むことによってより手になじむ変換エンジン（「Trigram/SLM (Statistical Language Model)」）を新たに採用していることと好対照であるといえよう。
- [5] Microsoft Office IME 2007 の手書き入力機能は、依然として JIS X 0208, JIS X 0212 のみ対応。ちなみに、Microsoft が提供する中国語用 IME の微软拼音输入法の手書き入力では、更に多くの漢字が認識できる。
- [6] 実際には、本特集記事で紹介する CHISE IDS FIND や今昔文字鏡などの漢字検索ツールを使って検索した方が早いのだが…
- [7] Andrew West 氏の開発による Unicode 文字パレット。ドネーションウェア（日本語ではカンパウェアという呼び方もある）での公開。<http://www.babelstone.co.uk/Software/BabelMap.html>
- [8] <http://furukawablog.spaces.live.com/blog/cns!156823E649BD3714!9079.entry>>><http://furukawablog.spaces.live.com/blog/cns!156823E649BD3714!9079.entry>
- [9] <http://www.atok.com/try/>
- [10] [http://www.justsystems.com/jp/products/atok\\_teigaku/](http://www.justsystems.com/jp/products/atok_teigaku/)

# ことえり

上地 宏一 (かみち こういち)

## ◆ 良くも悪くも Mac OS 標準 IME

「ことえり」と言えば Mac OS を利用していなくても、多少なりともマックに興味があれば知っているぐらいに有名な Mac OS 標準の IME である。筆者は長年 Windows を利用していたが、ここ数ヶ月程度 MacBook を使い続ける機会があり<sup>[1]</sup>、文字入力の際に受けた感覚の違いも含めてレポートできればと思う<sup>[2]</sup>。

## ◎ ことえりにはバージョンがあるのか？

Windows や Office の場合、製品のバージョンが非常に重要な要素となり、95、Me、XP、Vista (Windows) やオフィス 97、オフィス 2002、オフィス 2007 という具合に呼ばれることが多い。一方で、Mac についても OS X 以前については OS 8 や OS 9 と呼んでいたが、Mac OS X になってからは「X」(ローマ数字で 10) というメジャーバージョンは固定し、10.1、10.2 といったマイナーバージョンが上がっていく製品ラインナップとなり、(少なくとも筆者やそのまわりでは) Mac OS X のことを話すときにバージョンをしっかりと認識する機会が少なくなったように思う<sup>[3]</sup>。具体的にバージョンを区別する際は Tiger (10.4)、Leopard (10.5) といった名称を用いるため、Windows 利用者にとっては文化が違うと感じるものがある。

話がそれてしまったが、一方で「ことえり」についての歴史は実は長く、漢字 Talk7.1 (1992 年) に標準搭載されたのが始めであるという<sup>[4]</sup>。「漢字 Talk」についての説明は Mac における日本語

処理の歴史そのものとなるのであるがここでは省略する。筆者は気がついていなかったが、その後バージョンアップを重ね、現在は「ことえり 4」というバージョン番号がついている。ただし初出である Panther (10.3) のことえり 4 と現在の Leopard (10.5) のことえり 4 では中身が違うようだ<sup>[5]</sup>。

## ◎ Apple の言い分と利用者の感覚

「ことえりの変換能力は低い」と言われていることは有名であるが、一方で Apple の売り文句では、「業界最高水準の変換能力を誇る」としている<sup>[6]</sup>。このギャップは何なのだろうか。筆者が思うには、ここ数年で着実に実力を上げてきているが、表立ってアピールされていないことと、パー

図 1 Leopard (10.5) 日本語環境の紹介ページ。ことえりについてはたったの 5 行。







### ●部品からの文字検索

ことえりのもう一つの独自機能に、部品からの文字検索がある。ATOKのような専用のウィンドウを使うのではなく、その場で必要な文字に含まれる部品（あるいはその部品を含む別の漢字）を2、3個入力し、それらを選択した上で「Ctrl + 1」を入力すると、選択した部品を用いる漢字のリストが表示される。この際、各部品同士はORの関係となるため、必要な部品単体を入力することができずに他の不要な部品と組み合わせさせた漢字を指定しても検索できることも特徴だろう<sup>[10]</sup> (図3)。

### ●文字種の AdobeJapan1-5 相当への対応

すべての文字ということではないと思われるが、辞書収録単語において早くから JIS X 0208 よりも広い範囲での文字種に対応していた。「臨淄」<sup>りんし</sup>「潯東」<sup>ぼくとう</sup>「川芎」<sup>せんきゅう</sup>などの変換が可能である。JIS X 0208 に収録されない文字を使っている場合は、Windows Vista の「機種依存文字」表示と同様の注意が示されるが、「JIS では定義されていない文字が含まれています」と「新 JIS で追加された文字が含まれています (JIS X 0213 の文字が使われている)」の2種類があることが興味深い。残念なのは注意のマークがやや小さくて見づらいことである (図4、5)。

### ●文字パレット

ことえりの優れている点に文字パレットの充実が挙げられる。文字コード表とヒラギノフォントとの組み合わせは、何時間眺めていても飽きない。それはさておき、記号(約物)や各言語・文字が上手にカテゴリー化されていることや(図6)、漢字が領域(Ext. A、B、URO)を区別せずにまとめて部首画数順で調べられること(図7)、漢字の読み(日本語だけでなくピンインも)や筆画(中国の五筆)でも検索できる(図8)。加えて、コード表は Unicode 順だけでなく Big-5 や GB18030 などでも表示可能である。さらに「グリフ順表示」も可能である。これはフォントファイルに含まれているグリフの中で、Unicode や Adobe-Japan1

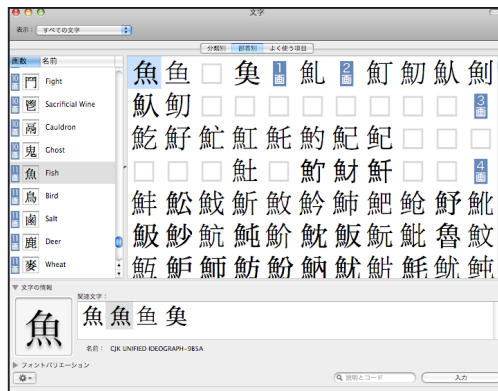


図7 魚編の漢字が画数順に並ぶ。空白はフォントがないため

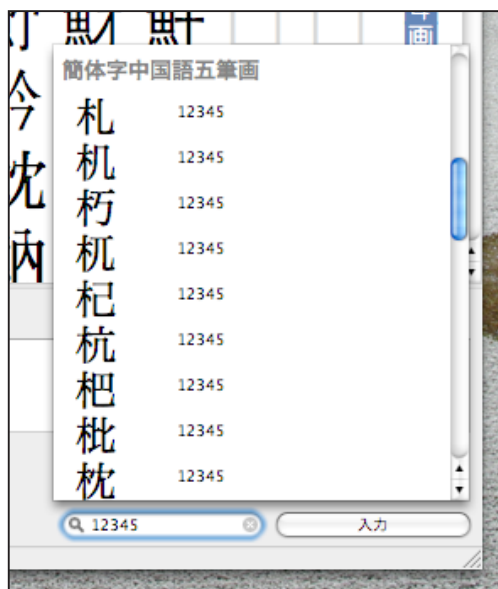


図8 五筆番号で「12345」を検索した例

番号に割り当てられていない文字についてもすべて表示され、実際にテキストエディットなどに入力することが可能となっている。たとえば「Arial Unicode MS」というフォントは漢字 20,902 字がすべて埋まっていることで有名であるが、基本的には日本の規格で定義されている文字は日本のデザインに合わせてある。ところが、フォントファイルの内部には中国(大陸)および台湾・香港デザインのグリフも隠れて用意されている。通常は隠れたグリフを使うことはできないが、ことえりの文字パレットを用いると可能なのだ(図9)。



図9 Arial Unicode MS に隠されているグリフ。上段から通常のグリフ、中国（大陸）デザイン、台湾・香港デザインとなっている。右の番号はフォントファイル内部でのグリフ番号

この隠れグリフは RTF 形式で保存すると正しく保存できる。それ以外の形式ではうまくいかないようだ。

## ◇ 共通タスクの変換結果

ここで与えられたテストデータの変換能力を確認することにしよう。まずは単語編から。

- 腱鞘炎
- 謝礼
- 社歴
- 上記を参照して
- 感じの誤変換（漢字の誤変換）
- 使用感
- 電源遮断
- 葬式会場
- 男子生徒
- 確かに
- 怪現象
- キーの長尾氏（キーの長押し）
- 映像クリエイター
- 漢字がこう変換されます
- ご苦笑（極小）
- 教室にて補修（教室にて補習）

## × 再会合格（最下位合格）

結果としてはまずまずであろうか。まず「漢字」と「感じ」の使い分けは他の IME でも同様に難しいようで、特に筆者の場合、両者の使用頻度は「漢字」>「感じ」のため、誤変換で困るケースが多い。同様に「件」（メール冒頭で多用）と「研」（大学の〇〇研で多用）の使い分けもできずにストレスを感じることがある。「長押し」という語句はことえりにはないようで、「キーを長押しする」でも変換できなかった。「ごくしょう（極小）」も辞書にはないようで「きょくしょう」で変換できた。一方で「教室にて補習」は事務方であれば「補修」も可能性があるわけで、「教室にて補習授業」や「教室にて補習を受ける」とすれば正しく変換できた。最後に「最下位合格」については「最下位で合格した」でも変換できなかった。ジョブズの辞書には「最下位」はないようである<sup>[11]</sup>。

次に文章編に移るが、あまり結果は芳しくなかった。下線部はうまく変換できなかった部分である（出典はいずれも中島敦『名人伝』より）。

ちょうど彼らの真上、空の極めて高いところを一羽のトビが悠々と輪を描いていた。

（ちょうど彼等の真上、空の極めて高いところを一羽の鳶が悠々と輪を画いていた。）

その人の前に出ては我々の技のごとき宜に等しいといった市の言葉が、彼の自尊心に答えた。

（その人の前に出ては我々の技のごとき兎戯に等しいといった師の言葉が、彼の自尊心にこたえた。）

問題があるとすれば「兎戯に等しい」「自尊心にこたえる（堪える）」は正しく変換できてほしいところである。

## ◇ まとめ：ことえりとうまく付き合うコツ

変換精度を上げるために取り入れられている文脈解析や学習機能をうまく活用するためには、前述のような単語ごとに入力するのではなく文単位で入力すると精度が上がると考えられる。たとえば「はたけ」と変換すると「畑」となるが、「はたけひさし<sup>[12]</sup>」と変換すると人名であるとの推測されるのか「畠久志」と変換される。また誤変換を修正するときは、誤ったまま確定してから該当部分を修正するのではなく、一度未変換状態に戻して修正し、正しい一文で確定することが重要になる。

個人的にはことえりには括弧記号の対応機能（たとえば「**【**」を入力した際、次に入力した閉じ括弧は対応する「**】**」が第 1 候補となる機能）が抜けていると感じる。また標準設定では半角カナを入力するためのショートカットが用意されていない<sup>[13]</sup>。このように細かい点で気になることはあっても、決して「ことえりは使えない」と感じることはなかった。OS 標準の IME として十分及第点に達していると言えよう。

### 注

- [1] 白状すると、過去 3 回ほど新しい iBook が発売されるたびに手を出してはすぐに手放していたりするのであるが、MacBook になってようやく安心して使えるようになったと感じている。
- [2] 本来であれば Mac（ことえり）をより長く利用しているエキスパートの方に執筆をお願いするべきであったのかもしれないが、特に Windows 利用者に対して「こ

とえり」をアピールするという意味で小生が担当することになった。なにとぞご容赦いただきたい。

- [3] ただしこれは Mac OS X のライフサイクルが短いことに起因するようにも思う。
- [4] <http://ja.wikipedia.org/wiki/ことえり>
- [5] Apple の公式な情報は確認できなかったが、変換精度が上がっているという Web の書き込みは多い。辞書が拡充されたとも考えられる。
- [6] <http://www.apple.com/jp/macosex/features/japanese.html>
- [7] Wikipedia の記事のほか、  
<http://www.apple.com/jp/articles/panther/20031107/detail.html>  
<http://www.apple.com/jp/articles/jaguarokotoeri/>  
<http://allabout.co.jp/computer/macosex/closeup/CU20050429A/index3.htm>を参照した。
- [8] 筆者は Mail 標準のスパムフィルタは使わず、独自のフィルタを用いているため、スパムメールも一般のメールフォルダに格納される。このため、ことえりの自動学習機能がスパムメールを排除しているかどうかは不明である。
- [9] Leopard (10.5) では Spaces の機能が「Ctrl + 1」および「Ctrl + 2」を使うため、別の数字に変更する必要がある。もしくはマウス操作でも可能。
- [10] その反面、候補が多くなるという欠点も生じる。
- [11] もちろん冗談である。
- [12] ちなみに「はたけ」と「ひさし」は Leopard 発売当初のことえりで変換するとフリーズすることで話題になった単語である。
- [13] 知っている人であれば、すぐに設定可能。ちなみに某オンラインバンキングサービスでは口座振込の際の情報入力に半角カナの使用が強制されていた。

# ウェブベース日本語 IME

秋山 陽一郎 (あきやま よういちろう)

## ◆ なぜウェブベースなのか？

日本語 IME (日本語入力メソッド) といえば、この特集中でも紹介されている Windows の MS-IME や ATOK、Mac OS X のことえりなどが一般的によく知られている。これらの定番 IME は、周知の通り、オンライン状態 (インターネットやローカルネットに繋がっている状態) でもオフライン状態でも動作する。

それに対して、ここで紹介するウェブベース (サーバーサイド) の IME は、インターネットに繋がっていることが前提になっていて、原則としてオフラインでは利用できないサービスが中心だ。

このように説明されると、「オフラインで利用できないのに、使い物になるのか」と疑問に思うのが自然だろう。しかし、このウェブベースの IME には、ちゃんとした使いみちがあるのだ。

## ◎ 主な需要は海外の端末での日本語入力

本会会員や本誌の読者には、出張や調査旅行などで海外に飛ぶ機会のある方が少なくないと思う。海外に自分の端末を携行している場合はともかく、もし何らかの事情で、現地の端末でメールの読み書きをしなくてはならなくなった場合には、どうするだろうか？

海外の端末では MS-IME やことえりが利用できることがまれ<sup>[1]</sup>な上、キーボードも全角 / 半角キーや、英数・かなキー、変換キーがまずついていない。

こうした状況に陥った際に、絶大な効果を発揮するのがウェブベースの IME だ。ウェブベース

の IME なら、ソフトウェアをローカルディスクにインストールする必要がなく、OS やアプリケーションの設定も変更せずに済ませられる。もちろん、キーボードも英数字と Space バーと Enter キーなどがあれば事足りるので、現地の端末に備えつけられているものをそのまま使えばいい。ローカルディスクを圧迫しないので、EeePC や OLPC といった、低スペックなミニノートでもインターネットに接続されてさえいれば全く問題ない。

普段使いのための手段ではないが、知っておくといざという時に思わぬ恩恵を受けられる。本節では、そんなウェブベース日本語 IME を取り上げる。

	クライアント IME	ウェブベース IME
オフライン	可	不可
オンライン	可	可
インストール (+ ローカル ディスク領域)	必要	不要
レスポンス	使用する端末の スペックに依存	サーバースペック と回線に依存

## ◆ Ajax IME

本誌でもこれまでに何度か紹介しているが、Ajax IME は、工藤拓氏 (Google) が 2005 年 2 月に公開した形態素解析器 MeCab を変換エンジンにしたウェブベースの日本語 IME だ。

### ● Ajax IME

<http://ajaxime.chasen.org/>

[使い方]: まず上記のページにアクセスして [IME On/Off] ボタンをクリックする (図 1)。入力用フォームの中が淡い水色に変わったら準備完了。これで英数字入力モードのまま、入力用フォームに日本語を入力できるようになったはずだ。(日本語環境下で試す場合は、使用中の IME の日本語入力モードがオフになっていることを確認すること。) 文章の入力が完了したら、入力した文章をコピーして、目的のウェブページやアプリケーションに貼り付ける。

◎ どこでも Ajax IME

その都度、Ajax IME のページに移動して、コピー & ペーストするのは面倒だという人のために、「どこでも Ajax IME」という、任意の入力用フォームで Ajax IME を利用できるようにするブックマークレットがある。(ブックマークレットとは、ウェブページの URL の代わりに、短い JavaScript を記述したブックマーク=お気に入りのこと。簡単なミニプログラムを実装・実行することができる。)

[登録法]: Ajax IME のページ下部に「どこでも Ajax IME」というセクションがある。この中にある [Ajax IME] というリンクを右クリックして、Firefox なら [このリンクをブックマーク]、IE なら [お気に入りに追加] する。この時、ブックマークツールバー (IE ならお気に入りバー) に登録しておくことで起動が楽になる。

[使い方]: 任意のテキストフォームで、ブックマーク or お気に入りの [Ajax IME] をクリックすると Ajax IME が有効になる。あとは普通のローマ字変換入力と同じ要領で日本語を入力していくだけ。

なおこのブックマークレットは、サーバーやブラウザのセキュリティ設定によっては利用できないことがある<sup>[2]</sup>。その場合は、やはり Ajax IME のページにアクセスして、その都度、入力用フォームに入力するしかない。



図 1 Ajax IME

● Ajax を使った手書き文字認識

<http://chasen.org/~taku/software/ajax/hwr/>

このほか、手書き文字入力用のインターフェイスも用意されている。Ajax IME の辞書には登録されていない、人や場所の名前などに使われる文

図 2 Ajax 手書き文字入力



字を入力したい場合に活躍するだろう。ただし、認識可能な漢字は JIS の第 1 水準までのようだ。

[使い方]：淡い水色の領域に、マウスで漢字を書くと、右側に認識候補が 10 字リストアップされる。目的の文字がリストにある場合は、目的の文字をクリックすると、下のテキストフォームにその文字が出力される。ない場合は [やり直す] ボタンを押して文字を書き直すか、[教える] ボタンをクリックして学習させることもできる。ただし、第 2 水準以降の漢字や、記号などの特殊な文字は、今のところ学習させても利用できない。本格的に漢字を探したい場合は、後述の CHISE IDS find を利用した方が効率が良いだろう。

### ◎ Ajax とは？

Ajax (エイジャックス) とは、Asynchronous JavaScript and XML の略。端的に言えば、JavaScript の非同期通信機能 (XMLHttpRequest クラス) を利用して、サーバーサイドプログラムへのリクエスト結果や、ほかのファイルの一部ないし全部を、ページを再読み込みさせることなく、動的に表示中のページに反映する仕組みのことだ。命名者の Jesse James Garrett 氏によれば、

- 標準に準拠した XHTML + CSS プレゼンテーション。
- DOM (Document Object Model) を利用した双方向的な動作と動的な表示。
- XML と XSLT による、データの変換や操作。
- XMLHttpRequest による非同期的なデータの取得。
- 以上を JavaScript によって統合させること。

という 5 つの技術を組み合わせたものと定義されるが<sup>[3]</sup>、現在の実勢は必ずしもこの定義の通りというわけではない。たとえばリクエストするデータは XML 形式である必要はなく、プレーンテキストや、HTML、CSV、JSON といった形式であることも多い。

実装の一例を挙げるなら、別のページに切り替えることなく、同じページの中で地図をどこまでも延々とスクロールし続けることができる、Google マップがわかりやすいだろう。非同期通信を使っているのも、必要なタイミングに、必要な分だけの情報を要求し、表示中のページに追加や上書きすることができるのが Ajax の特長だ。

Ajax は、今やウェブアプリケーションを語る上で外すことのできない技術のひとつで、本節で取り上げるウェブベース IME の多くも、この技術を利用している。具体的には、文字をテキストフォームに入力するたびに、入力した文字列をサーバーに送信し、マッチする変換候補を入力者に返すといったことをバックグラウンドでしている。

## ◆ Sumibi.org

### ● Sumibi.org

<http://www.sumibi.org/>

Ajax IME と類似のサービスとして Sumibi.org がある。変換候補をリストボックスから選択させられたり、SKK の影響だという独特の分かち書き入力<sup>[4]</sup>が必要だったり、Ajax IME のようなブックマークレットがなかったりと、一見、使い勝手は Ajax IME の方が優れているように感じられるが、実は細かく見ていくと、Ajax IME にはない特長が色々あって、これも選択肢として勝るとも劣らない魅力がある。

[使い方]：

1. Google 検索モードと長文作成モードの 2 種類のフォームから目的に合う方を選ぶ。
2. 単語や助詞をスペースで区切りながら、文章や語句をローマ字で入力する。なお、区切り文字として入力したスペースは、変換候補を確定した時点で取り除かれる。
3. 漢字変換しない語句については、平仮名なら末尾に .h、片仮名なら末尾に .k、アルファベットなら末尾に .e (もしくは .l) を

	Sumibi.org	Ajax IME
語句解析	単文節区切り (利用者自身による分かち書き入力)	形態素解析 (MeCab <sup>[5]</sup> )
動作確認済ブラウザ <sup>[6]</sup>	IE, Firefox, Opera, Safari, Google Chrome	IE, Firefox (Opera は変換の際、一時的に表示が乱れる。)
変換辞書	Emacs への組み込みを行った場合や、自前の Sumibi Server を設置した場合は自作・カスタマイズ可能。	カスタマイズ不可。
その他	GPL2 ライセンスのオープンソース・ソフトウェア。SumibiWebAPI を利用して独自にウェブアプリケーションを開発することも可能。Emacs に組み込むことも可能。(sumibi.el)	任意のフォームで Ajax IME を呼び出せるブックマークレットあり。

つけると確実に意図した通りに変換される。ただ、変換候補の中に平仮名・片仮名・アルファベットの候補が必ず入っているので、この入力規則を覚えなくても心配はない。空白文字 (スペース) は、スペースを 2 つ続けることで入力できる。

4. フォームの直下に表示される変換候補の中から適切なものを選択。この時、実際に手動で選択するのは、ハイライトされている第 1 候補が不適切な語句だけで良い。(つまり、図 3 のように第 1 候補にすべて適切な変換候補が並んでいる場合は、わざわざ選択しなおす必要はない。)
5. [確定] ボタンをクリックすると、変換されたテキストが、フォーム中に出力される。
6. (長文作成モードの場合) 文章の入力が完了したら、入力したテキストをコピーして、目的のページやアプリケーションに貼り付ける。

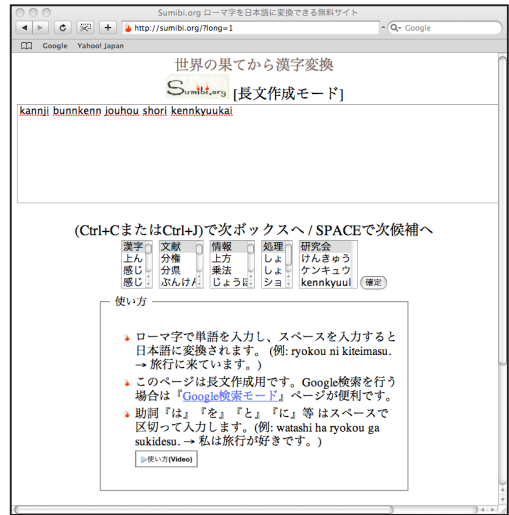


図 3 Sumibi.org (長文作成モード)

だ<sup>[7]</sup>。

●CHISE IDS find

<http://mousai.kanji.zinbun.kyoto-u.ac.jp/ids-find>

CHISE IDS find は、本誌でもたびたび紹介されている CHISE (Character Information Service Environment) プロジェクト<sup>[8]</sup> から生まれた成果のひとつで、漢字を構成部品の組み合わせから検索するためのウェブサービスだ。

IDS (Ideographic Description Sequence) とは、ISO/IEC 10646 で規定されている、漢字の部品構成を記述する方法で、記述には u2FF0 ~ u2FFF にある

◆ CHISE IDS find

海外の端末で日本語を入力する際に、必ずといって良いほど直面する問題が、人や場所といった固有名詞に使われる漢字の入力だ。この点、これまで紹介した Ajax IME や Sumibi.org の辞書やユーザーインターフェイスでは、漢字を探すには全く向いていない。唯一、向いてそうな Ajax 手書き文字認識も、JIS 第 1 水準程度の漢字しか扱えないため、これまた実用性には限界がある。

こうした場面で活用したいのが CHISE IDS find

## CHISE IDS 漢字検索

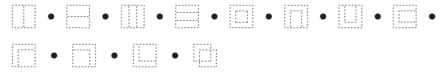
Version 0.24.1 (Last-modified: 2008-09-18 20:32:29)

部品文字列

指定した部品を全て含む漢字の一覧を表示します。

CHISE で用いられる実態参照形式 (例: &M-00256;) で部品を指定する事もできます。

図 4 CHISE IDS find



の 12 種類 の IDC (Ideographic Description Characters) を利用する。CHISE IDS find は、CHISE の文字データベースに登録されている文字素性情報のうち、この IDS 情報を検索する。

## CHISE IDS 漢字検索

Version 0.24.1 (Last-modified: 2008-09-18 20:32:29)

部品文字列

- 勅 U+52C5 (link map) 束力 ⇒[唐代拓本]
  - 勅 U-000208A6 (link map) 勅正
  - 勅 U+9D92 (link map) 勅鳥
  - 勅 U-0002A0E0 (link map) 勅鳥
- 勅 U-0002A19F (link map) 勅力鳥
- CB02246 CB02246 束力
- &HZK12-CA76; 勅力&B-B3BE;

Copyright (C) 2005, 2006, 2007, 2008 MORIOKA Tomohiko

Powered by XEmacs CHISE 0.24 (Kasagi).

図 5 CHISE IDS find で「勅」を検索

[使い方]: [部品文字列] フォームに、探したい漢字の構成部品を (一部ないし全部) 入力し (海外で利用する場合は、Ajax IME や Sumibi.org 経由で部品となる漢字を入力してコピー & ペースト)、[検索開始] ボタンをクリックすると、入力した構成部品を含む漢字の一覧が表示される。

あとは、表示された一覧の中から目的の文字をマウスで選択してコピーすれば良い。

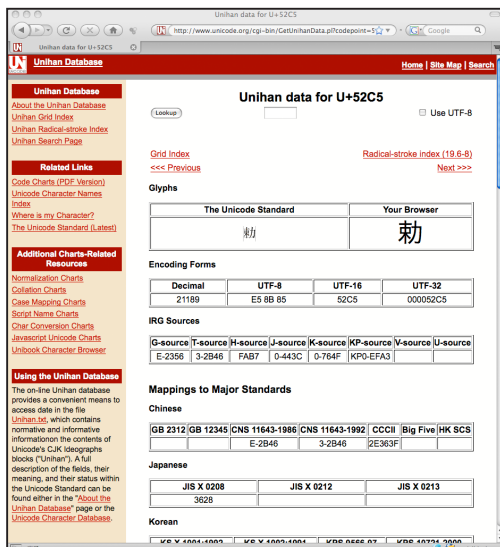
ちなみに検索結果として出力されるのは、左から順に、見出し字 (“勅”)、文字コード (“U+52C5”)、CHISE 漢字連環図へのリンク (“link map”)、IDS (“束力”)、拓本文字データベースへのリンク (“⇒ [唐代拓本]”) となっている。

見出し字のリンクをクリックすると、部首と部首部分を除いた残画数、総画数、紙媒体辞書における漢字番号、各種文字コードのコードポイント、異体字関係情報といった「素性」情報が表示される。

また、文字コード部分のリンクをクリックすると、Unihan Database<sup>[9]</sup> の当該文字データに飛ぶようになっている。Unihan Database は、Unicode 規格書内の字形を Glyphs - The Unicode Standard の欄にスキャン画像で示しているの、文字化けなどの理由で CHISE IDS find のページ上で見出し字の字形が確認できない場合にも便利だ。

情報の充実度はまだまだ不十分だが、漢字の類字 (≡異体字) 関係を Graphviz で図示した、上地宏一氏の CHISE 漢字連環図 (図 7。CHISE Link Map)<sup>[10]</sup> も、利用シーンによっては、漢字を探す上で非常に有効なツールとなる。これも CHISE IDS find の検索結果の “link map” と書かれているリンクをクリックすることで表示できる。

図 6 Unihan database





◎ CHISE IDS find の優位性と弱点

CHISE IDS find は、漢字を構成部品から検索できる類似のシステムのうち、ATOK よりも多くの漢字を検索することができ、今昔文字鏡よりも豊富な情報に無料かつ自由にアクセスできる。ウェブベースのサービスであるため、プラットフォームを選ばず、専用ソフトのインストールが必要ない手軽さも大きな強みだ。さらに中国の五筆输入法や台湾の倉頡输入法のように、キーボードの字根(構成部品)の配列を覚える必要もなく、誰でも予備知識なしに直感的に漢字を探すことができる。

反面、構成部品を細分化しすぎたり、元の文字から字形が変化した構成部品(「足」と「趾」、「心」と「忄」など)を含む場合などに、期待した検索結果が得られないことがある。この場合は、考えられるほかの構成部品で再度検索を試みるしかない。また極端に画数の少ない字や、曲線を含むなどの変則的な字の検索も不得手だ。このうち画数の少ない字については、その字を構成部品として含む漢字を検索し、その IDS 情報から探し当てるといった方法があるが、画数の多い漢字と比べると、やはり探すのにより時間を要するケースが多い。

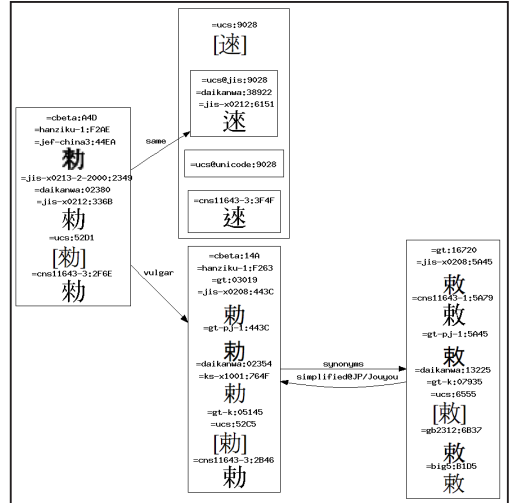


図7 CHISE 漢字連環図(「勅」字)

語彙が収集されやすいとは言えるだろう。

◎ ツールバー

- ① IME 選択ボタン
- ② 入力文字種選択ボタン
- ③ 予測変換モードボタン
- ④ 単語登録ボタン

◎ 予測変換モード

文字パレットの類いはないが、ショートカットキーも含め、基本的な操作は MS-IME に準じている。ユーザーインターフェイスの面で特筆すべきは予測変換モードだ。

ツールバーの「予」というボタン(図8-③)をクリックするか Ctrl + 半角全角キーで、予測変換モードの on / off を切り替えられる。(このモードが off の状態でも、入力中に Tab キーを叩くこ

◆ Social IME

Social IME は、奥野陽氏が(慶應義塾大学理工学研究科)開発した、変換辞書と学習データに集合知を取り入れた Windows 用日本語入力メソッドだ<sup>[11]</sup>。

● Social IME

<http://www.social-ime.com/>

Ajax IME や Sumibi.org と違って、ソフトウェアのインストールを必要とし、ウェブブラウザ以外のクライアントアプリケーション上でも動作するが、基本的にオンライン状態での利用が前提となっており、オフライン状態では利用できない。ただウェブブラウザ以外のアプリケーション上でも入力できる分、インターネット外で使用される

図8 Social IME ツールバー



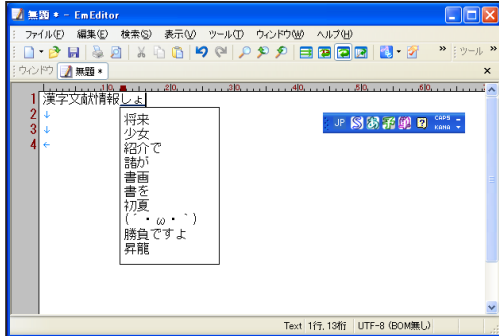


図9 Social IME

とで、任意のタイミングで予測変換候補を表示できる。)

### ◎ 変換候補は利用者層に依存

さて、肝心の集合知による変換精度だが、予測変換をオンにしてしばらく試用してみた印象では、図9にも端的に現れているように、流行語（いわゆるオタク系の語彙）や顔文字の類が多く、今のところかなり偏りがあるようだ。少なくとも、学術方面での実用性はこのままでは見込めないだろうが、ある偏ったユーザー層における変換ログが蓄積・反映されている痕跡は確かに認められ、試みとしては面白い。

変換候補の偏りは、利用者層の偏りを示すものにほかならない。より幅広い層のユーザーがつけば、変換候補の傾向もより均質化するものと思われるが、IMEが常用性の高いツールであることを考えると、変換候補が偏っている現状は、おそらくより一段、ユーザー層の偏りを招いてしまうものと予想される。Google マップにおけるマップレットのように、やはり分野ごとに辞書を分けて、利用者ごとに任意の辞書を選択できた方が、(研究テーマとしての面白味は別として) 実用性は高まったであろうと思われる。

また、たとえば文語を多用する層と、口語を多用する層とでは、おのずと使用する語彙の傾向が異なってくるように、選択する変換辞書（すなわち、そこから導き出されるユーザー層）によって、まったく性質の異なる学習結果が得られよう。この意味で、すべての単語・学習データを画一的に

一個所でまとめてしまっているのは、もったいない気がする。

## ◇ IME とプライバシー

### ◎ 集合知=ユーザーの入力ログの収集

さて、ここでどうしても注意を喚起しておきたいことがある。それは、集合知（もしくはパーソナライズ）を謳うウェブアプリケーションを利用する際は、必ずプライバシーやセキュリティに留意すべしということだ。

たとえば前項の Social IME は、変換辞書を全ユーザーが共有することから、個人の住所やメールアドレスを辞書登録するのは御法度である。さらに、実は辞書に登録するまでもなく、個人情報を入力しただけでもリスクを負うことになるという点が、意外に見落とされやすい<sup>[12]</sup>。これは決して大げさな話でも何でもなし。現に利用規約にも以下のように書かれている。

Social IME では、変換される文章はサービス提供者の運営するサーバに送信され、サーバサイドで変換が行われます。その際、変換のログをサーバ上に残します。また、単語登録した辞書もサーバ側で管理します。(利用規約 - プライバシーポリシー)

つまり、その気になれば「サービス提供者」は、IP アドレスを情報として含む変換ログを解析し、個人情報や、ID・パスワードを収集することもできてしまうわけだ。(特に予測変換がオンの状態で、これらの情報を Social IME で入力した場合。) またサーバに送信される入力ログが、Social IME の場合は平文のまま暗号化されずに送信されるため、「サービス提供者」に良心があったとしても、悪意ある第三者によって盗み見られてしまうリスクもある。

利用者はこうしたリスクを、サービスの仕組みを通して充分理解しているべきで、リスクの内容に応じて利用するシーンを意識的に限定したり、

サービスの利用そのものを回避するなどの判断を各自でするようにしていただきたい。

### ◎ Ajax スキーム

そもそも Ajax で頻繁にくり返される以下のスキーム自体が、利便性とリスク両面の根源的な要因となっている。

1. キーボードで文字 (列) を入力。
2. 変換候補をサーバーに問い合わせ。(XMLHttpRequest)
3. サーバーサイド変換エンジンが候補を返す。
4. 変換候補を (DOM+CSS で) 元のページに出力。

これは、ここで取り上げた Ajax IME や Sumibi.org のようなウェブベース IME に限らず<sup>[13]</sup>、Google サジェストなど、あらゆる同種のスキームを持つウェブサービスについても言えることだ。ただ、ウェブ検索と比べ、IME はプライバシーに関わる情報を入力するシーンが比較にならないほど多い。この点が大きなネックなる。

最近、IE 8 beta・Google Chrome beta・iTunes 8 など、利便性向上のために、ユーザーごとの情報や入力ログを同じ仕組みで収集する機能を持ったアプリケーションが相次いで出てきている。

アプリケーションが個人情報を収集したり送信したりする場合、必ずユーザーに明示する必要があります。収集、送信される情報が何であるかを明確にした上で、ユーザーに、それらの情報の送信許可の確認を行うべきです。(Google ソフトウェア原則)<sup>[14]</sup>

実装者側には、たとえば上に掲げたようなルールの共有や遵守が求められるのと同時に、利用するユーザー側にも、性悪説を前提としたリテラシー向上がますます必要になってきている感がある。

### 注

- [1] MS-IME (Windows) なら [コントロールパネル] (Control Panel) → [(日付、時刻、) 地域と言語のオプション] (Clock, Language, and Region | Regional and Language Options) → [言語] (Languages) → [テキストサービスと入力言語] (Text services and input languages) から、ことえり (Mac OS X) なら [システム環境設定] (System Preferences) → [言語環境] (International) → [入力メニュー] (Input Menu) から追加できるが、企業や組織のネットワークに繋がっている端末だと、管理者権限がないとこれらの変更を受け付けないようにしていることが多い。  
なお、非東アジア圏 Windows の場合は、上記の手順で MS-IME を追加する前に、さらに [Text services and input languages] ダイアログの [Languages] タブ下部にある、[Install files for East Asian languages] と書かれているチェックボックスにチェックを入れる必要があるが、この際に Windows のインストールディスクが必要になることがある。
- [2] 特に SSL などのセキュリティで保護されたページや、XSS (クロスサイト・スクリプティング) 対策が有効になっているブラウザで利用できない可能性が高い。
- [3] "Ajax: A New Approach to Web Applications"  
<http://www.adaptivepath.com/ideas/essays/archives/000385.php>
- [4] 漢字や漢熟語の頭を大文字で入力しないなど、厳密には SKK の分かち書きとは違う。文節ごとに入力者自身がスペースで区切る分かち書き入力は、慣れない人は抵抗を覚えると思うが、形態素解析のような区切り間違いが発生しにくく、変換精度がかなり向上する。SKK は、佐藤雅彦氏 (京都大学大学院情報学研究所) が開発したオープンソースの日本語入力メソッドで、現在は SKK Openlab が開発を引き継いでいる。あくまで筆者個人の印象だが、独特の入力方式にさえ慣れてしまえば、BMP 外の拡張漢字の辞書登録や入力もまったく問題なく、この SKK は、日本語形態素解析をあてにできない古漢語 (いわゆる漢文) の入力には最も向いている日本語入力メソッドかもしれない。  
<http://openlab.ring.gr.jp/skk/index-j.html>  
<http://www.sato.kuis.kyoto-u.ac.jp/~masahiko/>

- [5] この形態素解析エンジン MeCab (めかぶ) も工藤氏の代表的な実装成果のひとつ。  
<http://mecab.sourceforge.net/>
- [6] 動作確認は、Windows XP SP4 と Mac OS X 10.5 の、IE 7, IE 8 beta, Google Chrome beta (以上は Windows のみ)、Firefox 3, Opera 9.5, Safari 3.1 (以上は Windows と Mac の両方) で行った。
- [7] CHISE IDS find については、上地宏一氏のレビュー (本誌第 6 号、2005) も参照されたい。また山田崇仁氏のオンラインマニュアルも有用なリソースなので覚えておきたい。  
[http://www.shuiren.org/chuden/toyoshi/syoseki/chise\\_ids.html](http://www.shuiren.org/chuden/toyoshi/syoseki/chise_ids.html)
- [8] 守岡知彦・江渡浩一郎・苫米地等流・宮崎泉・師茂樹「CHISE Project」(本誌第 4 号、特集 2 漢字処理技術の最新動向、2003) など。  
<http://www.jaet.gr.jp/archives/journal/jj04.pdf>  
CHISE プロジェクト  
<http://www.kanji.zinbun.kyoto-u.ac.jp/projects/chise/>
- [9] <http://www.unicode.org/charts/unihan.html>
- [10] [http://fonts.jp/chise\\_linkmap/](http://fonts.jp/chise_linkmap/)
- [11] 「Social IME: サーバサイド日本語入力とログ活用サービス」(未踏ソフトウェア創造事業 2007 年度第 I 期) の成果物。詳細は下記リソースを参照。  
<http://www.ipa.go.jp/about/jigyoseika/07fy-pro/mito/2007-0469a.pdf>  
<http://www.ipa.go.jp/about/jigyoseika/07fy-pro/mito2007-0469d.pdf>  
<http://www.ipa.go.jp/jinzai/esp/2007mito1/mdata/4-2.pdf>
- [12] 現在は機能そのものが削除されているが、当初は入力した内容が Twitter (ミニブログサービス) に自動投稿される、「公共モード」という恐るべき仕様があって、しかもその「公共モード」はデフォルトで有効になっていた。
- [13] Sumibi.org は、変換時に入力された文字列情報を記録

せずに破棄していると明記しているし、自前の Sumibi Server を立ち上げることもできる。しかし、ここで大事なのはプライバシーの保護が、個人情報保護法とサービス提供者の善意 (いうなれば性善説) に依存しているという状況で、悪意があれば、入力した内容のログをいつでも収集することができるという状況それ自体なのだ。

- [14] [http://www.google.com/intl/ja/corporate/software\\_principles.html](http://www.google.com/intl/ja/corporate/software_principles.html)

ただし、この文書には同時に「比較的最近のアプリケーションの中には、この原則に完全に順守していないものや、ここに記載されていない事例もあります」という注記がある。

実際、Google Chrome の利用規約に、当初「コンテンツを送信、投稿、表示することにより、ユーザーは、本サービスで、または本サービスを通じて送信、投稿、または表示したコンテンツを再生、改作、改変、翻訳、公表、公開、配信できる恒久的かつ取り消し不能で、使用料が発生しない非排他的なライセンスを Google に付与することになります」という条文があって、ちょっとした騒動になった。なお、この条文はその後削除されている。

[追記] なお、脱稿後、以下のリソースの存在に気づいた。

◆ Zinnia : 機械学習ベースのポータブルな手書き文字認識エンジン :

<http://zinnia.sourceforge.net/index-ja.html>  
<http://chasen.org/~taku/blog/archives/2008/09/zinnia.html>

◆ かな漢字変換 API (Yahoo! デベロッパーネットワーク)

<http://developer.yahoo.co.jp/jlp/JIMService/V1/conversion.html>

# 特集2

## 文献画像のデジタル 化をめぐる諸問題

近年、デジタルカメラが格段に安価になり、研究者が資料調査などに利用することも珍しいことではなくなってきた。一方、文献画像をウェブで公開する例もどんどん増えており、古典籍から研究論文まで、かなりの量の文献が閲覧できるようになっている。かつて大規模テキストデータベースと言え、台湾中央研究院の漢籍全文資料庫のような電子テキストのシステムであったが、現在では Google ブック検索のように画像が表示されるものを想像する人の方が多いかもしれない。本特集はこのような現状をふまえて企画された。

寄せられた4本の報告は、個人レベルから研究機関レベルまで、研究や出版を前提としたハイレベルな撮影手法からウェブでの共有・公開を前提としたデジタル化まで、現在の文献画像デジタル化をめぐる主な論点を網羅しているのではないと思う。もちろん、カバーしきれなかった論点や、より深く掘り下げべき問題など、今後の議論に委ねるべき部分はたくさんある。読者諸賢のご批判をいただければ幸いである。

なお、本特集の企画にあたっては、漢情研 BBS での議論がベースとなっている。議論に参加して下さった会員各位に感謝申し上げます。

### Contents

より包括的な「漢字文献情報処理」をめざして 特集「文献画像のデジタル化をめぐる諸問題」によせて	師 茂樹	62
寺院文化財調査におけるデジタル写真撮影 花園大学歴史博物館の調査活動を事例に	志水 一行	66
文献のデジタル画像化 ——『周作人書簡集』影印本製作（失敗の巻）——	小川 利康	71
報告：学会誌論文のデジタル化	山崎 直樹	80
ハーバード大学における蔵書デジタル化支援体制 ——韓国古典籍デジタル化事業を中心に——	江上 敏哲	83

# より包括的な「漢字文献情報処理」をめざして

特集「文献画像のデジタル化をめぐる諸問題」によせて

師 茂樹 (もろ しげき)

## ■ 本特集のねらい

### ＊今さらながら「漢字文献情報処理」とは

本誌はその名の通り「漢字文献」の「情報処理」について「研究」することを目的として発刊されている。しかしこれまで、そもそも「漢字文献情報処理」なるものがどのようなものなのかについての明確な合意や定義のないまま、本誌9号まで発刊されてしまっている。とは言え、研究会員の間で「漢字文献情報処理」についてのコンセンサスがまったくないのかと言えば、そうではない。「漢字文献情報処理研究会設立趣意書」<sup>[1]</sup>には、次のように書かれている。

…こと東洋学分野における情報化は、他分野に比べて立ち遅れた状況にある。それは、従来、コンピュータの日本語環境における多言語処理・漢字処理が困難であったこと、及び、コンピュータへの理解不足に起因するものと思われる。

現在ではハード及びソフトの急速な発展により、ごく一般的なパーソナルコンピュータで簡単に数万字の漢字や中国語を

扱えるようになってきている。一方、台湾・中国・米国などでは、急速に東洋学分野の電子化が進展しており、我が国の東洋学は世界的潮流からも取り残されつつある。しかし、我が国では、このような事実すら知られておらず、その状況は危機的とも言える。

下線部を見ればわかるように、本研究会設立時における「漢字文献情報処理」とは、文字コードとフォントによって表現される（主に漢字で書かれた）電子テキストを、入力・編集したり全文検索やテキスト処理をしたりすることが意図されている。本研究会による入門書『電腦中国学Ⅰ』『電腦中国学Ⅱ』（好文出版）も基本的には同じ考え方だし、本誌のバックナンバーの記事にも電子テキストに関するものが多い。したがって、本研究会における「漢字文献情報処理」は、概ね電子テキストが対象であると言ってよいだろう。

### ＊画像だって「文献」だ

一方、近年、人口に膾炙されるようになった「デジタルアーカイブ」<sup>[2]</sup>における「文献」のデジタル化は（きちんと統計をとったわけではないが）、画像データが中心であると言ってよいのではないかと思う。Google や世界各国の大学・図

書館などが文献の画像化と公開に取り組んでおり、日本においても国立国会図書館や国立公文書館をはじめとする多くの機関が所蔵資料を次々に画像データとして公開しているのは本誌レビューなどで既報の通りである。

本研究会が発足した約 10 年前は台湾を中心とした大規模テキストデータベースの開発・公開が大きな話題であり、一方で Unicode が徐々に普及して多漢字処理環境が大きく発展しようとしていた時期であった。逆に現在は、パソコンの性能向上やブロードバンド普及などが画像中心の「デジタルアーカイブ」を後押ししている側面がある。しかし、このような時代状況を差し引いても、本会の「漢字文献情報処理」に対する認識に偏りがあることは否めないように思われる。

筆者は、本研究会の BBS で「【雑談】文献撮影用のカメラの選び方」というスレッド<sup>[3]</sup>を立ち上げ、それが縁となって本特集の企画にも携わったが、その背景にはこのような本研究会と現況とのずれが問題意識としてあった。

## ■ 目的・方法論による「文献」概念の違い

### ＊包括的であれば良い、というものでもない

とは言え、画像処理について研究するようになればそれで OK とは考えていない。本意を言えば、画像データにも目を向けることで、何となく電子テキスト中心でやってきた我々の意識を相対化し、再考するきっかけになるのではないかと密かに期待しているのである。

安岡孝一氏は、漢字情報学の構築を目指し「テキスト情報階層モデル」を提唱し、イメージ層・テキスト層・シンタクス層・セマンティクス層という階層（それぞれ画像データ、テキストデータ、マークアップ、辞書データなどがゆるく対応する）と、OCR・組版・自然言語処理などの処理を包括した叩き台的なモデルを提示（安岡氏曰く「デッチあげ」）している<sup>[4]</sup>。このモデルで重要なのは、単に様々なデータ形式を包括しているだ

けでなく、それと処理との対応関係をモデル化しようとしている点である。例えば、OCR はイメージ層からテキスト層への変換処理と考えられるが、その際シンタクス層やセマンティクス層は関係がない、という具合に、処理に応じて対応するデータも変化する。もう少し人文学よりの言い方をすれば、研究分野の目的に応じて、つまり研究方法に応じて、同じ「文献」でも必要となる部分（層）が違うということである。逆に言うと、電子テキストやデジタル画像、あるいはそれらにもとづいたデータベースやデジタルアーカイブについて議論する場合には、まずもってその方法論について論じなければならないということである<sup>[5]</sup>。

実は本研究会においても、方法論についてのつっこんだ議論がなされたことがあった。それは、石岡克俊氏を招いた著作権講演会においてである。例えばこの講演会においては、著作権法で保護されるのは無体物としての文献、すなわち（おおざっぱに言えば）文献の内容であり、有体物としての文献（モノとしての文献）ではない、という法律上の原則をふまえて議論をしたり、校訂作業に著作権は発生するのか？という問題意識からそもそも校訂とはいかなる行為なのかという議論にまで発展したりした。これらの議論では、安岡氏による階層モデルのように、我々が漠然と考えていた「文献」というものが研究方法に依存した概念であることを明確にすることができたのである。

### ＊漢文研≒文献学？

本研究会の発足当初、電子テキストに偏った議論がなされ、また方法論についての議論をほとんどする必要がなかったのは、文学・哲学・歴史学など、本文批判や史料批判をベースとする文献（史）学的な研究をしている会員（あるいはそれらの訓練を受けた会員）が多いからだと思う。

文学作品や哲学書など文献学的に研究している者にとっての「文献」は、そこに書かれた内容の方に重心があり、モノとしての文献の位置づけは内容読解のために必要な前段階であってゴールではない、という意識が強いのではないと思われる。例えば文献学にとっての写本は、自筆本でも

ない限り作者が書いたオリジナルを不完全に写し取ったコピーであり、本文批判によって残された写本からできる限りオリジナルに近づいたうえで、それを解釈し作者の真意を追求することが文献学の基本的な考え方である（いささかナイーブな説明だとは思ふが）。本文批判においては写本自体から推測する「外的証拠」だけでなく、内容から推測する「内的証拠」も重視されるから<sup>16)</sup>、モノとしての文献の価値は、文献学全体から見れば部分的なものでしかない。書誌学や古文書学が文献学や文献史学の「補助学」<sup>17)</sup>という位置づけで分類されていることから、モノとしての文献の位置づけはわかるだろう。

そしてこのような内容重視の姿勢が電子テキストの重視につながっているのではないと思われるのである。

#### ＊ 文献学以外の研究分野における「文献」

もちろん、「文献」を扱うのは文献学者だけではない。図書館や公文書館、博物館や美術館などでは、テキストの内容もさることながら、まさにモノとしての文献の諸要素——色や形状、材質などの情報が重要である。

また、これらの機関においては、文献の保存と公開が課題としてあるが、古典籍や貴重書などを保存・公開するためにデジタル化する場合と、公文書館が公文書を保存・公開する場合は、当然のことながらデジタル化の方法が異なる。

他にもいろいろな場合が考えられるが、いずれにせよ、「文献」には方法や目的に応じた様々な側面（層）があることをふまえて、適切な「漢字文献情報処理」を模索しなければならないのではないだろうか。

## ■ デジタル画像をめぐる諸問題

### ＊ 撮影・スキャン

では、文献のデジタル画像化には、どのような問題があるのだろうか。簡単にながめておこう。

文献のデジタル画像化において、技術的にもっ

とも大きな部分はやはりデジタルカメラやスキャナを使った撮影・スキャン作業であろう。そしてそのなかでもっとも関心が払われるのが、どのような機材を選ぶか、という点であろう。

機材には大きく分けてカメラとスキャナの二種類がある。さらにカメラには、直接デジタル化するデジタルカメラと、フィルムのカメラで撮影したものをフィルムスキャナで読み込むという二種類の方法がある。スキャナの場合、一般的なフラッドヘッドスキャナのほか、本を傷めずに効率的にスキャンするためのブックスキャナなどもある。

機材の選定にあたっては、解像度や色の再現性、（カメラの場合）レンズによるゆがみなどへの配慮が必要である。文献学者であれば、とりあえず文字が読めればよいという程度の要求しかしないかもしれない。また、ウェブで公開して不特定多数の人々の閲覧に供するというような場合も、インターネット回線の制約上、それほど高い解像度は求められないだろう。

逆に、古文書学や美術史などの分野においては、紙の繊維が見えるぐらいの高解像度が求められたり、厳密な色の再現性が求められたりするであろう。また場合によっては、本特集の志水一行氏の報告にあるように、赤外線などを使った特殊な撮影方法も必要になってくる。

なお、色について一言しておくが、存在論・認識論に関わる哲学的な議論がなされる一方で、あまりにもありふれていて問題意識が持たれないという意味で、電子テキストにおける文字論に匹敵するぐらいに奥が深い問題領域ではないかと筆者は考えている。

### ＊ メタデータ

撮影・スキャンされデジタルデータとして蓄積された画像は、放置しておくとも再利用が難しい巨大なデータの山になってしまう。電子テキストであれば grepなどで全文検索をすれば、見失ったデータでも容易に探し出すことができるが、画像の場合、検索できるのはせいぜいファイル名や撮影日時ぐらいである。本特集の各報告には、画像データの管理の重要性と目録データベースや



カタログニングなどについて述べられているが、いつ、どこで、何を撮影したのか等々の情報は、検索したり共有したりする際に非常に重要である。

このような情報は一般にメタデータとよばれる。デジタルアーカイブにおけるメタデータの重要性や方法については多くの論考で議論されているのでそちらにゆずり、ここではひとつだけ付け加えておきたい。これまで述べてきたように、文献（だけに限らないが）のデジタル化においては、それがどのように利用されるのか、という処理に依存する。学術研究に限定して言えば、研究の方法によって、デジタル化の要求が異なるのである。デジタル画像のメタデータを作成する場合、これがどのような目的で撮影されたのか、どのような方法で利用されることが前提になっていたのかなどをメタデータとして記録することは、処理の場や研究領域を超えた共有と利用を念頭に置いた場合、重要なことではないかと思われる。

#### ✳ 所蔵者との折衝、著作権処理など

本特集の各報告を読んでもわかるように、文献の撮影や公開においては、所蔵者との折衝や、著作権などの処理が大きな問題となっている。特に、大量に複製品があるわけではない一点ものの写本などの場合、文献の持ち出しに関する交渉、持ち出せない場合は保管場所内での撮影場所の確保などが必須であるし、文献の状態によっては撮影前後の修復が要求される場合もある。また万が一、撮影対象を破壊したり紛失したりした場合の補償なども、考慮に入れなければならないだろう。

文献の著者の著作権が切れていない場合には、当然のことながら著作権の処理が必要である。では、著作権の保護期間がとっくの昔に終わっている古典籍などはどうであろうか。もちろんその場合、著作権による保護はないが、だからと言って撮影し放題というわけでもない。モノとしての文献（を含む諸資料）の撮影などを所蔵している博物館などが制限することができるのは、一般に言

われている著作権ではなく、民法に規定された所有権にもとづく、というのは、本研究会の著作権講演会において石岡克俊氏が明らかにされたことであるが<sup>18)</sup>、このような法律面の知識は（その重要性が認識されているにも関わらず）驚くほど共有されていない。

## 最後に

以上、思いつくままに書き連ねてきた。このほかにも様々な論点があるが、紙幅の都合とご了承いただきたい。画像データをめぐる議論が、「漢字文献情報処理」全体の議論を豊かにする方向で発展することを祈念しつつ、ひとまず筆を措きたいと思う。

## 注

- [1] <http://www.jaet.gr.jp/guiding.html>
- [2] 人文科学におけるコンピュータ利用の研究史における「デジタルアーカイブ」概念の変遷については、永崎研宣「デジタルアーカイブの弁証法」（『情報処理学会研究報告』Vol. 2005, No. 105 (2005-CH-068)、2005年10月）に詳しい。
- [3] <http://www.jaet.gr.jp/JAET-BBS/contents.cgi?room=ac4&mes=94>（本研究会会員のみ閲覧可能）
- [4] 安岡孝一「テキスト情報階層モデル」（『漢字と文化』第2号、2004年2月）
- [5] データベースやデジタルアーカイブの開発が、その研究分野の方法や史料の見方を記述することになる、ということについては、拙稿「情報歴史学の教育に挑む」（『歴博』第140号、2007年1月）で簡単に述べた。
- [6] バート・D・アーマン（松田和也訳）『捏造された聖書』（柏書房、2006年6月）、170ページ。
- [7] Amazon ではつい最近まで、古文書学などの書籍を「歴史補助学」というカテゴリで分類していた。
- [8] 石岡克俊「収蔵作品へのアクセスと法」（『漢字文献情報処理研究』第5号、2004年10月）

