数理的研究

石綿敏雄

数理的研究の分野で扱われるものは、その範囲の規定が必ずしも容易でないのにもかかわらず、毎回の展望で大体一定した範囲のものが扱われている。それは前回の展望でも中野がいっているように、数理的な方法を用いたものを扱うからであろう。ただ、数字、数量でとらえたものはすべて数理的研究の範囲で扱うことになるかというと、それは不可能である。ためしに、一九八〇、八一両年の「国語学」誌に載った論文を調べてみると、少なくともどこかで計量的な方法を用いているものを含めるという扱いをするならば、その過半というよりも大半に近い論文が計量的な方法を用いている。ある意味で現代の国語学では、計量的な方法が普及し、珍しいものではなくなったといえよう。そこで、計量的な扱いをするものについては、一段進んだ方法を考えているもの、あるいは計量的な扱いそのものを目的としているものに限定してゆくことにする。

数理的な方法は統計的、計量的なものと、非計量的なものに大別される。そのような方法で扱われる対象は、音韻、文字、文法、語彙、意味、文体、文章（談話）などである。さらに目的・用途として言語学的な記述、情報処理への応用ということがある。以上のように、この項で扱うものは方法、対象、応用の三つの観点からみることができる。従来この「数理的研究」の項では応用を分類記述の重要な基準にしてきたが、今回は副次的なものとみなすことにした。機械翻訳、かな漢字変換など、それぞれまとまった応用分野ではあるが、そのなかでは共通する部分も少なくない。かつそれが言語研究のどのような分野と関連をもつのか、あるいはそれ自身言語研究そのものではないか、というみかたもまた必要であろう。ここではこの立場で扱う。応用分野ではいくつもの分野を同時にとりあげることが普通であるが、言語研究の分野でもたとえば文法と語彙など、いくつかの領域の関連を十分考えてみる立場がある。

情報処理の分野などがそのよい例であるが最近の言語研究では、従来の言語学、国語学などの系列以外の人が、数多く研究に従事しており、多くの論文を発表している。ここでは論文名のすべてを挙げることは不可能である。そのなかのいくつかを例としてとりあげてその傾向をみることにしたい。

統計的手法によるもの

まず、中野洋「『分類語彙表』の語数」（『計量国語学』、昭56・3）、野村雅昭「常用漢字の音訓」（同、昭56・6）、同「週刊誌の

漢字含有率」（同、昭55・6）、横山晶一「国語辞典データベースの基本データ」（同、昭56・12）のような各種データの量的な記述を目的としたものが注目される。この種のデータ処理にもコンピュータが使用され、最後のでは三省堂の明解国語辞典の見出し語の数、品詞の分布、語頭音別語数、語尾音別語数などを調べている。別にアクセント型などの統計も発表されているようであり、この種の知識が容易に得られるようになったのは喜ぶべきことであろうと思う。釘貫亨「上代オ列甲・乙母音対立崩壊の一要因」（『国語学』、昭56・12）で使用されたマルチネの機能負担量のようなみかたも、各時代の言語音全体についての計量的な記述を基礎としてそれぞれの音韻変化が説明されれば、より強力なものとなろう。

次に、社会言語学的観点から、言語使用の社会的広がりを調査したものがある。筑波大学バイリンガリズム研究会「日本におけるバイリンガリズム」（昭55・3、文部省科学研究費、以後科研費と略記）、柴田武・鈴木たか「生徒の成長に伴う親の名称の発達、『母』というようになるまで」（昭55・2、科研費）、野元菊雄・水谷修・沢木幹栄・石井久雄・志部昭平・真田信治・日向茂男・杉戸清樹・田中望「日本人の知識階層における話しことばの実態」（日本語教育のための言語能力の測定、科研費、代表者野元菊雄、昭55・3）などはその調査の規模からいっても計量的な手法を十分に使ったものとみることができる。しかしこの類のものとして最大なのは、国立国語研究所『大都市の言語生活　分析編』（昭56・3、三省堂）同資料編である。昭和四九年度から始めた、東京および大阪で総計一五〇〇名の市民を対象とした大調査である。従来の中小都市を対象とした地域社会の調査とは異なる領域を開拓したものとして、画期的なものといえよう。その成果も豊富であるが、たとえば「見られる」と「見れる」の分布など、きわめて興味深い結果が得られている。

この『大都市の言語生活』の調査では、そのなかの「言語行動の類型化」の部分で、数量化理論が使用されていることも、これまでの国語研究所の地域社会の調査になかったものであろう。そこには数学側の前進もあったわけで、社会言語学の研究が、言語学内部の理由だけでなく、それをサポートする数学側の前進や画期的な用具の使用に支えられて、やはり画期的な前進をすることが予想あるいは期待される。

