INFORMATION AND KNOWLEDGE NEWS

情報知識学会ニューズレター

1995 2.1

30

情報知識学会事務局 発行 〒110 東京都台東区台東1-5-1 (凸版印刷㈱内) TEL03(3835)5692 FAX03(3837)0368 ISSN0915 1133

画像を読む

東京国立文化財研究所写真資料研究室 鈴木廣之

長年にわたって美術品を撮影した４×５インチのフィルムが万の単位である。これをデジタル画像に移して活用しようというのが私の職場での大きな課題だ。最近のコンピュータの性能なら不可能ではない。画像を圧縮してＣＤに収める。試みに電卓で必要枚数をたたいたら、表示された数字にうんざりした。当分は実験にとどめておこうと思う。

いまは必要なデータがパソコンにいれてあるので、フィルム検索に不便はない。でも画像データベースが将来できたときのことを考えると、文字から画像を検索するだけではつまらないと思う。どうせなら画像から画像を検索して未知の画像に遭遇する楽しみを味わいたい。似た画像、対照的な画像などいくつもあるにちがいない。もちろん専門家のあいだではとっくに研究テーマになっているだろう。いくつかの事例を耳にすることもある。だが、人間の眼の生理機能を画像検索のモデルにしようとする例にあたると、本当だろうかと思わず首をかしげてしまう。私の本業が美術史なのでこんなことを平気でいえるのかもしれないのだが。

絵巻物をしらべていると、画中のひとまとまりの人物がそっくりそのまま別の絵に転用されていることがよくある。当時の絵師たちは剽窃を少しも気にしなかったらしい。しかも物語の筋と無関係に切り取られているところをみると、彼らなりのパターン認識のしかたがあったようだ。彼らは切り取ったパターンに頭のなかでラベルをはりつけていたのではなかろうか。それを自在にあやつるのが絵師たちの腕の見せどころだったかもしれない。「絵はイメージだ」などというのは彼ら一流のウソで、裏でやっていたのはこんなことだろうと私はにらんでいる。

「体験談ほどあてにならないものはない」といわれればそれまでだが、私自身の研究の日常はどうやら絵師たちの逆をやっているらしい。徹底的に画像を分解し、部品化し、一つひとつに名付け

(次頁へ)

目　次

をする。全体から受ける印象を感じ取ろうとすることは少ない。「画像を読む」といった方がぴったりする。

コンピュータにこれをやらせようと思ったら、画像のなかに印をつけて、橋なら橋とキーワードをふってやればかなり役立ちそうだ。しかし、絵巻に登場する人物を「右手をななめ上へのばして、左手を頭の上に、右足は……」などといちいち書いていたらきりがない。山水画に描かれている山の形の特徴をコトバにするのはもっとやっかいだ。大雑把でいいのだから、形のもつ特徴をパターンに検出し、これをキーにして画像を検索できないだろうかと考えてしまう。

どんなアルゴリズムで動いているのか私にはわからないが、親近感をもつのは文字読み取りのＯＣＲである。ランダムな点の集合のなかから意味のある形を必死に読み取ろうとするソフトに、同業者によせる共感のようなものさえ感じることがある。もちろん意味を解読するのは人間の役目で、コンピュータは手掛かりになる候補を答えるだけなのは分かっているつもりだが、画像のなかから特定のパターンを選び出し、ラベルを張り付ける発想を生かせないものか。専門家のご意見をうかがいたい。

よくいわれるように人間の脳がおそろしく効率のよい画像処理を実行しているのはたしかだろう。だが、ふつうイメージ処理というとコトバと正反対のものに受け取られるのは、ひどい誤解のように思えてならない。美術史家は「線の美しさ」などということがあるけれど、よく考えてみればこれだって彼なりのパターンにどう分類できるのか、判断を言い換えているにすぎないと思う。それが「美しい」という言い方になるから惑わされるのだ。夢とはコトバから逆方向に再生されたイメージだという説もある。言語機能のなかに人間のイメージ操作のなぞを解くカギがあるように思うのだがどうだろうか。

学会カレンダー(Ver. 1.0, '95)

