

## データ整理のためのパソコン用ソフトの使い方

田原広史

### 1 はじめに

筆者は、パソコンを用いたデータ整理を 20 年近くおこなっている。方言を中心に研究をおこなっている関係上、整理するデータの内容は、初期は方言調査をおこなった結果を集計するという作業が中心であったが、最近では、それだけではなく、方言音声データを編集・整理した上でデータベース化したり、方言地図をパソコン上で作成するといった方面も扱うようになってきた。また、卒業論文を指導する上で、方言調査だけではなく、新聞・小説・映画のタイトルといった文字に見られる表記上の特徴や、字種・語彙の統計などについても、パソコンを用いて整理することが多い。このように、パソコンで扱うデータの範囲が拡大してきた経緯には、筆者の研究への興味に移り変わったことや、広がったことが原因ではあるが、当然、パソコンのハードウェア、ソフトウェア両面の発展が前提となっている。

かつて、本雑誌において、「わたしのパソコン言語学」という特集があり、1993 年 8 月号に「リレーショナルデータベースソフトを中心としたデータ整理」と題したものを書いたことがある。当時、筆者がおこなっていたパソコンによるデータ整理の方法と、データ整理に対する考え方を書き記したものである。それから約 7 年が過ぎた。この機会に、当時と、21 世紀の最初の年である現在とを比べて、パソコンを取り巻く環境も含め、どのような点が変わり、どのような点が変わっていないのかを確認してみたいと思う。

また、この特集の趣旨は、日本語学の分野において数量的手法を用いた卒業論文を書く学生に役立つように、基本的な考え方や方法論を示すことだと聞いている。この点について考慮するために、筆者がおこなっている卒論指導を念頭においた説明を心がけたいと思う。

以下では、まず 7 年前と現在の状況との違いを簡単にまとめることから始める。なお、本稿のタイトルは「データ整理のためのパソコン用ソフトの使い方」であるが、具体的な操作方法を細かく書くことは、紙数からして無理であるし、読者がそれぞれのソフトに接して学んでもらうしかないと考えるので、ここでは、データ整理をおこなうにあたって、それぞれのソフトを使うと、どのようなことができるのか、なぜそのソフトを使った方が良いのかといったことを中心に述べたいと思う。

### 2 パソコン環境と使用ソフトの変化

この 7 年でパソコン環境は劇的な変化を遂げた。7 年前は MS-DOS (エムエスドス) の時代だったが、今は Windows 2000 の時代である。Windows へ移行により、パソコンでグラフィックが新たに扱えるようになり、グラフ、方言地図などの処理、整理が格段におこ

ないやすくなった。さらにマルチメディアへの対応により、パソコンによる方言音声データベースへの道も開かれた。

7年前に筆者が使用していたソフトと使用目的を簡単に振り返ってみる。データベースソフトの「桐」でデータの入力・編集・保存・管理をおこなっていた。エディタソフトの「MIFES」(マイフェス)あるいはワープロソフトの「一太郎」を用いて、データ加工・集計結果の編集等をおこなっていた。集計用ソフトは「GLAPS」(グラプス)を用いて単純集計・クロス集計・データ変容(コードの付け替えなど)・方言地図化をおこなっていた。また、集計結果のグラフ化は自作の BASIC プログラム(「GDP」と名付けていた)を用いていた。

現在は、データの入力・編集・保存については全面的に「エクセル」、集計結果のレイアウト、印刷などは一太郎に一本化している。集計については現在も GLAPS を使っており、グラフ作成は「The Graph」という専用ソフトを新たに用いている。方言地図作成については、エクセルと「ファイルメーカー Pro」を組み合わせることにより、作図とデータベース化を同時におこなうことが可能になった。

こうしてみると、道具立てについては、多少の変化はあるが、使い方の基本はさほど変わっていないことが分かる。データ処理の基本部分は単純なだけにならないということでもある。しかし、処理の正確さ、速さという点については格段の改善が見られる。また、表、グラフ、地図等の見栄えの良さは飛躍的に進歩しており、版下原稿を個人が直接作成することが当たり前になった。これは、単にパソコンのスピードが上がったからではなく、プリンタの性能の向上、ソフトの進歩により、より高度な使い方が可能になってきていることによるものである。