# 寺院文化財調査における デジタル写真撮影

## 花園大学歴史博物館の調査活動を事例に

志水一行（しみず いっこう）

### ■ はじめに

文化財調査での調査写真がその後の研究において貴重な資料となることは周知の通りである。調書をはじめとする記述形式では記録することのできない膨大な情報を記録することができる写真は、調査・研究資料として重要な位置を占め、これまでもにおいても文化財との関係の中でその重要性が語られてきた<sup>[1]</sup>。

したがって、当然のことながら調査者・研究者には写真のノウハウおよび撮影技術が問われる。そのノウハウ・技術に関しては、しばしば博物館・美術館の研究者、あるいは文化財写真の撮影に携わる専門家によって現場における事例が報告されている<sup>[2]</sup>。また、研究者を対象に文化財写真の撮影技術について解説された、鈴木昭夫・吉田成・岡宮誠一・田口栄一・鷺野谷秀夫『研究者のための資料写真の撮り方』<sup>[3]</sup>は、文化財撮影の必携マニュアルとしてその拠りどころとされている。

しかし、近年の文化財調査における調査写真撮影の現状として、これまでの撮影方法に大きく変化した点がある。それはデジタルカメラを使用しての撮影である。印刷物用としてなどの一部の撮影を除き、ここ数年における文化財調査写真はデ

ジタルカメラでの撮影が一般化している。もはやフィルムカメラからデジタルカメラへの移行は、その過渡期を過ぎたといっても過言ではない。フィルムカメラに代わってデジタルカメラが用いられる理由はいくつかあるが、①撮影した画像をその場で確認できること、②大量のカット数が撮影可能というコスト面での利点にあるといえよう。さらにデジタル画像は調査後の画像管理をスムーズにさせるとともに、PC上で各種アプリケーションソフトを用いることにより、新たな視点に立った研究の可能性を研究者に与えてくれる。

このように、デジタルカメラの普及によって、調査写真の撮影方法および撮影後の写真管理はここ数年で大きく変化してきていることは明らかである。こうした状況を踏まえ、文化財調査における写真撮影の現状を報告するとともに、デジタル撮影は撮影およびその後の研究にいかなるメリットを与えてくれるのか、その可能性について考えてみたい。

なお本稿においては、あらゆる文化財調査のなかでとりわけその機会が多い寺院調査における撮影についてとりあげたい。その際、花園大学歴史博物館における事例として、当館にて開催された2008年度春期企画展「春日局ゆかりの寺 麟祥院展」<sup>[4]</sup>に際しての調査・撮影事例<sup>[5]</sup>を紹介す

ることによりその報告にかえたい。

まず、本展覧会の概要を述べておく。本展覧会は妙心寺（京都市）山内塔頭のひとつである麟祥院の協力のもと、当院が所蔵する文化財、なかでも絵画作品を中心に展覧するものである。

当院は徳川3代将軍家光の乳母である春日局の菩提寺として著名である。局が帰依した単伝士印の法孫碧翁愚完を開祖として寛永11年(1634)に創建され、現在にいたるまで数多くの文化財を伝えている。とりわけ、狩野探幽筆「春日局像」をはじめとする局ゆかりの品が所蔵されている。

寺院の文化財調査はその目的によって種々の方法があるが、本調査においては、当院に収蔵されるすべての文化財の目録制作が最優先事項としてあり、展示対象となる書籍・絵画分野に関しては悉皆調査となった。本調査における撮影までの作業過程は次のようである。

先述したように、当院の場合は所蔵文化財目録が未制作であったこともあり、まずは様々な機能・役割を有する境内すべての建物から文化財として認められるものを抽出する作業から始まった。第2の作業としては、文化財の分野（書籍・絵画・彫刻・陶磁・金工・染織など）ごとの分類があり、第3に形状（絵画を例にあげてみれば掛軸装・卷子・屏風・障壁画など）ごとの分類をおこなった。

## ■ デジタル撮影

### ＊(1)写場

寺院の文化財調査の撮影において、その進行を大きく左右するのが撮影場所である。博物館・美術館には写場・スタジオなどの設備が整っているが、当然のことながら寺院にはそのような施設はなく、図録用写真の撮影など一部の事例を除き、スタジオを組むことは少ない。いずれにせよ、寺院の一室を借りて撮影することになるのであるが、まず如何なる部屋が撮影場所に適しているか選択しなければならない。

通常、調査写真の撮影は調書作成と同時進行で行うケースが多く、調査部屋と同室で撮影するこ

ととなる。本調査においても限られた時間のなかで多くの文化財を調査・撮影する必要があったため同室にて行った。その際、どの部屋が撮影場所に適しているか選ぶ条件としては、①屏風などをはじめとする大型作品の撮影が可能であること。②撮影するにあたり必要最低限の光量が得られることなどである。

だが、ここにあげた条件を満たしていたとしても、必ずしも撮影がスムーズに進行するとは限らない。先の条件②を例にあげるならば、比較的光量を得られる明るい部屋を撮影場所として選んだとしても、太陽光を受け入れやすい部屋は時間帯によって差し込む光量は変化し、一日を通しての撮影場所の明るさが安定しない。したがって、露出面——絞り、ISO感度、ホワイトバランスなど——の調整が常時必要となる。これは一例にすぎず、寺院での撮影にあたってはあらゆる状況の想定と、それに応じた準備が必要とされる。

繰り返すようであるが、限られた調査時間内での本撮影においては、撮影機材などのハード面を充実させるとともに、それを活用したソフト面を連携させることによってその問題をカバーすることを試みた。

### ＊(2)カメラ

本撮影では4台のデジタルカメラを用いた。使用したカメラはデジタル一眼レフカメラ〈OLYMPUS E-330〉・〈Canon EOS 5D〉と、〈SONY Cyber-shot T30〉・〈SONY Cyber-shot DSC-H9〉である<sup>16)</sup>。なお、調査時間の関係上、三脚は使用せず手持ちによる撮影をおこなった。

本撮影において2台のデジタル一眼レフカメラを使用したのは、限られた時間内での撮影において標準ズームレンズとマクロレンズの交換の手間を省くためであり、標準ズームレンズ〈ZUIKO DIGITAL 14-45mm F3.5-5.6〉を装着した〈OLYMPUS E-330〉と、マクロレンズ〈EF100mmF2.8 マクロ USM〉を装着した〈Canon EOS 5D〉を用意した。標準ズームレンズを装着した〈OLYMPUS E-330〉は被写体の全図撮影用、マクロレンズを装着した〈Canon EOS

5D) は細部撮影用という使い分けである。また、〈OLYMPUS E-330〉は内蔵ストロボを、〈Canon EOS 5D) はマクロ撮影に必須のリングライト〈MR-14EX) を使用し、ストロボにおいてもレンズの特性に応じて使い分けた。

なお、〈OLYMPUS E-330〉はフルタイムライブビュー機能を搭載した可動式液晶モニターを備えていることから、ファインダーを覗くことが困難な箇所の撮影に適しており、本撮影においてもその特性を活かした。大幅作品の高部の撮影、そして複写台撮影の要領を応用しての冊子・巻子の撮影である。無理な姿勢での撮影を少なくし、安全かつ正確に撮影することができるメリットがある。

さらに、〈SONY Cyber-shot T30) を使用したのは、本 Cyber-shot T シリーズに搭載されてい

る「拡大鏡モード」撮影機能を活用するためである。書籍・絵画の調査の場合、細部撮影より近接しての款記・印章の撮影が必須としてあり、一般的には細部撮影と同様の撮影方法をとるが、マクロレンズを使用した場合でも撮影距離には限度がある。その問題を解決し、より鮮明に印章を撮影することを可能としたのが「拡大鏡モード」撮影であり、本機能を使用した場合、1cm まで近接しての撮影ができる<sup>[7]</sup>。それに加え、本シリーズ純正アクセサリである白色 LED が採用されたリングライト〈SONY HVL-RLS) を本機に装着した場合、内蔵ストロボ使用時に起こるハレーションが抑えられ、適確な露出による撮影が可能となる。調査時の印章および細部撮影を飛躍的に簡易にするものといえよう。

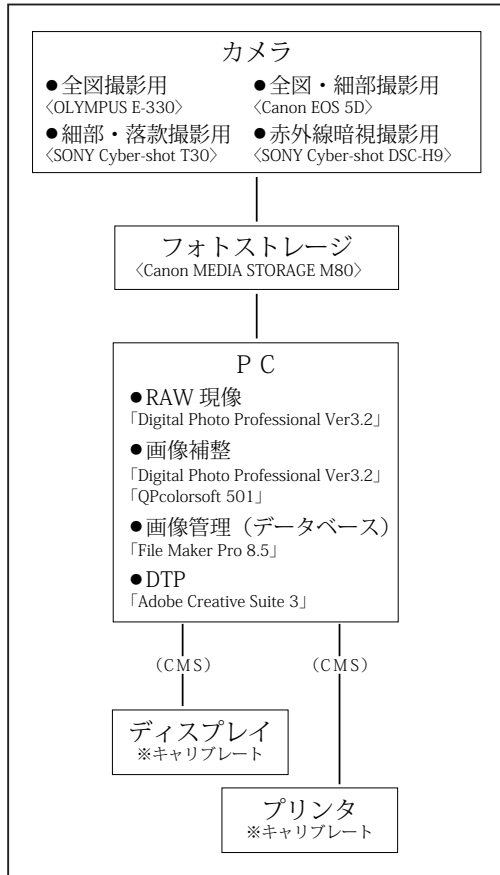
そして、〈SONY Cyber-shot DSC-H9) であるが<sup>[8]</sup>、本機は赤外線カットフィルターを外した赤外線暗視撮影機能、いわゆるナイトショット機能が搭載されている。赤外線写真はこれまでも頻りに文化財調査で用いられてきた撮影方法のひとつである。赤外線写真撮影には種々の調査目的があるが、一般的には肉眼で確認することが難しい墨書、例えば肖像画の賛、款記、箱書などの墨書を判読するための補助の役目を果たしてくれる<sup>[9]</sup>。したがって、本機能がデジタルカメラに搭載された意義は大きい。すなわちデジタルカメラの場合、調査現場で撮影した画像を確認することができるため、調査現場において画像を実物と照らし合わせて判読する理想的な調査が行えるのである。調査の記録写真としてのみならず、調査現場においても一資料となりえる。

### ＊(3)周辺機器

ここで、本撮影における記録ファイル形式について触れておく。本撮影では4台ともJPEG形式にて記録し、〈OLYMPUS E-330)・〈Canon EOS 5D) はともにRAW撮影を行った。〈Canon EOS 5D) を例にあげるならば、RAW + JPEG (ラージファイン) という記録形式である。

先にデジタル撮影の利点のひとつとして、大量のカット数が撮影可能であることをあげたが、

花園大学歴史博物館 画像管理システム構成図



RAW 撮影の場合、1 ファイル 10M を超える大きなサイズで生成されるため、一時的にデータを保存・バックアップする大容量のストレージが必要となる。そこで、本撮影においては大容量のフォトストレージ〈Canon MEDIA STORAGE M80〉にてバックアップを行った。バックアップストレージとしては、この種のフォトストレージを用いなくても PC で補えるのであるが、撮影現場に持ち込んだ PC が後述する RAW 現像ソフトをもちあわせていない場合、RAW 画像を表示・確認することができない。その点、RAW 画像を表示・確認できる〈Canon MEDIA STORAGE M80〉は<sup>11)</sup>、撮影現場においての画像管理及び詳細情報の閲覧が可能であり、バックアップストレージとして以上の機能をはたす。

#### ✳(4)ソフトウェア

これら撮影したデジタル画像、すなわち RAW ファイルは各メーカー純正の RAW 現像ソフトによって現像作業を行う。つまりこの場合、PC が暗室となるわけである。当館では、RAW 撮影での大容量の画像データを円滑に処理することができる高性能 CPU を搭載した PC を使用し、データ処理の高速化を図るとともに、24 インチワイド TFT 液晶ディスプレイにて画像閲覧におけるストレスを少なくした環境を整えている。

大抵の RAW 現像ソフトにはコントラスト・ホワイトバランス調整機能が備わっており、露出面の補正がおこなえ、そこで生成された画像データには ICC プロファイルが付加される。さらに〈Canon EOS 5D〉の使用例として、メーカー純正の RAW 現像ソフト「Digital Photo Professional Ver3.2」には上記の画像補正機能のほか、レンズ収差補正機能として歪曲補正機能がある。実際に目にした実物と同様の情報を記録することが求められる文化財写真において、レンズ収差・歪曲はそれを異にする大きな要因であり、文化財写真を扱う研究者において今後注目される機能のひとつになりそうである。

また、画像補正に関して付け加えておこなうならば、本撮影においては QPcard AB 社のデジタル用カ

ラーチャート〈QPcard201〉を被写体に添えて撮影した。これまでのカラーチャートとさほど変わりはないが、〈QPcard201〉用アプリケーションソフト「QPcolorsoft 501」を用いれば、画像のカラーバランス・ホワイトバランス調整が簡易に行える。したがって、博物館・美術館・研究機関に完備されたハードウェア（ディスプレイやプリンタ）などにキャリブレーションがなされていれば CMS——Color Management System——を構築することができ、独自に DTP——Desktop prepress——を行うことができる。当館においても CMS を構築した上で、「Adobe Creative Suite 3」にて広報物（ポスター・チラシ類）、図録のデザイン制作を独自に行っている<sup>11)</sup>。

そして、これらの画像の管理であるが、「File Maker Pro 8.5」による画像データベースを構築し画像管理をおこなっている。まず調査画像のサムネイル画像を作成し、主テーブルのオブジェクトフィールドに OLE オブジェクトとしてサムネイル画像を埋め込み、一方で元画像をリンクさせる形式をとった。さらに、展覧会準備のデータ（キャプション・解説テキスト・出品目録など）とリレーションシップを組むことにより、調査から準備までのデータを一括管理している。

また、デジタル撮影が画像管理の面に与えた影響は大きく、それに関するソフト類も他に比して充実している。本調査では用いなかったが、GPS ユニットキット〈SONY GPS-CS1KSP〉を同社製デジタルカメラと併用することで、「Picture Motion Browser」・「Super Mapple Digital Ver.7 for Sony」にて地図上に撮影画像をマッピングすることができる。画像補正同様、今後新たな視点での画像管理も可能となりそうである。

## ■ まとめ

近年、寺院での文化財調査においてデジタル撮影が一般化している現状を踏まえ、その現状として花園大学歴史博物館の調査事例を報告するとともに、デジタル撮影が撮影および調査後の研究に与えるメリットについて考えてきた。

デジタル撮影のメリットとしては、先述の通り撮影時に画像を確認できることと、大量のカット数が可能ということがあげられる。すなわち、「質」と「量」の双方に利点があり、それらの利点が結びつくことによって、目的に応じた良質な資料写真をこれまで以上に多く撮影することができるのである。あらゆる状況を把握して撮影に望む必要がある寺院での撮影において、撮影環境面におけるリスクを軽減し、より確かな研究資料として記録することを可能にした意義は大きい。

また、調査後の画像整理・管理をスムーズにし、画像補整もストレス無く簡易に行えるようになった。本稿においては、一部のアプリケーションソフトを用いた事例を紹介するにとどまったが、ソフト面と連携することで最大の効力を発揮するデジタルは、調査後の作業および研究に新たな可能性を与えてくれる。

しかしその一方で、文化財調査におけるデジタル撮影の一般化は、博物館・美術館あるいは研究者が独自に画像調整・管理を行う必要性をこれまで以上に大きくしたと思われる。デジタルはハード面とソフト面が連携して初めてその特性が最大限に生かされる点、ハード面の充実とソフト面を柔軟に活用する一貫したシステムの構築を独自に行う必要がある。したがって、写真のノウハウおよび撮影技術のほか、撮影から管理までのシステム構築が文化財調査においてデジタル撮影を行う際の課題<sup>[12]</sup>としてあることを最後に付け加えておきたい。

## 注

- [1] 岡塚章子「写された国宝——日本における文化財写真の系譜」(『写された国宝——日本における文化財写真の系譜』東京都写真美術館、2000年)、『写真と文化財の関わり』(日本写真学会、2004年)、『月刊文化財』(「特集 写真と文化財」517号、2006年)
- [2] 前掲注1『写真と文化財の関わり』・『月刊文化財』517号
- [3] 鈴木昭夫・吉田成・岡宮誠一・田口栄一・鷺野谷秀夫『研究者のための資料写真の撮り方』(理工学社、1991年)
- [4] 2008年度花園大学歴史博物館春期企画展「春日局ゆかりの寺 麟祥院展」、会場：花園大学歴史博物館(無聖館4階・第2展示室)、会期：2008年4月2日(水)～6月28日(土)、<http://www.hanazono.ac.jp/museum/2008syunki>
- [5] 調査は2007年4月26(木)～28日(土)に麟祥院にて行った。調査構成員は花園大学歴史博物館学芸員および花園大学文学部文化遺産学科・花園大学歴史博物館協会(ボランティア組織)の学生である。
- [6] 本撮影に用いたカメラを含め後述する周辺機器・ソフトウェアのスペックに関しては、各メーカーのホームページ上の製品情報を参照されたい。
- [7] 光学ズームはW側に固定。
- [8] 本撮影で用いた〈SONY Cyber-shot DSC-H9〉はオーバーサイズモデルである。国内においては、〈SONY Cyber-shot DSC-H50〉がナイトショット機能搭載モデルとして現行発売されている(2008年7月現在)。
- [9] 佐多芳彦氏による、赤外線暗視撮影機能を搭載した〈SONY Cyber-shot DSC-F828〉を用いての調査報告がある。佐多芳彦「長林寺所蔵「長尾政長像」について——中世武家服制再考の素材として——」(『栃木史学』19号、2005年)
- [10] ただし、Canon製のデジタルカメラで撮影したRAWファイルのみしか表示されない。
- [11] 花園大学歴史博物館2007年度秋期企画展図録『宇治人形——知られざる茶の木人形の世界——』(花園大学歴史博物館、2008年)、なお花園大学歴史博物館2008年度春期企画展図録『春日局ゆかりの寺 麟祥院展』も現在制作中である。
- [12] このほか銀塩とデジタルとの間で様々な問題点があり、これまでもデジタル撮影における問題点についての指摘がある。

# 文献のデジタル画像化

## ——『周作人書簡集』影印本製作（失敗の巻）——

小川 利康（おがわ としやす）

### ■ デジタル事始め

この数年来、筆者は種々の縁に恵まれ、中国人作家・周作人（1885-1967）とその翻訳者・松枝茂夫（1905-1995）との間で交わされた書簡（戦前 65 通、戦後 107 通）の整理校訂に取り組み、排印版として「周作人・松枝茂夫往来書簡」（戦前篇 (1)(2)(3) 及び戦後篇）を発表した。中国側でも筆者の提供するデジタル画像に基づいて独自に整理校訂する形で排印版が刊行されている<sup>[1]</sup>。

だが、残念ながら、念願の影印本自体の刊行はならず、そのために作成した画像データ（90GB 弱）は現在に至るまでハードディスクに死蔵されている。いずれ日の目を見ることを期待しているのだが、小文では、その作業での経験で役に立つと思われる部分（失敗談ともいう）を記し、読者諸兄の参考に供したい。

この仕事は、2001 年に北京で中国人研究者から故松枝茂夫所蔵の周作人書簡が閲覧可能か質問されたのが縁で始まった。その際のやりとりで、書簡が百通以上あると知った彼の地の研究者は是非カラーの影印本を出版してはどうかと提案してくれ、さらに影印の許諾を取るために遺族を紹介してくれた。早速会って話してみると、遺族も乗り気では非中国で出版しようという。その上、遺族宅で放置されていた大量の書簡から松枝茂夫のものを発見し、提供して下さった。文革中は「抄

家」で一旦は没収されたものの、返還を受けた資料であるという。この発見はまことに望外のものであり、遺族の松枝ナヲ女史（松枝茂夫夫人）も大変喜んでくださり、出版計画にもご了解いただけたので、さっそく出版準備を進めることになった。

ご承知の方も多いかも知れないが、従来は古典でも相当良い版本でもない限り、カラーの影印本などは出版されなかったものだが、近年は近代文学の作家の日記や書簡も多数影印刊行されている。周作人関連だけでも二種類ある。

#### ◆『周作人兪平伯往来書札影真』

（北京図書館出版社、1999 年 6 月版）

紅学の大家として著名な兪平伯との往復書簡 350 通余りを全て影印したもの。倣線装本で 2 冊本（箱装）。多色コロタイプ印刷により、毛筆の繊細な筆致ばかりか精緻に描かれた便箋の絵柄までも忠実に再現している。680 部の限定版だ。

#### ◆『江紹原蔵近代名人書札』

（江小蕙編、中華書局 2006 年）

中国民俗学のパイオニアとして知られる江紹原が交わした書簡 159 通を影印したもの。魯迅、林語堂らの書簡も含むが、実は周作人の書簡が全体の 3 分の 2 を占めている。こちらは洋装本オフセット印刷ながらフルカラーで書簡の雰囲気をも十分に表現している。



『周作人俞平伯往来書札影真』

近年の中国の印刷技術向上は目覚ましく、ひと頃には考えられなかったような影印本が多数刊行されている。惜しむらくは価格が高く、刊行部数が少ないことだが、有り体に言って、中国でも買うのは一部の好事家だけなのだから、致し方ないのかもしれない。

なにより周作人書簡の影印本を日本で出すとしたら自費出版以外不可能だが、中国では商業出版の望みが十分あるという。ならば、その可能性に賭けてみることに否も応もない。筆者は美しい影印本が刊行されるのを夢見て、是非にとお願いした。思えば、これが坎坷の道の始まりだった

## ■ 著作権問題

これまで本誌でも議論されてきた問題だが、デジタル化する以上、そこに著作権にかかわる問題を処理しなければ前に進めない。今回のケースでは以下ようになる。

- 周作人書簡（136通、執筆時期 1936～

1965年）

所有者：松枝茂夫遺族

著作権者：周作人

- 松枝茂夫書簡（36通、執筆時期同上）

所有者：周作人遺族

著作権者：松枝茂夫

周作人側は故人長男の夫人・張茨芳女史が代表者として、松枝側は故人の夫人・松枝ナヲ女史が代表者として、それぞれ著作権に関する委任状を書いて下さった。

委任状の内容は、それぞれ日本における出版及び印刷に関する権利に関しては筆者に一任し、その代わり中国における出版及び印刷に関する権利は周作人遺族に一任するものとした。

筆者は周作人遺族に日本で出版できたとしても現実的に経済的利益は見込めないこと、また松枝家としても故人の残した書簡を出版することで経済的利益を望んでいないことを伝え、周家にも了解を求めたのである。周家としても現物は長年松枝家にある以上、多くを望む気持ちはなく、了解していただけた。ただ、中国で刊行する場合には一任して欲しいとの申し出があり、同等条件なので了解した次第である。実際の結果としても、前述通り影印本は刊行できていないので、経済的利益は皆無に等しい。もしも実利を伴うものだったら、書簡の所有比率も異なるのだから、こんなに簡単に事は運ばなかっただろうと思う。

仲介者となった筆者としては、利益の見込めない話に快く応じてくださった双方の遺族に対して心から感謝申し上げたい。

著作権問題は簡単に解決できた。だが、その先の技術的な問題はなかなか容易ではなかった。

## ■ 自力でデジタル化(1)

### ✳ デジカメで撮る

中国で影印本を出版するにしても、その素材となる画像データを作らねばならない。作業のうえで一番楽なのは、書簡を中国の出版社に預けてデ



ジタル化してもらふことだ<sup>[2]</sup>。だが、この方法はリスクが大きすぎる。筆者が北京までオリジナルを運び、スキャンが終わるまで一通たりとも紛失せぬよう監視し、安全に日本まで持ち帰らねばならない。よしんばその労力を厭わぬにしても、万が一紛失した場合の責任までは負いきれない。周作人遺族側もその点は承知しており、早々に諦めていた。

そこで筆者は自分でデジタルカメラ（以下、デジカメと略す）を使って接写することにした。周作人遺族の側でも特段妙案もないので了解してくれ、とりあえずやってみるようになった。

筆者にしても、決してカメラについて精通しているわけではなかったが、書籍の接写に関しては相当場数を踏んでいる。まだコピー機が自由に使えなかった 80 年代末の上海図書館で随分フィルムカメラを使って解放前の雑誌や単行本を接写したことがあるのだ。デジカメはマクロ撮影モードがついており、フィルムカメラなどより簡単に接写できる。手元には生前の松枝茂夫先生から頂いた書簡集の白黒コピーがあったが、これでは文字の判読すら難しいので、さしあたってパイロット版を作るぐらいのつもりで挑戦することにした。

2003 年の春休みの某日、松枝邸に出向き、朝から夕方まで黙々とシャッターを押し続けた。正確な記録は残していないが、1 通につき数葉にわたる書簡もあり、日付の消印を確かめるために封筒も撮影するため、撮影枚数は 360 枚を越えた。

以下に書き記すのは、写真撮影について専門教育を受けたことのない者が経験的に身につけたノウハウなので、その点を十分了解のうえ読まれることを希望する。

#### ✳ デジカメ選び：RAW 現像、歪曲収差

通常の研究目的のデジタル化ならば、最近発売のデジカメのどの機種を使っても問題ない。普及機のコンパクト・デジカメの場合、画像ファイル形式は JPEG が一般的だが、ネット上で公開する程度の用途ならば、それでも十分な画質だ。

だが、より高画質を追求するならば、RAW 形式の画像<sup>[3]</sup>で記録することをお勧めする。RAW

は、カメラの CCD が捉えた電気信号を、文字通り生のまま非圧縮で記録しているので、撮影後の加工調整（フィルムカメラの紙焼きの連想から現像と呼ぶ）の自由度が高く、高画質の撮影には必須のファイル形式だ。一般に普及機種で利用される JPEG は、RAW を不可逆圧縮（冗長性の多いデータは間引きされる）したものなので、映像を加工する度に、画質が劣化する。

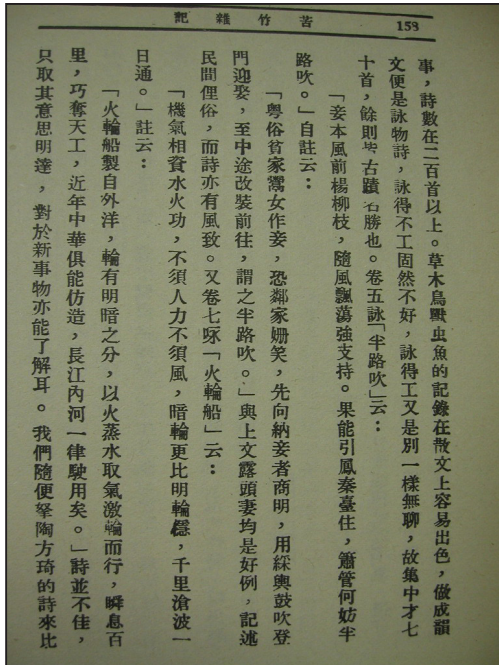
以下で紹介する 2003 年当時の撮影では、記憶メディア容量も足りなければ、RAW をサポートするデジカメも高級一眼レフに限られていたため、利用しなかった。だが、近年は、大容量の記録メディアも安価に入手できるうえ、普及機種でも RAW をサポートしているものが増えた。もし予算など条件がゆるすのなら、RAW 撮影できる機種を選んだ方がよい。

更に文字資料を撮影する際に考慮するポイントは、接写機能だろう。接写機能を一般にはマクロと呼び、接写距離が短いことを売り物にするメーカーが多いが、ほとんどの場合、花や昆虫の撮影を考慮して設計されているため、本当に書籍接写に適しているかどうかは分からない。実際に試し撮りして「歪曲収差」がどれくらい出るか調べる必要がある<sup>[4]</sup>。ここで「歪曲収差」について詳しく説明するゆとりも能力もないが、端的に言えば、被写体が広角レンズなら樽型に広がり、望遠レンズなら逆に細く映ることをいう。

デジカメの場合、普及機を中心にはほとんど広角レンズを採用しているので、程度の差はあれ、みな樽型になる欠点を抱えている。広角レンズは焦点距離が短く、接写に向いているが、書籍資料の場合、直線的な形状であるため、この欠点が通常よりはっきり出やすい。接写距離が近ければ近いほど出やすいので、レンズの能力ぎりぎりまで被写体に寄らず、ある程度光学ズームを活用して、距離をあけて撮影した方が歪みも目立たない。ソフトウェア的に「歪み補正」する機能のあるカメラもあるので、幾つか試して選ぶと良いだろう<sup>[5]</sup>。

#### ✳ デジカメ接写：撮影環境

接写に限らないかも知れないが、撮影で配慮す



歪曲収差が明瞭に出た撮影例

べき事は以下の二点に尽きる。

◆明度：太陽光を十分確保する

技術がなくても、普通のカメラと十分な明度があれば、文字を読むうえで問題ない撮影が出来る。十分な明度を確保するために撮影は昼間に行う方がよい。多少の曇りでも問題ないが、必ず南向きの窓際の明るいスペースを確保する。ストロボで明度を補おうとすると、一枚ごとに仕上がりに差が出てしまい、調整に苦勞するので止めた方がよい。



◆安定：三脚でカメラを固定する

効率よく確実に撮影するために、三脚を使い、カメラが撮影対象にまっすぐ向き合うように調整しよう。明度が確保できていても、接写で手ぶれしたら致命的だ。手ぶれ補正機能を当てにするよりも三脚を使った方がよい。一人で撮影する場合は、本をめくるなど、同時にやるべき作業があるので、なるべく手はあけておいた方がよいのだ。

三脚でカメラを固定したら、何枚か試し撮りしてレンズが撮影対象に対して真っ直ぐに向き合っているかどうか必ず確認しよう。歪曲収差が出ていないかの確認も必要だ。机の上で本が多少斜めになっても、ソフトで調整できるが、斜めから撮影した写真は通常補正できない<sup>16)</sup>。

✳ デジカメ接写：撮影作業

◆用意する機材

ここから実際の作業に入る。

2003年春、この作業に取り組んだ際に使用した機材は OLYMPUS C-40 ZOOM (400万画素、光学 2.8 倍ズーム) であった。安定した画質で評価が高かった名機で、JPEG 画像のほか、コンパクト・デジカメとしては珍しく非圧縮の TIFF でも撮影できる仕様だった。当然画質を優先するならば、TIFF で撮影した方が高画質で撮影できることは分かっていたが、一枚撮るごとにスマートメディア (以下 SM、最大 128MB) への書き込みで時間がかかり、なおかつ 300 枚以上の撮影には頻りにメディアを交換せねばならぬ事を考えると、余りにも無理があった。さらに後述するように、最高画質の TIFF で撮影したとしても、2272×1704 ピクセル (24bit) が最高なので、大判の影印本の品質にはどのみち達していなかったと考えられる。

結果として、選択したのは 1280×960 ピクセル (24bit) の JPEG 画像だった。これならば途中 2、3 回画像を PC に待避させれば全て撮り終える事が出来る。

撮影当日、松枝邸に持参した機材は以下の通り。

- デジカメ本体

- 予備バッテリー、予備 SM
- 三脚 (カメラ固定用)
- 黒の画用紙 (書簡撮影用の台紙)
- 白い上質紙 (色見本用)
- 手袋 (資料保護)
- ノートパソコン (撮影画像転送用)

余り説明は要らないだろうが、貴重書を扱う以上、気休めかも知れないが、手袋は一応必要だ。黒い画用紙は撮影する書簡の下に敷く台紙として用いた。必ずしも黒である必要はないが、透かし絵のある便箋もあるので、裏に当てておけば透かし絵も撮影できる。新聞のように裏の字が透けてしまう資料にも黒紙は有効だ。ほかに白い上質紙を用意するのは、ホワイトバランスを決めるための色見本である。色調豊かな撮影を行う場合はカラーの色見本が必要だが、この場合はホワイトバランスさえ間違えなければ大丈夫だ。

#### ◆ ホワイトバランス調整

上記機材を配置して、スタンバイできたら、まずホワイトバランスを決める。日中の太陽光が十分ある状態で、上質紙の白がどのようにカメラに認識されるかを基準として決めておき、その後の撮影の色あいを決める判断とさせる。これを怠ると同じ光でも、太陽光や室内灯では色温度が異なるので、赤みがあった白やら、青みがあった白になってしまう。これを色かぶりと言い、その問題を回避するのがホワイトバランスだ。

書籍接写の場合、撮影する対象は白と黒の中間色に相当するので、このバランスは大切になる。プロ用に 18 パーセントグレーの色見本が売られているので、それを使用しても良い。その白の基準となる上質紙を撮影し、マニュアルでホワイトバランスを補正しよう。ほとんどの機種で蛍光灯、太陽光などとプリセットの値が用意されているが、マニュアルで補正した方がよいようだ。もし撮影途中に急に光量が変わるようなら再度補正する。

ここまで準備が整えば、後は一気呵成に撮るだけだ。ストロボは当然発光禁止にセットして、太陽光だけで撮る。光学ズームを利用して良いが、

一度固定したら変えてはならない。さもないと仕上がった画像の大小比率が揃わず、資料としての価値が半減してしまう。

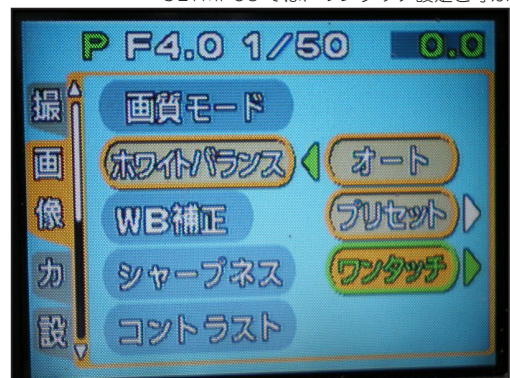
途中、何回か画像を PC に転送し、撮り漏らしがないかどうかチェックしよう。取り直す機会が二度と無いケースがほとんどのはずだ。JPEG では大幅な色合いの変更は出来ないので、撮影に失敗がないかどうか確認が必要だ。特に白飛びと呼ばれる露出オーバー (明るすぎ) には注意しよう。一般にアンダー気味 (暗すぎ) に撮影したものはソフトウェアで補正できるが、白飛びしたら補正は不可能だ。

#### ＊ 撮影後の作業

##### ◆ ファイル整理の重要性

完成したファイルは撮影時の記憶が残っているうちに整理した方がよい。周作人書簡の場合は年ごとにフォルダを作成し、書簡の執筆年月日が確定できるものは、年月日でファイルネームを付けた。例えば、19431215.jpg といった形式である。確定できないものは 1943a.jpg のように、アルファベットで通し番号を付け、校勘作業のなかで順次問題を解決していった。大量のファイルに一括して名前を付ける際には専用のソフトウェアがある<sup>[7]</sup>ので、煩雑な作業もかなり省力化できる。ファイル名が決まったら、ファイル名一覧をエクセルにでも書き出し、内容一覧を作っておこう。ここをこまめにやらないと混乱の元になる。実は筆者の場合、整理が進むにつれてファイル名も変

OLYMPUS では、ワンタッチ設定と呼ぶ



更するのをしばしば怠ったため、行方不明になるファイルが続出した。ファイル管理は綿密にやった方がよいというのが経験的な教訓である。

筆者が撮影したのは周作人書簡の方だけで、松枝茂夫書簡の方は周作人遺族宅の方で撮影したものを添付ファイルで送ってもらった。とりあえずフラットヘッドスキャナで特に何か設定をいじったりせず取り込んだとのことだった。彼らの方で取り込みの際に画像をトリミングしたためかサイズは均等でないが、平均して1280×960ピクセル(24bit)で、十分鮮明な画像だった。北京にある松枝書簡は本当に影印本を刊行するとなれば、簡単に出版社へ持参してスキャンし直せばよい。当面の文字入力には、この画像で十分用が足りる。

## ◆文字入力・校訂

書簡の影印本を刊行するにしても、読みやすい排印版はどうしても必要である。印刷品質も低く、読みにくい『周曹通信集』(香港南天書業公司一九七八年刊)がネット上でテキスト化されているように、現実的には影印本だけでは研究に使用できないのだ<sup>18)</sup>。とはいえ、筆者ひとりで全部文字起こしをしていたら、とても間に合わない。

そこで、2003年末頃から整理が終わったファイルをA4判に印刷すると、順次親しい中国人留学生、日本人学生に文字入力を依頼した。

中国人でも読めない文字は幾つもあるだろうと予想されたので、分かる文字だけ入力し、分からないところは■で印を付けておくよう依頼した。この時、もっと厳格な文字入力の方針を立てておくべきであったが、その時点では単純に中国語は繁体字(正字)で入力し、フォントはMingLiUとすること、日本語は旧仮名遣いのまま忠実に入力するという方針しか伝えていなかった。

この結果、仕上がってきた入力原稿には幾つかの問題点があり、それぞれ筆者が確認作業をするなかで修正していったが、判断に迷うものも少なくなかった。

- 問題点:繁体字が原則だが、略字体も多く、簡体字で入力される場所が多数あった

→すべて繁体字に修正。

- 問題点:文中の書名や引用は出典に当たり、その表記に従うようにしたが、誤記や転記ミスなどが相当数存在する。

→テキストを忠実にどこまで残すか判断が難しく、統一性に欠けた。

- 問題点:日本語の旧仮名遣いは維持したが、文中の漢字表記には中国語(繁体字、簡体字)、日本漢字(旧字、新字)が混在している。

→全て統一するのは無理があると考え、「書いてある状態に近い」書体で入力することにした。このため「東京大学」と『大學』など、文中で言及する書名や固有名詞などの字体がバラバラになってしまった。

結論から言えば、表記のぶれを認めた方がテキストへの忠実さは高まるが、検索性は犠牲となる。その点からすれば、オリジナルから大きく乖離するとしても、日本語は現代仮名遣いに改め、漢字表記は繁体字に統一してしまった方が良かっただろう。この問題を十分に解決せぬまま、排印版を作ってしまったのは大いに反省するところである。

書簡の校訂作業は文字表記の問題だけでなく、他にも色々な問題をはらんでいるが、ここでは技術的な問題だけに絞り、影印本用に行った画像スキャンの話に移りたい。

## ■ 自力でデジタル化(2)

### ✳ スキャニング

#### ◆ 高品質の画像作成のために

2003年末から、筆者が文字入力と校訂作業を進める一方で、周作人遺族は複数の出版社との交渉を進めてくれていたが、なかなか刊行への展望

が開けなかった。一つの原因は周作人自身の歴史の評価の問題であり、漢奸という評価によって出版が忌避されやすいこと、もう一つの原因は言うまでもなく採算そのものが取れない出版物であることが災いしていた。2004年は足踏み状態のまま終わり、ほとんど諦めかけたところで、2005年の春先になって、朗報が入ってきた。上海の某出版社と話し合いを持った結果、実現の方向で話し合いも順調であるとの由で、影印本の作成に必要な画像として使えるかどうか、筆者の作成した画像ファイルも関係者に見てもらったという。

そのうえで彼らが伝えてきた反応は予想通り影印本のクオリティにはほど遠いので、再度撮影が必要であるというものだった。また、デジカメでは必要なクオリティが確保できないので、フラットヘッドスキャナを使い、TIFFのファイルを出して欲しいとの要請だった。

スキャナは、デジカメよりも必要な機材が多くなり、出先で気軽に使えるものではないが、メリットも少なくない。スキャナは上記のデジカメで取り上げた光量や歪曲収差の問題を考慮する必要が無く、昼夜を問わず作業ができて、高画質という点ではスキャナの方がはるかに優れている<sup>[9]</sup>。スキャナでも広く利用されているのはJPEGだが、デジカメの項でも触れたように不可逆圧縮であるため高画質の撮影には適さない。スキャナの場合は印刷業界でも利用されている非圧縮のTIFFがデファクトスタンダードであるので、出版社からの要請も当然のものだった。

とはいえ出版社側から提示されたスキャンの条件は極めてハードルの高いものだった。

- 光学解像度 1200dpi (24bit)
- サイズ：原寸 100%
- ファイル形式：非圧縮 TIFF

影印本には高い品質が求められるのは当然としても、この条件は通常のスキャンのレベルを遙かに超えるものである。ここでは資料がA4判であるとして、単純計算すると、

$$\begin{aligned} & (\text{原稿サイズ} \div 1 \text{ インチ}) \times \text{解像度 dpi} \\ & = \text{横} (210\text{mm}) \div 25.4\text{mm} \times 1200 = 9921 \\ & = \text{縦} (297\text{mm}) \div 25.4\text{mm} \times 1200 = 14031 \\ & = 9921 \times 14031 \text{ pixel} \end{aligned}$$

という巨大ファイルができあがる計算となる<sup>[10]</sup>。

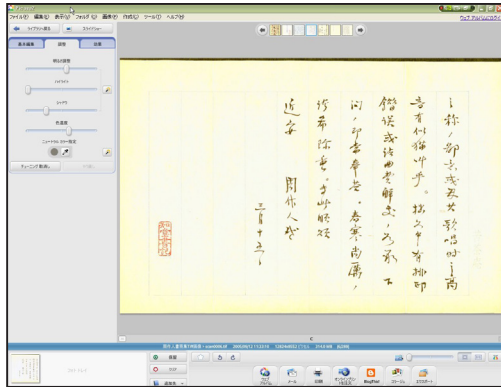
実際にスキャン作業を試みに行ってみると、非圧縮のTIFFの場合、果たせるかな30MB近くのファイルができあがった。1枚あたりのスキャン所要時間も1~2分かかる。むろん所要時間は機種に依存するので、一概に言えないだろうが、2005年当時使用していたHP Photosmart 2610(複合機)では光学解像度(最高2400×4800dpi)をある程度まで上げると、スキャンが格段に遅くなってしまふ。このため書簡一枚一枚セットする時間も計算に入れると、1枚こなすのに3分強かかり、360枚を休みなしにやるとしても、18時間かかる(!)という計算になる。全ての画像を収めるには36GB以上のディスクスペースも必要だ。普通のプリンタで打ち出す程度ならば、300dpiで原寸のままスキャンする程度でも十分な画質(2480×3508 pixel)が得られるはずで、この条件には当惑せざるを得なかったが、高画質であることに越したことはない。腹を括って作業に取り組むことにした。

#### ✳ スキャニング：昼夜兼行

2005年夏、再び松枝邸に伺って、スキャンの作業をさせていただくことになった。今回は重装備ではあるが、作業はむしろデジカメよりも単調である。

- ノート PC 本体
- スキャナ本体 HP Photosmart 2610
- 外付けハードディスク (250GB)
- 黒の画用紙 (書簡撮影用の台紙)

今回はホワイトバランスも気にしなくても良い。スキャナ内部のCCD自体が調光するので、極端な話、真っ暗な部屋でもスキャンは出来る。問題



Picasa2

はスキャンに要する時間である。筆者は二日間朝10時から夜8時まで作業をして、何とか全てのスキャンを完了させた。

簡単な段取りだけでも書き出してみると、スキャナ付属のソフトウェア（この場合はHPディレクタ）から画像のスキャンをメニューから呼び出し、「ファイルに直接保存」を選び、上記設定でスキャンする。HPのソフトウェアの仕様では連続して何枚でもスキャンが出来るようになっていたが、不具合が心配なので、10枚程度スキャンしたところで、必ずファイル保存をするようにした。保存メニューが出たところで、ファイル形式としてTIFFを指定する。ソフトウェアによってはTIFFのマイナーなオプションもサポートしているようなので、非圧縮でフルカラー（RGB）のTIFFを選ぶように注意しよう。後は根気強くスキャンするだけである。一日で百数十枚を八時間かけてスキャンした後の疲労感（徒労感？）は経験した者でないと分からないものがある。

かくて完成したTIFFは、RAWと異なり、現像作業は基本的に必要ない。筆者は早速ハードディスクごと複製を作成し、ちょうど別件で日本に滞在していた周作人の遺族に託した。これで一件落着のはずだった…だが、現実はそう甘くなかった。

## ✦ スキャンング：レタッチ

苦労してスキャンした画像は版下用に作成したので、日常的に参照する必要はない。だが、その後も出版計画は進まぬなか、ハードディスク上で

無用の長物と化したファイルを再利用しようと、やはりレタッチすることにした。1枚100MB近くもあるファイルは開くだけで時間がかかり、いかに画像が精細といっても実用にほど遠かったのである。

レタッチソフトは高機能高価格の製品からフリーウェアまで極めて多くのソフトがあるが、ここではPicasa2（Google）を紹介しておきたい。機能も決して豊富ではないが、最低限の調整が出来て、そのうえRAW現像も一応できる。

Picasa 2はインストールが済むと直ちにローカルディスク上の全てのメディアファイルを検索登録する機能がある。レタッチするうえで不可欠ではないが、ライブラリを開く時に軽快な動作を望むならば、空き時間に検索登録させておいた方がよい。ディスク全体ではなく、一部だけに限定したい場合は、[ツール]から[フォルダマネージャ]を開いて、監視対象フォルダを限定することも出来る。

Picasa2のライブラリ画面で該当の画像ファイルをクリックすると、レタッチ画面に入る。サムネイルがあれば、あらかじめ縮小画像が内部で作成されているので軽快に開けるはずだが、画像ファイルが大きい場合は、一括処理ソフト<sup>[11]</sup>で別途縮小版を作成しておいた方が楽だ。

左のメニューから「調整」を選び、「明るさ」「色温度」などを調整することで、画質を調整できる。ニュートラルカラーはホワイトバランスのことで、画像のなかで基準となる白をポイントすることで自動調整できる。画像の調整が済んだら、そのまま書き保存するか、エクスポートを選択して、JPEGとして出力すればよい。Picasa本来の機能として、Google上のPicasaウェブアルバムにアップロードすることも出来るので、ウェブ上で公開したい資料なら、一連の作業がワンストップで出来るので便利だ。

## ■ 二重の遺憾

2006年10月ごろ、周作人遺族から最後の契約書調印の段階で出版社が出版契約を取り下げた

旨、連絡が来た。今後も継続して出版社を捜す努力は続ける意志はあるようだったが、これまでの経緯から考えれば、状況が相当変わらない限り、望みは薄そうである。遺憾な結果というほかない。だが、それ以上に深刻なのが、遺族の松枝ナヲ夫人の体調悪化だった。この頃すでに幾つかのアクシデントが重なって、すっかり衰弱されて、入退院を繰り返す状態となっており、これ以上悠長に影印本を待つわけには行かなかった。

そこで周作人遺族に事情を説明し、日本で排印版を刊行することにしたと伝え、了解を得た。これは当初の合意事項の範囲内であるが、当然の礼儀としてお知らせした。そのうえで急遽作成した原稿を12月初旬に入稿し、翌2007年3月に「周作人松枝茂夫往来書簡」が公開できる見込みがあった。ところが、2007年1月13日、ようやく初校が出た頃、夫人は逝去された。とうとう間に合わなかった。残念至極というほかない。

## ■ まとめ

以上、周作人書簡集の画像データ作成にまつわる体験談を書き連ねてきた。近年、研究資料のデジタル化は急速に進みつつあるのは慶賀の至りだが、いざ自らが取り組む側に回ってみると、戸惑う事柄も多い。小文もまた試行錯誤の結果報告であって、明快な方向性を示すものではない。読者諸兄のために玉を引く磚となれば幸いである。

## 注

- [1] 『文化論集』(30～33号、早稲田商学同攻会刊2007年3月～2008年9月)。早稲田大学商学部のウェブ([http://www.waseda.jp/w-com/outline/publications/w\\_thesises01.html](http://www.waseda.jp/w-com/outline/publications/w_thesises01.html))からPDFをダウンロードできる。また、中国では「周作人と松枝茂夫通信」と題して『中国現代文学研究叢刊』(2007年4期～2007年6期)に連載された。ただし、収録されたのは周作人側の書簡のみ。
- [2] 後、日本の業者に頼む手もあることに気づいたが、個人で負担するには躊躇われる金額であったため、当時は断念した。
- [3] RAW画像とはデジカメの感光体(CCD、CMOS)の電気信号をそのまま記録したものだ。『デジタル「写真の学校」』(キットタケナガ著、雷鳥社2005年刊)などを参照することをお勧めする。
- [4] 書物でなくても、壁やタバコケースなど直線で構成されている者で撮れば良い。
- [5] カメラ内部でソフトウェア的に自動補正する機種も多い。自分で補正したい場合はレタッチソフトImageFilter(庭師芝達氏作、フリーソフト、<http://homepage2.nifty.com/niwashiclub/index.htm>)を利用すると良い。
- [6] 現在は一部のカメラには斜め撮り補正機能があるらしいが、筆者は未検証なのでお勧めしない。
- [7] Exif Renamer (Software factory 作、フリーウェア、<http://softwarefactory.jp/japan/products/exifrenamer/index.html>)はWebアルバムの自動作成機能も備えた秀作だ。Exif情報の取り込みに対応したソフトを利用しよう。
- [8] 肖毛——看云居：[http://5352919.blog.hexun.com/9669238\\_d.html](http://5352919.blog.hexun.com/9669238_d.html)を参照。
- [9] デジカメは画素で表示され、スキャナはdpiであるため、単純な画質の比較は難しい。
- [10] 「Cyber Librarian 図書館員のコンピュータ基礎講座」(<http://www.asahi-net.or.jp/~ax2s-kmt/n/index.html>)「参考資料・画像のスキャニング・サイズ」に明快簡潔にまとめられている。
- [11] BatchGOO! (LNsoft Entertainment、<http://www.lnsoft.net/>)など。当然ながらある程度の劣化は避けられないが、オリジナルは別に残しておけば問題ない。

# 報告： 学会誌論文のデジタル化

山崎 直樹 (やまざき なおき)

## 1 はじめに

日本中国語学会では、その機関誌『中国語学』に掲載された論文のうち、特に優秀なものに対し、各年度毎に「学会奨励賞」を授与している（通例、1～2篇）。そして、これらの学会奨励賞受賞論文はデジタル資料化され、同学会の公式ウェブサイトで全文を閲覧できるようになっている（画像1参照）<sup>1)</sup>。

筆者は、同学会のウェブサイトを管理するウェブソース委員会のメンバーとして、この奨励賞論文のデジタル化とウェブサイトへの掲載の業務に関与している。

この文章では、この作業の概略を報告したい。

画像1



## 2 デジタル化の方式に関する一般的な事柄

紙媒体で発表される論文をデジタル化するには、大雑把に考えて、次の2種類の方式がある。

- A) 電子組版データを、直接、PDF に変換する。
- B) 紙媒体の機関誌を裁断して、各ページを光学的にスキャンしてデジタル画像にする。

(A)の方式は、最も簡便でコストがかからない。これを実現するために必要な条件は、電子組版のシステムを構築する際、PDFに出力可能な組版ソフトを用い、PDFに埋め込み可能なフォントを用いることであろうか。

なお、余談ながら、学術団体が、将来、機関誌に掲載した論文をデジタル資料化することを考えているのなら、機関誌の組版を発注する業者を選定する際、この条件は非常に重要である。

欠点は、閲覧環境によっては、紙面を完全に再現できるかどうか、保証ができないことであろうか。

(B)の方式は、(A)の方式が適用できない場合の代替措置と考えられる。適用できない理由としては、以下の3点が考えられる。

- (i) 業者が電子組版を採用していない。
- (ii) 電子組版を行う業者が、データをPDFに



出力できないソフトを採用しているか、PDFに埋め込めないフォントを使用している。

- (四) 過去に遡って電子化を行う場合、過去の業者にコンタクトが取れない。あるいは、業者にデータが残っていない。

(B)の方式の欠点は、(A)に比べ読みづらいこと、文字列の検索がしにくいこと、などであろうか。長所は、閲覧環境による再現性に差がないことである。

なお、(B)の方式は画像を使うため、ファイルの容量が大きくなるのが懸念されるかもしれないが、現実には、フォント埋め込みのPDFと、容量においては大差ない。

### ■ 3 中国語学会学会奨励賞受賞論文のデジタル化について

中国語学会の「学会奨励賞論文デジタル化」は、前章で述べた(A)(B)2つの方法を併用した。以下の通りである。

- B方式（画像化）…246号（1999年）の受賞論文（※第1回）～251号（2004年）の受賞論文
- A方式（組版データからの出力）…252号（2005年）の受賞論文～253号（2006年）の受賞論文（※最新の受賞論文）

直近の252号・253号を発注した業者は、組版データからのPDF出力に問題ないシステムを採用していたので、A方式によった。

参考までに付言すると、まだ受賞論文が決定していない254号（2007年）と、現在、編集が進行中の255号（2008年）も、この業者に発注しているので、同様に組版データから出力したPDFが得られる見通しである<sup>[2]</sup>。

251号以前に関しては、複数の業者が関与しており、試みに、ある業者に問い合わせたところ、PDFに出力できないシステムを用いていたこと

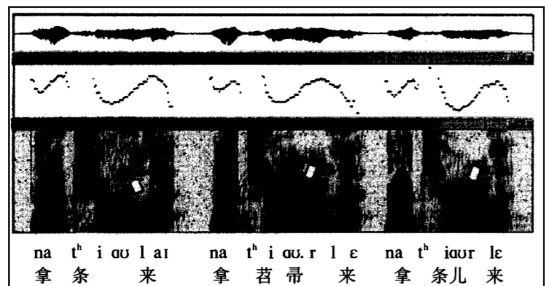
が判明した（技術的に表現すると、文字のアウトライン情報を抽出できないOCFフォントを使用するシステムであったため、レイアウトを崩さずにPDFを作成することができなかった、ということになる）。時代的に考えて、おそらく他の業者も同様であろうと考え、また、逐一、問い合わせる煩を厭い、残りは一括してB方式（画像化）を採用することにした。

### ■ 4 画像の仕様

光学的にスキャンする方式では、解像度：400dpi、白黒2値で画像化した。解像度がやや高めのようなのだが、この機関誌は、中国語学の多岐に亘る分野の論文が掲載されるので、下記の(1)(2)点を考慮し、複数の解像度設定でスキャンしたサンプルを検討し、この解像度に決定した。

- 1) 馴染みの無い文字（僻字、方言字、構造の複雑な非漢字…etc.）や、字形の微細な差異が重要な文字（国際音標字母など）が使われることもあるので、その識別性を高めるため。
- 2) オリジナルの複雑な画像（詳細な方言地図、スペクトログラム、研究対象の文献の複写画像…etc.）の可読性を高めるため（画像2はその一例。ただし、このサンプルは、将来の画像化のことも考えて特に選んだもので、この論文自体は、奨励賞受賞論文ではない）。

画像2 朱春躍・2004。「北京話声母及音节脱落现象的实验分析」、『中国語学』251, pp.1-13より



画像3は、組版データから出力したPDFを倍率200%で表示し、その画面をキャプチャーしたものである。

また、同じく画像4は、画像化によりデータ化したPDFを倍率200%で表示し、その画面をキャプチャーしたものである。

## ■ 5 学会内の理解と今後の展望について

当学会の近年の役員（会長、理事、事務局……）は、学会の有するリソースのデジタル化とその公開には、非常に積極的であるので（予算上の措置の面で特に）、上述の作業を遂行するにあたっては、常に理解と協力が得られた。

これまでの流れから考えるに、日本中国語学会では、今後は、学会誌に掲載された論文すべてをデジタル資料にして、何らかの形（閲覧制限あり／ない、有償／無償…etc.）で公開していくことが検討されるであろう<sup>[3]</sup>（紙媒体の機関誌の販路とも関係するので、実現までには、まだ多くの時間と議論が必要であろう）。

