

## 方言研究とコンピューター

著者	沢木 幹栄
出版者	長野県ことばの会
引用	ことばの研究 6: 39-44(1992)
発行年月日	1992-03-25
URL	<a href="http://hdl.handle.net/10091/00022392">http://hdl.handle.net/10091/00022392</a>

## 方言研究とコンピューター

沢木 幹栄

最近になってコンピューターは我々の生活のいたるところで使われるようになって来ました。言葉の研究の世界でもそれは例外ではありません。コンピューターがこの世に出現した直後からコンピューターは言語の研究に使われていますが、最近ではコンピューターを使うのは限られた研究者ではありませんし、また、言語研究のなかの限られた分野だけで使われているわけでもありません。

今日は、言語研究のなかでも特に方言研究でコンピューターがどのように使われているか、また、コンピューターを使う事によってどんな局面が開けてくると考えられるかについてお話ししたいと思います。というのは、私がここ何年かコンピューターを使って主に方言に関する研究を行なって来たということと、今年の夏に開かれた国際方言学会議に参加してこの分野での世界の研究の動向に接する事ができたということで、皆さんに少しは新鮮な話題を提供できると思うからです。

ところで方言研究の話に入るまえにコンピューターの得意とするのはなにか、それをちよっと考えてみましょう。

コンピューターのできる事としては、つぎの4つが考えられます。

1. 大量の計算を高速に正確に行なう
2. 大量のデータを記憶して任意の順番でとりだしたりする
3. 他の機械をコントロールする
4. 通信

1. はコンピューターという言葉でみなさんが真っ先に連想することだろうと思います。円周率の計算はその代表的なものです。いまはコンピューターを使って何百万けたという精度で円周率が計算されています。コンピューターができるまえは円周率の計算に一生を捧げた人がいるくらいです。シャンクスという人の墓には彼が一生かかって計算した円周率の数字が刻み込まれているそうですが、実はその数字は途中から間違っているのだそうです。それくらい大変な計算がいまやあっという間にできるようになっているのです。もっと身近な例としては天気予報が例としてあげられます。最近では、けたはずれな計算能力をもつコンピューターを使って天気予報を出すことが行なわれています。沢山の観測地点での風向きや風力温度湿度などを組合せた巨大な方程式を解くと何時間後の天気がわかるという仕組みです。観測地点を増やせばそれだけ正確な予報が出せるわけですが、そうすると方程式が大きくなりすぎてコンピューターの能力を越えてしまうとか、計算に時間がかかりすぎるとかの問題がでてきます。それから、予報というのは、その現象が起きる前に知らせなければ意味がありません。24時間後の天気を計算するのに24時間の計算時間がかかったら、天気予報にはなりません。天気予報につかわれるコンピューター

には莫大な量の計算を短い時間で行なう能力が必要とされるわけです。現在の天気予報の精度は、一面で現在のコンピューターの限界を表しているということになるかと思えます。

2. の例としては銀行の口座管理があげられます。現金の引き出し機にカードをいれて決められた操作をするとお金がでてきて、しかもその分がちゃんと口座からひかれているということが普通に行なわれています。今はあらゆる金融機関でこれが可能ですが、10年以上前には、郵便局でできなかったし、さらに10年前では普通の銀行でもできなかったことです。大量のデータを記憶して管理するコンピューターが出現して初めて可能になったわけです。

郵便貯金を考えてみれば分かりますが、膨大な数の口座を管理するというのはたいへんなことです。しかも、ちょっとでも誤りがあれば、たちまち社会問題になってしまいます。

似たような例としてJRの座席予約システムがあります。これもどれだけ大量のデータを処理しているか、またどれだけの信頼性が要求されているかちょっと想像してください。

3. の例ですが、最近マイコンを組み込んだ製品を私達の身の回りでよく見かけます。自動炊飯器にはマイコンが使われるのが当たり前になりました。このマイコンというのは、昔、釜で炊いていた頃に「はじめチョロチョロなかパッパ、赤子泣いても蓋とるな」と言っていた火加減を、自動炊飯器で行なうようにしているものらしい。マイコンなるものが発熱部をコントロールするわけです。マイコンというのは一個の小さな部品にすぎませんが、コンピューターの種類であることには変わりはありません。いまや、日本のメーカーは、我々の身の回りのあらゆるものにマイコンを使おうとしているようです。最新の家庭電気製品にはたいていマイコンが使われていますし、自動車もそうです。