1995年3月27日～31日	EACL-95, 7th Conference of the European Chapter of the Association for Computational Linguistics. University College Dublin, Belfield, Dublin, Ireland. Contact: Allan Ramsay, Department of Computer Science, University College Dublin, Belfield, Dublin 4, Ireland. Phone:(353)-1-7062479, Fax:(353)-1-2687262. E-mail: allan@monkey.ucd.ie
1995年4月10日～13日	KB&KS'95, Second International Conference on Building and Sharing of Very Large-Scale Knowledge Bases, University of Twente, The Netherlands Contact: KB&KS'95, c/o Knowledge-Based Systems Group, Dept. of Computer Science, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands, Fax: +31 53 339605, E-mail: kbks95-org@cs.utwente.nl
1995年4月19日～22日	PACLING '95, Pacific Association for Computational Linguistics 2nd Conference. The University of Queensland, Brisbane, Queensland, Australia. Contact: Roland Sussex, Centre for Language Teaching and Research, The University of Queensland, Queensland 4072, Australia. Phone:+617 365 6896, Fax:+617 365 7077. E-mail:sussex@lingua.cltr.uq.oz.au
1995年5月29日～30日	Tutorial on Automatic Dictionary Making, Brussels, Belgium Contact: Mrs. V. Vienne, ILMH, CTB, 11 rue d'Arlon, B-1040 Brussels, Belgium

ニューズレター原稿募集（１９９４年４月更新）

１９９１年度より情報知識学会のニューズレターの発行が年６回になり鮮度の高いニュースを掲載しております。

つきましては、会員の皆様の原稿を募集します。内容は自由自在、"情報"を題材にしたものから、"情報"に関係の無いもの迄、特に指定はありません。

なお現在電子編集を行っておりますので一般原稿はフロッピィでお送り下さい。（出来ればドラフトとして出力原稿を添付して下さい。）また学会の研究会やセミナー等の案内については一部オフセット印刷を併用しております。ワープロ・Ａ４サイズで出力されたものを、そのまま御郵送ください。いずれの場合も原稿の長さは２段組で１ページ２０文字×４０行×２段（１６００文字）となります。段組なしの場合も１ページあたり１６００文字を目安としてください。

これまで通り、以下の記事は常時募集します。執筆ご希望、又はどなたか推薦したい方など御紹介下さい。

巻頭言（タイトル込み　４４文字×２２行）
研究紹介、人物紹介
会員の随想、書評
学会のニュース・カレンダー
対談記事・インタビュー、学会出席報告
関連学会の開催案内、国際会議紹介
会社紹介・情報関係開発商品紹介
役に立たないミニ情報・役に立つミニ情報

*** なお執筆者は現在のところ会員に限りますので、記事を書きたい方は「情報知識学会」への入会をお奨め下さい。
*** 法人会員の広告も掲載致します。編集委員に御相談下さい。
*** 締め切りは変わりません。これまで同様、発行前の奇数月１５日です。
*** 下記の事項は必ずフロッピィと一緒に文書としてお送り下さい（フロッピィには書き込まないで下さい）。

掲載希望日：第　　号　　年　　月　　日　発行

氏　　名：

連 絡 先：〒

Tel　　Fax

問い合わせ・原稿送付先

〒167　東京都杉並区上荻4-4-5-101
長瀬　真理
Tel: 03(3395)8168　Fax: 03(3395)8608

万葉集テキストデータの作成と流布

山口大学　吉村　誠

コンピュータ用の万葉集テキストが出来上がった。これは万葉集研究に従事するものにとっては長年の夢と言ってもよいであろう。例えば「妹背」という言葉がどの歌にあるかといったようなことが、たちどころに表示できる。

語句の検索は、従来は索引を利用するしか方法がなかった。もちろん索引が出来るまでは、全てを記憶に頼ったか、最初から1首づつ読んで探していたという苦労話も聞いている。万葉集は、古典文学の中でも源氏物語と並んで、詳しい索引などがあるもっとも整備の進んだ作品である。昭和28年万葉集大成の中に正宗敦夫による総索引が初めて姿を現した。それまでは各句索引が注釈書などに付けられていたことがあったが、各単語の検索が出来るものとしては画期的であった。これは品詞別になっているし、単漢字の所在も調べることが出来る。その後平凡社から独立して出版され、現在でもなお活用されている索引である。しかしこの索引は当時流布本として中心的存在であった寛永版本の本文や訓読を基準にしており、そのために現在の西本願寺本を底本とした校訂本文や訓読からは、時代にそぐわなくなってきている。