7年前、本誌で筆者が唱えていたソフト間のデータのやりとりについても、筆者の思惑通り、それぞれのソフトが歩み寄りを見せ、また、役割分担を積極的におこなうようになってきた。「これ一つで何でもできます(でも多少の使いにくさは我慢してください)」の時代から、「この部分についてはおまかせください。どこにも負けません。」という時代になってきたということである。このことは言い換えれば、用途に応じて、幾つかのソフトを組み合わせることにより、より良いもの、あるいは個性的・独創的なものを生み出していける可能性が出てきたということでもある。

具体的に言えば、7年前本誌で取り上げた桐は、入力とデータ管理、作表、作図等を一手に引き受ける統合化ソフトと呼ばれるものだった。しかし、現在の方向としては、入力、作表であればエクセル、データ管理・データベース化なら「アクセス」(筆者はファイルメーカー Pro を愛用している)、簡単なグラフ化ならエクセル、より精密で大量のグラフを作る場合は The Graph、といったように使い分ける方が、それぞれの特徴を生かし、それぞれの作業の効率が上がるということである。

それでは、筆者が具体的にどのようにデータ整理をおこなっているのかを紹介する。

### 3 調査データ集計作業の概要

筆者が、日常おこなっているデータ整理の作業は、次の三つに大きく分けられる。一つは、「調査データ集計作業」で、方言調査などの結果をコードに置き換え、集計をおこな

い、結果を表やグラフにするものである。二つ目は、「テキストデータ加工作業」で、文献等から抜き出したテキストデータを蓄積していき、そのデータを用いて文字列を検索したり、加工したりして、字種や語彙の統計をとっていくものである。三つ目は、「データベース作成作業」で、収録後パソコンで編集した音声を用いて音声データベースを作成したり、方言調査で得た語形に記号を与えた上で方言地図データベースを作成するといったものである。ここでは、紙数の関係上、一つ目の「調査データ集計作業」についてのみ紹介する。

ここで扱う作業の流れは、以下の通りである。

調査	コード化	入力	集計	グラフ化	印刷
----	------	----	----	------	----

### 【コード化】

調査票に記入されている調査データはさまざまな形態をしている。話者が答えたままを記入したような場合もあれば、あらかじめ選択肢が設定されていて回答が幾つかの数にまとまっている場合もある。自由記入のデータについては、回答を分類するところからデータ整理が始まる。調査によっては、この二つが混在するようなものもあるだろう。選択肢についても、語形等をいったん「コード化」(数字に置き換えること)する作業が必要になる。語形のままで、集計の途中で統合・分類等がおこないにくいからである。その他の語形が多い場合は、コード化の段階で、新たに出てきた語形とコードとの対応ができるように、一つの調査項目につき一枚のカードを用いたコードブックを作る。

整理上の工夫としては、予想される語形にはコードを割り振っておき、調査票にもあらかじめ印刷しておくことが考えられるが、アンケート形式のものでは、記入者が誤解したり、気が散ったりしないように配慮する必要がある。また、「使う」「使わない」、「あてはまる」「あてはまらない」といった、単純で、その他の回答が出てこない選択肢の場合は、直接入力作業に移ることができる。

### 【入力】

現在、「入力」作業は表計算ソフトであるエクセルを用いている。エクセルでは、入力をおこなうにあたり、さまざまな入力支援機能が用意されており、効率的かつ正確に入力作業がおこなえる。例えば、通し番号をすべてのレコード(行)に一気に入力したり、住所のように同じものが繰り返し出てくる場合、一気にコピーするといった「オートフィル」と呼ばれる機能や、同じカラムに関しては、同じものが出てきた時は、識別できる文字を入力した段階でそれ以降を自動的に補完してくれる「オートコンプリート」と呼ばれる機能がある。

入力方法も、文字通り縦横無尽で、横方向にずっと入力することも可能だし、縦方向に入力することも可能である。その際のカーソル移動も、ブラインドタッチでおこなったとしても、スムーズに流れるように設計されている。当たり前のように聞こえるかもしれないが、エディタやデータベース系のソフト(桐やファイルメーカー pro など)では、行列の観念がはっきりしすぎていて、入力は行(横)の単位が基本であり、縦への入力はでき

ない、あるいは非常にしにくいのである。

データの確認という面からもエクセルは傑出している。プリントアウト時のレイアウトについて、罫線を入れたり、ヘッダ・フッタを設定したり、すべてのページに行や列のタイトルを印字したりすることが可能であるし、何ページ以内に納めたいという指定をすれば、自動的に縮小率が設定されるなど、見やすさも重視されている。完成したデータは、そのまま卒論、報告書等の資料として使用できる。

