

# 数理的研究

草 薙 裕

## 一、はじめに

「言語の数理研究」ということを、言語データの数量化による研究と、意識するかしないかは別として言語（正確には文字列）を数値化する（すなわちコンピュータの処理をする）研究の両面から見ることにする。

昭和59年と60年は、言語のコンピュータ処理では画期的な年であった。前の年までは、自然言語処理に関する著書もあまりなく、研究活動もごく限られた範囲の研究者の間で行われていた。ところが、この期間には、いわゆる人工知能や第五世代コンピュータの研究の煽りを受け、自然言語（言語学という言語）のコンピュータ処理に関する著書や翻訳書、しかも、一般向けのものが、かなり出た。

前回の展覧で、荻野綱男氏は、その期の特徴として、科研費特定研究「言語の標準化」のスタート、新世代コンピュータ技術開発機構（ICOT）の活動開始、およびワード・プロセッサの急激な普及をあげている。それに対し、今期は、言語学者と工学者の研究交流、自然言語処理や人工知能に対する関心、および、ワード・プロセッサのみならずパーソナル・コンピュータの急激な普及があげられ

よう。

言語の研究に何らかの形でコンピュータを関与させたり、数量的な研究をすることを考えると、三種類に分けられるだろう。

まず第一は、今までと同じような言語の研究にコンピュータを用いることである。すなわち、言語の研究を内省的に行い、その結果をコンピュータで検証しようというものである。この場合、コンピュータは単なる道具になる。近年のコンピュータの高性能化、小型化、さらに、低価格化は言語学者の間へのパーソナル・コンピュータの滲透を促し、二、三年前には予想もつかなかったような人が、現在、コンピュータやワード・プロセッサと親しんでいる姿を見る。かなり年配の研究者がコンピュータの前に坐つて、キー・ボードを叩いているのを見ると、言語の研究も質的変化が起こるかも知れないという錯覚に囚われがちだ。ところが、この方法は従来の言語の研究方法を根本的に変えるものではない。規則などの検証の時間は縮まろうが、内省的な研究であることには変わりない。

大量のデータをコンピュータで統計的に処理したり、コンピュータを言語に用いる、いわゆるCAIも、真の意味の自然言語処理ではない。

コンピュータを用いる言語の研究の意義が発揮されるのは、コンピュータ上に言語のモデルを構築することにより言語の研究を行うとするものである。これは前述のコンピュータを道具とする研究とは異なり、研究の枠組み自体がコンピュータ上で築かれるということが目的である。

この種の言語研究は、多量のデータに対し、適切な規則化が必要である。データが多く、したがって、その処理は当然コンピュータにさせなければならぬ。そのためには、その多量のデータが処理できる規則の構築になる。これは従来の言語研究とは異なつて、データの処理が内省的なものばかりでなく、多量のデータによる検証あるいは反駁が、人間の行う研究というよりコンピュータによる客観的なものであるということではなければならない。今までは、予算的にも、人間の数でも、この辺りは言語学の分野には苦手であつたものが前述のようにパーソナル・コンピュータの普及で、かなり研究の見通しが立つて来た。

さらに、このような言語研究とともに、言語の数量化も、はじめからアンケートなどで数値を得て、それを統計処理をするのではなく、数量には全く関係のないように見える言語データに対し、何らかの理論や仮説を立て、それを数量化する研究がある。これも、言語データに対する直観的な評価に対し、何らかの客観的基準による分析が可能になる。この面でも徐々に研究が成果を上げていく。

最後の種類に、機械翻訳や、コンピュータによる言語理解に関わる自然言語処理があげられよう。これはコンピュータに文を入れ、その翻訳文を出させるようなシステムや、コンピュータに命令するのに現在のようなコンピュータ言語を使うのではなく、日本語や英

語のような日常言語でコンピュータとやりとりが出来るようになるものである。この分野は、コンピュータのハードウェアやソフトウェアにかかわる部分も大きく、今まで工学者が主にかかわつて来た。この分野は、いわゆる第五世代コンピュータへ向けての研究や、人工知能の研究が最近大幅に伸びてきた。言語学の知識が少ない工学者だけでは効率のいい言語の処理は望めない。そこで言語学者の応援なり、共同研究が望まれる。まず、これから見よう。