このような試みが実行に移される際、近年の号に掲載された論文については、上述の通り、即座

にPDF化することが可能であるが、過去の誌面をデジタル化するにあたっては、上記のB方式（画像にしてPDF化）を採用することになろう。その時には、今回の経験が生かせるはずである。

### 注

- [1] <http://www.chilin.jp/journal/award.html>
- [2] 2009年以降の業者は未定であるが、ウェブリソース委員会としては、§1で述べたように、デジタル資料としても公開することを前提とした組版ができる業者を選定してもらいたいと考えている。
- [3] 権利問題について付言するなら、近年は、学会誌への投稿の際に、デジタル資料化してウェブ公開することへの承諾を得ているが、それ以前の誌面のデジタル化に際しては、著者に対し、個別に承諾を取る作業が必要である。

参考までに、現在の「『中国語学』投稿規程（2007年5月改訂、2007年11月一部修正）」の関係箇所を抜粋して、以下に掲げる。

#### 六. ウェブでの公開

採用論文の本誌掲載に当たっては、将来これを学会がウェブ上で公開することに著者が同意したものとみなす。その際、学会は以下の二点を遵守する。

- (1) 著者の氏名を付記する。
- (2) 著者の了解なしに論文（サマリーを含む）の内容に変更を加えない。（<http://www.chilin.jp/journal/contribution.html>より抜粋）

画像3 宇都健夫. 2006. 「“就是”と“只是”——補足注釈機能について——」, 『中国語学』253, pp.314-334より

現代中国語の“就是”は、様々な機能・用法を持つ語である。その中で、本稿が特に興味深い用法として注目するのは、以下の例(1)、(2)のようなものである。

- (1) 你父亲可真是好人哪, 就是不大会做生意, 啊, 不大会做生意。  
老舍《二马》<sup>1)</sup>  
(お前の親父さんは本当に良い人じゃが、ただ商売が不得手じゃな。)

画像4 東ヶ崎祐一. 2003. 「『説文解字繫伝』にみられる反切下字混用——梗撰入声と曾撰入声、および外転一等韻と二等韻の間の——」, 『中国語学』250, pp.32-49より

反切下字の混用は韻母の合流あるいは音声的近似を反映しているものである。梗撰と曾撰の韻尾は、唐末～五代期にはともに硬口蓋的な $-ŋ, -k$ であったと考えられている（花登1974）。そして昔韻も職韻もともに三等韻であり、口蓋化介音 $-i-$ をもつ。2つの韻の違いは、主母音が広い $e$ であるか（昔韻）、狭い $ie$ であるか（職韻）だけである。

# ハーバード大学における 蔵書デジタル化支援体制

——韓国古典籍デジタル化事業を中心に——

江上 敏哲（えがみ としのり）

## ■ はじめに

筆者は2007年3月から1年間、当時在籍していた京都大学附属図書館からの派遣で、ハーバード大学に図書館業務実地研修のため滞在し、現地の様々な図書館活動・運営の実例を調査してきた。ハーバード大学をはじめとするアメリカの大学図書館では、図書館蔵書、特に著作権の切れたものや古典籍、写真、図版資料について、そのデジタル化とオンライン公開が非常に活発である。ハーバード大学では図書館内に撮影やデジタルデータの保存等を請負う専門の部署が設けられており、学内の蔵書デジタル化事業を支援する“インフラ”として機能している。

本稿では、筆者が滞在していた Harvard-Yenching Library（以下 HYL）での韓国古典籍コレクションのデジタル化事業を例に、ハーバード大学図書館における蔵書デジタル化事業の実際を紹介したい。

なお、このデジタル化事業をはじめとする本研修での見聞・調査については、京都大学図書館機構 web サイト上で「ハーバード日記」<sup>[1]</sup>と題して随時報告してきた。本稿と併せてお読みいただければ幸いである。

## ■ 1.ハーバード大学図書館の概要

本論に先立ち、ハーバード大学とその図書館について概説しておきたい。

ハーバード大学はボストンに隣接したケンブリッジ市にある。アメリカ最古の大学であり、開学は1636年。学生約20000人、教員約2300人である。

学内には大小あわせて80以上の図書館・図書室があり、全学で約1600万冊の書籍を持つ。全館が部局・分野の枠を超えて Harvard University Library と呼ばれる図書館群を構成している。館長以下、ハーバードの旗艦図書館である Widener Library、学部生用学習図書館である Lamont Library、筆者の研修先である HYL 等が属する。また、全体の図書館業務をカバーする各種事務局・施設・プロジェクトが組織されている<sup>[2]</sup>。Weissman Preservation Center（古典籍・特殊資料専門の保存・修復部署）、Harvard Depository（郊外保存書庫とその管理組織）、Harvard-Google Project（Google Library Project を担当）等。

HYL<sup>[3]</sup>は、Harvard-Yenching Institute 内の図書館として1928年に設立された、東アジア研究図書館である。蔵書数は約120万冊で、東ア

ジア資料の図書館としては、北米では議会図書館、UC バークレーに次ぐ規模である。選書・レファレンス担当の各言語サブジェクトライブラリアンのほか、古典籍専門のライブラリアン等がいる。この図書館で開発された「イェンチン分類」は、初代館長の裘開明が独自に開発した、中国語・日本語資料整理のための分類法である。

## ■ 2. 韓国古典籍コレクションデジタル化事業

現在 HYL で進められている韓国古典籍コレクションのデジタル化事業は、「National Library of Korea - Harvard-Yenching Library Korean rare book digitization project」と称されている。これは、HYL と韓国国立図書館<sup>[4]</sup>との共同プロジェクトであり、HYL に所蔵されていて韓国本国にはない古典籍約 500 タイトルを、デジタル化するものである。

もともと韓国国立図書館では、韓国国内にある韓国古典籍ばかりでなく、日本・中国等で所蔵されている韓国古典籍のマイクロフィルム化・デジタル化を継続的に行なっている。本事業はその延長であり、アメリカでの初のプロジェクトであるという。なお、HYL 以外にも米国内の同様のコレクションをデジタル化するプロジェクトが、計画として進行しつつある、ということだった。(2008 年 2 月当時)

2006 年 11 月に HYL におけるプロジェクトの計画と交渉がスタートし、実作業は 2007 年夏から 3 年間。撮影・デジタル画像作成はハーバード大学と HYL で行ない、年間約 6 万ドルと見積もられる費用は韓国国立図書館が負担する。結果として得られた画像データは、双方がそれぞれ所有・公開する、という契約である。

HYL が持つ韓国古典籍のうち、どの資料をデジタル化の対象とするかについては、韓国国立図書館側が選定を行なった。当館刊行の冊子体目録である『The annotated catalogue of Korean rare books at Harvard-Yenching library, Harvard university』<sup>[5]</sup>を丹念にチェックした結果、選ば

れたタイトルの多くが写本であったとのことである。

実作業は、おおまかに言って以下のような流れで行なわれる。

保存状態確認 → 撮影・デジタル化 → 格納  
→ 公開 及び 目録登録

## ■ 3. 保存状態確認——Weissman Preservation Center

HYL の韓国セクション担当者によって撮影対象資料が準備されると、撮影前に、Weissman Preservation Center (以下 WPC)<sup>[6]</sup>から資料保存の専門家 (conservator) が派遣され、現物の状態をひとつひとつチェックする。

WPC はハーバード大学図書館における貴重資料専門の保存部署である。学内の古典籍や特殊資料 (貸出用図書ではないもの) について、物理的状态や保存状態をチェックしたり、修復・保存のための処置を施したりといった活動を行なっている。

蔵書デジタル化事業が行なわれるにおいては、どのプロジェクトにおいても、上記のような conservator による資料のチェックが行なわれる。チェックの結果、適切な撮影方法や機材がアドバイスされることもあれば、事前に修復が行なわれることも、場合によっては撮影が断念されることもあるという。

このように蔵書デジタル化事業においては、資料保存活動やその部署との関わりが非常に重要視されている。「複製・媒体変換による資料の保存が、デジタル化の目的である」ということだけでなく、「撮影・デジタル化にあたってその資料が傷むことのないよう管理する、対策をとる」ということもその大きな理由のひとつである。そもそも、後に紹介する Imaging Service 自体、この WPC 内の一部署としてスタートしたという経緯があり、現在も両部署は Preservation & Imaging Department というひとつのグループとして活動している。

Conservator のチェックの結果、破損のためそのままの状態では撮影に耐え得ない場合には、WPC に一度持ち帰られて修復処理がされたり、製本がきつくてページが充分に開けない場合には、WPC でいったん製本を解いてから撮影されたりといった対応がなされる。なお、本事業においてはこれまでに撮影断念の勧告がされるほど保存状態の悪い資料は出ていないとのことであった。これには、そもそも韓国の古典籍は総じて紙が丈夫である、ということもあるようである。

#### 4. 撮影・デジタル化——Imaging Service

資料の撮影は、ハーバード大学図書館内の資料撮影専門部署である Imaging Service<sup>[7]</sup> によって行なわれる。

Imaging Service はハーバード大学図書館内の資料撮影・デジタル画像作成についての専門部署である。蔵書の写真撮影、マイクロフィルム作成、デジタル画像作成や、そのための保存修復・目録登録・メディア変換なども必要に応じて行なう。Widener Library の地下にあり、暗室・撮影スタジオ等の専用スペースが設けられている。マイクロフィルム用、デジタルスキャン用、オーバーヘッドカメラ等の撮影用カメラのほか、書籍資料用のブックスキャナー、貴重書専用スタジオ、地図等大型資料専門スタジオもあり、様々な種類の資料に対応している。今回紹介している韓国古典籍のプロジェクトには用いられないが、近代出版物のデジタルスキャンでは、OCR によるテキストデータ作成もこの Imaging Service で行なわれている。そのほか、デジタル化後のオンライン公開に欠かせないインデクス、チャプター、ページ情報といった管理用メタデータもここで取り扱われている。

特に古典籍の場合、学内での撮影が理想的である。この Imaging Service には約 40 人の技術者・カメラマン・画像専門家が専属職員として勤めており、質の良い写真・デジタル画像を外注より低価格で作成することができる。

本事業では、撮影はデジタルカメラで行なわれ、

マイクロフィルム作成は行なわれない。TIFF や JPEG による画像データが結果として作成される。担当者によれば、作成された画像の質には充分満足しているが、撮影者が韓国語を理解していないためか、ページのとぼしやダブリがどうしても生じてしまう、ということであった。

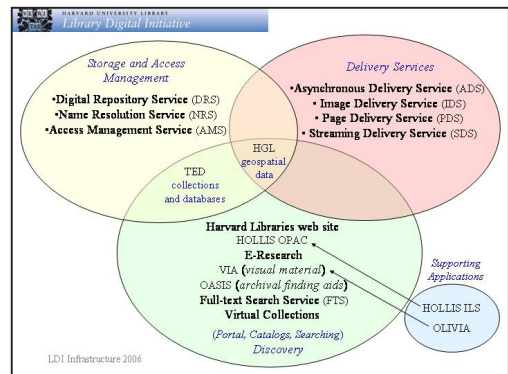
#### 5. 格納——Digital Repository Service

作成されたデジタル画像は、Digital Repository Service(DRS)<sup>[8]</sup> が提供するデジタルデータ保存システムの中に格納される。

DRS はハーバード大学図書館の Office Information System(OIS) が管理・提供しているサービスである。画像や音声などのデジタルデータを格納し、永年保存するとともに、検索機能やブラウジング機能によってオンラインで利用者に提供する。ハーバード内の図書館や機関・プロジェクトであれば、これを利用することができる。韓国古典籍の画像データのほか、Google Library Project によってスキャンされたハーバード大学蔵書の画像データも、この DRS に格納されている。

DRS は “デジタル資料における Harvard Depository (ハーバード図書館の郊外保存書庫)” を意図して構築されており、デジタルデータであればどのような内容の資料でも、この

デジタル資料の格納・公開・目録を示した図  
Office for Information Systems web サイトより引用  
<http://hul.harvard.edu/ois/systems/>



サービスによって保存・提供することができる。但しそこには、「利用者に提供することを意図したデジタル資料でなければならない（少なくとも学内に。将来的公開のためのアーカイブでも可）」、「著作権・知的所有権を持っているか、処理されていなければならない」、そして「データベースやwebページなど何らかのかたちで目録が提供されていなければならない」といった条件がつけられている。すなわち、単なるデータ置き場としてのみこのDRSというサービスがあるのではなく、利用者によるアクセスを促進し保証するためのものとして機能している。データは永年保存が保証されており、必要に応じてマイグレーションも行なわれる。こうしたメンテナンスにかかるコストをまかなうため、このサービスの利用には、1Gにつき年間20ドルという使用料が課せられている。と同時に、そのデータのオーナーである図書館または組織が、将来にわたってデータを監督する義務を負う。

## 6. 公開——Page Delivery Service ほか

格納されたデジタル画像を閲覧する際に使用されるのが、Page Delivery Service(PDS)というシステムである。書籍形態、すなわち、ひとつの資

料が複数のページからなる画像を持つタイプのデジタルデータについて、ページごとにブラウジングしていくためのインタフェースである。特別なビューワーのダウンロード等を必要とせず、一般のwebブラウザでそのまま閲覧することができる。ページ指定、表示サイズ切り替え等の基本的な機能がついているほか、資料によっては目次情報や全文テキストデータの閲覧や検索、印刷用PDFの作成も可能である。但し、そこまでの機能が活用できるのは、OCRでのテキストデータの生成が容易なラテン文字・近現代出版物がほとんどである見え、本事業の韓国古典籍の画像についてそこまでの機能は実現していない。

下記はその一例である。

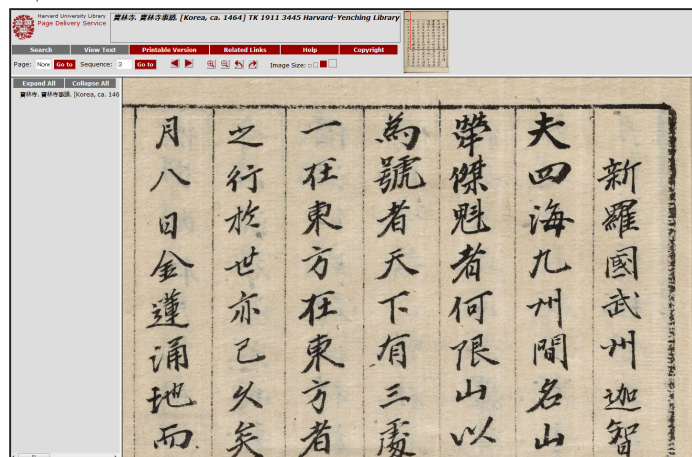
ここで表示されている簡易書誌は、イェンチン図書館の韓国資料担当者が準備したものである。また、PDS上で画像を提供する際にはページ・ファイル番号などの管理用メタデータが必要となり、今回の例ではImaging Serviceがそれを作成している。

本事業による韓国古典籍の画像データは、ハーバードの学内外から自由にアクセスが可能である。が、資料やプロジェクトの目的によっては、学内メンバーのみに公開しているもの、特定の部局やゼミの受講生のみにアクセスを許可しているもの等もある。ハーバード大学には、学内メンバー全員に付与されたIDとパスワードで、学内ネット

ワークやコンテンツ、webページ等へのアクセスの許可・不許可を管理する認証システムがある。このシステムが用いられ、アクセスのコントロールが行なわれる。

實林寺事蹟

<http://nrs.harvard.edu/urn-3:FHCL:1187878>



## 7. 目録登録 —— cataloging

公開された韓国古典籍の画像データへは、ハーバード大学図書館の蔵書検索データベースであるHOLLIS<sup>9)</sup>からアクセスできる。HOLLISで該当する韓国古典

籍を検索すると、その詳細データ中に画像ページの URL が記述されており、リンクをたどってアクセスすることがきる。プロジェクトの名称（ここでは「National Library of Korea - Harvard-Yenching Library Korean rare book digitization project」）も書誌レコード内に記述されることになっており、特定のプロジェクト下でデジタル化された資料を網羅的に検索することも可能である。

ハーバード大学での蔵書デジタル化事業においては、目録・検索ツールの整備が重要視されている。先に紹介した、DRS に格納するデジタルデータについては何らかのかたちで目録が提供されていないと条件がついているのもその一環である。図書館の蔵書である書籍資料を検索するための HOLLIS のほかにも、写真・絵画等のビジュアル資料が検索・閲覧できるデータベース・VIA<sup>[10]</sup> や、文書資料のリストやファインディングエイドを検索・ブラウジングできるデータベース・OASIS<sup>[11]</sup> 等が提供されている。プロジェクトによっては、独自のリストやデータベースを自館 web サイトから提供する場合もある。

カタログングを実際には誰がどう行なうかについては、それぞれのプロジェクトによる。各図書館所属のカタログガーが目録作業のすべてを行なう、Imaging Service に所属するカタログガーに依頼する、プロジェクトのために外部のカタログガーを期限付きで雇用する、等。WPC には写真資料のセクションがあり、そこに所属する写真専門のカタログガーに依頼することもできる。

本事業では、撮影対象である韓国古典籍はすべて HOLLIS に書誌レコードを登録済みであった。このため、実際にカタログングとして行なわれた作業としては、HYL 韓国目録担当者による手動でのリンク情報・所蔵情報の編集のみで済んでいる。当初は、各資料の目次情報を作成してナビゲーションのひとつとして提供する予定だったというが、人材確保が難しく、実現していないとのことであった。

## 8. 学内インフラとしての支援体制

ハーバードの各図書館が蔵書のデジタル化のためのプロジェクトを行なう場合には、今回紹介した各種の専門部署・サービスによるサポートを受けることができる。それは実作業の請負だけにとどまらない。

先に紹介した Imaging Service には Librarian for Collections Digitization と呼ばれる専門のライブラリアンが所属している。デジタル化事業を行なおうとする図書館は、プロジェクトそのものの計画や実際の進行について、このライブラリアンに相談し、アドバイスや提案、他部署との連絡調整といったサポートを受けることができる。例えば、撮影方法や画像データの技術的仕様はどれを採用すべきか、資料保存部署からはどのようなサポートを受けることができるか、目録やメタデータ等の作成は誰がどう行なうのがよいか、等。貴重資料のコレクションか、貸出用一般図書かによって、そのとるべき手法等も異なってくる。

このような支援体制や各部署・サービスは、ここ 10 年のうちに急速に整備されてきたものであるという。それ以前の蔵書デジタル化は、個々の図書館がそれぞれのプロジェクトとして、独自に、その場その場で行なわれていた。これを、できるだけ組織化・集中化させ、専門の部署に専門の技術者やライブラリアンを配置させることで、個々のプロジェクトに対して質の高いサポートを効率的に提供できるようになる。機材や技術だけでなく、人材・サービスや蓄積された知見など、プロジェクトの企画から公開までをトータルでサポートしており、個々の図書館やその担当者は一からすべてを自分で準備する必要がなく、スムーズにスタートさせることができる。専門の部署が組織されることによって、学内での蔵書デジタル化事業が一過性のものとして終わるのではなく、必要な限りその支援体制が保証されることになる。彼らがこの体制を“インフラ”形成として位置づけていることの意味が、そこにある。

HYL には、今回のデジタル化事業の対象となった資料以外にも、数多くの韓国古典籍が所蔵されている。これらについても別途デジタル化を実現させることが望まれており、その基金獲得を試みている、とのことであった。

## 注

- [1] 「ハーバード日記：司書が見たアメリカ」 <http://www.kulib.kyoto-u.ac.jp/modules/wordpress/>
- [2] 「Harvard University Library」 <http://hul.harvard.edu/>
- [3] 「Harvard-Yenching Library」(HCL Libraries, hcl.harvard.edu) <http://hcl.harvard.edu/libraries/#hyl>
- [4] 「국립중앙도서관」 <http://www.nl.go.kr/>
- [5] 尹忠男, 金成煥編著.  
하바드燕京圖書館韓國貴重本解題 = The annotated catalogue of Korean rare books at Harvard-Yenching library, Harvard university. 景仁文化社, 2005.6. (Harvard-Yenching library bibliographical series. 11)
- [6] 「Weissman Preservation Center」(Library Preservation at Harvard) <http://preserve.harvard.edu/wpc.html>
- [7] 「Imaging Services」(hcl.harvard.edu) <http://hcl.harvard.edu/info/imaging/>
- [8] 「Overview: DRS & Delivery Systems」(Office for Information Systems) <http://hul.harvard.edu/ois/systems/drs/>
- [9] 「HOLLIS Catalog」 <http://hollis.harvard.edu/>
- [10] 「Visual Information Access」 <http://nrs.harvard.edu/urn-3:hul.eresource:viaxxxx>
- [11] 「OASIS: Online Archival Search Information System」 <http://nrs.harvard.edu/urn-3:hul.eresource:oasisxxx>



# フリーソフトで出来る画像処理

山田 崇仁 (やまだ たかひと)

## ◎ はじめに

通常、画像によるデータベースを行う場合、ハードウェアとソフトウェアの負担が必須となる。もちろんメモリは搭載すればするほど作業は快適になるし、やはり Adobe Photoshop は値段相応の性能を見せてくれる。

しかし、論文の抜き刷りがかさばるからスキャンして画像で保存したい、といった用途には、ハイエンドのハードウェアと Adobe Photoshop の組み合わせは、パワーの無駄遣いにも見えてしまう。また、そのために数十万の投資をするには、あちこち（財布や家族）の同意を得る必要もあるだろう。

しかし、インターネット上を探せば、様々な画像処理機能を提供してくれるフリーなソフトが存在する。ここでは、それをいくつか紹介してみることにする。筆者が Windows ユーザーであるため、Windows 環境で動作するソフトウェアを中心に紹介する。

## ◎ 画像の編集

### ● GIMP

<http://www.gimp.org/>

### ● GIMP2 を使おう (日本語版のダウンロード先)

<http://www.geocities.jp/gimproject/gimp2.0.html>

GIMP (GNU Image Manipulation Program) は、ビットマップ画像の編集や加工を行うためのソフトウェアである。Adobe Photoshop ライクな高機能の画像処理機能の実装を目指して開発が進められ、多くのプラグインも公開されている。

### ● Inkscape

<http://www.inkscape.org/>

Inkscape はベクトル画像の編集を行うためのソフトウェアである。Adobe Illustrator の機能をフリーで使いたい方（無論全く同じ性能ではないが）におすすめである。

### ● Picasa

<http://picasa.google.co.jp/>

Picasa は、Google が提供するデジタル写真を管理する目的で開発されたソフトウェアである。本誌特集 2 の小川氏による記事でも紹介されているが、Picasa は単純な画像閲覧機能だけではなく、簡単な編集機能や、Google アカウントとリンクしたオンラインアルバム機能などを備えている。RAW 画像にも対応しているのは便利だろう。Mac OS X 版はまだ提供されていないがほぼ同様の機能を備えた iPhoto に対応したプラグインが提供されている。

### ● Picnik - 写真を簡単に、ブラウザからオンラインで編集

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

今まで紹介したものは、基本的にローカルな環境でアプリケーションを起動するタイプだが、Picnik は、Web サービスとして提供されている点が特徴である。Flash ベースのインターフェイスで構成され、ハードディスクにあるファイルを使用して画像修正を行うことも可能である。Firefox から画像を Picnik に送るアドオンも公開されている (<https://addons.mozilla.org/ja/firefox/addon/4889>)。)

●その他の画像処理ソフト

その他、総合的な機能を提供する画像処理ソフトウェアとしては、JTrim (<http://www.woodybells.com/jtrim.html>)・PhotoFiltre ([http://photofiltre.free.fr/frames\\_en.htm](http://photofiltre.free.fr/frames_en.htm)) などがある。

また、個別の画像処理機能を提供するソフトウェアもある。例えば、「藤-Resizer- (<http://www.vector.co.jp/soft/cmt/win95/art/se132646.html>)」や「縮小専用。 (<http://www.vector.co.jp/soft/win95/art/se153674.html>)」は高機能な拡大縮小を行うソフトであるし、「こみこんぶ (<http://www.vector.co.jp/soft/win95/art/se278069.html>)」は、スキャンした画像の傾きの補正や裏写りの消去などを行ってくれるソフトであり、抜き刷りの画像での保存に威力を発揮するだろう。また、少々古いが、高機能な減色機能を持つ Padie (<http://www.vector.co.jp/soft/win95/art/se063024.html>) は筆者にとって Web 版レジュメ作成に欠かせないソフトウェアとなっている。

◎ 画像閲覧ソフト

作成した画像は、閲覧ソフトを使わないとみることができない。一応、OS には標準で画像閲覧ソフトは実装されているが、それより便利なソフトはいくらでもある。

●IrfanView

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

●IrfanView32 日本語版

<http://www8.plala.or.jp/kusutaku/iview/>

画像のみならず、音声やビデオの視聴にも対応した老舗のマルチメディアビューワ。簡単なフォトタッチ機能に加え、プラグインにも対応している多機能ぶりである。

●XnView

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

海外のソフトウェアである。400 種類以上の画像に対応した閲覧機能に加え、簡単なフォトタッチ機能も搭載されている。初期状態で日本語にも対応しているのは便利。

●Susie for Win32

<http://www.digitalpad.co.jp/~takechin/>

日本で開発された老舗の画像閲覧ソフトである。本ソフトの特色は単なる閲覧機能ではなく（本体のみでは BMP 形式の画像しか閲覧できない）、対応画像を増やすための規格である Susie Plug-in にこそその真価がある。この規格はその仕様が公開されているために、第三者による多くの形式に対応した Susie Plug-in が開発されており、Susie 以外にも Susie Plug-in に対応した閲覧ソフトが多数公開されている。

●Leeyes

<http://www3.tokai.or.jp/boxes/leeyes/>

国産の見開き画像ビューワー。Susie Plug-in に対応しており、多くの画像形式に対応している。単ページ化された画像を見開き形式で閲覧できたり、ループ機能で画像を拡大して見られるなど非常に便利である。

◎ おわりに

以上、ざっとフリーで使用可能なソフトウェア（細かい使用条件については、各ソフトウェアの紹介ページを熟読していただきたい）について紹介した。

それ以外にも多くの画像処理ソフトが存在する。まずは定番の VECTOR (<http://www.vector.co.jp/>) や新しもの好きのダウンロード (<http://www.macsoft.jp/>) 辺りから探してみてもいいかだろうか。



# 実践レポート

近年、大規模な e-learning システムの導入で学生・授業マネジメントは効率化したのに、柔軟性を欠く「教材作成」機能のため、教師の創意工夫が十分発揮できていない例が散見される。Adobe が提唱する e-learning プラットフォームは、Acrobat や Flash のような汎用性の高い技術をベースにしており、比較的容易にリッチな教材コンテンツを制作できる。このシステムによって、教師は教材を「デザインする」力を我が手に取り戻せるのかどうか。Acrobat の最新バージョンと Captivate による教材作成・授業利用の実践を紹介する。

## 日本中国語 CAI 研究会について

日本中国語 CAI 研究会は、2007 年をもって研究会活動を休止した。本コーナーも、研究会の編集協力ではなく、有志の企画となる。なお、情報交換のためのメーリングリストは引き続き運用しているので、興味のある方は世話役:田邊 (ttanabe@iic.hokudai.ac.jp) まで連絡してほしい。

### Contents

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！ .....	清原 文代 ..... 92
Adobe Captivate を用いた教材作成 Flash ムービー教材を手軽に作る方法 .....	鈴木 慎吾 ..... 103

# 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！

清原 文代 (きよはら ふみよ)

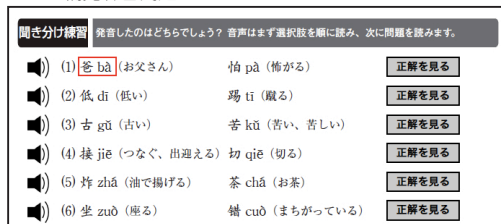
## ※ 1 はじめに

今年度勤務校に新たに CALL 教室が設置され、また Learning Management System<sup>[1]</sup> の全学的な運用が始まった。これらで使用するコンテンツが必要となり、私はそのファイル形式として PDF を選んだ。その大きな理由は 2 つある。

- 1 つのドキュメントを CALL 教室と普通教室の両方で使えるようにする。
- テキストや音声を 1 つのファイルにまとめる。

ご存知のとおり PDF は Windows や Mac OS、Linux といった OS の違いを超えてドキュメントを同じレイアウトで表示できる規格である。それ

図 1 「ネットで get! ピンイン聞き分けトレーニング」の補充練習問題



ゆえ公的な性格を持つ文書の配布によく使用されるが、PDF の用途はそれだけにとどまらない。テキストや図表・画像以外に、音声や動画、更には JavaScript を組み込むこともできる。本稿では筆者の実践例を紹介しつつ、PDF を教材を配布する器として活用する方法について提案したい。

## ※ 2 音声付き PDF を授業で使う

筆者が勤務校で主に担当するのは初修外国語としての中国語初級クラスである。最初の 1 ヶ月余りはピンインを憶えることに費やされるが、この時点で記号と音声の関係をしっかりマスターしておく必要がある。そこで音声付きの PDF を使用した。

### ☞ 2.1 『中国語ジャーナル』の連載記事「ネットで get ! ピンイン聞き分けトレーニング」

筆者は 2007 年度『中国語ジャーナル』（アルク）という中国語学習者向けの CD 付き月刊誌に「ネットで get ! ピンイン聞き分けトレーニング」という連載をしていた。この連載の特徴は誌面と連動して補充の練習問題を無料でネット配信（ポッドキャスト配信）していたことである。このネット配信した練習問題は現在でも以下の Web サイトから入手可能である。

<http://pinyin-training.cocolog-nifty.com/>

補充練習問題は音声付きの PDF で、ごく簡単な JavaScript が組み込んであり、どの選択肢が正解であるかボタンをクリックして表示させることができる<sup>[2]</sup>。

更に今回の授業にあたって、『中国語ジャーナル』の誌面の連載記事についても『中国語ジャーナル』編集部の方了承をとった上で<sup>[3]</sup>、誌面をスキャンして PDF にし、付属 CD の音声 を MP3 に変換して PDF に貼り付けたものを作成した。こちらは雑誌のバックナンバーが商品として販売中ということもあり、公開していない。

## ☞ 2.2 CALL 教室の授業での PDF 使用方法

まず『中国語ジャーナル』の誌面連載記事の PDF の解説部分を学生 2 人につき 1 台設されているセンターモニタと教室前面にあるスクリーンに表示しながら、ピンインの読み方の規則や注意点を音声で再生しつつ説明する。

次に「自主練」と称して正解表示ボタンのついた補充練習問題の PDF を各学生のパソコンに配布し、一定時間ピンインの聞き分け練習を行う。補充練習問題の主な出題形式は音声聞いて選択肢を選ぶものである。

例 音声はまず選択肢を順に読み、最後に問題を読みます。発音したのはどれでしょう？

ǐ(yǐ)    ǔ(wǔ)    ǔ(yǔ)

PDF には音声を再生するためのボタンがあり、何度でも音声を聞くことができる。正解を見るとこのボタンをクリックすれば、正解の選択肢が赤い□の枠で囲まれる。私の能力では学生に答えを選択あるいは入力させて正解がどうかを判定し、点数を計算するような JavaScript を組み込むことができなかつたが、このような単純なものでも概して学生は有用であると感じていたようである(後述)。

「自主練」が終わった後は、再び『中国語ジャーナル』の誌面連載記事の PDF に戻る。こちらに

も解説の後に聞き分けを中心とした練習問題が載っているが、こちらは問題の音声は貼り付けてあるだけで、正解表示ボタンはない。学生のパソコンに転送して各自に問題を解かせ、正解の選択肢だけを紙に書かせる。解答時間が終了したら、別に正解を書いた PDF をセンターモニタとスクリーンに表示して、自己採点させた。

授業終了後は授業中に使用した PDF と、授業には使用しなかつたが関連のある練習問題の PDF を LMS にアップロードしておいた。学生は学内のパソコンから、或いは自宅からインターネット経由で好きな時間に LMS に接続し、PDF をダウンロードして使用することができる。

## ☞ 2.3 履修者の感想

筆者の担当するクラスのうち、履修者が重複しない 2 クラスにおいて簡単なアンケートを実施したところ、計 65 名(理系学部 22 名、人間社会学部 13 名、経済学部 30 名)から回答が得られた。

設問 1: 音声付き PDF は従来の教科書と CD に比べて pinyin の習得に役立ちましたか? 1 つを選んで○をつけてく

役に立ったと思う。	66.2%
どちらかと言えば役に立ったと思う。	27.7%
どちらかと言えば役に立たなかつたと思う。	1.5%
役に立たなかつたと思う。	0%
わからない	4.6%

ださい。

設問 2: 音節付き PDF の良い点、改善すべ

自分のペースでできる。	9 名
何度でも繰り返せる。	9 名
自分の聞きたいところだけを聞ける。	7 名

き点について教えてください。

良い点については 52 名がコメントを寄せ、そ

パソコンのトラブルやエラー	7名
練習問題の音声をもっと細かく区切るか、巻き戻しができようにして欲しい。	3名
音声を最初に再生する際に出るメッセージを出ないようにして欲しい。	2名

のうち特に多かったものは以下のとおりである。

改善すべき点については22名がコメントを寄せ、そのうち多かったものは以下のとおりである。

改善すべき点であるが、トラブルやエラーについては、PDF教材自身の問題ではなく、本年4月からCALL教室の運用を開始したため、CALL教室のシステムが落ち着くまで数ヶ月を要し、また私自身がCALL教室に慣れるまで余裕がなかったためである。

音声の区切りの件については、正解表示ボタンのついた補充練習問題の方は枝間ごとに音声再生ボタンがついているのだが、『中国語ジャーナル』の誌面をスキャンしたPDFについては、付属CDのトラック1つに枝間が数問入っている形になっているのをそのまま貼り付けている。こちらは補充問題で慣れた上で、誌面の練習問題では止めずに一気に数問解いてもらおうということでこうしたのだが、それが不評だったようだ。

音声を再生する際のメッセージというのは、これは音声などマルチメディアを含んだPDFを開けて初めて再生する際に再生してよいかと尋ねるダイアログが表示されることである。キャンセルボタンをクリックしてしまうと、一旦PDFを閉じてもう一度開き直すまで音声を再生することができない。再生ボタンを選ぶように絶えず指示していたのだが、ダイアログの日本語の表現がややこなれていないこともあって、わけがわからないからとりあえずキャンセルしてしまう学生が学期が始まってからしばらくは毎回いた。だが、慣れるにしたがって減っていき、学期後半にはこの手のトラブルはなくなった。

#### ☞ 2.4 ローテクな e-learning 教材でも学生の

#### 役に立つ

上記の結果から学生の9割以上が音声付きPDFに有用性を認めており、その有用性はテキストと音声が入った1つのファイルにまとまっていて、クリックするだけで自分の好きなように音声を聞けるということに起因することがわかる<sup>14)</sup>。筆者が今回使用したPDF教材は要は音の出るテキストに毛が生えたようなもので、採点機能もなければ学習履歴をとることもできない。e-learning教材としては「ローテク」、原始的とも言える。しかし、完全に対面授業をなくしてしまい、e-learning教材だけで学習するのであれば別だが、学生と教員が教室という空間、授業という時間を共有するのであれば、上記の履修者の感想からわかるように、この程度のものでじゅうぶん学生の学習を支援することができるのである。

#### ☞ 2.5 PDFは教員にとって2度おいしい

学生が良い点として挙げた項目については、特にPDFを媒体にしたからというわけではなく、テキストと音声の両方を同時に扱うe-learning教材全般に対して言えることである。しかし、授業に使う教室が全てCALL教室になることは考えづらい。HTML+JavaScriptやFlashで時間と手間をかけて教材を開発しても、黒板と机と椅子しかない所謂普通教室ではその威力を発揮することはできない。HTML+JavaScriptにしるFlashにしる紙に印刷することは基本的には考慮されていないフォーマットであるからだ。それに対してPDFはそもそも紙の文書の再現を意図したものであり、印刷に適している。CALL教室で使う時は音声付きPDFをそのまま学生に配り、普通教室ではPDFを印刷してプリントとして配って音声はCDの形で利用するということが可能である。教員にとっては教材をPDFにしておくことは、言わば一挙両得の旨味があると言えよう。ワープロソフトを使って授業用のプリントを作ることはたいていの教員が日常行っている作業である。そこ

にもう一手間かけてPDFにし、音声を貼り付けることによって、普通教室で使えるプリント教材と、Windows、Mac OS、LinuxといったOSの壁を越えて同じように使えるe-learning教材が同時にできるのだ。

## ✿ 3 PDFの作り方

ご承知のとおり、WindowsであれMac OSであれLinuxであれ、PDFの閲覧・再生・印刷は、無料で配布されているAdobe Readerがあればできる<sup>[5]</sup>。

Mac OS Xの場合はテキストと画像だけといった単純なPDFであればOS同梱のアプリケーションソフトで閲覧・印刷ができる<sup>[6]</sup>。Mac OS Xは更にOSそのものがPDF書き出し機能も備えていて単純なPDFであれば書き出せる<sup>[7]</sup>。

しかし、Windowsの場合はPDFの作成のためには別途ソフトウェアが必要だ<sup>[8]</sup>。PDF作成ソフトの代表的なものはPDFという規格を作ったAdobe社のAcrobatである。PDFの規格自体は公開されており、他社製の安価なPDF作成ソフトもあるが、値段はやや高いけれども機能の豊富さという点ではやはりAcrobatがお勧めである。Mac OS Xの場合もOSの機能だけでは作れないPDFを作るためにはAcrobatが必要だ。幸い教員はアカデミックディスカウントでAcrobatを購入することができる。Acrobatの現時点の最新バージョンは9で、機能の差により、Standard（Windows版のみ）、Pro（Windows版とMac OS X版）とPro Extended（Windows版のみ）の3種がある<sup>[9]</sup>。

以下、Acrobat 9を使ったPDFの作り方を紹介する。

### ☞ 3.1 音声付きPDF

#### ■ 3.1.1 テキスト

##### ● WordからPDF

Windows版のWord・Excel・PowerPointを使用する場合は、AcrobatをインストールするとPDF Makerがインストールされ、Word等の



図 2 Word2007にインストールされたPDF Maker

ツールバーにPDFを作成するためのボタンが追加されるので、それをクリックすればよい。PDF Makerを使ってPDFを書き出すと、Wordの文書内のリンクを保ったままPDFに書き出せる<sup>[10]</sup>。

WindowsでPDF Makerが対応していないソフトを使用する場合や、Mac OS Xの場合は、印刷の際に通常のプリンタの代わりに、Adobe PDFを選んで印刷すれば、PDFが書き出せる<sup>[11]</sup>。その際にPDFの保存場所を尋ねられるので、保存場所を指定する。

##### ● スキャナからPDF

既存の教材をスキャナで読み込むには、Acrobatを起動し、文書メニューから「スキャナからPDFを作成」を選ぶ。「検索可能にする(OCR実行)」にチェックをつけておくと、テキストを画像としてスキャナで読み込んだ後、自動的に文字認識が行われ、テキストとして検索できるようになる。初期設定ではOCRの認識する言語は日本語であるが、オプションボタンをクリックすれば中国語（簡体）や中国語（繁体）も選べる<sup>[12]</sup>。

#### ■ 3.1.2 音声

##### ● CDからMP3

音声はWindows標準の音声ファイル形式であるWAVEか、またはMP3を用いる。筆者はダウンロード時の利便性を考え、PDFの大きさを小さくするためにMP3を使っている。MP3の圧縮率と音質は反比例するが、人の声の主である外国語学習教材では128kbpsくらいで十分である。128kbpsの場合、1分間の音声でおおよそ1MBくらいの大きさになる。

CDをMP3に変換するソフトウェアはいろいろあるが、今回『中国語ジャーナル』のCDをMP3に変換するにあたってはiTunes（Windows版・Mac OS X版、フリーウェア）<sup>[13]</sup>を使用した。

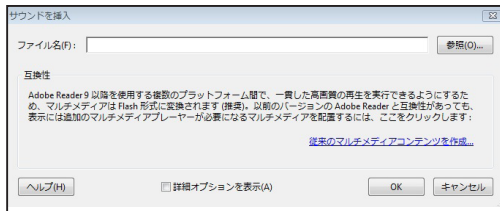


図3 音声を貼る範囲を指定すると現れるサウンドツールのウィンドウ

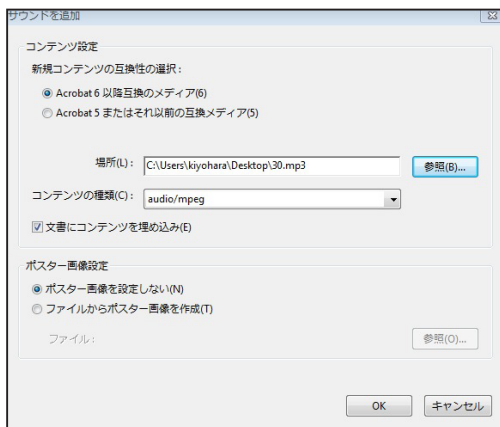
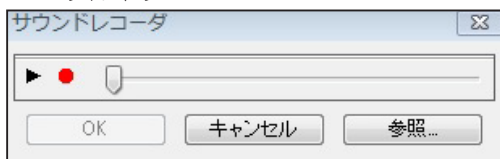


図4 以前のバージョン用のサウンドツールのウィンドウ

- (1) iTunes は初期設定では音声を AAC に変換するので、「編集メニュー」→「設定」→「一般」→「インポート設定」で「インポート方法」を MP3 エンコーダに変更する<sup>[14]</sup>。
- (2) iTunes に CD を読み込む。通常は CD を iTunes ライブラリに読み込みますかと尋ねてくるので OK する。もしくは詳細メニューの「MP3 バージョンを作成」を用いる。
- (3) 変換が終わったら CD を取り出し、iTunes のウィンドウの中にできた当該のトラックをデスクトップにドラッグすれば、MP3

図5 音声注釈をつける場所を指定すると現れる録音用のウィンドウ



ファイルがデスクトップにコピーされる。

### ●PDF にサウンドツールで音声を貼る

- (1) Acrobat で音声を貼り付けたい PDF を開き、「ツールメニュー」から「マルチメディア」→「サウンドツール」を選ぶ。
- (2) 十字カーソルが出てくるので、音声を貼り付けたい箇所をドラッグして指定する。
- (3) 初期設定では Acrobat9 及び Adobe Reader 9 以上に対応する形式での音声を貼り付ける設定になっている。相手が Adobe Reader 9 を使っているのであれば、選択ボタンを押して MP3 ファイルを選ぶ。

もし相手が以前バージョンの Acrobat や Adobe Reader を使っている場合は、「従来のマルチメディアメディアコンテンツを作成」<sup>[15]</sup>をクリックする。「Acrobat6 以降互換メディア」を選び、選択ボタンをクリックして MP3 ファイルを選択する。「文書にコンテンツを埋め込み」が ON になっていることを確認したのち、OK をクリックする。

既存の音声を貼り付けるのではなく、音声の録音や編集を専門とするソフトウェアを別途使用して MP3 形式で録音し、上記のサウンドツールを使って PDF に貼り付けることもできる<sup>[16]</sup>。

録音・音声編集ソフトとしては、例えば Audacity (Windows・Mac OS X・Linux/Unix 用、フリーウェア)<sup>[17]</sup> や Sound it! (Windows・Mac OS X 用、有料パッケージソフト) といったソフトウェアがある<sup>[18]</sup>。

### ●PDF に直接録音する

サウンドツールとは別に Acrobat 自身が簡単な録音機能を持っている。

- (1) 「注釈メニュー」→「注釈ツール」→「音声ファイルを注釈として添付」を選ぶ。
- (2) スピーカーのアイコンがついたカーソルが出てくるので、音声を貼り付けたい箇所をクリックして指定する。



- (3) 録音のための小さなウィンドウが開くので、録音ボタンをクリックして録音する。この録音は上書き録音なので、失敗したら再録音すればよい。
- (4) 録音が終わったら、OK をクリックする。

### ☞ 3.2 動画付き PDF

外国語学習教材に不可欠なマルチメディアと言えば音声であるが、「百聞不如一見」、動画を教材に使いたいことがある。しかし、様々なパソコン環境で動画の再生を保証することは案外難しいことで、音声は WAVE か MP3 にしておけばまず間違いなく再生できるが、様々なコーデックが存在しファイルフォーマットとコーデックの関係が一律でない動画ではそうはいかない。この問題の解決方法の 1 つが Acrobat 9 の 3 つのエディションのうち、Acrobat 9 Pro Extended を使って PDF に動画を貼り込むことである。Acrobat 9 Pro Extended は様々な動画ファイルを Flash (FLV 形式) に変換して PDF に埋め込むことができるのだ。再生には Acrobat 9 ないしは Adobe Reader 9 が必要になるが、とにかく相手に最新版の Adobe Reader を入れてくれというだけで済むので、制作者側の負担が軽くなる。また、埋め込んだ動画の中でキーとなる画面をポスターフレームとして指定しておけば、印刷したときにも静止画ではあるがそこだけは見せることができる。

- (1) Word または PowerPoint で教材ファイルを作る。
- (2) Word または PowerPoint の「Acrobat メニュー」から「ビデオを埋め込む」を選ぶ。
- (3) 参照ボタンをクリックして動画ファイルを選ぶ。
- (4) 動画の下にあるスライダーバーを動かしてポスターフレームにしたい場面を表示する。
- (5) 「現在のフレームからポスター画像を設定」をクリックする。
- (6) OK をクリックする。

### ☞ 3.3 Web ページへのリンク付き PDF

#### ■ 3.3.1 使用例

今回初級中国語の授業で使用した PDF には音声を貼り付けただけだったが、PDF には Web ページへのリンクを含めることができる。例えば両替に関する課文では中国人民銀行の Web サイトの URL を書いておき、外貨のレートを調べさせるといった活動ができる<sup>[19]</sup>。また課文に料理が出てくれば Google イメージ検索や百度图片で料理の写真を探したり、YouTube でその料理の作り方の動画を探したりして、そこへリンクを貼っておくということができる。画像や動画そのものではなくリンクなので著作権上の問題を避けられ、ファイルが重くなることもない。

#### ■ 3.3.2 リンクの貼り方

Web ページの URL を含んだ文書を PDF に変換したのち、「アドバンスドメニュー」から「文書処理」→「URL からリンクを作成」を選ぶと、自動的に文書中の URL を認識し、Web ページへのリンクが貼られる。

上記の方法でうまくリンクが貼れなかった時は、「ツールメニュー」→「高度な編集」→「リンクツール」を使って、リンクを貼りたい箇所を選択し、リンク先として「Web ページを開く」を選び、URL を記入する。

なお、PDF に貼られた Web ページへのリンクを初めてクリックすると、Acrobat や Adobe Reader がセキュリティ警告を表示する。許可をクリックしないと Web ページには接続できないので、その旨学生に指示しておかねばならない。

### ☞ 3.4 JavaScript を PDF に

「ネットで get ! ピンイン聞き分けトレーニング」の補充練習問題 PDF で選択問題の正解を表示している仕組みは単純で、赤い枠で中身が透明のボタンを正解の上に重ねて置いておき、最初は見えない状態にしておく。それを「正解を見る」と書かれたボタンに組み込んだ JavaScript によって見えるようにしているだけである。正解表示ボ

タンは、正解の赤枠が表示されていなければ表示し、赤枠が表示されていればそれを消すという動作をする。単に選択肢を選ぶだけでなく、選択肢の上に正解を上書きするような問題もあるが、それは赤枠のボタンに正解を書いた画像を貼り込んであるだけである。

- (1) 「ツールメニュー」→「高度な編集」→「ボタンツール」を選び、正解の選択肢の上にボタンを作る。
- (2) フィールド名にボタンの名前を入力し、メモしておく。

(例) seikail

- (3) ボタンツールが選ばれた状態で、ボタンをダブルクリックする。
- (4) 「一般」タブで「表示と印刷」を「隠す」にする。
- (5) 「表示方法」タブで「境界線の色」を「赤」に、「塗りつぶしの色」を「カラー無し」にする<sup>[20]</sup>。これで正解を示す赤枠ができあがる。
- (6) 正解を表示させるためのボタンを作り、ボタンをダブルクリックする。
- (7) 「アクション」タブの「アクションを選択」で「JavaScript を実行」を選ぶ。
- (8) 「追加」をクリックし、以下の JavaScript を入力する。"seikail" の部分には先ほど作った赤枠のボタンの名前を入れる。

```
var f = getField("seikail");
if(f.display == display.
visible)
f.display = display.hidden;
else f.display = display.
visible
```

- (9) 「オプション」タブ→「レイアウト」をラベルのみにし、「ラベル」のところに「正解を見る」と入力する。
- (10) 以下、赤枠のボタンの名前を変えながら、問題の数だけ同じ作業を繰り返す。
- (11) 赤枠を隠した状態で PDF を保存する。

Acrobat を使って PDF に組み込む JavaScript については Adobe 社の以下の Web ページに詳細な情報がある。

<http://www.adobe.com/devnet/acrobat/javascript.html>

JavaScript を使いこなせる方であれば、もっとスマートに複雑な機能を PDF に組み込むことができるであろう。

### 3.5 採点機能付き小テストを PDF で

筆者が上記の JavaScript 入りの PDF を作ったときには Acrobat のバージョンは 8 であったが、Acrobat 9 になって、スクリプトを一切書くことなく、問題文と答えを用意するだけで採点機能付きの小テストを作成し、PDF に書き出すことができるようになった。この機能は Acrobat 9 の 3 つのエディションのうち、Acrobat 9 Pro Extended に同梱されている Adobe Presenter を使用する。Presenter で作った PDF を使用するには Acrobat 9 ないしは Adobe Reader9 が必要である。

小テストの形式は以下の 5 種類である。

- 複数選択（選択肢の中から 1 つないしは複数の正解を選ぶ）
- 正誤（選択肢は 2 つのみ、どちらが正しいかを選ぶ）
- 穴埋め（文章の空欄に適語を入力、ないしはあらかじめ用意された選択肢から適語を選ぶ）
- 短文記述（答えを自由に入力する）
- 照合（二つのリストの各項目の間を結びつける）

これらに加えてアンケート用のリッカート（評価尺度）がある。

まずの用語が若干わかりにくいので説明しておく。「クイズ」と「質問」という似た用語があるが、両者の関係はクイズが上位、質問が下位である。質問(問題)が集まったもの、それがクイズ(小テスト)である。1 つのファイルに複数のクイズ

(小テスト) を含むこともできる。

- (1) Presenter は PowerPoint に組み込まれて動作するので、PowerPoint で教材ファイルを作る。
- (2) PowerPoint の「Adobe Presenter メニュー」から「新規クイズを追加」を選ぶ。
- (3) クイズの設定を決める画面が現れるので、初期設定のものでよければそのまま OK をクリックし、必要であれば適宜設定を変更して OK をクリックする [21]。

この設定は「Adobe Presenter メニュー」の「管理」ボタンをクリックしてクイズマネージャーを起動し、「クイズ」タブでクイズの「編集」を選ぶことにより後から変更できる。

- (4) 「質問を追加」をクリックして、個々の問題を作っていく。
- (5) 質問を作り終わったら、OK をクリックする。後から質問をつけ足す場合は、「管理」ボタンをクリックしてクイズマネージャーを起動し、「クイズ」タブの「質問を追加」をクリックする。
- (6) Presenter は LMS への出力に対応している。初期設定ではそれが ON になっているので、必要がなければ、「管理」ボタンをクリックしてクイズマネージャーを起動し、「レポート」タブで「プレゼンテーションのレポートを有効にする」のチェックをはずしておく。
- (7) 作業をすべて終わったら、「Adobe Presenter メニュー」の「パブリッシュ」をクリックし、「Adobe PDF」タブを選ぶ [22]。
- (8) 「選択」ボタンをクリックし、PDF を保存する場所と PDF の名前を決める。
- (9) 「パブリッシュ」ボタンをクリックして PDF を書き出す。

Presenter はテキストだけでなく、音声や動画を使用することもできる。「Adobe Presenter メニュー」には既存の音声ファイルや動画ファイル



図 6 PowerPoint2007 にインストールされた Adobe Presenter

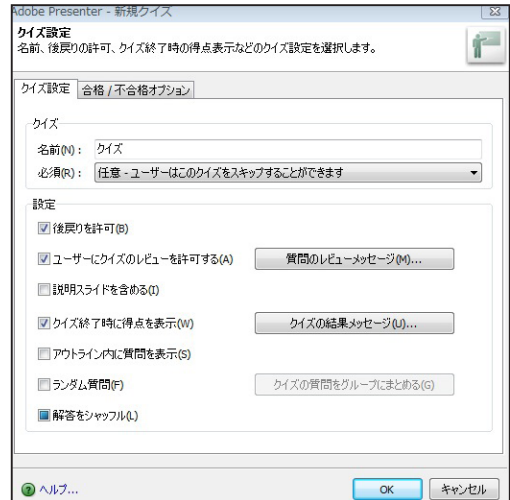


図 7 クイズの設定

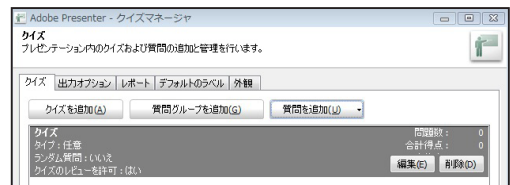
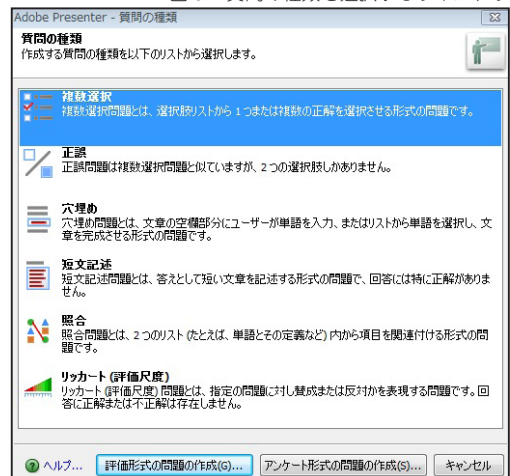


図 8 クイズに質問を追加するためのウィンドウ

図 9 質問の種類を選択するウィンドウ



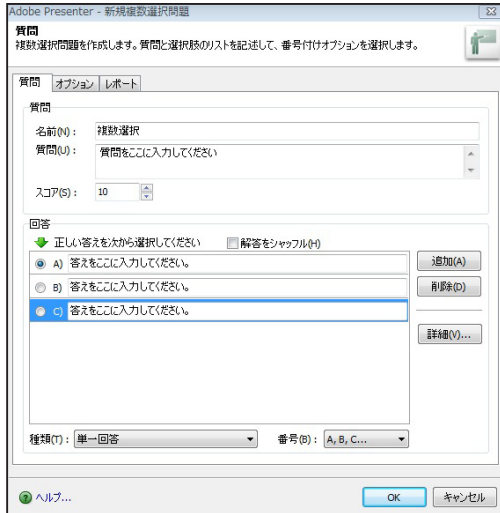


図 10 選択問題を作るためのウィンドウ、質問に問題文を入力し、「追加」ボタンをクリックして選択肢を入力する。

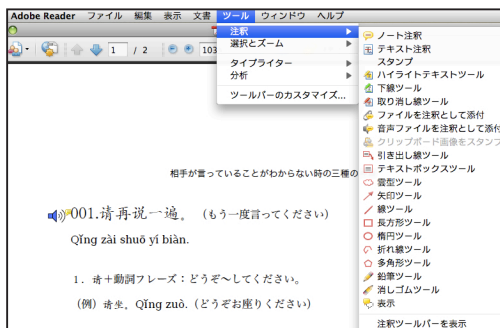
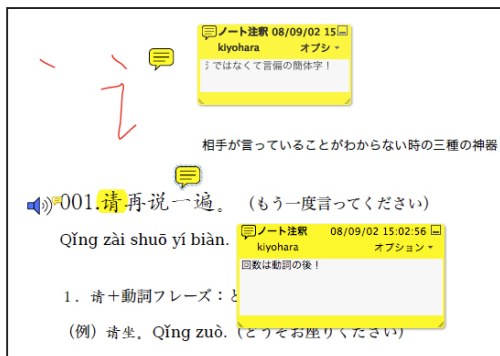