エクセルには、データ入力に関する利点以外にも、データ管理・データ保存という観点からも優れた点がある。一つのファイルの中に複数のシートが作成できることも優れた点の一つである。入力段階においても項目ごとに、あるいは世代ごとに分けてデータを作成していくことが可能である。複数の人間で手分けてして入力作業をする場合にも、シートの分割・統合がドラッグ&ドロップでできる手軽さは魅力の一つである。複数のシートを活用することにより、一つのプロジェクト（例えば調査）ごとに、一つのファイルで管理がおこなえることは、データ整理上、大きな利点と言えるだろう。

### 【集計】

集計作業については、GLAPS を用いておこなっている。筆者の使っているソフトの中で、ワープロソフトの「一太郎」と並び、この GLAPS は 10 年を越えて使い続けている数少ないソフトである。集計ソフトに関しては、一般向けのものが普及しないことからすると、集計という作業が日常においては一般的ではないのだということが分かる。統計用ソフトとしては「SPSS」があるが、個人で買うには高価であるし、構造が複雑で、どうも気軽に使いにくい。

GLAPS の使い勝手のよさは、仕組みが単純で分かりやすいことである。ユーザプログラムという命令用のファイルを作る必要があるが、見本をいくつか見れば、それを真似てすぐに作れる程度のものである。また、作業の結果がこのファイルに残ることから、何年か経って見たときも、データの形式・構造や、以前におこなった分析をすぐに理解することができるという点も特長の一つである。ただし、この GLAPS は元々大型計算機用のプログラムを MS-DOS 上で動くようにしたものなので、今後、いつまで使えるか不安である。いずれは、別のソフトに移行する日も来るのだろう。

集計方法には、異なったコードを割り出し、人数と全体に占める割合を算出する単純集計と、さらにそれを別の変数（男女や世代）でグループ分けした後におこなうクロス集計の二つが基本になる。クロス集計については、GLAPS では同時に各グループの集計結果を別のファイルに書き出す機能を備えており、そのデータを再びエクセル上に読み込んで加工することにより、グラフを作成することができる。また、結果を出力する際に、コードに名前を付けることができるので、確認のために元の語形を名前として付けておく方がよい。ユーザプログラムで設定した名前は、そのまま受け渡していき、最終的にグラフの凡例としても使うことができる。

エクセルで入力したデータの GLAPS への受け渡しは、CSV 形式と呼ばれるコンマ区切りデータの形でおこなう。ただし、上で述べたように GLAPS は昔の構造をひきずっているため、一行を 80 桁以内に納める操作が必要である。これにワープロソフト等で作成した GLAPS 用のユーザプログラムをリンクさせて、GLAPS を走らせるという手順である。

学生への卒論指導の際は、この部分は昔風のパソコンの使い方であり、わかりにくいので、学生には直接やらせず、筆者が学生の目の前で説明しながら集計することになっている。学生にとっては、卒論を書き終えたら二度とおこなわない作業だと考えるので、使い方を覚えてもらう必要がないからである。

#### 【グラフ化】

GLAPS で集計し、出力されたクロス集計表は、テキストファイルとして保存し、一太郎で読み込んでレイアウトを整えた後、そのまま資料として打ち出す。グラフについては、上で述べたように GLAPS から出力された数値データを再度エクセルに読み込んで、グラフ用の形式に整える。エクセル上でグラフ化することも可能であるが、調査結果の分析にあたっては、同じ形のグラフを大量に作ることがふつうなので、グラフのレイアウトを登録しておき、それを利用して効率よく大量のグラフを作ることができる The Graph を使用している。The Graph はデータの扱い方、配列の形式などについて、エクセルにほぼ準拠しているので、エクセルのシート上で整えたデータを、そのまま The Graph のデータシートにコピー＆ペーストして使うことができる。

作成したグラフは、そのまま印刷して資料として使う場合が多いが、グラフィックデータとして、一太郎やワードといったワープロソフトの文書内に貼り付け、見やすくレイアウトすることも可能である。このあたりの使い勝手の良さも、この5年くらいの間に格段に進歩した点であり、これからもさらに改善されていくであろう点である。

