

〈資料・情報〉

読みたい人に読まれるために

古田 啓

キーワード：『国語年鑑』 キーワード データベース

一、情報が多すぎる

現代は情報が多すぎるとはよく言われる。情報の絶対量が多すぎて処理しきれない。処理しきれないので、必要な情報が必要とする人の所に届かない場合も少なくない。これは、^(注1)国語学の世界においても言えることだと思われる。

ではどれほどの量であるのか。『国語年鑑』に載っている論文数で考えてみる。この『国語年鑑』『雑誌論文一覽』の中で、国語(学)、国語史、音声・音韻、文字・表記、語彙・用語、文法、文章・文体の分野を、昭和二十九年版から昭和五十八年版までの三十年分集めてみると、約二万二千本になる。年々増加の傾向にあるから、右の分野の雑誌論文の総計は、現在、一年で一千本を越しているはずである。

もちろん全部の分野を読むとは限らない。しかし、分野別に三十年分の合計を見ても、文法や文体の分野では約三千本、語彙の分野に至っては約五千本ある。しかもこれは雑誌論文の数であり、この外に講座類や論文集に収録されている論文もある。したがって分野別に見ても、二日に一本は読んでいないと、毎年『国語年鑑』に登録されている論文の全部に通すことは出来ない計算になる。そして卒業論文を書くような人にとつては、これらを何十年分もまと

めて読まなければならぬ。したがって、もし明治以降の論文を全部読んでいたら、自分の論文を書いている暇がなくなる結果になる。そこで、論文の検索と絞り込みがどうしても必要となる。

論文の検索については、筆者の場合、本格的に検索する時には次のような方法を採用している。まず、『国語年鑑』を端からめくり、その分野の中で自分のテーマに関係しそうな論文を探す。『国語年鑑』は、年ごとに分類が異なっていたり、複数の分野にまたがる論文はどの分野に入れられているか分からなかったりするから、注意しなければならぬ。同時に、著者名索引でそのテーマについて書きそうな人の論文を探す。『国語年鑑』は、昭和二十九年版以降、三十冊以上でているから、右の作業を三十回以上繰り返して行なう。そして得られた論文を読んで、芋づる式に他の論文を探す。

絞り込みについては、手近にない雑誌の場合、他の論文に引用されているかどうかをはじめとして、著者や掲載誌によって読まずに済ませるかどうかを判断する。

ただし、これは本格的に検索する場合であり、かなりの労力を費やすので、必ずしも主テーマにかかわらない事柄については、もっと略式の方法を採用することも多い。

右の方法は多くの人が採っている方法であると思われるが、このような状況では、必要とする人にその論文が確実に読まれているという保証はない。むしろ、不幸にして、その論文の存在を知らないまままで終わることも、現実には多いであろう。

なんとか論文の検索が安全かつ容易に出来ないものか。筆者が卒業論文を書く時はそう思っていた。まさか自分がそのための索引作成にかかわることになるとは、夢にも思わなかったが。

以下は『国語学研究文献総索引』作成にかかわっている間に、個人的に思った事柄である。

二、 しかるべき所へ送ろう

国立国語研究所編の『国語年鑑』には、確かに落ちている論文もあれば誤植もある。しかし、その具体的な編纂過程を聞いた者としては、よくもあれだけ少ない人員で、あれだけの年鑑が出来るものだと思うざるを得ない。現在のところ、『国語年鑑』は論文検索において最も頼りになる。さらに頼れるものにするためには、各著者の協力が必要となる。『国語年鑑』で論文題目の上に*印がついているものは、その雑誌が入手できなかったという意味であるから、*印のついている紀要などは、なるべく国立国語研究所に寄贈すべきである。

また、各種の展望は、その数多い論文の中でとりあえず読むべきものを探す指針となる。ところが、展望の執筆者がその論文を集めるのは一苦勞である。かつて筆者は、さる執筆者が三百本の論文のコピーをとるのを見たことがある。展望の執筆者が良心的な場合、謙譲の美德は負担をかけることになるから、やはり紀要に載つたものなどは送るべきである。このことについては、『国語学』百四十二集の百三十八頁を参照せられたい。

三、 分かりやすい題目・引きやすいキーワード

現在でも、一体何について書いてあるのか分からない題目を見ることがある。このような論文は正しい位置に分類されないことが多いし、題目を見ても自分が読みたいことについて書いてあるのかど

うか分からない。著者名から判断することになるが、学部¹の学生にその判断を求めるのも酷だ。論文の題目は何について書いてあるのか分かりやすいものであつてほしい。

さらに、「上代(特殊)仮名遣い」「陳述」などのキーワードで、そのことについて書いてある論文を検索できれば大変便利である。そのためには各論文にキーワードがついていなければならない。早くから雑誌『計量国語学』では、投稿論文にキーワードを3〜10個つけるよう規定してあり、各著者はこの小文の冒頭につけたような形で、各論文につけている(もちろん、冒頭の例は、例として示すためだけにつけたものであり、実際は、このような片々たるエッセイにいちいちキーワードがついていたら、検索する例が迷惑である)。

厳密に言うと、『計量国語学』ではこのキーワードを「ディスクリプタ」という。「ディスクリプタ」とは、同じことを意味するキーワード(の語形)が何とおりもある場合、それらを標準化したキーワードのことである。『計量国語学』には、このディスクリプタによる八十号までの索引があり、検索に大変役立つている。

