



つくばね vol.29 no.4

目次

- 1 知的資源の基盤整備と活用
- 3 開学30周年特別企画「本学図書館所蔵の貴重書」
殿村本『水滸後伝』：識語が伝える本書の来歴
- 5 Ask Us としょかんミニガイド
- 6 開学30周年特別企画「活字と歩んだ筑波大学の30年」
筑波フォーラムのあゆみ
- 8 筑波大学図書館実務研修を終えて
- 9 本学教官寄贈著書紹介
- 10 私の一冊
- 11 掲示板
- 12 とびっくす

知的資源の基盤整備と活用

板橋 秀一

1. 電子ジャーナルの保存

20世紀を一言で表現すれば機械文明の時代といえると思うが、それに対して21世紀は情報あるいは「知」の時代になるといわれている。図書館は「知」の宝庫であるが、今後もそうあり続けることができるだろうか。現在進行している電子化をどのように進めていけばよいのであろうか。

電子化によって図書および図書館の利用形態が変化することは避けられない。図1に示すように、印刷雑誌の場合は図書館で雑誌を購入してそれを保管し、利用者が借りたり閲覧をしたりしている。それが電子ジャーナルになると、もちろん図書館が介在することはできるが、原理的には、利用者が出版社の電子ジャーナルアーカイブに直接アクセスすることが可能である。

この方式が広く行き渡ったとき、仮に出版社が倒産したような場合、電子ジャーナルのアーカイブを誰が維持・保存するのであろうか。印刷雑誌の場合は、印刷した冊子が世界中の利用者の手元にあるので、出版社が倒産してもその雑誌は図書

館や個人の手元に残り、利用できる道が残されている。しかし、電子ファイルの場合は誰かがどこかで電子ファイルを維持しないとデータを読み取ることができなくなってしまう。その点から、電子ジャーナルの場合も、やはり図書館で電子ファイルを購入して保管しておく必要があるように思

印刷雑誌			
出版社	図書館		利用者
電子ジャーナル			
出版社	(図書館)		利用者

図1 雑誌の利用形態の変化

われる。幸い、現在は冊子体の雑誌を購入しないと電子ジャーナルの契約ができないことが多いので、冊子体での保存が可能であるが、将来、冊子体なしで電子版だけのジャーナルが出版されるようになると、このような問題が生じてくることが予想される。もちろん、このような事態を避けるために種々の方策が考えられてはいるようであ

る。

一方、記録媒体についても、以前はオープンリールの磁気テープやカートリッジテープをよく見かけたが、近頃は殆ど目にするのがなくなり、最近ではハードディスク、CD-ROM、DVD等新しい媒体が次々と利用されるようになった。古い媒体に記録されたデータは、新しい媒体に移していかないと読み取ることができなくなってしまふ。また、媒体が変わらないとしても、情報の記録・読み取り方式が変わった場合にはやはり似たような問題が生じる。このように、電子化データの場合は、情報の可読性を維持することが重要な問題になってくる。

2. 大量音声データの検索

最近では電子メールなしでは大学の研究・教育が殆どできないのではないかとさえ思えるほどである。特に外国との連絡の場合は電話と違って相手が不在でもよく、時差を気にせずじっくりと考えながらメールを書くことができるという点で、外国語を話すことが不得意な人にとっては便利なものである。しかし、人間同士のコミュニケーションの基本はやはり音声言語による会話である。音声自動認識と言語自動翻訳それに音声合成技術を組み合わせれば、いま電子メールにより拙い英語で行っていることが、時差と不在の問題を除けば電話で容易にできるようになる。

テレビやラジオのニュース、インタビュー、学会等の講演、学校の授業なども音声で行われている。最近公開されたNHKアーカイブスには大量の(音声付)映像データが収録されている。このように音声を録音しておけば後々利用できるわけであるが、音声のままではどこに何が入っているかが分からないので、後で利用するのが難しい。そこで、音声を自動認識して内容を文字化し、検索できるようにして利用の便宜を図ることが求められるようになった。第2次大戦中の体験のインタビューを音声認識して文字化するプロジェクトがアメリカを中心として進められており、ヨーロッパでは、過去の膨大な放送記録を音声認識によって文字化しようというプロジェクトが進められ

ている。

3. 音声情報処理とデータ

音声自動認識や音声合成の研究が進展し、その一部は実用化されるようになったが、任意の人が発声した会話音声の自動認識あるいは任意の人の声で音声合成を行うことのできるシステムの開発は今後に残された大きな課題となっている。例えばアナウンサーが原稿を読み上げているような音声なら95%程度の精度で自動認識できるが、自由に発声した音声については70%程度になってしまうのが現状である。