そういったこともあって、最近、新たに万葉集の索引が出版された。底本は西本願寺本であることはもちろんのことながら、作者などの付加情報も付けられている。もちろん作成にあたってはコンピュータが活用されている。しかし出版の版下となった関係か、或いは大型機でのデータ作成であるのか詳しい事情は不明であるが、データそのものは機械の中で眠ったままである。

コンピュータを利用しない人は、最近のこのすぐれた索引は有効なものとなるであろう。しかしコンピュータによる検索方法を一旦知ってしまった人にとっては、その索引がどんなに工夫されているとしても、コンピュータデータの活用には及ばないと判断してしまう。まず検索方法。我々はgrepなどのすぐれた検索プログラムの持つ正規表現というものを知っている。かざ[さしすせ]と検索語を入力することによって、風下（かざしも）や、飾る（かざる）という余計な言葉と区別して「かざす」とその活用形を検索できる。また「山」と「川」とが一首中に同時に出ている歌とか、「山」または「川」のどちらかがある歌とかといったものも機械が勝手に探し出してくれる。そればかりでなく、仮に誤植があったとしてもデータであればすぐに手直しが出来る。もちろん索引は出版されている以上、厳正な校正が施されているとは思うが、それでも古典的な正宗敦夫の総索引ですら、私の知る限り数例の誤植は認められる。

また必要な情報を後からでも付け足すことが出来る。作者、歌の質、詠まれた場所、といったようなもの。そしてそれ自身も検索対象になる。例えば羈旅であって望郷の歌で高市黒人が詠んだものという条件の歌を、たちどころにつかまえることが出来る。

何よりも違うことは、検索結果をファイルに格納して、そのまま論文へ引用するとか、注釈書で調べた付加情報を書き込むとか、分類して並べ直すとかが可能である。ノートへ手書きで歌番号を書き出し、それをまた書き換えてという旧来の方法からすれば、データ整理としては雲泥の差が出るであろう。もちろんそれがきらいという人は強制するわけではない。

万葉集テキストはあくまで機械で使用するということを前提にして作成している。いまさら索引を作るわけでもないし、ましてその版下用というわけでもない。しかも和歌集には、散文と異なって国歌大観番号による整理が存在する。それは特定の活字校本のformatに依拠する必要のないことを示している。このことは機械データとしてもそのまま本文テキストの属性と

して使用が可能である。ただ本テキストは画面表示を念頭に置いたために、短歌以外は1行80バイトを1レコードとして改行コードで区切っている。そのために特に長歌のand検索には支障をきたす。最近はパソコンでも画面表示の変更が可能であるので、このあたりのことはまだ改良の余地はあろう。

テキストは、原文、仮名漢字まじりの訓読文、完全な訓読を示した仮名書き文の3つのファイルをセットとして成り立っている。この3つのファイルを歌番号でリレーションすることによって、仮名訓読の検索語入力から原文を出力したり、原文の特定の用字の訓読を探し出したりということが可能である。或いはこの3種類のファイルを一つにしてawkなどで検索してもよい。また上にも述べたように、データ集を作成することによって、作者や歌内容から検索するということも出来るようになっている。リレーションには、grepやsed、awkといったすぐれたtool類が必要であったり、自分でプログラムを組むという作業が必要であるが、使い込むことによって、様々な利用が工夫出来るであろう。またデータそのものも自分で利用しやすいものに改変するということも許されている。実際に、このデータを見たユーザーの方が、原文と仮名書きとを対比した各句ごとの区切れデータをsedで作られ、提供していただいている。

本データの本文、訓読は、万葉学の成果を配慮したものの、特定の活字校本に依拠したものではない。従って、たびたび問題になる印字本との著作権や版権からは離れていると判断してよい。ただし流通過程における全くのfreeにはしていない。内容の責任上GNU準拠という形態をとっている。この条件を承認していただける方は、どなたでも無償で利用できる。人文系のコンピュータ利用の場合は、基礎となるテキストデータは必需品である。しかし古典文学に限って言えば、現在のところ共有できるデータは、印字本に比べて極めて少ない。またわずかに存在するものも印字本からの利用規制がかかっているものが多い。

出版文化によって、我々は多くのすぐれた古典作品を共有してきた。それと同様に今後、機械可読のテキストを多く共有することは、学問の発達にとって必須の条件であると考える。そういった意味で、本万葉集テキストが草分け的存在の一つとなれば、作成の労が報われたと言ってよいであろう。