4. はますますぴんと来ないかもしれませんが、最近コンピューターの仕事としてはどんどん重要性をましてきているものです。電話局にある自動交換機はコンピューターそのものですし、さきほどの現金引き出し機にしても、銀行の本店にあるコンピューターと通信してデータをやりとりしています。

最近クロネコヤマトがFAXサービスを始めるということを新聞で読みました。普通のFAXだとAさんのFAXからBさんのFAXに直接原稿がおくられてくるだけだったのですが、クロネコのはAさんのFAXの内容をいったんセンターのコンピューターにためておいて、BさんのFAXから要求があったときだけこれを送信する。

Bさん、つまり受け手のがわからずれば、メールボックスをのぞいて自分あての手紙がないかどうか調べるのと同じような感覚でFAXを受け取ることができるし、自分専用のFAXでなくても、出先でも自分にあてられたFAXを受け取ることができるわけです。

これもコンピューターが通信に使われる一例です。

以上の1. 2. 3. 4. がコンピューターの得意とする機能です。コンピューターのしていることのほんの一部を述べただけですが、今の世の中がコンピューターなしではたちもさっも行かなくなっていることがよく分かります。しかも、それにもかかわらずコンピューターの存在を意識することはほとんどなくて、「これもコンピューターを使ってい

るんだよ」と言われてはじめて「あっそうか」ということになるわけです。

ここで本題にはいることにしましょう。方言研究でのコンピューターの使われ方も分類すればいままで述べてきたのとおなじように考えることが出来ます。

まず、1.の「大量の計算を高速に正確に行なう」ですが、コンピューターが手軽に使えるようになったときにまず最初に考えられたのがこの能力を生かす事でした。

数量化とかクラスター分析とか言われる数学的な方法があります。たとえば、A地点とB地点がどの程度方言的に隔たっているかを知りたいとします。ほかには、C地点、D地点などがあり、これらの地点を同一線上で比較する事はできないのが普通です。語形の共通度が地点間の言語的な距離を表すとしたら、ABCの各点が一本の直線上に並んでいると話が簡単なのですが、実際にはABCという三角形の頂点のようになっているのが普通です。

このような場合に威力を発揮するのが数量化などの方法です。うまくやれば、沢山の地点を、距離が近いもの同士いくつかの大きなグループにまとめることができます。ところが、これを実行しようとする膨大な量の計算をこなさなければなりません。コンピューターが使えるようになって、はじめてこの方法が使えるようになったわけです。

私は今年の夏、ドイツのバンベルクで開かれた「国際方言学会議」に行ってきましたが、海外でもこのような数量化の方法を用いた研究が盛んである事を知りました。

コンピューターの1.の機能を生かした研究として私も孤例の研究というものを行ったことがあります。コンピューターがくりかえし大量の計算をするということがどんなことなのか、説明するのにちょうどいいので、どんなことをどんな手順で行なったのかちょっとくわしくお話ししましょう。

まず、言語地図のなかには多くの地点で使われる語形と、わずかの地点でしか使われない語形があります。ひとつひとつの地図について、1地点だけで使われる語形の数ほどれくらいか、2地点で使われる語形の数ほどれくらいかということを調べていきます。これは語形が伝播していくときにどんな現象が起きるかということと関係があります。

つぎに、1地点だけでしか使われない語形に注目します。こういう語形を孤例と呼びますが、30枚の地図を通して見たときに、孤例が多く出現している地点を探します。そういう地点を地図にプロットすると、ある地図ができます。孤例が多いというのは、その地点が変わっている、特異なところだということを示しています。琉球にはそうした地点が多いのですが、それは偶然ではないでしょう。この地方では集落ごとに言葉が違うと言われているからです。

これだけのことを、実際のプログラムではつぎのような手順で処理しました。まず、地図ごとに使用されたすべての語形が何回使われたかを数えます。つぎに孤例の語形がどれであるかを調べ、その語形が使われた地点を記録します。以上の手順を地図の枚数だけくりかえして各地点で孤例が何回出現したかを調べます。