## 二、言語学と工学の協力

従来、機械翻訳や自然言語処理の研究を行なっている工学者達は、元来の言語研究の成果は即、自然言語処理には使にくい。したがつて、工学と言語学の学際研究が望まれるといつてきた。

今期は、前期における工学者と言語学者の協力の成果が公刊されたり、さらに密な協力が行われたことが特筆できよう。

今期の初頭に出された、長尾真編「講座現代の言語7、言語の機械処理」(三省堂)がその典型であろう。これは昭和52年度から三年間継続した文部省科学研究費特定研究「言語生活を充実発展させるための教育に関する基礎的研究」のうち言語の機械処理に関するものの成果を、分かりやすく解説したものである。発行がやや遅いくらいがあるが、言語の解析(分析)、機械翻訳、言語表現と論理表現、言語理解談話、言語習得など、ほとんどの分野をカバーしている。言語学的研究には、山梨正明・寺津典子・稲田俊明「意味と文脈」がある。さらに、工学的な自然言語処理には興味がない言語学者にも辻井潤一の「言語処理のソフトウェアとデータ」は大いに参考になる。KWIN(文字データ検索システム)やソーティング(データの並

へ換えシステムなどの解説に加えて日本で現存するプログラムやデータに関する情報も出ている。

このような言語学者と工学者の共同研究の場合は昭和57年度から三年間続いた同じく科研費特定研究「言語の標準化」があり、さらに今期の範囲外だが、昭和61年度から、やはり特定研究「言語情報処理の高度化のための基礎的研究」とどんどん広がって行く。

田中幸吉編『知識工学』（朝倉書店）も昭和54年度より三年間継続した科研費総合研究「知識工学の基礎とその応用に関する研究」の成果をやはり解説的にまとめたものであり、工学者主体の自然言語・知識処理に関する知見が豊富だが、石綿敏雄「日本語動詞格支配の類型」では格文法、結合価文法、語彙文法を統合し、日本語の動詞の文型の記述を行っている。また吉田将「言語情報による様態概念を含む知識の表現」は日本語の科学技術論文から約二百六十の様態表現を抽出し、分類をこころみたものだが、内省的研究ではなく、数多くのデータを処理する工学的手法が、我々言語学の研究に示唆することも多い。前の二書と同類のものとしても一冊、長尾真監修『日本語情報処理』（電子通信学会）をあげておこう。

さらに、この時期は科学技術調整費で行われた機械翻訳システムの開発である。「日英科学技術文献速報システムに関する研究」が日本科学技術情報センター、電子技術総合研究所、京都大学を中心に着実に行われ、その報告書も毎年出されている。この研究の研究推進委員会も工学者と言語学者の「対話」の場になっている。「対話」といえば、日本学術振興会でも「言語の知能的処理に関する研究開発専門委員会」が昭和59年度より設けられ、活発な研究報告や討論が交わされている。

— 同様な委員会の報告としては、「知能科学技術に関する調査研究報告」（新技術振興渡辺記念会）もある。

### 三、数量化

次に、言語の数量化を見よう。言語の数量化に関する研究では水谷静夫氏が相変わらず目を引く。「戦時下流行歌の用語」（『計量国語学』14巻5号）は流行歌の用語を数量化理論を用いて出現情況の類似度という観点から考察したもののだが、単なる数量化および統計的処理にとどまらず、するどい考察が光る。言語の数量化はいろいろ可能性があるが、やはり言語学の枠組の中で行われるものは、その言語学的意義を十分認識した上で、結果に対する説明も十分になされなければならない。そこに言語学者が行う研究の意味がある。水谷氏の研究は言語の使用を説明するという意味からも範を垂れているといつていいだろう。同じ流行歌の語彙の統計を求めた調査に中野洋「流行歌（昭和58年）の語彙」（『計量国語学』14巻4号）がある。さらに、芳賀純「大学生による要約文と漢字使用率」（『計量国語学』14巻7号）も言語使用に関する研究である。

フィールド・ワークをすると大量のデータが集まる。そこでその処理を求めて数量化への検討が進む。そういう意味で方言研究などで数量化が発展するのも当然だろう。荻野綱男「敬語使い分けの段階数——札幌における敬語調査から⑥——」などそのいい例であろう。