最終的には、キーワードを用いたオンライン検索ができればもっと便利であろう。例えば「ハ」と「陳述」が同時にキーワードとしてついている論文を探せと命令すると、その条件をみたま論文が一覧表となつてでてくるシステムである。このようなシステムはオンライン・データベース・システムと呼ば^{は?}れる。

しかし、このようなデータベースを作成するためには、いままでもキーワードがついていなかった論文に、標準化されたキーワードをつけなければならないが、これは実に大事業である。

四、キーワードの自動切り出しについて

その大作業を少しでも楽にできるようにと、筆者は、題目からキーワードの候補となるものを切り出すプログラムを作成した。このプログラムは、漢字・カタカナなどの文字の種類によって候補となる文字列(単語とはとも言えない)を切り出すものである。現在は、漢字・カタカナ文字列と、「 」で囲まれたものをキーワードとして採っている。「国語字用語シソーラス」とでもいったものがあるれば、もっと気のきいた方法もあるが、現時点ではこのようなものしか出来ない。

この方法では、題目にある文字列でさえ、キーワードとして採るべきものを落とす可能性もあれば、キーワードとして採るべきでないものを採る可能性もある。例えば、「使役の助動詞す・さす・しむ——源氏物語に見られるものを中心として——」という題目であれば、「使役／助動詞／源氏物語／見／中心」と切り出す。「す・さす・しむ」が落ちる一方、「見／中心」が入っている。

また、この方法では「ドチリナ・キリシタン」と『吉利支母教義』が、全く別物として扱われ、相互参照の手立てもない。

そして、本来、題目の字面にあらわれたものしか採れない。題目にあらわれてこないものを推定して採る能力はない。ただし、できるいいシソーラスがあれば、採ったものから推定することは可能である。例えば、『天草本平家物語』に「時代／室町時代」という情報がついていれば、『天草平家物語』における「ごさる」「おぢやる」「おりやる」という論文に、「室町時代」というキーワードをつけることは出来る。しかし、右にあげた二例は分かりやすい題目のもの

であるが、全部ひらがなで書いてあるような題目や、極めて文学的な題目からは、シソーラスがあっても何も採れない。

結論としては、無いよりはましであり、このプログラムの出力を台紙にして加除を行えば、最初から書くよりは手間が省けるかも知れない、その程度のものである。結局、人間がシソーラスを作り、人間が個々の論文を見なければならぬ。機械は、当然のことであるが、人間の行なう仕事の手助けしかできないということである。

ただし、人間が行なうという、その苦勞はいかほどのものであろうか。論文の題目・著者名・発行年月日などを原本校正するの、一人が一日に百本が限度である(このことについては中野洋氏と筆者による報告を参照せられたい)。中身を読むとなつたら、どれほどの手間が必要であるのか、考えてみるだけで慄然とする。枚数などによる、ある程度の取捨選択はやむをえないかも知れない。

五、コンピュータの手助けと人間の苦勞

確かにコンピュータは大いに助けになることがある。例えば、原本校正は、『国語年鑑』の分野別発表年代順に並んでいた二万二千本の論文を、雑誌名順に並べ直したファイルを対象に行なっている。並べ直すのが約二分間、結果の印刷が約一時間で済んだ。二万二千本を手作業で並べ直すとしたら、いうまでもなく大変なものである。ましてや、雑誌名順に並んでいない台紙を原本で校正することなどは、到底不可能である。

いついかなることにしても手作業が大切であるという意見もある。しかしコンピュータを使う方が早く正確に出来ることまで、人間が手作業で行なわなければならないとすれば、それこそ非人間的

である。

とはいえ一方、コンピュータがなかったら、原本校正自体、行なわなかったかも知れない。いやそもそもこの事業自体が沙汰やみになっていた可能性も大きい。こうして考えてみると、コンピュータは、ある程度人間の負担を軽くするが、また別の負担を押しつけているのかも知れない。

コンピュータは人間の手助け以上は出来ないので、結局は人間の苦勞が必要である。しかし、特定の数人にばかり、その負担の行くことがないように、これからは注意すべきであろう。

六、読みたい人に読まれるために

話が横道にそれたが、結局、読みたい人に読まれるためには、次のような工夫が考えられよう。

- ・ 国立国語研究所や、展望の執筆者にも、抜刷などを送る。
- ・ 分かりやすい題目をつける。
- ・ 雑誌の方針にもよるが、キーワードをつける。

既に、これまでの論文にキーワードをつける計画が始まっているが、コンピュータで出来ることはおのずと限界がある。各分野の専門家に依頼する計画であるが、その場合、一部にあまり負担をおかけしないような配慮が必要であると思われる。

注1 生化学の分野でどのような状況であるかについては、アイザック・アシモフ「たった一兆」(山高昭訳、昭和六十年九月、早川書房、ハヤカワ文庫NF27)の中の「10 息をはずませて」が入手しやすく読みやすい。

注2 オンライン検索については、山崎昶「知的検索の技術——新しいもの探しの科学——」(昭和五十七年七月、講談社、ブルーバックス

B-505)が丁寧に解説してある。

——茨城大学専任講師——

(昭和六十一年一月五日 受理)