それを立証するかのように、多変量解析の方法を方言分布に適用し、地理的、年齢的な分布パターンを総合し、そこから計量的な方言区画を試みるものが現われる。まだある地域のデータを扱ったものであるが、井上史雄「方言の地域差・年齢差のパターン」（『言語研究』、昭56・11）がそれである。この種の手法が確立されていけば、方言区画だけでなく、これまで主観的な扱いに終始していた国語学国語史などの他の類似の操作にも適用されるようになってゆくと考えられる。

柴田武監修、荻野綱男・藤田克彦・尾崎秩子・御園生保子「都市の敬語の社会言語学的研究」（社会関係・場面に応じた言語行動、研科費、代表者柴田武、昭55・2）、尾崎秩子・御園生保子・藤田克彦・荻野綱男「話者の属性からみた敬語の使い分け　札幌における敬語調査から1」（『国語学』、昭55・3）なども大きな社会調査である。この調査に関連して、荻野綱男「敬語における丁寧さの数量化」（『国語学』、昭55・3）において点数化における新しい計算法

を提案した。「荻野法」については水谷静夫「分割項目数量化の荻野法に寄せて」(『国語学』、昭56・3)がある。荻野綱男「敬調表現の長さと丁寧さ　札幌における敬語調査から3」(『計量国語学』、昭55・9)は、従来から予測されていた、「語形の長い敬語形は丁寧である」ことを実証したものである。

変数が複雑にからみあっている統計量は、社会言語学だけに存在するわけではなく、文化現象の多くに見られるものであろう。水谷静夫「流行歌の用語を数量的にみる」(『科学朝日』、昭55・10)では数量化理論Ⅲを、同「用語類似度による歌謡曲仕訳」(『計量国語学』、昭55・3)では、二言語作品の用語の類似度についてのものさしとして、宮島達夫のCに対し水谷静夫のDを提案し、数量化理論Ⅴを用いて吟味している。

中嶋久「和独辞書の訳語の重なりによる見出し語の分類」(『計量国語学』、昭55・9)では、語彙分類の指標として、各語の訳語に共通しているものがあることに注目し、その数を対称行列にまとめ、多変量解析によらず、最小モーメント法でならべかえる方法をとっている。同「非対称行列を見易くする法」(同、昭56・12)では対称・非対称の別にかかわらず適用できる方法を提案している。

安本美典「『シフト検定法』の検討1」(『計量国語学』、昭56・6)、同「同2」(同、昭56・9)はさきに提案されたシフト検定法を検討したもので、その批判についてこたえたり、類似の方法について論じたものである。

単語を文字の連続とみたばあい、そのなかのどの位置にどんな文字がくるかを調べるという必要が生じることがある。たとえば現代かなづかいでは「を」は格助詞「を」にしか使われない。小書きの「っ」は自立語の先頭には立たない。「者」は(シャと読むときは長い単位の)語の終に位置するなど。田中卓史 "Statistical analysis of Japanese character" (COLING 80)はこれを視覚的にわかりやすく見せてくれる。荒木啓介・金子明夫・高野文雄・日夏健一「日本語文論文タイトルからのキーワード自動抽出システム」(情報処理学会自然言語処理研究会資料、以下自然言語処理と略記、昭56・7)は日本語の、分かち書きのない普通の文表記から単語を自動的に切り出してキーワードを抽出するシステムであるが、単にこの種の単語と文字の関係の追求だけでなく、辞書の適切な使用も行っている。この種の研究にも機能負担量に当るような巧みなとらえ方を考えてゆくべきであろう。

道具としてのコンピュータ

統計的な解析であってもなくても、数理的な方法を用いるものは、多くのばあいコンピュータを使用している。そのようにコンピュータの使用法、プログラムを中心とした提案もいくつか存在する。大量のデータの処理の報告もここに入れよう。

『計量国語学』誌では「プログラム欄」でそのような報告を扱っている。その一つを例としてあげれば、水谷静夫「語の使用率分布関係の算定」(昭56・3)がある。これはかつて水谷が提案した使用率の分布関数の近似としての射影函数について、同じ仮説と同じデータでも曲線のあてはめ方によって結論に影響が出ることを考えたが、かつては電動計算器で計算に時間がかかったため、確かめることができなかったことが、今日コンピュータが使えるようになり、確かめることができるようになったと述べている。コンピュー

タは道具であるには相違ないが、しかし新しい道具の使用は新しい視界の獲得につながることを立証している。この意味で水谷静夫編『文科学生による言語データ処理　国語研究の道具としての計算機』（昭56・7、東京女子大学日本文学科）は有益な報告である。荻野綱男「コンピュータが描く言語地図」（『言語』、昭56・10）は言語地理学への応用である。これらの言語研究にコンピュータがどのように役立つかをまとめたのが、草薙裕「コンピュータによる言語研究」（『言語』、昭56・10）であり、日本語の文のなかで品詞がどのように並ぶかを使いやすい形で表にしたのが『新編日本語品詞列集成』原文篇（昭55・12）解説篇（昭56・3）である。