おわび

12月 1日発行の第29号ニューズレターにおきまして、投稿者の方のお名前に誤りがありました。おわびして訂正いたします。正しくは以下の通りです。

正	誤
楢崎　久武	楢崎　久

ＣＵＩ環境とテキストデータベース

藤井 素彦(美術史)

次世代のオペレーティングシステムの登場が相次いでいる。既に発売されている OS/2 やリリースの待たれる CHICAGO など、いずれも MS-DOS を必要としない、独立した OS である。640kb というプログラムの主記憶領域の限定がなくなり、マルチタスクが実用的なレヴェルで実現されるだろう。今度こそ本当に、パソコンは変わろうとしている。

趨勢の向く所は、GUI(Graphic User Interface)一辺倒である。しかし、私たちのパソコンの使用環境に大幅な変更が強いられる事は未だ当分あるまい。どんなシステムが導入されようとも、DOS におけるテキスト形式は、今後も汎用的な規格として用いられるはずだ。また、特に文書入力の場合、キーボード上から文字を入力するというインターフェースは変わりようがない。

もとより、多くのユーザーがコンピュータを運用する目的は、文字データの処理に止まってはいないか。そして、その限りにおいては、現在の DOS ベースでの運用で大した不足はない(なくし得る)はずだ。パソコンは CUI(Character User Interface) 環境としては既に完成されている(完成され得る)のである。

ここでは、優れたテキストエディタである VZ Editor の使用例の紹介を通じ、CUI 環境でのデータ管理のあり方に僅かながら触れたい。VZ は恐らく DOS を使える道具にしてくれる最強のツールであり、そのインターフェースは CUI 環境の標準形として、今後も継承され得るものではないか。このソフトと MS-DOS の関係を通して浮かび上がるヴィジョンは、未だに豊かなものに思われる。

VZ の最も大きな特色として、非常に柔軟なマクロ機能がある。所定の文法に従ったプログラムを施す事で、文書管理・編集機能の大幅かつ細部に渡る変更・追加を可能とする。ユーザーの数だけ VZ は化けると言っても誇張にはなるまい。ネットワーク上にアップロードされる VZ 用のマクロプログラムは数多く、アマチュアプログラマーの裾野は文系・理系の区別を越えて幅広い。マクロは多種多様であり、論文の註番号を自動的に振ってくれるマクロや、文書内の数値の加減乗算をする表計算マクロなど、文字列を扱う限り、殆ど出来ない事はないかのようである。CUI における統合環境を構築し得ると言っても過言ではない。

このマクロによって DOS コマンドを登録しておき、キー一発で実行する事も可能である。例えば DOS のユーティリティに FIND.EXE というものがある。検索したい文字列が含まれる行をテキストファイルから抽出してくれるプログラムであるが、これを VZ との連携で用いる事が出来る。編集画面でキーワードを入力し、範囲指定する。特定のキーをタイプすると、キーワードが検索語としてパラメータに含まれた FIND コマンドがコマンドラインに送出される。作業終了後、抽出結果がリダイレクトでまとめられたファイルが、自動的に編集画面にオープンされる。

筆者は美術史の文献データを、[1956 PANOFSKY] や [1960 GOMBRICH] に始まる書式で蓄積している。このようなデータが FIND によって複数抽出されたとする。DOS には SORT.EXE というユーティリティもあり、これは行頭から n 文字目の文字順に各行を並べかえる機能を持つ。マクロによって、VZ の編集画面で「1」にカーソルをおいて起動すれば刊年順、「P」や「G」で起動すれば著者名順のソート結果を編集画面に渡すといった事が出来る。即座に文献表が出来上がる訳である。

FIND.EXE よりも更に利用性の高い検索ツール GREP は、フリーソフトウェアとして入手

出来る。様々な種類があるが、中でも YGREP は、正規表現によって大文字・小文字、全角・半角、ひらがな・カタカナを同一視する曖昧検索、二行に跨る検索、スペースを無視した検索が可能である。また、LHA による圧縮ファイルからも検索する WGREP は、AND 検索、OR 検索も出来る。また、UGREP は検索語のある行を含む複数行を、「#」等の「指定記号の間の行」という単位で抽出する、いわばカード型の検索を行ってくれる。