このプログラムで一番多く行なわれているのは単純な足し算ですが、それを何重にも繰り返して行なっているのです。人手では不可能なことが単純な計算の繰り返しで可能になる、これはコンピューターを使った仕事の一つの典型だと思います。

面白いのは、このプログラムは最初、大型計算機で動くように作られていたのです。大型計算機というのは大学の計算センターなどで見るような、名前のおり大きな部屋にやっと入るぐらいの大きなものです。ところが、そのプログラムを我が家のパーソナルコンピュータで動くようにしてみたところ、計算時間が大型の倍ぐらいで済んでしまったのです。あとで、もっと上等なパソコンを買って、同じプログラムを動かしたら今度は、大型よりもずっと早く計算が終わってしまいました。私が以前勤めていた国語研究所というところで1年前まで動いていた大型計算機よりも、このプログラムを実行するという点においてはずっと速くて使いやすいコンピュータが私の机の上のっているというわけです。ちょっと前なら個人で使う事は夢にも考えられないような、すばらしい計算能力をいま我々は手にしているのです。

孤例の研究が今後どのように進んで行くべきかについては、私自身考えてはおります。ひとつはここで見られた傾向がほかの色々な規模の言語地図でも同じように見られるのか確認すること、もうひとつは語形の伝播、ひろがりがある一定の数式に従っていてそれがこのような統計的な結果を産むと考えているのですが、この仮説を証明する事です。このどちらにもコンピュータをさらに活用する事になるでしょう。

2. の「大量のデータを記憶して任意の順番で取り出したりする」機能ですが、これはパーソナルコンピュータではデータベースという形で利用されています。

さきほどのバンベルクの学会でも話題になったのは方言データのデータベースをどうやって作るか、そのデータベースにどのような機能を持たせるかでした。我々が言語地図を作る時には、いろいろな語形をひとつにまとめて、ある記号をあたえるということをしています。そのあとで地図をかいて、まとめかたが正しかったか、記号の与えかたが適当だったかなどを検討し、満足がいかなければもういちど地図を書き直すということをしているわけです。

この一連の作業をすべてコンピュータで行なう事にすれば、非常に能率的であろうというのが方言データベースの出発点になっています。つまり、1. の場合と違ってコンピュータを使わなければならないような新しい事をするのではなく、いままでやってきたことをより能率的に行なうために利用するという考え方です。こうして作ったデータベースで、さらに等語線を引くとか、数量的な処理を行なうとかの機能を加えてさらに便利にするということを考えているようです。バンベルクの発表をきいた限りではかなり満足のいくシステムがいろいろな所で作られているという印象でした。実は同じようなものは国語研究所の前川さんも考えています。

ただ、これから本格的に使うためには機械の側にまだ色々な問題があります。機械のほうの問題である以上、5年以内にはすべて解決されるでしょうが。

ひとつは、ディスプレイの表示がまだ粗いということです。日本言語地図程度の複雑で細かい地図をディスプレイに表示するのは今の機械の性能ではまだ無理です。バンベルクでは「しゃべる言語地図」という発表までありました。これは調査した語形がどんな音声だったかを機械が記憶していて、必要が生じた時にいつでも声が出てくるというものです。言語地図を見て実際の発音を知りたいと思う事がよくありますが、このシステムでは簡単に知る事ができるわけです。ただ、これは現在の技術ではまだ制約が多く、発表者も

完成は3年後だと言っていました。

また別の発表者は「言語地図の意味は変わった。言語地図はデータベースの出力する形式のひとつに過ぎない」という威勢のいい発言をしていました。

データベースに近いものとしては『東京語アクセント資料』のデータがあります。ただ、このデータは出版には利用しましたが、まだ未整備な部分があって研究に利用する状態になっていません。

3.の「他の機械をコントロールする」にあたるものはコンピューターにつながっているプリンターをコントロールして言語地図を書き出したり、音声記号を打出したりすることでしょう。

コンピューターで言語地図を打ち出したものでひとことふれる必要のあるのは『糸魚川言語図巻』でしょう。これはすべての地図をXYプロッターという機械を用いて描いています。

