日本語を研究する

第４回

日本語研究における コンピュータの利用

――日本語教師のための日本語の用例の検索――

筑波大学文芸・言語学系教授　草薙　裕

1．はじめに

最近、パーソナル・コンピュータ（パソコン）が非常に進歩、普及し、日本語教育の現場でも、パソコンがないところはないくらいです。そして、どこででもパソコンをどういうように日本語の研究に利用したらいいかという質問を受けます。そこで、本稿では、日本語研究とコンピュータの関わりを簡単に考察した後、海外の日本語教師がコンピュータを利用するのに手っ取り早く簡単な方法を述べることにします。

2．日本語・日本語教育とコンピュータ

日本語・日本語教育へのコンピュータの利用を考えると、その目的から、つぎのように分けられるでしょう。

（1）日本語教育への支援

（2）日本語研究への支援

（3）コンピュータ処理のための日本語研究

（1）はコンピュータを教育に用いるためのシステムのことで、コンピュータの特性を生かして、個別教育を行う可能性がある有意義なコンピュータ利用の一つですが、教育を目的にしているので、本稿ではこれ以上、考察しないことにします。次の（2）は日本語の研究あるいは日本語教育者のデータ収集のためにコンピュータを利用するものです。これには、つぎのようなことが考えられます。

（2.1）日本語研究、日本語教育のためのデータを、データ・ベースから取り出して利用するもの

（2.2）日本語の形式化の検証のためにコンピュータを用いるもの

（2.3）日本語に関して、なんらかの基準で数量化し、それを統計的に処理するもの

に分けられます。（2.1）については次項で方法も含めて解説します。

（2.2）は日本語の、例えば文法の規則を分析し、その規則が正しいかどうかをコンピュータで大量のデータを使って調べるもので、これは文法の規則によって処理が違いますから、研究者が自分でコンピュータのプログラムを実際に作らなければなりません。

（2.3）の数量化とは、元来、数量的ではない言語を、何らかの基準で数的に解釈することです。この分野を計量言語学と呼びますが、文の長さ、特定の品詞が使われている割合などを数量化して調べることで、ある作品の著者を推定したり、ことばの使い方の実態を解明するなどの研究があります。この分野の研究はほとんど統計的処理を容易にすすめるためにコンピュータを利用するというものです。

最後の（3）はコンピュータに日本語の処理をさせること、言い換えれば、コンピュータがあたかも日本語を「理解」し、日本語文を「作り出す」ようにするための研究で、コンピュータ言語学と呼びます。これはコンピュータのための言語研究として、また、コンピュータの本格的な処理ということで、コンピュータ工学との学際研究など、かなりのコンピュータの専門知識が要求されます。

3．日本語教師のためのコンピュータ利用

本誌の読者の多くが海外にいる日本語教師であると思うので、日本語研究及び日本語教育における道具としてのコンピュータ利用について考えようと思います。

なんといっても言語研究、言語教育にはデータとしての言語表現が必要です。いろいろな語や文法現象がどのように使われるのかは語彙研究や文法研究には欠かせません。データを集めたデータ・ベースがある限り、調べたい表現の使用例をコンピュータで検索するのは比較的簡単です。

目指す表現がどのような文脈で用いられているかを調べるため、目的の語とその文脈を検索するＫＷＩＣ（Keyword in Context）はかなり前から用いられています。ただし、日本語の動詞のように活用する語は、すべての活用形をキーワードにしなければなりません。それを解決するためには、一つは、データ・ベースそのものに文法的な情報を付加情報としてあらかじめ付けるという方法がありますが、いろいろ問題があります。

ここでは、日本語を研究したり、日本語を教えたりする際、コンピュータのプログラムなどができなくても手軽に、研究や教育にコンピュータを利用する非常に簡単な方法を述べることにします。

データが小説とか随筆、さらに論文など実際に使われた表現であれば、研究者の言語直観とは関わりなく、その表現が立派に使われたということであり、問題なく言語データとして使うことができます。また、日本語を教える教師の母語が日本語でないなら、ある表現が適切かどうかの判定は、自分が聞いたこともない、読んだこともないものであれば、難しいでしょう。そんなとき、データ・ベースを検索し、その表現を使った例があるかどうかを調べることができるわけです。

そこで、日本語の文を集めたデータ・ベースが使える限りそれをコンピュータの文書処理ソフト（ワープロ）に呼び出し、「検索」という機能を使って、キーワードで探せば目指す語や文法の表現が出るようになっています。

では、日本語のデータ・ベースはどのようにしたら手に入るのでしょうか。例えば、小説や１年を単位とした新聞記事などがCD-ROMで市販されていたり、いろいろな日本語のデータが出ていますから探してみてください。

それから、もう一つ、海外で日本語教育に携わっている方（日本語が母語でも母語でなくとも）へ是非お勧めしたいのは、いま流行りのInternetです。これは、パソコン、Internet用のソフト、電話回線を利用してコンピュータ通信を行うモデムがあれば、Internetに繋げる業者に加入することで、いろいろな情報が非常に安く手に入ります。その日の日本の新聞を海外で読むことさえ、なんでもないことになりました。そして、これを読み出すとともに、自分のデータに保存することで、自分の日本語データ・ベースを作ることができるのです（注１）。いろいろ違ったコンピュータで使えるWindows95を使うのが一番いいと思いますが、必ず、日本語に対応したもの（注２）を使ってください。

（注1）
最後に簡単なデータ・ベース作成と検索の手順を記しておきましょう。（アンダーラインはソフトのコマンド<命令>を指す。）
1．Windowsの中でInternet用のソフトを起動する。
2．新聞社や会社、大学の研究室のホームページを探す。
3．画面の文章の範囲を指定して、編集、コピーで文章をクリップボードへコピーする。
4．ワープロのソフトを起動する。
5．編集、貼り付けでクリップボードから前記の文章をコピーする。これを繰り返すことで自分のデータ・ベースが作れる。
6．データ・ベースに対して、ワープロの検索機能を使って目当ての表現が入った文を探す。
なお、こうして入手したデータを利用するにあたっては、著作権上の問題が生じないように、十分留意する必要があります。

（注２）日本語対応のソフトを使うとは
日本語に対応していないコンピュータ・ソフトでは日本語を入力したり表示することができません。本文で述べた日本語研究のためのコンピュータ利用には最低でもWindows95のようなシステムソフトの他にワープロ・ソフト（これはWindows95に付属しているソフトでも代用できないことはありませんが普通のワープロ・ソフトの方がはるかに使いやすいと思います）、それにInternet用のソフト（Netscape NavigatorやMS Internet Explorer）が必要ですが、すべて日本語に対応していなければなりません。日本語で発売されているもの（それぞれのパソコンの機種に対応していることはもちろんです）が一番いいのですが、日本の国外でも発売会社か販売の業者にきけば手に入ると思います。

参考文献

草薙裕（1983）『コンピュータ言語学入門』大修館書店
田中章夫他（1983）『朝倉日本語新講座６ 運用Ⅱ・人文系研究のための言語データ処理入門』朝倉書店
『月刊言語』（1996.9）Vol.25、No.9（特集 パソコンの言語学）大修館書店
『日本語学』（1995.7）Vol.14、No.8（パソコンを使う日本語研究）明治書院