音声研究を進める上で音声データが必要なことは言うまでもない。その音声データは、多種多様(性別・年齢・方言・人数等)であることが求められる。最近では、各種の統計的手法の発達により、大量の音声データが音声処理システムの学習のために必要とされるようになった。従来は各研究者が、必要に応じて音声データを収録し、保管・利用していたが、より体系的なデータ収集とその効率的な利用が求められるようになった。

一方、音声情報処理システムの研究・開発を行うためには、分析・合成・認識の各種の手法を適切に比較・評価することが必要であるが、これを行う方法としては現在のところ、共通の音声データを用いてこれらの処理を行い、その結果を比較するという方法以外は知られていない。

このようなことから、共通利用可能な各種・大量の音声データを収録し、保管・公開することは研究・開発過程での利用および認識装置の性能評価の両面から求められている。このような目的に利用される音声データを一般に「音声データベース」あるいは「音声コーパス」と呼んでいる。音声情報処理の分野で「データベース」というときには、データベースシステムよりも「大量の音声データの集積」そのものを指すことが多い。そのため最近では、それを意味する「コーパス」を使うようになった。音声コーパスの必要性やその意義については近年広く認められるようになってきた。

これまで情報処理研究は情報の内容よりも容器

を対象としてきたが、最近ではコンテンツという用語がしばしば用いられるように、次第に情報の内容の扱いに目を向けるようになってきた。音声コーパスの構築は、その一つの現れと見ることもできる。

なお、上述のことは自動翻訳に代表される自然言語処理研究と言語（テキスト）コーパスや画像処理分野の研究と画像データ等についても同様に考えることができる。

4. 音声・言語データの活用

日本では(社)日本電子工業振興協会(現在、電子情報技術産業協会)、ATR音声言語通信研究所、(社)日本音響学会、文部(科学)省重点領域研究やCOEプロジェクト等によって、各種の音声・テキストコーパスあるいは電子化辞書等が構築されてきた。このように個別のデータや知識の電子化はある程度進んでいるといえる。これをさらに発展させて、大規模知識資源の構築と体系化さらにはその活用基盤を整備しようというプロジェクトが日本でも始められている。

音声・テキストコーパス等の知的資源の収集・

構築・保管・配布を行う機関として、アメリカでは90年代の初めに言語データコンソーシアムが、欧州ではヨーロッパ言語資源協会が90年代半ばに設立され、音声・自然言語処理研究に多大な貢献をしてきた。21世紀に入ってから、韓国や中国にも同様の組織ができている。一方、日本では音声コーパスの整備は比較的早く始められたが、上記のような組織の整備は遅れて、99年に予算の裏づけのないまま言語資源協会が設立され、2003年にNPOとして再発足してこれから活動を開始しようとしているところである。

このような動きは図書館とは少し違った方向を向いているように見えるが、知的基盤の整備という点では同じ方向を目指している。音声・テキストに限らず、画像・映像を含めた知的情報資源の整備とその活用基盤の構築がこれからますます重要になってくるものと思われる。図書館がそのようなものどのかかわっていくのか、考えていく必要があるのではないだろうか。

(いたばし・しゅういち 電子・情報工学系教授)

開学30周年特別企画「本学図書館所蔵の貴重書」

殿村本『水滸後伝』：識語が伝える本書の来歴

大塚 秀明

洋の東西を問わず名作は続書を生む。『水滸伝』にも数種の続書があるが、続書の常として作品の評価はあまり高くない。しかし『水滸後伝』だけは例外である。むしろ『水滸伝』を越えることはないが、『水滸伝』で生き残った三十二人が遺児や二世や曾ての敵将と一緒に船で外国に渡る物語は出色である。とくにわが国では馬琴の『椿説弓張月』の素本として知られ、数ある中国の続書のうちで邦訳が2種類もあるのは『水滸後伝』だけである。そして本学の蔵本は、その馬琴と関係があるようだ。

『水滸後伝』(八巻四十回)は清初に書かれた長編小説。古宋遺民著とあり、序には明の萬曆の年号があるが仮託であり、明の滅亡(1644)後、遺民として生きた陳忱の作とされている。原刊本とい

われる英国博物院蔵本には「遺經堂蔵書」という書舗とともに清の康熙三(1664)年の刊年が記されている。原刊本はのちに「紹裕堂新刻」として重刻され、近年「古本小説集成」に華東師範大学蔵本の影印本が収められている。さらに後年には人物画像八葉を入れた有図紹裕堂刊本もある。この原刊本に対して蔡元放が改修を加えた乾隆三十五(1770)年の序を持つ刊本があり、所蔵も多数報告されている。今日広く流布しているのは改修本を排印したもので、亜東図書館本(1924)や上海古籍出版社本(1981)などがある。上述の邦訳2種とは森槐南が改修本を訳し、鳥居久靖が原刊本を訳した邦訳をいう。