非統計的手法によるもの

この分野のなかには代数的なもの、計算手順的なもの、生成と解析などさまざまなものがあり、互いに重なりあってうまく分離できない。ここでは非統計的なものとしてまとめ、言語研究の分野によって基本的に分け、応用はそのなかで言及する。

文字については筆順のアルゴリズム的なとらえかたのものがある。石井久雄「カタカナの筆順」（『計量国語学』、昭56・9）、下村武 “Science of the stroke sequence of kanji” (COLING 80) など。

文法のうち形態論に関するものとしては、山中光一「日本文の表記と『切れ』『連なり』から見た活用の構造」（『計量国語学』、昭56・9）がある。活用を記述要素、機能要素、ヨミ上の連結要素の三成分からなるとするみかたである。これは清瀬義三郎則府「連結子音と連結母音と」（『国語学』86）と似た考えかたであるが、立場に相違がある。松沢利彦・西野博二「拡張 LINGOL 上での用言の形態素処理」（自然言語処理、昭56・7）は、用言と助動詞、助動詞相互の接続の処理をすることができるようにしたものである。

統語論とそれに対応する意味論の分野は、もっとも話題の多い分野である。これについてはまず、田中穂積・佐藤泰介・松本裕治・内田ユリ子・横山晶一・板橋秀一・三国一郎・池田尚志・大谷木重夫・寺津典子・山梨正明『自然言語処理技術と言語理論』（電子技術総合技術研究所報告、昭56・4）をあげるべきだろう。情報工学と言語学の研究者が共同で両分野の相関連する理論、技術について展望を試みている。B5版三二ページで、内容は言語学でいえばヨーロッパの言語学から生成文法、モンタギュー文法までを含むが、自然言語処理技術が言語学者からみて他の領域の研究ではなく、また単なる応用でもなく、実は言語研究そのものであることを感じさせる。処理技術のすべてが言語研究だというわけではないが、その主要部分は言語研究である。少なくとも新しい形の言語研究であろう（柴田武「総記」『国語学』一二一にも関連した記述がある）。同様のことは、田中穂積『計算機による自然言語の意味処理に関する研究』（電子技術総合研究所報告、昭54・7）についてもいえる。ここで扱われている一例をあげてみよう。

C1　45度ノオ湯
C2　45度ノウォツカ
C3　45度ノ棒
C4　壁ニ対シテ45度ノ棒
C5　温度ガ45度ノ棒
C6　45度ノ失敗

これらの表現では「45度ノ」が共通しているが、それぞれ内容が異なる。C1では湯の温度、C2ではアルコール濃度を指すのが普通であろう。これらの表現を「理解」するためには「湯」「ウォツカ」の項に「温度」「アルコール濃度」を書きこむような欄をもつ記憶の様式が用意してあることが必要である。人間がこの表現を理解する場合、ふろの湯かげんが四〇度くらい、ウォツカのアルコール濃度が約四五度くらいという事物に関する知識（筆者のことばでいえば百科事典的知識）があり、これがそれぞれ温度、アルコール濃度と判断する助けになると考えられる。C3は「角度」か「温度」か不明であるが、もしわかればC4、C5と同様に扱うことができる。「C6は、後続語「失敗」の意味的性質から、『45度』は回数をあらわしていることが分かる」という。このように文脈を検討することにより、数詞を含む表現の意味を理解し、抽出するというのが、この研究の一部になっている。