図 11 Adobe Reader でも注釈ツールが使えるようになった PDF

図 12 ノート注釈・ハイライトテキストツール・鉛筆ツール・音声注釈を使った PDF の例、ノート注釈は邪魔であれば非表示にできる。



を読み込む機能、音声を録音する機能、パソコンに繋いだビデオカメラから動画をキャプチャする機能がある。録音した音声やビデオカメラから取り込んだ動画に対して簡単な編集を行うこともできる。したがって、音声や動画を見ながら聞き取った内容を、短文記述問題の解答として入力させたり<sup>[23]</sup>、穴埋め問題として聞き取った内容の一部を入力させたり、音声や動画の内容にふさわしい選択肢を選ばせたりすることも可能だ。

### 3.6 Adobe Reader で PDF にメモや音声をつける

紙の教材の良いところは余白に学生がメモを取れることである。PDF は紙の再現を目指しただけあって、PDF に注釈という形でメモやコメントをつけることができる。ただ学生が使用するパソコンに Acrobat がインストールされていることは少なく、たいてい Adobe Reader がインストールされているだけであろう。Adobe Reader では基本的には注釈をつけることはできないが、Acrobat 9 Pro または Acrobat 9 Pro Extended を使用すれば、Adobe Reader でも注釈をつけてそれを保存することができる PDF を作成できる。

操作は簡単で、Acrobat の「注釈メニュー」で「Adobe Reader で注釈を有効にする」を選んで、元の PDF のファイルとは別の名前でも保存するだけである（Adobe Reader で注釈を有効にするのは不可逆的な変更なので、オリジナルの PDF を上書きせずに別に保存した方がよい）。

このようにして保存した PDF は、Adobe Reader で開けても、Acrobat と同様にツールメニューの中に注釈メニューが現れるようになる。

学生はこの注釈メニューを利用して、ノート注釈機能で PDF 教材にメモを付け加えたり、重要なところをハイライトテキストツールで黄色にマーキングしたりして保存することができるようになる。鉛筆ツールをつかえば PDF 上に自由に線を引くこともできる。更に Adobe Reader で音声注釈をつけられるようになるので、あらかじめモデルになる音声を貼り付けておき、学生には音声注釈で自分の声を録音させ、それぞれ再生し

て聞き比べるといった練習も可能である。

## ✿ 4 まとめ

文書の配布手段として知られる PDF であるが、そこにはテキストや画像以外に、音声や動画といったマルチメディア、更には採点機能付きの小テストを入れることもできる。その一方印刷にも適している。ワープロソフトやプレゼンソフトで教材を作ったら、PDF にすることによって e-learning 教材への道が開ける。逆に e-learning 教材を PDF で作れば、普通教室でのプリント配布にも対応できる。いろいろなものが入る便利で大きな器、それが PDF である。

## 注

- [1] 筆者の勤務校では NEC の i-Collabo.LMS が導入されている。  
[http://www.i-collabo.jp/lms\\_1.html](http://www.i-collabo.jp/lms_1.html)
- [2] 毎回の練習問題の内容及び JavaScript の原型は私が書いたが、連載に当たって PDF の作成は『中国語ジャーナル』編集部の服部浩之氏にお願いした。ここに記して御礼申し上げる。
- [3] CALL 教室や LMS 用のデジタルコンテンツを作る際に問題になるのが、権利処理の問題である。既存の教科書をデジタルコンテンツ化するためには権利者の許諾が必須である。今回は筆者が著者である連載以外に、一部既存の教科書を使ったコンテンツも作成したが、履修者には必ず当該の教科書を買わせ、デジタル化したものは授業中に配布するか、授業を履修している学生のみがアクセスできる LMS で配布する、教科書の全てをデジタル化しない等といった条件で許可を得ることができた。もちろん著者や出版社によって対応は様々であるが、最初から無理だとあきらめるのではなく、まずは権利者に連絡をとってみる価値はある。
- [4] 設問 2 の記述の中には CD だとトラック番号を合わせなければならないのに比べて PDF は便利である、目で見てすぐ聞きたい音声を選べるといったようなコメントがあった。現在の中国語教科書の多くは CD 付きであるが、CD の封を切らないままの学生を見かける

ことは決して珍しくない。昨今の学生にとっては教科書を見ながら CD プレーヤーを操作してトラックを選び音声を聞くということすら面倒なことになりつつあるのかもしれない。

- [5] Adobe Reader  
<http://www.adobe.com/jp/products/acrobat/readstep2.html>
- [6] Mac OS X では OS に同梱されているプレビューというアプリケーションソフトで PDF を表示することができるが、PDF に組み込んだ音声や JavaScript を再生するためには Adobe Reader が必要である。  
Mac OS X は初期設定ではプレビューで PDF を開くように設定されているので、常に Adobe Reader で開くようにするためには以下のようにして設定を変更する。
  - (1) PDF ファイルのアイコンをクリックして cmd キー + i を押す。
  - (2) 「このアプリケーションで開く」のところを Adobe Reader に変更する。
  - (3) 「類似したすべての書類を開くときにこのアプリケーションを使用します。」の下にある「すべてを変更」をクリックする。
- [7] Mac OS X は OS 自身が PDF の書き出し機能を持っているが、PDF に音声や動画、JavaScript を組み込むためには Acrobat が必要である。
- [8] Windows で Word2007 を使用している場合は、マイクロソフト社の Web サイトから無償のアドインをダウンロードしてインストールすることにより、Word から PDF の書き出しが可能になるが、この場合も PDF に音声や動画、JavaScript を組み込むためには Acrobat が必要である。  
<http://office.microsoft.com/ja-jp/help/HA101675271041.aspx>
- [9] Acrobat9 の各エディションの機能比較  
<http://www.adobe.com/jp/products/acrobat/matrix.html>
- [10] Word 固有の機能をどう変換するか、PDF のセキュリティ設定をどうするかは Word に追加された Acrobat メニューの環境設定を使って設定する。
- [11] Mac OS X は、印刷に際に OS 自身が持っている PDF 書き出し機能を使っても、プリンタドライバとして

Adobe PDF を指定しても PDF を書き出すことができるが、Adobe PDF を使うとフォントの埋め込みやできあがりの品質などについて設定された複数のプリセットから PDF の書き出し設定を選ぶことができる。また Acrobat に同梱されている Acrobat Distiller を使用すれば、オリジナルの PDF 書き出し設定を保存することもできる。

[12] この OCR 機能は既存の PDF (テキストを画像として取り込んでいるもの) やスキャンした画像ファイル (TIFF・JPEG) にも有効である。当該のファイルを Acrobat に読み込み、「文書メニュー」→「OCR テキスト認識」→「OCR を使用してテキストを認識」を選ぶ。その際、編集ボタンをクリックすれば OCR の言語を切り替えることができる。

[13] <http://www.apple.com/jp/itunes/download/>

[14] Mac 版では「iTunes メニュー」→「環境設定」→「一般」→「読み込み設定」→「読み込み方法」

[15] Acrobat9 及び Adobe Reader9 以上に対応する形式で音声を貼り付けると再生コントロールボタンは自動的に表示されるが、従来の形式で音声を貼り付けた場合は設定を変更しないと再生コントロールボタンは表示されない。

(1)「サウンドツール」が選ばれた状態で、貼り付けた音声をダブルクリックする。

	音声による注釈	サウンドツール
録音機能	○	×
MP3	×	○
再生コントロールボタンの表示	×	○

(2)「設定タブ」で「レンディションを編集」を選ぶ。

(3)「再生設定タブ」で「プレーヤーのコントロールを表示」にチェックをつける。

(4)「再生場所タブ」で再生場所を「文書内」に変更する。

[16] 音声による注釈とサウンドツールの主な機能の違いは以下のとおりである。

[17] <http://audacity.sourceforge.net/>

mp3 で録音するためには別途 LAME MP3 encoder が必要であるが、これもフリーウェアとして配布されており、上記の Web サイトからリンクが貼ってある。

なお、Audacity は日本語インターフェースを持っているが、不具合があるらしく、英語インターフェースま

ま使用した方がよいようだ。

[18] <http://www.ssw.co.jp/products/sit/index.html>

[19] 2007 年度にポッドキャストを使って清原文代・顧春芳「大阪府立大学中国語ポッドキャスト旅行会誌編」をインターネット配信したが、そこで配信した音声付き PDF には課文に関連する Web サイトのリンクも含めている。

<http://www.las.osakafu-u.ac.jp/podcast-lang/zh/travel/index.html>

[20] 赤枠で囲うのではなく、正解を選択肢の上に上書きする場合は、「オプションタブ」で「レイアウト」をアイコンのみにし、「アイコンの選択」であらかじめ画像として作っておいた正解を選択する。

[21] この設定画面には多くの設定項目があるが、特に変更する可能性が高い主な設定は以下のとおりである。

「クイズ設定」の「後戻りを許可する」——前の問題に後戻りして解答することを許可するかどうか。

「クイズ設定」の「ユーザーにクイズのレビューを許可する」——レビューを許可した場合、採点後に「クイズのレビュー」ボタンをクリックすると、自分の答えが表示され、不正解の問題では自分の答えとともに正解が表示される。

「合格／不合格オプション」——合格点を決める。

「許可する試行回数」——不合格の場合に何回までやりなおしを認めるか。

[22] Presenter は PDF 以外にも Web ブラウザで再生できる Flash 形式で小テストを書き出すこともできる (「パブリッシュ」で「マイコンピュータ」タブを選ぶ)。但し、PDF で書き出すとできるファイルは 1 つであるが、Flash 形式では複数のファイルをフォルダに入れた形になる。Web サイトにアップロードするのであればそれでもかまわないが、学生に直接ファイルを渡す場合は、ファイルが 1 つで済む PDF の方が便利でトラブルも少ないであろう。

[23] 聞き取った文を入力させる設問では、「姥爷」と「老爷」のような漢字表記そのものの揺れや、「,」を使うかどうか、「,」は全角か半角かといった表記の揺れが採点の際に問題になるが、短文記述形式の問題でも正解を複数指定することができるので、この問題はある程度回避できる。

# Adobe Captivate を用いた 教材作成

## Flash ムービー教材を手軽に作る方法

鈴木 慎吾 (すずき しんご)

### ✿ 1 Flash ムービーで教材を作りたい!

Flash ムービーといえば一般に、ウェブサイト上でグリグリ動く、派手なアニメーションを再生するフォーマットとして知られている。これは、単に決められた動画を見せるだけでなく、インタラクティブな仕掛けを組み込むこともできるため、工夫次第で教材への応用も可能なものである。この Flash ムービーは通常 Adobe Flash<sup>[1]</sup> を使って作成する。ならば、いざ作成! といきたいところなのだが、Adobe Flash は色々なことができる反面、操作にはある程度の習熟が必要な上、インタラクティブなコンテンツを作るにはプログラミングの技術も必要であり、その点でちょっと敷居が高い。そのため、我々が Flash 教材を作ろうとする場合は Flash に関してある程度のスキルを持った人を捕まえて作ってもらうか、あるいは業者に頼むことになるケースが多い。そのことが、さらに敷居を高くしている。

Flash ムービーでちょっとした教材を作りたい、

という場合、Adobe Captivate (以下、Captivate) で作るという選択がある。Captivate を使うと、プログラミングの特別な知識がなくともソフトウェアのデモや、トレーニング用のシミュレーション教材、小テストなどを簡単に作成することができる。

### ✿ 2 Captivate で教材作成

昨今は中国語教師も情報系の授業を受け持つことが多々あるが、この場合往々にしてパソコンの操作そのものが教授内容に入ってくる。Captivate で作る教材の第一は、このパソコン操作に関する教材である。Captivate はその名前から連想されるように、パソコン画面のキャプチャー機能を中心とするソフトウェアであり、パソコン操作についての教材作成はその最も得意とするところである。

もちろん中国語教師としては本職として中国語を教える教材も作りたいわけで、これが本稿で紹介する教材の第二である。Captivate には小テスト作成機能があるので、これを活用して教材を作成することができる。

#### ☞ 2.1 Captivate の概要

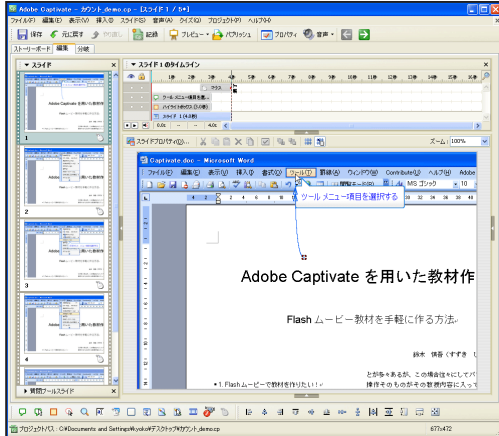


図1 Captivateの編集画面(右上にタイムライン表示)

メインとなる編集画面の構成は PowerPoint に似ているが、一枚一枚のスライドにタイムラインが存在しており、より込み込んだスライドを作ることができる(図1)。また、あらかじめ複数のシナリオを用意しておいて、再生中に何らかの条件によってスライドを分岐させることができるのも大きな特徴である(図2)。また、スライドに小テストを組み込み、評価を表示する機能も備える。作成したスライドは Flash ムービーファイル(.swf)で出力することができる<sup>[4]</sup>。

## 2.2 Captivate で作るさまざまな教材

### 2.2.1 ソフトウェアのデモ

Captivate を使うと、パソコンの操作をアニメーションとして記録することができる。これは、ソフトウェアの操作方法を教えるための教材を作る際に大いに役に立つ。もし画像を使って説明するとすれば何枚もの画像を使って解説しなければいけないところ、アニメーションを使えば極めて簡潔かつ直感的な説明をすることができる。

Captivate の操作記録機能は、教材作成を強く意識した作りになっており、単に操作を記録するだけでなく、操作に関する解説文をキャプションの形式で自動的に挿入してくれる機能がある(図3)<sup>[5]</sup>。この機能のおかげで、ちょっとした操作説明デモであればあっという間にできあがってしまう。我々の業界であれば、中文入力やメール作成などの操作デモを細かくモジュール化して作っておけば適宜授業に活かすことができる<sup>[6]</sup>。

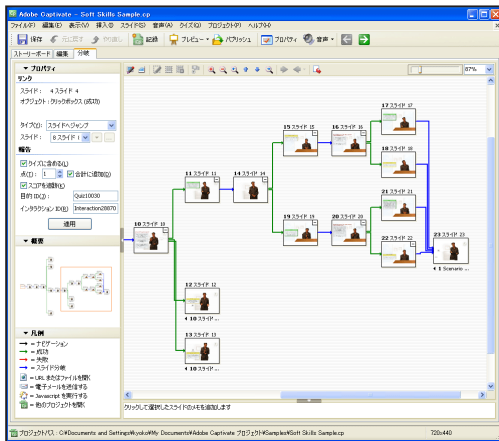


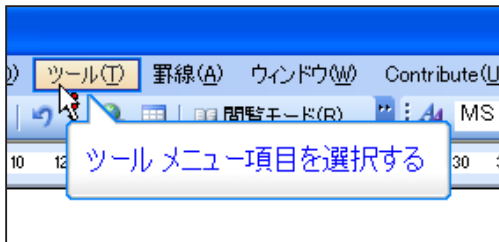
図2 Captivateの分岐シナリオ表示画面

### 2.2.2 ソフトウェアのシミュレーション

Captivate は、ソフトウェアの操作をデモの形式で単に見せるだけでなく、学習者にソフトウェアの操作を実際に体験させることによって学習を行わせるシミュレーション教材を作成することもできる。このタイプの教材は、実際の体験を通じて単なる知識を活きた技能として定着させる効果が期待できる。

Captivate で作るシミュレーションにはトレーニングシミュレーションと評価用シミュレーション

図3 自動挿入されたキャプション



現在、Captivate の最新バージョンは3で、Windows版のみが出ている。操作は割合シンプルで、いわゆるレイヤーの概念<sup>[2]</sup>さえ理解していればおおよその操作はできると思われる<sup>[3]</sup>。



ンの二つのモードがある。前者は学習者をヒントで導きながら操作を学習させるモードであり、後者は学習者にヒントを出さずに操作させ、操作を間違えたときにはキャプションを出して誘導し、最後に学習に関する評価を表示するモードである。

シミュレーション教材はデモ教材と違って学習者に対して積極的な活動を促しつつ進行するタイプの教材であるので、ソフトウェア操作の特に重要な部分を教える教材として使用すれば効果的だろう。

### ■ 2.2.3 小テスト

操作画面の記録（キャプチャー）とは直接的な関係がないようにも思えるのだが、Captivate は教材作成用ソフトウェアとして進化してきたため、小テストを作成する機能も備えている。この機能はソフトウェア操作に関する教材に限らず、例えば我々であれば中国語教育の場面で広く応用することができるものである。

Captivate で作成できる小テストのタイプは以下の通りである。

- 多肢選択：いわゆる「選択問題」。
- ○ × 形式：「多肢選択」と似ているが、こちらは二択式。
- 空欄埋め：文中の空欄を答えさせる問題。選択肢を与えるか、自由に記述させるかを選ぶことができる。
- 短い答え：自由記述問題。「短い」答えに限定されているのは、マッチングによって正解判定を行う必要があるため。
- 照合：二つのリストの項目を関連づけさせる問題。よくある「正しい組み合わせとなるよう線で結びなさい」というタイプ。
- ホットスポット：画像を示して、そのうち質問に合致する場所を答えさせる問題。
- 順序：いわゆる「並べ替え問題」。
- 評価尺度（リッカート）：質問に対してどの程度同意できるかを五段階評価などの形式で答えさせる問題。アンケートに用いる。



図4 「ホットスポット」を使ったトレーニング例<sup>[7]</sup>

Captivate には「質問プール」という仕組みがあって、ここに問題をたくさん用意しておいて、そこからランダムに出題するようにすることもできる。

さて、Captivate のこの「小テスト」機能を使って実際にどんな教材を作るかは、作成者のアイデアとセンスによるところが大きい。ただ、単なる確認用の小テストといったようなものであれば割合アイデアは思いつきそうだから、ここでは特に記憶力や反応力の強化といったトレーニングに目的をおいた初級教材のアイデアをいくつか紹介しておく。

#### ● 2.2.3.1 「ホットスポット」を使って…

中でも特に応用範囲が広そうなのがこの「ホットスポット」である。例えば単語の音声を聞かせて、対応するイラストをマウスクリックで選ばせる問題は、反復して出題する事で学習者の反応を高める効果が期待できる。制限時間を設けることもできるので、3～5秒ぐらいに設定してやるとテンポよくトレーニングできる。質問は名詞だけでなく、動詞、形容詞、方位詞などのほか、フレーズや文など、色々工夫が可能である（図4）。

#### ● 2.2.3.2 「順序」を使って…

単語の音声を3つか4つほど連続して聞かせ、聞いたとおりに単語を並べ替える練習。ただし、普通に答えさせるだけではあまり面白くないので、

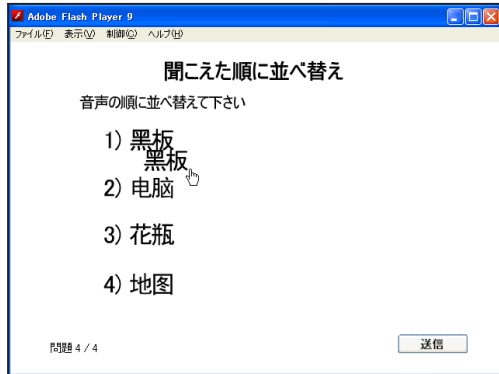


図5 「順序」を使ったトレーニング例（アイテムはドラッグアンドドロップで並べ替えることができる）

音声を聞かせるときには画面には何も表示しないで、3秒ほど経ってからはじめて単語のリストを表示するようにすると、学習者に一時的な記憶を求めることになるのでトレーニング効果が高まる（図5）。

### ●2.2.3.3「多肢選択」を使って…

2.2.3.2の方法が難しければ、聞かせる単語を一つだけとし、「多肢選択」を使って選択肢から選ばせるだけでもよいだろう。ただし、この場合も選択肢を「後出し」する点が重要である。

## ✿ 3 おわりに

Flashムービーファイルを再生するには通常Flash Playerが使われるが、このFlash Playerにはさまざまなプラットフォームで動作するバージョンがあって、それらは無償で配布されているため、結果的にFlashムービーファイルの汎用性は極めて高くなっている。このことは我々が

Flashを用いる上での大きな利点である。

Captivate 3の販売価格は90,300円（税込）と高額だが、アカデミック版があって、教育機関所属者ならば実質3万円台で購入できる。興味のある方は30日限定の無償体験版があるので、まずはそちらで遊んでいただければよいと思う<sup>[8]</sup>。

この手のソフトウェアは、はまる方はきっとはまると思うので、ぜひお薦めしたい。

### 注

- [1] 以前はMacromedia Flashという名称だったが、MacromediaがAdobeに買収されたために今ではAdobe Flashに名称が変更されている。Adobe Captivateも同様。
- [2] 画像などのパーツを透明なシートの上に配置してそれらを重ねていくような形で全体を構成する方法。グラフィック系ソフトで広く取り入れられている。ここでのパーツは画像に限らず、テキスト、音声、動画など、さまざまなオブジェクトがあり得る。
- [3] CaptivateにはCaptivateそのもので作成されたチュートリアルムービーが用意されており、これによって操作方法を学習することもできる。
- [4] このほか、Windows上で単体で再生できる.exeファイルとして出力したり、Word形式で出力したりすることもできる。
- [5] 自動挿入されるキャプションは日本語以外の言語も選べるので、色々遊ぶこともできる。
- [6] SHARPのMebiusのサポートページにはCaptivateで作成された操作デモムービーが大量に公開されており、参考になる。[http://support.sharp.co.jp/mebius/menu/m\\_faq/index\\_top.shtml](http://support.sharp.co.jp/mebius/menu/m_faq/index_top.shtml)
- [7] 作例中のイラストは国際交流基金日本語国際センター作成の「みんなの教材サイト」より拝借した。<http://momiji.jpf.go.jp/kyozai/index.php>
- [8] <http://www.adobe.com/jp/products/captivate/>

# ソフトウェア レビュー

2008年度のソフトウェア・レビューをお届けする。

昨年は、Windows VISTA と Office2007 のリリースという大きな話題があったが、本年はパソコン用ソフトウェアに関する大きな話題は無かった。むしろ、iPhone や ASUS の EeePC といった小型ノートパソコンなどのハードウェア方面の話題が豊富な年であったといえるだろう。

そこで本号では、少々毛色を変え、各 OS にインストールされているフォントの整理や、iPhone や iPod touch といった今時のデバイスについて採り上げてみた。今日、多言語を前提とした情報環境（インターネットやアプリケーションソフト）の整備は、本誌読者としては至極当たり前の作業となっているだろう。その一助あるいは確認のための資料として、本誌の記事がお役に立てれば幸いである。

また、本年も本誌読者向けのアプリケーションを中心に紹介した。導入を考えておられる読者諸賢には、各レビューの記事が参考になれば幸いである。

## Contents

アプリケーションソフト	iPhone で中国語.....	内田 慶市	108
	iPhone & iPod touch 2.0 Software .....	秋山陽一郎	111
	WWW ブラウザ .....	上地 宏一	120
	Office 2008 for Mac.....	師 茂樹	124
	フォント .....	上地 宏一	127
	楽々中国語 V5 .....	金子 真也	132
	Google ピンイン IME .....	千田 大介	134
	一太郎 2008 .....	山田 崇仁	135

# ❖ アプリケーションソフト

## iPhone で中国語

内田 慶市

### ❖ はじめに

昨年アメリカで発売された iPhone だが、この 7 月によろやく日本でも発売された。昨年すでに初代の iPhone を入手し、SIM ロックを解除して海外での出張の際には現地の SIM を入れて使用していたが、あいにく日本では電話機能は使えずにいた。今回晴れて日本でも正々堂々と使用でき、しかも 3G 対応である。

これまで W-ZERO3 や X01HT をはじめ色々なスマートフォンを使ってきたが、iPhone はやはりひと味もふた味もそれらとは違っている。洗練されたデザインだけでなく、タッチパネルに代表されるスマートな操作方法、そして何よりも標準で多言語に対応しているのが嬉しい限りである。さすが Apple ブランドである。Windows 系列には決して真似のできぬものである。

すでにご存知だと思うが、先ず iPhone でできる一通りの機能について簡単に触れておく。

#### ①電話機能

アドレス一覧や履歴、キーパッド、留守電が視覚的に秘められる

#### ②メール機能

マルチアカウントが使える機能的なメーラーであり、iSoftbank メールや MobileMe 以外に Gmail、Yahoo メール、その他普通のパソコンメールが扱える。と言うよりむしろ、パソコンのメール機能そのものである。当然、HTML 形式も送受信可能であるし、添付ファイルも PDF、Word、Excel、PowerPoint などビジネスアプリケーションのファイルが一通り閲覧も可能である。

#### ③WEB 閲覧

Safari を搭載しており、パソコンのブラウザと同じ

ように閲覧できる。文字が小さい場合には、マルチタッチによって拡大／縮小が可能である。また、液晶画面を縦横に向きを変えることで画面そのものが回転する（この機能は他のアプリケーションでも同じ）。

#### ④iPod 機能

音楽、ビデオ、PodCast 等々、iPod そのものである。

#### ⑤Wi-Fi 機能

無線 LAN の環境下では、Wi-Fi を使ってネットにアクセスすることができる。iTunes Store から音楽を購入したりする場合には、3G ではなくて、Wi-Fi を利用することになる。

#### ⑥カレンダー機能

iCal と同期させて、スケジュールを管理する。

#### ⑦カメラ機能

200 万画素ではあるが、カメラ機能が付属しており、iPhoto と同期させる。

#### ⑧YouTube ビデオの閲覧

#### ⑨Google map

いわゆる GPS 搭載 Map であり、GPS や Wi-Fi、携帯基地局を利用して現在位置を測定する。検索機能も充実しており、周辺にある目的場所を容易に調べることができる。車に搭載しておけば、ナビ代わりに使用することも可能である。

#### ⑩App Store

iPhone 用のアプリケーションが配信されており、便利なアプリケーションが目白押しである。なお、iPhone SDK が無料で配布されており、これによって独自のアプリケーション開発も可能となっている。

この他にも天気、SMS、時計、計算機、株価などの機能も備わっている。

### ❖ iPhone での中国語の入力方法

さて、iPhone での中国語環境であるが、これは至っ

て簡単である。

まず、入力であるが、デスクトップの「設定」(図1)→「一般」→「キーボード」(図2)→「各国のキーボード」(図3)で、中国語(簡体字)を選択すればよい(図4)。ここで更に、「手書き」と「Pinyin」の両方を「オン」にしておく(図5)。繁体字の場合も同様である。これでたとえば、「メモ」アプリを起動させ入力してみる。

ソフトウェアキーボードを使ってピンイン入力が確実ではある(図6)が、手書き入力の精度はなかなかのもので十分使用に耐えうるものである。(図7)

なお、各国のキーボードを見ればわかるように、日本語、英語、中国語はもちろんのこと、デンマーク語、ドイツ語、スペイン語、フィンランド語、フランス語、イタリア語、韓国語、ノルウェー語、オランダ語、ポーランド語、ポルトガル語、ロシア語、スウェーデン語などにも対応している。

デスクトップ等の表示言語を完全に中国語に変更することも可能である。キーボードと同じように、「設定」→「一般」→「言語環境」で、「簡体中文」を選択すれば(図8)、図9～図12のように、全ての表示が簡体字中国語に変わる。

もちろん、iPhoneはUnicodeを採用しているので、言語環境は日本語あるいは他の言語のままでも、何ら不都合はなく、WEBブラウザ(図13～14)でもメール(図15)でも中国語は正しく表示される。

上述のようにiPhoneのアプリケーションは、語学学習ソフトやビジネスソフト、辞書(図16)、ゲーム、ユーティリティ等々、数多くのジャンルのものが現在までにすでに1000以上がアップされており、App Storeからダウンロードしてすぐに使用することができる。中には無料のものもある。しかも、iPhone SDKを使って自作が可能であり、筆者も現在、Appleと中国語関係のものを共同開発することを考えている。

## ❖ おわりに

最後に、現在までのところ日本で発売されたiPhoneはSIMロックがかけられており、他のキャリアはもちろんのこと、海外で現地のSIMを使用することもできない仕様になっている。香港やイタリアなどでは、SIMフリーとして売られており、海外で使用



図1



図2



図3



図4



図5



図6



図7



図8

したい人は、少々割高にはなるが、それらを購入するのも一つの手である。ただし、その場合は日本では電話機能が使えないことを念頭に置いて欲しい。裏技はなくもないのだが、それは「違法」となることも承知すべきである。

いずれにせよ、これまでの携帯電話や Windows Mobile マシンとは一線を画す iPhone は今後も進化を続けていくはずである。



図 9



図 10



図 11

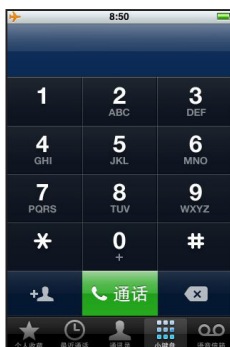


図 12

図 13

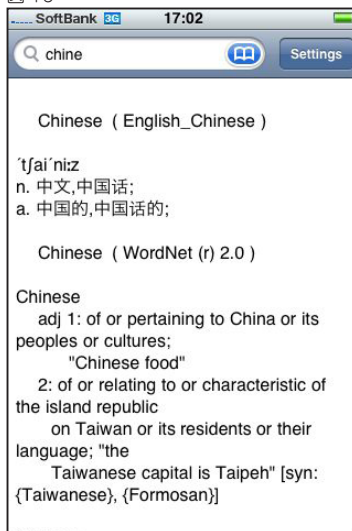


図 14



図 15

図 16



# iPhone & iPod touch 2.0 Software

秋山 陽一郎

## ❖ iPhone 2.0 Software (for iPod touch)

7月11日、満を持して登場したiPhone 3G<sup>[1]</sup>。本誌読者なら、多言語ガジェットとして興味をひかれた方も少なくないのではないだろうか。

筆者も強い関心を持った一人だが、いまだ他のキャリアのユーザーで、iPhone 3Gは手にしていない。ただ幸いすぐに最新のファームウェア（基本ソフト）が公開され、手元のiPod touchでもiPhone 3Gに近い環境を持つことができるようになった。

本レビューでは、そのiPhone/iPod touch 2.0ファームウェア（以下、2.0）の多言語ガジェットとしての潜在力を探ってみようと思う<sup>[2]</sup>。

### ■ iPhone と iPod touch の違い

まず、レビュー用端末にiPhoneではなく、iPod touchを主として利用するため、先にiPhoneとの大まかな違いを、ごく簡単に確認しておきたい（下表）。

これ以外で、iPhone 3Gが持っている主要な機能（App Storeや追加キーボードなど）の大半は、iPhone 2.0 Software Update for iPod touch<sup>[3]</sup>を導入することで、iPod touchでもiPhoneと同様に動作する。

	iPhone 3G	iPod touch
無線通信	WiFi, Bluetooth, 3G (HSDPA)	WiFi
位置情報	A-GPS, 携帯電話基地局, WiFi ホットスポット	WiFi ホットスポット
電話	有（追加アプリケーションでボイスレコーダ化も可能。）	無（2Gはボイスレコーディングに対応。）
カメラ	200万画素相当。 (GPSによるgeoタグの付与可。ただし無線経由で送信できる画像は640×480まで。)	無
センサー	加速度センサー 近接センサー 環境光センサー	加速度センサー
音声出力	ヘッドホン 内蔵スピーカー	ヘッドホン (2Gのみ内蔵スピーカー搭載)
SMS <sup>[3]</sup>	有	無



図1: iPhone 2.0ソフトウェアアップデート (iTunes)

## ❖ 日本語・中国語キーボード

### ■ 利用可能なキーボード

iPhone/iPod touchで利用するソフトキーボードは、ホーム画面から「設定」→「一般」→「キーボード」or「言語環境」→「(各国の)キーボード」で選択できる。

2.0から待望の（簡体字・繁体字）中国語キーボードが、それもQWERTYキーボード（ピンイン入力）と手書き入力キーボードの二点が一気に実装され、日

図2: 日本語・中国語のキーボード





図3：日本語テンキー入力

本語も従来の QWERTY キーボード（ローマ字入力）に、新たにテンキーキーボードが加わっている。

## ■ 2.0 の日本語テンキー入力 [2.0+]

日本では、すでにインターネットへアクセスするデバイスの軸が PC から携帯電話に移行しつつある。とくに若年層では、QWERTY キーボードよりもテンキーによる日本語入力に、より親しみを感じている層が年々拡大してきていて、その意味で 2.0 がテンキー入力というスタイルを採用したのは、必然の流れなのかもしれない。

だが、2.0 のテンキー入力は、従来の携帯電話のテンキー入力とは明らかに一線を画した、画期的なインターフェイスに仕上げられている。

### ● 入力の手順

1. QWERTY キーボードが表示されたら、日本語テンキーキーボードが出てくるまで、地球アイコン（通常、左下に位置する）をタップする。

図4：日本語 QWERTY キーボード



2. 一例として「れ」を入力したい場合は、「ら」をタップ&ホールド（長押し動作）する。
3. ら行のほかの文字が表示されたら、「ら」の右隣にある「れ」の方に指をスライドさせ、「れ」の上で指を離す。
4. 馴れてきたらフリック（払い動作）を使うと劇的に入力が速くなる。たとえば、「ら」を起点に右方向に指をフリックする（払う）だけで「れ」が入力される。「れ」が画面に表示されるまでいちいちホールドしてる必要はない。
5. 予測変換を活用すると、さらに入力が速くなる。変換候補中に目的の語がない場合は、右端の矢印をタップして一覧から選ぶ。

もともと iPhone / iPod touch には、タッチパネル型 QWERTY キーボード（図 2）が実装されていて、日本語のローマ字変換ができた。しかし iPhone / iPod touch の QWERTY キーボードは、わずか 3.5 インチの画面に収めているため、入力しにくいとの批判が第一世代 iPhone の発売当初から絶えなかった。

筆者はこれに代わるインターフェイスがあるとすれば、手書き入力になるのではと想像していたが、よもやテンキーが採用されるとは思いもよらなかった。テンキーが想定外だった最大の理由は、タッチパネルだと、キーストロークのある物理的なボタンと違って「入力した」という実感に乏しく、連続タップが誤入力を誘発しやすい点にある。

だが、2.0 のテンキーキーボードは、そうした負の印象を、タッチパネルならではの手段で見事に覆してしまった。

### ● 日本語テンキー入力の特長

- テンキーだと QWERTY キーボードと比べて表面積が広く、タップしやすい。
- 50 音 = 10 行 × 5 段という日本語の音韻モデルを、巧みに組み込んだインターフェイスデザイン。
- タッチパネルならではの、タップ&ホールドやフリックといったアクション<sup>[4]</sup>を活用することで、濁音・拗音・促音をのぞけば、50 音すべてについて、1 音 = 1 アクションで入力できるようにした。これにより従来の携帯電話のよ



うなボタン連打が必要なくなった。

- もちろん、予測変換機能も搭載。
- 上記の結果、(慣れの度合いにもよるが) 従来のソフト QWERTY キーボードや、携帯電話のテンキー入力と比べて、総合的に誤入力率がかなり減った印象。

なお、QWERTY キーボードもテンキー入力も 2.1 で大幅に動作速度が改善したほか、新たに連文節変換に対応した。

### ■ 中国語ピンイン入力 [2.0+]

中国語もまた、何らかの省入力機構がない限り、QWERTY キーボードによる全拼入力では、「中国」の「中」一文字だけで、5 回もキーを叩かねばならないなど、いちじるしく不便さを強いられる<sup>[5]</sup>。

方式	入力内容	「中」入力に要するキー
全拼	ピンイン全音入力。	[z]+[h]+[o]+[n]+[g]
双拼	声母と字母を入力。	[v](=zh)+[s](=ong)
簡拼	声母のみ入力。(通常は語句単位の入力。)	[z](+[h])
五筆	字根の組み合わせ。	[k]
倉頡	字根の組み合わせ。	[l]
注音	注音字母を入力。	[5]+ [j]+[ / ] (大千式)

2.0 の中国語 QWERTY キーボードは、基本的には簡体字・繁体字ともに、この面倒な全拼(ピンイン全音)入力が採用されているが、実はひそかに簡拼(不完整輸入)が利用できるようになっていた。たとえば「清华大学」なら、“Qing Hua Da Xue”の声母を取って、“q”, “h”, “d”, “x” と入力すると、その声母の組み合わせにマッチする語句である「清华大学」が変換候補に出てくるといった具合だ<sup>[6]</sup>。

これなら、双拼・注音・五筆・倉頡方式のように、特定キーへの字母や字根の割り当てに依存することもなく、また入力するキーの回数も大幅に減らすことができる(中国語に馴染みがあれば、通常は後述の手書きキーボードよりも断然速く入力できる)。

しかし、先行するサードパーティ製の iCosta などと比べると、アップル純正の簡拼入力は、残念ながら今の段階では語彙力が弱く、提示できる候補がまだまだ少ないのが悩ましいところだ<sup>[7]</sup>。



図 5 : 簡拼入力 (左:アップル純正, 右:iCosta)

### ■ 中国語手書き入力 [2.0+]

このほか 2.0 では、これまたタッチパネルならではの手書き入力(手写輸入)インターフェイスに、予測変換機能を加えたキーボードが用意されている。中国語に不慣れな外国人に良いかもしれない。

試しに短めの時事系ニュースをいくつか入力してみたが、「表」字の例のように、多少なぐり書きや続け字気味になっても、認識精度は比較的良好で、サクサク”書き進める”ことができた。日常的に頻出する範囲の語彙であれば、まずまずの出来といつて良いだろう。

しかし反面、「北」など一部の字について、楷書でいねいに書いても、なかなか正しく認識してくれないケースがあった。ここで気になったのは、右側の欄で認識候補が4つしか示されない点で、これは正直にいつてかなり痛い。おそらく、この辺が評価の分かれ目になるうかと思う。

また先行するサードパーティ製の HW Pen (汉王手

図 6 : 中国語手書き入力モード





図7：連絡先データから予測変換候補に



図8：Macのアドレスブック

写)が対応していなかった予測変換(連想)機能についても、登録語彙がまだまだ少なかったり、候補の提示が語句単位(「奥」→「奥运会」)ではなく文字単位(「奥」→「運」→「会」)だったり、改善の余地がかなりある。こちらも併せて今後のバージョンアップに期待したい。

図9：CJKV漢字表示テスト



## ◆ 連絡先を簡易ユーザー辞書に

本誌の読者には、メールや、メモ、カレンダー(スケジュール表)の入力に、専門語彙を登録したユーザー辞書を利用したいという方もおられるだろう。ただ残念ながら、iPhone/iPod touchには今のところユーザー辞書機能がない。

そこで、当面の代替策としてiPhone/iPod touchの連絡先(アドレスブック)をユーザー辞書代わりに利用する方法を紹介しておく。

実はiPhone/iPod touchは、連絡先に登録されている名前を予測変換候補として優先的に扱う仕様になっている、この仕様を利用する。

連絡先の「姓」に単語を、「姓の読み」に単語の読みを登録する。新旧両字体あるいは簡体字・繁体字を同時に登録したい場合などは、同じカードの「姓」「名」を利用して、2語を同時に登録すると管理しやすい。

Macユーザーなら、単語の登録はiTunes経由で同期が可能なアドレスブックで編集・管理するといいたいだろう。グループもiPhone/iPod touch内でちゃんと保持されるため、ユーザー辞書用のデータをアドレスデータと区別しておける点が良い。

Windowsユーザーでも、アドレス帳やOutlook Express、Outlook(2007以降)のほか、GMailやYahoo!メールのアドレスデータをiTunesからインポートできるようになっているが、こちらは振り仮名の取り込みに対応していなかったり、文字化けしたりと、まだまだ問題が多く、現実的な選択肢とはいえない。

## ◆ Safari(ウェブブラウザ)

### ■ CJK漢字出力

結論からいえば、基本的にiPhone/iPod Touchは、本節で取り上げるSafariも含め、CJK統合漢字拡張Bも含めた、Unicodeの基本多言語面(u0000-uFFFF)外の文字まで扱うことができる。(Leopardと同等のUnicode 4.1.0相当?)

おまけにプリインストールされているSTHeitiフォントのおかげで、CJK統合漢字領域はもちろんのこと、CJK統合漢字拡張A領域の漢字も、すべて初期設定の状態でも出力できる。これは多言語PDAとして活用したい方には朗報といっていいたいだろう。

また拡張 B 領域の漢字についても、Hiragino Kaku Gothic ProN や、STHeiti シリーズに収録されている字に限定されるものの、その一部の表示が可能だ。iPhone / iPod touch 用の表示テストページを用意したので、端末をお持ちの方は興味があれば試してみたい<sup>[8]</sup>。

なお、2.0 で CJK 漢字を収録しているフォントは、以下のとおり<sup>[9]</sup>。

- Arial Unicode MS
- Hiragino Kaku Gothic ProN W3
- Hiragino Kaku Gothic ProN W6
- STHeiti J (Medium / Light)
- STHeiti K (Medium / Light)
- STHeiti SC (Medium / Light)
- STHeiti TC (Medium / Light)

#### ■ 任意の拡張 B フォントの表示は可能か？

2.0 で CJK 統合漢字拡張 B の文字を扱うために目下必要なのは、何といてもフォントだ。現状、iPhone / iPod touch で拡張 B 漢字の表示を実現するには、以下の 2 つの手段が選択肢として浮かぶ。

- フォントを端末にインストールする。
- CSS 3 Web Fonts を活用する。

#### ● 端末にフォントのインストール

iPhone / iPod touch のフォントは、  
/System/Library/Fonts/  
に保存されている。この領域は基本的に Jail Break なくしてはアクセスできない上、root 権限（管理者権限）でないと一切編集できない。技術的には、おそらくインストール可能だと思われるが、そもそも手動によるフォントの設定変更自体が非常にリスクが高く、やすやすと勧められる手段ではない<sup>[10]</sup>。

アップル自身による正規の手段（もしくはアップルの検閲を経た App Store 系のアプリケーション）が提供されるまで、静観しておくのがベストだろう。

#### ● CSS 3 Web Fonts

端末への手動インストールよりも、はるかにリスクもハードルも低い選択肢が CSS 3 Web Fonts<sup>[11]</sup> を活用する手段だ。2.0 の Safari はバージョン 3.1.1 がベ



図 10 : Web Fonts 表示テスト

スになっているが、実は Safari はバージョン 3.1 からこの Web Fonts をサポートしている。

Web Fonts は、サーバー上に設置した ttf や otf などのフォントファイルをダウンロードして、一時的にブラウザ上で適用させるようにする CSS (Cascading Style Sheet)<sup>[12]</sup> の仕様だ。

```
@font-face { /* フォントを定義する */
  font-family: myFont;
  src: url("http://domain/font.ttf");
  format("truetype");
}
p { /* 定義したフォントを適用する */
  font-family: myFont;
}
```

この Web Fonts が利用できれば、iPhone / iPod touch の Safari で拡張 B 漢字に限らず、任意のフォントを閲覧者の環境や設定を選ばずに表示できるところだったが、どうやらアップルはこの Web Fonts のサポートを、iPhone / iPod touch 用の Safari から外したらしい。考えられる理由は、以下に挙げるようにいくつかある。

- ハード上の制約。(ファイルサイズの大きいフォントをキャッシュすると、動作が極端に不安定になるなど。)
- インフラ上の制約。(トラフィック量を抑えるためなど。)
- フォントの知的財産権の問題。(Microsoft は、この点から EOT (Embedded Open Type) 形式の標準化を今春から推進している<sup>[13]</sup>。)

数MB～数十MBにも及ぶフォントを、閲覧の都度ダウンロードするのが、はたしてスマートな方法といえるかどうかは検討の余地があるかもしれないが、原文テキストを保持しつつ、表示面も同時にサポートできるメリットは小さくない。特にフォントの入手が困難なマイナー言語や、少数民族の文字、古文字の利用に大きな効果が期待できるだろう。

## ■ その他の Safari の特記事項

まず本誌読者の関心領域として挙げるべきは、文字化けの問題だろう。Mac OS X の Safari や、その他の WebKit 系ブラウザ<sup>[14]</sup> 同様、iPhone / iPod touch の Safari もまた、言語・文字コード判定やフォントとの関連づけが甘い<sup>[15]</sup>。WebKit も、もはや Mac だけで動作する描画エンジンではなくなったのだから、いい加減に何とか改善して欲しいところだ。

iPhone / iPod touch をスマートホンや PDA として見た場合により致命的なのは、コピー & ペーストができないという点だろう。これは Safari に限らず、iPhone / iPod touch 内のすべてのシステムやアプリケーションについてもいえる。

たとえば iPhone / iPod touch でメールを書き込んで、あるウェブページの情報を知人と共有したいと思っても、対象となるページの本文も URL もソフトウェアキーボードで直接入力するしかない。これはかなり快適さを損なっている。

幸い、アップルの関係者によれば、コピー & ペーストは優先順位の関係で実装が遅れているだけとのこと、対応するつもりがないわけではなさそうなので<sup>[16]</sup>、将来のファームウェアアップデートでの対応に期待したい<sup>[17]</sup>。

オフラインといえば、実は iPhone / iPod touch の Safari は、HTML 5 のオフラインストレージ<sup>[18]</sup>に対応している<sup>[19]</sup>。となると、ウェブアプリケーションも、実装次第では今後オフライン状態で部分的な機能が利用できるようになるかもしれない。

## ❖ その他のアプリケーション

### ■ マップ (iPhone/iPod touch 版 Google Maps)

iPhone / iPod touch には、機能をスリム化した Google Maps 専用のビューワーアプリケーション

「マップ」がプリインストールされている。

ビューモードはマップ、航空写真、地図+写真、リストの4種類。マップレット<sup>[20]</sup>は利用できない。

iPhone / iPod touch 版の独自機能としては、GPS による現在地特定機能がある。ただし、iPod touch は Wi-Fi ホットスポットを利用して位置情報を取得するので、iPhone と比べると精度はかなり甘くなるし、付近にホットスポットがなければ、そもそもこの機能自体利用できない。

なお、この「マップ」、本誌8号の山田崇仁氏のレビューでも指摘されていたように<sup>[21]</sup>、中国国内の情報密度が著しく薄いことは覚えておきたい。

- 道路や建物の区画や名前を表示するマップビューは、中国国内分は情報不足で事実上、使い物にならない。これは北京や上海といった都市部も含めてだ。PC版とは使っている地図が異なるらしい。(なお台湾と香港については、中国本土よりも詳細なマップデータが利用できる。)
- 中国の測地系 (BJ-54) と、Google Maps が採用する米国測地系 (WGS84)<sup>[22]</sup> との違いから、航空写真ビューや、地図+写真ビューでも、地図と座標に 100m 前後のズレが生じている。

### ■ カメラ +GPS 機能

前項の GPS 機能は、iPhone に内蔵されているカメラとも連動しているので、たとえば調査旅行で撮った写真を地図と結びつけることも簡単にできる。ただ、今のところ次に挙げるような問題があるので、各自の目的を満たせるかどうかは、必ず事前に一考しておくべきだろう。

- 画質はあまり期待できない<sup>[23]</sup>。特に四隅のボケやカラーノイズは、ある程度覚悟が必要。もちろんオートフォーカスや、手ブレ防止といった自動補正機能もついていない。
- 画素数は 200 万画素あるが、無線経由で転送できる画像サイズは 640×480 まで。
- iPhone / iPod touch が採用している測地系も WGS84 で中国国内と異なる。(中国測地系の地図データにマッピングすると誤差が生じる。)

## ■ App Store

1.1.4以前のiPhone/iPod touchでは、アップルが用意したアプリケーションしか使えず、アップル以外の第三者に、独自のアプリケーションを開発する途は開かれてはいなかった。例外的な存在として、なかば不正にiPhone/iPod touchのロックを解除するJail Breakと呼ばれる処理を施した端末を対象にした、いわば裏のアプリケーション市場があるのみだった。

しかし、今年3月にアップルがソフトウェア開発キット(iPhone SDK)をリリースしたことで、Jail Breakに依らずとも利用できる、正規のアプリケーションが開発できるようになった。このiPhone SDKで開発されたアプリケーションを販売、配布する唯一の窓口として、アップルが3G iPhoneの発売と同時に用意したのがApp Storeだ。このApp Storeは、2.0以降のファームウェアを導入したiPhone/iPod touchなら、本体もしくはPC側のiTunesからアクセスできる。

### ● App Store アプリの特徴

App Storeで入手できる、つまりiPhone SDKで開発されたiPhone/iPod touch用アプリケーションを見てみると、iPhoneのハード面の機能を活かしたアプリケーションが目立つ。

より具体的にいえば、iPhoneのマイクを利用したボイスレコーダー、加速度センサーを利用した水平器やゲーム、GPSを利用したトラッキングログや、現在地付近で撮影された写真のスライドショーといった感じだ。

しかし、コピー&ペースト機能の未実装に加え、iPhone/iPod touch内のデータアクセスや、アプリケーション間のデータの受け渡しに著しい制限が設けられているといった事情もあって、キラーアプリが出ていないのが現状だ。

たとえばWeDictという、App Storeから入手できる英中辞典(51,000語)を標準同梱した辞書アプリがある。実はこのアプリケーション、本来はJail Break済み端末向けに配布されていたもので、StarDict形式<sup>[24]</sup>の辞書ファイルを追加することで、ほかの人が作成した辞書や発音音声、自作の辞書などが利用できた。こうした拡張性が、今のiPhone、iPod touchの実装や窓口がApp Storeに一本化されたシス

テムでは期待しにくい。

こうした携帯端末用アプリケーションの開発や配布がオープンにできるという意味では、今後、GoogleのAndroidや、NokiaのSymbian OSが、iPhoneに対抗する競争力をつけてくる可能性があって、目が離せない。

## 注

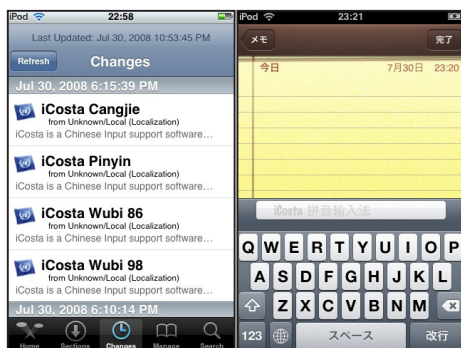
- [1] iPhone自体はこれが第二世代にあたり、ここでいう“3G”は、いわゆる第三世代携帯電話(UMTS=W-CDMA方式)を意味している。
- [2] なお、日本時間9月10日に、第二世代iPod touch(以下2G touch)とiPhoneソフトウェア2.1(以下、2.1)が発表された。音楽のプレイリストに集合知の要素を採り入れたGeniusという新機能を除けば、バグフィックスやパフォーマンス改善がメイン。本レビューでは、2.1で変更があった部分のみ、その旨を言及し、基本的な話題はバージョン1.xに対するメジャーアップグレード版である2.0を軸とする。
- [3] Short Message Serviceの略。ユーザー間で短いメッセージを、チャットのようにやり取りできるサービス。ちょうどMacのiChatのようなものだが、コミュニケーションのペースはそれよりも緩やか。
- [4] アップルのサイト(iPhone 2.0ソフトウェアアップデート/iPhone 2.0 Software Update for iPod touch)とiTunes Music Store(iPod touch 2.0ソフトウェアアップデート)とで表記が異なっているが、いずれも同一のもの。iPhone/iPod touchを動かす、組み込み型の基本ソフトであるため、しばしばファームウェア(firmware)やOS X iPhoneとも呼ばれている。バージョン2.0のファームウェアは、iPhone 3Gについては導入済みの状態で出荷されているが、iPod Touchの利用者は、このiPhone 2.0ソフトウェアをiTunes Music Storeより1,200円で購入する必要がある(図1)。ファームウェアのバージョンが1.1.4(2.0の1つ前のバージョン)以下の場合、App Storeや本レビュー中で紹介する追加キーボードなどの機能は利用できない。  
<http://www.apple.com/jp/ipodtouch/whatsnew.html>
- [5] iPhone/iPod touchの主要操作アクション

タップ (画面を 1 回叩く)	
ダブルタップ (画面を 2 回連打 or さわる。)	
ホールド タップ&ホールド (画面の 1 点に指を固定。いわゆる長押し。)	
ドラッグ (画面に触れたまま指をスライドさせる。)	
フリック (1 方向に指を払う or はじく。)	
ピンチイン (2 本の指を画面に接したまま、つまむように内向きにスライドさせる。)	
ピンチアウト (2 本の指を画面に接したまま、押し広げるように外向きにスライドさせる。)	

[6] 従来 (1.1.4 以前) のファームウェアでは、以下の 3 種類の選択肢があった。いずれも非公式ながら、(馴れの程度にもよるが) 入力効率では 2.0 の純正入力方式に勝るとも劣らないものが多い。このうち、iCosta、WeFIT については、2.0 にも対応している。

iCosta: 拼音・五筆・倉頡などの入力方式に対応。アップル純正の拼音入力よりも強力な熟語 (词组) 変換や、予測 (联想) 変換機能も搭載。

<http://iphone.freecoder.org/iCosta.html>



WeFIT: 拼音 (全拼, 双拼, 全双混合)・五筆・五筆拼音混合・笔画入力に対応。双拼の韻母の配列も、既存形式の中から選択できる。

<http://www.weiphone.com/thread-64422-1-1.html>



HW Pen (汉王手写): 手書き入力のツール。予測変換機能はないものの、田字格 (井形) で区切られた特定の区画 (左半分、右半分、上半分、下半分など) でアルファベットや、数字、标点符号などを手書き認識させられる点はユニーク。

<http://iphonecake.com/soft/hwpen-%E6%B1%89%E7%8E%8B%E6%89%8B%E5%86%99>

ただ、これらは iPhone や iPod Touch のセキュリティロックを解除した、いわゆる jail break (脱獄) を行った端末でのみ動作する。つまり、所有する iPhone や iPod Touch が故障する大きなリスクを伴い、アップルのサポートも受けられなくなることを意味する。導入に踏み切る場合は、よくよく熟考の上、自己責任で利用したい。自分自身で関連情報を調査・理解でき、ある程度の問題解決と責任能力に自信がなければ、導入は控えるよう強くお勧めしたい。

[7] 音節を声母と韻母に分けるというモデルは、中国では反切という音韻表記法の形で古くからある。最近の日本でも、たとえば 2ちゃんねる用語などで、「くわしく」を "kwsk" と子音表記する事例が見られるが、これも簡拼とまったく同じメソッドといえる。

[8] たとえば、「故宫博物院」の場合、iCosta が "ggbwy" で一発変換できるのに対し、純正繁体字版は "gg" + "bwy"、純正簡体字版にいたっては "gg" は「故宫」に変換されるものの、"bwy" は変換できなかった。(一応、「博物館」なら "bwg" で変換できる。)

[9] <http://www.karitsu.org/iphone/jj9.html>

[10] なお、スクリーンショットをご覧になってお気づきになった方もいるかもしれないが、OS X iPhone は、表示したい文字のグリフが、指定したフォント中に含まれていない場合、その文字のグリフを含む別のフォントで代替表示するような仕様になっている。(これは Mac OS X も同様。)

[11] たかがフォントと思いがちだが、最悪、起動不能もありうる。一応、筆者の手元の端末で、/System/Library/Fonts/ 以下に HAN NOM B を置き、CGFontCashe.plist で font name, family name を定義してパスを通す、という方法でインストールを試みてみたが、今のところ変化は見られない。

[12] <http://www.w3.org/TR/css3-webfonts/>  
ほかに、Downloadable Fonts、Embedded Fonts、@font-face などと呼ばれることもある。