現状では、コンピューターで作った地図は機械のほうの制約があってもうひとつ評判がよくないようです。

音声記号を打ち出すことは、バンベルクでもかなり取り上げられていました。実は私も国語研究所にいたときにやっていたことです。『方言文法全国地図』第1集の解説書では地図として印刷された60項目の全データをそのままの形で—というのは調査者が報告してきた表記そのまま—収録しました。データはコンピューターで読める形になっていたため、プリンターで打ち出したのですが、プリンターをうまくコントロールすることで音声記号の印刷もできるわけです。

調査のデータを公開するのは、いままであまり行なわれていないことですが、公開することによって議論が深まっていくこともあります。どしどし公開するほうがよいと思います。

このデータを利用して、たとえば見出し語形の索引のようなものが簡単に作れます。第一集は助詞に関する項目を集めてありますが、この索引を使うと、ある見出し語形がどんな項目に現われるのか、あるいはどんな意味のひろがりを持っているかなどが分かってきます。もちろん、孤例の研究などもできます。

それから、このようなデータの形にしておくと、出版の形態が変わってくる可能性があります。コンピューターのデータというのはコピーが簡単にできます。フロッピー一枚には100万字以上の情報がはいつていますが、これをコピーするのに一分もかかりません。ですから流通させるのは印刷したものよりも簡単です。地図は地図として生のデータがあったほうが自分でいろいろな研究ができるからいいという人もいるはずですが、たとえば、組合せの地図などは作るのが大変だったのが、こうしたデータをコンピューターで処理するとずっと楽にできるようになる。

それはともかくとして、データの形で公開すれば、理科系の学問でよく言う「追試」が可能になって、業績が客観的なものになるということを言う人もいます。

それから、データがすでに出来ていれば、簡単なプログラムを付けるだけで非常に原始的なデータベースとして利用することもできます。第2集と第3集は活用を扱う予定ですが、活用形主体、語形主体と視点を変えた一覧表を2種類作ると地図づくりをする際に大

変有益な情報が得られるはずで。

話が前後しましたが、コンピューターを使って発音記号を打ち出すのはプリンターの外字機能というのを使いました。バンベルクでいくつか音声記号の印刷システムが発表されましたが、それらと比較しても印字の品質はかなりいい部類だったと思います。レーザープリンターを使えば、またさらにずっとみやすい印字になるでしょう。これも機械が進歩することによって解決される問題かもしれません。

しかし、今になって振り返ってみると、大変なのはプログラムよりもデータをつくることでした。私と研究室の白沢さんの二人がこれだけのために何年か費やしたような印象があります。なにしろ、全体のデータ量は800万字ぐらいになるのです。

余談ですが、データ入力も機械の進歩でここ何年かですいぶん能率が上がってきました。いい例が漢字入力で、5年前に『東京語アクセント資料』のデータを作ったときは漢字のデータを作るのに大変な苦勞をしましたが、今ではコンピューターで漢字が扱えるのは当たり前のことなのでこの変化というのは大変なものです。

最後の4.の通信ですが、これは方言研究に直接役に立たないけれど、研究の刺激にはなる可能性があります。

バンベルクで会った欧米の学者にはEメールのやりとりをする習慣を持った人がたくさんいました。Eメールというのは先ほどお話したクロネコFAXとよく似た仕組みで手紙のやりとりをするものですが、これのいい点は、世界中の人とリアルタイムでしかも非常に安い料金で情報の交換ができることです。地方は中央にくらべて情報が遅いということがよく言われますが、コンピューターを使った通信が普及すれば全国どこでも同じように情報が受け取れるし、沖縄と北海道の人が共同研究をするのも容易になるでしょう。

以上ざっとコンピューターと方言研究の関わりについて述べてきました。コンピューターを使うことによって方言研究にも新しい局面が開けてくる論旨で話をすすめてきたつもりですが、ちょっと心配なこともあります。コンピューターが面白すぎるために、データが元来どんな素性のものだったかを忘れて、コンピューターの世界にのめり込んでしまうことです。方言研究で扱われるデータは生身の人間が生身の人間に接して得たものなのに、データを何重にも加工するとどんどん抽象性が増していきます。これはこれで数学の世界にも似て面白いものなのですが、おおもとに生々しいデータがあったということを忘れてはいけません。データをもてあそんで単なる知的な遊びに墮してしまうことがないように自らを戒めなければなりません。

(信州大学人文学部助教授)