以上述べたことを、ソフトの使用との関連でまとめると以下ようになる。

調査	コード化	入力	集計	グラフ化	印刷
手作業		エクセル	GLAPS	エクセル	The Graph

#### 4 おわりに

以上、「データ整理のためのパソコン用ソフトの使い方」と題して、最も基本的な調査データの入力からグラフ化までの工程を簡単に説明した。残念ながらここでとりあげることができなかった、テキスト処理に関するエクセルの使い方については、近いうちにまとめておきたいと考えている。データベース作成作業に関しては、ファイルメーカー Pro を使った論考がすでに幾つかあるので、参考文献にあげておいた。

最後に、筆者がパソコンを使ってデータ処理をおこなう場合に心がけていることを述べて終わりにしたい。どのようなソフトを使って、どのような手順をとるかということは、単純なようでいて、なかなか難しい。筆者が心がけていることは、その方法が、他の人（多くは学生）にとって、使いやすい環境にあるものであり、なおかつ、単純で間違いを犯しにくいものを、ということである。そのためには、多少効率が落ちてでも仕方がないと考えている。Macintosh のハイエンドユーザならスクリプトを使ってあっという間にできることが、Windows では一々変換していかなければならない、といった類の話は、パソコンの世

界ではよくある。しかし、遅くても分かりやすく確実に結果にたどりつける方法があるなら、迷わずそちらを採用すべきだと思っている。これは、研究の方法論を普及させ、受け継いでいくという観点からも大切な点である。

もう一つ心がけている点は、パソコンでおこなうすべての作業を、常に手作業の延長線上にとらえることである。言い換えれば、決してパソコンを過信、盲信してはならない、間違ってもパソコンに振り回されたり、ましてやパソコンに使われてはならないということである。パソコンはあくまでもユーザの良きアシスタントであるべきで、それ以上を期待すべきではない、というのが筆者の信念である。

それぞれのソフトの画面のイメージ、印刷見本等を示すことができなかつたので、分かりにくかつたかもしれない。想像力を駆使して、再度お読みいただければ幸いである。パソコンを活用するにあたっては、「想像力」ということばが重要なキーワードでもある。

(注記)

「MIFES」はメガソフト社、「一太郎」はジャストシステム、「エクセル」「アクセス」はマイクロソフト社、「ファイルメーカー Pro」はファイルメーカー社、「The Graph」はアスキーサムシング社、「SPSS」はエス・ピー・エス・エス社、のそれぞれ登録商標である。また、「GLAPS」は荻野綱男氏、「GDP」は筆者の作成したプログラムである。下記の参考文献のうち入手しにくいものは、筆者(tahara@x.age.ne.jp)までお問い合わせいただきたい。

#### 【参考文献】

- 荻野綱男(1994)『アンケート調査分析用ソフトウェア GLAPS の使い方 第2版』(私家版)
- 荻野綱男・内田伸子・田中ゆかり・田原広史・齋藤孝滋(2001)「人間を対象にした調査や実験の問題点」(『第7回 社会言語科学会研究大会 予稿集』)
- 岸江信介・中井精一・鳥谷善史・石田祐子(2000)「エクセルとファイルメーカープロを利用した言語地図の作製 『徳島県言語地図』作製を例として」(『大阪樟蔭女子大学日本語研究センター報告』第8号)
- 田原広史(1991)「データの収集と処理」(『新・方言学を学ぶ人のために』世界思想社)
- 田原広史(1993)「リレーショナルデータベースソフトを中心としたデータ整理」(『日本語学』12巻9号、明治書院)
- 田原広史(1996a)「GLAPS をとりまくパソコン環境の現状について」(『パソコン版 GLAPS の開発』平成7年度科研費 総合研究(A) 研究成果報告書)
- 田原広史(1996b)「河内方言調査について」(『フィールドワークを歩く』武蔵野書院)
- 田原広史・中村一夫(1998)「国語国文学分野におけるデータベースソフトの評価 『ファイルメーカー Pro』と『桐』の比較を中心に」(『日本語研究センター報告』Vol.5 大阪樟蔭女子大学日本語研究センター)
- 田原広史(2000)「方言録音調査に基づく音声データベース作成の試み CD-ROM『道教え文字化・音声資料』」(『大阪樟蔭女子大学日本語研究センター報告』第8号)

(たはら・ひろし 大阪樟蔭女子大学助教授)