実は過去に 1997 年に Netscape 4 と IE 4 が Web Fonts を実装した実績があり、後者は今でも IE 6, 7 で利用できるが、eot (Embedded Open Type) という、独自の形式にフォントを変換する必要がある。一方の Netscape も pfr (TrueDoc Portable Font Resource) という独自形式への変換が必要で、Safari のように直接 ttf, otf ファイルを利用できるわけではなかった。なお、9 月 10 日に Firefox 3.1 のテストビルドが ttf, otf の直接読み込みに対応した。Safari, Opera (Acid3 用ビルド) に続き、これが 3 例目の実装になる。

[13] CSS は、ウェブページの外観を整えるためのスタイル言語。

[14] <http://lists.w3.org/Archives/Public/www-archive/2008Aug/att-0010/EOT-charter-draft-1.html>  
<http://www.w3.org/Submission/2008/01/>  
<http://blogs.msdn.com/ie/archive/2008/07/21/font-embedding-on-the-web.aspx>  
<http://cwilso.com/2008/07/23/fonts-embedding-vs-linking/>

EOT フォントの特徴は、バンダーが著作権をある程度コントロールできる点とサブセット化 (部分抽出) ができる点 (特に後者はファイルサイズが大きくなりがちな東アジア圏のユーザーには大きなメリットといえるかもしれない)。ただし、ttf や otf を eot 形式に変

換するには、Microsoft が無償配布する WEFT (Web Embedding Fonts Tool) を介する必要がある。

<http://www.microsoft.com/typography/WEFT.msp>

[15] たとえば今年 9 月初頭にベータ版が公開された Google Chrome が WebKit ベース。詳しくは、本誌ブラウザレビューを参照。

[16] W3C の Richard Ishida 氏のテスト結果を見ても、この辺の処理について Safari が突出して悪いことは一目でわかる。

<http://www.w3.org/International/tests/results/results-lang-and-cjk-font>

[17] <http://www.engadget.com/2008/07/14/apple-says-turn-by-turn-gps-coming-to-iphone-copy-paste-not-a/>

[18] なお、使い勝手は快適と言いたいものの、iCopy と呼ばれる、Safari でコピー & ペーストを実現するブックマークレットが公開されている。

<http://www.biocow.com/iCopy/index.php>

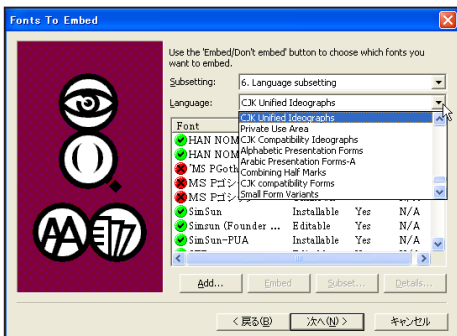
ブックマークレット (ブックマークに URL の代わりに短い JavaScript を記述することで、ブラウザの機能を拡張できるミニプログラム) であるため、Safari 内でしかデータの受け渡しができない。しかし、アップルがコピー & ペースト機能を実装するまでのしぎの選択肢としては貴重な存在といえる。

[19] HTML 5 - 4.11 Structured client-side storage (W3C)  
<http://www.w3.org/TR/html5/structured.html#structured>

[20] <http://jsgt.org/mt/archives/01/002157.html>

JavaScript から SQL を使い、1 データベースあたり最大 5MB の iPhone / iPod touch ストレージ領域を利用できるとのこと。

[21] ここでいうマップレットとは、個人や団体・企業などが独自に公開している情報を Google Maps 上に、レイヤーとして追加表示するサービスのこと。たとえば、



Flickr にアップロードされた geo タグ (位置情報) つきの写真を、Google Maps 上に出力する Flickr Gmap Show や Flickr Photo Maplet など。

この機能が iPhone や iPod touch といった小型携帯端末用にも提供されるようになれば、かなり利便性は高まると思われる。Google は現在、携帯電話向け OS の Andorid と、その Android に載せるウェブブラウザの Google Chrome の開発を進めており、こうした小型携帯端末向けのサービス拡充も、今後期待していきたい。

[22] 山田崇仁「Google マップ」(本誌 8 号、2007)

[23] Google Maps も米国の WGS84 測地系を採用している。日本国内では日本測地系から世界測地系 (ITRF 測地系) への移行が進みつつあるが、最近ではこの ITRF と WGS84 との間の差は、ほとんどなくなっていると

のこと。

[24] 画質が気になる方は、以下の Flickr のページを参考にするといいだろう。

<http://www.flickr.com/cameras/apple/iphone/>

(Flickr に投稿された写真の中から、iPhone で撮られた写真だけを動的に抽出したページ。カメラの種類は、写真に埋め込まれている Exif データから判別されている。)

[25] <http://stardict.sourceforge.net/HowToCreateDictionary>

StarDict は、Linux や Windows で動作するフリーの辞書アプリケーション。ネット上から多数の辞書データを落として拡張できる点が強み。

<http://en.wikipedia.org/wiki/StarDict>

<http://stardict.sourceforge.net/download.php>

## WWW ブラウザ

上地 宏一

### ❖ ブラウザ戦国時代再来か

毎年恒例のブラウザシェアの紹介ではなく、今年はいきなり結論を述べてしまうと、まさにブラウザ戦国時代の再来を思わせるようなニュースが満載の 1 年であった。特に検索大手の Google が独自ブラウザ「Google Chrome」を発表し、ベータ版が公開されたことは、ブラウザ業界への黒船到来とも言えるだろう。

#### ■ Netscape Navigator の終焉

一方で、第 1 次<sup>[1]</sup> ブラウザ戦国時代の雄であった Netscape Navigator が、2008 年 2 月の最終版 (9.0.0.6) をもって幕を閉じた<sup>[2]</sup>。2 月末にてサポートも終了し、現在は Firefox やその派生である Flock へ移行するようにアナウンスされている。かつてブラウザの代名詞であった「ネットスケープ」はついになくなってしまったのである。

### ❖ ブラウザ別概況

#### ■ IE (Windows Internet Explorer)

2008 年 6 月末で販売が終了するはずであった

Windows XP がそれ以降も OEM (DSP) 版として継続販売されることになった。また ASUS の EeePC、ACER の Aspire one、MSI の Wind Netbook、Lenovo の IdeaPad といった製品が次々と登場する低価格ミニノート PC の多くにおいて Windows XP が搭載されていることもあり、IE 7 を標準搭載する Windows Vista の普及が足踏み状態の中、結果として IE 6 から IE 7 への移行スピードも鈍化しているようである。2 月から開始された IE 7 への自動更新も強制ではなくユーザーがインストールの有無を明示的に選択できるようになっているため、読者の中には「インストールしない」を選択した方も多いただろう。

#### ● IE 8 ベータ第 2 版リリース

そんな中、早くも次バージョンである IE 8 のベータ第 2 版が 2008 年 8 月に公開された。IE 7 では標準的な CSS への準拠が強化されたため、その反動として既存の IE 6 向けに作成された Web ページの表示に問題が生じる現象が多発している。そこで IE 8 ではさらなる CSS への準拠を目指しつつ、IE 6 および IE 7 に対応した表示機能も搭載することで互換性を高めているとしている。新しい機能としては閲覧記録等を残さないプライベートブラウズが挙げられるが、多



漢字・多言語表示等に関して目新しい点はない模様である。

### ■ Firefox (Mozilla Firefox)

昨年の記事では Firefox 2 について触れているが、さらに開発が進み 2008 年 6 月に Firefox 3 が登場した。その際、全世界同時リリースを行い 24 時間以内に 8 億ダウンロードを記録してギネス記録として認定されている<sup>[3]</sup>。第 2 版の系列について 2008 年中は引き続きセキュリティ面の修正が行われる。

Firefox 3 についても多漢字・多言語機能の目新しい点はないが、強いて挙げるならばフォントが用意されていない文字について、中黒(・)や四角(□)ではなく具体的な符号位置が表示されるため、どの文字が表示できていないかを簡単に知ることができて便利である(図 1, 2)。

### ■ Safari (Apple Safari)

#### ● Windows 版の正式公開

昨年の記事で Windows 用 Safari のベータ版公開を紹介したが、その後 2008 年 3 月の第 3.1 版公開時に Windows に正式に対応した。ところが 5 月末に Microsoft は Windows XP / Vista と Safari の組み合わせにおいて重大な危険性が生じるため、Safari の使用を控えるようにとの発表を行った<sup>[4]</sup>。6 月中旬の第 3.1.2 版でこの問題は解決したが、結果的に Windows 版 Safari の不安定さが目立つことになった。加えて、開発元 Apple 社の音楽総合管理ソフトである iTunes のアップデート時に自動的に Safari が一緒にインストールされるなどマイナスイメージが先行しているのが現状である。さらに筆者の環境では Ext. B 領域の文字表示が確認できていない。

また Windows 版固有の特徴として、文字の描画を Windows の標準機能 (GDI+) ではなく独自のエンジンを用いていることが挙げられ、Windows の ClearType とは異なったアンチエイリアス機能を用いた、Mac OS X に近いフォント表示がなされる。残念ながら標準で用いられるフォント「MS ゴシック」は文字線のウェイトが太いため全体的に重い(黒い)画面になることが一部では不評である<sup>[5]</sup>。

#### ● CSS3 のリモートフォント指定

Safari の一つの特徴は CSS への積極的な対応である。

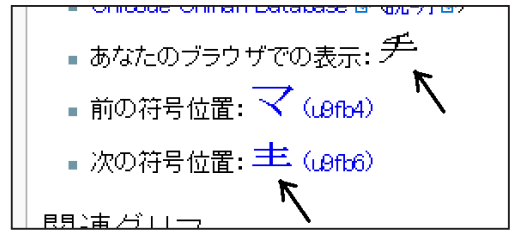


図 1: フォントが表示されている例

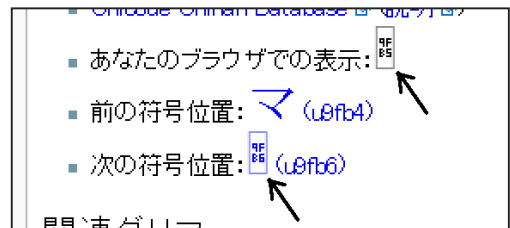


図 2: フォントがなく符号位置が表示されている例

その中でも CSS3 の「@font-face」はネットワーク上のフォントリソースを用いた描画を指示できるようにする記述であるが、Safari では第 3.1 版からサポートしている。詳細は後述する。

### ■ Google Chrome

Google 社による独自ブラウザ Chrome (クロム、クローム) が発表され、Windows 用のベータ版が 2008 年 9 月 2 日より先行配布された。いわば「黒船到来」とも言えるこのブラウザは Safari と同じ Apple 社が中心に開発を行っている WebKit を HTML 描画エンジンとして利用しているため高速なページ表示が可能となっているほか、ページ内での処理を高速に行うことができる JavaScript エンジンについては Google の独自開発となっている。

操作インターフェースや機能面においては他のブラウザで有用とされる機能を採用しているほか、今までのブラウザにはない特徴も持っている。次に実際にベータ版 (Official Build 1583) を触ってみた感想を紹介する。

#### ● 一見普通のタブブラウザだが…?

Chrome は最近流行のタブブラウザである。しかしよく見るとタブの位置とアドレスバーの位置が逆になっている(図 3, 4)。つまり、タブごとにアドレスバーが存在している。さらにアドレスバーは Google の検索バーを兼ねているのである。

また Web ページを開き、アプリケーションとしてのショートカットを作成すると、アドレスバーを伏せた状態でページが開かれるという機能を備えている。先号では Google コミュニケーションツール群について紹介したが、これらのブラウザ上で動作するツールを Chrome で実行するとあたかも 1 つのアプリケーションが動いているように見せることが可能となるわけである。その恩恵を一番受けそうなのが Gmail である。Google 謹製のスクリプトエンジンを搭載しているためか、非常にきびきびと動作する (図 5)。

## ●多言語周りは…残念

残念ながら Ext. B の表示については文字送りなどの表示が崩れるため、不適であると言わざるを得ない。また Safari と同じ WebKit を使っていながら「@font-face」には対応していない。IE と同じくフォント (グリフ) が用意されていない文字について別のフォントで代替する機能がない。このあたりは Firefox に大きく水をあけられているといえる。

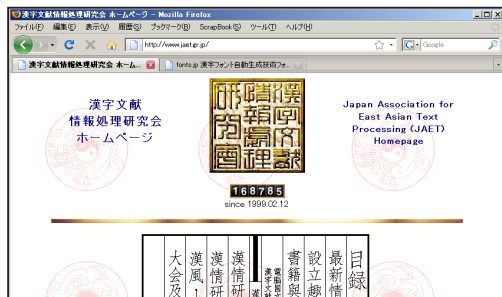
## ●今後の展開に期待

細かい機能については実装されていないほか、セキュリティに関する不具合も早速報告されている。

図 3 : Google Chrome によるページ表示例



図 4 : Firefox による同ページ表示比較



ベータ版の使用はあくまで評価用と考える必要がある。とはいえ、Gmail などの Google サービスを利用するだけでも便利であるほか、既存のブラウザとは異なる新たな展開に期待したいところである。

## ❖ CSS3 の @font-face と多漢字処理への期待

現在 Safari でのみ実装されている機能として CSS3 に準拠した @font-face 記述があるが、本稿ではこの機能に注目してみようと思う。

@font-face 記述はたとえるならばリモートフォント機能であり、Web の表示においてインターネット上にある <sup>[6]</sup> フォントファイルを指定することで、そのフォントを用いた表示を可能とするものである。実は @font-face は CSS2 から存在していたが、実際に利用できたのは IE 4 以降のみであり、フォントファイルも通常の TrueType ではなく WEFT と呼ばれるツールによって変換された Web ページ埋め込み専用ファイルを用意する必要があったため、ほとんど普及せずに CSS2.1 では仕様からいったん削除されたが、CSS3 で復活すると見られている。@font-face 記述はとくに Web のデザインを追及する上で追加された機能と考えられるが、見方を変えれば多漢字 Web にこの機能を使うことが有用であると思われる。たとえば Web 上であるテキストデータを公開する際に、Ext. B 領域の文字を使っていたと仮定すると、Windows XP や Mac OS X では ISO/IEC 10646 の Ext. B 領域に完全対応したフォントが含まれていないため、すべての文字を表示することができない。また文字コードに含まれない文字を外字フォントとして用意することも考えられる。従来、閲覧側がフォントファイルを入手して自分の PC にインストールする必要があったが、ここにリモートフォント機能を活用すれば、閲覧側が特に操作をすることがなく自由に不足する文字をネットワーク上のフォントファイルを用いて表示させることが可能となる。

具体的な例を見ていただきたい (図 6)。

ISO/IEC 10646 に 2008 年に追加された CJK Strokes (CJK 筆画) 領域や、漢字 20,902 字の後に追加された 30 字、あるいは「功」や「吉」に異体字を割り当てたものや外字領域を用いた文字を画像ではなくテキストとして表示することが可能となっている。

この Web ページは <style> タグに以下の @font-face 記述を施している。

```
<style type="text/css">
@font-face {
font-family: "gottani";
src: url('http://glyphwiki.org/font/gw035360.ttf') format("truetype");
}
.sample {
font-family: gottani, 'M S 明朝',
serif; font-size: 175%;
}
</style>
```

src 属性で続けて指定した URL のフォントファイルを指定し、font-family 属性で指定した名称で実際に表示対象を指定する。このサンプルの場合、<div class="sample"></div> でくくられた範囲について、このフォントが用いられることになる。Windows 版 Safari ではサロゲート領域を用いることができないので、実際には外字領域 (U+E000 以降) を活用するのが無難であろう。

### ■ グリフウィキとの併用

手前味噌ではあるが、この @font-face 記述は漢字字形データ管理 Web システムであるグリフウィキ<sup>[7]</sup>と親和性が高い。グリフウィキは任意の漢字字形を登録し、また他者が登録した漢字字形から必要なものを選んだ漢字集合を任意のコードポイント (たとえば外字領域) に割り付けることができる。割り付けた漢字集合は直ちに TrueType 形式のフォントファイルに変換することが可能で、変換されたフォントは同時にインターネット上で公開された状態となる。このフォントを @font-face で指定した Web コンテンツを作成すればよい。グリフウィキに登録されているデータはすべて自由に利用できるのでフォントファイルの権利処理も不要である。

### ■ 現状での問題点

紹介した @font-face 記述は多漢字 Web の有効な切り札になると期待できる。しかし現状でこの機能を利用できるのは残念ながら Safari に限られる。また



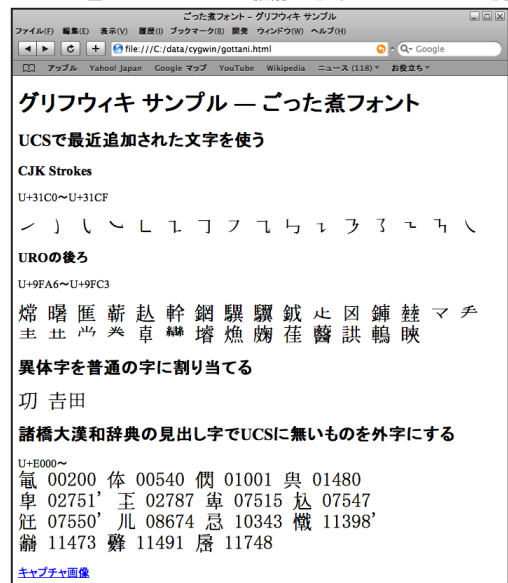
図 5: Gmail をアプリケーションとして表示

Windows 版 Safari では Ext. B 領域の利用ができない模様である。このように今すぐに活用することは難しいが、いずれ標準機能として各ブラウザに搭載されることが予想される。

また文字に対して通常とは異なる文字を割り当てることが可能となるため、テキストデータの情報と視覚化される情報を異なったものにもできる (たとえば「一」というコードポイントに「十」という漢字字形を割り当てれば、そのフォントでは「一万円」というデータは「十万円」と表示されてしまう) わけであり、新たなセキュリティ面での懸念材料ともなり得る。この欠点も知った上で正しく活用することが求められるとも言える。

近い将来、@font-face 機能が多漢字処理の 1 つの解

図 6: @font-face 機能を活用した Web ページ例



決方法となることを期待するものである。

## 注

- [1] ここでは Netscape と Internet Explorer によるブラウザ競争時代を第 1 次と表現している。
- [2] The Netscape Archive <http://browser.netscape.com/>
- [3] Mozilla Japan プレスリリース <http://mozilla.jp/press/releases/2008/07/02/>
- [4] マイクロソフト セキュリティ アドバイザリ (953818) <http://www.microsoft.com/japan/technet/security/advisory/953818.mspx>
- [5] 反対に Windows のアンチエイリアス機能や ClearType による描画に不満を持つ人が作成した、フォント描画を Mac OS X に近い描画機能で代替するアプリケーションも公開されている (<http://drwatson.nobody.jp/gdi+/> あるいは <http://sourceforge.net/projects/inkstone/>)。また、Safari を独自描画エンジンではなく Windows 標準の GDI+ を用いた描画に変更する設定方法が Web で公開されている。 <http://webkit.org/blog/168/gdi-text-on-windows/>
- [6] 実際はインターネット上だけでなく PC 内のフォントを指定することもできる。
- [7] グリフウィキ <http://glyphwiki.org/>

## Office 2008 for Mac

師 茂樹

### ❖ 最新の Mac 用 Office

#### ■ Apple の戦略における Microsoft Office

Microsoft の Office 2008 for Mac<sup>[1]</sup> は、Mac OS X 用に開発された最新のオフィススイートである。Apple には iWork という独自のオフィススイートがあり、そこには Pages (ワープロ)、Numbers (表計算)、Keynote (プレゼンテーションソフト) などのソフトウェアが収録されている。近年、各所で Microsoft Office 離れが進みつつあるとは言え、やはり大局的なシェアで言えば Microsoft の優勢は揺るぎないというのが現状であろう。Windows 用の Office 2007 が出て以来、Office 2004 for Mac のバージョンアップも待たれていたが、1 年遅れての登場である。

ところで Apple はここ数年、Windows との比較広告を展開したり、Mac OS X 用の Internet Explorer の同梱をやめたりするとともに、Windows 用の Safari や iTunes を公開したりするなど、Microsoft への対抗を示唆するような動きを見せている。一方、Mac のハードウェアで Windows を起動するための Boot Camp を Mac OS X に同梱したり、ウェブサービスである MobileMe (かつての .Mac) を Windows に対応させたりすることで、Windows との共存を強調するようなメッセージも発信している。

Office 2008 for Mac が後者の戦略にとって重要なものであることは間違いないし、そのような文脈を離れても、完成度の高いきわめて有用なソフトウェアであることは間違いない。しかしながら、そもそも少なからぬ Mac ユーザにとって、Office の存在意義は必須のアプリケーションというよりは Windows ユーザが作ったファイルのビューワないしコンバータというものではないかと思われるうえ、Apple が Google との連携を強めながらウェブサービスを強化したり、iWork が着実にバージョンをあげている状況などを見ると、「Microsoft がいない世界」を実現するまでの「つなぎ」なのではないか、という穿った見方をしてしまう。今後の Apple の動きに注目したい。

	通常版	Special Media Edition	ファミリー & アカデミック
Word	○	○	○
Excel	○	○	○
PowerPoint	○	○	○
Entourage	○	○	○ <sup>[2]</sup>
Messenger	○	○	○
Expression Media	×	○	×

## ■ 収録されているソフトウェア

さて、Office 2008 for Mac には、通常版 (Office 2008 for Mac) と「Special Media Edition」、3 台までインストール可能な「ファミリー&アカデミック」版の 3 種類のパッケージがある。ファミリー&アカデミックにはアップグレード版はなく、またこのほかにも Word や Excel の単体パッケージ (通常版+アップグレード版) がある。それぞれの収録ソフトウェアは下表の通りであるが、見ての通り大きな差はない。

今回ラインナップに追加された Expression Media について、ここで簡単に紹介しておこう。これは、デジタルカメラの写真や動画、音楽ファイルなどを効率よく管理するためのソフトウェアで、Apple の iTunes や iPhoto などと似たような機能を持つソフトウェアである。しかし、Microsoft 製品全体のなかの位置づけとしては、Office といっしょにバンドルされるべきものというより Visual Studio などの開発系に近い。すなわち、アドビシステムズの Flash や AIR、Google の Ajax を中心とした開発環境などに対抗して Microsoft が公開しているクロスプラットフォームの リッチ・インターネット・アプリケーション実行環境 Silverlight のための統合開発環境 Expression Studio の一部と位置づけられているのである。

## ■ 他の Office との互換性

上に述べたように Office 2008 for Mac が「Windows ユーザーが作ったファイルのビューワないしコンバータ」としての性格を持っているならば、他の Office 製品との互換性、特に Windows 用の Office 2007 との互換性は気になるところである。以下、主立ったところを列挙しておこう。

- Office 2007 で導入された Office XML 形式 (拡張子が docx、xlsx など) に対応。
- インターフェースは Office 2007 の新しいインターフェースとは異なる (メニューはかつての Office を踏襲)。→図 1
- Windows で作成された文書のレイアウトの再現性が向上。→図 2、図 3
- VBA 非搭載 (後述)。
- PowerPoint のスライドテーマには、Windows がないフォントを使った独自のものが多い

(Office 2004 for Mac と同様)。

- フォントについては次節。

上にあげたもののうち、ネット界限で特に問題視されているのが VBA の非搭載である。Microsoft の提供する Q & A によると、

Excel 2008 には VBA は搭載されておりません。そのため、VBA を使用するソルバーや分析ツールも提供しておりません。こちらについては、ユーザーの皆様からのご要望が大きいく、マイクロソフトでは代替案を調査しているところです。しかしながら、どうしても VBA や VBA を使用するソルバーなどのツールを使う必要がある場合は、当面の間は Excel 2004 をお使いいただきますようお願いいたします。VBA で組んだマクロをお使いの場合は、VBA を AppleScript に変換するための情報を英語版 mactopia web サイトより提供しております。<sup>[3]</sup>

ということであるが (なお、VBA 非搭載は Excel 2008 だけではなく Office 2008 全体である)、VBA から AppleScript への変換コストは小さいとは言えない (ちなみに、上の引用では英語版のサイトへ誘導しているが、日本語のサイトも存在する<sup>[4]</sup>)。VBA よりも AppleScript の方が優れている場合もあるだろうが、気軽に作業を記録してマクロ化する機能や、これまで蓄積した VBA マクロを使い回せないのは、やはり大きな問題であろう。Mac は Mac らしく AppleScript を使うべきだと考えたのだろうか、それとも単なる開発コストの問題なのだろうか、Microsoft がなぜこのような判断をしたのかはわからない。

## ◇ CJK フォントの収録状況

Office 2008 for Mac をインストールする際、同時にインストールされる CJK フォントは下表の通りである。

### ■ JIS X 0213:2004 への対応

この表のなかでまず注目されるのは、メイリオの収録であろう。メイリオは Windows Vista に標準で搭載されたフォントで、JIS X 0213:2004 における字



図1：Word 2008 for Mac のツールバー

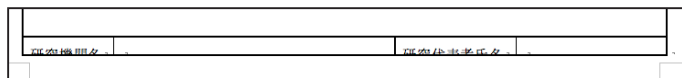


図2：Windows で作った Office 文書を Office 2004 for Mac で開いた場合（レイアウトがずれている）

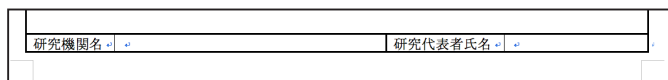


図3：図2と同じファイルを Office 2008 for Mac で開いた場合

体の変更に対応したことで話題となった（図4）。これにあわせてMS明朝・MSゴシックなども、JIS X 0213:2004 に対応した字体変更がなされている。

Mac OS X は Leopard の段階で JIS X 0213:2004 に対応したフォント（ヒラギノ明朝 ProN など）を搭載しているが、Windows Vista との互換性という意味では Microsoft の新しいフォントが搭載された点は評価できるだろう（もちろん、過去の Windows、Office との間には字体の非互換が発生する）。

## ■ Extension B は部分的対応

一方、本誌の読者にとって残念だと思われるのは、SimSun-ExtB や MingLiU-ExtB などの CJK Extension B 領域をカバーするフォントが提供されていないことだ。Extension B の文字のうち、ごく一部は JIS X 0213 にも収録されているので、先に紹介したメイリオなどで表示その他ができるが、それ以外の大部分については別途フォントを入手しなければならない。

フォント名	バージョン
Batang	1.50
Gulim	1.70
メイリオ	5.02
MS 明朝	5.02
MS ゴシック	5.02
MS P 明朝	5.02
MS P ゴシック	5.02
PMingLiU	4.55
SimSun	2.92

そのほかにも、中国語のフォントの種類が少なかったり、プロポーショナルではない無印の MingLiU が収録されていないなかったりと、日本を除く東アジアの言語環境を軽視しているのではないかと、という印象がある（日本以外の東アジア各国における Apple のシェアの問題もあるのかもしれない）。Mac OS X にバンドルされているフォント（蘋果儷中黒など）である程度はカバーできるが、他の Office で作った文書が書体化けしないためには、各 Office のフォントセットはできるだけ共通化してほしいと思う。

## ◇ 最後に

以上、簡単に Office 2008 for Mac について紹介してきた。冒頭にも述べたように、所謂「あちら側」（Web 側）にストレージやアプリケーションを置く Google 的なモデルが成功例として注目されている現在、Windows という特定のプラットフォームとともに歩んできた Microsoft Office の立ち位置はかつての盤石な時代と比べると安定したものではないし、Mac OS X 版はさらに不安定なものだと言える。Apple がプラットフォームを超えて iTunes を提供したことにより iPod の成功を導いたように、Microsoft Office も各 OS 間の違いを最小限にとどめたクロスプラットフォーム戦略（どんな環境でも Office があれば同じように仕事ができる）を展開すべきなのかもしれない。しかし、VBA の非搭載ということが端的に表しているように、Office 2008 for Mac は Windows 版との差別化、すなわちプラットフォーム依存性をむしろ強めてしまっている。Google Docs でも Office 文書が開けるようになっている現状を考えると、Windows との間をとりもつという Office 2008 for Mac の重要な存在意義が失われてしまうのではないかと懸念してしまう。杞憂で終わればよいのだが。

## 注

- [1] 製品のパッケージには「Microsoft Office:mac 2008」と表記されているが、本レビューでは「Office 2008 for Mac」という表記で統一する。
- [2] ただし Exchange 非対応。
- [3] <http://www.microsoft.com/japan/mac/products/office2008/qanda.mspx#006> (2008年9月10日最終確認)
- [4] <http://www.microsoft.com/japan/mac/library/articles/asforoffice.mspx> (2008年9月10日最終確認)

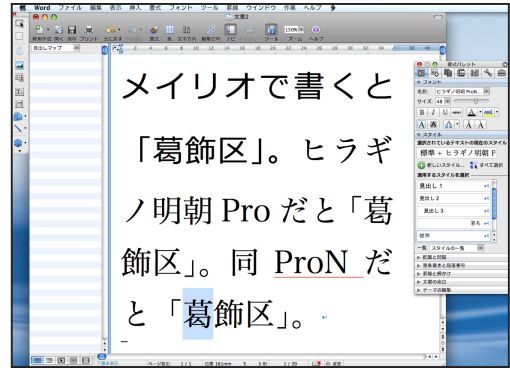


図 4: メイリオとヒラギノ

## フォント

上地 宏一

### ❖ 文字コード、フォントがなければただの表

現在の一般的なパソコンには文字コードとして Unicode が採用され、多くの PC で Ext. B 集合を含めた 7 万漢字を処理することが可能となっている。実際には、内部で 7 万字の処理が可能となってもフォントが用意されていなければ画面表示や印刷はできないため、「使える」とは到底言えない。

また、ネットワークを通じて情報を公開する場合には自分の PC だけでなく想定される閲覧者の PC のことも考慮する必要がある。具体的には、インターネットでテキストデータを公開する際、どの文字コードを利用して、さらにどの文字集合を用いるかについては計画段階で十分に検討することが求められる。

#### ■ UTF-8 の採用

従来、テキストデータの公開に用いる文字コードは複数の候補があり、それぞれ一長一短であった。もっとも標準的な JIS X 0208 を用いると必要な漢字が足りなく、台湾・香港で用いられていた Big-5 が古典テキストに適していると言われたこともあった。さらに過去には今昔文字鏡や TRON コードなどの選択肢もあった。

Unicode (ISO / IEC 10646) の普及によって文字

種不足の問題はおおむね解決したと考えられ、また Unicode を扱うツールも充実してきたことは本誌第 7 号の「特集 1: 極める! デジタルテキスト」で紹介したとおりである。また Unicode の符号化については UTF-8、UTF-16 などがあるが、Web ブラウザ等で標準的に処理可能な UTF-8 を使うことが妥当であろう<sup>[1]</sup>。

### ❖ 一般の PC で使えるフォントとは

次に重要なのは、Unicode に収録されている漢字のどの集合を用いてテキストを作成するかの検討である。すべての PC が 7 万漢字をフルに利用できるフォントを搭載しているとは限らない。すると一般に普及している OS に標準搭載されているフォントがサポートしている文字集合のみを対象とすることが考えられる。たとえば Windows XP に標準搭載されているフォントについて考えた場合に、日本の文字デザインによるフォントに限定する場合と、中国(大陸)デザインなど、多少日本とは異なるデザインのフォントも含めて検討する場合とでは利用できる文字集合が大きく異なる。さらには、無償で入手できるフォントや、Microsoft Office に代表される、多くの PC にインストールされているであろうアプリケーションに添付されるフォントも考慮することで利用できる文字集合が大きくなることも考えられる。

そこで本稿では、一般的な PC において用意されているフォントが収録する漢字集合をまとめ、インターネットでテキストデータを公開する際に、Unicode のどの程度の文字集合を用いることが妥当であるかの目安を示したいと思う。

## ■ 本稿の調査対象

本稿では OS として現在販売されている Windows XP、Windows Vista、Mac OS X (Leopard) を調査対象としている。なお Windows XP は SP2 Professional Edition を対象としたが、現行の SP3 や Home Edition でも同様と思われる。同様に Windows Vista については無印の Home Premium を対象としたが SP1 や他のエディションでも同様と思われる。Mac OS X については、Microsoft Office を導入している利用者も多いと考え、今回入手できた「Office 2004 for Mac」を調査対象として追加した。現バージョンの「Office 2008 for Mac」については本号に同名のレビュー記事があるのでそちらを参考にしてほしい。このほか無

償で入手できる大規模漢字フォントならびに Ext. B 領域を初めて多数収録するというインパクトのあった「Simsun (Founder Extended)」を加えた明朝体およびゴシック体フォント合計 41 種類<sup>[2]</sup>を調査した。

## ■ 調査結果

実際に 41 種類のフォントを調査した結果は表の通りである。グループ化されたフォントごとにフォント名、フォントのバージョン、書体、どの地域のデザインが施されているか、そして実際の収録漢字数を示した。収録漢字数は 5 つのグループ (Ext. A 領域 U+3400 ~ U+4DB5、CJK 統合漢字領域 U+4E00 ~ U+9FC3、CJK 互換漢字領域 U+F900 ~ U+FAD9、Ext. B 領域 U+20000 ~ U+2A6D6、CJK 互換拡張領域 U+2F800 ~ U+2FA1D) に分けている。また漢字部品 (CJK 部首、康熙字典部首) や漢文記号、CJK 字画は対象外とし、また外字、Adobe Japan1 集合、および OpenType フォントにおいてフィーチャとして収録され UCS に変換されないものは対象外とした。

Windows XP の標準フォント

フォント名	バージョン	書体	デザイン	収録漢字数				
				Ext.A	統合	互換	Ext.B	互換
MS 明朝	2.31	明朝	日本		12,204	34		
MS ゴシック	2.30	ゴシック	日本		12,204	34		
SimSun	3.12	明朝	大陸	52	20,902	21		
SimHei	3.02	ゴシック	大陸		20,902	21		
MingLiU	3.21	明朝	台湾香港		20,902	302		
Batang	2.21	明朝	韓国		7,476	268		
Dotum	2.21	ゴシック	韓国		7,476	268		

Windows Vista の標準フォント

フォント名	バージョン	書体	デザイン	収録漢字数				
				Ext.A	統合	互換	Ext.B	互換
MS 明朝	5.00	明朝	日本	164	12,579	98	304	
MS ゴシック	5.00	ゴシック	日本	164	12,579	98	304	
メイリオ	5.00	ゴシック	日本	202	10,755	359	339	45
SimSun	5.00	明朝	大陸	6,582	20,910	21		
SimSun-ExtB	5.00	明朝	大陸				42,711	
SimHei	5.01	ゴシック	大陸	6,582	20,902	21		
YaHei	5.00	ゴシック	大陸	6,582	20,909	21	8	
MingLiU	6.02	明朝	台湾香港	6,582	20,916	302		
MingLiU-ExtB	6.02	明朝	台湾香港				42,711	11
JhengHei	5.00	ゴシック	台湾香港	6,582	20,902	302		
Batang	2.21	明朝	韓国		7,476	268		
Dotum	2.21	ゴシック	韓国		7,476	268		
Malgun Gothic	5.00	ゴシック	韓国					



Windows XP 向けにダウンロード可能なフォント

フォント名	バージョン	書体	デザイン	収録漢字数				
				Ext.A	統合	互換	Ext.B	互拡
MS 明朝	5.00	明朝	日本	164	12,579	98	304	
MS ゴシック	5.00	ゴシック	日本	164	12,579	98	304	
メイリオ	5.00	ゴシック	日本	202	10,755	359	339	45
SimSun-18030	2.06	明朝	大陸	6,582	20,902	21		

Windows Vista 向けにダウンロード可能なフォント

フォント名	バージョン	書体	デザイン	収録漢字数				
				Ext.A	統合	互換	Ext.B	互拡
MS 明朝	2.50	明朝	日本	164	12,579	98	304	
MS ゴシック	2.50	ゴシック	日本	164	12,579	98	304	

Windows 用 Office、Proofing Tools に付属するフォント

フォント名	バージョン	書体	デザイン	収録漢字数				
				Ext.A	統合	互換	Ext.B	互拡
Simsun (Founder Extended)	1.00	明朝	大陸	6,582	20,902	21	36,862	
Arial Unicode MS	1.01	ゴシック	混在		20,902	302		

Mac OS X の標準フォント

フォント名	バージョン	書体	デザイン	収録漢字数				
				Ext.A	統合	互換	Ext.B	互拡
ヒラギノ明朝 Pro W3	7.11	明朝	日本	202	10,755	110	339	45
ヒラギノ明朝 ProN W3	8	明朝	日本	202	10,755	110	339	45
ヒラギノ角ゴ Pro W3	7.11	ゴシック	日本	202	10,755	110	339	45
ヒラギノ角ゴ ProN W3	8	ゴシック	日本	202	10,755	110	339	45
Osaka	4.4	ゴシック	日本		6,356			
STSong	6.0d4e1	明朝	大陸	6,582	20,910	21	6	
STXihei	6.0d5e1	ゴシック	大陸	6,582	20,910	21	4,241	
STHeiti	6.0d4e1	ゴシック	大陸	6,582	20,910	21	6	
Hei	5.0d1e1	ゴシック	大陸		6,763			
Apple LiSung	6.0d1e2	明朝	台湾香港		13,060	2		
LiSong Pro	6.0d4e1	明朝	台湾香港	511	17,607	4	1,640	11
Apple LiGothic	6.0d1e1	ゴシック	台湾香港		13,060	2		
LiHei Pro	6.0d4e1	ゴシック	台湾香港	511	17,607	4	1,640	11
AppleGothic	6.0d11e1	ゴシック	韓国		4,620	268		
Arial Unicode MS	1.01	ゴシック	混在		20,902	302		

Mac 用 Office に付属するフォント

フォント名	バージョン	書体	デザイン	収録漢字数				
				Ext.A	統合	互換	Ext.B	互拡
MS 明朝	2.52	明朝	日本	164	12,572	98	303	
MS ゴシック	2.52	ゴシック	日本	164	12,572	98	303	

Web で配布されているフォント

フォント名	バージョン	書体	デザイン	収録漢字数				
				Ext.A	統合	互換	Ext.B	互拡
HAN NOM A	2.0; 2005	明朝	混在	6,582	20,902	362		
HAN NOM B	2.0; 2005	明朝	混在				42,711	542

## ■ OS 別の標準的な想定漢字集合

この調査結果をまとめ、OS ごとに標準的な漢字集合を想定した。

### ● Windows XP

日本デザインに限定すると JIS X 0208 の 6,356 字 (JIS 規格で非漢字とされる「全」を含む) および JIS X 0212 の 5,801 字に加えて IBM 拡張漢字 360 字のうち重複しない 81 字の合計 12,238 漢字に限定されるが、Microsoft の Web サイトから JIS X 0213 部分が追加された「MS 明朝」、「MS ゴシック」および「メイリオ<sup>[3]</sup>」を入手することができる。これにより日本の文字コード規格をすべて網羅する 13,145 漢字<sup>[4]</sup>が利用できる。

またデザインにこだわらなければ「SimSun-18030」の入手によって Ext. A 領域はすべて網羅でき、標準の「SimSun」や「MingLiU」によって CJK 統合漢字についても 20,902 字をカバーする。

残念ながら Ext. B 領域および CJK 互換拡張領域については JIS X 0213 が割り当てた 304 字あるいは AdobeJapan1-5 集合で割り当てられる 384 字しかサポートされず、Windows XP の利用者を対象とする場合は Ext. B 領域全面的利用は推奨されない。ただし Office Proofing Tools の購入によってその一部である 36,862 字を収録する「Simsun (Founder Extended)」の利用も考えられる<sup>[5]</sup>。36,862 字の内訳は中国 (大陸) の文字コード規格や資料を典拠とする 30,521 字<sup>[6]</sup> + その他 6,341 字であり康熙字典・漢語大字典の掲出字の範囲であれば問題ない。

### ● Windows Vista

日本デザインのフォントに限ってしまえば状況は Windows XP と変わらないが、中国 (大陸) デザインおよび台湾・香港デザインによる Ext. B をすべて収録したフォントが標準で付属することは前号「特集 1: 漢情研的 VISTA & Office2007」で紹介したとおりである。また CJK 統合漢字の末尾に追加された 30 字についても 22 字については収録された<sup>[7]</sup>。まとめると日本の文字コード規格の漢字すべて、あるいは Unicode の漢字すべてが想定される標準漢字集合となる。

### ● Mac OS X

ヒラギノフォントを搭載する Mac OS X については JIS X 0208 および JIS X 0213 の両規格に対応しているが、標準では JIS X 0212 に対応したフォントを搭載していない。

ただし Office for Mac を購入していると想定することで JIS X 0212 を全収録したフォントがインストールされるので、Windows XP と同じ状況となる。またデザインにこだわらなければ Ext. A 領域および CJK 統合漢字 20,902 字についてはすべてサポートしている。CJK 統合漢字の末尾部分の追加 30 字については 20 字が収録されている。一方、残念ながら Ext. B 領域について全面収録されたフォントは搭載していない。ところで文字パレットで Ext. B 領域を見ていると、ヒラギノフォントのほかに時々文字が埋まっていることに気がつく。これは「STXihei (宋体細黒)」フォントによるもので、Ext. B 領域の 4,241 字をカバーする。これはベトナムの文字コード規格を典拠に追加された文字集合であり<sup>[8]</sup>、ベトナムのチュノム文献をテキスト化する際に有効と考えられる<sup>[9]</sup>。このほか、香港の文字コード規格を典拠とする 1,640 字<sup>[10]</sup> (Ext. B 領域) を収録する「LiSong Pro」「LiHei Pro」フォントがある。

まとめると想定される標準漢字集合は日本デザインにおいては JIS X 0208 および JIS X 0213 であり、Office をインストールしていると仮定すると Windows XP と同様に JIS X 0212 にも対応する。あるいは Ext. A 領域および CJK 統合漢字に対応したフォントも搭載していることになるが、Ext. B 領域については特殊な一部を除いて対応していない。

## ■ 補記: HAN NOM フォントについて

今回の調査対象に HAN NOM フォントがある。本誌第 6 号「Unicode 対応フォント」でも紹介したこのフォントは無償で配布されているが、収録される文字のデザインについて Unicode コンソーシアムの配布する文字リストの字形とよく似ていると言われ、その出自が不明であった。本稿の調査に際し、HAN NOM フォントのグリフデータと Unicode コンソーシアムの PDF データを比較したところ、興味深いことが判明したので補記しておこうと思う。

Unicode の PDF データ<sup>[11]</sup>を良く見ると、ある拡大率において、文字の黒塗り部分が白抜けで表示され

ることがある。フォントは漢字字形の輪郭をデータ化したものであるが、筆画ごとに輪郭が独立している場合、輪郭が重なったときに黒塗りの白抜けが生じることがある（通常のフォントは1つの輪郭だけで表現できるように輪郭同士の結合処理がなされているが、本フォントでは文字によって部分的に結合されずに残っている）。ところでPDFデータで使われている Ext. B のフォントは漢字の筆画とは関係ない奇妙な輪郭の分割を施しているケースが見られる。この分割の状態が HAN NOM フォントの多くで共通しているのである（図1。本来ならば Unicode 側の図形も並べて示したいが権利上差し控える）。

すべての文字について分割の状態が一致するのではないが、少なくとも HAN NOM 側で輪郭が結合されずに残っている状態であれば Unicode 側でも同じ分割の状態となっている。これは明らかに同じデータではないだろうか。また、細かいケースでは図2のように、輪郭の内部に残ってしまった明らかに不要な座標点が、HAN NOM フォントと Unicode のPDFとで一致するのである。

完全に同じデータであるならば、両者の図形を拡大率の変更によって数値的に完全一致させることができるはずであるが、微妙にずれていることも分かった。おそらくこれはなにかの時点で小数点などの丸め処理を行ったものと考えられる。

ここで言える結論としては、HAN NOM フォントと Unicode が配布する PDF 内のデータは根を同じくすると言ってほぼ間違いないということである。HAN NOM フォントの作成者が権利関係を明確にクリアしているとは思えず<sup>[12]</sup>、現時点ではこのフォントを積極的に「利用してください」と言うべきではないだろう。

## ❖ まとめ

費やした調査時間に対して月並みな結論になってしまいが、日本デザインフォントに限定するのであれば、JIS X 0208、JIS X 0212、JIS X 0213、IBM 拡張漢字の集合が標準的な漢字集合と考えられる。この集合に対応するフォントは「MS 明朝」「MS ゴシック」のバージョン 2.5 および 5.0 系列である。またデザインにこだわらないのであれば Ext. A、CJK 統合漢字 20,902 字についても利用しても良い。一方で Ext. B 集合については標準とするには時期尚早である。

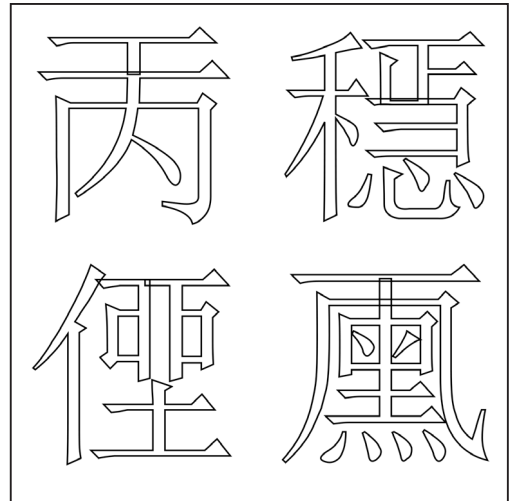


図1 不自然な輪郭分割の例。HAN NOMとUnicodeで一致する

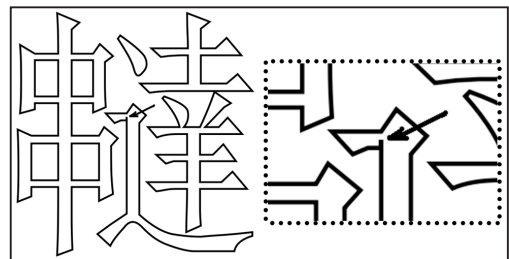


図2 両者で同様に取り除かれなかったゴミ座標（右は部分拡大）

実際には Ext. B に収録される 42,711 字をすべて利用するようなテキストは現実的ではなく、必要な文字種は 1,000 字程度であろう<sup>[13]</sup>。グリフウィキ<sup>[14]</sup>などを用いて必要な字種のみを収録した専用フォントを作成し、テキストデータ公開と合わせて配布することも一つの方法であろう。

今後、Ext. C などの更なる漢字集合の追加が予定されている。文字コードの拡張と OS によるフォント実装の時間差はまだ続くことになり、我々にとって頭の痛い日々が終わることはないのかもしれない。

## 注

- [1] UTF-8 符号化は漢字 1 文字に最低でも 3 バイト必要になってしまうので、漢字文献のテキスト化には不適なものであるが。
- [2] 本稿では同名のフォントでもバージョンの異なるもの

は区別している。

- [3] メイリオは JIS X 0212 自体は収録しない
- [4] 12,238 漢字 + JIS X 0213:2000 非重複 897 字 + JIS X 0213:2004 追加 8 字 + その他 2 字
- [5] Office 2007 より Proofing Tools の提供形態が変更されたが、現在提供される Language Pack に収録されるフォントについては手元に情報がない。
- [6] なぜか中国（大陸）ソースのうち U+20089、U+215D7 の 2 字は収録されていない
- [7] 残りの 8 字については日本の AdobeJapan-1 に由来するため、将来的に「メイリオ」でサポートされると思われる。
- [8] 厳密にはベトナムソース 4,228 字（なぜか 4 字だけ足りない）+ 中国ソース 12 字 + U+2146D
- [9] 筆者はベトナムの文字コード規格がチュノム文献の漢字集合をどの程度網羅しているか不明なため、あくま

で推測である。

- [10] 厳密には Ext. B 領域における香港ソース 1,682 字のうち 42 字は収録されていない。
- [11] <http://www.unicode.org/charts/PDF/U20000.pdf>
- [12] [http://vietunicode.sourceforge.net/fonts/fonts\\_hannom.html](http://vietunicode.sourceforge.net/fonts/fonts_hannom.html) の記述に現れる 3 名が Unicode コンソシアムと関係があるとは思にくい。明らかにデザインが異なる CJK 互換拡張領域部分の作成について 2 年を費やしたということであろうか。
- [13] 筆者が関わったテキストデータのケースでは、30 種類弱の典籍 400 万字のうち、Ext. B 領域相当が 1,000 種類強であった。
- [14] GlyphWiki <http://glyphwiki.org/> 収録されているデータのうち Ext. B 領域は 9,569 字（2008 年 9 月 17 日現在）となっている。もちろん不足する文字は自由にデザインして登録可能である。

## 楽々中国語 V5

金子 眞也

### ❖ はじめに

「楽々中国語 V5」が 2008 年 3 月 20 日に発売された。マイクロソフト社製 IME が Windows 本体に実装され、Google からも無料で中国語 IME がダウンロードできる現在<sup>[1]</sup>、日本国内で中国語入力ツールが商用ソフトたりうるためには、1) 変換効率が優れていて操作性も日本人に無理なく受け入れられること、2) 付加機能が優れていること、の二点が重要となる。そこで、本稿でもこの二点にしぼって検証していく。OS による動きのちがいをチェックするため、動作検証には Windows XP SP3 マシン（新規購入のいわゆる Netbook）と SP1 を適用した Windows Vista (Home Premium) マシンの 2 機種を用いた。

### ❖ 白水社『中国語辞典』搭載

「楽々中国語 V5」の最大の特徴は白水社『中国語辞典』を搭載したことである。「楽々中国語」(cWnn)の変換効率の良さはもともと定評がありバージョン・

アップによってさらに磨きがかかったようだが<sup>[2]</sup>、バージョン・アップの真価は実はそんなことにはなくて白水社の辞書搭載という点にある。搭載する辞書が増えて、白水社『中国語辞典』、講談社『現代中国語新語辞典』、ナツメ社『言いたいことがすぐに見つかる中国語会話辞典』、小学館『日中辞典』『中日辞典』第二版となった。

「楽々中国語 V5」には『中日大辞典』は搭載されていないものの、白水社『中国語辞典』が使えるメリットは大きい。

### ❖ 便利な「辞書ブラウザ」

「楽々中国語 V5」には「辞書ブラウザ」という辞書引き機能がある。特に便利なのが「全文一括辞書引き」機能で、辞書引きしたいウィンドウを指示すると、自動的に文章を解析して辞書引きを実行する。結果をリスト表示することもできる。図からわかるように、これは中国語初心者で Web 上の中国語を読む人にとって、たいへん便利で親切な機能といえよう。ただし「楽々中国語」による文章の解析力に過度の期待

を抱いてはいけない。

また、「全文一括辞書引き」機能を利用して中国語ウェブサイトの辞書引きを行う場合、サポートするブラウザが事実上 Internet Explorer 6 に限定される<sup>[3]</sup>ので注意が必要である。IE 以外のブラウザに慣れたユーザにとっては寂しい対応だ。また、IE7 との相性にはかなりの問題がある<sup>[4]</sup>。修正モジュール等の公開が待たれるところだ。

IE7 との相性問題は、おおざっぱに言うと、「Windows Vista + IE7」の組み合わせでは若干の制限がつくものの使用 OK、「Windows XP + IE7」の組み合わせは不可である。「Windows XP + IE7」環境でも Microsoft Word あたりにコピーしてから辞書引きを行うことはできるが、面倒なので推奨しない。

なお、すでに IE7 を導入してしまった Windows XP ユーザが「楽々中国語 V5」を使う場合は、「IE を 6 に戻す」ことをお勧めする。IE7 を削除すると IE6 に戻るが、うまく行かない場合は Microsoft サポートオンラインの「Internet Explorer 7 を削除する方法<sup>[5]</sup>」に載っている手順で IE を 6 に戻すことができる。

### ❖ 文例ナビ

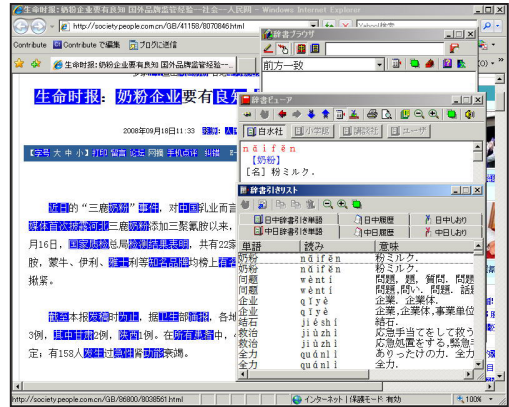
外国人が中国語で文章を書く場合の強力な「助っ人」が「文例ナビ」だ。前バージョンにも搭載されていたが、今回のバージョンでナツメ社・白水社・小学館中日・日中の 4 種の辞書に対応するようになった。日本語を検索窓に入力すると、対応する例文を表示する、中国語を検索窓に入れて検索することはできない、という作り込みじたいは前バージョンと同じである。

中国語を検索窓に入れて検索できたらうれしいのだが、そういうユーザは購入層としておそらく想定されていないのだろう。残念なことである。

### ❖ その他機能てんこもり

「楽々中国語 V5」には前バージョンになかった中国語手書き入力機能や中国語音声合成エンジンが搭載されている。機械翻訳機能もついているし、「読み振りツール」で中国語にルビを振ることもできる。

中国語学習ツール「陳さんのやさしい中国語教室」も健在だ。発音の基礎(第 1 課・第 2 課)と基本表現(第 3 課～第 6 課)は CD-Rom からの起動、第 7 課～第



12 課まではオムロンソフトウェアのサイトへのリンクが貼られている。挿絵はきれいだし教材の作り込みは良い方だ<sup>[6]</sup>。

「楽々中国語 V5」は中国語初心者向け機能満載で、お買い得感にあふれているといえよう。

### ❖ 注意すべき点など

「楽々中国語 V5」では声調つきピンインを入力できるのは cWnn の独自コードのときだけで、GB で打っているときには声調つきピンインの入力ができない<sup>[7]</sup>。従って中国語学習者あるいは中国語教員がピンインつきデータを他人と交換する場合には、相当の注意が必要である。

### ❖ 「楽々中国語 V5」は買いか？

「楽々中国語 V5」は初心者にやさしいソフトとしてよく練られたソフトである。中国語初心者の学生が買うのなら、まあお買い得かもしれない。中国語教員が教材作成用に購入するのも、まあ、悪くないだろう。

ただし、「楽々中国語 V5」を使用する上で絶対外せない条件が一つある。それは、「楽々中国語の独自コードで作ったピンインつきデータには互換性がないことをよく理解すること」である。

データを USB メモリに入れて手軽に持ち運びができる時代だが、せっかく持ち運んでも同じソフトの入っていないマシンで読むとデータが文字化けしてしまう<sup>[8]</sup>。

「楽々中国語 V5」で作ったピンインつきデータは

互換性の問題があるので出先ではなく自宅で印刷する、それさえ守れば「楽々中国語 V5」はお買い得で便利なソフトとってよい。

## ❖ ChineseWriter9 と比較して

「楽々中国語 V5」はアカデミック価格で 18690 円、これに対して ChineseWriter9 の Master でない方はアカデミック価格で 21800 円。価格差は約 3000 円である。ChineseWriter9 のアカデミック版はダウンロード販売だと 19800 円になるので、両者の差は約 1000 円まで縮まる。

「白水社の辞書」(楽々中国語 V5) を取るか「ピンインの互換性 + 中国語単語からの例文検索」(ChineseWriter9) を取るか、難しい選択である。

ちなみに、筆者は両方使っている。「楽々中国語」の以前のバージョンの登録ユーザには、優待販売の利用<sup>[9]</sup>をぜひお勧めしたい。

## 注

[1] <http://tools.google.com/pinyin/>

[2] ただし、cWnn は変換候補として中国要人の固有名詞

への対応を想定していないと考えられる。hujintao や wenjiabao はもとより maozedong も一発で変換できない。ChineseWriter ではこれらの人名は簡単に変換できるので、筆者からみると不便に感じてしまう。