さらに、柴田武氏の方言研究における「ネットワーク法」はその最たるもので、方言区画を立てるための、「東条区画論」や「等語線方式」に対して、すべての地点間で相互にいくつの言語的特徴を共有するかを数えて共有度のマトリックスを得、このマトリックスに

対して数量的処理を施し、地域的対立を浮かび上らせるという方法であり、コンピュータを用いた処理によって、複雑なデータの要素まで取り込んで、既存の理論へ挑戦した研究の好例であろう。この「ネットワーク法」は柴田武・熊谷康雄「言語的特徴による地域分割のための「ネットワーク法」——特にNTT<sup>(注)</sup>について——」(国語学)一四〇号に詳しい。このようにコンピュータを利用することにより今までの方法の修正あるいは挑戦という研究の新しい方法は大いに評価の対象になる。

#### 四、ワープロとCAI

コンピュータに特に関わりを持たない言語学者もかなりワード・プロセッサを操作しはじめているようだ。ワード・プロセッサの中味(ソフト・ウェア)がどうなっているか知らなくとも、いろいろの操作をしているうちにソフト・ウェア上の言語の規則に問題があることに気が付く。機械翻訳や言語理解と同様、ワード・プロセッサの仮名あるいはローマ字漢字変換の形式化の開発に実際に言語者が関わっていることはほとんどない。

ここでも言語学の知見が利用されればシステムの効率化が図れることが予想される。雑誌「日本語学」(明治書院)は昭和59年の7月号でワードプロセッサの特集を組んでいる。その中で石綿敏雄、かな漢字変換の原理と応用——構文解析の観点から——は、漢字変換における言語学、国語学の問題点を論議している。また吉田将「機械辞書について」はワード・プロセッサの辞書の問題を論じている。

計量国語学会でも昭和60年9月の大会で、「ワードプロセッサに望む」と題してパネル・ディスカッションを組み、言語学者のワ

ード・プロセッサに対する言い分を聞くセッションを特集し、辞書や文法に対する注文が出された。

コンピュータは言語教育とも結び付く。これも本格的な自然言語処理の研究を行なう研究者とまでは行かなくともコンピュータを用いて言語を教えようという発想は、はるか二十数年近く溯る。パーソナル・コンピュータの普及、教師の不足(予算的に)、学生の質の多様化などの要素がコンピュータを使った語学教育が出来ないものかという研究を推進する。日本語教育学会の機関誌「日本語教育」が昭和59年10月の54号で「コンピュータと日本語教育」の特集を組んでいる。これには、構島忠夫「コンピュータと言語教育」、村上温夫・佐藤純子「日本語教育CAIの将来性に対する私見と入門期の文字教育の一例」、草薙裕「日本語教育とコンピュータ」、深谷哲「コンピュータを使った語学教育」、清田潤「コンピュータを使った学習過程」の設計」、神田紀子・深尾百合子・加藤容子・大坪一夫「名古屋大学日本語教育におけるマイクロコンピュータ利用の実態と可能性」、原土洋「日本語ワードプロセッサの日本語教育への利用——書くことの反省と訓練——」、山本直三「日本語ワードプロセッサと言語教育」など、日本語教育へのコンピュータ利用の構想や実施報告が盛りこまれている。

さらに、白井英俊・荒井卓也・内藤佳有「日本語教育用CAIシステムの試作——文字と語彙の学習——」(情報処理学会資料42「自然言語処理研究会」もマイクロコンピュータ上で外国人に日本語の漢字を学習させるためのCAIシステムの開発を扱ったものである。

## 五、自然言語処理

これだけコンピュータが普及し、人工知能や機械翻訳が盛んになってくると、それを支える、言語学者による自然言語処理の研究がますます必要になってくる。これはコンピュータ処理の特徴をふまえた上で、自動的な処理をするには言語のどこをどのように形式化すべきかという発想で初まる。したがって、普通の言語学と同じ、言語の構造や使用の解明という目的を持ちながら、焦点の置き方が異なるといえよう。

「朝倉日本語新講座」(朝倉書店)は全六巻で、計量国語学会で活躍するコンピュータ寄りの言語研究者が集まって、言語の研究に関する論文をまとめたものだが、今期には、その第4巻、「文法と意味II」が出た。この中では、草薙裕「文法形式が担う意味」、南不二男「質問文の構造」、中野洋「語義記述法の問題点」がコンピュータ処理を頭においた言語表現、意味の形式化の問題を扱っている。

