

国立国語研究所学術情報リポジトリ

特集・電子化資料による日本語研究

メタデータ	言語: jpn 出版者: 公開日: 2019-03-25 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	https://repository.ninjal.ac.jp/records/2068

特集・電子化資料による日本語研究

パソコンを使って日本語研究をする場合、最低限必要なのは、「パソコン」と「プログラム（ソフトウェア）」と「電子化資料」である。もちろん、これらすべてに関する方法全般について本誌で扱う余裕はない。そこで、この3項目のうち、大事なことでありながら、これまで見過ごされがちであった「電子化資料」という点に焦点を絞って特集を組むことにした。

電子化資料を用意する方法としては、第三者が作成したものを入手する方法と、研究者自身が作成する方法とがある。

横山氏ほか3名の「新聞漢字調査の現状と将来」は、朝日新聞社が作成した朝日新聞の記事のCD「CD-HIASK '93」を利用したものである。ところが、文字の計量的研究に利用しようと、いろいろ調査してみると、CDで供給された電子化資料の記事内容と、それがもとになった原記事の内容とが文字単位で一致していないところがあることが判明し、それを指摘したものである。紙を媒体とした文献学では「本文批判（テキスト・クリティーク）」はもっとも基礎的な作業として行われるのが常識である。電子化資料の場合も文学作品では本文批判が厳密に行われている場合が多いが、新聞などの電子化資料を利用した研究ではその点が看過されている場合が少なくない。そのような現状に警鐘を鳴らすという点でも本論文の指摘は重要である。

小磯氏ほか6名の「『日本語話し言葉コーパス』における書き起こしの方法とその基準について」は、自分たちで作成した電子化資料を利用するものである。これまでの電子化資料は書きことばの文字データが中心であったが、この研究グループは話し言葉の音声データを中心とする点で新しさがある。ただ、音声データだけでは検索などで問題が出てくるので、その音声データを文字データに書き起こす必要が出てくる。この論文はその書き起こし基準について紹介したものである。今後、文字データだけではなく、音声データも電子化資料として活用されていくことになるのは時代の趨勢であるが、その早い段階の書き起こし基準として、今後の電子化資料の作成に大きな影響を及ぼしていくことになるだろう。

田野村氏の「サ変動詞の活用のゆれについて」は、横山氏のグループと同じ朝日新聞の電子化資料6年分を利用して、実際にサ変動詞の活用のゆれについての調査を行い、分析と考察とを加えたものである。電子化資料による日本語研究の模範例として、益するところ大である。

電子化資料の蓄積という点では、イギリスとアメリカの動きが早く、すでに1960年代からアーカイブが開設されたり、コーパスが作成されたりしている。日本でも国立国語研究所が中心となって、1960年代には新聞の電子化資料が作成されたりしたが、コンピュータでは漢字はまだ使うことができなかった。この漢字というハンデがあったため、本格的に日本語のコーパスが作成されるようになったのは、1980年代後半からといってよい。しかし、現在は膨大な数の電子化資料が利用できる環境にある。

ちなみに近年、パソコンを使った日本語研究が当たり前のものとなってきているが、このよう

な状況になったのは、ほんの数年前からのことである。これまで日本語研究とパソコンがどのような関係にあったのか、そして、現状がどのように位置付けられるのかがよく分からないという人も少なくないと思う。そこで、日本語研究の道具という観点から、パソコンがどのような発展と普及とを遂げてきたかを振り返ることにしたい。

日本で本格的なパソコンが登場したのは1979年である。それまでは、大型計算機や中型計算機をもつ大学や研究所に所属している、ごく一部の研究者だけが日本語研究に計算機を活用できたのである。といっても、全角の漢字が使えたわけではなく、半角のアルファベットや片仮名による「電子化資料」を自分たちで作って分析していた。計算機自体の使い勝手もいまのパソコンとは比べものにならないほど悪かったので、研究の道具として気軽に使えるものではなかった。

パソコンが市販されたからといって、すぐに普及したわけではない。何よりも値段が高かった。パソコンシステムをそろえると、100万円は軽く越えた。たとえ、科学研究費で購入できたとしても、まだすぐには使えなかった。なにしろ市販されていたソフトはゲームしかなく、ワープロも表計算ソフトもデータベースも何もなかったのである。そのため、パソコンを仕事で使いたい場合は、自分でプログラミングをする以外に方法がなかった。このため、大金を出して購入したものの、結局は使えないで、ホコリをかぶってしまったパソコンのなんと多かったことか。

パソコンがJIS第2水準の漢字まで、本格的に使えるようになったのは、1983年に16ビットパソコンが登場して以後のことである。応用ソフトもワープロ、表計算ソフト、データベースとそろえるようになり、日本語研究のためにパソコンを利用するための環境は整ったといえる。しかし、その後、10年は普及しなかった。これはまだパソコンシステムが高価であったことと、OSがMS-DOSであったため、キーボードに拒否反応を示す一般の人にはまだ使い勝手が悪かったのである。パソコンを日本語研究に活用できた人は「パソコンに強い人」だけだったわけである。

この流れを変えたのが、Windows 95の登場と、日本でもIBM互換機が標準になったことである。この二つのことは、パソコンの操作性を向上させ、また、価格を引下げることとなった。結局、現在のようにパソコンが普及しはじめたのは、1996年以降と見てよい。

さて、以上のように日本におけるパソコンの発展・普及の流れを見てみると、現在は、ようやく誰でもパソコンを日本語研究に使える時代が到来して、まだ間がない時点にあることが分かる。

しかし、大学の授業で言語研究のためにどのようにパソコンを使えばよいか、あるいは、どのような点に注意しなければならないかを教えているところは極めて少ない。このような現状では、パソコンを日本語研究に使いたい気持はあるものの、その方法が分からないという研究者や学生が相当に多いものと推測される。

そのような方々のために、本特集が良質の影響を与え続けることを願うものである。

編集委員長 伊藤雅光