[3] IE6 以外のブラウザへの対応についてはマニュアルに特に明記されていない。念のため Firefox (3.0.1) や Google Chrome (0.2.149.29) で試してみたが、きちんと動かなかった。「動く」とも「動かない」とも言っていない以上、メーカの責任は問えないものの、やや寂しい。

[4] <http://www.omronsoft.co.jp/SP/win/raku2cv5/support/cyui.html#vista>

[5] <http://support.microsoft.com/kb/927177/ja>

[6] ただし、調音点や調音方法を図示するといった配慮は一切ない。

[7] 「読み振りツール」経由等の裏技は存在するが面倒である。

[8] ChineseWriter でも独自コードを使用した場合に同様のことは起こりうるが、ChineseWriter の場合には「独自コードを使わないでピンインを入力する」選択が可能である。後者の方法をとれば文字化けは回避できる。

[9] [http://mars.mc-webshop.com/omron/event/index.php?main\\_page=index&cPath=20](http://mars.mc-webshop.com/omron/event/index.php?main_page=index&cPath=20)

# Google ピンインIME

千田 大介

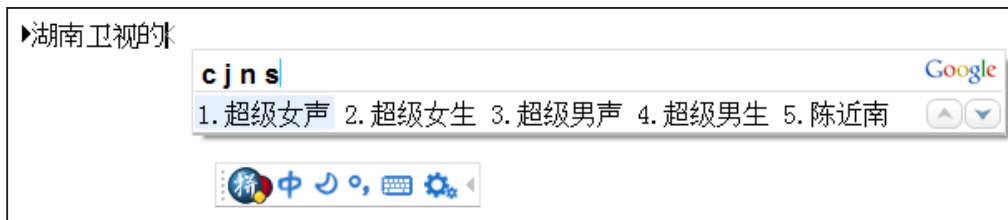
Google 拼音輸入法 (以下、Google ピンイン) は、Google 中国、穀歌 (谷歌) が無償配布している簡体字中国語 IME である。β 版の公開は 2007 年の春、正規版の配布開始が 7 月、それから一年あまりの間に頻りにアップデートを繰り返しており、執筆時点の最新バージョンは 2008 年 8 月 8 日公開のバージョン「1.2.30.71」である。

Google ピンインはごくごくオーソドックスなピンイン入力 IME である。そこから逆に、Google 中国の本気を感じとることができよう。ピンイン入力 IME としては、Windows 標準の微軟拼音輸入法 (以下、MS ピンイン) が広く使われているが、それに対して Google ピンインがいかなるアドバンテージを持って

いるのか検証してみよう。

Google ピンインは、Google 中国サイト (<http://www.google.cn/>) からダウンロードする。穀歌トップページの「更多」の右の「▼」をクリックして、Google のサービス一覧に移り、そこで「拼音輸入法」をクリックすると、ダウンロードページに移動するので「下载谷歌拼音输入法」をクリックして IME をダウンロードする。セットアップはダウンロードしたファイルをダブルクリックするだけだ。日本語環境ではダイアログが英語で表示されるので、迷うことは無かる。

Google ピンインのピンイン入力方法は、MS ピンインと近く、カーソルキーで変換候補を選択し、スペースバーで確定する。単語の区切り位置変更が上下カー



ソルキーというのも同様だ。ただし、MS ピンインでは、ピンインをずらずらと入力してから一度スペースバーを押し、未確定状態のインライン入力された文字に対してカーソルキーを使って変換候補を選択し確定させていくのに対して、Google ピンインの方はシステムライン入力であるので、Google ピンインのボックス上で文字列を確定させなくてはならない。このため、長いフレーズは一気に入力せずにこまめに確定させていくことになる。MS ピンインから乗り換える場合、慣れるのに多少の時間が必要かも知れない。

また、細かな入力設定にも違いがある。両者共に、ピンインの声母だけを入力して確定させる短縮入力に対応しているが、MS ピンインは、声母単位で入力しないとしないのに対して、Google ピンインでは冒頭の一文字だけで短縮入力できる。つまり、「中国」の場合、MS ピンインでは「zhg」と入力しなくてはならないのに対して、Google ピンインでは「zhg」はもとより、「zg」だけでも変換できるのである。ピンインの綴りから発想する短縮入力としては、Google ピンインの方がより利便性が高いのではなかろうか。

Google ピンインは、繁体字入力にも対応している。ただし、文字パレットは装備していないので、難読文字の入力には難儀する。この点は改善を求めたい。

Google ピンイン最大のアドバンテージは、辞書にあると言えよう。筆者が試用した限りでは、MS ピン

インに比べて収録されている語彙が格段に多く、かつごく最近の新語や流行語、話題の人物名などへの対応に優れているようだ。たとえば、「liuxiang」と入力すると、MS ピンインでは第一候補、かつ唯一の熟語の変換候補が「流向」であるのに対して、Google ピンインでは一発で「刘翔」に変換できる。「超級女声」(スーパーガール (オーディション番組))、「李宇春」(クリス・リー)、「颐达」(日産 TIIDA)、「愿景」(ビジョン)などの MS ピンイン未対応語彙も、Google ピンインは収録している。このため、中国語の変換効率も相当に高くなっている。

もう一つの特徴的な機能が、ユーザー辞書のオンライン同期機能だ。この機能は、Google 中国のアカウントを取得すると使えるようになる。ネットを介してユーザー辞書を共有できるので、家と職場、デスクトップとノートなど複数の PC を使い分けているユーザーにとっては嬉しい機能だ。

以上のように、Google ピンインの完成度はなかなか高いので、中国語入力を頻繁に使う人であれば、十分に試してみる価値があろう。なにより、これまで MS ピンインの独壇場であった AI 変換対応ピンイン IME の分野に、Google ピンインという有力な対抗馬があらわれたことは、今後のピンイン入力 IME の発展のためにも好ましいことであると言えよう。

## 一太郎2008

山田 崇仁

### ❖ 文章をいかに作成しやすくするか？

毎年バージョンアップを重ねる一太郎。今年の 1

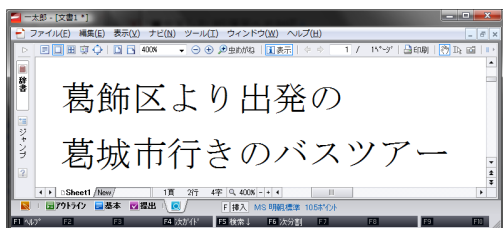
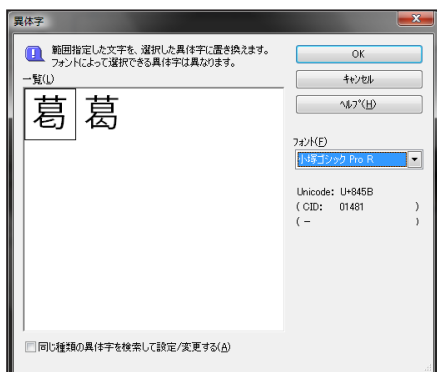
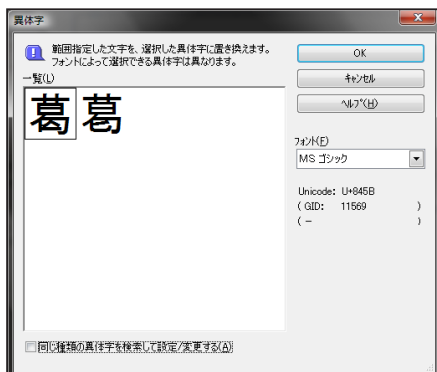
月に、数少ない Mac OS X 用日本語ワープロソフトであり、いち早く Adobe-Japan1-5 の完全対応を果たした EGWORD が市場からの撤退を行い、ますますこの分野が Word 一色 (と、まだまだ実用にはほど遠

い Web サービスとして提供されるワープロや、UTF-16 のサロゲートペア対応が未だ実装されていない (OpenOffice.org Writer は存在するが) に染まっている感がある中、一太郎はまだまだその存在感を示し続けている。

## ❖ 充実した JIS 漢字への対応

以前のバージョンから、一太郎は UTF-16 のサロゲートペアに対応しており、Unicode の CJK 統合漢字拡張領域 A、B などの BMP 以外の面の文字が使用可能であった。そのため、JIS X 0213 の漢字は既に使用可能だったのだが、今回のバージョンアップでは単に入力だけでなく、プラス  $\alpha$  の機能が付け加わった。

それが、「異体字フォントの入力に対応」機能である。



## ■ 異体字フォントの入力に対応

2000 年の国語審議会による「表外漢字字体表」の答申と、それを受けた JIS X 0213:2004 の改正に伴って、Windows が標準で添付する日本語用フォントのデザインが変更されることになった。これまでの JIS 漢字の文字デザインを使う必要がある (例: 葛城市) と新しいデザインを使う必要がある (例: 葛飾区) 場合を混在したい場合 (例: 葛飾区より出発の葛城市行きのバスツアー)、Windows 環境ではフォントを一文字ごとに切り替えるか、Adobe InDesign のような、OpenType の異体字切り替え機能に対応したソフトウェアと使う必要があった。

一太郎は、今回のバージョンアップで、後者の方法に対応した。そのため、一太郎の文書内では「葛飾区より出発の葛城市行きのバスツアー」が表示可能となったのである。

但し、フォントによっては、異体字の関係が MS 明朝・ゴシック・メイリオなどと逆になってしまうので注意が必要である。また、当然他のソフトウェアとの互換性もないので、Word や HTML 形式で保存した場合、これらの異体字情報はすべて消去されてしまうのは残念ではある。

## ❖ おわりに

駆け足で一太郎 2008 を紹介した。ますますディープなユーザーにとって使い勝手がよくなったであろうし、異体字変換機能は魅力的なことは確かである (ATOK と連動してくれれば言うこと無かった)。

バージョンアップ前、リボンインターフェイスへの対応をどうするのが興味があったのだが、一太郎は従来から使われてきた使用状況に応じたフェーズと呼ばれる複数の編集画面切り替え機能を洗練させ、加えてマウスジェスチャーにも対応するという答えを出してきた。

これは、一太郎のテーマである「日本語文章の作成しやすさ」にはリボンインターフェイスがまだまだ洗練されていない、即ち一太郎の過去の蓄積こそが洗練の証ということを宣言したのだと判断したい。来年はどのようなレビューをお届けすることができるのか、あと半年ほどで新バージョンが出るはずなので、今から楽しみに待っておこう。



# 学術リソース レビュー

毎号、メンバーが固定化されているような気がしたので、本号は例年より多めに新しい人に記事をお願いすることにした。突然の依頼に快く応じてくれた執筆陣に誌面を借りて御礼申し上げたい。

冒頭の記事をお願いした岡本氏は、言わずと知れたメールマガジン「アカデミックリソースガイド」の編集者であり、おそらく日本でもっとも人文系サイトに精通した方であろう。他の方もいずれおとらぬ、その業界では名を知られていた人である。そのため、送られてきたレビューはそれぞれの知見の蓄積から厳選されたエッセンスそのものといえ、大いに知的好奇心を満たす内容になったと自負している。

読者諸賢におかれては、斯界の達人たちが厳選したこれらの Web サイトを足がかりに、それぞれのお気に入りサイトを増やしていただきたい。また、それらのサイトがまた本会で紹介されていなければ、是非とも BBS にてご紹介いただければ幸いである。

また、学術ソフトも、基本工具書である『漢語大詞典』の最新バージョンと、唐代長安の復元をデジタル 3D 化したソフトである『唐長安の数碼重建』について採り上げた。こちらも併せてご参考いただきたい。

## Contents

学術サイト	RSS リーダーで購読すべき人文系のブログ・Web サイト.....岡本 真 138	
	中国哲学.....齊藤 正高 142	中国史.....佐藤 仁史 145
	大蔵経データベース.....石井 公成 149	
	東大寺文書目録データベース.....照井 武彦 153	
	中国のネットと IT 事情.....千田 大介 158	
	図書館・OPAC・漢籍..小島 浩之 162	中国語教員必見サイト.金子 眞也 165
	「龍源」(Dragon source) —— 中国語雑誌オンライン購読サービスを試す —— .....小川 利康 167	
学術ソフト・製品	『漢語大詞典』 CD-ROM ver.3.....千田 大介 172	
	唐長安の数碼重建 (A digital reconstruction of Tang Chang'an) 小島 浩之 174	

## ❖学術サイト

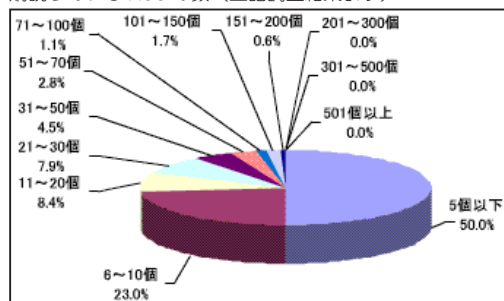
### RSSリーダーで購読すべき人文系の ブログ・Webサイト

岡本 真

#### ❖ はじめに—RSSリーダーの利用実態—

2007年4月の調査では<sup>[1]</sup>、インターネット利用者のRSSリーダーの認知度は約40%に、利用率は約20%に達し、RSSリーダーとRSSが徐々に普及していることがうかがえる。とはいえ、まだRSSの利用が一般化したとは到底言えないことも事実だろう。RSSリーダーは、一度使いだせば便利さを実感し、手放せなくなるツールの一つであるが、その最初の一歩までが意外に遠い。また使い始めても、購読するRSSの数がこれまた意外に少ないらしい。その実態を先の調査が明らかにしている。

購読しているRSSの数（上記調査結果より）



国立国会図書館



ご覧の通り、購読しているRSSが5個以下というRSSリーダー利用者が実に50%もいる。購読数が少なければ、RSSリーダーはその威力を発揮しにくい。5個程度のRSSであれば、気になるサイトを個別に巡回するのときほど大差はないからだ。だが、購読数を10個、20個、30個と増やしていくと、RSSリーダーの便利さを実感できるはずだ。様々なサイトやブログで発信されている情報が時系列に沿って一覧表示されると、新たな発見があり、これまで見えなかった関係が見えてくる。こういった発見は人文科学に限らず、科学全般に関わる人々にとって欠かせない作業だろう。

とはいえ、まずは購読するRSSを増やさなければ始まらない。そこで、筆者が購読する約1000のRSSの中から、人文系の関心を有する方々に購読を勧めたいRSSを紹介していこう。本稿を参考にまずは購読するRSSを増やし、RSSリーダーの醍醐味を味わってほしい。

#### ❖ 基本の3点+α

人文系の図書館、公文書館、研究機関では次の3ヶ所は最低限抑えておこう。

##### ● 国立国会図書館

<http://www.ndl.go.jp/>

国立国会図書館サイトの更新情報が配信される。更新情報に掲載されるのは、近代デジタルライブラリー等へのデータ追加情報や、『国立国会図書館月報』『カレントアウェアネス』『アジア情報室通報』といった刊行物の発行情報。また「テーマ別調べ案内」の更新情報が掲載されることもある。国立国会図書館の電子化事業が目覚ましく進展しているだけに、有用な情報が得られる。だが、すべての更新情報がここに集約

されるわけではないことに注意したい。たとえば「アジア情報室のホームページをリニューアル」のように図書館向け更新情報として、「図書館へのお知らせ ([http://www.ndl.go.jp/jp/library/library\\_news.html](http://www.ndl.go.jp/jp/library/library_news.html))」だけに掲載されていることもある。

#### ● 国立公文書館

<http://www.archives.go.jp/>

公文書館ニュースに掲載された情報が配信される。国立公文書館の運営に関する情報が多いが、国立公文書館が運営するアジア歴史資料センター (<http://www.jacar.go.jp/news.html>) や国立公文書館デジタルアーカイブの更新情報 (<http://www.digital.archives.go.jp/news/>) も含まれている。ただし、両サイト独自の更新情報に比べると、RSS で配信される情報はだいぶ要約されている。この点には注意が必要だ。

#### ● 国立国語研究所

<http://www.kokken.go.jp/>

2007年のサイトリニューアルに際してRSSの配信が始まっている。配信されるのは更新情報を中心。更新頻度はそれほど高くないが、逆に定期的に巡回するサイトではない分、RSSリーダーでの購読に向いている。なお、「言語データベースとソフトウェア」で新たなリソースやツールが公開された際のニュースは、ほとんどこのRSSには含まれない。コーパスや辞書、ソフトウェアに関心がある場合は、「言語データベースとソフトウェア <http://www.kokken.go.jp/lrc/>」のRSSも購読したほうがよいだろう。

人文学にも情報学の影響が増していることを考えると、これら3つに加えて国立情報学研究所 (NII) も推しておきたい。

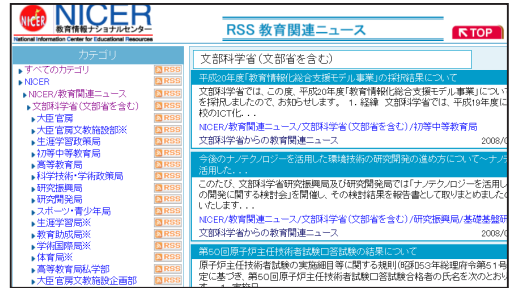
#### ● 国立情報学研究所 (NII)

<http://www.nii.ac.jp/>

ニュースとイベントの情報がRSSで配信される。なお、ニュースとイベントはそれぞれ別々のRSSが提供されている。また他の機関と同様、国立情報学研究所 (NII) のRSSもすべての情報を含んでいるわけではない。特に学術機関リポジトリ構築連携支援事業のような学術情報資源関係のサイトの更新情報は、別のRSSが提供されている。

### ❖ ニュース

ニュース関係では大規模なニュースサイトから小規



NICER

模なニュースクリップまで、幅広くRSSを購読しておこう。

#### ● サイエンスポータル

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

科学技術振興機構 (JST) が提供する科学技術情報ポータルのニュース記事とレビュー記事のRSSが配信されている。当然、話題は人文系に限定されない。むしろ科学技術に関する話題が多いが、文教・科学関係のニュースが幅広く扱われている。

#### ● 文部科学省からの教育関連ニュース

<http://www.nicer.go.jp/lom/program/search/bunkyouyousei.php>

#### ● RSS 教育関連ニュース

<http://www.nicer.go.jp/lom/program/nrss/reader.php>

文化庁を含む文部科学省の更新情報を購読できる。提供しているのは教育情報ナショナルセンター (NICER)。配信されるのはニュース記事ではなく、報道発表等の一次資料的なものを中心だが、文部科学省や文化庁のサイトではRSSが提供されていないだけに非常に有用。

#### ● トラカレ!

<http://torakare.com/>

「知と情報を繋ぐ人文系ニュースサイト」を謳い、人文系の書籍情報やサイト情報を掲載している。運営するのは『ウェブ炎上』や『ネットいじめ』等の著書を持つ荻上チキ氏。荻上氏の専門である批評に関するニュースが多く、人文学の現在を幅広く見渡せる。

#### ● 笠間書院

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

いまもっともウェブを活用している出版社のサイト。日本語・日本文学の研究書を専門とする笠間書院らしく、言語・文学関係の情報を集中的に掲載している。

日本語・日本文学の研究書を刊行する出版社、笠間書院のブログ。

## 笠間書院 kasamashoin online

〒101-0064 東京都千代田区猿樂町2-2-3 NSビル302 Tel.03-3295-1331 Fax.03-3295-1332

[Go to Top Page](#)

**新着記事 (最新 10本)**

- ◎ 2008奈良絵本・絵巻国際会議 大塚大会(2008年8月30日(土)～31日(日)、慶應大坂(ハートランドキャンパス))
- ◎ 古代文学会9月例会(第355回)近藤 信義 氏「歌人としての桓武」(2008年9月6日(土)、立正大学)
- ◎ 『国語と国文学』平成20年9月号(『東京大学国語国文学会』で小塚和明著『陀羅尼文字の書影』(作者・荒木浩氏))
- ◎ 笠間書院『9月刊行予定の本』
- ◎ 明治大学・日本古代文化における文学・図像・伝承と宗教的探究「企画展『国語と国文学』(2008年8月21日(木)～9月16日(水))
- ◎ 第202回 大塚大学古代中世文学研究会(2008年8月30日(土)午後2時～、法・経大学院総合研究棟(郵便局隣)3F L2教室)
- ◎ 国際日本文化研究センター・第43回学術講演会(平成20年9月18日(木)、日文研ホール)
- ◎ 斎宮シンポジウム「藤原・白河の宮」(斎宮)を語る(平成20年9月23日(祝・火曜日)、横浜歴史博物館 講堂)

★月部(15～30日)短評★  
2007の最新情報、新年賀詞

笠間書院

**CA Current Awareness Portal**  
図書館に関する情報ポータル

CA-R CA-E CA 調査研究レポート 研究ツール 雑誌新刊目録

カレントアウェアネス・ポータルは、図書館界、図書館情報学に関する最新の情報をお知らせする、国立国会図書館のサイトです。

検索の方法  
このサイトについて  
サイトの更新履歴 (2008-04-25)

**カレントアウェアネス-R 新着ニュース**

- 「情報リテラシーのロゴコンテスト受賞者が決定」(2008-08-22)
- 「ウラゲツ☆ブログ」の図書館員による活動が顕著に」(2008-08-22)
- 雑誌マックスブック協会、PLoSのOACI賞を協会賞に」(2008-08-22)
- 「情報管理」5月号(2008年)を公表」(2008-08-22)
- これからの図書館の在り方検討協力会会誌(第10号)の配布資料」(2008-08-22)
- IFLA 2009年1月にウェブサイトのデザインを立案」(2008-08-22)

カレントアウェアネス・ポータル

## ウラゲツ☆ブログ

urag.exblog.jp

**【公開版】今週の注目新刊(第10回(第159号))：08年8月10日**

グローバル体験とグローバルビジネス—仕事と生きがい  
ミハイ・チクセントミハイ(1934-)著 大森弘毅監訳  
世界思想社 08年6月 2,415円 461頁(312頁) ISBN 978-4-7907-1351-7

■販売元が又よし学生からビジネスリーダーまで必読の書！「ロー(読み)を体験することで、仕事は自己実現と社会貢献の場となりうる。幸福と成功への道はビジネスとは何か、豊富なデータと世界のトップ企業リーダー三九人へのインタビューから読み解く。

★チクセントミハイは、イタリア・フィレンツェ(現在はクアトロ・リノ)に生まれ、アメリカの心理学者、シカゴ大学心理学部・教育学部教授、カリフォルニア州クレアモント大学大学院教授などを歴任。全米教育アカデミー、全米レジャー・科学アカデミー会員。

ウラゲツ☆ブログ

特に学会や研究会等のイベント開催情報が充実している。

●SENTOKYO ブログ

<http://blog.goo.ne.jp/sentokyo>

●実業史研究情報センター・ブログ「情報の扉の、そのまた向こう」

<http://d.hatena.ne.jp/tobira/>

●レファレンスクラブ ブログ

<http://reference.blog.shinobi.jp/>

それぞれ専門図書館協議会、実業史研究情報センター(渋沢栄一記念財団)、日外アソシエーツが運営している。運営者の性格上、図書館関係のニュースが多いが、広く人文系全般に有用だろう。

## ❖ 学術資源情報

ウェブの普及によって日々新たな学術情報資源が登場している。最新の動向を把握するには、次の3ヶ所を活用しよう。

●カレントアウェアネス・ポータル

<http://current.ndl.go.jp/>

国立国会図書館による「図書館界、図書館情報学に関する最新の情報」。速報性を重視した「カレントアウェアネス-R」、解説に重きを置いている「カレントアウェアネス-E」や「カレントアウェアネス」等の更新情報を配信している。RSSは個別に購読することもできるし、カレントアウェアネス・ポータルの全コンテンツをまとめて購読することもできる。図書館と銘打ってはいるが、扱われている話題は広い。どの専門分野も「情報」と無縁ではいられなくなった以上、必読のRSSと言えるだろう。

●情報管理 Web ニュース (β版)

<http://www.johokanri.jp/news/>

科学技術振興機構(JST)の月刊誌「情報管理」の編集部によるニュースクリップ。記事の内容はカレントアウェアネス・ポータルに近いが、情報の整備・流通・活用をテーマとする雑誌が母体であるため、知的財産等の話題も扱われている。

●ACADEMIC RESOURCE GUIDE (ARG)

<http://d.hatena.ne.jp/arg/>

手前味噌と思われるだろうが、本稿筆者が発行するメールマガジン ACADEMIC RESOURCE GUIDE (ARG) のブログ版。インターネットの学術利用をテーマに新たに公開されたサイトや開催予定のイベント情報を掲載している。

## ❖ 新刊情報・古書情報

人文系にとって書籍は依然として重要なメディアの一つである。新刊・古書情報のRSSを紹介しよう。

●ビーケーワン (bk1)

<http://www.bk1.jp/>

あくまでオンライン書店のビーケーワン (bk1) が入荷したという限定はあるものの、日本語書籍の新刊情報(新入荷一覧)をRSSで配信している。

●ウラゲツ☆ブログ

<http://urag.exblog.jp/>

哲学や現代思想等の人文書の版元である月曜社のブログ。他社の刊行物を含め、幅広く注目の人文書を紹介している。

●スーパー源氏

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

スーパー源氏は古本・古書の販売・買取専門のオンライン古書店。サイト自体の更新情報もRSSで配信されているが、特筆すべきは商品目録の更新情報。条件を決めてRSSを購読しておけば、探している書籍が入荷次第、通知してくれる。

❖ 個人ブログ

最後に個人によるブログのうち、今後の人文学にとって重要なテーマを扱っているものをいくつか挙げておこう。

●Daily Searchivist

<http://d.hatena.ne.jp/searchivist/>

アーカイブズ学を専攻する坂口貴弘氏のブログ。記録の管理と保存のあり方は焦眉のテーマであり、その重要性は分野を問わない。

●Copy & Copyright Diary

<http://d.hatena.ne.jp/copyright/>

末廣恒夫氏による著作権をテーマにしたブログ。非論理的な主張も入り乱れる著作権論議にあって、非常に論理明快な解説・主張を続けている。

●やまもも書齋記

<http://yamamomo.asablo.jp/blog/>

人文情報学の話題を中心にした當山日出夫氏のブログ。分野に関わらず、「情報」との関わりを考えるきっかけを与えてくれるだろう。

❖ 人文系 RSS の課題

以上、RSSリーダーでの購読を勧めたい20のサイトとブログを紹介してきたが、最後に一言だけ人文系RSSの課題を述べておきたい。

冒頭でふれたように、伸長しているもののRSSリーダーの認知・利用は決して高いものではない。この原因の一つはRSSを配信していないサイトの多さにあるだろう。日本国内には人文学を掲げる研究所が多数あるが、そのほとんどはサイトが貧弱でRSSの配信



Daily Searchivist



MAFFIN News Feeds Center

もしていない。まずはこの状況を解決する努力を関係者に求めたい。

その上で主要なRSSをひとまとめに購読できるようなサービスの登場を期待したい。たとえば、農林水産学の分野では、農林水産研究情報センターが、

●MAFFIN News Feeds Center

<http://www.afrc.go.jp/ja/rss/>

を提供している<sup>[2]</sup>。

ここでは農林水産系の各試験研究機関のサイト更新情報を一覧でき、さらにRSSも配信されている。人文学においても、このような取り組みの登場が大いに待たれるところである。

注

[1] 「RSS 定期リサーチ (14)」(japan.internet.com 編集部、2007年4月6日)

<http://japan.internet.com/research/20070406/1.html>

[2] 林賢紀、宮坂和孝「RSS (RDF Site Summary) を活用した新たな図書館サービスの展開——OPAC2.0へ向けて」(「情報管理」49-1、2006年1月)

[http://www.jstage.jst.go.jp/article/johokanri/49/1/49\\_11/\\_article-char/ja](http://www.jstage.jst.go.jp/article/johokanri/49/1/49_11/_article-char/ja)

# 中国哲学

齊藤 正高

今回のリソースレビューは中国語圏の経学・諸子学を中心とした紹介から趣向をかえ、アメリカの電子図書館における漢籍公開状況を紹介します。中国哲学と関連の深い中国医学、比較的特殊な分野である明清天主教(カソリック)文献の公開状況を紹介してみたい。

## ●プロジェクト・グーテンベルク

<http://www.gutenberg.org/catalog/>

プロジェクト・グーテンベルク(以下PG)は1971年、合衆国イリノイ大学の学生であったマイケル・ハートがはじめた電子図書館計画<sup>[1]</sup>である。

1971年といえば、WindowsやMS-DOSは言うに及ばず、NECのPC-8001も発売されていない。当時画期的であった「マイコン」TK-80が発売されたのが1976年であり、個人所有のコンピュータ「パソコン」が出現する以前の話である。

ハート氏は当時非常に高価であったXeroxの汎用コンピュータを使用できるアカウントを得て、何か役立つことをしようと、『アメリカ独立宣言』を入力した。これがこのプロジェクトにおける最初の電子テキストとなったそうである。

発足当初から現在にいたるまで、PGは一貫してボランティアによって運営され、著作権保護期間が満了となったパブリック・ドメイン<sup>[2]</sup>のテキストを公開している。1994年、ローマのピエトロ・ディ・ミケーリ氏によって最初のウェブページが作られ、現在はノース・カロライナ大学のサーバーで公開されている。冒頭にあげたURLを参照すると、50種以上の言語

## プロジェクト・グーテンベルク



で書かれた約2万5千冊を公開しており、オーストラリア・ヨーロッパ・ドイツ・フィリピンなどの提携プロジェクトをあわせれば、電子テキストの総数は10万冊にのぼる。言語別にみれば、2008年8月末の時点で、英語22,492冊、フランス語1,225冊、ドイツ語542冊であり、これらが上位3位を占める。ちなみに、日本語で公開されている電子テキストは、松尾芭蕉『奥の細道』・『マルチン・ルターの小信仰問答書』・芥川龍之介『羅生門』の3冊のみである。

ところで、この最古参の電子図書館で近年、中国語電子テキストの公開が活発になってきた。大学の演習でよく読まれている顧炎武『日知録』38巻も2008年4月に簡体字で公開されている。

現在、カタログに登録されている中国語電子テキストの総数は320冊<sup>[3]</sup>であり、いずれもUTF-8・BIG-5の標点符号つきプレーンテキストとしてダウンロードができる。また、ジェイムズ・レグ(1815-1897)の『論語』注釈など、英文による中国関連書が25冊ほど公開されている。

このサイトで公開されている中国語文献は、明清白話小説など文学に分類されているテキストが約半数を占める。これらをのぞけば、まず、十三経・先秦諸子などの基礎的文献が公開されている。しかし、十三経においても『尚書』『儀礼』『公羊傳』『穀梁伝』を欠いており、組織的電子化が行われているわけではない。そのかわりに、おそらくボランティアの個人的関心によって、比較的珍しいテキストが公開されており、とくに、ほかの無料データベースには少ない宋以後の文献にこの特徴がみとめられる。

たとえば、孫光憲『北夢瑣言』、歐陽修『帰田録』、蘇洵『蘇洵集』、蘇軾『東坡志林』、劉基『鬱離子』、王守仁『王陽明全集』、呂坤『呻吟語』、黃宗羲『明夷待訪録』、王士禎『池北偶談』などが公開されており、孟元老『東京夢華録』、徐霞客『徐霞客遊記』などの風俗地理書や、馬欽『瀛涯勝覽』、費信『星槎勝覽』などの明代の外国地理書、また、沈括『夢溪筆談』、李杲(金代の医家)『脾胃論』、宋応星『天工開物』など科学技術に関係する著作、明清交替期の蜀の唐毅を記録した彭遵泗『蜀碧』なども公開されている(目録

では shu bi と記述)。

唐以前のテキストは、すでに台湾中央研究院などで電子化され、無料公開されているテキストが大部分を占めるが、『説苑』『新序』といった劉向の著作が比較的充実しており、『韓詩外伝』『孔叢子』などのテキストも公開されている。

一般に漢籍の電子化には、異体字の問題・校正の問題・標点符号の著作権法上の問題<sup>[4]</sup>など、多くの問題があるが、これらはPGのデータにもいえる。とくに(一部をのぞき)版本情報の欠如している点は、テキストデータを研究に用いる場合には問題であり、紙媒体で確認をとる必要があるが、あらしらべの道具としては使用できるであろう。また、PGのテキストデータにはプロデューサーの名前が明記してあり責任の所在が明記されているが、他サイトのコピーが混入する可能性も否定できない。

とはいえ、さまざまな個人プレーをまとめ、特徴のある文献を敷居の低い形で万人に公開している点は、異文化に対するアクセスの確保として評価できるだろう。

なお、PGではオーディオブックも公開されている。音声データはhuman readingとcomputer generatedに分かれ、mp3などの形式でダウンロードできる。英語の音声を中心であり、462冊のhuman readingのパフォーマンスが公開されているが、例外的に中国語のパフォーマンスもある。『唐三百家詩』巻1(Three Hundred Tang Poems, Volume 1)がhuman readingで公開されているのである。

英訳漢籍のオーディオブックには『孫子』(The art of war)、『老子』(The Tao Teh King レッグ訳)が公開されている。

これらのデータは、視覚障害者による中国文化へのアクセスを確保するだけでなく、語学教育の教材としても活用できるのではないだろうか。

#### ●中国医学——中医世家

<http://www.zysj.com.cn/index.html>

2006年に立ち上げられた西安在住の王氏のサイトである。氏は上のサイトで父、王提仁氏を紹介しており、先祖は中国近代四大名医の一人に数えられる王紹業であるとしている。

王氏は97年ごろから中国医学のデータベースを作りはじめ、個人的に使用していたが、2006年友人のすすめによりウェブサイトを持ち上げたそうである。



また、サイト上では薬剤を売買せず、いかなる診療所も宣伝しないというポリシーを掲げ、中国医学の学習および資料の収集をサイトの目的としている。

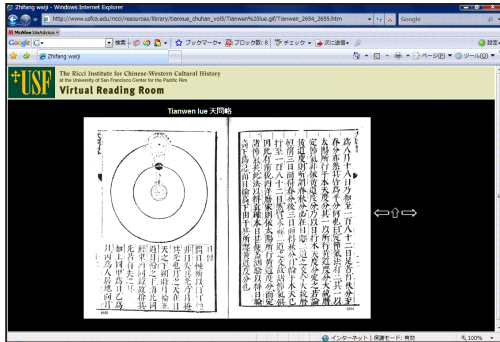
このサイトで紹介されている中国医学の書物は350冊と非常に多い。「中医書籍」のページをみると医学書は16分野に分類されており、『黄帝内经素問』『黄帝内经灵枢』『黄帝内经太素』『難経』『類経』などの中国医学原論(経論)、『傷寒論』『金匱要略』などの流行病救急医学(傷寒・金匱)、『脈経』などの診断法(診治)、『雷公炮炙論』『本草綱目』などの薬学(本草)をはじめ、内科・外科・処方・産婦人科・小児科・眼科及び耳鼻咽喉科(五官)・鍼灸科などの書物が、簡体字のテキストデータとして公開されている。

テキストは章ごとに分割されたHTMLファイルで、基本的にブラウザをつかい読むように作られている。文中の分かりづらい薬名などにはリンクが施しており、ほかの文献の解説を参照できるテキストもある。

「最も古い医学概論」<sup>[5]</sup>である『難経』(『黄帝八十一難経』)を参照すると、81の問答をすべて掲載しており、一難~二十二難までを「論脈」とし、以下、「論経絡」(23~29)、「論臟腑」(30~47)、「論病」(48~61)、「論穴道」(62~68)、「論針法」(69~81)と、問答の主題によって整理が加えられている。

『本草綱目』を参照してみると、公開されているテキストデータは現代語による部分訳のようである。手もとのテキスト、『本草綱目』(中国書店1988年、底本光緒11年重校本)で、「水部」の冒頭を比較してみると、以下のようになる。

露水【釈名】露者陰氣之液也。夜氣著物而潤沢於道傍。(露とは陰気の液である。夜気が物について、道端を潤すのである。)



リッチ・インスティテュート

(中国書店本 巻5 水部・第3条)

露水【釈名】在秋露重的时候，早晨去花草間  
 收取。(秋の露が多いとき、朝草むらにゆき採  
 取する。)

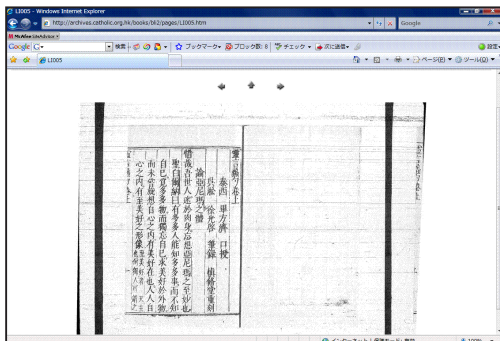
(中医世家 水部・第1条)

中医世家の【釈名】の文章は、本来【主治】の項  
 目にある「秋露繁時、以槃收取」(秋の露が多いとき、  
 皿でとる)を分かりやすく現代中国語に訳したもので  
 であろう。王氏のよったテキストがどのようなものかは  
 明記されていないが、とにかく『本草綱目』を整理し、  
 現代語訳したものであることはまちがいないであろう。

ところで、中国哲学では、医学書から当時の身体観  
 や自然観を探る分野があるが、基本的には医学書を歴  
 史史料として扱い、診断法や治療法が現在も妥当であ  
 るかどうかについては、実験を行い検証するわけでは  
 ない。中国医学の情報発信についても、このサイトの  
 ように現代にアレンジして発信していることは、やは  
 り、中国医学が現代もいきている医学であることを感  
 じる。

ともあれ、このサイトの中国医学書のデータベース

## 香港天主教区



は非常に多く、『難経』のように原文とつきあわせても、  
 一致するものも多い。中国医学や中国医学史を学ぶ上  
 でも役に立つサイトであることはまちがいない。

なお、「中医世家」では、「中華人民共和国薬品管理  
 法」や、2008年5月12日に起こった四川大地震の  
 際に作られた災害時の疾病対応マニュアル(「地震災  
 後常見病多発病中医薬治療手冊」)など、現代の中国  
 医学の文献も公開されている。

## ●天主教文献1——リッチ・インスティテュート

[http://www.usfca.edu/ricci/resources/library/  
 online\\_reading.htm](http://www.usfca.edu/ricci/resources/library/online_reading.htm)

サンフランシスコ大学の The Ricci Institute は、1984  
 年エドワード・J・マラテスタ神父(1932-1998)が、  
 中国に伝道したイエズス会士、マテオ・リッチ(1552  
 -1610)を記念して設立した研究機関である。

このサイトは、マテオ・リッチのほかに、ジュリオ・  
 アレニ(艾儒略)、サバティエーノ・ド・ウルシス(熊三拔)  
 など、イエズス会士の中国語著作を画像データで公開  
 している。

研究所サイトの online reading room を見ると、  
 2005年に更新された目録がでてくるが、天主教文献  
 の叢書である『天学初函』の大部分が画像で公開され  
 ている。内容は以下のとおり。

- 『西学凡』『景教流行中国碑頌』
- 『畸人十篇』『交友論』
- 『二十五言』『天主実義』
- 『職方外紀』『泰西水法』
- 『渾蓋通憲図説』『幾何原本』
- 『表度説』『天問略』『簡平儀』

以上の著作は神学・地理学・水利技術・天文学・測  
 量学・数学などの著作をふくむ。ただし、イエズス  
 会士の伝えた学術は、ガリレオ式望遠鏡を紹介した  
 『遠鏡説』などをのぞき、概ね西洋の古典学術であり、  
 17世紀のいわゆる科学革命の成果を紹介したもので  
 ない。だが、明末清初思想・科学技術史・中西文化  
 交流・現代学術用語の形成などの分野で貴重な史料で  
 ある。

また、同サイトでは、『天主教東伝文献三編』に含  
 まれる以下の著作も公開されている。

- 『速友篇』『正学鏐石』『聖教信証』



『五十言余』『勤修一鑑』

さらに、康熙 59 年 11 月 18 日の日付のある論旨「論西洋人」も公開されており、朱筆による補正も見られる。

●天主教文献 2——香港天主教区

<http://archives.catholic.org.hk/books/index.htm>

1841 年設立されたカソリック香港教区の歴史文献を公開しているサイトである。設立当初の歴史文献や、日本占領期（1941-1945）をふくむ 1841～1969 年までの史料目録などが公開されている。

コンテンツとしては約 80 種の明清天主教文献が画像で公開されているのが特徴である。マテオ・リッチやジュリオ・アレニの著作のほかに、アダム・シャル（湯若望）、フェルビースト（南懷仁）、アルフォンソ・ヴァニョーニ（高一志）、エマニュエル・ディアス（陽瑪諾）、ニコラス・ロンゴヴァルディ（龍華民）など、来華イエズス会士の著作のほか、徐光啓や楊廷筠など奉教士人の著作もある。

一般的に布教に関わる宗教書が中心であり、ミサや懺悔の方法を述べた著作が目をひくが、アダム・シャルの『主制群徴』は、その一部でガレノス医学を紹介しており、また、フランシスコ・サンビアシ（畢方濟）が口述し、徐光啓が書き起こした『靈言蠡勺』は亜尼瑪（アニマ：靈魂）を論じたものとして大変興味ぶかいものである。

注

- [1] Gutenberg:The History and Philosophy of Project Gutenberg by Michael Hart (<http://www.gutenberg.org/wiki/>)
- [2] ヒトゲノム計画の解析結果なども公開している。
- [3] Advanced Search で言語別に絞り込んだ結果を用いた。なお、『近思録』が 2 種あるなど、重複も存在する。
- [4] 「東洋学研究と著作権」本誌 6 号。2005 年を参照。
- [5] 石田秀実『中国医学思想史』。東京大学出版会。1992 年を参照。

## 中国史

佐藤 仁史

### ❖ 進展する学術情報の電子化

本誌第 6 号において本欄を担当して以来 3 年が経過した。その間、研究情報の収集において電子化された学術情報の利用はごく普通のこととして行われるようになった。例えば、CNKI や Cinii、JSTOR などのサイトにおける学術論文の検索や本文のダウンロードは日常的に行われている作業であろう。また、CNKI や台湾国家図書館「全国博碩士論文资讯网」で最新の博士論文を確認することも多いであろう<sup>[1]</sup>。電子化された学術情報の充実に伴い、研究活動においてそれらに常に目を配り、その内容を吟味する能力は不可欠となろう。以下では、筆者の専門領域である中国近現代史に関わるサイトの動向や取り組みについて、主に一次史料に着目しつつ紹介したい。

### ❖ 図書電子化プロジェクトにおける研究文献

●中英文図書数字化国際合作計画（CADAL）

<http://www.cadal.zju.edu.cn/>

第 10 期 5 カ年計画期における「211 工程」（国家の経済建設や社会発展の基盤を構築するために、一部の大学と学科において優秀な人材を育成するプロジェクト。主要内容に、基盤 100 校の選出と重点学科の形成、高等教育公共サービスシステムの構築などがある）の一環として開設されたサイトである。CADAL の構築には浙江大学と中国科学院研究生院という 2 つの「技術中心」と浙江大学をはじめとする全国 14 大学の「資源中心」が参与したという。デジタル化された文献は 100 万冊にものぼり、うち半分は英文文献によって、残りは中文文献によって構成されている。

検索画面には検索キーワードを入力するテキスト



CADAL

ボックスがあり、その下部には「古籍」「民国図書」「民国期刊」「現代図書」「学位論文」「絵画」「視頻」「英文」という分類に即したチェックボックスが配置されている。「古籍」では様々な文集、地方志などの閲覧が可能である。それ以外については、当プロジェクト参加大学・研究機関のネットワーク内で閲覧する必要がある。たとえ、全文をみることができなくても、「民国期刊」ではキーワード検索によって、電子化されている雑誌の書誌情報を得ることができるので便利である。

なお、日本国立公文書館「近アジア歴史資料センター」と同様に、画像形式としてDjVuを採用しており、全文画像の閲覧にはブラウザプラグインをインストー

中国国家図書館「民国期刊」



ルする必要がある。

● 超星数字図書館

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

デジタル図書館の「老舗」として有名な超星数字図書館は一時期運営方針がぶれていた印象があるが、現在では安定しているようである。例えば、一時期閲覧が不可能であった地方志や文史資料も現在では一部閲覧が可能である。旧地方志は中国地方志集成所収のものが、文史史料は1980年代後半以降のものが公開されている。江蘇省と浙江省、上海市の新編地方志は一部の鎮志をのぞいて公開されていない。かつて公開されていた内容との差異が少なくないように思われる。

現在では多くの附加的機能も設けられているが、デジタル図書館としての主要機能は「会員図書館」「免費閲覧室」の2つである。文献の閲覧やダウンロードといった前者のサービスを利用するには閲覧カードを購入する必要がある。後者はその名が示すとおり閲覧カードを購入していないユーザも閲覧可能な書籍が置かれている。有用な研究書も数多く含まれていて便利である。

❖ 国家機関における一次史料

● 中国国家図書館「民国期刊」

<http://www.nlc.gov.cn/>

中国国家図書館による近年のデジタルアーカイブの充実は目をみはるものがある。貴重な一次史料群は「館蔵珍品」の欄において閲覧をすることができる。内訳は甲骨文や碑文の拓本などから民国期の雑誌に至るまで多岐にわたり、近現代史研究者にとって有用なのは「民国期刊」「民国図書」「民国法律」であろう。「民国期刊」は館蔵のマイクロフィルム資料のうち、3,337種の中文雑誌をデジタル化した本格的な一次史料データベースである。他の分野に比して進んでいるとは言い難い中国近現代史における史料のデジタル化も実用に耐えるレベルに達してきたといえよう。本稿執筆時点においては2,756種が公開されている。閲覧に際しては「NLCReader」という独自のビューワーをインストールする必要がある。

「民国図書」は当該図書館の「民国図書資源庫」が所蔵する民国期の図書6,229種、6,453冊のデジタル画像を公開したものである。書名や著者、分野、主題などに即して検索をすることができる。「民国法律」は、

辛亥革命直後の中華民国臨時政府を初めとする諸政権、北洋政府期、南京政府期、満洲国、汪精衛政権、中国共産党の革命根拠地政府といった様々な政権の法律関連文献のデジタル画像データベースであり、4,568篇、16,650頁におよぶ。当時の制度を確認するのに至極便利である。“民国図書”“民国法律”ともに閲覧にはAdobe Readerが必要である。

上記以外にも注目すべきコンテンツがある。例えば、「館蔵珍品」の欄には年画の画像コレクションが公開されている。年画は民間文化を考える際に重要な非文字史料の1つであるが、漫然とながめるのも楽しい。なお、当該サイトには「近代文献聯合目録」という検索システムがあり、国会図書館と上海図書館が所蔵する1900年～1949年の文献を横断検索することができる。

#### ●台湾中央研究院“漢籍電子文獻”

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

中国史研究における研究情報デジタル化の老舗的存在である台湾中央研究院の取り組みは着実に進展している。中国近現代史に深く関連するデータベースには「史語所蔵内閣大庫檔案」「近史所檔案館藏函目彙編查詢系統」などがあるが、既に広く知られているものである。以下では最近の変更点について紹介する。最も有名な「漢籍電子文獻」「漢籍全文文獻資料庫」に改良が加えられ、インターフェイスも装いを新たにした。異体字や同義語を含めた検索や付箋・メモ機能が備わっており、テキストの徹底的な検索・読解の便に供している。しかしながら、日本国内では利用することはできず、利用には中央研究院に直接赴く必要がある。「漢籍電子文獻」の“台湾文獻叢刊”や“中華民國史事日誌”、“清代經世文編”は日本からでもアクセス可能である。

#### ●日本国会図書館“近代デジタルライブラリー”

<http://kindai.ndl.go.jp/>

日本の国家級機関におけるデジタルアーカイブの取り組みとして本サイトが挙げられる。国会図書館所蔵図書のうち、著作権の失効したものを順次JPEG画像としてデジタル化しているものである。コンテンツも充実しており、例えば、“歴史・地誌”だけでも12,206件にものぼる。画像閲覧はJPEGとJPEG2000のいずれかを選択することができるが、後者の場合プラグイン(JuGeMuPlayer ver4.5)をインストールする必要がある。画像を保存する場合、PDF

ファイルに出力できるが、一度の操作につき10頁分に限定されている点は改善を期待したい。

#### ●国立公文書館“アジア歴史資料センター”

<http://www.jacar.go.jp/>

国立公文書館、外務省外交史料館、防衛省防衛研究所所蔵の公文書をデジタル化して公開している国立公文書館の「アジア歴史資料センター」も、前回の本誌における紹介以降も堅実に作業を進展させている。日本発信の歴史史料デジタルアーカイブの中でも不動の地位を占めている。

ところで、国家機関におけるデジタルアーカイブの整備が着実に進展しているのに対して、その他の機関、例えば各大学が所蔵するアジア歴史資料の整理・デジタル化・発信の必要性があることを筆者は言及したことがある。当該サイトには、「日本国内所在の主要アジア歴史資料に関する調査報告書」平成18年度調査分の最新バージョン(Ver1.2)が掲載されており、全国の概況がわかり便利である<sup>[2]</sup>。東亜同文書院の大旅行誌は言うまでもなく、旧商科大学・旧高等商業学校学生の旅行誌や調査報告書などは貴重な一次史料群である。しかしながら、それらは旧商大・旧高商の流れを汲む大学の図書館に必ずしも十分に活用されずに眠っているのが現状である。日本からの学術情報の発信という観点からもこれらの史料のデジタル化は急務であろう。

### ❖ 地方文献情報

近年では市・県級の所蔵機関における地方文献の電子化・公開が進展しつつある。以下では筆者がフィールドとする江南地方の状況を紹介する。

#### ●呉江市図書館

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

「古籍閲覧」において、『垂虹識小録』など呉江県の旧県志10種と、『民国同里郷土志』など郷鎮志や郷土志14種が公開されている<sup>[3]</sup>。

#### ●常熟図書館

<http://www.cslib.cn/>

トップページから「徐兆璋日記大型清史工程」に入ると、『徐兆璋日記』の総称される日記群のうち、光緒20年(1894)、光緒24年(1898)、光緒25年(1899)の3年分の全文テキストが公開されている<sup>[4]</sup>。徐兆璋(1867-1940)は清末の進士で、日本考察に赴い



徐兆璋日記

たこともあった。民国初に衆議院議員に選出されるも曹錕の賄選に反対して帰郷して後は、郷土の公共事業に尽力した。その内容は学校や常熟図書館の建設、水利工程の推進、『重修常昭合志』の編集など多岐にわたる。『徐兆璋日記』は光緒20年2月12日(1894年3月18日)年から民国29年6月12日(1940年7月16日)の間、途中の中断時期を除いて42年間にもおよぶ。その経歴を考えると、中央政治史の領域にとっても、地域社会史の領域にとっても稀有な価値を有する貴重な史料である。聞くところによれば、近年進められている清史の編纂にも一部が供されており、影印本も出版の予定であるという。『徐兆璋日記』の公開がさまざまな領域における研究の進展に寄与することを願う。

なお、「常熟古籍」の「常熟地方文献」では、『重修常昭合志』と『海虞文徵』の2種が画像形式で公開されている。

●湖州檔案信息网

<http://huzda.zj001.net/>

檔案目録情報の公開では、檔案の閲覧とリンクしたサービスを提供している上海市檔案館が群を抜いており、少なからぬ外国人研究者がその恩恵にあずかっている。市・県檔案館でも檔案目録情報を公開する所が現れており<sup>[5]</sup>、湖州市檔案局・檔案館による当サイトもそのひとつである。どの程度の歴史檔案が目録化・オンライン化されているのか現段階では定かではないが、実際に検索してみたところ、興味深い内容の歴史檔案もヒットした<sup>[6]</sup>。オンライン目録検索によって実際に当該檔案館に赴く前に具体的な当たりをつけておけることは便利である。また、現地に赴いたならば、配架されている目録とオンライン目録の異同を調べてみる必要があろう。

●嘉興市図書館

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

地方文献閲覧室の充実度は近隣の市級図書館の中では屈指であるが、サイトにおいても「文史資料庫」や「歴史文化名城」という欄を設けて地方文献情報や当地民俗の概況を紹介している。「文史資料庫」の中にある“地方文献”“嘉興著作目録”では歴代の著作の書誌情報が掲載されている。「歴史文化名城」の中にある“民俗民間文化”では蚕桑に関する習俗や年中行事など嘉興の民俗の概況が紹介されており、フィールドワークをはじめとする現地調査に赴く前に現地の特徴を把握しておくのに有用である。

◇ 研究情報一般

●愛問 (iask) 「共享資料」

<http://ishare.iask.sina.com.cn/>

愛問 (iask) は百度や搜狗などと並んで著名なポータルサイトである。Google を相当意識していることはサイトのデザインやサービス内容から容易に看取できる。とはいえ、当該サイト独自のサービスも行っている。「共享資料」はその一つであり、文字通りファイルを共有できるようにしたものである。トップページに行くと、「教育資料」「娯楽・生活」「辦公文書」「專業資料」などの6カテゴリーに分類されている。そのうち、「教育資料」内の“学术研究”や“專業資料”内の“教育、出版”などには、有志が入力した様々な史料や研究文献などがアップロードされている。

各種史料のダウンロードには会員登録が前提となっている。実際にダウンロードする際には資料によって異なるポイントが必要になるが(ポイントの要らないものも多い)、ポイントは当該サイトへのアクセスや自らの電子資料のアップロードなど「参与」の度合いによって蓄積される仕組みになっている。

しかし、問題点がないわけではない。個人で作成したファイルを交換しているだけであれば問題はないが、例えば筆者がスキャナで取り込まれた『費孝通全集』全巻の pdf ファイルを見つけたように、著作権に抵触したのも少なくないようである。サイトの利用に際してはこの点に留意する必要がある。

## ❖ 結びにかえて

以上、例示的にいくつかのサイトにおける学術情報の現状について紹介した。本稿で紹介したウェブサイトの史料のみでも本格的な研究の利用に耐えうるコンテンツが充実していることが実感できよう。今後コンテンツの充実化が確実に進展していくと、近現代史研究においてもオンライン上に溢れる史料や研究情報をどのように効率的に処理していくかが早晚問題として浮上してこよう。このような状況に対応するには、有用な情報を選別する問題意識を常にとぎすますという極めて常識的な方法に勝るものはないように思われる。

## 注

- [1] 台湾国家図書館「全国博碩論文資訊網」(<http://etds.ncl.edu.tw/theabs/index.jsp>)

- [2] 「日本国内所在の主要アジア歴史資料（平成 18 年度調査分 報告書 Ver.1.2）」2008 年 4 月 16 日 ([http://www.jacar.go.jp/houkoku/H18\\_report\\_honbun1\\_2.pdf](http://www.jacar.go.jp/houkoku/H18_report_honbun1_2.pdf))
- [3] なお、これら 24 種の地方志の電子データを収録した『縦覧呉江——呉江 500 年古代地方志匯編』（江蘇電子音像出版社）も刊行されている。
- [4] 常熟檔案館所蔵の地方文献については、佐藤仁史「近代中国の地方文献についての覚書」アジア研究情報ゲートウェイ (<http://asj.ioc.u-tokyo.ac.jp/html/037.html>) も参照されたい。
- [5] 例えば、青浦区檔案局・館による青浦檔案信息 (<http://files.shqp.gov.cn/>) においても目録情報の検索が可能である。
- [6] 土地改革に関する檔案を検索すると、「关于处理山区土地改革中山林分配遗留问题和群众共有山林权益问题通知」という共有林の分配に関する興味深い檔案の情報を得ることができた。