テキストファイル検索ツールが充実する中、VZ をデータベース化するマクロプログラミングが近年の流行となっているようだ。例えば、FEP の略語登録のような感覚で、複数行をも登録出来るマクロがある。より本格的なものでは、編集中の文書から任意の箇所を切り出して登録し、また検索結果を文書中に貼り込み、あるいは結果に編集を加え更新登録出来るものもある。そこでは、テキストの編集とデータの生成・蓄積とが一体になった環境が提供される訳である。こうした環境をテキストデータベースと言う事がある。

学術論文というテキストは、特定の文字列の再現率が比較的高い。基本文献の書誌データなどは、幾度も表れる文字列の好例だろう。そうした文書に関るデータをテキストと同じ形式で蓄積し、執筆環境とデータベースとが直結する事の意義は大きい。

論文執筆とは、例えば情報カードに表現されるように、気付きやデータの断続する線を連続的なものへと繋いでいく作業である。その過程で、或る箇所を取り敢えずペンディングしておいて、別な所での再利用を考える事はよくある。そのような「取り敢えず」のデータの取得に、執筆環境から離れ、データベースソフトを使う事は面倒である。あるいは意味がなく、更には不可能だろう。しかし、テキストデータベースにおいては、範囲指定の後に登録を行えば良いだけだ。分類の必要はなく、あらゆる文字を検索対象とし得る事で「項目」の概念が殆ど意味をなくす。

これは極めて重要な事である。項目とは、大規模にデータを管理する場合、構造的に要請されるものに他ならない。だが、常にその規模のデータを必要とするとは考えにくいし、参考書誌のデータ個々の項目数を一致させる必要などもないだろう。項目化されたデータ構造が支援し、あるいは規定するのは、ルーティンワークに止まる。我々の仕事の実際、更に言えば思考の自由な過程は、そこに収まるものではない。より必要なデータ形式は、最も汎用性と可塑性の高い形式、即ち単純なテキスト形式ではないか。真に我々の思考過程の反映となり得るのは、その他をおいてない。データベースにも変身する DOS テキストエディタの意義がここにある。

果たして私たちは DOS による CUI 環境を玩味し切ったと言えるのだろうか。DOS と VZ とを通じて見える未来もまたパソコンの未来であり、しかも手に届く未来である。

■DOS による簡易データベースについては、次の名著に教えられた。
林晴比古著『パソコン書斎整理学』(ソフトバンク、1990年)
■以下の二冊は数多いマクロを付属ディスクに収録する。CUI 環境をめぐるユーザーの熱気が伝わる本である。GREP も数種が収められている。
兵藤嘉彦他著『VZ 天国』(ビレッジセンター出版局、1992年)
見米快介編著『続 VZ 天国』(同、1994年)
■筆者の使用するシステムは以下の通りである。
【ハード】PC-9801 NS/L (メモリ8MB、HD80MB)このセットが中古で10万円前後/PC-9801 NL (メモリ8MB) 中古で7万円前後
【ソフト】MS-DOS Ver.5/[FEP]WXII+Ver.2.7 (エー・アイ・ソフト、定価¥9,800、WXIIIが近日中に発売される) /VZ Editor Ver.1.6 (ビレッジセンター、定価¥9,800)/[印刷ソフト]PRTII (同、定価¥9,800)

「日本語会話コーパスの構築と談話分析」プロジェクトについて

福岡工業大学　上村　隆一
九州共立大学　田吹　昌俊
国際基督教大学　村野　良子

1. はじめに

近年、ようやくわが国でも学術用データベース作成と共同利用環境づくりの必要性に対する認識が深まってきたように思われる。特に、技術情報の蓄積や論文検索を主体とした、どちらかというと理科系中心の利用環境から文学作品、歴史資料、文化財等の文科系寄りの利用環境へと広がりを見せてきたのが最近の特徴である。　また、欧米の研究資料を受動的に利用するだけでなく、独自のデータベースを作成し、インターネットを通じて国際的な共同利用を可能にしようとする能動的な気運も徐々に高まってきた。とりわけ、日本語・日本文化等に関するデータベース作成はわが国に対する国際社会の理解を助け、同時にわれわれ自身が自国の言語・文化を理解し、評価する際の基礎資料としても重要な意味をもつ。