この種の取扱いに格文法がしばしば使われるのは理由のあることだと思われる。そこで格文法のわくぐみにしたがって問題点を整理してみる。まずS－M＋Pという派生ルールにしたがってMの部分を扱ったものをとりあげると、内藤昭二・島津明・野村浩郷「日本語文における法情報の解析」（自然言語処理、昭56・7）がある。法情報に含まれるのは、相、時、判断、態度、様態などである。これらMの内容を自然言語処理の立ち場から整理している。たとえば判断は、否定、可能、必要性、推量、意図、使役ほか、に大分類しそのなかを小分類し、語形との対応を示している。吉田将「言語情報による知識表現のあいまいさ」（知識工学の基礎とその応用に関する研究、以下知識工学と略称、代表者田中幸吉、昭55・3、56・3）は格文法による日本語の文の解析、〝様相〟の種類とその表現を扱う。草薙裕 "A model of natural language processing of time-related expressions." COLING 80. は時間関係表現の処理を考えたもので認識レベルにおけるフォーカスを導入している。フィルモアの格文法のPの部分を扱ったものに、岡田直之 Conceptual taxonomy of Japanese verbs for understanding natural language and picture patterns" COLING 80、村木新次郎「日本語の機能動詞表現をめぐって」（国語研究報告集2、昭55・3）がある。田中穂積・荻野孝野「形容詞もしくは形容動詞の修飾先の名詞を決める原則について」（『計量国語学』、昭55・6）は形容詞が含まれる複雑名詞句のなかでの形容詞のかかり先にアンビギュイティがある場合のきめかたを研究したものである。西田富士夫ほかの "English-Japanese translation through case-structure conversion" および長尾真・辻井潤一・平川秀樹・三田村敬・久米雅子 "A machine translation system from Japanese into English"（両者ともCOLING 80）は格文法を用いた文の解析を行っている。後者については、石綿敏雄「機械翻訳はどこまで進んだか」（『言語』、昭56、10）で言及した。絹川博之・松岡潤・木村睦子 "Japanese sentence analysis for automatic indexing" COLING 80 は格構造を用いた自動索引語づけ、日高達、吉田将 "A syntactic parser based on dependency grammar and its efficiency." COLING 80 も格構造を用いた構文解析プログラムである。仁田義雄『語彙論的統語論』（明治書院、昭55・3）は格文法の延長線上にある研究である。

文法としてモンテギュー文法を用いる傾向もこの分野では顕著である。池谷彰「モンテギュー文法入門」が『情報処理』に三回連載

（昭56・3～5）されたということが関心の深さを示すものであるが、淵一博のいうようにモンテギュー文法と新しいプログラム言語の間に類似性のあることもその一方の原因であろう。石本新「モンテギュー文法の代数的定式化Ⅲ」（『計量国語学』、昭55・9）、同「モンテギュー文法の代数的定式化Ⅳ」（同、昭55・12）はモンテギュー文法とレスニェウスキー存在論とを結びつけ、代数的定式化を提唱している。西田豊明・堂下修司 "Hierarchial meaning representation and analysis of natural language documents" COLING 80 は自然語文書の分析と意味の表現にモンテギュー文法を使用したものである。

語彙に関するものとしては、中野洋、土屋信一、靍岡昭夫 "An automatic processing of the natural language in the word count system" COLING 80、田中康仁「専門用語の解析と応用」（『計量国語学』、昭56・3）、同「漢字列長単位語の抽出」（同、56・6）、木村睦子「自動索引法概説」（計量計画研究所報告、'80）、村木新次郎「日独両語の語彙体系の対照比較」（『国立国語研究所論集』、昭56・3）、同「日本語とドイツ語の『基本語彙』をくらべる」（『計量国語学』、昭56・3）などあがる。

ディスコースに関するものとしては、白井英俊「文章理解と意味結合関係」（『計量国語学』、昭55・12）が、文章のなかの意味結合関係と理解のプロセスを扱って野心的である。照応については桃内佳雄「『その』に関する二、三の考察」（『計量国語学』、昭55・3）、寺津典子・稲田俊明・山梨正明「日本語における照応現象について（その二）」（計算機による日本語談話行動の総合モデル化、談話行動と略記、科研費、代表者石綿敏雄、昭55・1）がある。後者は、日本語の照応を代名詞で示す有形照応と、従来省略といわれてきた無形照応に分け、どんな条件で両者が生起するかを詳細に論じたものである。ディスコースについてはいわゆる SCRIPT 形式で談話理解のプロセスに接近した、溝口文雄「人間の長期記憶内の知識利用について」（談話行動、昭55・1）がある。石綿敏雄「文章解析のアルゴリズム化への試み（続編）」（談話行動、昭55・1）はそれまでの考えをまとめたものである。野村真木夫「現代日本語体験話法序説」（『国語国文研究』、昭56・7）はテキスト学の立場から日本語体験話法研究の立場を措定したもので、アルゴリズミックな展望をもっている。古郡廷治「地理の知識と道案内文の生成」（『計量国語学』、昭55・6）では、心理地図を方向のあるグラフとしてDとし、道路、場所、建物などの対象をAとし、二項間の関係をRとすれば、D＝∧A, R∨ これを言語で表現するとき、テンプレートとして動詞を考えると直進動詞として「…ヲ…ニムカッテ…カラ…マデ」の型があり、曲進動詞として「…ニアル…ノ…ヲ…ニ」がある。このような考えで道案内をする文をつくるプログラムをつくっている。一九八〇年東京で開かれた計算言語学国際会議（COLING 80）ではソ連のシャリアピナが日本語を扱って "Problems of formal representation of text structure from the point of view of automatic translation" を発表した。テキストの構造を語彙文法論的構造、意味、統語論的構造、アナフォリックな関係と線条化特性からみて論じている。

——茨城大学教授——