# 大蔵経データベース

石井 公成

## ❖ SAT（大蔵経テキストデータベース研究会）

### ■ 大正蔵大蔵経電子化の完成

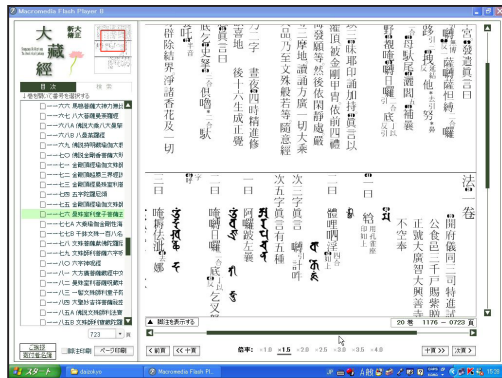
2007 年 7 月 30 日に、「大蔵経データベース化完成記念大会」が、SAT（大蔵経テキストデータベース研究会。代表は下田正弘東京大学教授）と大蔵経データベース化支援募金会（事務局長は奈良康明駒澤大学元学長）の共催によって開催された。大正新脩大蔵経の電子化プロジェクトが 10 数年の困難な過程を経て、ようやく完成したことを祝い、その便利さと意義を、諸研究者による講演とシンポジウムによって様々な面から明らかにしようとする会である。

会場では、2 枚の CD に収められた大正蔵のテキストデータベースが配布されたほか、後日、高額の寄付者に対して、報告書と CD が郵送された。SAT では 1998 年に大正蔵の一部をインターネット公開して以来、大正蔵の入力と公開を進めてきており、数年前か

ら 85 巻全体のデータをダウンロードできるようにしてきたが、今回は校正を重ね、校異も入力された。

悉曇を含めて大正蔵の外字がすべて埋まったデータが公開されたのは、今回が世界で初めてであろう。外字表記については、文字鏡番号方式から、GT 番号方式に変更され、GT がない外字については、SAT の独自番号が新たに付けられた。GT に無い外字のフォントは、コンピュータ学校の学生など、10 数人のスタッフが大正蔵の字形に基づいて作成した。悉曇については、大正大学の密教学の院生など若手の専門家たちが、美しく正確なフォントを作成してくれた。

配布された CD では、Flash を採用した SATViewer と称するプログラムによって、大正蔵どおりに表示し印刷することが可能となっている。つまり、3 段組による縦書きで、割注も本文どおりの形で表示し、図も画像で入り、校異も表示できるようになったのである。このため、ノートパソコンにインストールすれば、横書きのテキストファイルでなく、大正蔵どおりの画面を世界中のどこでも眺められる環境が整ったことにな



SATViewer

る。

表示は1倍から0.5倍きざみで4倍まで拡大できるようにしており、上の図は、1.5倍のものである。拡大すると頁の一部しか表示されなくなり、どの辺りなのか分かりにくくなるため、画面左上に頁の全体を示し、現在はその頁のどの部分を表示しているかを赤枠で表示するとともに、その赤枠にカーソルを重ねて動かすことによって、「頁の下段の真ん中やや左側に移動する」といった操作が簡単にできるようになっている。

## ■ 超高速検索サイトの公開

CD版では、ごく簡単な検索機能しかついていなかったが、2008年4月にはSATの新たな検索サイト (<http://21dzk.l.u-tokyo.ac.jp/SAT/>) が公開され、大正蔵85巻を超高速で検索することが可能になった。検索は、句点抜きと句点付きの検索を選択できるようになっている。アクセスが集中する時間帯などでなければ、大正蔵全体を検索しているにもかかわらず、検索

## SATの検索サイト



結果は1～2秒で表示され、その速さは感動的ですらある。

このため、筆者は、著者に関して盛んな論争がなされている三経義疏を大学院の講義でとりあげるにあたり、教場を変更してPC教場で行うこととし、問題のありそうな語句をすべてSAT検索ページで検索しながら読解を進めている。特にNGSMと組み合わせて検索を行うと、きわめて効果的であり、多くの発見をすることができた<sup>[1]</sup>。ただ、そのためには、仏教漢文や変則漢文の語法に関する基礎知識、大正蔵における漢字や句点などの特徴、研究における問題意識、そして検索する際の多少のコツが必要となる。また、それぞれの文献を実際に読み、その特徴を理解していないと、踏み込んだ検索ができないことは、言うまでもない。

永崎研宜氏が開発した独自の検索システム<sup>[2]</sup>は、新字・旧字・異体字を自動的にカバーしているほか、従来のgrep検索では、該当文字列を含む1行だけが検索対象となることから、句点の切り方の誤りなどによって検索が不十分になってしまうことを考慮し、1行だけに限らず、ひとまとまりの範囲の中で該当文字列が何度登場するかを検索し、その回数を表示するよう改められた。その結果、より完全な検索がなされるようになったのである。結果表示については、該当文字列がKWIC方式で用例の中央部分に赤字で示されるようになったため、見た目にもわかりやすく、また文脈の把握も簡単になった。むしろ、検索結果から該当する本文の箇所へ飛ぶこともできる。

本文に飛ぶと、INBUDS(印度学仏教学論文データベース)にリンクが張られているだけでなく、A.C.Muller氏が主催する英文のオンライン漢字辞書・仏教辞書とも連動している。このため、本文のうちの必要な部分を範囲指定すると、そこに含まれている漢字・術語・句のうち、Muller氏の辞書に収録されているものリストが画面右下に自動的に表示され、それをクリックすることにより、それらの漢字・術語・句にたいする諸国語での発音や英語による語義説明などを見ることができる。

このため、SATのこの検索サイトは、英語圏の研究者にとっては既に不可欠な補助ツールとなっており、アメリカその他の多くの大学の学科などが利用契約を結んでいるという。むしろ、日本の研究者にとっても、この辞典連動の機能は便利であって役に立つ場合が多

い。

この検索サイトは使い方が簡単であるうえ、きわめて高速であるため、大いに歓迎されるだろう。今後は、仏教学だけでなく、歴史学、文学、美術史その他の多くの分野の研究者が、これまで以上に研究対象の仏教面に注意するようになることが期待される。SATでは、これからの方向としては、木版本や写本の画像とのリンクや、現代語訳とのリンクを計画しているため、この検索サイトは、今後さらに便利になっていくものと思われる。

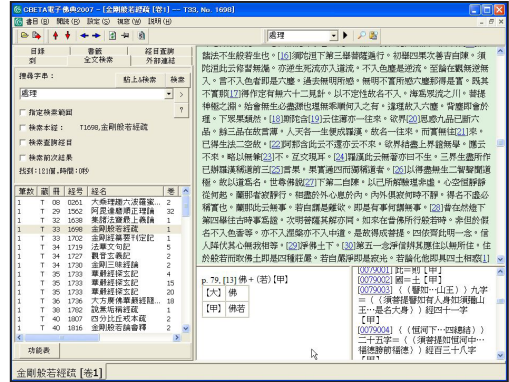
なお、CD版では『一切経音義』は外字表記が不十分であったうえ、最初の10巻以後は画像表示となっていたが、検索サイト版ではすべての外字が表記されるようになった。ただ、大正蔵の『一切経音義』は印刷が悪く、不鮮明な部分や明らかな間違いもあるため、字形を定められず、■のまま残した場合もかなりある。研究のためのテキストとしては、版本や写本に基づいて新たに校訂しなおした電子テキストを作成する必要がある。

◆ CBETA (中華電子仏典協会)

■ 卍新纂続蔵経の電子化

CBETAでは、日本撰述部を除く大正蔵の入力が完成すると、『卍新纂続蔵経』88巻の入力に取り組み、2007年3月に完成させた。大正蔵の分と合わせると、総漢字数は1億5千万字を超える。その成果は、『CBETA 電子仏典集成 2007』と題するCDに収録されて配布されるとともに、CBETAのサイト (<http://www.cbeta.org/index.htm>) でも公開されている。CBETAでは、新しいバージョンが出るたびに大々的に配布しており、世界中のかかなりの数の研究者に利用されているものと思われる。日本語 Windows でも、むろん利用できる。

『CBETA 電子仏典集成 2007』では、CBETA Reader というプログラムによって、多国語環境での多様な処理を可能にしている。中国語繁体字・簡体字、英語、日本語による説明があり、丁保福『仏教学大辞典』『南山律学辞典』ほか複数の仏教辞典や漢字辞典とリンクしている。校異も画面右下に表示されるばかりでなく、特定の校異を指定すれば、「宋本」「元本」「明本」などの違いが、対照表の形で一覧表示されており、わかりやすい。

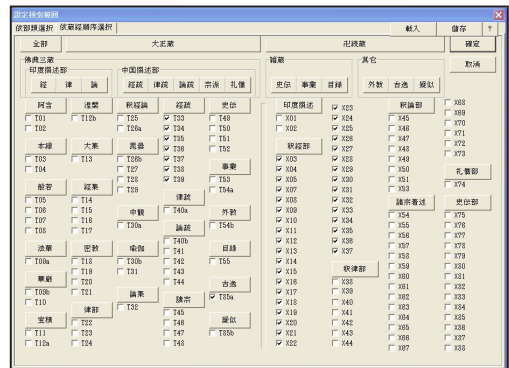


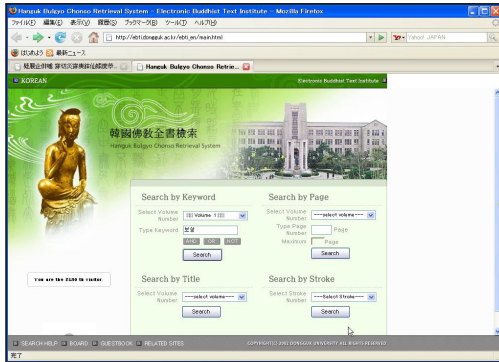
CBETA Reader の画面

CBETAは、独自の大蔵経の構築を予定しているため、大正蔵や続蔵の明らかな漢字や句点の間違いについては、訂正した版も作成しており、その数は2007年の段階で85部317巻に及んでいる。現在は、さらに大幅に増えていよう。しかも、CBETA Readerを立ち上げる際は、大正蔵どおりのテキストにするのか訂正版にするのか、両方の検索を可能にするのか、選択することが可能になっている。検索にあたっては、Big5以外のコードの漢字を自動的にBig5に変換して検索している。

大正蔵と続蔵は、別々に、また両方にまたがった形で検索できるようになっており、巻による指定もできる。題目による検索その他、様々な検索が可能であり、検索は高速であるうえ、全文検索においては、and、or、before、nearの指定、およびワイルドカードの使用により、柔軟な検索が可能になっている。また、本文のうちの特定の箇所を見たい時には、巻・頁・段・行などを指定し、ジャンプすることもできるなど、ユーザーの様々な要望を考慮した作りになっている。

CBETA Reader での検索範囲設定画面





EBTC の英語版検索ページ

## ■ XML の活用

花園大学国際禅学研究所の電子化プロジェクトを受け継いだ面がある CBETA では、XML によるマークアップにも早くから力を入れており、最先端のマークアップを試みている。その結果、原文どおりの改行をすることが、行頭情報を表示するか、その他様々な表示方式を選択できるようになっている。

また引用コピーも工夫されている。引用コピーする場合は、国書刊行会の正統蔵版だけでなく、旧版、および広く用いられている台湾の新文豊書店版の巻・頁・段・行などの情報も示される。さらに、引用する際には、經典名・品名・巻数などや校異を表示することも、自分の好みで指定することができるなど、きわめて柔軟な作りになっている。

## ❖ EBTC (電子仏典文化コンテンツ研究所)

高麗大蔵経は、高麗大蔵経研究会によって早くに電子化されたが、韓国仏教の著作について論ずる際は、東国大学出版部から刊行されている韓国仏教全書を用いるのが通例である。本全書は、新羅時代篇 3 冊、高麗時代篇 3 冊、朝鮮時代篇 4 冊、補遺篇 4 冊の計 14 冊から成る。2006 年に開校 100 周年を迎える東国大学では、その記念事業の一環として、EBTC を組織し、韓国仏教全書の電子化に取り組んできた。

EBTI (Institute Electronic Buddhist Text & Culture Contents) は、他の国際組織との名称の重なりを避け、また「仏教関係の文物」の電子化という点を強調するために EBTC と改名したが、理系の研究者の協力も受

け、韓国仏教全書の電子化を 2007 年に完成した。成果は CD で無料配布されているほか、EBTC のサイト、([http://ebti.dongguk.ac.kr/ebti\\_en/main.html](http://ebti.dongguk.ac.kr/ebti_en/main.html)) でも検索ができる。

このサイトには英語ページも用意されているが、検索は、ハングルで入力すると自動的に漢字に変換されて検索する方式であるなど、研究者向けでないため、大幅な改善、またテキストファイルでの公開が望まれる。

## ❖ 大蔵経データベースの今後

これまで紹介してきたように、大蔵経の電子化は一段落し、協力と競争によって使い勝手の良さを高める段階に入ったが、これらはすべて版本に基づく大蔵経であった。今後は、所蔵するスタイン文書を中心にして敦煌文物の画像データベース化を進めている大英図書館の The International Dunhuang Project: The Silk Road Online (<http://idp.bl.uk/> 日本語ページは、<http://idp.afc.ryukoku.ac.jp/>) や、国際仏教学大学院大学が「デジタル版・復元唐仏教一切経」作成のため、学術フロンティア事業として進めている奈良平安古写経の電子化などが重要になってこよう。むろん、中国でも様々な版本や写経の電子化が進行中である。

漢文以外では、パーリ語大蔵経は既に複数の版が公開されており、チベット語大蔵経も複数の版の入力が進んでいる。サンスクリットの大蔵経は現存しないが、佚文を集めて公開しようとする試みが韓国でなされており、いずれは漢文大蔵経との比較もできるようになるだろう。

大蔵経以外の仏典電子化も盛んである。たとえば、花園大学の国際禅学研究所では、早くから電子化テキストの公開に努めてきたほか、駒澤大学図書館の電子図書館 (<http://www.komazawa-u.ac.jp/~toshokan/el/>) では、明治以前の禅籍の貴重書をすべて画像化しおわり、詳細な書誌を添えて公開している。SAT との連携も計画されているため、近年中に、大正蔵の電子テキストで禅籍を眺めていて疑問に思う箇所は、電子図書館で画像が公開されている禅籍については、画像に飛んで確認できることになろう。

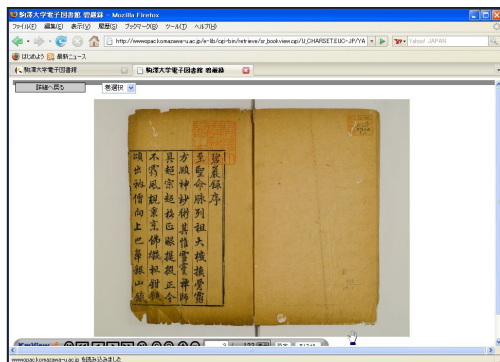
ただ、このように電子化が進み、便利な環境が整いつつあるものの、それにふさわしい研究成果が数多く出ていないのも事実である。これは、いずれの国でも



漢文の力や古典の素養が弱くなっていることに加え、先に触れたように、検索のコツなどを含めた活用法が知られていないことも一因となっていよう。今後は、書誌学などとならば基礎学の一つとして、電子データの特性に関する知識や検索法などを教える必要があるだろう。

注

- [1] 石井公成「三経義疏の語法」(『印度学仏教学研究』第57巻に掲載予定)。
- [2] 永崎研宣「人文科学のためのデジタルアーカイブにおけるコンテンツのサイクル」(「東洋学へのコンピュータ利用：第19回研究セミナー」京都大学人文科学研究所、2008年3月21日発表資料)。



駒澤大学図書館電子図書館

究所、2008年3月21日発表資料)。

# 東大寺文書目録データベース

照井 武彦

❖ はじめに

国立歴史民俗博物館(歴博)がサービスしている東大寺文書目録データベースについて紹介する。まず、歴博のウェブを紹介する。「れきはくホームページ」に入り、「データベースれきはく」、「東大寺文書目録データベース」の各入口を経て、検索の様子を見る。そのあと、元データの東大寺文書目録について、および東大寺文書について簡単に述べる。

❖ れきはくホームページ

「れきはくホームページ」<sup>[1][2]</sup> にアクセスすると、展示と催し、研究、大学院教育、施設案内、利用案内、などなどのエントリが並び、研究機関であり、博物館である歴博の性格を実感できる。ウェブデザイナー(非常勤)が居て、細部は頻繁に更新されている。

❖ データベースれきはく

データベース種別として、



図1 れきはくホームページ

- データベースれきはく 1  
インターネットですぐに検索できる
- データベースれきはく 2  
利用登録の申請が必要
- 来館者利用データベースれきはく  
利用登録が必要で、来館して利用する

の3種類がある。

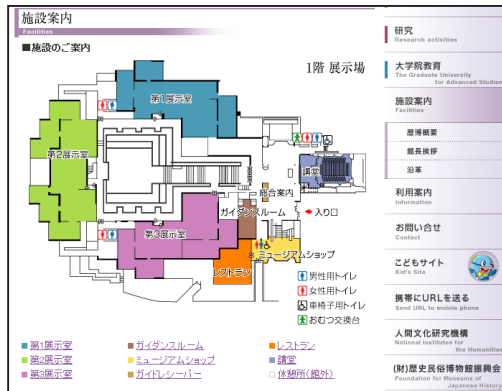


図2 施設案内のページ

データベース一覧を見るとデータベース1と2が31種類、全文記録類データベースが8種類ある。

以上について、詳しくは次節の手順に従って画面を見ていただきたい。

なお、全文記録類データベースはその名のごとく日記などのテキストデータベースである。

データベースれきはくとは別に、館蔵資料画像データベースがある。この中には文書、典籍、記録類も含まれているので、こちらも眺めていただきたい。

## ❖ 東大寺文書目録データベース

東大寺文書目録データベースはあいにく利用登録が必要なので、まず、そのための手順を示す。



図3 子どもサイトのページ

- (1) れきはくホームページ 研究をクリック
- (2) 研究のページ データベースをクリック
- (3) データベースのページ データベースれきはく2の「利用申請」をクリック

別窓に登録方法のページが出るのでその説明にしたがって登録申請していただきたい。数日すると利用承認書が送られてくる。

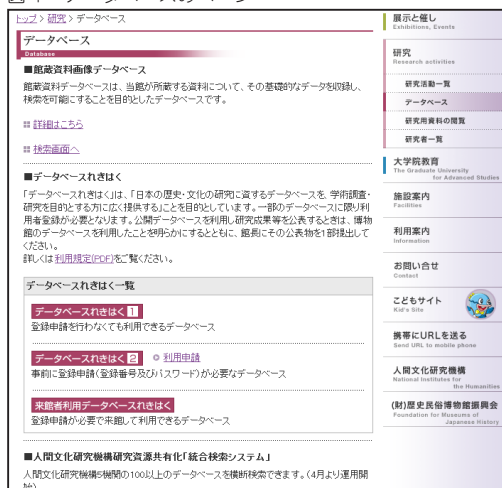
以下、検索のページにいたる手順を示す。

手順(1)(2)のあと、

- (4) データベースのページ データベースれきはく2をクリック
- (5) 別窓：データベースのページ スクロールダウン、公開データベース一覧の東大寺文書目録の概要説明をクリック

ここにデータ項目の一覧があるので見ておく。

図4 データベースのページ



- (6) 戻るをクリック
  - (7) (5)の東大寺文書目録をクリック
  - (8) データベースれきはくの入り口ページ 登録番号、パスワードを入れ、開始をクリック
- これで検索をはじめることができるが、ここでは東大寺文書目録データベースに慣れていない人のためにヘルプなどを先に見ておくことにする。
- (9) 別窓：東大寺文書目録データベースの検索ページ 検索語例をクリック、戻る
  - (10) 同上 ヘルプをクリック
  - (11) 別窓：検索条件設定画面 文書番号についてをクリック、戻る

(12) 同上 目録第1～5巻目次 文書番号表示付をクリック、戻る

目録凡例、目録例1～6も閲覧していただきたい。以上で、検索の手引きをひとつとおり眺めたことになる。冊子の手引きなら、パラパラめくって当たりをつけたところである。

検索には検索語例から検索語を選ぶと分かりやすい。たとえばフリーワードで「美濃国」を検索してみる。検索結果で「番号」をクリックすると、1件の詳細が表示される。なお、検索結果の「文書番号」の( )の中の数字は、冊子目録の巻(1桁)頁(3桁)である。このデータベースの開発経緯は以下のとおりである。

1981	歴博の機関設置
1982	図書システム開発、館蔵資料データ起し、3分野の研究データベースの選定など開始
1983	歴博開館、展示オープン
1986	東大寺文書目録のデータ入力を開始、試入力(第1巻の頭70頁)
1987-88	入力
1989-92	校正
1991	公開準備
1992	データベース公開 <sup>[2][3]</sup> 汎用機 HITAC660、検索は ORION、通信は電話回線・学術情報ネット併用
1999	データベース更新 サーバーへ乗せ換え、運用・検索は ORACLE、通信はインターネット

割り書き、傍注、ミセケチなどを含め1行に並べるために漢字記号のタグを採用した。タグは当時普及しつつあったTeXの考え方に沿ったものである。冊子目録の書式通りにプリントするに十分な、多様なタグ記号を使った。このデータベースでは、それらのタグを//に置き換えてある。

この//や、外字を表す■、その他漢字記号は一般の漢字と同じく検索にかけることができる。

### ❖ 東大寺文書目録<sup>[1]</sup>

第1巻の「序」に、この目録の作成事情がつぎのように述べられている。

この『東大寺文書目録』は、奈良国立文化財

■公開データベース一覧

データベース名	概要説明 最終更新 データ件数	データベース名	概要説明 最終更新 データ件数
館蔵資料(画像付き)	概要説明 2008年7月更新 209,053件	歴博図書目録	概要説明 随時更新
民俗誌	概要説明 1990年4月更新 5,203件	旧高田領取調帳	概要説明 1990年4月更新 97,359件
陶磁器出土遺跡	概要説明 1994年12月更新 7,992件	日本荘園	概要説明 1998年2月更新 8,975件
荘園関係文獻目録	概要説明 1998年2月更新 4,611件	土偶	概要説明 2001年10月更新 10,641件
館蔵中世古文書	概要説明 2003年3月更新 1,925件	近世農業遺跡	概要説明 1999年1月更新 1,317件
近世農業関係主要文獻目録	概要説明 1999年1月更新 1,904件	日本民俗学文獻目録	概要説明 2007年9月更新 33,458件
城館城下発掘	概要説明 1999年10月更新 3,348件	弥生石器遺跡	概要説明 2000年7月更新 3,459件
自由民権運動研究文獻目録	概要説明 2007年8月更新 4,253件	東大寺文書目録	概要説明 1999年11月更新 12,099件
棟札	概要説明 2003年7月更新 1,060件	古代・中世都市生活史(物備)	概要説明 2004年7月更新 37,253件

図5 データベース(一覧)のページ

データベースれきはくの入り口

データベースれきはくへようこそ。

当該サイトの利用者にパスワードを入力、開始ボタンを押していただき、[クリア]ボタンを押すと、入力した項目が消えて、空白の状態になります。

データベースれきはくでは、ログイン時にJAVAScriptが必要となります。ブラウザの設定でこの機能が無効になっている場合、JAVAScriptが導入できない可能性があります。以下のサイトから入手していただく(無料です)。 <https://java.com/ja/>

利用希望ID:

パスワード:

開始

利用希望IDをお持ちでない方は登録が必要です。詳しくは、当館のWWWページの「データベースれきはく登録方法」をご覧ください。当館の情報管理へお問い合わせください。

お問い合わせ先:  
〒840-0202 奈良県 住吉市 城下町117  
国立歴史民俗博物館 管理部 総務課 情報管理係  
Fax:0478-408-4125  
E-mail: [info@ml.cb.nihon.go.jp](mailto:info@ml.cb.nihon.go.jp)

National Museum of Japanese History. All rights reserved.  
<http://www.nihon.go.jp>

図6 データベースれきはくの入り口ページ

東大寺文書目録データベースの検索

検索語を入力し、[検索]ボタンを押してください。( [検索語例](#) )

[検索]

文書番号:

文書名:

次頁:

フリーワード:

結果表示件数:

(1)このデータベースは、奈良文化財研究所発行の「東大寺文書目録」を原資料として作成したものです。  
(2)このデータベースのいかなる部分も奈良文化財研究所の書面による事前許可無しに印刷形態や機械可読形態での複製はできません。  
(3)このデータベースの著作権は、国立歴史民俗博物館が所有します。

[れきはくホームページへ](#) [データベース選択へ戻る](#) [ヘルプを表示](#)

National Museum of Japanese History. All rights reserved.

図7 検索のページ

研究所が文化庁の依頼をうけて実施してきた東大寺文書の調査報告である。東大寺文書は現存する我が国の古代中世文書中でも質量ともに白眉の存在で、その史料的価値の高さについては今さら事新しく述べる必要もない。これまでに東大寺成巻文書その他が重要文化財に指定されているが、なお未成巻文書・薬師院文書その他、未指定のままとなっている文書も八千点を越えている。しかも中には破損により修理を要するものも少なくないが、その数量はきわめて大きく、現状を正確に把握することは容易でない。そこで文化庁は東大寺文書全体の保存を考えるのに必要な資料を得るため、当研究所に調査を依頼した。この調査は昭和四十九年以来実施し、文化庁の協力のもとに現在もお調査を継続中である。しかしこの調査報告の作製が急がれるため、調査完了をまたずに、調査終了部分から逐次目録を刊行して行くこととした。

その計画のとおり、調査と目録の刊行が進み、左下表のように A5 版 6 巻にまとめられた。

このように、データベースの原稿となったのは第 1 巻から第 5 巻までである。年代は、奈良時代から江戸時代までであるが、日付が月日だけで年号・年のない文書が多数あり、推定できるものもあり、不可能なものもある。後者は、編年目録の各時代の末尾に置いてある。

## ❖ 東大寺文書

東大寺文書の全体像を知るには 2001 年に刊行された「東大寺文書を読む」<sup>[5]</sup>が便利である。以下にその要点を記すので、詳しくは同書を紐解いていただきたい。

同書<sup>[6]</sup>によれば、全体は成巻文書と未成巻文書から成る。

成巻文書については、明治期に一部の文書が 96 巻の卷子本に成巻され、その後成巻された 5 巻を加えて 101 巻がある。

96 巻のうち 1 巻（東大寺奴婢見來帳）が単独で重要文化財に指定され（昭和 27 年）、他はのちに昭和 36 年重要文化財に指定された。後者を百巻文書と呼ぶ。

未成巻文書は大正以降昭和の初めにかけて、中村直勝氏らによる整理が行われた。その結果は「寺領部」等々上記のように反映されている。つまり、東大寺文書目録は、この未成巻文書の目録であり、これが完成したことにより昭和 59 年「東大寺文書 八千五百十六通」（未成巻文書）として重要文化財に指定された。

以上は東大寺に現に所蔵されている文書であるが、ほかにも多くの東大寺文書が存在する。これらについては「東大寺文書を読む」の「序」<sup>[7]</sup>につきのように述べられている。

東大寺では創建時代以来から太政官（だいじょうかん）を始めとする官庁が発給した文書や、寺領荘園に関する重要な文書を印蔵（いんぞう）に収納し、印蔵文書として大切に伝えてきた。今日東大寺に伝わる成巻文書百巻（九百七十九通）と、未成巻文書八千五百十六通のほかにも、明治五年（一八七二）八月に皇室に献上されて正倉院に保存されている東南院文書百十二巻（八百六十七通）や、内閣文庫・東京大

	頁	発行年	所載
第 1 巻	488	1979	東大寺文書 第 1 部（寺領部）第 1 ～ 第 23
第 2 巻	384	1980	東大寺文書 第 1 部（寺領部）第 24 ～ 第 26 東大寺文書 第 2 部（寺法部）
第 3 巻	360	1981	東大寺文書 第 3 部（文書部）第 1 ～ 第 9
第 4 巻	387	1982	東大寺文書 第 3 部（文書部）第 10 ～ 第 12 東大寺文書 第 4 部～第 6 部
第 5 巻	487	1983	東大寺文書 第 7 部～第 10 部、第 104 部、 第 141 部、第 141 部 B、 第 142 部 東大寺文書（宝庫文書） 薬師院文書第 1 部、 薬師院文書第 2 部、 薬師院文書第 4 部
第 6 巻	—	1984	編年目録 成巻文書 指定品

学（文学部・法学部・史料編纂所）・京都大学文学部・早稲田大学図書館・天理大学図書館・根津美術館・大東急記念文庫・大倉集古館・お茶の水図書館・エール大学・真福寺などに所蔵されているものを含めた東大寺文書の総数は二万通に達すると思われる。

東大寺文書の活字化された史料集のうち、小口雅史氏の史料集成<sup>[4]</sup>は、その書名のとおり冊子とCDROMからなり、東南院文書などが使われている。日本古代特有の漢字を取り扱った先進的労作である。

## ❖ おわりに

東大寺文書を読む<sup>[5]</sup>が刊行されたのは、このデータベースが動いてからずいぶん後であるが（筆者歴博退職後）、上記の経緯やデータの意義が理解でき、筆者個人にとって、まことにありがたいことであった。

このデータベースは田中稔氏<sup>[8]</sup>の強い要望からはじまり実現した。歴博に赴任して来られるまでは、上記文化庁調査および目録刊行の推進役であったそうで、データが一通りできた時にエディタの単漢字検索を教えたところ、喜んで使っておられたのが印象的であった。急に亡くなられたため、データベースは間に合わなかったのが残念であった。

このデータベースで不十分な点は多々あるが、なかでも外字と新しい漢字コードへの対応が気になる。

使用した漢字コードはJIS X 0208 1983年版、その外字は上に述べたように■と表示される。表示された■がどの漢字に当たるかは、その都度6巻の冊子目録を参照しなければならない。

また、最近新しい漢字コードが普及してきた。その1は、史料集成<sup>[4]</sup>のように今昔文字鏡などを利用する方式である。これによれば殆どの外字を解消することが期待できるが、パソコン上のデータベース個別開発によるもので、歴博データベースのようにサーバーに載せるには、一からソフト開発しなければならない。

その2はJIS補助漢字（JIS X 0212, UTF-8）を取り

込むことである。外字がどの程度救われるか、改めて調査してみなければわからない。こちらは技術的に、よりサーバーになじむものと思われる。

以上、最近の漢字コード事情については勉強不足のままの論であるから間違っているかも知れない。ご教示いただければ幸いである。

最後に、当時一緒に構築に当たった方々、種々お教え下さった方々に改めて感謝し、この稿を終わる。

## 注

[url-1] 国立歴史民俗博物館

<http://www.rekihaku.ac.jp/index.html>

[url-2] 東大寺

<http://www.todaiji.or.jp/>

[url-3] 奈良国立文化財研究所

<http://www.nabunken.go.jp/>

[1] 東大寺文書目録（全6巻）、奈良国立文化財研究所編、同朋社、1979～1984

[2] データベース検索の手引き（東大寺文書目録データ）国立歴史民俗博物館、1992（執筆担当：著者）

[3] 東大寺文書データについて、照井武彦、国立歴史民俗博物館研究報告、第12集、1993

[4] デジタル古文書集 日本古代土地経営関係史料集成——東大寺領・北陸編——、小口雅史編著、同成社、1999

[5] 東大寺文書を読む、綾村宏・永村真・湯山賢一編、思文閣出版、2001

[6] 注[5]一、文書の伝来 概説 綾村宏氏 本文2～3頁

[7] 注[5]序 堀池春峰氏 目次1頁

[8] 田中稔氏

調査開始当時、奈良国立文化財研究所・歴史研究室長  
1981年、歴博・歴史研究部長

1984年、歴博・情報資料研究部長

1991年、逝去

稲荷山古墳出土鉄剣銘文などの解読者でもある。

## 中国のネットとIT事情

千田 大介

### ❖ 3G と通信業界再編

#### ■ 中国の通信業界再編

2008年5月、かねて噂されていた、中国の通信キャリアの再編が発表された。通信キャリア主要6社、中国移动通信（移動、チャイナモバイル）・中国电信（電信、チャイナテレコム）・中国網絡通信（網通、チャイナネットコム）・中国聯合通信（聯通、チャイナユニコム）・中国鐵路通信（鉄通、チャイナレールコム）・中国衛星通信（衛通、チャイナサットコム）がシャッフル・統合され、新たに新移動・新電信・新聯通の3社が発足することになった。このようなダイナミックな再編は、これらの企業が国有企業であるからこそ可能なのである。

中国の通信キャリアは、旧郵電部系の中国电信一社体制からスタートした。1994年、中国政府は英国的な二大通信企業体制の構築を企図して聯通を発足させ、後には鉄通・衛通を設立するなど、通信市場に競争原理を導入しようとした。しかし、以前は電信、現在は移動が市場で圧倒的な力を持っており、理想的な競争的市場は形成されなかった。そこで、各社の業務と規模とをできるだけ平準化し、対等な競争関係を生み出そうというのが、今回の再編の主眼であろう。

中国国内のこの再編の評判は、必ずしも芳しくない。しかし、突出したシェア・収益を上げている移動を解体するわけにいかない以上、固定電話、2G・3Gの携

帯電話といったサービスを均等に備え、かつ企業の規模を接近させるには、この組み合わせがベストに近いことも確かである。

#### ■ 移動と TD-SCDMA

情報産業部（2008年6月に工業と情報化部に改組）が、各社の統合作業の進展を見て3G営業免許を発給することとしていることから、今回の再編が3G導入と密接に関連するものであることは確かだ。

国際的な携帯電話の3G規格としては、WCDMA（日本ではドコモ・ソフトバンク）とCDMA2000（同じくAU）があり、特に前者はGSM規格からのアップグレードで幅広く採用されている。中国では、これらに加えて、独自の3G規格としてTS-SCDMAが開発されている。TD-SCDMAは大唐グループがドイツのシーメンス社の協力を得て開発したもので、提唱から9年を経た2008年4月、さまざまな紆余曲折の末、ようやく試験サービスにまでこぎ着けた。

北京・上海・広州などの都市で始まった移動によるTD-SCDMAの試験サービスは、しかし評判が芳しくない。報道によれば、通話が繋がらないことが多い、テレビ電話機能ではマスの大きなモザイクしか見られない、メールがかなりの高確率でロストするなど、評判は散々である。これは通信網整備の遅れが主な原因だとされ、移動が本心ではWCDMAをやりたいがためにサボタージュしているのではないかという穿った見方もあった。

しかし、今回の業界再編で、新移動の3GサービスがTD-SCDMAになることが確定した。近年、独自のイノベーションを提唱し、世界の下請け工場からの脱却を目指す中国政府は、新聯通・新電信に比べて倍近い収益とユーザー数を誇る新移動に、独自規格3Gの普及と発展という重責を与えたことになる。

#### ■ 新聯通・新電信のチャンス

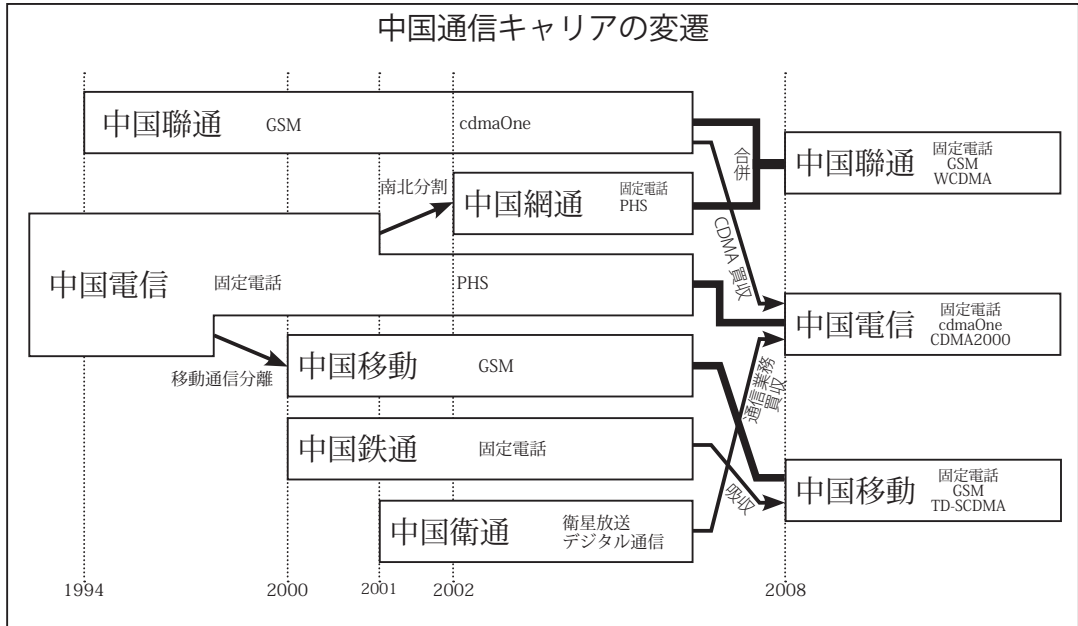
この再編で最も大きな利益を得たのは、おそらく聯通であろう。聯通は、設立当初より中央政府上層に後押しされてきたが、旧中国电信系が主流を占める通信行政の現場では逆に、冷遇され続けてきた。今回、聯

中国移动通信



大唐グループ





通は GSM と cdmaOne という二重のインフラ整備の責務から逃れたのみならず、WCDMA という最も美しい 3G 規格を手にすることができた。

現在、中国で販売されている携帯電話端末は、大多数がノキア、モトローラ、ソニー・エリクソン、サムスン、LG といった海外系メーカーの製品で、ハイエンド機の大半は WCDMA と GSM のダブルバンド対応である。だから、既にそうした高級機を持っている人々にしてみれば、3G 化にあたって端末を買い換える必要のない WCDMA のメリットは大きい。また、iPhone 3G も WCDMA 対応であるから、中国で発売されるとしたらキャリアが新聯通になる可能性が非常に高くなった。

新電信にとっても、comOne から CDMA2000 へのアップグレードは比較的容易であるとされるので、3G サービスを開始する上で恰好のプラットフォームを得たことになる。既存ユーザーをそのまま 3G 環境へ移行でき、また海外メーカー製端末をそのまま導入できるのも強みだ。

両社にとっては、組織や業務の統合作業を迅速に進め、こうしたアドバンテージを如何に業績に結びつけることができるかが、ポイントとなる。

■ 中国国産端末と“山寨機”

現在のところ、TD-SACDMA には、中国国産の端末しかない。海外系メーカーも端末の開発はしているというので、TD-SCDMA が軌道に乗れば投入されるのだろうが、しばらくは国産メーカーに頼らざるを得ない。

ところがその中国国産携帯メーカーが、近頃、全面崩壊の様相を呈している。2000 年代前半には、シェアの半ばを占めたこともある国産端末は、年々シェアを落としており、2007 年には市場シェアの 8 割以上を海外系 5 大メーカーが占めるまでになっている。これは、製品の品質やアフターサービス、さらには海外系製品の模倣デザインといった問題が大きく響いてユーザーの信頼を失ったゆえであり、TD-SCDMA がそれら国産メーカーの起死回生となり得るかは予断を許さない。

一方、中国国内では、“山寨機”と呼ばれる国産ノーブランド端末が、燎原の火のように勢いを得ている。中国で携帯端末を生産するには、国の免許を受けなくてはならないが、携帯端末そのものは、既製パーツを組み合わせるだけでも生産できるので、技術的なハードルは高くない。特に近年、台湾メディアテック（聯発科）社が、携帯端末用チップ・ボード・ソフトウェア

アのパッケージを安価で大量に供給するようになったことから、町工場がメーカー製品を模倣した筐体にメディアテックのパッケージを組み込んだ携帯端末を違法に生産し、闇市場に流すようになったのである。

メディアテック社のパッケージが高性能で安定していることから、“山寨機”は国産メーカー製よりも高品質で、海外系メーカー端末よりも遙かに安価であるとして、農村や地方都市で人気を集めており、都市部にも進出しつつある。買う方も、メーカーサポート一切無しの安価な端末として、割り切って使っているようだ。

“山寨機”も、政府の市場参入規制に起因する需要と供給の乖離、イノベーションに乏しい地方部の法規を無視した場当たり的な営利主義、目先の安価を追求する消費態度などなど、中国の社会・経済が抱える問題の落とし子であるという点で、海賊版DVDなどと同様の存在である。それゆえに、公式統計の枠外で、これからもシェアを伸ばしつづけていくことだろう。

## ❖ 狂乱の“人肉サーチ”

### ■ 人肉サーチとは

中国では、インターネット検索を通じて人物や個人情報を探し出すことを、人肉サーチ（人肉搜索引擎）という。事件の犯人や生き別れた肉親に関する情報提供をネットで呼びかけるようなもの、例えば2008年8月に話題となったiPhoneガール探しなども人肉サーチと呼ばれるが、人肉サーチの人肉サーチたる所以は、ネット上で話題となった道徳や愛国主義などの点で問題ありとされた人物の本名・住所・経歴等のリアルな情報を、徹底的なネット検索で探しだし、更にそれをネット上に晒すことで、本人の现实生活に影響が及ぶほどの集中攻撃を加える、かかる溺れた犬を袋叩きにするかのようなネット上の“祭”的イベントである点にある。“人肉”というカニバリズムを想起させる語を冠するのは、ブラックであるが、ある意味、本質を突いていると言える。

人肉サーチの起源は2001年に遡るが、社会問題に浮上したのは2006年のことである。2月には、子猫をハイヒールで踏みつける女性の映像がネット上で物議を醸し、またたく間に容疑者が特定され、謝罪に追い込まれた。また4月に勃発した銅須事件では、妻がオンラインゲームのオフ会で“銅須”というハンド

ルネームの男と関係を持った、という夫の書き込みにネットワークerたちが反応し、銅須の本名や電話番号・住所・職場・出身校などを調べ上げてネット上に晒し、本人のみならず家族や知人までもが脅迫にさらされるなど、大きな社会的事件になった。

近頃でも四川地震をめぐり、各種オンラインエンターテイメントサービスが休止していることにむかつ腹を立てて地震被災者を侮辱する発言を収録したビデオをYouTubeにアップした女性が謝罪させられるなど、人肉サーチ事件は頻発している。

人肉サーチは、一方では犯罪者の特定など警察的機能を発揮している面もあるものの、その横暴ぶりで近年の中国ネットに暗い影を投げかけている。文化評論家として知られる同済大学の朱大可教授は、こうした行為を引き起こすネットワークer群を「道徳民兵」と名付け、中国のインターネットの特色であるとしているが、モラルや愛国をネタに特定の人物をさらし者にする行為そのものは、日本のいわゆる「ネットイナゴ」と本質的に異なることはないと言える。

### ■ 人肉サーチと個人情報保護

ただ中国では、「革命無罪」や「愛国無罪」という言葉が端的に示すように、動機が政治的に正しければ、行為そのものの違法性は問われないという考え方がある。反日デモで日本料理屋が破壊されても、個人が賠償責任を問われることが無かったことは、この考え方の一つのあらわである。人肉サーチやそれに続くパッシングが、脅迫電話など、犯罪レベルにまでエスカレートするのも、こうした考え方に大きな原因がある。

人肉サーチ問題にはもう一つの側面がある。2008年、中国で初めての人肉サーチ裁判が起こされ、原告側はプライバシーの侵害を訴えているが、しかし被告弁護側は、BBSなどに晒された個人情報は、いずれもネット上で検索し発見されたもので、プライバシーの侵害に当たらないと主張している。確かにたび重なる“人肉サーチ”事件で晒された個人情報は、大半が検索の名手によってネット上から見いだされたものである。ここから、企業や機関、学校などの個人情報保護意識の欠如が、人肉サーチ事件のもう一つの元凶であることが見て取れる。

中国では、2008年9月現在、『個人情報保護法』の草案が既に国務院に送られているので、正式に公布される日も近いと思われる。同法は、人肉サーチ裁判



に法的根拠を与えるものだと期待されている。しかし、法律が制定されることと法律が理解され遵守されることは別の問題であり、法治意識の希薄さ、公権と私権との関係、個人の権利の軽視といった中国社会に根ざした諸問題と密接に結びついているだけに、同法が施行されたとしても、人肉サーチ問題の解決にはまだまだ時間がかかるものと思われる。

## ❖ 動画投稿サイト規制

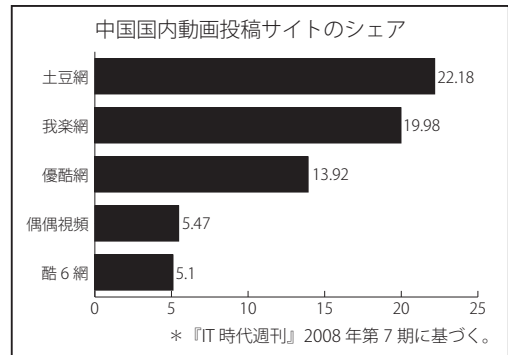
### ■ 中国国産動画投稿サイトの隆盛

我が国でもここ2・3年、YouTubeやニコニコ動画といった動画投稿サイトが人気だ。今や、テレビのドラマやバラエティは、専らそれらの動画投稿サイトで見るという人もザラである。一方、コンテンツホルダと投稿サイトとの間で著作権問題をめぐるいざこざも絶えない。

動画サイトは中国でも高い人気を集めているが、いずれも中国国内のサービスであるという特徴がある。海外で圧倒的に人気のWebサービスが生まれると、Googleに対する百度、eBayに対する淘宝のように、中国では国内に後追いサービスが生まれて市場を押さえる、という傾向が見られるが、それには理由がある。

まず、海外サイトへの中国国内からのアクセスは、往々にしてグレートファイヤーウォールの干渉を受けがちである。動画投稿サイトについても、YouTubeでは海外の中国の負の面を伝えるニュースやドキュメンタリーなども視聴可能であるため、中国当局は国内からのアクセスを遮断している。また海外企業が中国国内にサーバを設置しローカライズサービスを提供しようとしても煩瑣な手続きや審査が必要で一定の時間がかかるため、そのタイムラグが国内の後追いサービスに有利に作用する面もある。当局にしても、コントロールしやすい国内企業が市場を押さえてくれた方が好都合である。

かくて中国には、独創性に乏しいYouTube後追い動画投稿サービスが、雨後の竹の子のように林立することになった。数百を数えるそれらサイトの中で比較的大きなシェアを占めているのは、土豆網 (<http://www.tudou.com/>)・我楽網 (<http://www.56.com/>)などの5サイトである。それらのサイトには、国内外の芸能番組・ドラマ・アニメなどの映像が山と登録されており、しかも著作権管理がかなりゆるいため、著



作者の要請を受けてYouTubeで削除された動画が、中国のサイトでは視聴可能、ということもしばしばだ。日本のドラマ・アニメなどを視聴するために、中国国内動画投稿サイトに日本からのアクセスが殺到するという現象すら生まれているという。

このような状況に対して、中国当局の取締の不十分を批判する人がいるが、むしろ問題はそうした動画の登録を知らながら、法的措置などに及び腰な日本のコンテンツホルダの姿勢にある。ただでさえ中国のサイトは、著作権意識が希薄であると言われていたのだから、YouTubeの登録動画監視に払っているのと同様以上の力を注がずに、また法的手段も取らずに事態が好転するはずはない。

### ■ 動画サイト規制の背景

2007年の年の瀬も押し迫った頃、政府はあらたな動画投稿サイト規制を打ち出した。12月29日、情報産業部と国家広電総局（ラジオ映画テレビ総局）が



共同で『インターネット視聴覚番組サービス管理規定』を公表し、動画投稿サイトなどの動画や音声ファイルを配布するサイトは、国有企業でなくては運営できず、社会主義や人民のために奉仕しなくてはならないなどと定めた。

2008年3月20日には、『インターネット視聴覚番組サービスサンプリング調査レポート(1)』が発表され、25サイトにサービス停止を命じ、32サイトに警告を発するなど、厳しい処分がくだった。

今回の規制は、要するに、以後、動画投稿サイトを放送事業者に準ずる、映像や音声を伝達するメディアとして扱う、という宣言である。ネットワークの所管官庁である情報産業部のみならず、メディア規制を担当する国家広電総局の連名であることが、それを如実に物語っている。

ここ数年、広電総局はインターネットが未審査映像・音声を発表するメディアになっていることに神経を尖らせてきた。例えば、映画人がデジタルビデオ片手に撮影したドキュメンタリーや芸術映像は、DV映画などと呼ばれて、中国映像芸術の最先端を担っているが、それが審査を経ずにネット上に公開されることに対し、広電総局は以前から批判的だった。

また2006年、陳凱歌監督の迷作『PROMISE 無極』をパロったMADムービー、『一つのマントウが引き起こした殺人事件』がネットで大流行したときも、それが未審査映像であること、また自らが審査した国産大作映画への批判であることなどから、広電総局は不

快感を表していた。自由や民主といった価値観が埋め込まれたアメリカの国内未公開ドラマの大量の流入も、当局の警戒心を誘った。更に、動画投稿サイトには、政府にとって都合の悪いシーン、衝撃シーンなどの映像が大量にあふれ、また過当競争からセンセーションを追求するサイトも増えていた。

だから広電総局にとって、動画投稿サイトは国家プロパガンダ体制への脅威にほかならず、それ故に今回の規制は、社会主義体制の必然であるとも言える。

中国の動画投稿サイトの隆盛は、海外ベンチャーキャピタルの積極的な投資によるところも大きいとされる。今回の突然の規制は、国際的な経済状況の悪化によって新興国から資金を引き揚げる動きとあいまって、中国の動画投稿サイトのみならず、IT業界さらには経済全体に負の影響をおよぼす可能性が高い。

今回の動画投稿サイト規制が、北京五輪での偉大なる勝利を目指した中国政府の動きと密接に呼応したものであることは、論を俟たない。インターネット企業の経営状態などは、五輪成功という国家的至上命題の前には些細な事であると判断されたのだろう。

しかし、北京五輪という人民戦争モデルによる国家的祭典は終了した。もはや五輪の成功という目標を口実に、不平不満を押さえ込んだり、問題解決を先送りしたりすることはできなくなった。政治体制の維持・改革と経済発展、環境対策や所得格差・人権などの問題、いよいよ中国も正念場にさしかかっている。

## 図書館・OPAC・漢籍

小島 浩之

### ❖ はじめに

本誌昨年号(第8号)では、Google特集が組まれた。その中で筆者は「アーカイブ検索ツールとしてのGoogle」と題した一文を草し、Google scholarとGoogleブック検索の日本語β版について紹介等を試みた。それから1年を経過したが、いずれもβ版のままである。しかし、Googleブック検索については、コンテンツ、インターフェイスの両面で変化が見られ

た。そこで本年度のレビューはまずGoogleブック検索からとりあげる。次に国内の図書館のOPACや漢籍目録の状況について順次みてみよう。

### ❖ Googleブック検索 日本語β版

<http://books.google.co.jp/books>

#### ■ 出版社との連携のその後

昨年段階では、Googleブック検索で、日本語の

書籍のヒット数は少ないものであった。これは日本で  
の事業展開が開始されたばかりで、提携する出版社や  
図書館が少ないという事情があった。

しかし出版社については、この一年間で多くの提携  
が進んだようである。Google と提携している出版社  
一覧のようなものは公開されておらず、あくまで推論  
に過ぎないが、以下の方法でおよその傾向は確認でき  
る。すなわち、検索オプションで、言語を日本語、出  
版年を 2000 年から 2008 年、検索対象を「限定プレ  
ビューと全体表示」(本文等が一部分でも PDF で公開  
されているもの) の条件で検索すると、4342 件もの  
ヒット (2008 年 9 月 10 日現在) がある。

### ■ 慶應義塾大学図書館との連携のその後

図書館との連携については、2008 年 1 月 10 日に、  
慶應義塾大学図書館所蔵図書のうち、174 点が公開  
された<sup>[1]</sup>。日本の図書館における初めての、Google  
との提携プロジェクトであり、今回、公開されたのは、  
福沢諭吉および慶應義塾関係の図書である。『学問の  
すゝめ』や『西洋事情』、『慶應義塾百年史』、『慶應義  
塾図書館史』といったものから、慶應義塾発行の各種  
目録やパンフレット類などが閲覧できる。

検索オプションで検索対象を「限定プレビューと全  
体表示」に絞り、慶應義塾のキーワードで検索すると、  
444 件がヒットするので、1 月以降さらなる公開が進  
んでいるのであろう。

### ■ 利用者インターフェイスの変更など

本文閲覧画面のインターフェイスは、多少変更され  
ている。閲覧画面には「この書籍について」および「こ  
の書籍を読む」の二つのタブがある。前者には、当該  
資料の書誌と、本文から抽出されたキーワードが表示  
される。書誌によっては、所蔵元の図書館名が明記し  
てある。

図 1 に、「この書籍を読む」タブを選択した画面を  
示す。左フレームには本文が PDF で表示される。右  
フレームでは、書籍内の検索や、テキスト表示、ダウ  
ンロードも可能となっており、至れり尽くせりの感が  
ある。ただし、OCR を使ったテキスト表示は、日本  
語に関してはあまりあてにならない。

PDF 画像には、ページごとに「Google によるデジ  
タル化」とのクレジットが入るが、著作権問題をクリ  
アしたものは全てダウンロードできるのは、大変に

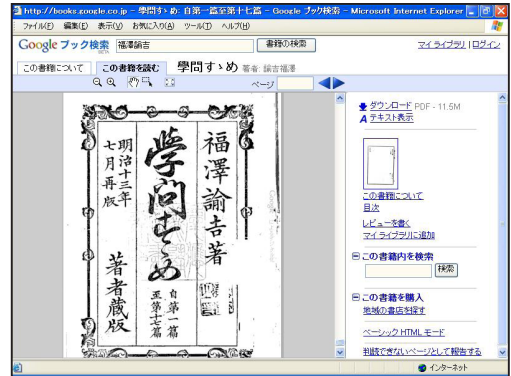


図 1: 本文閲覧画面



図 2: Google ブック検索トップページ

ありがたい。

また新たな機能として、マイライブラリ機能がある。  
この機能を使うには、Google アカウントを取得しな  
ければならないが、必要な書籍を登録して独自の電子  
図書館を構築できる。

マイライブラリが Google の機能に加わったことで、  
トップページも少し変更となっている。ページ上部に、  
ウェブ検索などへのリンクや、Google アカウントへ  
のログインリンクも貼られ、β 版といえども、正規  
版の様相を呈している。

この他、Google がらみの話題としては、アメリカで、  
新聞のデジタルアーカイブを開始したことだろう。こ  
ちらも近い将来、日本や中国に上陸するかもしれない。

### ● Google News Archive Search β 版

<http://news.google.com/archive/search>

### ◇ 近代デジタルライブラリー

Google ブック検索に対抗というわけではないが、  
国立国会図書館 (NDL) の近代デジタルライブラリー



図3：四庫提要の検索画面



図4：漢籍DB検索結果詳細表示（画像あり）

図5：Webcatでヒットした漢籍DB所収書誌の例



も健闘している。2008年8月26日に、明治期約400タイトル（約600冊）、大正期約4,200タイトル（5,100冊）が追加された。この結果、総数は約101,400タイトル（約148,200冊）となり、10万タイトルを突破した。

Googleブック検索のPDFはダウンロードできる点では便利だが、閲覧するには重く動きが緩慢である。また商業ベースの出版物も混在しているため、求めるデジタルアーカイブに辿り着くのに時間がかかる。また先述のように、テキスト化は現段階ではほぼ使いものにならない。したがってGoogleブック検索と近代デジタルライブラリーで、同じ書籍を閲覧するならば、現段階では、近代デジタルライブラリーの方が使い勝手が良いだろう。

もちろん、両者のコンテンツは全て同じではないし、共存の意義は大きい。あくまで同じ書籍を閲覧するという条件の下での評価である。

## ◆ 漢籍データベースとWebcat

京都大学人文科学研究所附属漢字情報研究センターの提供する漢籍データベース（漢籍DB）も、一昨年あたりから、徐々にリニューアルされている。

<http://kanji.zinbun.kyoto-u.ac.jp/kanseki>

まず、附随するツールとして、四庫全書総目提要がテキスト化され、全文の検索も可能となった。

### ●全国漢籍データベース 四庫提要

<http://kanji.zinbun.kyoto-u.ac.jp/db-machine/>

ShikoTeiyo/

また、典拠情報の整備ということで、書誌情報の典拠を示すとともに、蔵書印、題字、序跋などの追加情報をデジタル画像で示す作業が行われている。現在は、旧東方文化学院京都研究所蔵漢籍の史部について、図4のように検索結果に典拠画像が伴うようになっている<sup>[2]</sup>。