2. 研究経過

本研究プロジェクトは、日本語の談話構造に関して、日本語母国語話者(以下NS)と非母国語話者(以下NNS)の発話に含まれる言い誤りの類型を比較分析することを目的として、1991年度より試験研究が始められた。日本語の会話分析は、まだ体系的な研究が少なく、とくに分析対象としての一次言語資料（以下コーパス）の絶対量が少ない。

そこで、われわれはまず、インタビュー実験形式による会話データの収集と、それに基づくコーパスの構築作業から開始することにした。これまでに、２大学（国際基督教大学、成蹊大学）の教職員および学生の協力によって、NS, NNS計15人の会話データをビデオテープに収録し（被験者１人につき30分程度）、テキスト転写、画像・音声のディジタル化作業を経て試作版コーパスを作成中である。現時点で、転写されたテキストデータはNSの部分だけで被験者１人当り約100KB（ただし注釈等を含む）、ディジタル画像データ（動画、１秒間３０コマ再生）は被験者１人当り200～300MB（圧縮ファイル形式、圧縮比１：８程度）に達しており、追記型光ディスク（CD-Recordable）に保存している。　また、このコーパス作成と並行して、繋ぎ語や代名詞類の使用実態の分析（後者については本稿末尾の実例を参照）を同時に進めており、それらの研究成果の一部は、第１次中間報告の形ですでに国際学会 (*2nd Princeton Japanese Pedagogy Workshop 1994*)等で発表してきた。

3. 研究内容の特色

本研究プロジェクトにおいて作成する日本語会話コーパスは、従来のテキスト・データベースと異なり、文字・画像（動画）・音声データを統一された使用環境(GUI)で同時に利用可能にする、いわゆるマルチメディア型データベースである。われわれのコーパスの主な特徴は、①分析対象が話し言葉である場合、文字化しにくい音調、強勢、ポーズなどの諸特徴をそのまま音声情報の形で提供できる。その結果、テキストデータに特殊な音調記号等を付与する必要がなくなる。②画像（動画）データを提供することにより、非言語情報（身ぶり、手ぶり、顔の表情など）をテキストと同時に利用できるようになり、会話の状況、話者の特徴、周囲の雰囲気などを分析の手がかりにすることができる。③画像・音声データ自身をディジタル化することにより、ランダム・アクセスが可能になるので、テキスト検索に加えて、画像・音声データの検索等を容易に実行できる、などである。

上記のコーパスの特性を十分に活用することにより、これまで研究対象から事実上除外されてきた非言語情報を含む会話分析が可能になる。また、従来のコーパスのように、転写記号等に関する専門知識を必要としないので、研究用資料としてだけでなく、外国人向けの日本語教育の資料・教材としても十分に活用できるであろうと思われる。

4. 今後の研究計画

これまでの試験研究の段階では、被験者の確保、実験条件の整備等の諸問題があり、結果的にNNSのデータがNSに比較して相対的に少なくなった。また、人種的、年齢的な偏り（大部分が欧米系で20歳代の留学生）および日本語能力の格差も相当にあり、標本分布の上でバランスを欠いた。さらに、NS,NNSの両者について、言語行動の国際比較という観点から、日本国内だけでなく、外国在住のNSおよびNNSにまで分析対象を拡大して比較検討することが重要である、と思われる。

そこで、平成７年度以降、以下の順序で追加データの収集、コーパス構築作業を行う。

1) 日本語教育関係者（日本語会話能力テスター

有資格者）と同出版社（株式会社アルク）および数校の日本語学校の協力を得て、非欧米系（アジア、中近東などの出身者）NNSの会話データを収録。

2)米国在住の日本人研究協力者（牧野成一プリンストン大学教授）の指導の下に、国内と同一の実験条件で現地在住のNS,NNSそれぞれの会話データを収録。

3)日本在住の欧米系NNSおよび日本国内のNSのデータ追加収録については、国際基督教大学で、村野の責任において実施。

4)収集した言語データ（ビデオテープに収録）について、ある程度の取捨選択を行った後、集中的にテキスト転写、ビデオ部分のディジタル化(MPEG又はQuickTimeMovie)を実施し、最終的にNS,NNS合わせて50人分の画像・音声データのディジタル化と編集作業を完成させる。