二年に一度開かれるコンピュータ言語学国際会議は、コンピュータ言語学の分野で世界的に権威のあるものだが、第十回大会が昭和59年6月にアメリカのスタンフォード大学で開かれている。日本からの参加もだんだん増え、本大会の報告書には日本から参加の論文が十二件掲載されているが、技術的な論文が大半で、言語学的なもの、名詞の意味範疇を扱った Yoshiyuki SAKAMOTO, Masayuki SATOH, Tetsuya ISHIKAWA の「Lexicon Features for Japanese Syntactic Analysis in Mu-Project—JE”および移動動詞の意味的統語構造をあつかった Yutaka KUSANAGI の「Some Linguistic Aspects for Automatic Text Understanding”からい

あった。

この種の研究は情報処理学会の自然言語研究会資料にいくつか見られる。そのいくつかをあげると、岡本哲也・清水俊明「語い関数を用いた日本語の言い替え系」(同資料45)、村木新次郎・青山文哲・六条範俊・村田賢一「辞書における格情報の記述」(同46)、村木新次郎「述語素による動詞の分類」(同48)、田中卓史「物語理解のメカニズム」(同50)、辻井潤一・山梨正明「格とその認定基準」(同52) というように若手の学者の活躍が目立つ。

さらに、水谷静夫「国語表現での否定・量化に関わる考察、文章理解の為に」(「計量国語学」14巻6号)や宮島達夫「ドアをあけたが、あかなかった」(「計量国語学」14巻8号)は文の理解にかかわる問題を扱ったものである。

語彙の研究では、学術用語における造語の実態を明らかにする目的で、機械工学用語のパターンを研究したものに石井正彦・野村雅昭「機械工学用語の語彙構造」(「計量国語学」14巻4号)があった。さらに文字コードに関する提案として油谷幸利「2バイト系ハンダグコードの提案」(「計量国語学」15巻1号)、荒木卓也「漢字検索のための漢字コードの改善」(「計量国語学」15巻3号)が見られた。国語学や国文学に携わるものにとっては、コンピュータでいろいろの漢字が使えるかどうかは大いに気になるところだろう。この問題を扱ったのが田嶋一夫「漢字ソーラーズの構想と課題」(「日本語学」3巻3月号)である。

「計量国語学」では「プログラム欄」を設け研究に用いられたコンピュータ・プログラムを掲載している。今期に載ったものを列挙すると、水谷静夫「かたりづけ雑俳」(14巻7号)、清田潤「語彙表作

成プログラム」(14巻8号)、水谷静夫「引用句を拾う」(15巻2号)である。自然言語処理に関する研究論文で、その処理のプログラムが公開されているのはごく少ない。その意味では、この欄は大いに奨励される。プログラムの公開と云えば、自然言語処理を機械翻訳の例まで広げ、それをすべてBASICで組んでプログラムを載せたのが、草薙裕「パーソナル・コンピュータによる自然言語処理」(工学図書)である。

## 六、おわりに

今期には、もう一つ、自然言語処理で特記すべきことがあった。それは、LISPおよびSNOBOL4といった文字列処理のためのプログラミング言語がパーソナル・コンピュータで使えるようになったことだ。特にLISPはアメリカで、これまでに作られた人工知能、自然言語処理の90パーセントに用いられているが、今迄、本格的に用いるには大型コンピュータを使わなければならなかった。それが、パーソナル・コンピュータの性能がよくなってきたため、かなり強力なものが個人レベルで使えるようになった。言語の研究者の多くが、すぐLISPを使って自然言語処理の研究をするようにはならないだろうが、望めば本格的な自然言語処理のシステムを作ることが出来る環境が個人の手の届くところに来ている。

このような環境のもとで、これから、この数理的研究の分野では、どんどん若手の研究者が育ってくるだろうし、前記のように、科学研究費研究や委員会での工学者と言語学者との対話、共同研究が成り立って来ている。

この分野は近い将来、大いに注目され、また大きな進歩が期待、

要求されるのではないだろうか。そういう意味で、今期は一つのス  
 テッピング・ストーンになっているような気がする。

—筑波大学教授—