さらに、平成18年度より、国立情報学研究所（NII）のWebcat（<http://webcat.nii.ac.jp/>）との連携も計られている。現在、Webcatでも京都大学人文科学研究所附属漢字情報研究センター所蔵分の漢籍6,517件について、検索が可能である。漢籍DB分の書誌については、検索結果一覧において、書名の前に、\*\*\*が付いている。Webcatの書誌詳細画面の例を図5に示す。

書誌にはURLが貼られており、これをクリックすると漢籍DBの検索結果詳細表示画面（図4）に移動する。逆に、漢籍DBの検索結果詳細表示画面にもWebcatへのリンクが貼られ、図4と図5は相互リンクとなっている。

漢籍だけを検索したい利用者にとっては、現代目

録法に準拠した Webcat は非常に使いづらい。しかし、漢籍 DB と Webcat の両者が連携することで、史料と関連文献が一度に検索できるという利点もある。したがって、このように個別検索と統合検索の両方の機能を提供してゆくことは、今後重要になってくるのではないだろうか。

### ❖ 東洋文庫 新収蔵漢籍検索

<http://61.197.194.11/open/KansekiQueryInput.html>

東洋文庫では、従来から漢籍オンライン検索 (<http://61.197.194.10/TBDB/KansekiQuery3.html>) により、既存の冊子目録所収の漢籍をデータベース化してきている。これに対し、今回公開されたのは、上記データベースに納められた冊子目録が刊行された後に収集された漢籍である。今後の増加分はこちらにデータ入力されてゆくのだろう。

### ❖ おわりに

以上、図書館の話題として、慶應義塾大学と提携した Google ブック検索、NDL の近代デジタルライブラリーを取りあげた。また OPAC からは Webcat と漢籍 DB のリンクを、漢籍目録からは、漢籍 DB と東洋文庫について話題を提供した。

これらは、みな既存プロジェクトの進展や、データベースの改良にともなう話題で、新鮮味に欠けるかもしれない。しかし、こういった重要なサービスを継続して評価することも本誌の務めだと言えるだろう。

### 注

- [1] <http://www.keio.ac.jp/pressrelease/080109.pdf>
- [2] これとは別に、漢籍 DB に収録された都立日比谷図書館の実藤文庫の漢籍データから、252 点の巻頭画像も同様に公開されている。  
<http://www.library.metro.tokyo.jp/12/12354.html>

## 中国語教員必見サイト

金子 眞也

### ❖ はじめに

このコーナーでは新しくできたウェブサイトに限らず、中国語教員にとって必見のサイト、何かと便利なお役立ちサイトを紹介していく。いずれも筆者が日頃からよく参照・利用しているサイトばかりである。網羅的な紹介ではないこと、そして「学術」からやや離れてしまうことをお許し願いたい。なお、ポッドキャスト関係は大阪府立大学清原文代氏のウェブサイト詳しく紹介されているので、そちらを参照されたい。清原氏のサイトには大阪府立大学中国語ポッドキャストや氏個人のポッドキャストへのリンクが張られている。

<http://www.las.osakafu-u.ac.jp/~kiyohara/>

### ❖ 中国語の元締め

中国語が中国の言語である以上、「日本でしか通用しない中国語」とか「〇〇大学でしか通用しない中国語」などというものは存在しないというのが筆者の考えである。筆者と異なる考えを持つ人士を筆者は存じあげているし、中国語が話されている国・地域は中華人民共和国に限らないことも事実であるが、「中国語の元締め」としては中国本土を考えるべきだと筆者は考える。そこで、まず“国家汉语国际推广小组办公室”（略称“国家汉办”）を紹介する。

[http://www.hanban.edu.cn/cn\\_hanban/](http://www.hanban.edu.cn/cn_hanban/)

“国家汉办”は中国政府の中国語普及政策の元締めにあたるところで、このサイトから“孔子学院”や“汉语桥比赛”へのリンクが張られているので見逃さない。中国の主催する中国語検定試験は最近数が増えつつあるが、オーソライズしているのは最終的には“国家汉

か”になるのだろう。

私見によれば、“汉语橋”世界大学生中文比賽の日本国内予選の参加者は、日本における中国語学習者の数に比べると、少なすぎるように思われる。コンテストの存在自体があまり知られていないのだろうか。語学を専門としない中国語教員の方にも“国家汉办”や“汉语橋”のサイトは訪問してほしいと筆者は考える。

## ❖ 教学系の学会

### ● 世界汉语教学学会

<http://www.shihan.org.cn/shihan/index.do>

世界レベルの中国語教育の学会であり、日本からも学界の重鎮の先生方が《世界汉语教学》の編集委員を務められておられる。

### ● 中国語教育学会

<http://www.jacle.org/>

日本国内の中国語教学に関する学会。会報や大会、研究会等を通じて、世界汉语教学学会に関する情報やHSKの動向など有益な情報を知ることができる。

### ● 日本中国語学会

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

日本国内の中国語学研究者の学会だが、最近では教学に関する発表も大会等で活発に行われるようになってきている。

## ❖ 中国語教育お役立ちサイト

ここでは「音声のあるサイト」のみとりあげる。

### ● パンダと学ぶ中国語

<http://saigusa.com/>

見て楽しむためになる教材を発信しているところといたら、その筆頭はまず三枝裕美氏の「パンダと学ぶ中国語」をおいてほかにない。ムービーやかわいいイラストもあり、楽しく学習できるサイトである。

### ● 游

<http://133.220.106.221/index.html>

成蹊大学中国語 e-Learning のサイトである。「游」の Home から「中国語音声教育 DB システム」をクリックすると以前からある「中国語音声データベース」へ、また「e-Learning システム“遊 (yóu)” (教材)」をクリックするとマルチメディア教材に飛ぶ。中国語担当の専任教員が全学で湯山トミ子氏一名だけという情況のも

とでこれだけのものを構築し GP に採択されたことは、驚嘆に値する。

### ● Chinese Station 2

<http://cnstation.cocolog-nifty.com/blog/>

関西大学の中国語ビデオポッドキャスト。ポッドキャストである一方、パソコン上でそのまま視聴できるようになっているので、iTunes を使わなくても楽しめる。データベースに納められた冊子目録が刊行された後に収集された漢籍である。今後の増加分はこちらにデータ入力されてゆくのだろう。

### ● 東外大言語モジュール (中国語)

<http://coelang.tufts.ac.jp/modules/zh/index.html>

中国人同士が会話する画像を見ながら中国語を学ぶことができる。

## ❖ 検定試験など

### ● HSK-中国汉语水平考試

<http://www.hsk.org.cn/>

汉语水平考試 (HSK) の公式サイトである。いわゆる「第二外国語」として中国語を学ぶ学生にとって、HSK はちょっと敷居が高いかもしれないが、中国留学を目指す学生にはぜひチャレンジしてもらいたい試験である。

### ● 日本中国語検定協会

<http://www.chuken.gr.jp/>

日本中国語検定協会 (中検) の公式サイト。「中国語検定試験」→「試験データ」とクリックしていくと第 59 回以降の試験問題と解答が公開されているので試験の感じをつかむことができる。ただし音声は公開されていないので、リスニングの練習はできない。

### ● 中国語検定過去問 WEB

<http://www.kodensha.jp/Chuken/>

高電社の運営する中検過去問の学習サイト。2008 年 9 月現在は β 版で無料だが、今後課金制に移行する見込みである。こちらは中検公式サイトとちがって「実際に音声を聞きながら問題を解く」ことができ、その点で利用価値が高い。

## ❖ 実用的なツール

最後に中国語教員にとってとても実用的なツールを

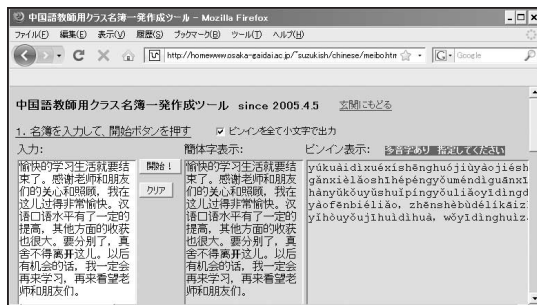
紹介する。鈴木慎吾氏のサイト「多言語実験室」におかれた氏作成の「中国語教師用クラス名簿一発作成ツール」である。

●多言語実験室

<http://homewww.osaka-gaidai.ac.jp/~suzukish/index.html>

このツールの優れている点は、「名簿作成」をうたいながら、図のようにかなりの長文の中国語でもピンインに変換してくれる点にある。

多音字にも対応している。ただし、普通の文章を変換する場合は轻声の部分は自分で手直しの必要がある。筆者は名簿作成用ではなく、長文のピンイン変換



用に利用させてもらい、重宝している。

ツールの作者の鈴木氏にとっては本意な使われ方かもしれないが、便利な物は便利なのである。

「龍源」(Dragon source)  
——中国語雑誌オンライン購読サービスを試す——  
小川 利康

◇ オンライン・データベースの充実

近年の中国のオンライン・データベースの充実ぶりについては多くを語る必要はないだろう。かつて本誌2号で紹介された中国期刊網(CNKI)、超星数字図書館<sup>[1]</sup>など、近年のオンラインデータサービスは、規模の大きさだけでなく、紙媒体では入手困難な資料まで豊富にそろえ、「大而全」の内容を誇っており、研究調査に欠くべからざる日常的ツールとなっている。また今日のオンラインデータサービスは「各有所長」であり、学術雑誌に特化した中国期刊網、デジタルブックを中心とする超星数字図書館と、分野別に特化し、棲み分けが出来ている。

今回、小文で紹介する龍源(Dragon source)は、中国国内で刊行される雑誌をタイムリーにデジタル化して提供するサービスに特化して発展してきた。かつて龍源について言及した拙文(2002年)では次のように紹介していた。

龍源期刊網は、現状では海外華人をターゲットにしているものの、将来的には有望なサイトである。



図1：龍源トップページ

このサイトが提供するの中国国内で刊行された雑誌記事であり、読者は費用を先払いすると、読みたい雑誌記事をオンラインで読むことが出来る。

この龍源期刊網の本体はトロント(カナダ)を本拠地とする中文書店であるが、オンライン書店を開設し、99年には雑誌のオンライン提

供を始め、現在では400種あまりの雑誌記事を有料提供している。(本誌2号、学術リソースレビュー「中国近現代文学」P.174)

ここで述べた龍源の紹介は今でも基本的に変わっていない。今でもオンライン決済で中国の雑誌が読めるサイトは他に余り例がない。さらに特筆すべきは初期の段階から、提供するデータはテキストを基本にしていたことであり、この点でも唯一無二といって良い存在であった。

現在、龍源が提供するデータは、雑誌2000タイトルまで拡大し、そのジャンルも時事問題から、社会科学、人文科学、映画、娯楽、トレンド、ファッションに至るまで多岐にわたり、その全てをほぼリアルタイムにオンラインで提供している。中国期刊網も雑誌を扱う点では共通し、重複する雑誌も若干あるが、龍源が得意とするのは一般向けの雑誌であって、学術雑誌

ではない。このため中国期刊網が大学など研究機関を中心にクライアントを獲得してきたのに対して、龍源の場合は超星図書館と同様、読者カードというプリペイド式カードを販売し、その購入者に一定期間アクセス権を与えるという定額購読方式であった。いわば超星図書館の雑誌版と考えると分かりやすいだろう。

今回、筆者がレビュー記事執筆に際して参照したのは、龍源が日本向けサービスを開始するに当たり開設したウェブページである。従って、本家のサイトの提供するサービスと必ずしも同一ではなく、「販売パートナー向けの機能デモ版」であるため、正式サービス画面では仕様変更される可能性があることを予めお含み置きいただきたい。

## ❖ 雑誌を読む:「電子版」と「原文原貌版」

龍源のデータはテキストファイルが中心である。これを「電子版」と呼んでいる。だが、近年新たに収録された娯楽・ファッション雑誌などはビジュアル重視の誌面構成であり、その特色を生かすため、現在では「原文原貌版」も存在する。こちらは誌面を画像として取り込んだもので、紙版のオリジナルの誌面を見ることができ、「電子版」では省かれた広告などもこちらで確認できる。

### ■ 原文原貌版

まずトップページを開くと、図1のように雑誌のカバー写真とタイトルが示される。上段のナビゲーションには、電子期刊の選択肢として、次の3つが出ている。それぞれ以下のような意味だ。

「整刊版」: 雑誌全体をテキスト化したもの。

「專題版」: 雑誌記事を主題別に抽出したもの。

「原文原貌版」: 雑誌全体を画像化したもの。

ここでは『瞭望東方周刊』のページを開いてみると、次のようになっている(図2)。最新号の表紙写真が出ていて、その下にはバックナンバーも表紙写真で選べるように並ぶ。この雑誌の場合、テキスト、グラフィック版のほかに音声で読み上げるサービスの「語音版」もある。これは龍源が独自に提供する朗読サービスのようなだ<sup>[2]</sup>。もし音声つきのデータが利用できるならば、中国語教材としての活用も可能であろう。

図2: 『瞭望東方周刊』原文原貌版のページ



図3: グラフィック版





説明によれば、Windows Media Playerなどで再生が可能になるというが、現時点ではまだ実験的なサービスのようで一部の雑誌にしかない。

画面のナビゲーションに従って、実際にグラフィック版を開いてみると、図3のように画像でページが表示される。その際、画像をページごとにダウンロードすることは出来ない仕様になっている。左クリックで読み進み、右クリックで画面を拡大し、スクロールボタンでページをスクロールできる。テキスト版で読み進むよりも操作が煩瑣になるのは確かだが、広告も含めてほぼ全て雑誌を丸のまま画像化しており、紙版のレイアウト・イメージが確認できる。どちらが良いかは利用目的次第だが、オリジナルの紙版の雰囲気を楽しみたいなら「原文原貌版」の方が適している。中国期刊網のように学術雑誌中心の場合、検索性に優れたテキスト版であることが重要になるが、龍源に求められているのは明らかに別のニーズであることがよく分かる。

## ■ 電子版

龍源の提供するデータは、出版社の提供するデジタルデータを利用して公開するケースもあれば、紙版の雑誌を受け取ってスキャナーで読み取り、データ化するケースもあるという。現在、北京を含む中国全土4拠点でスタッフが雑誌をテキスト化する作業をしている由である。当然デジタルデータの提供のある雑誌は公開が早く、そうでないものは、スタッフの手作業に依存しているので、公開が1ヶ月おくれるケースもあるという。自分の必要とする雑誌がどれくらいのタイムラグで更新されるかは龍源に直接確認した方が良いだろう。

また、最新号だけでなく、バックナンバーも龍源では利用可能だ。ほとんどのタイトルは2003年以降のようだ。なかには90年代以降のバックナンバーを全て揃えている雑誌もあるが、一定していない。

ここでは比較的古いバックナンバーから揃っている『百年潮』を見てみる。「電子版」を選択し、目次から該当の記事をクリックすると、図4のように、GB2312でエンコーディングされた標準的なHTMLファイルで表示され、長文記事の場合は数ページに分割して表示される。こちらについては自由にコピーして手元にテキストファイルを残すことが出来る。

一般のウェブ公開された雑誌や新聞からの転載記事



図4: 『百年潮』電子版



図5: 专题版

と異なり、龍源のスタッフが責任を持って校訂してあるだけにデータの信頼度は高いといえるだろう。

周知の通り、中国期刊網でも雑誌データが提供されている。だが、こちらのデータはすべてPDFかCAJファイル(専用Viewerが必要)になる。龍源ではオンラインで拾い読みしながら必要な箇所を直ちにコピー&ペーストできるのは、やはり便利だ。

## ■ テーマ別に読む「专题版」

テキスト版、グラフィック版のカテゴリとは別に提供される「专题版」はテーマ別に記事を提供するサービスである。このサービスは全てテキスト版で提供されるが、ジャンルごとに分類されてタイトルが表示され、興味関心に沿って雑誌を気軽に拾い読みできる(図

図6：検索画面

5)。

標準で提供されるテーマは時事、社会経済など大きなものなので、興味関心に沿う文章を見つけやすいとは言えない。だが、テキスト化された記事から検索によって記事を探し出すことが可能だ。これは記事の全文検索なので、キーワードによっては自分の読みたい文章をかなり高い確率で引き出すことが可能になる(図6)<sup>13)</sup>。

この検索自体は日頃誰もがGoogleでやっている検索を龍源の内部で行うに過ぎないのではあるが、検索対象が編集校訂作業を経て完成したデータに限定されるので、ジャンクデータは最初から排除されており、使い方によっては宝の山を掘り当てることが出来るはずだ。

ただ、遺憾ながら、現在の検索の仕様では複数の語彙を検索キーに指定しても、余り精度の高い結果が出てこない。高級検索でも、指定できるのは雑誌名、作者、タイトル、文中のキーワードだけである。Googleのように排除する単語やand/orの条件など、より細かい設定が可能にならないと、データ・マイニングは画

図7：電子図書

餅に帰してしまう。トップページにも全文検索用の入力欄は用意されているが、現状では実用に耐えないレベルである。検索速度の改善も含め、大いに改善が望まれるところだ。

## ■ 電子図書

龍源にとっては付随的なサービスながら、デジタル化された単行本も提供されている(図7)。文学関連が最も多く2500点を超えている。収録された書籍は基本的に画像として収録されたものだが、魯迅『中国小説史略』など定番といえる書籍が多数含まれており、専門家にとっては物足りないだろうが、急場をしのごう資料としては使えるだろう。

## ◇ まとめ

以上、駆け足で龍源の概要を紹介してきたが、筆者の雑駁な感想として、次の2点を指摘しておきたいと思う。

1. デジタル化が進むなかで、紙媒体の雑誌新聞を読むよりも、ウェブページからの記事参照、引用が増えているが、実は相当多数が掲載する際に省略や加筆が無断で行われ、杜撰なOCRでの処理により、誤字の多いデータが多数流通している。この事実気づいている人は意外に少ない。龍源のデータは著作権的にも対価を支払って提供され、発行元と読者に対する責任も明確なので、データにも厳密な正確さが期待できる。
2. ジャンクデータを大量に含む検索に時間を掛けるよりも、良質のデータを利用したデータ・マイニングの方が情報の精度も高くなる。中国期刊網とは採録対象が相当異なるので、研究資料とはならないが、現地調査の基本データとしての利用価値は十分あるだろう。

以上のような特長を考えるなら、①正確な原文データを必要とする研究者、②中国に関わるビジネスを手がける企業の企画開発担当者に一定のニーズがあると思われる。あとはニーズに見合う料金体系を日本で提示できるかどうかであろう。また、企業向けのほか、アカデミック版などの形で龍源は柔軟に対応する予定

があるというので期待したい。

いずれにせよ、対価を払う以上、全文検索のレスポンスに対する要求もシビアになると思われる。

さらに筆者からささやかなお願いだが、中国語教材づくりに腐心している大学教員向けにも、安価な従量制サービスを展開していただきたい。月額定額制では負担できないが、従量制なら十分ニーズはあると思われる。

## 注

- [1] 千田大介ほか「特集1：求められる学術研究情報の発信」、「超星数字図書館」(本誌2号、2002年刊)
- [2] 筆者の試した範囲では再生できなかった。
- [3] 筆者の試した範囲では何度か検索結果が帰ってこない

ことがあった。データベースが大きすぎるためだと思われるが、改善が望まれる。

## 編集部注

本レビューは、本文にも述べたように「販売パートナー向けの機能デモ版」を用いて書かれたものである。日本でのサービス提供に際し、開設されるWebサイトのURL等を日本でのサービスを提供されるビープラットフォーム株式会社に問い合わせたところ、日本では各クライアント毎に独自にURL等を設定するため、共通のWebサイトは存在しないが、代替として、日本語の龍源説明サイト(<http://www.dragonsource.jp>)を開設済みなので、参考にしていただきたい旨のお返事をいただいた。

## 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>

## ◆学術ソフト・製品

### 『漢語大詞典』CD-ROM ver.3

千田 大介

#### ◆ 待望の Unicode 対応

昨 2007 年、香港商務印書館から『漢語大詞典』CD-ROM ver.3 が発売された。ver.2 の発売が 2003 年であったから、およそ 4 年ぶりのバージョンアップということになる。

前バージョンでは、ver.1 にはなかった用例が収録されたことで、学術利用する上での利便性が大いに高まった一方、台湾 Big5 コードに基づくソフトであったため、収録される漢字の文字種が約 13,000 字に限られており、また日本語 Windows 環境で試用する際には、AppLocale の導入やロケールの切り替えが必要となり、また声調符号付きピンインが上手く表示できない、用例のコピー&ペーストでしばしば文字化けが発生するなど、さまざまな問題を抱えていた。

ver.3 の改善点として第一に挙げられるのが、Unicode への対応である。内部処理も含めて Unicode 化したので、ロケール切り替え無しに日本語 Windows 2000 / XP / Vista にインストールできるようになった。しかも、インストーラーには日本語インターフェイスまで用意されている。

辞書のインターフェイスそのものは、前バージョンとあまり変わっていない。しかし、ピンインなどの文字化けが発生しなくなったし、全く問題なく、本文をコピーしてワープロなどにペーストできるようになった。

ところで ver.2 では、インストール後 CD-ROM 内のファイルを丸ごと HDD 上にコピーし、プログラムフォルダ内の設定ファイルを書き換えてやると HDD 上で運用することができた。ver.3 では SecuRom によるセキュリティ機能が加わったため、CD-ROM の内容を HDD 上にコピーしてもプログラムを起動できなくなっている。

#### ◆ 収録文字種は増えたか？

Unicode に対応したからには、使用できる漢字の種類も増えたのではないかとの期待が膨らむ。マニュアル (p.6) には、ver.3 の見出し字は 18,014 字であると明記されているが、同ページ下の表には、Big5 の 13,069 字を収録と書いている。この数字の食い違いは、収録文字種は 13,069 字だが、発音・意味の違いなどによって重複して見出しに立てられた文字が延べ 5,000 字ほどある、という意味だと思われる。この数字を、ver.2 と比較してみよう。

	ver.2	ver.3	紙版
見出し字	18,013	18,718	約 22,700
熟語	336,385	336,706	約 375,000

残念ながら Unicode に対応した割には、見出し字・熟語とも収録数はさほど増えていない。ver.2 と ver.3 では、収録されているデータそのものに大差はなく、あいかわらず割ほどの熟語が未収録のままである。

だから、ver.3 でコードセパレート文字、「内」「説」などを検索してもヒットしない。簡体字も同様な。このため、検索したい字句をタイピングで入力する場合は、繁体字中国語の IME を使う必要があるし、C&P 入力でもユーザー自身がコードセパレートなどについて十分理解しておく必要がある。

このように、折角 Unicode に対応しながら、そのメリットを生かし切っていないのは残念なことだ。次期バージョンでは是非とも、簡体字や異体字の見出し字や熟語を収録して、紙版の完全再現を目指すとともに、異体字一括検索機能なども実装して欲しいものだ。

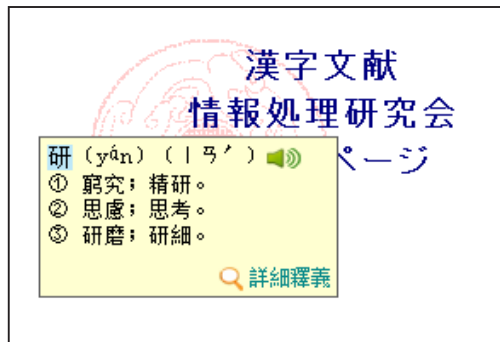
## ❖ 充実の検索機能

前にも触れたように、ver.3のメインインターフェイスは、前バージョンと大差ない。すなわち、左ペインで単漢字・熟語・成語のタブを切り替えて検索し、右ペインに表示された辞書本文を閲覧することになる。しかし、フォントサイズの切り替えや、見出しの文字コード番号などの情報表示に対応するなど、細かい使い勝手は上がっている。辞書本文の表示も、語釈は青、用例は黒に色分けしており（もともと、両者が入り乱れている場合には上手く色分けできない）、字句の提示画面もすっきりと見やすくなった。

ver.3で新たに加わった機能に、ポップアップ検索がある。IEやOfficeアプリなどの画面上で、漢字や熟語にカーソルを合わせると、その簡単な語釈をポップアップ表示してくれる機能だ。この機能自体は、既に多くの翻訳ソフトなどで実装されていて目新しくないし、表示されるのも語釈だけで用例は無いのだが、しかし、ウィンドウ右下の「詳細釋義」をクリックすることで、ver.3のメインウィンドウに当該字句を呼び出すことができる。つまり、一々検索語を入力して実行ボタンをクリックするという手間なしに、Webを閲覧しながら、あるいはワープロ上で翻訳やレジメを作りながら、随意に辞書を引くことができるのであり、作業の効率を大いに高めてくれる。

このポップアップ検索機能は、タスクトレイのアイコンを右クリックして選択することで、ON・OFFを切り替えられる。ONにしていると、パソコンのリソースがかなり食われてしまうので、作業時以外にはOFFにしておいた方がよさそう。

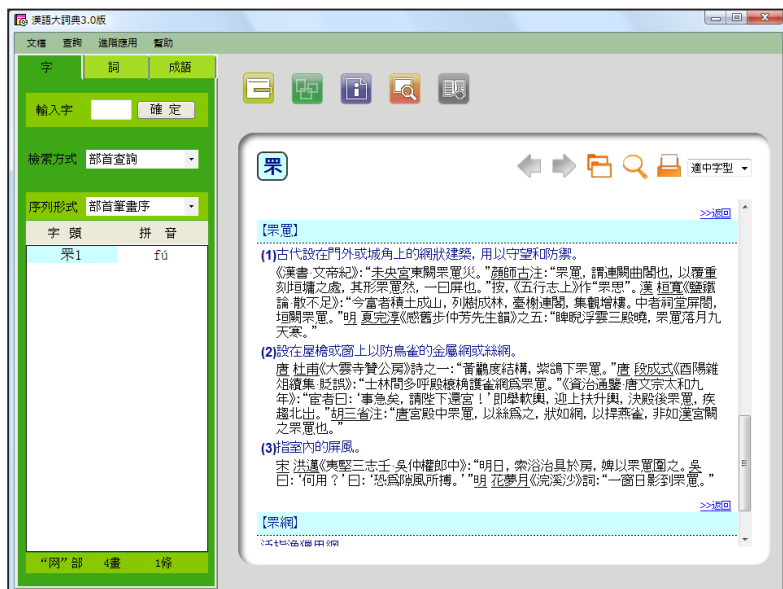
このポップアップ検索機能は、ver.3のメインウイ



ポップアップ検索

ンドウに対しても有効である。このため、『漢語大詞典』語釈中の字句の意味をただちに調べられることもできるし、また、項目から項目へとさまざまな関連性を見だしながら辞書内部を渡り歩くこともできる。なお、メインウィンドウ上では、画面上の字句を選択して上の虫眼鏡アイコンをクリックして、当該字句の項目にジャンプすることもできる。

総じて、『漢語大詞典』CD-ROM ver.3は、Unicode化による日本語環境での安定性向上、またポップアップ検索機能によるユーザビリティの向上など、前バージョンよりも格段に進化しており、収録文字種の拡大といった課題は残るものの、購入する価値は十分にあるものと考えられる。



# 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表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：長安の興衰
- 第2章 城市結構
- 第3章 数碼重建
- 第4章 明德門重建設
- 第5章 長安の創意
- 第6章 数碼長安——組合

第1章および2章は、長安城についての基礎的なデータを提供する。第1章では長安の自然、都市としての歴史を平易に概説する。続く第2章では、唐代長安城の都市構造（条坊制や市域の配置など）を概観した後、各地区（皇城、宮城、市、里坊など）ごとに詳細な解説がなされる。ここでは、各地区の社会的機能や社会生活一般にも触れられている。

第3章と4章が、本書オリジナルの価値を有する部分であり、著者の建築学研究者としての研究成果を文章の形で公表したものである。第3章は、都市景観や構造物をデジタルにより復原する際の方法論を中心に述べる。景観復原に際しては、長安の人口分布まで考慮されており、景観や環境といった空間的部分の歴史的復原の難しさが窺える。また建造物の復原については、第4章で1章分のスペースを割り、羅城門に当たる明德門の例を詳細に採り上げる。

第5章は、趣を変え、都市構造の比較史的研究に充てられている。『周礼』考工記に基づく伝統的な都城の思想から、日本、渤海など同時代の、元大都や清北京など後世の都城との比較が行われている。これらから導き出された各種の数値は、当然、デジタルによる復原研究に利用されている。

これら第1章から5章までを踏まえて、第6章では研究の総括が行われている。この研究の意義は、建築学の立場から「唐代長安城の設計過程の研究と他地域や他の時代の都城との比較研究により、比較相対的

図1：附録 DVD-ROM



かつ合理的な都城の構造の枠組みを推定でき」たこと、これら建築学に根ざした理論と文献史学、考古学との融合による研究成果が出せたことに尽きるであろう。

❖ 附録 DVD-ROM の内容と使用法

次に附録 DVD-ROM (以下、DVD と略) の内容について見てみよう。メインメニューには以下の4種が用意されている。図2のように実際のトップ画面には、序数表記はないが、いま便宜的に(1)~(4)の数字を振った。

- (1) 唐長安の数碼重建中文版
- (2) 唐長安の数碼重建英文版
- (3) 唐長安の数碼重建屏保
- (4) 唐長安の数碼重建图片欣赏

(1)(2)がこのDVDの主たるコンテンツである。以下、本稿では中文版を採り上げて解説するが、英文版も内容は全く同じである。内容は文字情報と音声情報(音楽および文字情報の読み上げ)、画像情報(動画、ムービー)から構成されている。

(1)をクリックすると音楽とともに影像が流れ、基本メニュー画面に行き着く(影像はスキップすることもできる)。

基本メニュー画面には、右下に概述、菜单、主页の選択アイコンがある。ここで概述を選ぶと、長安城の歴史、構造などが概説される。この内容はほぼ本編の第1章および5章をより簡略化したものとなっている。

菜单で詳細メニュー画面に移ると、図4のような画面となる。

ここでは、長安城内から13地点が採り上げられている。これらの地点は、政治、経済、社会、宗教に関する重要な場所で、都市長安の実態を表すのに相応しい場所が選択されている。この13箇所の詳細を以下に記す。

- ①麟德殿、②含元殿、③皇城和宮城、④寺廟群、⑤仏殿(西明寺)、⑥朱雀大街、⑦明德門、⑧東市、⑨小雁塔、⑩永寧坊(人口過密地域)、⑪靖安坊(人口密度中程度の地域)、⑫蘭陵坊(人口希薄地域)、⑬大雁塔

※ 数字は便宜的に付与した



図2: トップ画面(仮称)



図3: (1)の基本メニュー画面(仮称)



図4: 詳細メニュー画面(仮称)

①、②、⑤、⑦は歴史上の重要性に加え、近年の発掘調査により遺構の概要が明らかになっており、⑨や⑬は当時の建築物が残存している。このため、これらは、都市構造物の復原研究と言える。

これと対照的なのは、⑩から⑫である。この部分は、人口の多寡により、対象地点を設定しており、非常に興味深い。これらは、人口という要素を取り入れた、

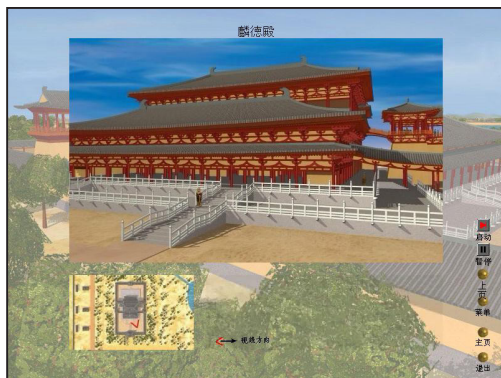


図5：麟徳殿の復原影像



図6：図5の俯瞰図部分の拡大図

都市景観、都市環境の復原研究なのである。

各地点の詳細画面では、原則として、当該地点について文字と静止画像、続けて動画により概要説明がなされた後、CGによる三次元復原影像が流される。このCG動画の例を図5に示す。図5は、麟徳殿といい、長安城の宮城北西側の大明宮における最大規模の宮殿である。

CG動画は、宮殿前の門をくぐり、麟徳殿を一周する様子を体感できるような構成となっている。同じ画面の下部には俯瞰図があり、視線の向きをくの記号（開いている方が視線の方向）で示す。このおかげで、視聴者は平面的な位置関係を同時に押さえることができる。図6として俯瞰図の部分拡大したものを例示しておく。

なお図2の残りのメニューのうち、(3)は前述のCG画像を利用したスクリーンセイバー（動作にはFlash Playerが必要）である。また(4)をクリックすると、画像が格納されているフォルダが開き、画像を直接閲覧・利用できるようになっている。

## ❖ 若干の問題点など

前章までで、本編およびDVDの特徴と簡単な使用

法について述べてきた。ここでは使用上の注意点や問題点などを示して、読者の便宜を図りたい。

まず、全体的な点では、本編、DVD、外箱のいずれにも動作環境についての言及が無い。そればかりかDVD操作のマニュアルやヘルプも無い。

OSについて言えば、筆者が今回試した限りでは、WindowsXP SP2 日本語版、Windows Vista Business 日本語版では問題なく動作した（WindowsXP SP3では確認していない）。ただし、端末によっては、CG動画が再生できないことがある。多くは「圧縮解除プログラム'vids:IV50'が見つからないため、ビデオを利用できません。」とのエラーメッセージが出る。この場合、WindowsXP SP2であれば、以下の手順をとられたい。

1. スタート→ファイル名を指定して実行
2. 名前欄に regsvr32 ir50\_32.dll と入力して、OK ボタンをクリック
3. ir50\_32.dll の DLLRegisterServer は成功しました。とメッセージが出れば処理終了

Windows Vista の場合は、「ファイル名を指定して実行」がデフォルトではスタートメニューに表示されないため、以下の処置をまずする必要がある。

- ① タスクバーを右クリックしてメニューからプロパティを選択
- ② 「タスクバーとスタートメニューのプロパティ」ダイアログの「スタートメニュー」タブをクリック
- ③ 現在使用しているスタートメニュースタイルの「カスタマイズ」ボタンをクリック
- ④ 一覧内の「ファイル名を指定して実行コマンド」にチェックを入れる
- ⑤ 各ダイアログをOKで終了させる

Windows XP SP1 や、上記の方法でもうまく動作しない Windows XP SP2 および Windows Vista の場合は、Microsoft Windows Update から修正プログラムを入手しなければならない。

附録とはいえ、DVDをセールス・ポイントにしている以上、動作環境は、顧客が商品を選択する上で欠かせない情報である。またマニュアルやヘルプは使用



者への利便性向上のため、最低限のものは必要だろう。再販のおりには再考されることを望む。

本編部分については、前述のように、よくまとまったものであるが、愛宕元氏や妹尾達彦氏をはじめとした、日本の長安城に関する研究への言及が無い点は問題だろう。この点、使用に際しては、愛宕元訳注『唐兩京城坊放』(平凡社東洋文庫,1994)の注釈および参考文献や、台湾中央研究院の公開する“隋唐五代長安研究文献目録外文部分”(http://www.ihp.sinica.edu.tw/~tangsong/papers/bibliography\_chang01.pdf)などにより必要な先行研究を補われたい。

ちなみに『唐兩京城坊放』は清の徐松の考証学的研究による、唐の長安と洛陽の復原である。唐代の長安・洛陽について歴史地理的研究をする場合に、まず繙かねばならない重要な書物と言える。本来であれば、本書の論評に当たっては、これら考証学的な研究成果についても、当然言及しなければならない。しかし、上記愛宕氏の訳注書に詳細な解題があるため、屋下に屋を架すことはしない。

また、長安の微妙な土地の起伏の差異による土地利用の相違や、長安城の都市計画における『周礼』以外の影響(例えば長安城のゆるやかな起伏が、『易』乾爻に基づくといった説など)への言及が無いのが気になった。景観復原を謳うのであれば、こういった自然地理的部分や、それに関係する思想にも配慮があってしかるべきではなかろうか。なお些末なことであるが、図版の出処一覧が、注や参考文献と同じ巻末ではなく、表紙折り返しにあるのは解りにくい。図書館などで、気づかずに再製本してしまえば、切り取られて解らなくなってしまう恐れもある。注に組み込むか、巻末に一覧とすべきであっただろう。

DVDについては、StarForceというコピープロテクトがかかっていることに注意されたい。このプログラム自体は、起動時にディスクが正規品かどうか確認するプログラムである。最初の起動時にはドライバのインストールのため再起動が必要となり、再起動後シリアル No. の入力を求められる。二度目以降は立ち上げるたびに確認の時間を要するのみである。

ただし Windows Vista Business では、UAC(ユーザーエージェントコントロール)の制御を停止しないとドライバがインストールできず、DVD が全く機能しなかった。UAC の停止方法は以下の通りである。

- 1) コントロールパネル→ユーザーアカウント→ユーザーアカウント
- 2) 「ユーザーアカウント制御の有効化または無効化」をクリック
- 3) UAC 有効のチェックを外して OK をクリック

UAC はプログラムの実行権限の制御をする機能なのだが、いろいろな情報を見る限りどうも多くの問題を孕んでいるようである。

また 1 のコンテンツにおいて、DVD の各画面では、主頁というアイコンがある。これをクリックすると、図 2 や図 3 ではなく、先に説明した図 2 から図 3 へ移動する影像の冒頭に移動する。コンテンツの大本である図 2 へ戻るためには、退出のアイコンをクリックしなければならない。このように、この DVD は、コンテンツの構成と各画面の位置づけが解りにくく、階層構造がとらえにくいことが難点である。

## ◇ おわりに

以上、簡単ではあるが、本書の内容とその使用方法、使用上の問題点について述べてきた。問題点といってもこれらは些末なことに過ぎず、本書の価値を貶めるものではない。

唐代長安の復原研究は、先述の徐松に始まり、現代に至るまで、考証学、フィールドワーク、考古学など種々のアプローチによる研究がおこなわれてきた。本書はこれら長期にわたる研究成果を建築学の立場から、集大成したものである。この点、唐代長安城の研究に関する最新の概説書、研究書としての価値を有している。また冊子体の本編に加え、附録 DVD によるデジタル情報の活用は、今後の歴史地理や考古学などにおける研究成果の公表方法に、影響を与える可能性があるだろう。

唐長安の数碼重建

A digital reconstruction of Tang Chang'an

王才強著

新加坡国立大学編

北京：中国建築工業出版社，2006.4

166p；23cm + DVD-ROM1 枚

ISBN：7-900189-28-9

## 『ICT を活用した外国語教育』

吉田晴世・松田憲・上村隆一・野澤和典 編著

CIEC 外国語教育研究部会 著

東京電機大学出版局 2008年7月 ISBN978-4-501-54400-3 3,000円+税

金子 真也

### ❖ 理論編と実践編

本書はその名に違わず「ICT を活用した外国語教育」について幅広く扱ったものとなっており、最初の第3章までが理論編、後ろの第4章～第12章が実践編と大きく二つに分かれている。理論編第1章「ICTへの道」（歴史）でLLからICTに続いていくこれまでの発展の経過について触れ、第2章「外国語教育・学習モデル」では理論的問題を総括的に扱っている。第3章「外国語学習デジタルコンテンツ」はいわば「デジタルコンテンツ作成概説」ともいべき内容で、理論編に含まれているものの、実質的には実践編の一部といってよい。

### ❖ 具体的な実践編

従来後半の実践編では、ICTの外国語教育への活用事例が個別に紹介されている。第4章から第8章までが英語教育関係で、第4章で「留学シミュレーション型CALL教材」、第5章で「アバター・チャット・システム」、第6章で「VOA番組の利用」、第7章で「Moodleによる教材開発・管理」を扱い、第8章が「小学校英語活動教材による教材開発・管理」である。第9章はドイツ語教育だがLMSを扱い、第10章がフ

ランス語、第11章が日本語である。言語だけ入れ替えれば他の外国語にも応用の利きそうな事例が満載である。

第12章と第13章が多言語を扱う章だが、その内容はというと、第12章は三枝裕美氏による「長崎・言葉のちゃんぼん村」の紹介である。長崎外国語大学における外国語学習サイトのいわば入り口になっているのが「長崎・言葉のちゃんぼん村」で、三枝氏作成の中国語学習定番サイト「パンダと学ぶ中国語」へのリンクも張られている。もちろん「パンダと学ぶ中国語」の紹介もある。

第13章は清原文代氏による大阪府立大学外国語学習Podcastの紹介である。Podcast受信方法の詳細で具体的な記述が初心者にはうれしい。

第14章は上村隆一・清原文代・三枝裕美の3氏によるPodcastでの教材作成配信入門となっている。

### ❖ まとめ

漢字文献情報処理研究会会員としての立場で本書を通読したときに感じるのが、内容の実用性の高さである。本書に収められたいわゆる「横文字言語」での個別活用例の中にも、ヒントとして活用できる取り組みがいくつも取り上げられている。外国語教育に携わる方々にはぜひとも一読をお薦めしたい書である。

## 『キーボード配列 QWERTY の謎』

安岡孝一 / 安岡素子 著

NTT 出版 2008 年 3 月 ISBN978-4-7571-4176-6 2,940 円+税

山田 崇仁

著者は、日本における文字コード関連の論壇の中心におられる方である。文字コードに関するご夫妻の著作については、『文字符号の歴史——欧米と日本編』（共立出版・2006 年）『文字コードの世界』（東京電機大学出版局・1999 年）を参照していただきたいが、今回の著作は、コンピュータに文字を入力するために欠かせない道具である「キーボード」の文字配列についての歴史である。

キーボードの文字配列の歴史は、巷間「今標準のコンピュータの QWERTY 配列はタイプライターのそれをまねたものである」「タイプライターの配列が QWERTY になったのは、タイピストがなるべく打ちにくいようなキー配列をデザインしたためである」とする説が唱えられてきた。かくいう私もそれを信じていた一人である。しかし著者は、後者を「都市伝説！」と断ずる。それを証明するために、その淵源であるタイプライターの歴史からひもとくのである。

結果として、本書は歴史の本となった。無論、著者のお二人とも専門の歴史家ではない。しかし、既存の歴史家がこの問題に関するコメントとして上記の（氏に言わせれば）都市伝説を標榜していることが許せなかったのだ。そう、歴史家ならば史料を丹念に調べ、それを時系列に並べ、その結果から結論を導き出すべきなのだ。

この本の内容はまさにそれを実践した結果として執筆されたものである。コンピュータのキーボードはタイプライターのそれが淵源である。だったら、そこから調べてみよう…

そして、それは結果として、アメリカ合衆国の経済史の一コマが映し出すものともなったのである。タイプライターの歴史上、多くの特許が発明され、それを独占しようと蠢く面々、その中での買収による市場の独占やそれに対する抵抗等々が絶え間なく起こり、

様々な人々が登場して消え、規格もまた同様に栄枯盛衰を遂げてゆき、現在に至る。

著者は本書の最後にこう慨嘆する。

QWERTY 配列の歴史は、結局、生産者側の押しつけの歴史に他ならない。

そう、キーボードの配列を決めたのは、単なる技術的制約を回避するための解決策ではなく、経済的・政治的な理由が根源にあったのだ。現在もおアメリカを中心に世界が振り回されているこの国の経済そのものこそが、キーボードの配列を結果的に規定することになってしまったといえることができるだろう。

本書は徹底的な調査から生まれ出たものである。この本の底には、「歴史を対象として文章を書くもの、決して二次三次の情報に惑わされず、徹底してよりオリジナルソースに近い情報源に当たるべし！」という著者のコメントが含まれているように思う。歴史を主たる研究領域にしている筆者としては、実に耳が痛い話であり、肝に銘じてゆくことにしたい。

著者の一人、安岡孝一氏はとことんまで調査をしないと気が済まない人であるらしい。以前、氏と「あきはばら」はいつ「あきはばら」になったのか？」について話した記憶があるが、貨物鉄道関連の史料をあさればいいのかあ～ 位しか答えられなかった記憶がある。その後、氏の調査がどうなったのか気になるどころだが、氏に言わせれば「気になったら自分でとことんしらべたらええやん」というところなのだろう。

本書についても同じだ。この本の巻末には膨大な史料一覧が掲載されている。これは要するに、「この材料でこの結論をだしただけだ。読者はこの本を孫引きするのではなく、自分でもこの材料+αを確認してからモノをを言おうね」という著者からのメッセージになっているのである。

# 漢字文献情報処理研究会彙報

2007.10～2008.9

2007年10月1日

会誌『漢字文献情報処理研究』第7号出版。

2007年12月22日

第10回大会、2007年度総会開催。

2008年7月19日

「ワークショップ：文字——（新）常用漢字を問う——」協賛。

## 第10回大会・2007年度総会

日時：2007年12月22日（土）

会場：二松学舎大学・九段キャンパス（東京都千代田区）403教室

共催：二松学舎大学21世紀COEプログラム「日本漢文学研究の世界的拠点の構築」

### ■ 第10回大会（13:00～16:50）

#### 1. 開会挨拶（13:00～13:30）

町泉寿郎（二松学舎大学21世紀COEプログラム事業推進担当者）

#### 2. 前半の部（13:30～14:30）：

- ネットピア、カ・フォスカリ大学とのネットによる漢文授業・研究会の実施報告  
山辺進（二松学舎大学21世紀COEプログラム事業推進担当者）・川邊雄大（二松学舎大学21世紀COEプログラム研究員）
- 二松学舎大学21世紀COEプログラムにおける日本漢文学コンテンツ公開構想について  
上地宏一（二松学舎大学21世紀COEプログラム研究員）・町泉寿郎（二松学舎大学21世紀COEプログラム事業推進担当者）

#### 2. 後半の部（14:50～16:50）

- 中国古典戯曲のデジタル化をめぐる  
千田大介（慶應義塾大学）・師茂樹（花園大学）
- 唐代行政地理情報のデジタル化についての諸問題

山田崇仁（京都大学人文科学研究所21世紀COE研究員）

- 簡繁変換と電子テキスト  
前寺正彦（日本IBM）

### ■ 総会（17:10～17:40）

- 2007年度事業報告、会計報告（会計監査：守岡知彦・山下一夫）
- 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（会計担当）：小島浩之  
幹事（名簿担当）：佐藤仁史  
幹事：田邊鉄  
幹事：千田大介  
幹事：二階堂善弘
- 2008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
- 会員の登録情報の更新・会費納入のお願いについて

## 2008年協賛講座

### ■ 協賛講座

題目：ワークショップ：文字

——（新）常用漢字を問う——

日時：2008年7月19日（土）13:30～

会場：花園大学（京都）自適館300教室

- 當山日出夫「（新）常用漢字表」（案）報道と現代社会の文字
- 師茂樹「一般キャラクタ論としての文字論の可能性」
- 安岡孝一「（新）常用漢字と人名用漢字と文字コード」
- 高田智和「常用漢字と「行政用文字」」
- 小形克宏「（新）常用漢字は本当に必要か」
- 全体討論

## 著者紹介

秋山 陽一郎 (あきやま よういちろう)

1975年生まれ。立命館大学文学部、同・大学院文学研究科で中国古代文献学を研究後、京都大学人文科学研究所の21世紀COEプログラムで、技術補佐員(2004年)・特別研究員(2005年)として唐代ナレッジベースプロジェクトに従事。2006年、花園大学非常勤講師。博士(文学)。

石井 公成 (いしい こうせい)

1950年東京生まれ。早稲田大学大学院人文科学研究科後期課程(東洋哲学専修)単位取得退学。駒澤大学仏教学部教授。博士(文学)。漢字仏教文化圏の宗教・思想・歴史・文学その他を研究。著書に『華嚴思想の研究』など。

内田 慶市 (うちだ けいち)

1951年福井市生。大阪市立大学大学院博士課程修了。博士(文学)。関西大学文学部教授。専門は中国語学、特に最近は近代東西言語文化接触研究。著書に『Macで中国語』(共著)、『近代における東西言語文化接触の研究』(単著)などあり。

江上 敏哲 (えがみ としのり)

国際日本文化研究センター資料課図書館職員。2008年3月までは京都大学附属図書館情報管理課に所属。2002年から約3年間、京都大学電子図書館「貴重資料画像」での古典籍資料の撮影・デジタル画像公開事業において、企画と実施に携わった経験がある。2007年3月から1年間、実地研修としてハーバード大学イェンチンに滞在。詳細は京都大学図書館機構webサイト「ハーバード日記：司書が見たアメリカ」(<http://www.kulib.kyoto-u.ac.jp/modules/wordpress/>)にて報告している。

岡本 真 (おかもと まこと)

1973年、東京都生まれ。国際基督教大学卒業後、編集者、校正者、翻訳家等を経て、1998年、ACADEMIC RESOURCE GUIDE (<http://www.ne.jp/asahi/coffee/house/ARG/>)を創刊。現在はインターネット関連企業に勤務。著書に『これからホームページをつくる研究者のために——ウェブから学術情報を発信する実践ガイド』(築地書館、2006年8月)。最近は図書館のウェブ活用の推進に力を入れているが、そろそろ自分の本当にやりたいことに専念したいと思いつつある。

小川 利康 (おがわ としやす)

1963年東京生まれ。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(1)(2)(3)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年東京生まれ。龍谷大学法学部教授。中国古典が専門だったが、最近はおかのことに関わりすぎているのを自覚している。

上地 宏一 (かみち こういち)

1976年生まれ。東京工科大学メディア学部助教。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・運営する自由な漢字字形Webデータベース「グリフウィキ <http://glyphwiki.org/>」は本号が出回るころにちょうど1周年を迎えているはずである。

## 小島 浩之 (こじま ひろゆき)

1971年、岐阜県生まれ。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。専門は東洋史学だが、本務としては図書館情報学、なかでも資料保存を得意分野?とする。それぞれ年に1本づつは、論文を発表することが当面の目標。近著に「唐代エリート官人の昇進経路の形成とその展開」(『東洋文化研究』10)、「日本の図書館等における蔵書の状態調査—その歴史と方法論」(『現代の図書館』46、矢野正隆氏と共著)などがある。

## 清原 文代 (きよはら ふみよ)

1964年大阪府生まれ。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。専攻は中国語教育、Podcastによる教材配信などモバイルラーニングが近年のテーマで、iPodと電子辞書が好物。最近の著作は『リズムで学ぶ三文字中国語 iPod徹底活用』(アルク、2007年)、『ICTを活用した外国語教育』(共著、東京電機大学出版局、2008年)など。Webサイトは<http://www.las.osakafu-u.ac.jp/~kiyohara/>

## 齊藤 正高 (さいとう まさたか)

1970年愛知県生まれ。地元の工業高校電気科卒業後、機械メンテナンス会社勤務をへて、愛知大学哲学科卒業、同大学院中国研究科博士課程単位取得満期退学。専門は中国哲学、目下の興味は方以智の自然学と思想。岐阜大学・愛知大学など非常勤講師。論文、「光肥影瘦論に就いて」(『東方学』104)、『『東西均』の反因説と水循環論』(『日本中国学会報』59)など。文字処理との出会いは20数年前の「ポケコン」(メモリー約500byte)からで、プログラムはノートに手書きして保存していた。  
<http://www.saitoma.net>

## 佐藤 仁史 (さとう よしふみ)

1971年愛知県生まれ。滋賀大学教育学部准教授。専門は中国近現代社会史。最近は江蘇・浙江農漁村におけるフィールドワークにはまっている。最近の研究成果に、太田出・佐藤仁史編『太湖流域社会の歴史学的研究』(汲古書院、2007年)、佐藤仁史・太田出・稲田清一編『中国農村の信仰と生活』(汲古書院、2008年刊行予定)がある。

## 志水 一行 (しみず いっこう)

1983年長野県生まれ。花園大学大学院文学研究科修士課程日本史学専攻修了。現在は花園大学歴史博物館研究員。専門は日本美術史。

## 白須 裕之 (しらす ひろゆき)

1959年、神奈川県生まれ。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)。現職は東京大学大学院人文社会系研究科次世代人文学開発センター研究員。現在は人文学における情報・知識・意味論に興味を持つ。

## 鈴木 慎吾 (すずき しんご)

1973年生まれ。大阪外国語大学大学院博士後期課程修了。博士(言語文化学)。現在、京都産業大学外国語学部助教。専攻は漢語音韻学、小学一般。

## 千田 大介 (ちだ だいすけ)

1968年東京都下生まれ。慶應義塾大学経済学部准教授。専門分野は、ここ一千年くらいの中国の大衆文化とその流通。著書に『北京なるほど文化読本』(大修館書店、共編)、訳書に『Chinese Culture Review』vol.1~5(好文出版、共監訳)など。  
<http://wagang.econ.hc.keio.ac.jp/>

## 照井 武彦 (てるい たけひこ)

1933年岩手県生まれ、東北大学工学部通信工学科卒、国産初期型電子計算機の開発、大学共同利用計算機システム、マイクロコンピュータシステムの研究などに従事、国立歴史民俗博物館の創設時から、博物館資料および日本史(歴史、考古、民俗)のデータベースを担当、1999年定年退職、同館名誉教授、盛岡市在住。

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「やまもも書齋記」

<http://yamamomo.asablo.jp/blog/>

[htoym@kcn.ne.jp](mailto:htoym@kcn.ne.jp)

二階堂 善弘 (にかいどう よしひろ)

1962年東京の下町生まれ。東洋大学文学部卒業、早稲田大学大学院文学研究科博士課程単位取得退学、博士(文学)。但し、その間は大学院と会社を行き来し、貿易事務やシステムエンジニアなども経験。東北大学大学院助手・茨城大学人文学部助教授を経て、現在、関西大学文学部教授。専門は中国の民間信仰研究で、著書に『封神演義の世界』(大修館書店)や『中国の神さま』(平凡社新書)などがある。実はベース弾きで、ギャルゲーおたくである。

サイトは「電気漢文箱」

<http://www2.ipcku.kansai-u.ac.jp/~nikaido/>

師 茂樹 (もろ しげき)

1972年生まれ。花園大学准教授。東アジアの仏教論理学(因明)の文献のデジタル化について、いろいろ模索中。

<http://moromoro.jp/>

山崎 直樹 (やまざき なおき)

1962年生。関西大学外国語教育研究機構所属。専攻は中国語学および中国語教育。言語学的知識をどう構造化して言語学習者に伝えるかが、最近の興味を中心である。目下、文構造の視覚化の方法や、学習者向け中国語辞書の記述方法を研究中。

山田 崇仁 (やまだ たかひと)

1970年、愛知県生まれ。立命館大学院文学研究科修了。博士(文学)。現職は立命館大学非常勤講師。現在は先秦諸子百家文献の成書時期や地域を弁別する研究が主テーマのはずだが、なぜか文字運用の歴史にはまっている気がする。<http://www.shuiren.org/>

## 編集後記

『漢字文献情報処理研究』第9号をお届けする。  
毎年のごとく、編集子の怠慢で編集作業が滞ってしまっている。それでも今年は前号のように、入稿後に出張先から再訂正を迫られるような事態にはならないはずなので多少はましと言うところか。

今号は、いつもと違う顔ぶれの方々に原稿を依頼した。いただいた原稿を拝読すると、「ああ、なるほど」とか、「こういう見方があるんだなあ」とか、非常に新鮮な知的刺激を受けた。刊行前にあれこれ原稿を読んで悦に入るのは、編集担当の数少ない役得である（もう一つ上げるとすれば、ギリギリまで締め切りを延ばせるところか…）。読者諸賢も編集子のような新鮮な体験ができれば幸いである。

本誌の刊行は、好文出版の尾方社長を始め、漢情研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。(♪)

## 漢字文献情報処理研究 第9号

発行日 2008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 真也  
上地 宏一 小島 浩之  
佐藤 仁史 田邊 鉄  
千田 大介 二階堂善弘  
師 茂樹 山崎 直樹  
デザイン 睡人亭：<http://www.shuiren.org/>  
DTP  
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。

ISBN978-4-87220-123-9

C3004 ¥2000E



9784872201239



1923004020005