完成したコーパスはCD-ROM上で利用可能な形にし、言語学、日本語教育および情報処理関係の学会でスピーチ・エラーの分析結果とともに研究発表する。また、テキストデータとビデオデータ（ディジタル・ムービー）はそれぞれインターネット上の分散型データベースサーバー（WWWなど）に登録し、国内外の言語研究者および日本語教育者に公開する。

5. おわりに

コーパスを用いた言語分析は、欧米ではすでに確立した研究手法であり、分析対象も文語体のテキストにとどまらず、会話内容を転写したデータから成る口語体テキストにまで及んでいる。日本でも、最近コーパスの重要性が認識され、欧米のLOB, London-Lund, Brown等の大規模コーパスを利用した英語の文法・語法などの研究例が増加しているが、日本語コーパスについては、いまだ欧米ほど確立した大規模なものがなく、まとまった研究成果も報告されていない。従って、われわれの研究は、日米間にまたがる大規模な日本語会話コーパスの構築プロジェクトとしては前例のないものであり、マルチメディア型データベースを使用した言語研究としても、先駆的な役割を果たすもの、といえる。

（参考）コーパス作成方法とデータ検索の実例

会話データの収集方法については、日本語会話能力検定の手法として知られるOPI(Oral Proficiency Interview) に準拠し、インタビュー形式の実験を行う。被験者の数はNS,NNS各25名程度である。実験者と被験者それぞれの視点の延長上および実験者と被験者の中間（実験状況を確認するための参照画面用）の計3カ所に高画質8ミリビデオカメラを据え、OPI実験の模様をビデオテープに記録する。

実験で収録したデータは、各大学の研究者・日本語教育関係者が分担して転写作業を行う。転写作業にあたっては、原則としてアナログ音声をディジタル変換したデータファイルを適宜分割し、音声編集可能なノート型パソコン（可搬性を考慮）に移植し、ワープロソフトを用いて行う。

転写作業完了以後のコーパス作成手順は、①実験データを収録したビデオテープから画像・音声データをディジタル形式に変換②データベース作成支援ソフト（*HyperCard*)を用いて、テキスト・画像・音声データをリンク③ディジタル化された画像・音声データを適当な単位に分割し、データ名を付与④ページ単位で画像と音声を再生するための注釈ボタンをそれぞれ設定⑤CDライタを用いて追記型光ディスク(CD-R）に全ファイルを保存、のような順になる。

最後に、試作版コーパスを用いた代名詞類の検索例を紹介する。文語体の場合に比べて、現実の発話においては、代名詞が単なる先行名詞の照応表現以上の機能を果たしていることが以下の実例によっても明らかである。なお、会話中の1は実験者、2は被験者を示す。また、カッコ内は聞き手のあいづち、下線部が代名詞（ここでは「あれ」）、*印で挟まれた部分が指示対象の名詞（句）を示している。

（後方照応による記憶の想起）

1：そうですね。あの、ご出身もー、<u>あれ</u>ですか？　あのー、*東京都*ですか？

2：ええ、東京です。

1：それは単に、あのー、<u>あれ</u>ですか、保谷から国分寺高校っていうのは、単に、その、*電車通学に憧れて[ということ]*？

（聞き手による指示対象の補完）

2：[夏休みは]取れなくないかもしれないんですけど（1：うーん。）それほどにまだ、（1：ああ。）あの、どこへ行くっていう、<u>あれ</u>もないですから。

1：ああ、そうですか。（2：ええ。）全然*予定*は立ててないっていう、うーん。

（話者交替による先行詞の引き継ぎ）

2：それで、特別にあと、プロジェクトをやったときには（1：うん。）一千万ぐらいはもらって、何かこういう*システム*を作ります。

1：え、じゃあ、すごい、<u>あれ</u>ですね。

2：そうですね。

情報知識学会平成７年度研究論文発表会
論文募集のお知らせ

情報知識学会では平成７年５月２７日（土）に、総会と併催で、研究発表会を実施する予定です。この発表会のための論文を募集いたします。研究発表会の内容については次頁をご参照ください。

１．公募するテーマ

基本的には自由論題としますが、以下のようなテーマを例として挙げておきます。

＜例＞

①フルテキストデータベース、電子出版、電子図書、マルチメディア
②著作権
③用語、ディスクリプター、シソーラス、電子化辞書
④機械翻訳とその適用
⑤コンピュータネットワークと分散データベース
⑥専門分野における情報と知識の表現法
⑦その他情報と知識に関する基礎的アプローチ

２．論文の要領

①研究発表会では質疑応答を含めて３０分で納まる内容
②予稿４頁を平成７年４月１４日（金）迄に提出出来ること（ワープロにて作成の事）
③予稿提出がないと発表は出来ません。また、予稿４頁迄は無料ですが、それ以上は有料となります。
④発表会は平成７年５月２７日（土）に実施予定。会場は凸版印刷（株）本社１階ホールを予定。

３．応募方法

下記申込用紙に記入の上、〒１１２文京区大塚３－２９－１　学術情報センター研究開発部　小山照夫宛て、郵送またはＦＡＸ（０３－５３９５－７０６４）にてお送りください。締切りは平成７年２月２８日（火）と致します。学会にて審査の上、発表者には平成７年３月１０日迄に予稿作成依頼のご連絡を致します。

情報知識学会平成７年度研究発表会に応募いたします。			
氏名		連絡先電話番号	
連絡先住所	〒		
応募テーマ			
論文題名			
論文概要			

平成７年１月１５日
情報知識学会事務局

平成７年度総会・研究発表会実施のお知らせ

情報知識学会の総会・研究発表会について、詳しい内容は未定ですが、およそ下記のようなものとなる予定です。

＜総会．研究発表会実施案＞
１．日時　平成７年５月２７日（土）　９：３０～１８：３０
２．会場　凸版印刷（株）本社一階ホール（予定）
３．当日のプログラム案

９：００	自由論題	自由論題	（午前中は公募セッション）
１０：００	研究発表５本	研究発表５本	２会場に分かれ分科会形式で実施
１０：３０			質疑応答を含めて３０分で１本の研究発表
１１：００			
１１：３０			
１２：００			
	昼食		
１３：３０			
１４：００	招待講演セッション	別会場にて	（午後は講演・パネル討論ほか）
１４：３０		デモ等	講演・パネル討論と製品デモ等
１５：００	コーヒーブレイク		
１５：３０	パネル討論		
１６：００	（論題検討中）		
１６：３０			
１７：００	平成７年度総会		
１７：３０	懇親会		
１８：００			
１８：３０			

４．スケジュール
①発表申込締切り
②発表者決定
③予稿原稿締切り

５．その他　予稿集、懇親会は有料とする。（参加費は無料）

情報知識学会通信

情報知識学会に入会を御希望の方は、このフォームをコピーして必要事項を御記入の上、事務局に郵送、又はＦａｘでお送り下さい。折返し入会案内、入会申込書等の書類をお送り致します。（現在入会金は1,000円、年会費は5,000円です。）なお現在ニューズレターがあります。御希望の方はお知らせ下さい。

Ｆａｘ：０３（３８３７）０３６８ 又は ０３（５６８８）４６９４

〒110　東京都台東区台東１丁目５番１号（凸版印刷内）

情報知識学会事務局　担当　五所　行

情報知識学会に入会したいので必要な書類をお送り下さい。

個人用　　法人用　　（どちらかを丸で囲んでください）

住　所：〒

(フリガナ)
氏　名：

電　話：

Ｆａｘ：

あとがき

この号の締切は、たいへんな事件の勃発と並行してしまった。神戸の大震災である。編集の田中さんに、原稿のメイルをパソコン通信で送付するかたわらニフティーをたちあげると、なくなられた方々のお名前が羅列されたファイルが、眼に飛び込んできた。東京で見ていても、刻々と亡くなって行く方々を確認しながら、震災発生後もっと早く何かができなかったのか、と深い悲しみとともに、激しい焦りを感じる。危機管理が叫ばれるなかで、ようやく首相が、通信の画面の存在を認知するような時代である。行政についての情報の大切さがわかり、そのアクセスにたけたプロが、なぜもっと判断の要に位置しないのか、メカだけが進み、情報についての哲学の進まない、わが国の文化の貧しさをいやが上にも感じてしまうのである。

木村三郎

本号では編集委員の方々に御尽力頂きましたが、残念乍ら十分に原稿が集まりませんでした。一般原稿も常時受け付けておりますので、３ページの執筆要項をご覧の上どしどしご投稿下さい。新しい企画も歓迎します。

最後に、この度の阪神大震災により、お亡くなりになられた方々のご冥福をお祈り申し上げますと共に、被害を受けられた多くの方々に謹んでお見舞い申し上げます。

長瀬真理